КРАТКО РЪКОВОДСТВО ЗА

РАБОТА С SPSS

Съдържание

Стр.

Глава 1. Въведение в SPSS	4
Глава 2. Как да поискаме помощ от SPSS	11
Глава 3. Опции	13
Глава 4. Файлове с данни	23
Глава 5. Редактор на данни	26
Глава 6. Подготовка на данните	38
Глава 7. Работа с множествени източници на данни	40
Глава 8. Преобразуване (трансформация) на данните	42
Глава 9. Работа с файлове	50
Глава 10. Работа с резултативния файл	66
Глава 11. Работа с таблици	81
Глава 12. Работа в команден режим	101
Глава 13. Анализ на честоти	106
Глава 14. Описание на метрични данни	110
Глава 15. Предварително изследване на данните	113
Глава 16. Таблици на спрегнатост	117
Глава 17. Обобщено описание на метрични данни	124
Глава 18. Погрупово описание на метрични данни	128
Глава 19. Т-тестове	132
Глава 20. Едномерен еднофакторен дисперсионен анализ	138
Глава 21. Обобщен едномерен линеен модел	144
Глава 22. Корелация	156
Глава 23. Линейна регресия	160
Глава 24. Непараметрични тестове	168
Глава 25. Анализ на надежността	185
Глава 26. Анализ на въпроси с повече от един отговор	189
Глава 27. Как да оформим резултатите	195
Глава 28. Графики	204

Това Ръководство е съкратен превод на оригиналния документ, който е на английски. Преводът е направен от студенти в НБУ и не е редактиран.

Глави 26, 27 и частично глава 28, не са преведени, а са дадени в оригинал.

Оригиналът на Ръководството за SPSS се намира на този сайт в папка E:\Applied Stat 2\ebox.nbu.bg\Base Guide\English.

Глава 1. Въведение в SPSS

SPSS за Windows е мощен инструмент за манипулации на данните и за статистически анализ в графична среда. Този инструмент е реализиран чрез описателни менюта и прости диалогови менюта, така че да бъдат удовлетворени една голяма част от практическите нужди за статистически анализ на данни, събирани по различни поводи и в различни области. SPSS е еднакво добър инструмент както в икономиката, така и в медицината, социологията, психологията и пр.

SPSS предлага два режима за работа:

- чрез менюта и
- чрез команден език.

1. Работа чрез менюта

Повечето задачи могат да бъдат решени чрез режима на менюта, които са лесно достъпни само чрез посочване и натискане на бутона на мишката. В допълнение на тези възможности за лесен достъп до статиситечския анализ SPSS предлага още:

Редактор на данни (Data Editor). Редакторът на данни е гъвкава таблично-подобна система за дефиниране, въвеждане, зедактиране и представяне на данне.

Редактор на резултати (Viewer). Този редактор облекчава разглеждането на резултатите, селективно показвайки или скривайки част от тях. Може да променя реда на показване на резултатите и да обменя таблици и графики между SPSS и други приложения.

Многомерни въртящи се таблици (Multidimensional Pivot Tables). Осигурява възможност вече създадена таблица да бъде преоформена чрез промяна на редовете, колоните или подредбата на слоевете. Позволява да се открият важни подробности, които иначе биха се загубили при една стандартна наредба. Дава възможност лесно да се сравняват групи визуално като разделя таблицата по начин всяка група да се показва отделно.

Графика с висока резолюция. Реализирани са всички най-често използвани графики за представяне на данните – стълбови и кръгови диаграми, диаграми на разсейването, 3-D графики, линейни графики и много други.

Достъп до бази данни. С помощта на Database Wizard могат да се четат данни от почти всички бази от данни.

Трансформации на данни. Тази възможност помага за по-бързата и коректната подготовка на данните за анализ. Предвидено е лесно да се извличат подмножества от данните, да се комбинират класове от данни, да се добавят, агрегират, разделят и транспонират файлове.

Online помощ. Подробни ръководства дават обширна информация за работата на SPSS. Има и контекстно насочена бързо достъпна помощ. В резултативния прозорец с посочването на даден термин и натискането на десен бутон на мишката се появява дефиниция на този термин. Има и т.нар. Statistics Coach (статистически инструктор), който може да бъде полезен при избора на метод за анализ. Допълнително има и база с примери, чрез които се обяснява как да се използват статистическите процедури и как да се интерпретират резултатите.

Команден език. Независимо, че по-голямата част от статистическите задачи могат да се решат чрез използването на менюта и прозорци, SPSS е снабден с мощен команден език, който позволява да се запазват и автоматизират задачи, които се налага да се решават периодично. Освен това, чрез този език могат да се решат и някои задачи, които иначе са недостъпни. Пълното описание на командния език е интегрирано в помощната система и отделно съпровожда пакета като PDF файл.

Прозорци

SPSS работи с няколко различни прозореца:

Прозорец „Данни" (Data Editor). Редакторът на данни показва съдържанието на файла с данни. Вие може да създаване нови файлове с данни или да променяте съдържанието на

съществуващи файлове. Ако сте отверили овече от един файл за всеки отделен файл се отваря собствен прозорец.

Прозорец „Резултати" (Viewer). Всички статистически резултати, таблици и графики се показват в този прозорец. Прозорецът „Резултати" се отваря автоматично при първото стартиране на процедура, която генерира резелтат.

Прозорец „Резултати в текстови вид" (Draft Viewer). В този прозорец резултатите **се** показват във вид на обикновен текст (освен интерактивните въртущи се таблици).

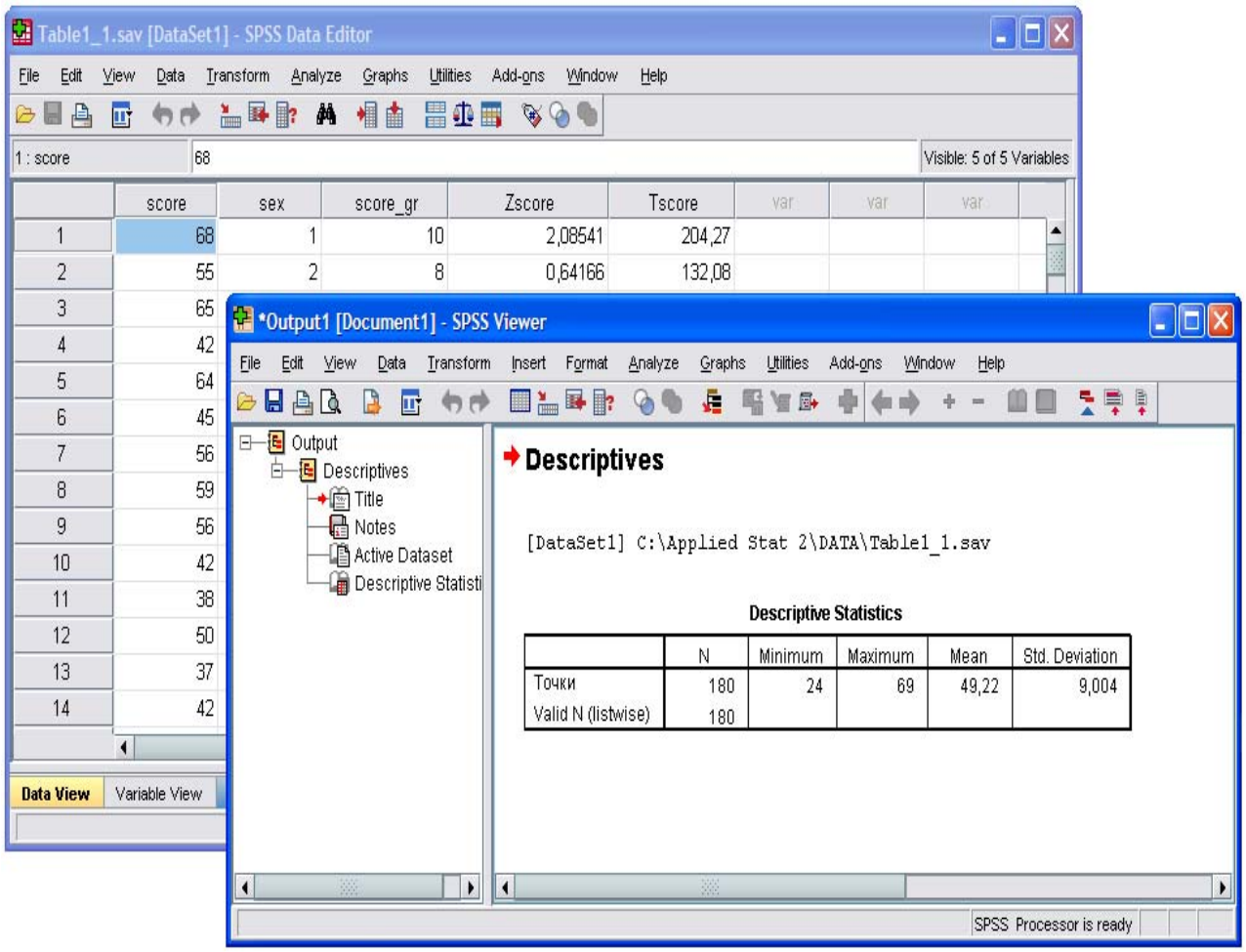
Редактор на въртящи се таблици (Pivot Table Editor). Резултатите, които се показват чрез въртящи се таблици могат да бъдат модифициране по много начини с помощта на този редактор, който се появява като отделен прозорец. Вие може да радактирате текста, да променяте местата на редовете и колоните, да добавяте цвят, да създавате допълнителни размерности на таблицата, избирателно да скривате колони или редове.

Графичен редактор (Chart Editor). В този самостоятелен прозорец вие можите да променяте графики с висока резолюция какот и да добавяте и премахвате елементи. Можете да променяте цвета, да избирате различни шрифтове и размери, да променяте хоризонталните и вертикалните оси, да въртите 3-D диаграми на разсейване и дори да променяте типа на графиката.

Редактор на тектови резултати (Text Output Editor). Резултатите, които не се показват като част от въртящи се таблици могат да бъдат променяни с помощта на редактора на текстови резултати. Вие можите да редактирате резултатите и да променяте характеристиките на шрифта (тип, стил, цвят, размер).

Синтаксис-редактор (Syntax Editor). Вие можете да създавате и променяте команди с помощта на този редактор. Можете да интегрирате синтаксиси и чрез директно използване на опцията Paste. Чрез синтаксис-редактора можете да ползвате специални възможности на пакета, които не са достъпни в диалогов режим. Всичко създадено с този редактор може да бъде запазено като тектови файл и използвано в други сесии на SPSS.

Скрипт редактор (Script Editor). Позволява да се автоматизира работата с SPSS.



Фиг.1-1 Прозорец „Данни” и прозорец „Резултати”

Менюта

Много от задачите, които SPSS може да реши са достъпни чрез избор на вградените минюта. Всеки прозорец в SPSS има свои собствен меню-бар с менюта, които са характерни за този тип прозорец.

Менютата „Analyze” и “Graphs” са достъпни от всички прозорци, което прави лесно генерирането на нови резултати без преминаване от един прозолец в друг.

Бар за състоянието (Status Bar)

Барът за състояние в дъното на всеки SPSS прозорец дава следната информация:

Команден статус. За всяка процедура или команда, която се изпълнява брояч на случаите, който показва докъде е стигнал процесът. За итеративните статистически процедури този брояч показва броя на итерациите.

Статус на филтъра. Ако сте извлекли случайна извадка или друго подмножество от измервания за анализ появавя се сйобщение “Filter on”, което показва, че някакъв вид филтър е активен в момента и не всички измирвания се включват в анализа.

Статус за претегляне. Съобщението „Weight on” показва, че данните са претеглени чрез съответната променлива с тегла.

Статус за разделяне на файла. Съобщението „Split File on” показва, че файлът е бил разделен на групи, които ще се анализират отделно. Това разделяне става чрез задаването на една или повече номинални групиращи променливи, например, пол.

Показване и скриване на бара за състояние

От менюто избираме: ***View → Status bar.***

Диалогови прозорци

При избора на повечето от менютата обикновено се отварят диалогови прозорци. Вие можете да използвате тези прозорци за да посочите променливите и опциите при анализа. Диалоговите прозорци на статистическите процедури и графиките обикновено имат две основни компоненти:

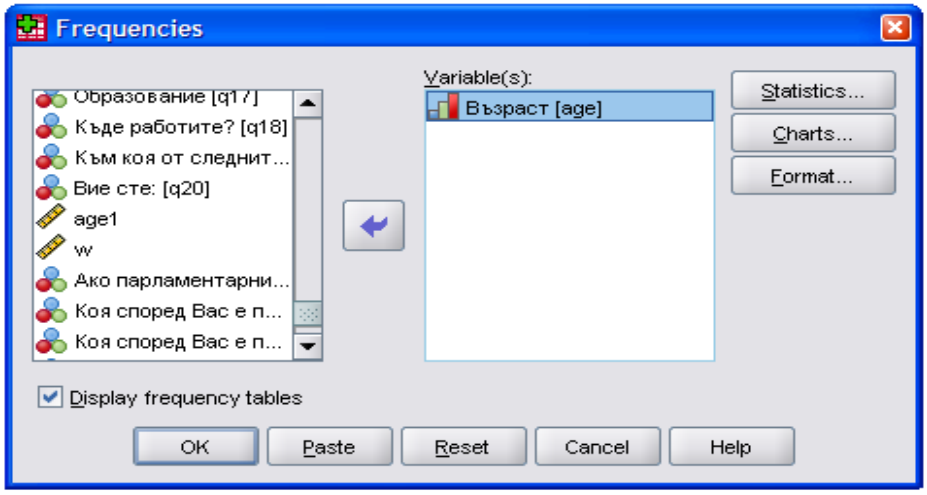
Изходен списък на променливите. Това е списък на променливите в активния файл с данни. В този списък се появяват само променливите, чиито типове, т.е. скалите на измерване, са подходящи за съответната процедура. Използването на къси и дълги стрингови променливи е ограничено от повечето процедури.

Целеви списък на променливите. Едни или повече списъци указват променливите, които сме избрали да участват в анализа, например списък на зависимите и на независимите променливи.

Имена и етикети на променливите в списъка на диалоговия прозорец

Вие можете да виждате или имената на променливите или техните етикети в списъка от диалоговия прозорец. За да контролирате кое да се покаже трябва да изберете от главното меню Edit→Options→General. След това избирате „Display Labels" или „Display Names" от правоъгълника с надпис ***Variable Lists***.

- За да дефинирате или промените етикетите на променливите използвайте Variable View от Data Editor.
- Когато данните се четат от база данни имената на полетата се използват като етикети.
- Ако променливата няма етикет за такъв се приема нейното име.



Фиг.1-2 Етикети на променливите както се ползват в диалоговия прозорец

Контролиране на диалоговите прозорци

Предвидени са пет стандартни начина за контрол в повечето диалогови прозорци:

OK. Натискането на тази контролна опция води до изпълнението на всяка една процедура. След като сте избрали вашите променливи и сте посочили допълнителните опции, посочене OK за да стартирате процедурата и да затворите диалоговия прозорец.
Paste (смесване). Генерира команден код от опциите, посочени в диалоговия прозорец като текстът се записва в прозореца на синтаксис-редактора. След това този текст е достъпен за модификации и добавяне на опции, до които няма достъп в диалогов режим.
Reset. Отменя всички посочени действие до момента и възстановява изходното състояние на диалоговия прозорец на съответната процедура без да го затваря.
Cancel. Отменя всички промени, направени в текущия диалогов прозорец и го затваря.
Помощ (Help). Дава контекстно-чувтвителна помощ. Посочването на Help ви препраща към помощно меню, което се отнася до отворения диалогово прозорец. Такъв вид контекстно-чувствителна помощ може да се получи и чрез натискането на десния бутон на мишката.

Поддиалогови прозорци

Тъй като повечето процедури предлагат множество възможности за избор не всичко може да се постави в един прозорец. Овновният диалогов прозорец обикновено предлага избор на най-важната (минималната) информация, необходима за работата на процедурата. Допълнителните възможности са оформени като подпрозорец. Ако в основния диалогов прозорец след дадено опция има мнготочие (...) това означава, че имате достъп до подпрозорец с допълнителни избори.

Избор на променливите

За да изберете отделна променлива просто я маркирайте в изходния списък на променливите и натиснете бутона със стрелкичката вдясно от този списъка. Тази стрелкичка трябва да е насочена на дясно. Ако имате само един изходен списък можете да изберете нужната променлива и с двукратно натискане на левия бутон на мишката върху променливата.
Можете също да правите и множествен избор на променливи:

- За да изберете едновременно няколко съседни променливи маркирайте първата променлива и след това с натиснат клавиш Shift посочете последната променлива от групата.
- За да изберете едновременна няколко променливи, които не са съседни постъпете така. Маркирайте първата променлива и след това с натиснат клавиш Ctrl маркирайте всички останали.

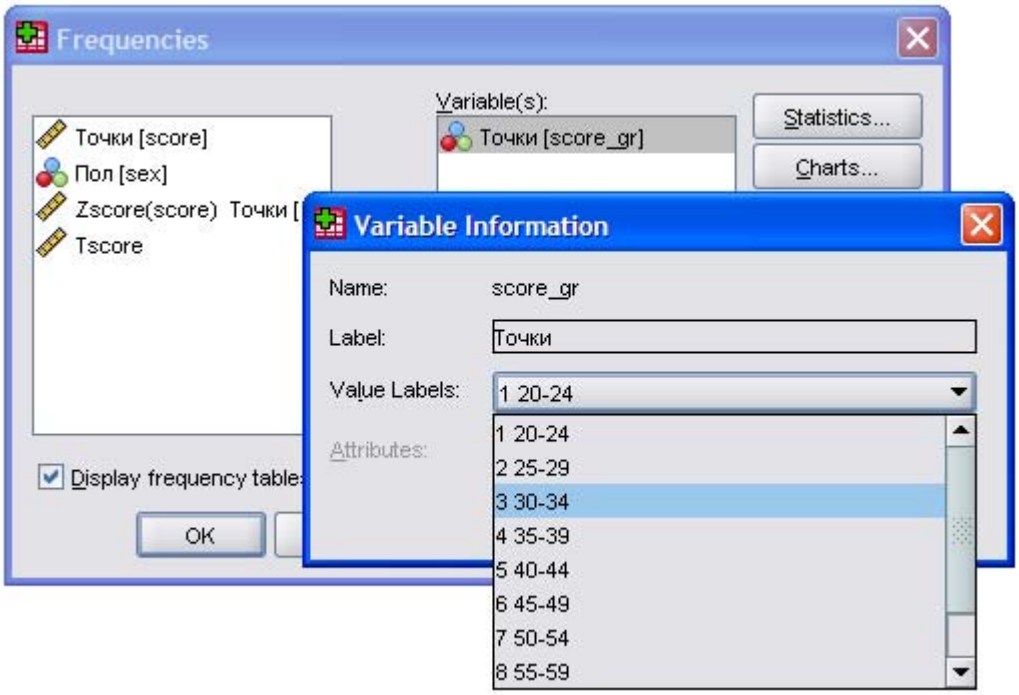
Икони в списъка на променливите

Иконката вляво на всяка променлива в диалоговия прозорец дава информация за типа на променливата и скалата на измерване.

Скала на измерването	**Тип на данните**			
	Числови	Нечислови (стрингови	Дата	Време
Метрична (Scale)		n/a		
Рангова (Ordinal)				
Номинална(Nominal)				

Достъп до информация за променливите в диалоговия прозорец

- Посочвате променливата и натискате десния клавиш на мишката.
- Избирате Variable Information.



Фиг.1-3 Информация за променливата

Основни стъпки при анализа на данни

Анализът на данни с SPSS е лесен. Всичко, което трябва да направите е:

Зареждане на данните в SPSS. Вие можете да отворите предварително запазен SPSS файл с данни, можете да прочетете таблица от външна програма (например, EXEL), база данни, тестови файл с данни, или да ги въведете директно с Редактора на данни.

Избор на процедура. Изберете процедура от менюто за да направите нужните изчисления или да създадете подходяща графика.

Избор на променливи за анализа. Променливите от файла с данни се показват в диалоговия прозорец на процедурата. От този списък изберете нужните.

Стартиране на процедурата и преглед на рузелтатите. Резултатите се появяват в прозореца „Viewer”.

Статистически съветник

Ако не сте на ти с SPSS или с някоя от необходимите ви статистически процедури Статистическия съветник може да помогне да започнете работа като ви задава несложни въпроси на неспециализиран език и показва примери, които помагат за избора на базовите статистически и графични подходи, които най-добре съответстват на вашите данни. За да използвате Статистическия съветник от менюто на всеки SPSS прозорец избирате: *Help → Statistics Coach*.

Статистическия съветник покрива само част от процедурите в базовата система. Той е планиран да предложи базова помощ за много от често използваните процедури.

Повече информация за SPSS

За обстойно описание на основните възможности на пакета може да се види онлайн тюториала. От всяко SPSS меню избирате: ***Help → Tutorial***.

Глава 2. Как да поискаме помощ от SPSS

Поискване на помощ

Помощта е достъпна по много и различни начини:

Меню „Help”. Менюто Help в повечето прозорци на SPSS осигурява достъп до главната система за помощ зплюс тюториали и шехнически справочни материали.,

Подменю „Topics”. Осигурява достъп до Съдържание, Индекс и Търсене, чрез които можете да откриете по-специални опции, които ви интересуват.

Подменю „Tutorial”. Дава стъпка по стъпка инструкции как да се използват много от основните възможности на SPSS. Вие не трябва да разглеждате целия тюториал от начало до край. Можете да изберете отделни теми за разглеждане, да прескачате теми или да разглеждането материала в произволен ред, да използвате индекс или таблица на съдържанието за да намерите определениа тема.

Подменю„Case Study”. Показва примери как да се създават различни типове анализи и кака да се интерпретират резултатите. Примерните файлове с данни, използвани в примерите също се показват за да можете да видети реално как точно се получават резултатите. Можете да посочите специфична процедура, която искате да разберете как работи, от съдържанието или да търсите чрез индекс.

Подменю „**Statistics Coach**”. Това е пощно маню, което ви дава възможност да намерите процедурата, което е подходяща за вошите данни. След като направите серия от избори Statistics Coach отваря диалогов прозорец на статистическа, графична или друга подходяща процедура, която най-добре съответства на избраните от вас критерии. Statistics Coach осигурява достъп до повечето статстически и графични процедури на базовия модул.

Подменю “**Command Syntax Reference**”. Подробна помощ за синтаксиса на процедурите в SPSS е достъпна по два начина: чрез общата Help система и като отделен PDF файл, до който е осигурен достъп в менюто Help под заглавие *SPSS Command Syntax Reference*.

Статистически алгоритми. Алгоритмите, които се използват в повечето статистически процедури са достъпни по два начина: чрез общата Help система и като отделен PDF файл, до който е осигурен достъп в менюто Help под заглавие *SPSS Algorithms*.

Контекстно-чувствителна помощ. На много места в потребителския интерфейс можете да получите контектна помощ.

Бутон за помощ в диалоговите прозорци. Повечето от диалоговите прозорци имат бутон за помощ “Help”, който ви дава дирекно помощ, свързана с този прозорец. Предлага се базова информация и насочване към свързана с тази процедура помощ.

Контекстна помощ чрез диалоговите прозорци. Много от диалоговите прозорци предлагат контексно-чуствителна информация относно специфични свойства на процедурата. Чрез натискането на десния бутон на мишката и избор на „What’s This?” на екрана се извиква помощ относно контролен бутон или указания за неговото използване. Ако текстът „What’s This?” не се появи, то този вид помощ не е достъпна за този диалогов прозорец.

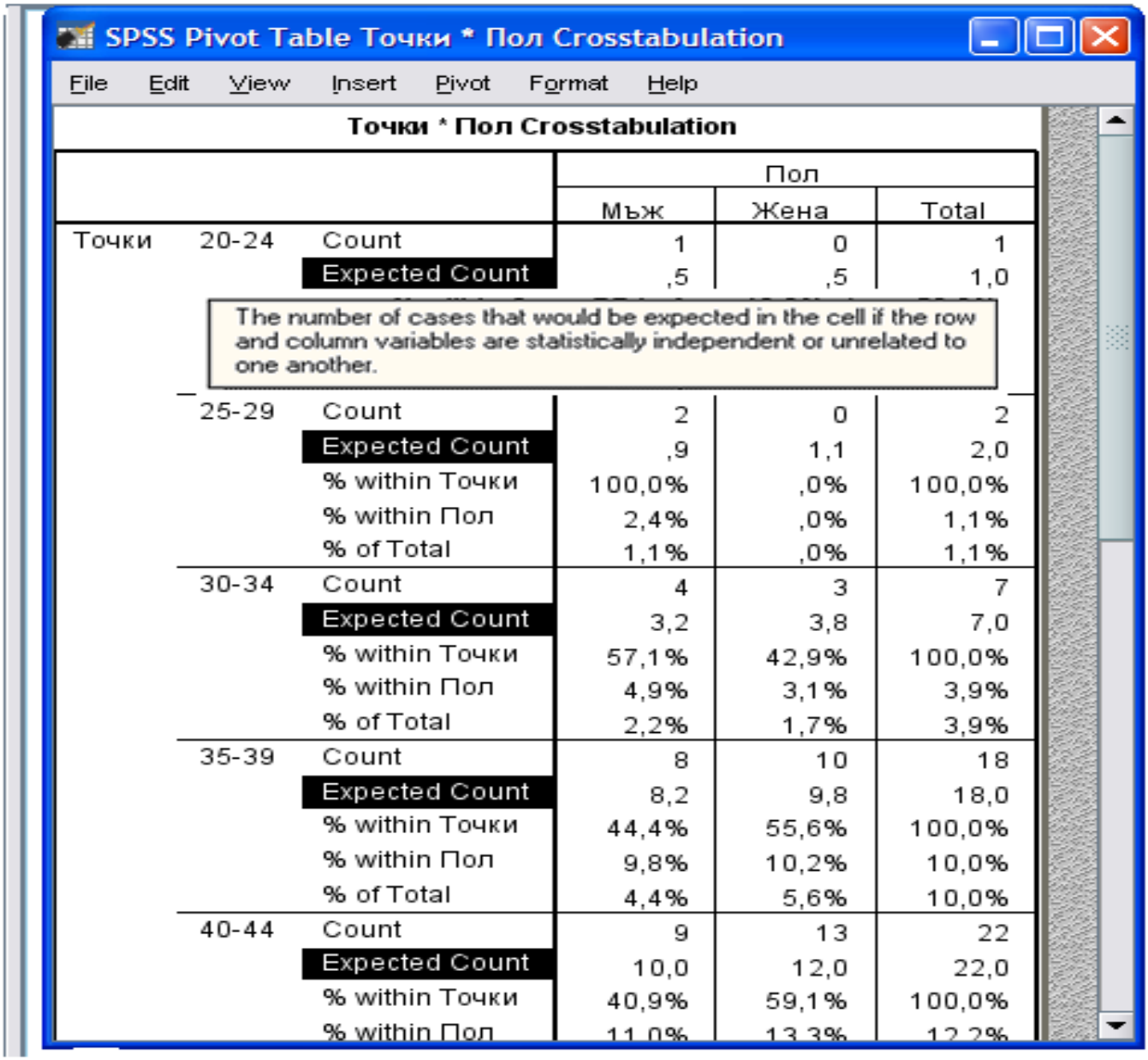
Помощ за синтаксиса. В синтактичния прозореца поставете курсора на мишката където и да е върху командния ред и натиснете F1 от клавиатурата. Ще се появи пълно описание на командата, както и връзки към допълнителна информация.

Помощ при тълкуване на рузултатите

За да видите определенията на термините във въртящата се таблица в прозореца Viewer трябва:

- да посочите два пъти таблицата за да я активирате;
- натискате десния бутон на мишката върху термина, за който искате пояснение;
- избирате „What’s This?” от контекстното меню.

Дефиницията се появява като допълнителен прозорец под термина.



SPSS Pivot Table Точки * Пол Crosstabulation

File Edit View Insert Pivot Format Help

Точки * Пол Crosstabulation

			Пол		
			Мъж	Жена	Total
Точки	20-24	Count	1	0	1
		Expected Count	,5	,5	1,0
	25-29	Count	2	0	2
		Expected Count	,9	1,1	2,0
		% within Точки	100,0%	,0%	100,0%
		% within Пол	2,4%	,0%	1,1%
		% of Total	1,1%	,0%	1,1%
	30-34	Count	4	3	7
		Expected Count	3,2	3,8	7,0
		% within Точки	57,1%	42,9%	100,0%
		% within Пол	4,9%	3,1%	3,9%
		% of Total	2,2%	1,7%	3,9%
	35-39	Count	8	10	18
		Expected Count	8,2	9,8	18,0
		% within Точки	44,4%	55,6%	100,0%
		% within Пол	9,8%	10,2%	10,0%
		% of Total	4,4%	5,6%	10,0%
	40-44	Count	9	13	22
		Expected Count	10,0	12,0	22,0
		% within Точки	40,9%	59,1%	100,0%
		% within Пол	11,0%	13,3%	12,2%

The number of cases that would be expected in the cell if the row and column variables are statistically independent or unrelated to one another.

Фиг. 2-1. Активирана помощ за термините (глосарий) на въртяща се таблица

Глава 3. Опции

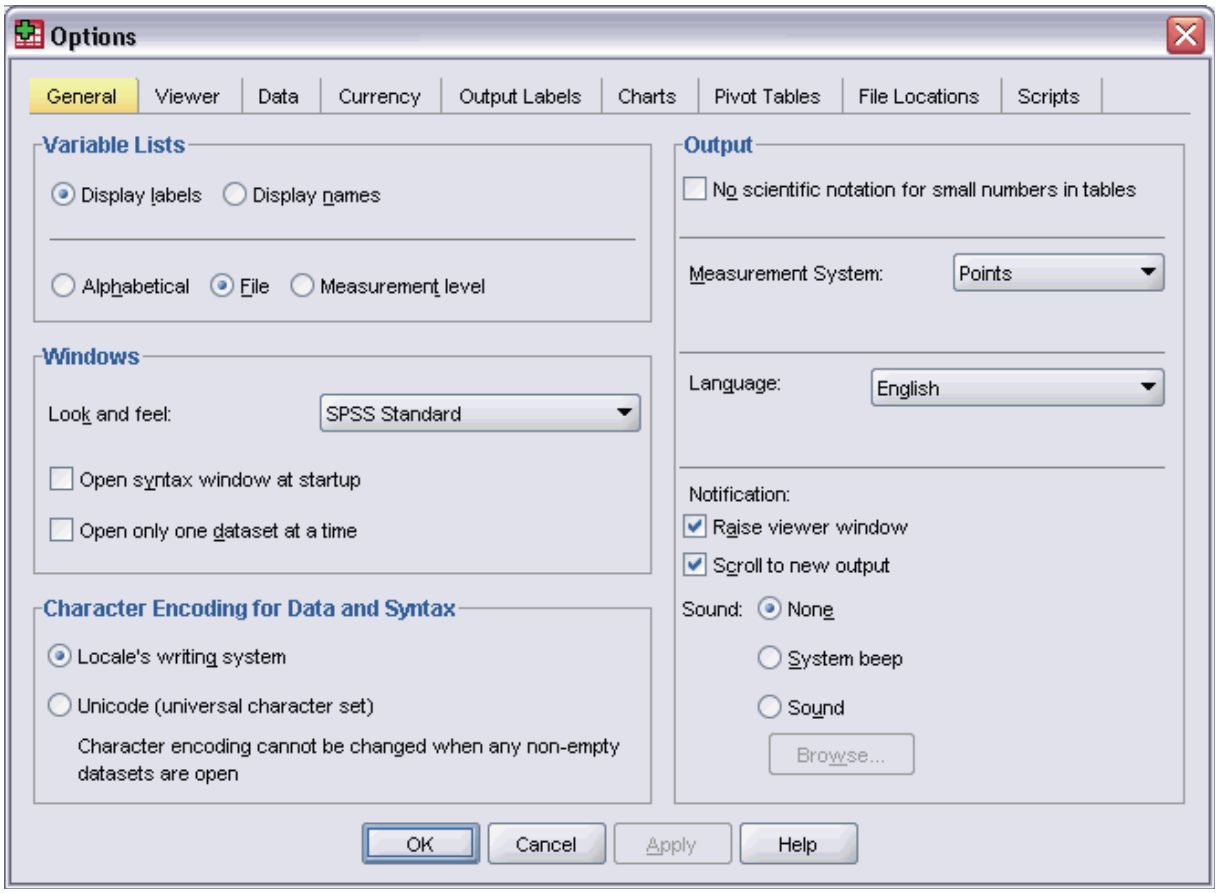
Опциите контролират широко разнообразие от възможности, включващи:

- Дневник на сесиите, който поддържа списък на всички команди, използвани при всяка сесия;
- Избор на подредбата на променливите в началните списъци в диалговия прозорец;
- Показване и скриване на елементи в резултативния прозорец;
- Нов изглед на въртящите се таблици;
- Стандартни формати за парични единици;

За да промените настройките на опциите изберете от менюто:

➢ Edit
Options…
Кликнете отметките на настройките, които искате да промените;
Променете настройките.

➢ Натиснете OK или Apply.



Фиг.3-1 Диалогов прозорец на опциите

ГЛАВНИ ОПЦИИ

Списък на променливите (Variable List)

Тези условия контролират дисплея на променливите в списъците на диалоговия прозорец. Можете да изобразите имената на променливите или етикетите на промеливите. Имената или етикетите могат да бъдат изобразени в азбучен или файлов ред, или групирани според техните размери. Редът на изображение влияе само на началните (source от англ. „начало, източник”) списъци на променливите. Списъците с обектните променливи винаги рефлектират на реда, по който се избират променливи.

Windows

Поглед и разбиране. Контролира основния вид на Windows и диалоговите прозорци.

Отваряне на синтактичен (syntax) прозорец при зареждане

Синтактичните прозорци са прозорци за текстови файлове, използвани за вход, редакция и за задаване на команди. Ако работите редовно с команден синтаксис, изберете тази опция за автоматично отваряне на синтактичен прозорец в началото на всяка сесия. Това е полезно главно за опитни потребители, които предпочитат да работят с команден синтаксис вместо диалогови прозорци.

Отваряне само едни файл от данни (Dataset) за съответна сесия

Затваря текущо отворения източник на данни всеки път, когато отваряте различен източник на данни, изполвайки менюта и диалогови прозорци. По стандарт, всеки път, когато използвате менютата и диалоговите прозорци, за да отворите нов източник на данни, този източник на данни се отваря в нов Прозорец за Редактиране (Data Editor Window). Тогава всички други източници на данни, отворени в други Прозорци за Редактиране (Data Editor Window), остават отворени и достъпни по време на сесията, докато не бъде изрично затворен съответния Прозорец за Редактиране.

Когато избирате тази опция, тя има моментален ефект, но не затваря никакви конфигурации от данни, които са били отворени по времето, когато настройката е променена. Тази настройка не указва ефект на източниците на данни, отворени чрез употребата на синтаксис, който се основава на DATASET команди, за да управлява многобройни конфигурации от данни.

Output

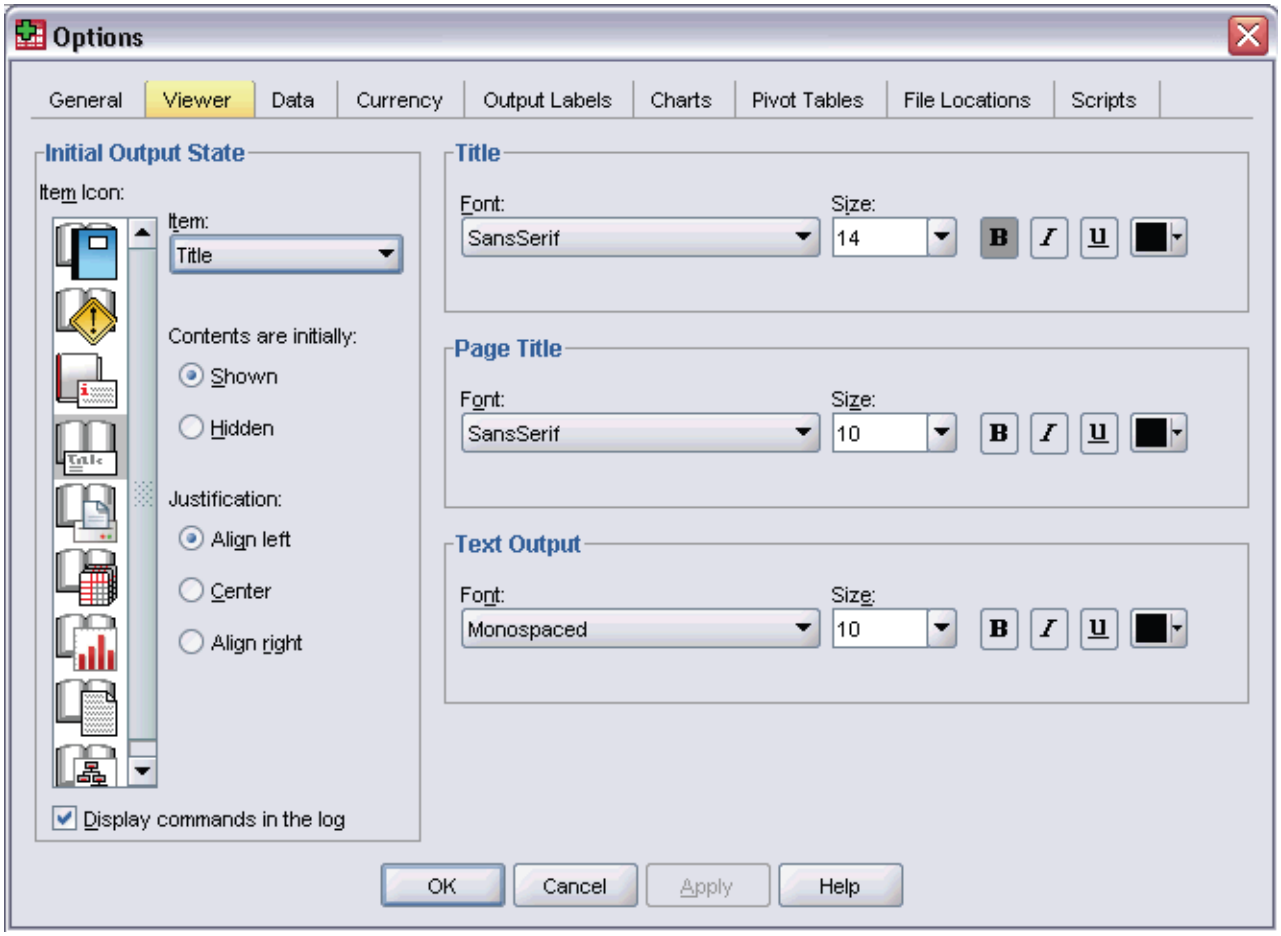
Отмяна на научната нотация на твърде малките числа. Променя формата на представяне на много малките числа от научен (scintific) към десетичен. Прекалено малките десетични стойности ще бъдат изобразени като 0 (или 0.000).

Измервателна система. Дефинира в какава система да се измарват дължините на низовете от символи: точки, инчове, или сантиметри.

Език. Контролира езика използван при изхода. Не важи за текстов изход. Списъкът на възможните езици зависи от текущите инсталирани файлове с езици.

ОПЦИИ НА VIEWER

Тези опции стават валидни след стартиране на нова сесия.

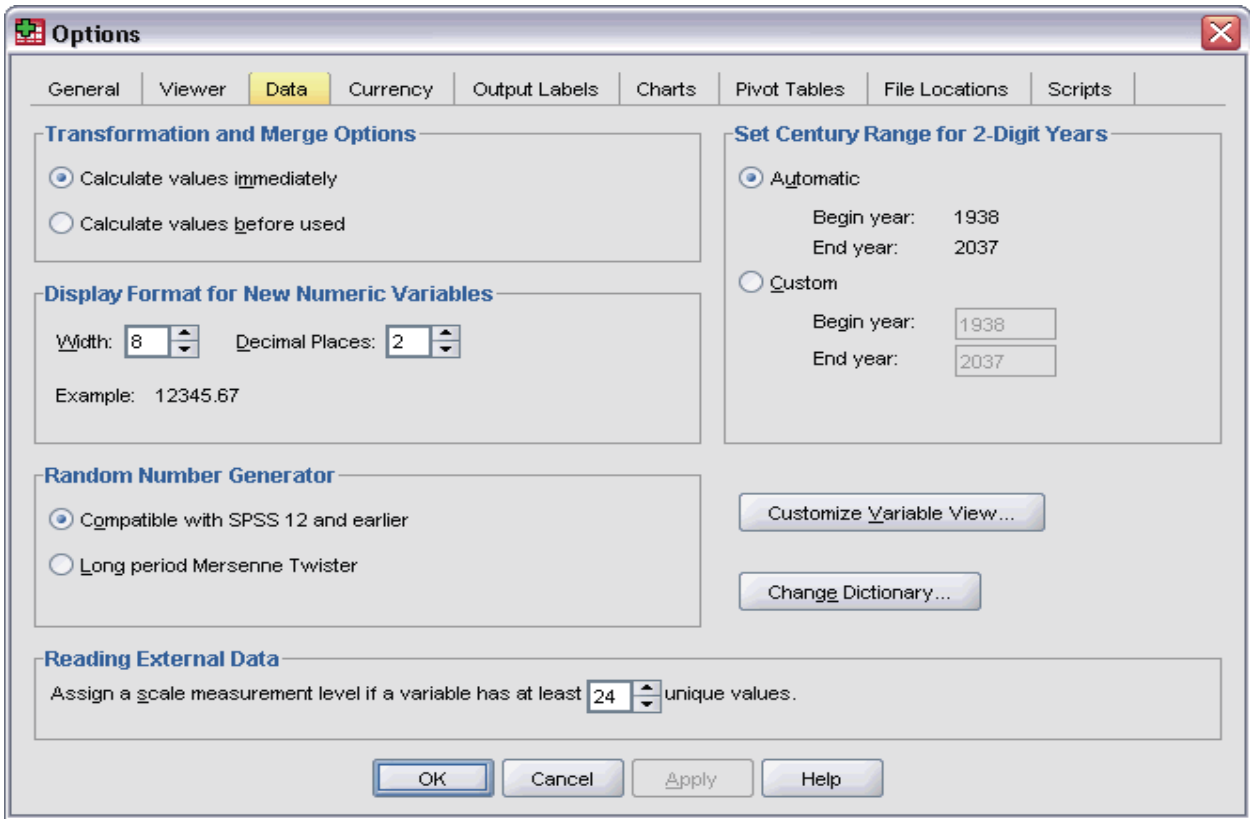


Фиг. 3-2 Диалогов процорец на опциите

Изходно състояние (Initial Output State). Контролира кои елементи да се показват или скриват всеки път, когато ползвате дадена процедура и как елементите да се подредятени, когато са инициализирани. Можете да контролирате изобразяването на екрана на следните елементи: log, предупреждения, бележки, заглавия, pivot таблици, charts, дървовидни диаграми и текстов изход. Също може да включите показването на командити в log-файла, след което можете да копирате командите и да гио запазите в syntax файл.

Заглавие. Контролира външния вид, големината и цвета за новите заглавия на страницата.

ОПЦИИ ЗА РАБОТА С ДАННИТЕ



Фиг. 3-4 Диалогов прозорец – данни

Трансформациои и сливане. Всеки път, когато програмата изпълнява команда, чете файловите данни. Някои трансформации на данни (като изчисляване и прекодиране) и файлови трансформации (като прибавяне на променливи и прибавяне на cases) не изисква отделна работа с данните, а изпълняването на тези команди може да бъде забавено докато програмата чете данни, за да изпълни друга команда, като например статистическа или графична процедура.

За големи файлови масиви, където четенето на данни може да отнеме известно време, можете да изберете Calculate values before used (Изчислете стойностите, преди да бъдат изполвани), за да забавите изпълнението и да запазите времето за обработка. Когато тази възможност е избрана, резултатите от трансформациите, които правите , използвайки диалогови прозорци като изчисляване на променливи, няма да се появят веднага в Редактора на Данни, новите променливи създадени чрез трансформации ще бъдат изобразени като липсващи данни, а стойностите на данните в Редактора на Данни не могат да бъдат променени, докато има висящи трансформации. Всяка команда, която чете данни, ще извърши висящите трансформации и ще обнови данните в Редактора на данни. Аналогично, можете да използвате Run Pending Transformation от менюто за трансформиране.

По подразбиране се изпълнява Calculate values immediately (Моментално пресмятане на стойности).

Формат за нови числови променливи. Контролира стандартното изображение на ширина и брой на десетичните места за нови числови променливи. Няма стандартен

формат за изображение за нови низови променливи. Ако стойността е прекалено голяма за специализирания формат на изображение, първо десетичните стойности се закръглят и тогава стойностите се конвертират към спецефична нотация. Изобразените формати не влияят на постоянните стойности на данните. Например, стойността 123456.78 може да бъде закръглена на 123457 при изобразяването, но оригиналната незакръглена стойност се използва в изчисленията.

Пускане на Обхват за век и двуцифрена година. Дефинира стартовата година, което ще се използва при формати, в които годината се изобразява само с послединте две цифри.

Генератор на случайни числа. Предлага избор от два различни случайни числови генератори. Първият е този, използван във версия 12 и предишни версии, а вторият е известен като „**Mersenne Twister".**

Четене на външни данни. За данни четени от външни файлови формати и по-стари SPSS- файлови данни (преди версия 8.0), можете да уточните минимален брой на стойностите на данните на числовите променливи, използвани за да класифицират променливата като номинална или метрична. Променливите с по-малко от уточнения брой уникални стойности са класифицирани като номинални.

Стандартен поглед върху променливите. Контролира стандартния дисплей и подредбата на свойствата в прозореца за преглед на променливите.

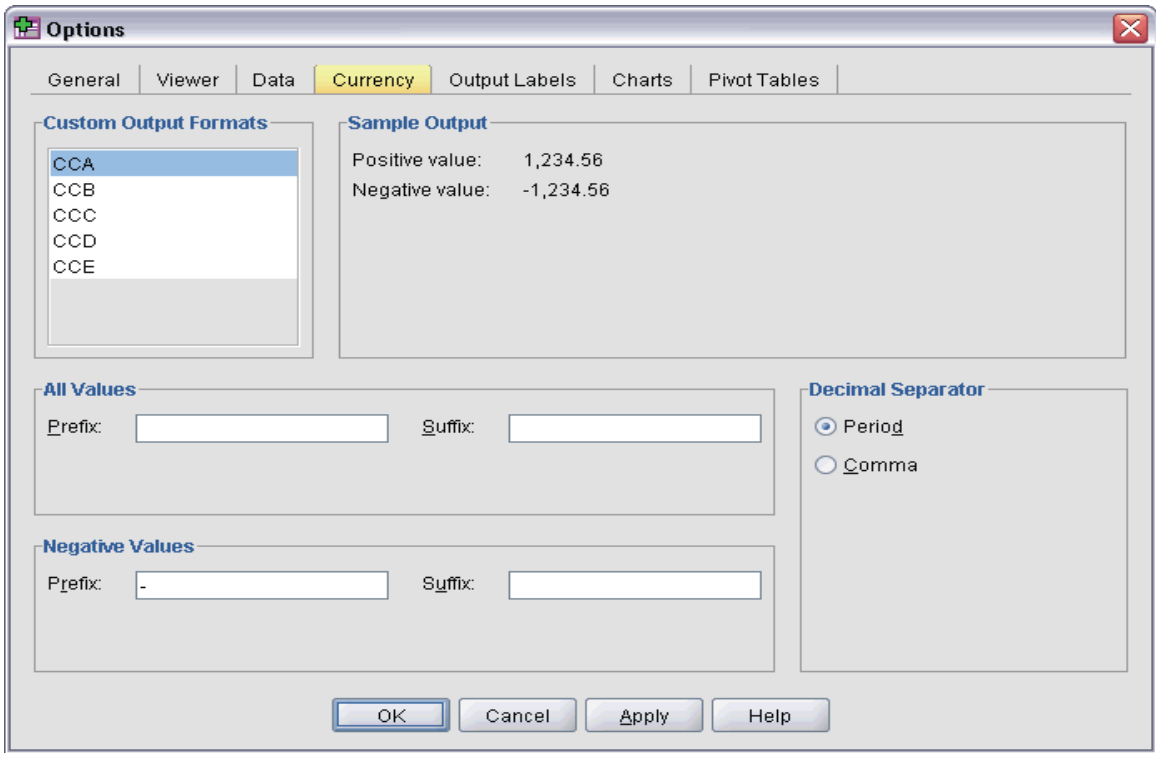
Промяна на директория. Контролира езиковата версия на речника използвана за проверка спелинга на елементите в Variable View.

Промяна на Default Variable View. Може да използвате Customize variable View, за да контролирате кои характеристики да бъдат изобразени стандартно в Variable View (например име, тип, етикет) и реда, по който да бъдат изобразени. Натиснете Customize Variable View.

ОПЦИИ ЗА ВАЛУТА

Можете да създадете до 5 стандартни формата за означаване на валута, които могат да включват специални букви за представки и наставки и специално изобразяване на отрицателни стойности.

Петте стандартни валидни форматни имена са CCA, CCB, CCC, CCD и CCE. Не можете да промените форматните имена и да добавяте нови. За да модифицирате стандартния валиден формат, изберете форматно име от списъка на източниците и направете промените, които желаете.



Фиг. 4-5 Диалогов прозорец: Опции за валута

Представките, наставките и десетичните индикатори, дефинирани за стандартни валидни формати са предназначени само за цели, свързани с изобразяването им. Не можете да въвеждате стойности в Редактора за Данни, използвайки стандартни валидни букви, шрифтове.

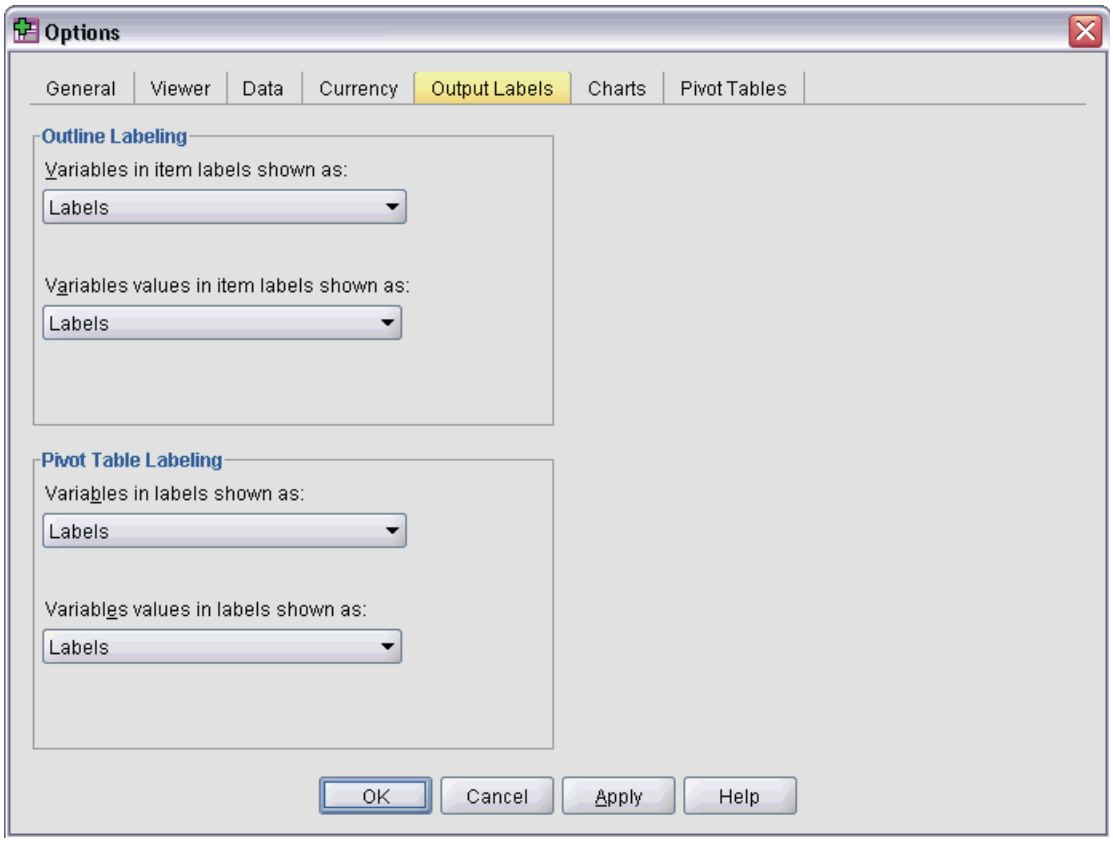
Създаване на стандартни валидни Формати:

- Натиснете Currency tab.
  - Изберете един от текущите формати от списъка (CCA, CCB, CCC, CCD and CCE)
  - Въведете префиксите, суфиксите и десетични индикаторни стойности.
- Натиснете OK или Apply.

ОПЦИИ ЗА ЕТИКЕТИТЕ

Изходните опции на етикетите контролират изображението на променливите и информацията за стойността на данните в резултативните таблици. Можете да изкарате на екран имена, етикети и актуални стойностти на данни, дефинирани стойности на етикети, или комбинация.

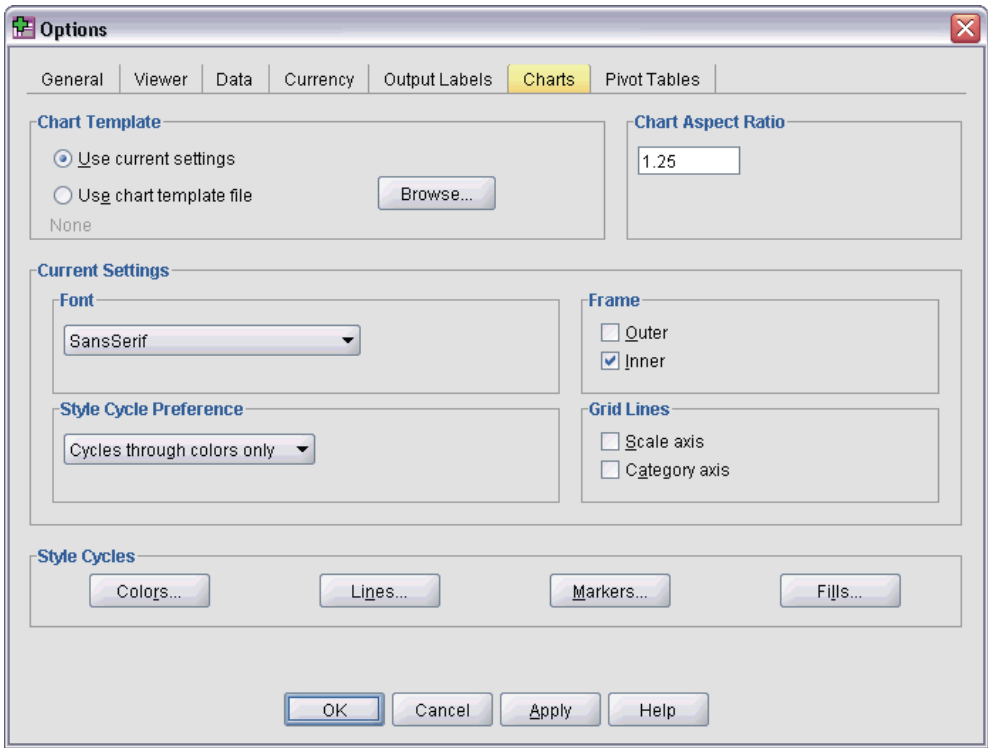
Описанията на стойностите и етикетите правят по-лесно интерпретирането на вашите резултати. Като цяло, дългите етикети могат да бъдат неудобни в някои таблици.



Фиг. 4-6 Етикети в резултатите

Изходните опции за етикета засягат само новите резултати, създадени след като сте променили настройките. Те засягат само променливите (pivot) таблици. Текстовият резултат не се влияе от тези настройки.

ГРАФИЧНИ ОПЦИИ



Фиг. 4-7 Графични опции

Chart Tamplates. Новите графики могат да използват или настройките, които са избрани тук, или настройките от графичния шаблонен файл. Натиснете Browse, за да изберете шаблонен от файл. За да създадете шаблонен файл, създайте таблица с характеристиките, които желаете и ги запазете като шаблон (изберете Save Chart Template от файловото меню).

Chart Aspect Ratio. Това е отношението на ширината и дължината на графиката. Можете да уточните коефициент ширина към височина от 0.1 до 10.0. Стойности по-малки от 1 правят таблиците по-високи, отколкото широки. Стойности по-големи от 1 правят по широки таблици, отколкото високи. Стойността 1 произвежда квадратна таблица. Веднъж създадена таблицата, а също и коефициентът, не могат да бъдат променени.

Current Settings (Текущи Настройки). Възможните настройки включват:

Font. Шрифтът се използва за целия текст в новите таблици.

Style Cycle Preference. Начално указване на цветове и символи за нови таблици.

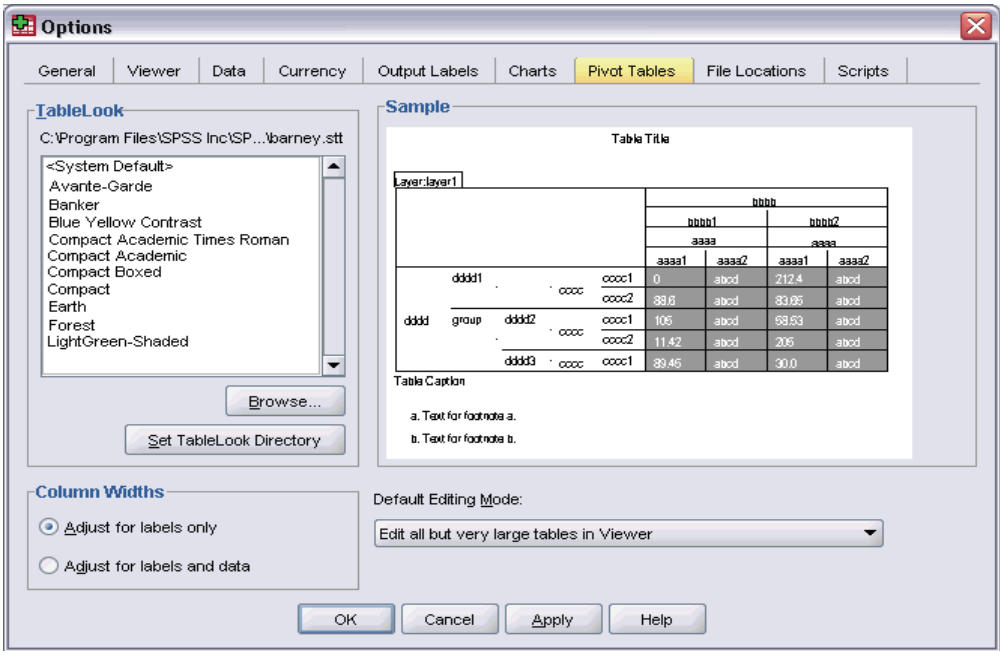
Frame. Контролира изображението на вътрешни и външни рамки в новите таблици.

Grid Lines. Мрежови линии. Контролира изображението на скалата и категорията на осевите линии на новите таблици.

Style Cycles. Задава цветовете, стила на линиите, маркиращите символи и попълва моделите за новите таблици. Можете да промените реда на цветовете и моделите, които са използвани, когато се създава нова таблица.

ОПЦИИ ЗА ОБОБЩЕНИ ТАБЛИЦИ (Pivot Tables)

Опциите за работа с обобщени таблици задават стандартния вид на новосъздаваните таблици. Оттук могат да се контролират всички характеристики на елементите на таблиците.

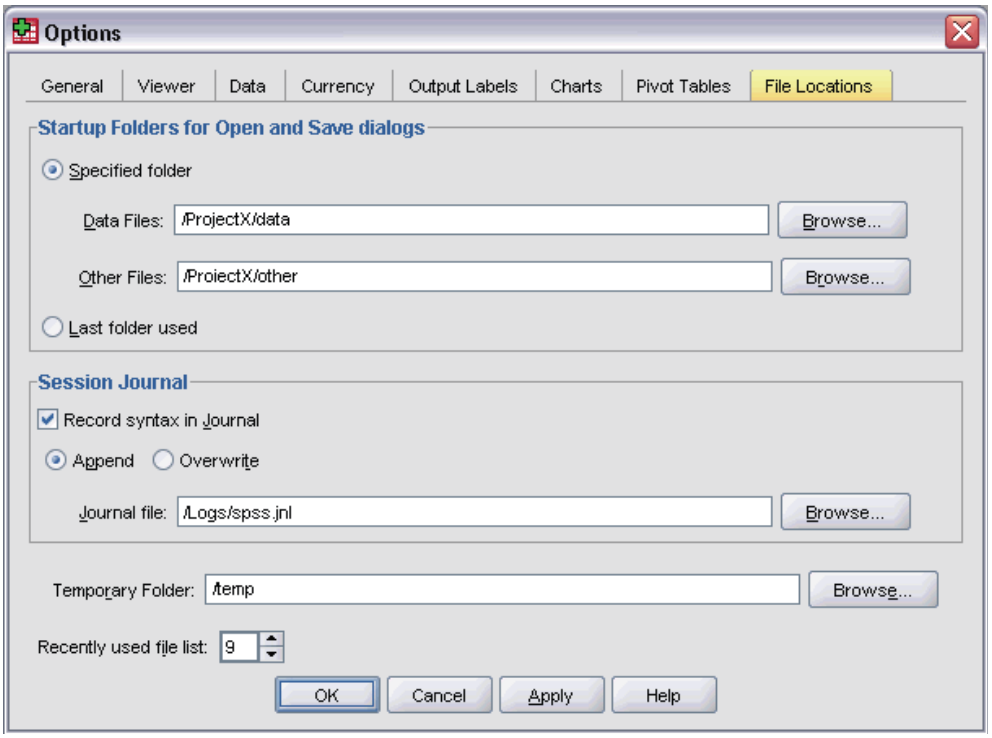


Фиг. 4-8 Диалогов прозорец на опциите: Етикет на обобщени таблици

Table Look (Изглед на Таблицата). Изберете от списъка от възможни изгледи. Можете да използвате стандартните SPSS изгледи или можете да създадете ваш собствен изглед, чрез Редактор за Обобщени Таблици.
Browse (Разглеждане). Позволява ви да изберете изглед на таблицата от различна папка.
Set TableLook Directory (Задаване на папка за таблични изгледи). Позволява промяна на стандартната директория .
Забележка: Таблиците, които са създадени с по-ранни версии на SPSS не могат да бъдат използвани във версии 16.0 или по-късни от нея.

ОПЦИИ ЗА ЛОКАЛИЗИРАНЕ НА ФАЙЛ

Тези опции контролират стандартното местоположение, откъдето ще четат и записват файловете в началото на всяка сесия, местоположението на файла на дневника, местоположението на временната папка, броят на файловете, които се появяват във файловя списък „Последно използвани”.



Фиг. 4-9 Диалогов прозорец на опциите: Локализиране на файлове

Глава 4. Файлове с данни

Файлове с данни

Файловете с данни имат широко разнообразие от формати. Този софтуер е проектиран да се справи с много от тях, включително:

- Електронни таблици, създадени с Excel и Lotus.
- Таблици от бази данни от много източници на бази данни, включително Oracle, SQLServer, Access, dBASE и други.
- Tab-delimited и други типове прости текстови фаилове.
- Файлове с данни в SPSS формат, създадени чрез други системи.
- SYSTAT файлове с данни
- SAS файлове с данни
- Stata файлове с данни

Отваряне на файлове с данни

В допълнение на файловете, запазени в SPSS формат, можете да отворите Excel, SAS, Stata, tab-delimited и други файлове без да трансформирате файловете или да въвеждате информация за дефиницията на данните.

Отварянето на файлове с данни активира набор от данни. Ако вече имате един или повече отворени файлове с данни, те си остават отворени и са достъпни за следваща употреба в сесията. Произволното кликване в прозореца Редактор на Данни за отворен файл с данни ще активира този набор от данни.

За да отворите файлове с данни:

- ➢ От менюто изберете:
  - File
    - Open
      - Data
- ➢ В отворения прозорец Open Data, изберете файла, който желаете да отворите.
- ➢ Натиснете Open.

Чрез опциите, можете да:

- Чете имената на променливите от първия ред от електронните таблици във файла.
- Уточнява диапазона на клетки за прочитане от таблиците във файла
- Уточнява таблица във файл на Excel за четене (Excel 5 или по-късна версия).

Типове Файлове с Данни

SPSS. Отваря файлове с данни, записани в SPSS формат, включително SPSS за Windows, Macintosh, UNIX, и също така продукт на DOS – SPSS/PC+.
SPSS/PC+. Отваря SPSS/PC+ файлове с данни.
SYSTAT. Отваря SYSTAT файлове с данни.
SPSS Portable. Отваря файлове с данни, записани в SPSS преносим формат. Записване на файлове в преносим формат отнема значително по-дълго време, отколкото записването в SPSS формат.
Excel. Отваря файлове на Excel.

Lotus 1-2-3. Отваря файлове с данни записани във формат 1-2-3 за версия 3.0, 2.0 или 1A на Lotus.
SYLK. Отваря файлове с данни записани в SYLK (символичен линк) формат, формат използван от някои приложения на електронните таблици.
dBASE. Отваря файлове dBASE - формат за или dBASE IV, dBASE III или III PLUS, или dBASE II. Всеки случай е документ. Променливите и стойностите на етиктите и липсващите стойности на спацификациите се губят, когато запазите файл в този формат.
SAS Long File Name. SAS версии 7-9 за Windows, дългата версия.
SAS Short File Name. SAS версии 7-9 за Windows, късата версия.
SAS v6 за Windows. SAS версия 6.08 за Windows и OS2.
SAS v6 за UNIX. SAS версия за UNIX (Sun,HP,IBM).
SAS Transport. SAS транспортируем файл.
Stata. Stata версии 4-8.
Текст. ASCII текстов файл.

ОПЦИИ НА ОТВАРЯНЕ НА ФАЙЛ

Четене на имената на променливите. За електронните таблици, можете да четете имената на променливите от първия ред на файла или първия ред на дефинираната област. Стойностите са конвертирани като необходими, за да се създадат валидни имена на променливите, включително конвертиране на интервали в долни черти.
Работен лист. Excel 5 или по-късни файлове могат да съдържат множество работни листи. Стандартно, Редакторът на Данни чете първия работен лист. За да прочете различен работен лист, изберете работния лист от падащия списък.
Обхват. За файловете с данни на електронни таблици, можете да прочетете областта с клетки. Използвайте същия метод за уточняване на областите на клетките, както бихте направили с приложенията на таблиците.

Четене на Excel 5 или по-късни файлове

Следните правила се отнасят за четене на Excel 5 или по-късни файлове:

- **Тип на данни и ширина.** Всяка колона е променлива. Типът на данните и ширината на всяка променлива се определят от типът и ширината в файла на Excel. Ако колоната съдържа повече от един тип на данни (например, дата и цифри), типът на данните се пуска като низ, и всички стойности биват четени като валидни низови стойности.
- **Празни клетки.** За цифрови променливи, празните клетки се конвертират към стойност липсваща от системата, идикирана чрез период. За низови променливи, празно поле е валидна низова стойност, а празните клетки се считат за валидни низови стойности.
- **Имена на променливи.** Ако четете първия ред на файл на Excel (или първия ред на определена област) като имена на променливи, стойностите, които не съответстват на правилата за наименование на променливи се конвертират във валидни имена на променливи, а оргоналните имена се използват като етикети на променливи. Ако не разчитате имена на променливи от файла на Excel, тогава биват възлагани имена на променливи по подразбиране.

Четене на по-стари файлове на Excel и други електронни таблици

- **Тип на данни и ширина.** Всяка колона е променлива. Типът на данните и ширината на всяка променлива се определят от типът и ширината в файла на Excel. Ако колоната съдържа повече от един тип на данни (например, дата и цифри), типът на данните се пуска като низ, и всички стойности биват четени като валидни низови стойности.
- **Празни клетки.** За цифрови променливи, празните клетки се конвертират към стойност липсваща от системата, идикирана чрез период. За низови променливи, празно поле е валидна низова стойност, а празните клетки се считат за валидни низови стойности.
- **Имена на променливи.** Ако не разчитате имената на променливите от таблицата, буквите на колоната (A, B, C, …) се използват за имена на проемнливи за Excel и Lotus файлове. За SYLK файлове и Excel файлове, записани в R1C1 формат на дисплей, програмата използва предхождащия номер на колоната с буквата C за имена на променливи (C1, C2, C3, …).

Четене на dBASE файлове

Файловете базаданни са логически много близки да файловете с данни в SPSS формат. Следните главни правила се отнасят за dBASE файловете:

- Имената в полето се конвертират във валидни имена на променливи.
- Колоните, използвани в dBASE полета с имена се превеждат в долни черти.
- Документите, маркирани за изтриване, но които не са всъщност почистени са включени. Програмата създава нови низова променлива, D_R, която съдържа звездичка за примеите маркирани за изтриване.

Четене на Stata файлове

Следните главни правила се отнасят за Stata файлове с данни:

- Имена на променливи. Stata имена на променливи се конвертират до SPSS имена на променливи във формат на чуствителност към малки и големи букви. Stata имена на променливи, които са идентични освен за case се конвертират във валидни SPSS имена на променливи чрез добавяне на долни черти и допълнителни букви (_A, _B, _C,…,_Z,_AA).
- **Променливи етикети.** Stata етикети на променливи се конвертират в SPSS променливи етикети.
- **Етикети на стойности.** Stata етикети на променливи се конветират в SPSS етикети на стойности, освен Stata етикети на стойности отнасящи се към „разширени” липсващи стойност.
- **Липсващи стойности.** Stata ”разширени” липсващи стойности се конвертират в системно-липсващи стойности.
- **Превръщане на данни.** Stata формата на стойностите на данни се превръщат в SPSS Date формат (д-м-г-) стойности. Stata „времевите серии” във вид на дата (седмици, месеци, quarters, и т.н.) се конвертират в прост цифров формат (F), запазвайки оригиналната, вътрешна целочислена стойност, която е номерът на седмиците, месеците, quarters, etc., от началото на 1960.

Четене на Фаилове от база данни

Можете да четете данни от всеки формат на база данни, за който имате програма за база данни. В локален режим необходимите програми трябва да са инсталирани на Вашия компютър. За тези операции се изискват известни професионални умения и затова не ги описваме в това ръководство.

Глава 5. Редактор на данни

РЕДАКТОР НА ДАННИ (DATA EDITOR)

Редакторът на данни предоставя удобен метод за създаване и редакция на данни под формата на електронна таблица. Прозорецът Data Editor се отваря автоматично, когато започнете сесия (период на работа).

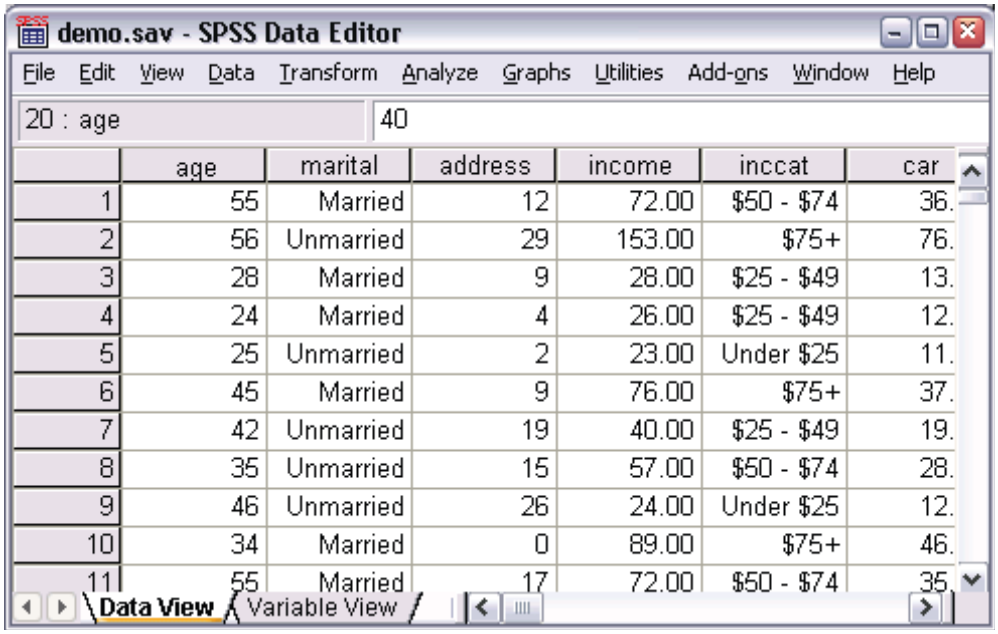
Редакторът на данни предоставя два изгледа на данните Ви:

Data View. Този изглед показва същинските стойности на данните или зададени етикети на стойностите.

Variable View. Този изглед показва информация за типа на данните, дефиниции на променливи и на стойности, тип на променливата (номинална, ординална или количествена) и стойностите за липсващите стойности.

И в двата изгледа могже да се добавя, променя и трие информация, съдържаща се във файла с данни.

Изглед на данните (Data View)



	age	marital	address	income	inccat	car
1	55	Married	12	72.00	$50 - $74	36.
2	56	Unmarried	29	153.00	$75+	76.
3	28	Married	9	28.00	$25 - $49	13.
4	24	Married	4	26.00	$25 - $49	12.
5	25	Unmarried	2	23.00	Under $25	11.
6	45	Married	9	76.00	$75+	37.
7	42	Unmarried	19	40.00	$25 - $49	19.
8	35	Unmarried	15	57.00	$50 - $74	28.
9	46	Unmarried	26	24.00	Under $25	12.
10	34	Married	0	89.00	$75+	46.
11	55	Married	17	72.00	$50 - $74	35.

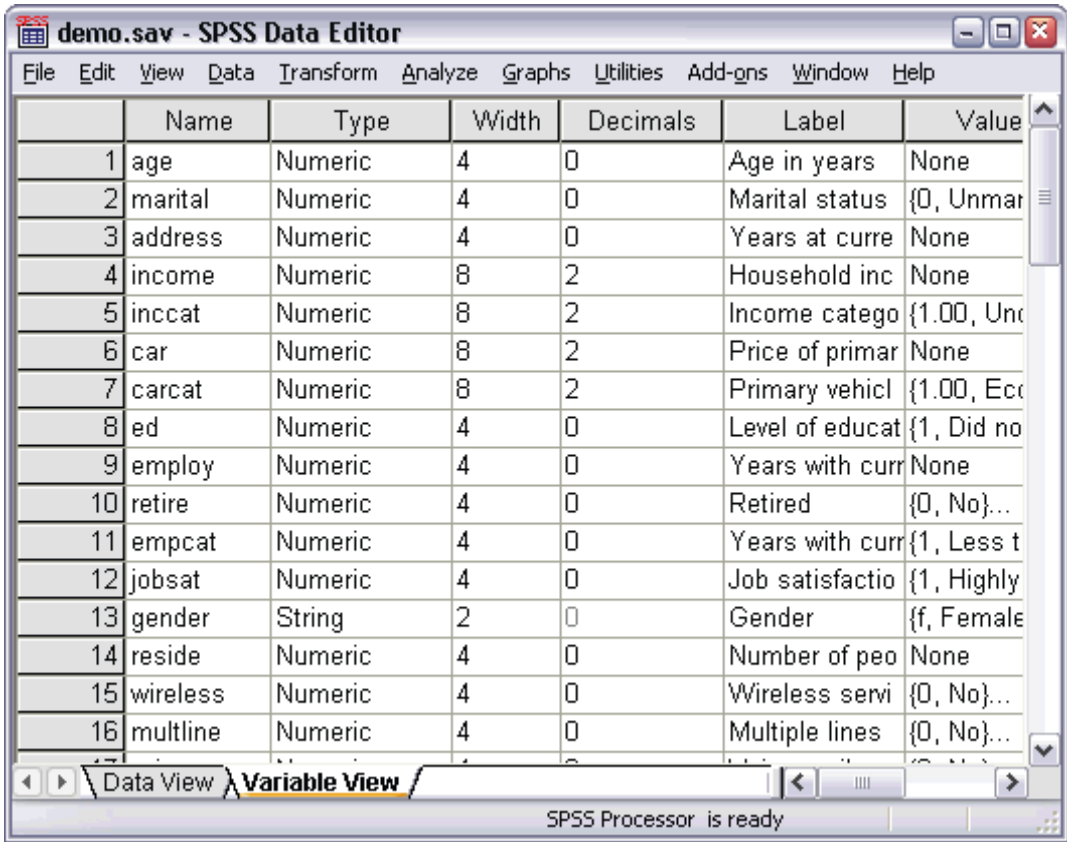
Фигура 5-1 Изглед на данните

Много от възможностите на Data Veiw са подобни на тези в програми за електронни таблици. Има няколко важни разлики:

- Редовете са случаи. Всеки ред е случай или наблюдение. Например всеки респондент отговорил на въпросник е случай.
- Колоните са променливи. Всяка колона е променлива или характеристика, която се измерва. Например всеки въпрос от въпросника е променлива.
- Клетките съдържат стойности. Всяка клетка съдържа една стойност на променлива за случай. Случай и променлива се пресичат в клетка. Клетките съдържат само стойности. За разлика от други програми за електронни таблици, Редакторът на данни не може да съдържа формули.
- Файлът с данни е правоъгълен. Размерът на файла с данни се определя от броя на случаите и променливите. Можете да въвеждате данни във всяка клетка. Ако въведете

данни в клетка извън границите на дефинирания файл с данни, правоъгълника с данни се разширява и вкючва редове и/или колони между клетката и границите на файла. Няма празни клетки в границите на файла с данни. За числови променливи празните клетки приемат стойността за системно определената стойност „липсва”. За променливи от тип низ се приемат низове с дължина нула – празни низове.

Изглед на променливите (Variable View)



	Name	Type	Width	Decimals	Label	Value
1	age	Numeric	4	0	Age in years	None
2	marital	Numeric	4	0	Marital status	{0, Unmar
3	address	Numeric	4	0	Years at curre	None
4	income	Numeric	8	2	Household inc	None
5	inccat	Numeric	8	2	Income catego	{1.00, Unc
6	car	Numeric	8	2	Price of primar	None
7	carcat	Numeric	8	2	Primary vehicl	{1.00, Ec
8	ed	Numeric	4	0	Level of educat	{1, Did no
9	employ	Numeric	4	0	Years with curr	None
10	retire	Numeric	4	0	Retired	{0, No}...
11	empcat	Numeric	4	0	Years with curr	{1, Less t
12	jobsat	Numeric	4	0	Job satisfactio	{1, Highly
13	gender	String	2	0	Gender	{f, Female
14	reside	Numeric	4	0	Number of peo	None
15	wireless	Numeric	4	0	Wireless servi	{0, No}...
16	multline	Numeric	4	0	Multiple lines	{0, No}...

Фигура 5-2 Изглед на променливите

Variable View Изгледа на променливите съдържа описания на атрибутите на всяка променлива във файла с данни. В изглед на променливите:

- Редовете са променливи.
- Колоните са атрибути на променливите.
  Можете да добавяте или триете променливи и да променяте атрибути като:
  - Име на променливата
  - Тип данни
  - Брой на цифрите или буквите
  - Брой на знаците след десетичната запетая
  - Описателни етикети на променливага и стойностите й
  - Потребителски зададени стойности „липсва”
  - Ширина на колоната
  - Ниво на измерване
  - Роля на променливата при анализа на данните.

Всички тези атрибути се запазват, когато запазите файла с данни.

Освен описания начин на дефиниране на свойствата на променливите има и други възможни подходи, например чрез използване на опцията Define Variable Properties в менюто Data на Data View. Този метод е особено удобен за категорийни променливи, които ползват числови кодове, за да представят категориите – например 0-мъж, 1-жена.

За да покажете или дефинирате атрибути на променливи

- Направете Data View данни активния прозорец.
- Натиснете два пъти името на променливата в горния край на колона или натиснете маркера за изглед на променливи.
- За да дефинирате нови променливи въведете име на променливата на празен ред в Variable View.
- Изберете атрибутите, които искате да дефинирате или промените.

Имена на променливи

Следните правила се ползват за имена на променливи:

- Всяко име трябва да е уникално; не се позволяват повторения.
- Имената на променливи могат да са дълги до 64 байта и първият знак трябва да е буква или един от знаците @, #, $. Следващите знаци може да са всяка комбинация от букви, цифри, точка (.) и непунктуационни знаци.
- Имена на променливи не могат да съдържат празно място (шпация).
- Знак # в началото на име на променливата дефинира scratch променлива. Можете да създавате scratch променливи само със синтаксис за команди. Не можете да ползвате # като първи знак на променлива в диалогов прозорец, който създава нови променливи.
- Знак $ в началото означава, че променливата е системна променлива. Знак $ не е позволен в началото на потребителски дефинирани променливи.
- Точка, долна черта и $, #, and @ могат да се ползват вътре в имената на променливи. Например *A._$@#1* е валидно име на променлива.
- Имена на променливи, завършващи с точка трябва да се избягват, защото точката може да се интерпретира като като завършваща команда. Можете да създавате променливи, които завършват с точка със синтаксис за команди. Не можете да създавате променливи, които завършват с точка в диалогови прозорци, които създават нови променливи.
- Имена на променливи, които завършват с долна черта трябва да се избягват, защото такива имена могат да са в конфликт с имена на променливи, създадени автоматично от команди и процедури.
- Резервирани думи не могат да се ползват като имена на променливи. Резервирани думи са: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO и WITH.
- Имена на променливи могат да се дефинират с всяка съвкупност от голем и малки букви, големината се запазва за четливост.
- Когато големи имена на променливи трябва да се пренесат на няколко реда, SPSS се опитва да ги прекъсва на долни черти, запетаи и където съдържанието се променя от малки на големи букви.

Тип на скалата на измерване на променливата

Можете да зададете нивото на измерване като количествено (числови данни върху интервална или пропорцифнална скала), ординално или номинално. Номинални и ординални данни могат да са низове или числа.

- **Номинална скала.** Променлива може да се счита третира като номинална, когато стойностите й представляват категории без ранк (например департамента в

компанията, в който работи служител). Примери на номинални променливи са региони, зип кодове и религиозна принадлежност.

- **Ординална скала.** Променлива се счита за ординална, когато стойностите й представляват категории с ранк (например нива на задоволство от обслужване, от много недоволни до много доволни). Примери на ординални променливи са също attitude scores отразяващи ниво на задоволство или сигурност, а също и рейтинг на предпочитания.
- **Метрична скала.** Променлива може да се счита за количествена, когато стойностите й представляват подредени категории със значеща метрика, така че сравненията на разстояния са добре дефинирани. Примери са възрастта в години и доход в хиляди долари.

Забележка: За ординални низови променливи азбучното подреждане на низовите стойности се счита, че отразява истинската подредба на категориите. Например за низова променлива със стойности *ниско*, *средно, високо*, подредбата на категориите се интерпретира като *високо, ниско*, *средно* - което не е правилната подредба. По принцип е по надеждно да се ползват числови кодове за представянето на ординални данни.

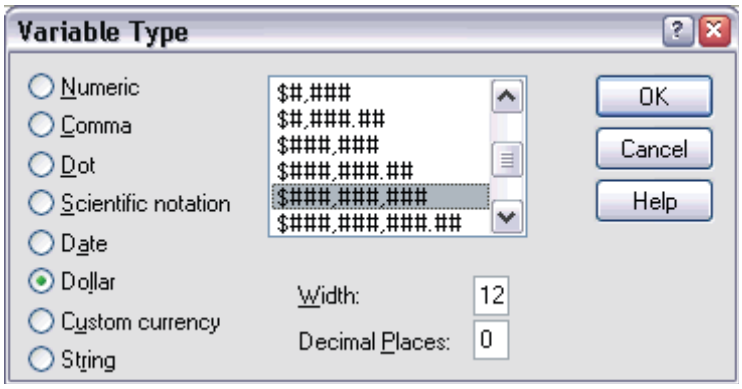
За нови променливи, създадени в процес на работа, данни прочетени от външни файлови формати и SPSS файлове с данни, създадени преди версия 8.0, присвояването на ниво на измерането по подразбиране се определя от следните правила:

- Числови променливи с по-малко от 24 уникални стойности и низови променливи се считат за номинални.
- Числови променливи с 24 или повече уникални стойности се считат за количествени.

Диалоговият прозорец, чрез който се дефинират свойства на променливите от Data View може да Ви помогне да изберете правилното ниво на измерване.

Тип на променливата

Тип на променливата определя типа данни на променливата за всяка променлива. По подразбиране всички нови променливи са числови. Можете да ползвате тип на променливата, за да промените типа данни. Съдържанието на диалоговия прозорец зависи от избрания тип данни. За някои типове данни има тектова кутия за настройки на ширина и брой на цифри след десетичната запетая. За други типове данни можете да изберете формата от листово меню със скролер.



Фигура 5-3 Диалогов прозорец за типа на променливата

Има следните типове данни:

- **Числов.** Променливи, чиито стойности са числа. Стойностите се представят в стандартен числов формат. Редакторът на данни приема числови стойности в стандартена или научна(scientific) нотация.
- **Запетая.** Числови променливи, чиито стойности се показват със запетаи на всеки три знака и точка за десетичен разделител. Редакторът на данни приема числови стойности за променливи със запетаи със или без запетаите или в научна нотация. Стойностите не могат да съдържат запетаи от дясно на десетичния разделител.
- **Точки.** Числови променливи, чиито стойности са разделени с точки на всеки три знака и запетая за десетичен разделител. Редакторът на данни приема числови стойности за променливи с точки с или без точките или в научна нотация. Стойностите на могат да съдържат точки от дясната страна на десетичния разделител.
- **Научна нотация.** Числова променлива, чиито стойности съдържат E е степен на десет със знак. Редакторът на данни приема числови стойности за такива променливи с или без степен. Степенният показател може да е след E или D с незадължителен знак или само със знак. Например 123, 1.23E2, 1.23D2, 1.23E+2, and 1.23+2.
- **Дата.** Числова променлива, чиито стойности се показват като един от няколко календарни и времеви формати. Изберете формат от списъка. Можете да въвеждате дати с наклонени черти, тирета, точки, запетаи или празни места за разделители. Обхватът на вековете за години с две цифри се определя от Вашите настройки (от меню редакция изберете настройки и кликнете етикета данни).
- **Долар.** Числова стойност, която се показва със знак долар отпред ($), запетаи, които разделят всеки три знака и точка като десетичен разделител. Можете да въвеждате данните с или без знак долар.
- **Ръчно зададена валута.** Числова променлива, чиито стойности се показват в ръчно зададени формати , които сте дефинирали в етикета валута в диалогов прозорец настройки. Дефинирани ръчно символи за валута не могат да се ползват при въвеждането на данни, но се показват в редактор на данни.
- **Низ.** Променлива, чиито стойности не са числови и следователно не се ползват в изчисления. Стойностите могат да съдържат каквито и да е знаци до позволената дължина. Прави се разлика между главни и малки букви.

За да дефинирате тип на променлива

- Натиснете бутона в клетката Type в прозореца Variable View за променливата, която искате да дефинирате.
- Изберете типа данни в диалоговия прозорец.
- Натиснете OK.

Как се въвеждат и как се показват данни

В зависимост от формата, данните в Data View може да се показват различно от начина, по който са въведени и съхранявани вътрешно. Това са някои основни насоки:

- За числови, запетая и точка формати можете да въведете стойност с какъвто и да е брой знаци след десетичния разделител (до 16) и цялата стойност се запазва вътрешно. Изглед на данни показва само дефинирания брой знаци след десетичната запетая и закръгля стойности с повече знаци. Въпреки това цялата стойност се ползва при всички изчисления.

- За низови променливи, всички стойности се допълват отдясно до максималната ширина. За низова променлива с максимална ширина три, стойност *Не* се запазва вътрешно като ‘Не ‘ и не е еквивалентно на ‘ Не’.
- За времеви формати можете да ползвате наклонени черти, тирета, празни места, запетаи или точки като разделители между стойностите за ден, месец и година и може да въвеждате числа, трибуквени съкращения или целите имена за месеци. Дати от обичайния формат *дд-ммм-гг* се изписват с тирета за разделители и трибуквени съкращения за месеца. Дати от обичайния формат *дд/мм/гг* и *мм/дд/гг* се изписват с наклонени черти за разделители и числа за месец. Вътрешно датите се пазят като брой секунди от 14 октомври 1582 г. Обхвата на века за дати с двуцифрови години се определя от настройките (меню редактиране, опции, етикет данни).
- За времеви формати можете да ползвате двуеточия, точка или празно място за разделители между часове, минути и секунди. Времената се показват с двуеточия за разделители. Вътрешно времената се съхраняват като брой секунди, които означават времеви интервал. Например 10:00:00 се пази като 36000, което е 60 (секунди в минутата) x 60 (минути на час) x 10 (часа).

Етикети на променливи

Можете да поставяте описателни шаблони на променливите до 256 знака. Етикетите на променливи могат да съдържат празни места и резервирани знаци, които иначе не са позволени в имена на променливите.

За да поставите етикет на променлива

- Направете Редактора на данни активен прозорец.
- Натиснете два пъти името на променливата отгоре на колоната в Data View или изберете Variable View.
- В клетката Values въведете описателния етикет на променливата.

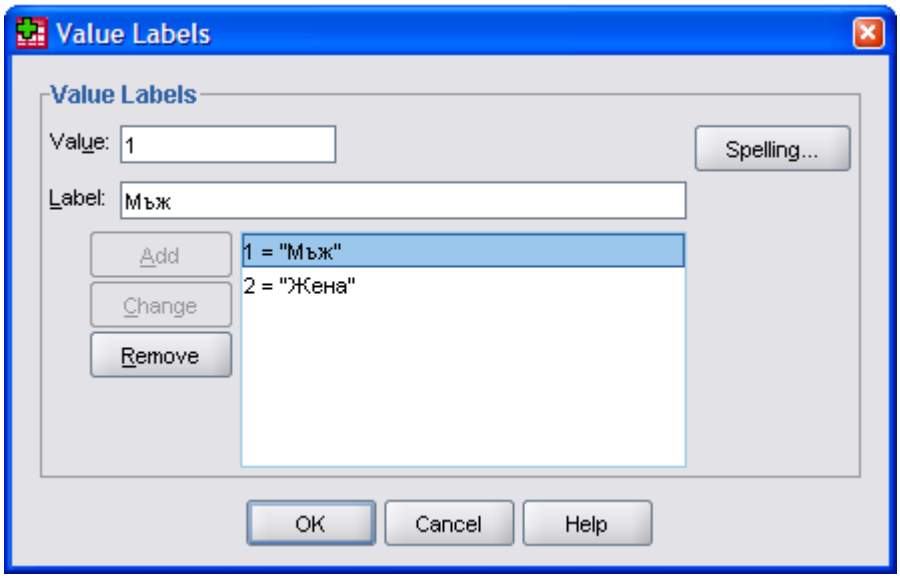
Етикети на стойности

Можете да поставяте описателни етикети за всяка стойност (клас) на неметрична променлива. Този процес е особено полезен, ако файловете с данни използват цифрови кодове за да означават нечислови категории (например кодове 1 и 2 за *мъж* и *жена*).

- Етикетите на стойности могат да са до 120 байта (символа).
- Етикети на стойностите не могат да се поставят за дълги низови променливи (низови променливи с над 80 знака).

За да се изберат етикети на стойности

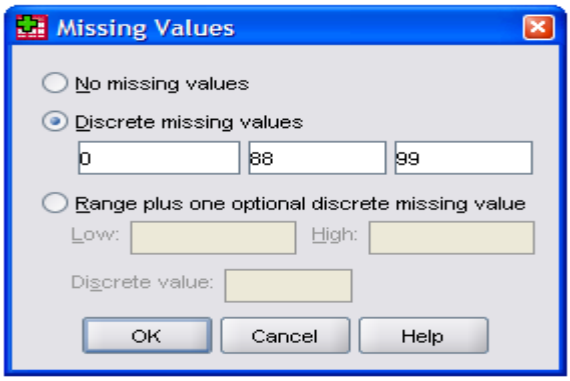
- Натиснете бутона в клетката Values на Variable Veiw за променливата, която искате да дефинирате.
- За всеки код въведете стойност и етикет.
- Натиснете Add за да въведете етикета на стойност.
- Натиснете OK.



Фигура 5-4 Диалогов прозорец етикети на стойности

Липсващи стойности

Missing Values дефинира стойности на данни, определени от потребителя като липсващи. Например, може да искате да различавате данни, които липсват, защото респондентът е отказал да даде отговор и данни, които не са приложими за този респондент. Стойности на данни, които са опеределени като липсващи се третират специално и са изключени от повечето изчисления.



Фигура 5-5 Диалогов прозорец липсващи данни

Можете да въведете до три дискретни (индивидуални) липсващи стойности, обхват от липсващи стойности, или обхват плюс една дискретна стойност.

- Обхвати могат да се определят само за числови променливи.
- Не можете да дефинирате липсващи стойности за дълги низови променливи (низови променливи, които са по–дълги от 80 знака).

Липсващи стойности за низови променливи. Всички низови стойности, включително празни или нулеви, се считат за валидни, освен ако не ги дефинирате като липсващи експлицитно. За да дефинирате празни или нулеви стойности за липсващи, въведете една шпация в едно от полетата в избор на дискретни липсващи стойности.

За да дефинирате липсващи стойности

- Натиснете бутона в клетката Missing на Variable View за променливата, която искате да дефинирате.
- Въведете стойностите или обхвата от стойности, които представят липсващите данни.

Ширина на колоната.

Можете да определите брой знаци за ширина на колоната. Ширината на колоната може да бъде променена и в изглед на данни като натиснете и влачите границите на колоната. Формата на колоната има ефект само на това как стойностите се показват в редактор на данни. Промяна на ширината на колоната не променя дефинираната ширина на променливата. Ако дефинираната и истинската ширина на стойност са по–широки от колоната се показват звезди (*) в изглед на данни.

Подравняване на променливата

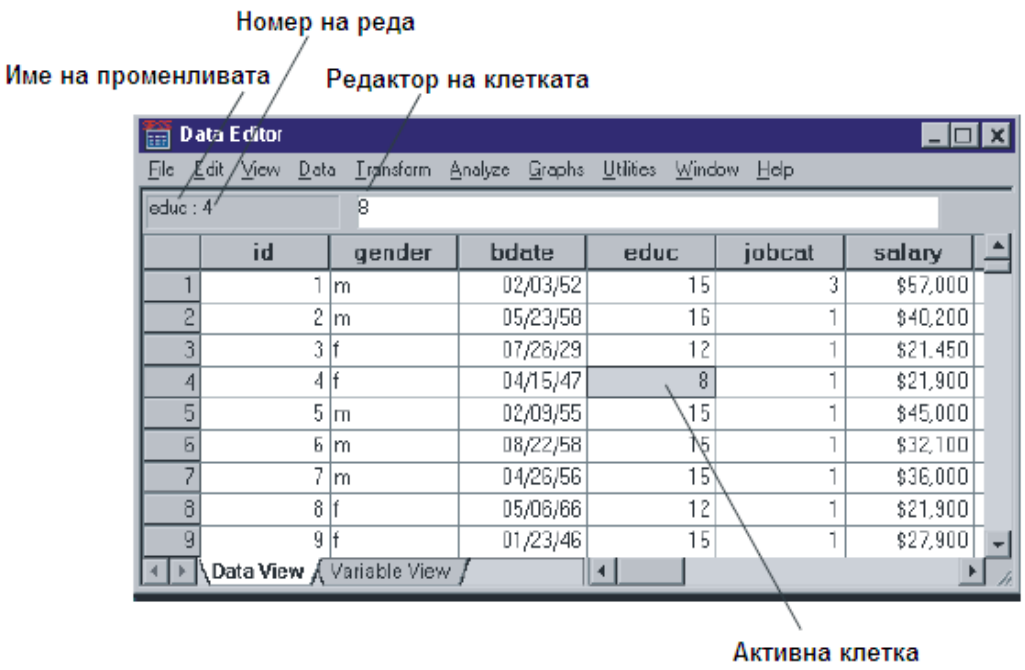
Подравняването определя как ще се показват стойности данни и/или етикети на стойности в изглед на данни. По подразбиране подравняването е дясно за числови и ляво за низови променливи. Тази настройка има ефект само за прозореца Data View.

Въвеждане на данни

В Data View можете да въвеждате данни директно в редактора на данни. Можете да въвеждате данни в какъвто и да е ред. Можете да въвеждате данни по случай или по променлива, за избрани райони или за индивидуални клетки.

- Активната клетка променя цвета си.
- Името и номера на реда на променливата на активната клетка се показват в горния ляв ъгъл на редактор на данни.
- Когато изберете клетка и въведете стойност, тази стойност се показва в редактора на клетка горе в Редактора на данни.
- Стойности данни не се записват докато не натиснете Enter или не изберете друга клетка.
- За да въведете данни, различни от числови, първо трябва да дефинирате типа данни.

Ако въведете стойност в празна колона, Редакторът на данни автоматично създава нова променлива и й дава име.



Фиг. 5-10 Активно множество данни в изглед на данни

За да въведете числови данни

- Изберете клетка в Data View.
- Въведете стойност. (Стойността се показва в редактора на клетки горе в Редактора на данни.)
- За да запишете стойността, натиснете Enter или изберете друга клетка.

За да въведете нечислови данни

- Натиснете два пъти име на променлива горе на колона на редактор на данни или натиснете етикета изглед на променливи.
- Натиснете бутона в клетката тип за променливата.
- Изберете типа данни в диалоговия прозорец тип на променливата.
- Натиснете OK.
- Натиснете два пъти номера на реда или натиснете етикета изглед на данни.
- Въведете данните в колоната за новодефинираната променлива.

Ограничения на стойности данни в редактор на данни

Дефинираният тип на променливата и дължината(ширината) определят типа стойност, която може да бъде въведена в клетката в изглед на данни.

- Ако въведете знак, който не е позволен от дефинирания тип на променливата, Редакторът на данни издава звук и не въвежда знака.
- За низови променливи не са позволени знаци извън дефинираната дължина.
- За числови променливи, целочислени стойности, които надвишават дефинираната дължина могат да се въвеждат, но Редакторът на данни ги показва или в научна нотация или със звезди в клетката, за да покаже, че стойността е по-широка от дефинираната.
- За да покажете стойността в клетката, променете дефинираната дължина на променливата.

Забележка: Като промените ширината на клетката не променяте дължината на променливата.)

Редактиране на данни

С редактора на данни можете да редактирате данни по много начини. Можете да:

- Променяте стойности данни.
- Изрязвате, копирате и поставяте стойности данни.
- Добавяте и триете случаи.
- Добавяте и триете променливи.
- Променяте реда на променливите.

Преобразуване на числови типове или дати в низ. Числови типове (например числа, долар, разделени с точка, разделени със запетаи) и формати на дати се пробразуват в низове ако се поставят в клетка на низова променлива. Низовата стойност е числовата стойност, както се показва в клетката. Например за променлива от тип долар, показаният знак за долар става част от низовата стойност. Стойности, които надвишават дефинираната ширина на променливата се изрязват.

Преобразуване на низ в числов тип или дата. Низови стойности, които съдържат приемливи знаци за числов тип или формат за дата за клетката приемник се преобразуват в еквивалентната числова или стойност за дата. Например низовата стойност 25/12/91 се преобразува във валидна дата, ако типа формат на клетката приемник е един от форматите за ден-месец-година, но стойността се преобразува в стойност ‘системно липсва’, ако типа на формата на клетката приемник е един от форматите месец-ден-година.

Преобразуване на дата в числов тип. Стойности на дати и времена се пробразуват в брой секунди, ако клетката приемник е един от числовите формати (например числов тип, долар, разделен с точка или запетая). Тъй като датите вътрешно се пазят като брой секунди от 14 окт. 1582, преобразуване на датите може да Ви да де някои особено големи числа. Например датата 10/29/91 се преобразува в числовата стойност 12,908,073,600.

Преобразуване на числа в дата или време. Числови стойности се преобразуват в дати и времена, ако стойноста представя брой секунди, които могат да произведе валидна дата или време. За дати, числови стойности по–малки от 86,400 се превръщат в стойността ‘системно липсва’.

Добавяне на нови случаи

Въвеждане на данни в клетка от празен ред автоматично създава нов случай. Редакторът на данни добавя стойност ‘системно липсва’ за всички други променливи за този случай. Ако има празни редове между новия случай и съществуващите случаи, празните редове стават нови случаи със стойности ‘системно липсва’ за всички променливи. Също можете да въвеждате нови случаи между съществуващи случаи.

За да въведете нови случаи между съществуващи случаи

Е В изглед на данни изберете която и да е клетка в случая (ред) под позицията, където искате да добавите нов случай.

Е От менютата изберете:

Редактирай

Добави нови случаи

Добавя се нов ред за случая и всички променливи приемат стойност ‘системно липсва’.

Добавяне на нови променливи

Въвеждане на данни в празна колона в изглед на данни или в празен ред в изглед на променливите автоматично създава нова впроменлива с име на променливата по подразбиране (префикс вар и пореден номер) и тип данни по подразбиране (числов).

Редакторът на данни поставя стойност ‘системно липсва’ за всички случаи за новата променлива. Ако има празни колони в изглед на данни ли празни редове в изглед на променливите между новата променлива и съществуващи променливи, тези редоев или колони също стават нови променливи със стойност ‘системно липсва’ за всички случаи.

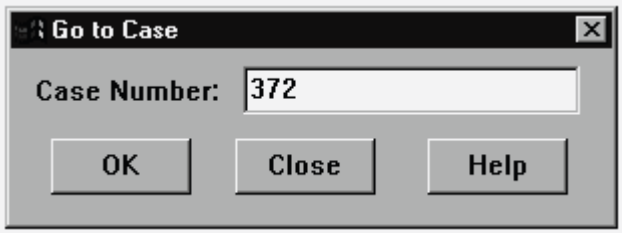
Можете да добавяте нови променливи между съществуващи променливи, както и да местите променливи.

За да смените типа данни

Можете да променяте типа данни за променлива по всяко време, като ползвате диалогов прозорец тип на променливата в изглед на променливите. Редакторът на данни ще се опита да преобразува съществуващите стойности към новия тип. Ако не успее, се присвоява стойност ‘системно липсва’. Правилата на преобразуване са същите като правилата за пренасяне на стойности данни към променлива с друг тип формат. Ако промяната на формата данни може да доведе до загуба на спецификации за изгубени стойности или етикети на стойности, Редакторът на данни показва алармиращ прозорец и пита дали искате да продължите или да отмените операцията.

Отиди на случай

Диалогов прозорец отиди на случай Ви позволява да отидете на указания номер на случай (ред) в редактор на данни.



Фигура 5-11 Диалогов прозорец отиди на случай

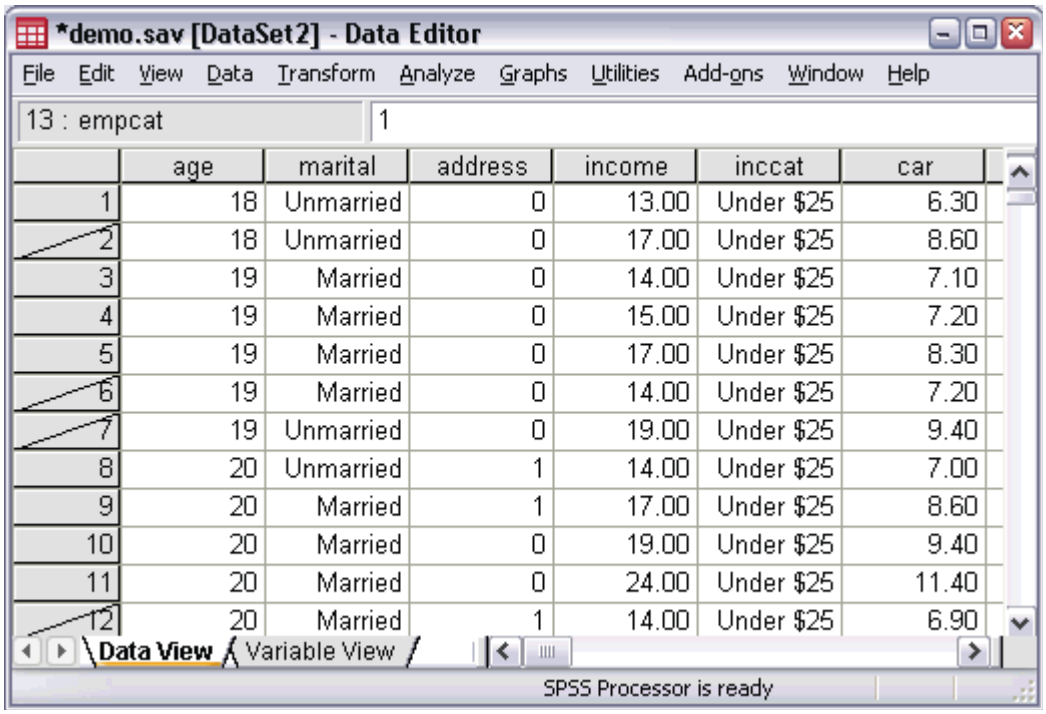
За да откриете случай в редактор на данни

- Направете Редактора на данни активния прозорец.
- От менютата изберете:
  Edit
    Go to Case...
- В диалогов прозорец Go to Case въведете номера за случай от редактора на данни.

По същия начин може да се отиде и на променлива.

Състояние на избраните случаи в Редактора на данни

Ако сте избрали подмножество случаи, но не сте махнали неизбраните случаи, неизбраните случаи се маркират в Редактора на данни с диагонална линия (наклонена черта) през номера на реда.



	age	marital	address	income	inccat	car
1	18	Unmarried	0	13.00	Under $25	6.30
2	18	Unmarried	0	17.00	Under $25	8.60
3	19	Married	0	14.00	Under $25	7.10
4	19	Married	0	15.00	Under $25	7.20
5	19	Married	0	17.00	Under $25	8.30
6	19	Married	0	14.00	Under $25	7.20
7	19	Unmarried	0	19.00	Under $25	9.40
8	20	Unmarried	1	14.00	Under $25	7.00
9	20	Married	1	17.00	Under $25	8.60
10	20	Married	0	19.00	Under $25	9.40
11	20	Married	0	24.00	Under $25	11.40
12	20	Married	1	14.00	Under $25	6.90

Фигура 5-12 Филтрирани случаи в редактор на данни

Работа с няколко източника на данни

Започвайки с версия 14.0, SPSS може да има няколко източника на данни, отворени едновременно, като това прави по-лесно да:

- Превключвате между източници на данни.
- Сравнявате съдържнието на различни източници на данни.
- Копирате и пренасяте данни между източници на данни.
- Създавате няколко подмножества случаи и/или променливи за анализ.
- Сливането на няколко източника от различни формати на данни (например работен лист, база данни, текстови данни) без да запазвате всеки източник първо във SPSS формат.

Глава 6. Подготовка на данните

ПОДГОТОВКА НА ДАННИТЕ

След като сте отворили файл или въвели данни в редактор на данни, можете да започнете да създавате доклади, диаграми и анализи без никаква друга предварителна работа.

Има обаче някои допълнителни възможности за подготовка на данните, които могат да Ви се сторят интересни, включително възможността да:

- Присвоявате свойства на променливите, които описват данните и определят как някои стойности ще се третират.
- Идентифицирате случаи, които може да съдържат повтаряща се информация и да изключите тези случаи от анализите или да ги изтриете от файла.
- Създавате нови променливи с няколко различни категории, които представят обхвати от стойности с голям брой възможни стойности.

Свойства на променливи

Данни въведени в Редактор на данни в изглед на данни или прочетени в SPSS от външен файлов формат (като таблица от EXEL или текстов файл с данн) нямат някои свойства на променлвите, които могат да са от полза, като:

- Дефинирани описателни етикети на стойности за числови кодове (например 0 – мъж и 1 – жена).
- Идентификация на кодове за липсващи стойности (например 99 за липса на отговор или невъзможност за отговор).
- Присвояване на ниво/тип на измерване (nominal, ordinal, or scale).

Всички тези свойства на променливи (и други) могат да се присвоят в прозорец Variable View изглед на променливи в редактор на данни (Data View). Има няколко инструмента, които могат да Ви помогнат в този процес:

- **Дефиниране на свойства на променливи** може да Ви помогне да дефинирате описателни етикети на стойности и липсващи стойности. Това е особено полезно за неметрични данни, които ползват числови кодове за означаване на категориите.
- **Копиране на свойства на данни** дава възможността да се използва вече съществуващ файл с данни в SPSS формат като шаблон за формата и атрибутите на променливи в работния файл. Това е особено полезно, ако често ползвате външни файлови формати на данни, които имат подобно съдържание (като месечни доклади в екселки формат).

Дефиниране на свойства на променливи

Дефиниране на свойства на променливи е предназначено да помогне в процеса на създаване на описателни етикети на стойности за неметрични (номинални, ординални) променливи. Дефиниране на свойства на променливи:

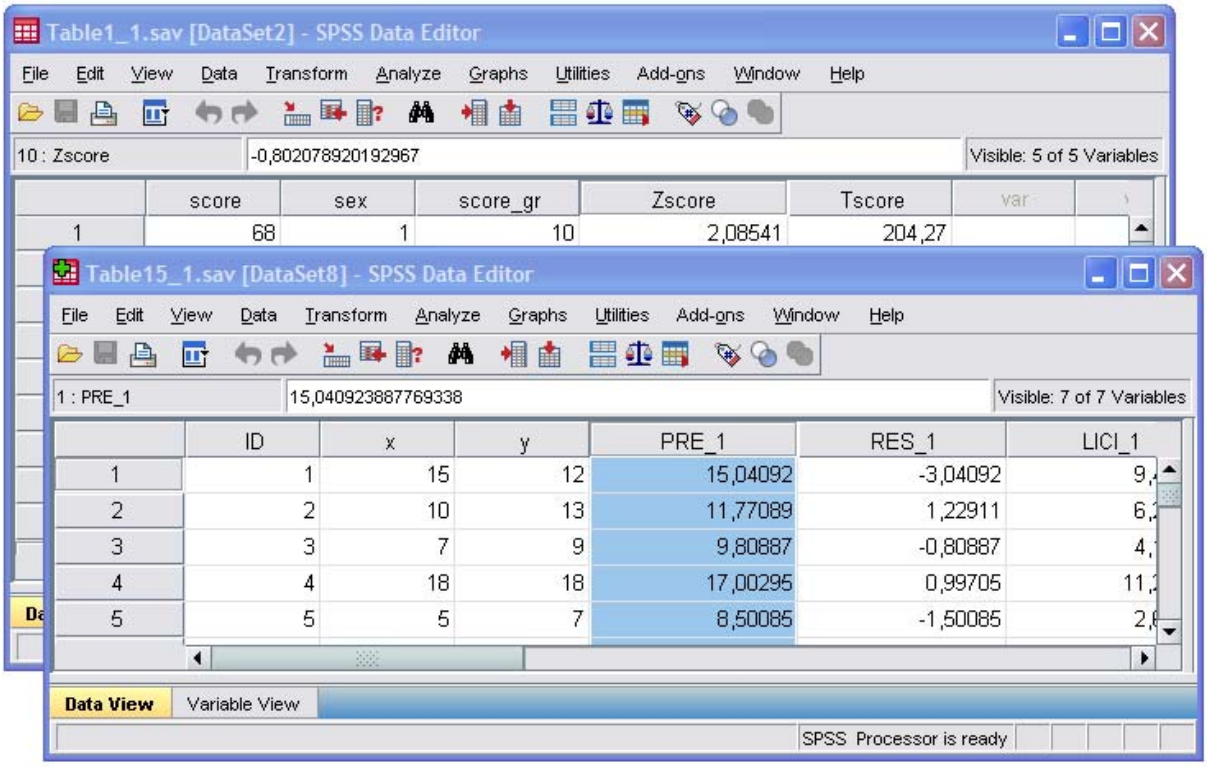
- Сканира същинските стойности на данните и дава лист с всички уникални стойности данни за всяка избрана променлива.
- Идентифицира стойности без етикет и предоставя функция за поставяне на етикети автоматично.
- Дава възможност за копиране на дефинирани етикети на стойности от една променлива в избраната променлива или от избраната променлива към няколко допълнителни променливи.

Глава 7. Работа с множество източници на данни

Започвайки с версия 14.0, няколко източника на данни могат да са отворени едновременно, което улеснява:

- Превключване между източници на данни.
- Сравняване съдържнието на различни източници на данни.
- Копиране и пренасяне на данни между източници на данни.
- Създаване на няколко подмножества случаи и/или променливи за анализ.
- Сливане на няколко източника от различни формати на данни (например работен лист, база данни, текстови данни) без да запазвате всеки източник първо във SPSS формат.

Манипулации с множество източници на данни

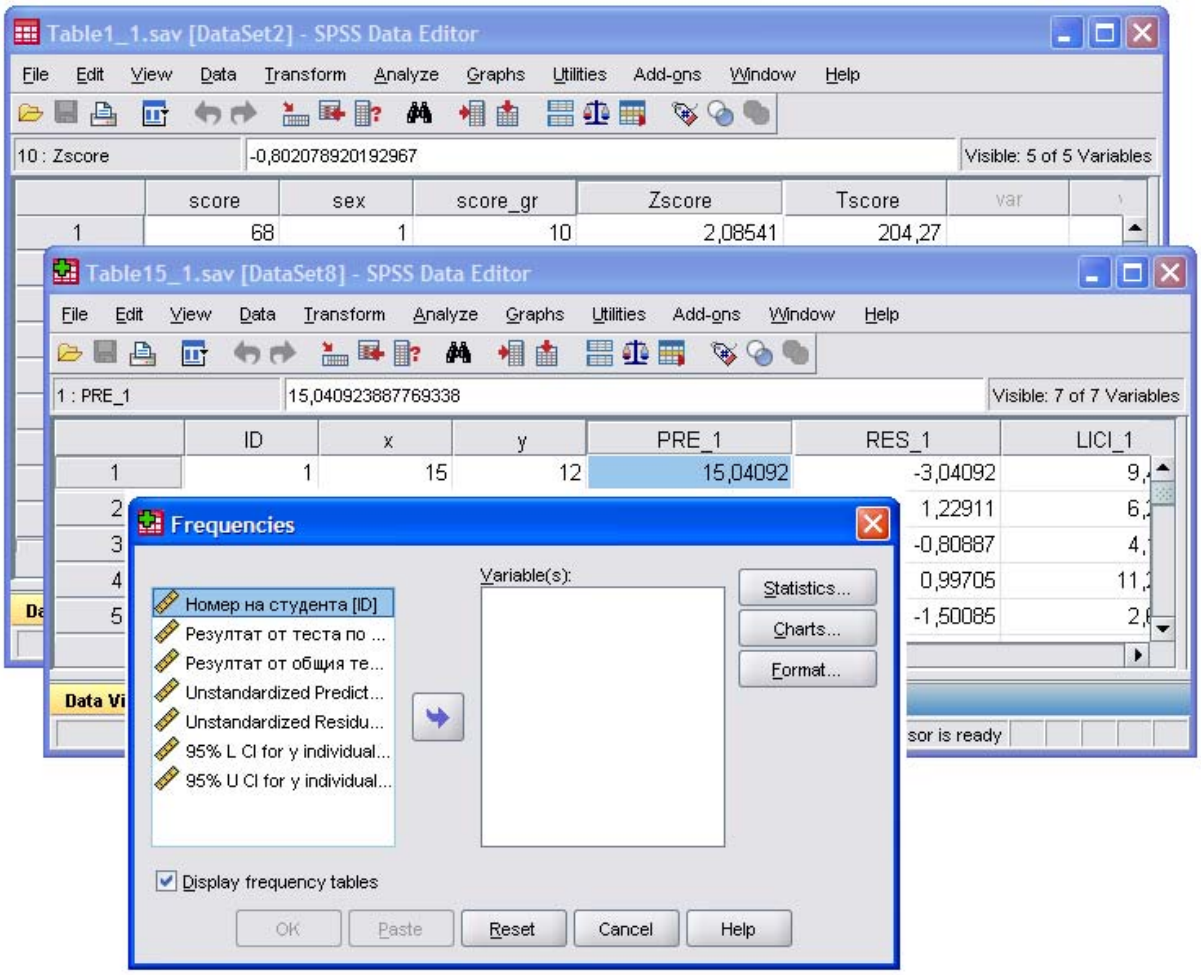


Фигура 6-1 Два източника на данни, отворени едновременно

По подразбиране всеки източник на данни, който отворите, се показва в нов прозорец на редактор на данни.

- Предварително отворени източници на данни остават отворени и може да се работи с тях.
- Когато за първи път отворите източник на данни, той автоматично става активното множество данни.
- Можете да промените активното множество данни като натиснете където и да е в прозореца на редактор на данни на източника, който искате да ползвате, или като изберете прозореца на редактор на данни за този източник от менюто прозорец.
- Само променливите в активното множество данни са на разположение за анализ.

- Не можете да променяте активното множество данни, докато някакъв диалогов прозорец, който има достъп до данните, е отворен (включително всички диалогови прозорци, които показват лист с променливите).
- Поне един прозорец на редактор на данни трябва да остане отворен по време на сесия. Когато затворите последния отворен прозорец на редактор на данни, SPSS автоматично излиза, като преди това Ви пита дали искате да запазите промените.



Фигура 6-2 *Лист променливи, който съдържа променливите в активното множество данни*

Копиране и пренасяне на информация между множества данни

Можете да копирате както данни, така и атрибути на променливи от едно множество данни в друго почти по същия начин, по който копирате и пренасяте информация вътре в един файл.

- Копиране и пренасяне на избрани клетки в изглед на данни пренася само стойности данни без атрибути на променливи.
- Копиране и пренасяне на цяла променлива в изглед на данни с избиране на името на променливата от горната част на колоната пренася всички данни и атрибути на променливи за тази променлива.

- Копиране и поставяне на атрибути на променлива или на цели променливи в изглед на променливите поставя избраните атрибути (или цялата дефиниция на променливата), но не пренася никакви стойности данни.

Преименуване на множества данни

Когате отворите източник на данни чрез менютата и диалоговите прозорци, всеки източник на данни получава име по подразбиране от типа *DataSetn*, където ***n*** е последователно цяло число. За да предоставите по-описателни имена на множества данни:
Е От менютата в прозореца на Редактора на данни, на който искате да смените името изберете:

- File
  Rename dataSet...
- Въведете ново име на множество данни, което отговаря на правилата за имена на променливи.

Забрана на едновременна работа с няколко множества данни

Ако предпочитате само едно множество данни и искате да премахнете функцията за няколко множества наведнъж:

- От менютата изберете:
  Edit
  Options ...
- Натиснете General
  Изберете Open only one dataset at a time.

Глава 8. Преобразуване (трансформация) на данните

Преобразуване на данни

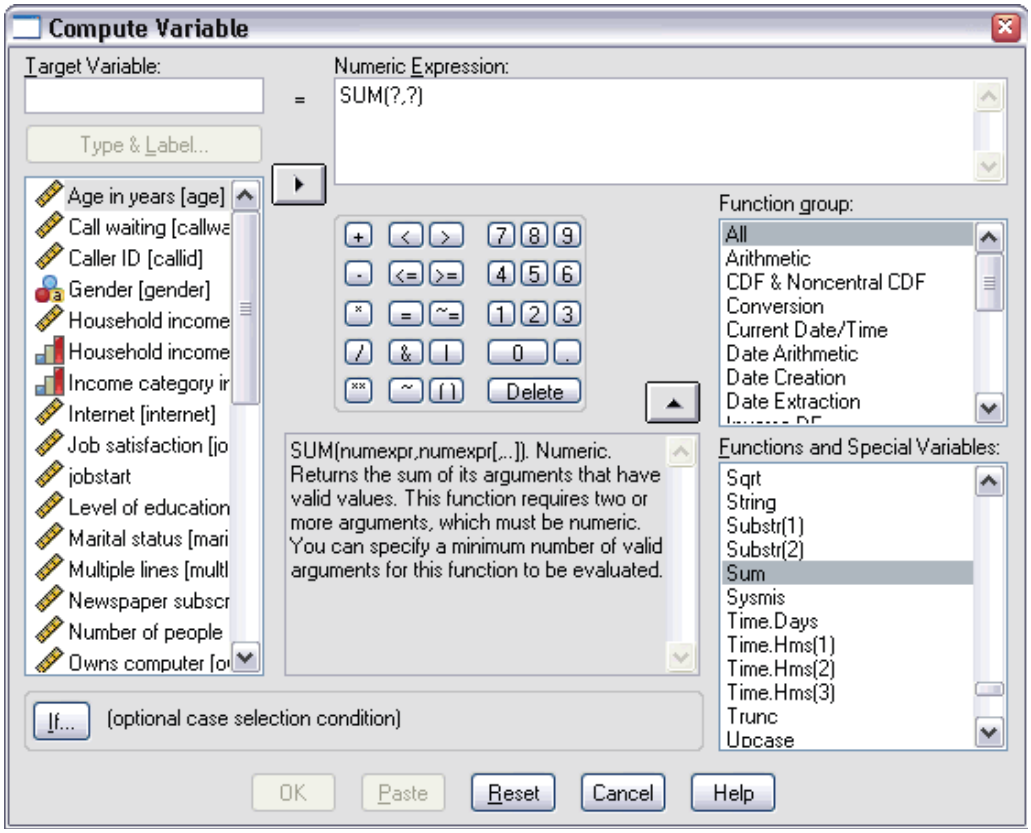
В идеалния случай, вашите необработени данни са напълно подходящи за вида на анализ, който искате да изпълните и всички отношения между променливите са или линейни или ортогонални. За съжаление случаят рядко е такъв. Предварителният анализ може да покаже необедителни схеми за кодиране или грешки при кодирането, или може да бъде изискана промяна на данни, за да се покаже истинската връзка между променливите.

Можете да извършите промяна на данни, вариращи от прости задачи, като например „*collapsing categories for analysis*", до по-сложни, като създаването на нови променливи въз основа на сложни уравнения и условни изрази.

Пресмятане на стойности

Използвайте „*Compute*" диалоговия прозорец за изчисляване на стойностите за дадена променлива, въз основа на числови преобразувания на други променливи.

- Можете да изчислите стойностите за числови или низови (буквено-цифрови) променливи.
- Можете да създавате нови променливи или да подменяте стойностите на съществуващите променливи. За нови променливи, можете също да задавате вида и етикета им.
- Можете да изчислите стойностите избирателно за подгрупи от данни, въз основа на логически условия.
- Можете да използвате повече от 70 вградени функции, включително аритметични функции, статистически функции, разпределитени функции, както и низови функции.



Фигура 8-1 „Compute Variable" диалоговия прозорец

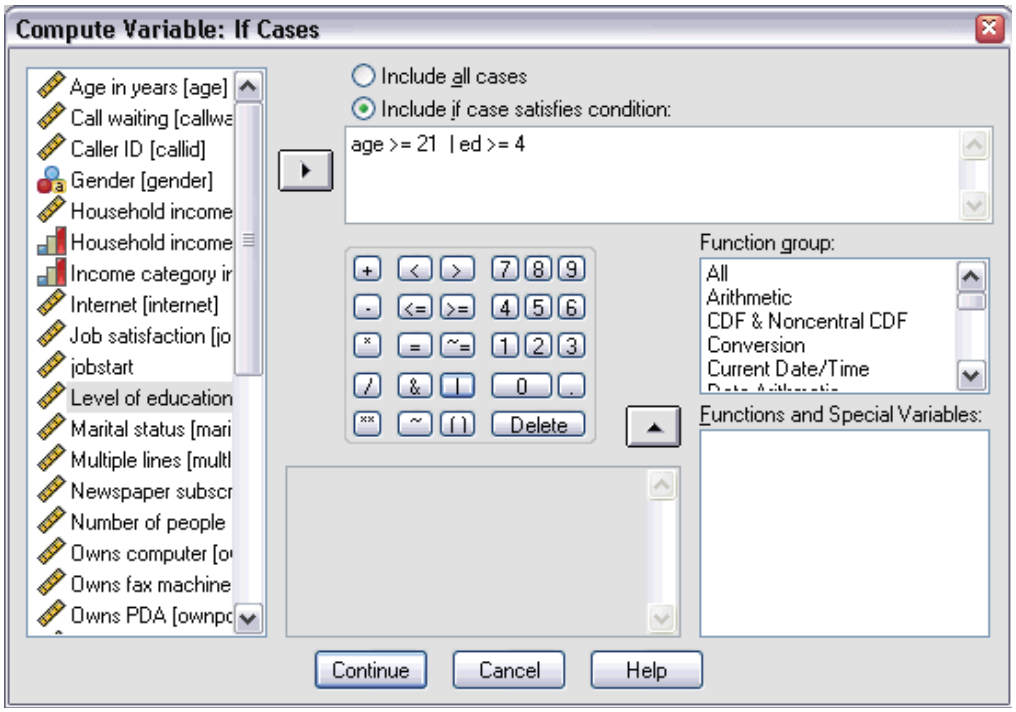
За пресмятане на променливите

- Изберете от менюто:
  „Transform”
  „Compute Variable...”
- Въведете името на желаната променлива. Тя може да е вече съществуваща или нова, която бива добавена в активната база данни.
- В „*Expression field*” създайте израза или вмъкнете компонентите директно.

- Можете да вмъкнете функции или често използвани системни променливи, като изберете групата от „*Function group*” списъка и щракнете двукратно върху функция или променлива във „*Functions*” и „*Special variables*” списъка (или изберете функция или променлива и щракнете върху стрелката в близост до „*Function group*” списъка). Попълнете всички параметри, отбелязани от въпросителен знак (важи само за функции). Функционалната група обозначена с „*All*” предлага показването на списък на всички налични функции и системни променливи. Кратко описание на текущо избраната функция или променлива е показано в отделен сектор в диалоговия прозорец.
- Низовите константи трябва да бъдат поставени в кавички или апострофи.
- Ако стойностите съдържат дробна част, точката (.) трябва да се използва като десетичен индикатор.
- За нови низови променливи, трябва да изберете „*Type & Label*”, за да определите типа данни.

„Compute Variables: If Cases”

„*If cases*” диалоговия прозорец ви позволява да се филтрират данни и да се избрат подмножества от случаите, използвайки условни изрази. Условният израз връща стойност „*true, false*” или „*missing*”.



Фигура 8-2 „Compute Variable If Cases” диалоговия прозорец

Ако резултатът на условния израз е „*true*”, случаят е включен в избраното подмножество.

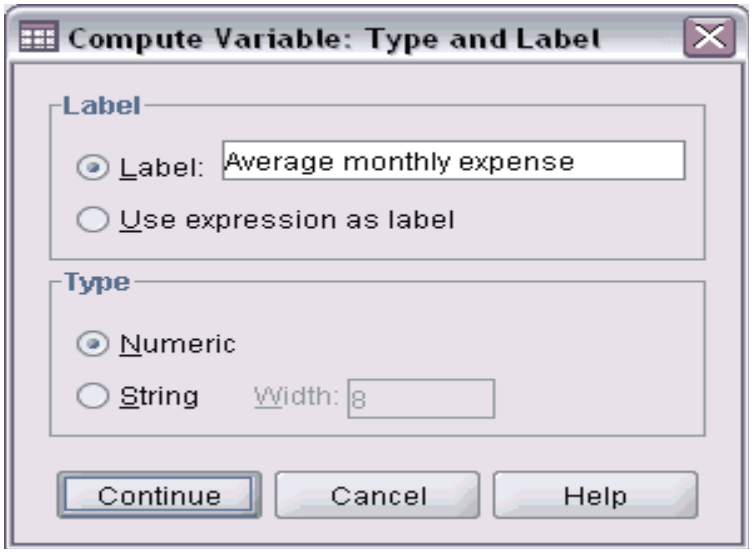
- Ако резултатът на условен израз е „*false” или „missing*”, случаят не е включен в избраното подмножество.
- Повечето условни изрази използват един или повече от шестте релационни оператори (<, >, <=, > =, =, ~=) на калкулатора.
- Условните изрази могат да включват имена на променливи, константи, аритметични оператори, числови и други функции, логически променливи и релационни оператори.

„***Compute Variable: Type and Label***”

По подразбиране новите изчислени променливи са числови. За да изчислите нова низова променлива, трябва да определите типа на данните и техния обхват.

„Label”. Задава етикет на променливата до 120 символа. Можете да въведете етикета или да използвате първите 110 знака на „*Compute expression*” като етикет.

„Type”. Променливата може да бъде числова или низова (буквено-цифрова). Низовите променливи не могат да бъдат използвани при изчисленията.



Фигура 8-3 „Type and Labe” диалогов прозорец

Функции

Предлагат се много видове функции, включително:

- Аритметични функции.
- Статистически функции.
- Низови функции.
- Дата и час функции.
- Многозначни функции.
- Произволни функции за променливи.
- Функции с липсващи стойности.
- Резултатни функции (само за SPSS сървър).

За повече информация и подробно описание на всяка функция, напишете „*functions*” в раздела „*Index Tab*” на помощната системата

Липсващи стойности във функции

Функциите и простите аритметични изрази обработват липсващите данни по различни начини.

В израза *(var1 + var2 + var3) / 3* резултататът е липсващ, ако случаят има липсваща стойност за която и да е от трите променливи.

В израза *MEAN(var1, var2, var3)* резултатът е липсващ само ако случаят има липсващи стойности за всичките три променливи.

За статистическите функции, можете да определите минималния брой на аргументите, които трябва да имат нелипсващи стойности. За да направите това, въведете точка, както и число след името на функцията, както в:
MEAN.2 (var1, var2, var3).

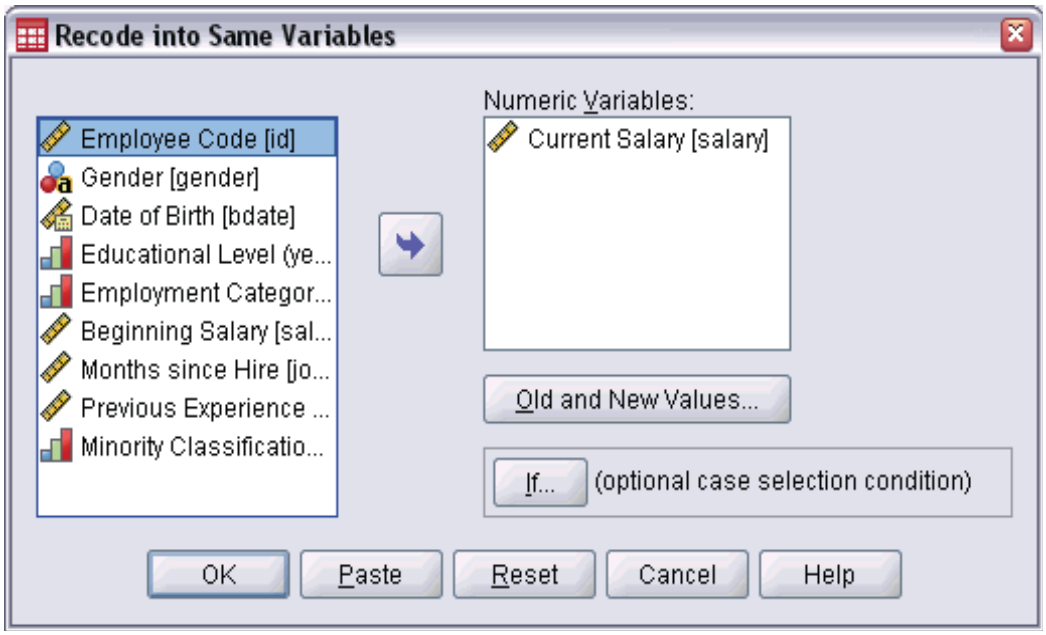
Прекодиране на стойности

Можете да променяте стойностите на данните, като ги записвате. Това е особено полезно за „*collapsing*” или “*combining*” категориите. Можете да прекодирате стойностите в рамките на съществуващи променливи или можете да създадете нови променливи, въз основа на записаните стойности на съществуващи променливи.

Прекодиране в същите променливи

„*Recode into Same Variables*” диалоговия прозорец ви позволява да пренаредите или да модифицирате стойностите на съществуващи променливи. Например, бихте могли да пренаредите заплатите в категории за обхват на заплатите.

Можете да прекодира числови и низови променливи. Ако изберете няколко променливи, всички те трябва да бъдат от същия вид. Не можете да прекодира числови и низови променливи едновременно.



Фигура 8-4 „Recode into Same Variables” диалогов прозорец

За да прекодирате стойностите на променливите

- Изберете от менюто:
  „Transform”
  „Recode into Same Variables...”
- Изберете променливите, които искате да прекодирате. Ако изберете няколко променливи, те трябва да са от същия тип (цифров или низов).
- Кликнете върху „*Old and New Values* ” е и определете как да се прекодират.

По избор можете да определите подмножество на случаи, които да прекодирате. „*If Cases*” диалоговия прозорец за това е същия като описания за „Count Occurrences”.

Прекодиране в същите променливи: „*Old and New Values*”

Можете да определите стойности за прекодиране в този диалогов прозорец. Всички стойностни трябва да бъдат от един и същ тип данни (цифров или низов), както променливите, избрани в главния диалогов прозорец.

„***Old Value*”. Стойност**(а/ите), които трябва да бъдат прекодирани. Можете да прекодира единични стойности, области от стойности и липсващи стойности. „*System-missing*” стойностите и диапазони не могат да бъдат избрани за низови променливи, защото нито една концепция не се отнася до низова променлива. Обхватите включват техните крайни точки и всякакви липсващи стойности въведени от потребителя, които попадат в обхвата.

„*Value*”. Съответната „*Old*” стойност се прекодира в нова. Стойността трябва да бъде от един и същи и същ тип данни (цифров или низов), докато променлив(а/ите) биват прекодирани.

„*System-missing*”. Стойностите, въведени от програмата, когато стойности във вашите данни са неопределени в съответствие с типа формат, който сте задали, когато числово поле е празно, или когато стойността, получена от променящата команда е недефинирата. Числовите „*system-missing*” стойности се показани като точки. Низовите променливи не могат да имат „*System-missing*” стойности, тъй като всеки знак е позволен в низова променлива.

”*System-„ или „user-missing*”. Наблюденията със стойности, които или са определени като „*user-missing*” или не са известни и са им дадени „*system-missing*” стойности, които се показват с точка (.).

„Range”. Включващ обхват от стойности. Не е наличен за низови променливи. Всеки „user-missing” стойности в границите са включени.

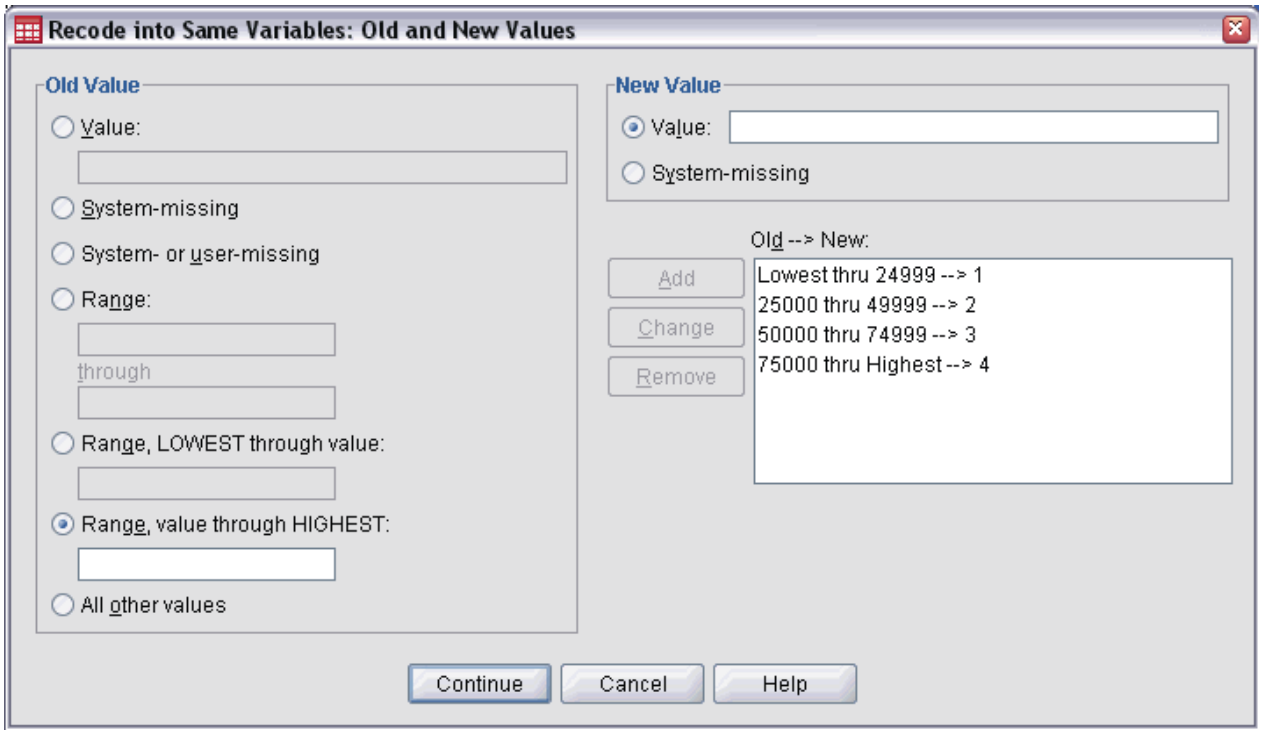
“All other values”. Останалите стойности, които не са включени в една от спецификациите на Old-New списъка. Това се изписва като ELSE на old-new списъка.

„New Value”. Стойността, която се присвоява на всяка стара стойност или диапазон от стойности се записва. Можете да въведете стойност или да присвоите „*system-missing*” стойност. Това става чрез:

- **„Value”.** Стойността, в които една или повече старии променливи, ще бъдат прекодирани. Стойността трябва да е от същия тип данни (цифров или низов), като старата стойност.
- **„System missing”.** Прекодира определена стара стойност, в „*system-missing*” стойност. ”*System-missing*” стойността не се използва при изчисленията, както и в случаите, когато „*system-missing*” стойностите са изключени от много процедури. Не е налично за низови променливи.

„Old-> New”. Списъкът със спецификации, които ще бъде използван за прекодирането на променливите. Можете да добавяте, сменяте, както и да премахвате спецификации от списъка. Списъкът автоматично се сортират, въз основа на спецификацията на старата

стойност, като се използва следния ред: единични стойности, липсващи стойности, обхвати и всички други стойности. Ако промените прекодира спецификация от списъка, процедурата автоматично пре-сортира списъка, за да се запази тази последователност.

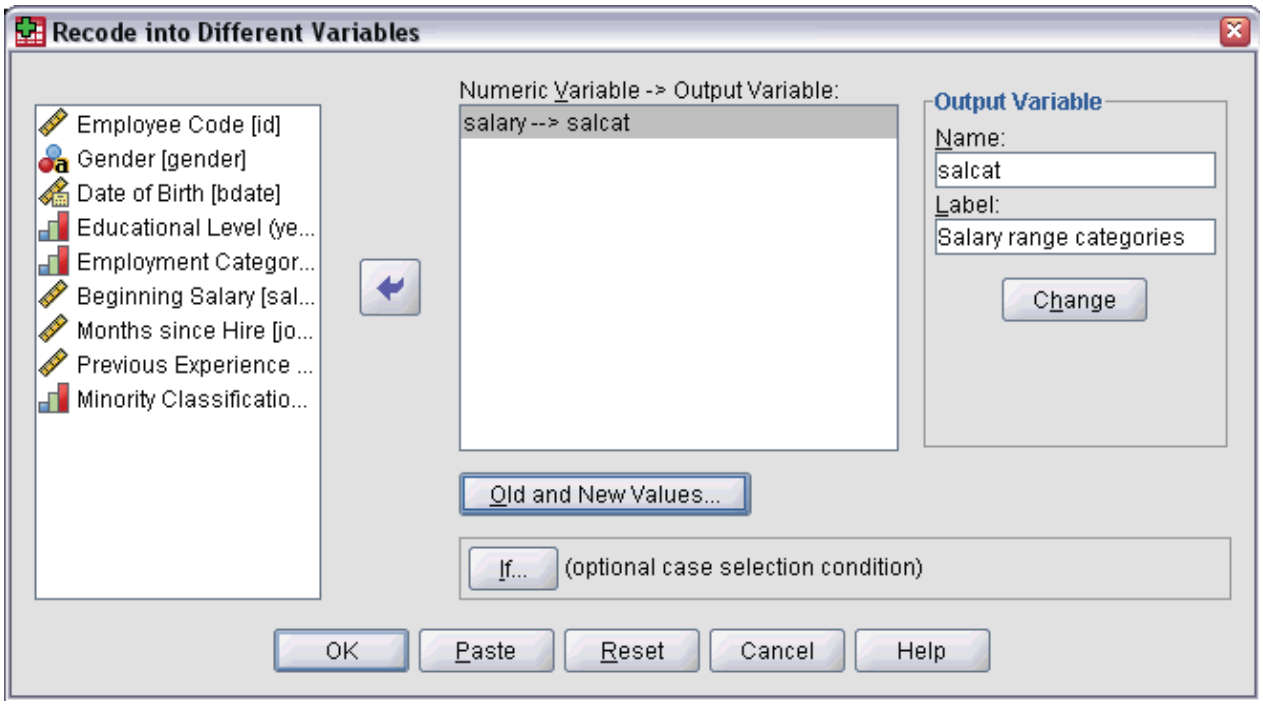


Фигура 8-6 „Old and New Values" диалогов прозорец

Прекодиране в различни променливи

„*Recode into Different Variables*" диалоговият прозорец ви позволява да модифицирате стойностите на съществуващи променливи като създадете нови променливи. Така например, бихте могли да пренаредите заплатите в нови променливи, съдържащи заплата-обхват категории.

- Можете да прекодирате числови и низови променливи.
- Можете да прекодира числови променливи в низови променливи и обратно.
- Ако изберете няколко променливи, всички те трябва да бъдат от същия тип. Не можете да прекодира цифрови и низови променливи заедно.



Фигура 8-7 „Recode into Different Variables" диалогов прозорец

За прекодиране стойностите на променливите в нови

- Изберете от менюто:
  Transform
  Recode into Different Variables…
- Изберете променливата, която искате да прекодира. Ако изберете няколко променливи, те трябва да са от същия тип (цифров или низов).
- Въведете „*output(new) variable name*" всяка нова променлива и щракнете върху „*Change*".
- Кликнете върху „*Old and New Values*" и определете как да се прекодират променливите.

По желание можете да определите подмножество от случаи за прекодиране. "*The If Cases*" диалоговия прозорец за това е същият като вече описаният.

„***Recode into Different Variables: Old and New Values*"**

Можете да определите стойности за прекодиране в този диалогов прозорец.

„*Old Value*". Стойността, която трябва да се прекодира. Можете да прекодирате отделни стойности, обхват от стойности, и липсващи стойности. „*System missing*" стойности и диапазоните не могат да бъдат избирани за низови променливи, защото никаква концепци не се прилага спрямо низовите променливи. Старите стойности трябва да са от същия тип данни (цифров или низов), както и оригиналните променлива. „*Ranges*" включват техните крайни точки и всички „*user-missing*" стойности , които попадат в обхвата.

„*Value*". Старата стойност се прекодира в нова. Стойността трябва да бъде от същия тип данни (цифров или низов), като променлива за прекодиране.

„*System-missing*". Стойности, въведени от програмата на местата където вашите са неопределени в съответствие с типа формат, който сте задали, когато числовото поле е празно, или когато стойност, в резултат на преобразуване е недефинирана. Числовите „*system-missing*" стойности са показани като периоди. Низовите променливи не могат да имат „*system-missing*" стойности, тъй като всеки знак е валиден в низова променлива.

„*System- or user-missing*”. Наблюдения със стойностите, които или са определени като „*user-missing*” данни или не са известни и са им присвоени „*system-missing*”стойност, която е показан с период (.).

„*Range*”. Включващ диапазон от стойности. Не е налично за низови променливи. Всички „*user-missing*” стойности в границите са включени.

“*All other values*”. Всякакви останали стойности, които не са включени в една от спецификациите на „*Old-New*” списъкс. Това изглежда като „*ELSE*” на „*Old-New*” списъка.

“*New Value*”. Стойност, в която всяка стара стойност или обхват от стойности се прекодира. Новите стойности могат да бъдат цифрови или низови.

„*Value*”. Стойността, в които една или повече използвани стойности, ще бъдат прекодирани. Стойността трябва да е от същия тип данни (цифров или низов), както старата стойност.

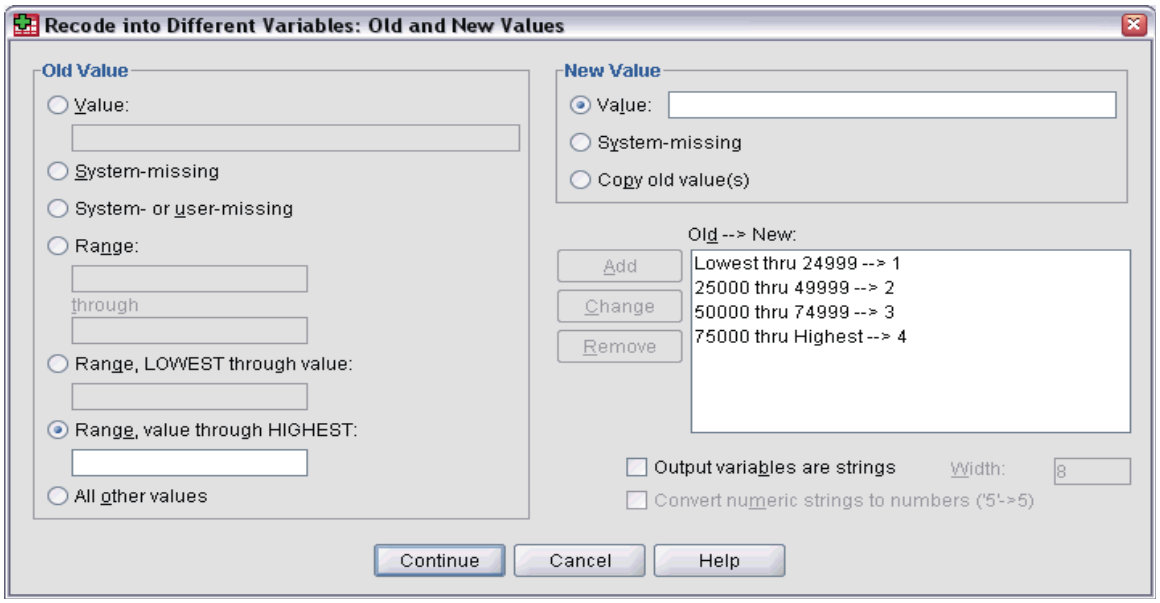
„*System-missing*”. Прекодира определени стара стойности, в „*system-missing*” стойност. ”*System-missing*” стойността не се използва при изчисления, случаите с „*system-missing*” стойност са изключвани от много процедури. Не е налично за низови променливи.

„*Copy old values*”. Запазва старата стойност. Ако някои от стойностите не се нуждаят от прекодиране, използвайте тази опция, за да включва старата стойност. Всички стари стойности, които не са посочени, не са включени в новата променлива и случаите с тези стойности ще бъдат присвоени с „*system-missing*” стойност за нова променлива.

„*Output variables are strings*”. Определя новата, прекодирана променлива като низ (буквено-цифрова) променлива. Старата променлива може да бъде цифрова или низова.

„*Convert numeric strings to numbers*”. Преобразува низовите стойности, съдържащ номера в числови. Низовете, съдържащи нищо друго освен номера и незадължителен знак (+ или -) са преписани на „*system-missing*” стойностите.

„***Old-> New*”.** Списъкът със спецификации, които ще бъде използван за прекодирането на променливат(а/ите). Можете да добавяте, променяте, както и да премахване спецификации от списъка. Списъкът автоматично се сортират, въз основа на спецификациите на старата стойност, като се използва следния ред: единични стойности, липсващи стойности, обхвати, както и всички останали стойности. Ако промените прекодирана спецификация в списъка, процедурата автоматично пресортира списъка (ако е необходимо, за да се запази тази последователност).



Фигура 8-8 „Old and New Values” диалогов прозорец

Глава 9. Работа с файлове

Обработка на файл и на файлови трансформации

Файловете с данни не винаги са организирани в идеална форма за вашите специфични нужди. Вие може да искате да комбинирате файловете с данни, да сортирате данните в различен ред, да изберете подмножество от случаи, или да променяте единицата за анализ чрез групиране на случаите заедно. Широка гама от възможности за трансформация на файл е валидна, включително и възможността за:

Сортиране на данни. Можете да сортирате случаи въз основа на стойността на една или повече променливи.

Преобразуване на случаи и променливи. SPSS прочита редовете като случаи, а колоните като променливи. За файлове с данни, в които тази подредба е обърната, можете да обърнете редовете и колоните и да прочетене данните в правилния формат.

Обединяване на файлове. Можете да обедените два или повече файла с данни. Можете да комбинирате файлове със същите променливи, ако са в различни случаи или едни и същи случаи, но с различни променливи.

Избиране на подмножество от случаи. Можете да ограничите вашия анализ до подгрупа от случаи, или да извършите едновременно анализ на различни погрупи.

Сумиране на данни. Можете да променяте единицата за анализ, като се сумират случаи въз основа на стойността на една или повече групиращи променливи.

Тегла на данни. Теглата на данните за анализи на базата на стойността на теглата на данните.

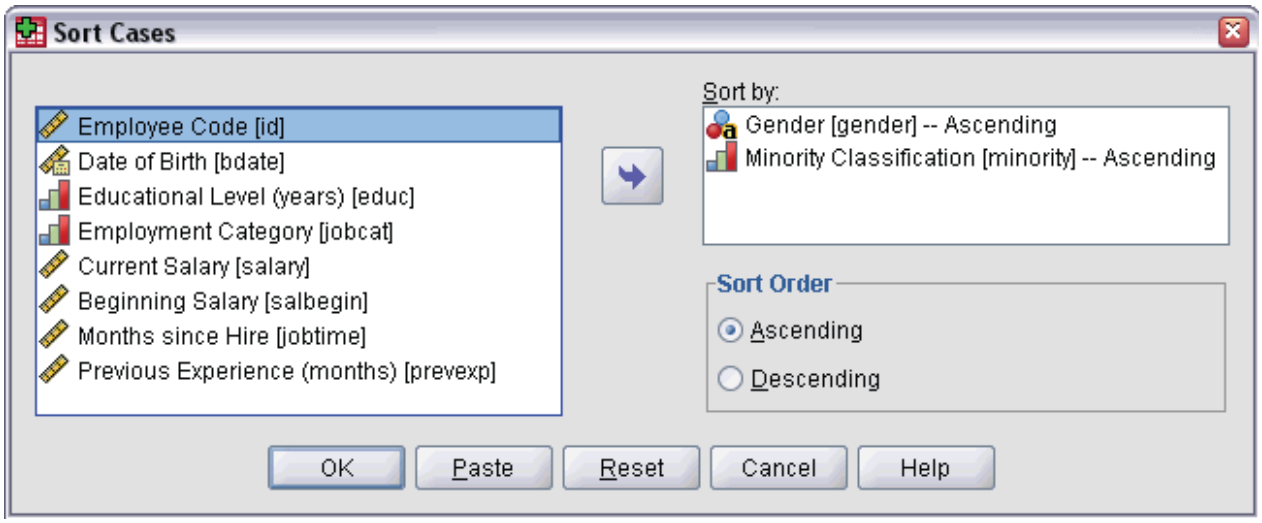
Преструктуриране на данни. Можете да преструктурирате данни, за да създаде единичен случай (запис) от множество случаи или да създадете множество случаи от единичен случай.

„Sort Cases” – Сортиране на случаи

Този диалогов прозорец сортира случаи (редове) на файла с данни въз основа на стойностите на една или повече сортиращи променливи. Можете да сортирате случаи във възходящ или низходящ ред.

- Ако изберете множество сортиращи променливи, случаите се сортират по всяка променлива в категории от предходните променливи на сортиращия лист. Така например, ако изберете „gender” като първата сортираща променлива и „*minority*” като втора сортираща променлива, случаите ще бъдат сортирани по малцинство в рамките на всяка „gender” категория.

- Сортиращата последователност се основава на определен по място ред (и не е задължително да е същата като цифровата подредба на кодове за характеристика). Подразбиращото се място е операционната система. Можете да контролирате мястото с езиковите настройки в „*General tab*” на „*Option dialog box (Edit menu)*”.



Фигура 9-1 "Sort Cases" диалогов прозорец

За да сортирате случаи

- Изберете от менюто:
  „*Data*"
  „*Sort Cases…*"
- Изберете една или повече сортиращи променливи.

„Sort Variables" – Сортиране на променливи

Можете да сортирате променливите в активния набор от данни въз основа на стойностите на всички променливи атрибути. (например, име на променлива, типа данни, измерване на нивото), включително и избираеми атрибути на променливи.

-Стойностите могат да бъдат подредени във възходящ или низходящ ред.

-Можете да запаметите оригиналния (предварително сортирания) ред на стойност в произволен атрибут на променлива.

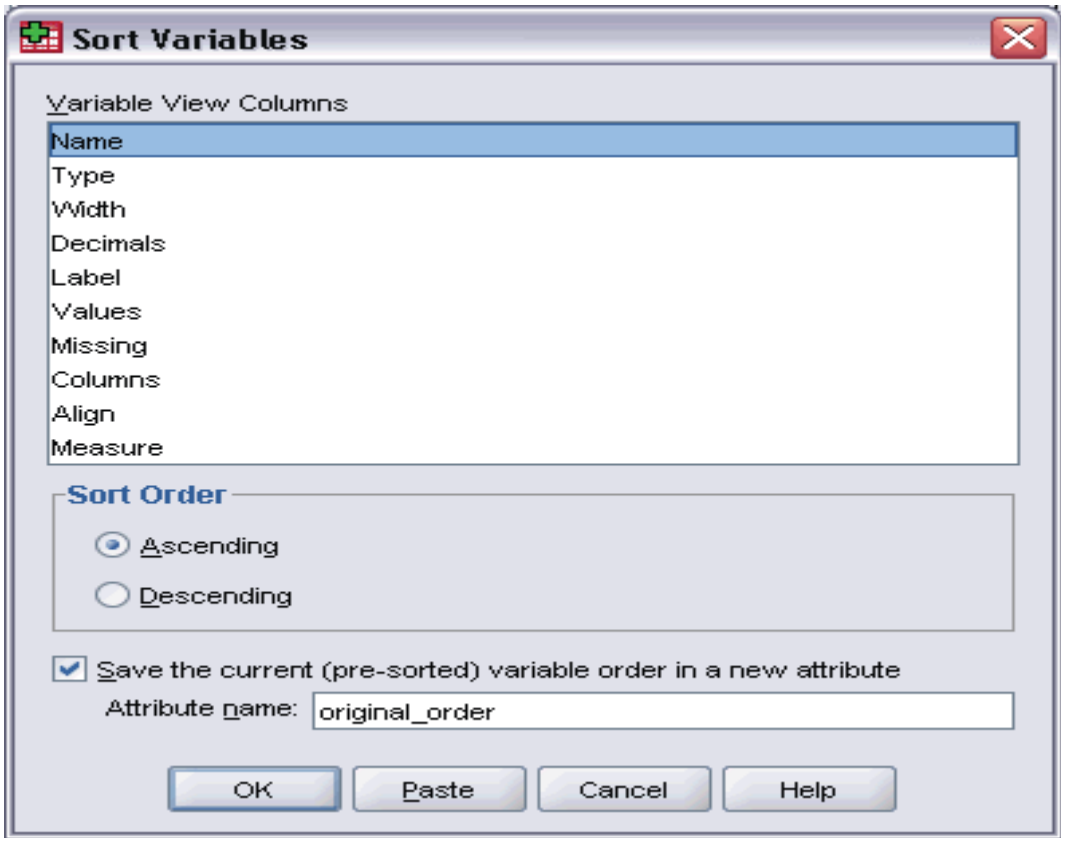
-Сортирането чрез стойности на произволни атрибути на променливи е ограничено до произволни атрибути, които са в момента видими в „Variable View".

За да сортирате Променливи

- В „*Variable View*" на „*Data Editor*":
  Щракнете с десния бутон върху заглавието на колоната за атрибути и от контекстното меню избирете „*Sort Ascending*" или „*Sort Descending*".

или

- От менютата на всеки прозорец, изберете:
  „*Data*"
  „*Sort Variables*"
- Изберете атрибута, който искате да използвате, за да сортирате променливи.
- Изберете сортиращият ред (възходящ или низходящ).



Фигура 9-2 „Sort Variables" диалогов прозорец

Сливане на файлове с данни

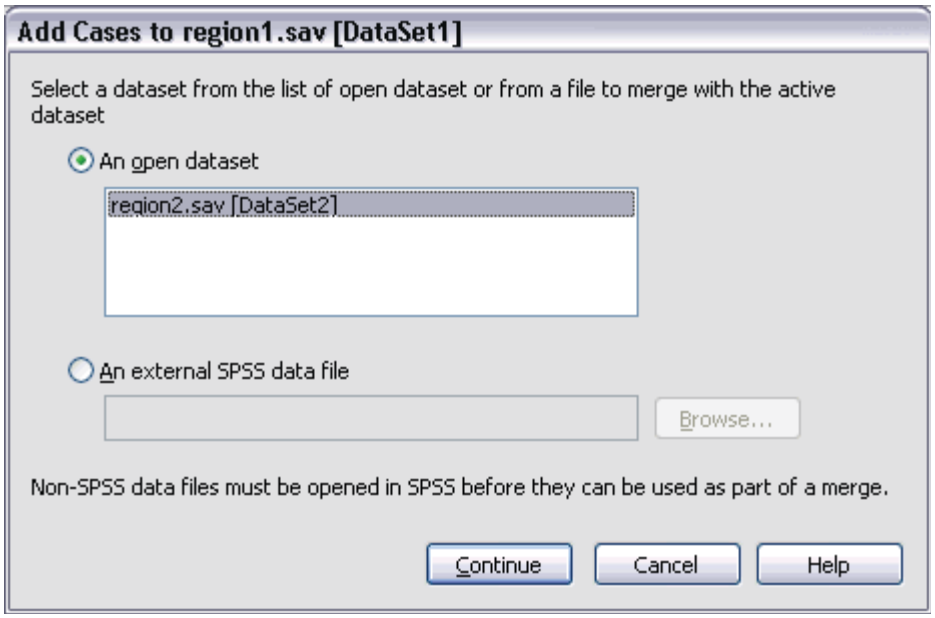
Можете да обедините данните от два файла по два различни начина. Можете да:

- Обединявате активния набор от данни с друг отворен набор от данни или данни от SPSS формат, съдържащи същите променливи, но различен брой случаи.

- Обедините активния набор от данни с друг отворен набор от данни или данни от SPSS формат, съдържащи същите случаи, но различни променливи.

За да обедините файлове

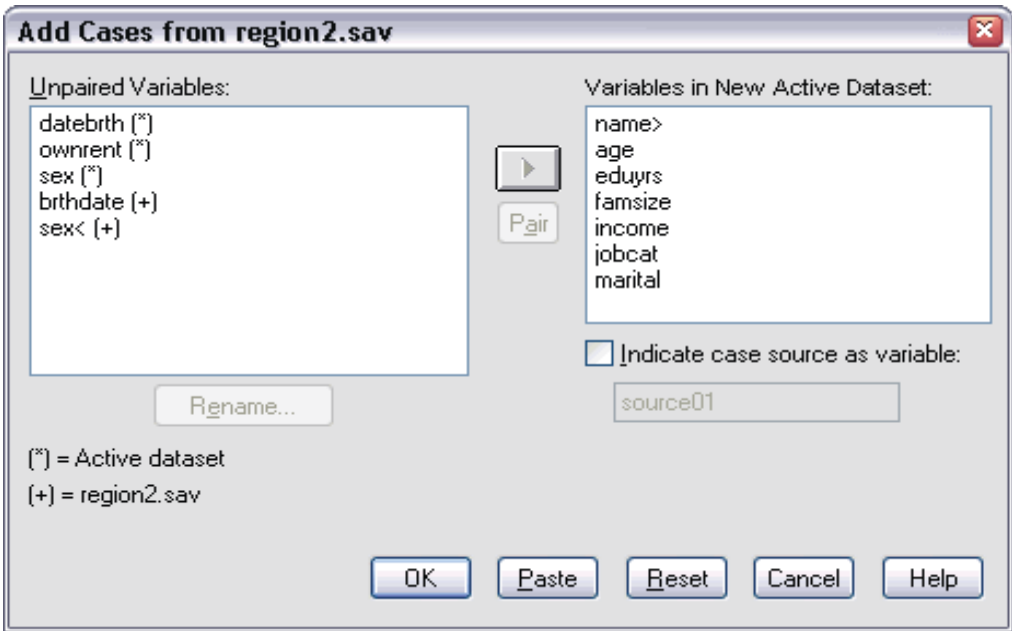
- Изберете от менюто:
  „*Data*"
  „*Merge Files*"
- Изберете или добавяне случаи или променливи.



Фигура 9-3 Избиране на файлове за сливане

„Add Cases" - Добавяне на случаи

„Add Cases" обединява активния набор от данни с втори набор от данни или SPSS форматен файл с данни, който съдържа същите променливи (колони), но различни случаи (редове). Например, можете да запишете същата информация за клиенти в два различни региони за продажби и да поддържате данните за всеки регион в отделни файлове. Вторият набор от данни може да бъде външен SPSS-формат файл за данни или валиден набор от данни в текущата сесия.



Фигура 9-4 „Add Cases" диалогов прозорец

Несъвпадащи променливи. Променливи, които биват изключени от новата, обединена база от данни. Променливи от активния набор от данни, се идентифицират със звездичка (*). Променливи от друг набор от данни, са идентифицират със знак плюс (+). По подразбиране в този този списък съдържа:

- Променливи от всеки файл с данни, който не съвпадат с имена на променливи от друг файл. Можете да създавате двойки от нечифтосани променливи и да включате в тях нови, обединени файлове.

- Променливи определени като числови данни в един файл и низови в друг файл. Числовите променливи не могат да бъдат слети с низовите.

- Низови променливи с неравна дължина. Определената дължина на низовата променлива трябва да бъде същата в двета файла с данни.

Променливи в нов активен набор от данни. Променливите, които биват включени в нов, обединен файл с данни. По подразбиране всички променливи, които съвпадат и по име и по тип на данни (числов или низов), са включени в списъка.

- Можете да премахвате променливи от списъка, ако не искате да ги включвате в обединения файл.

- Всички нечифтосани променливи, включени в обединения файл ще съдържат липсващи данни за случаите от файла, който не съдържат тази променлива.

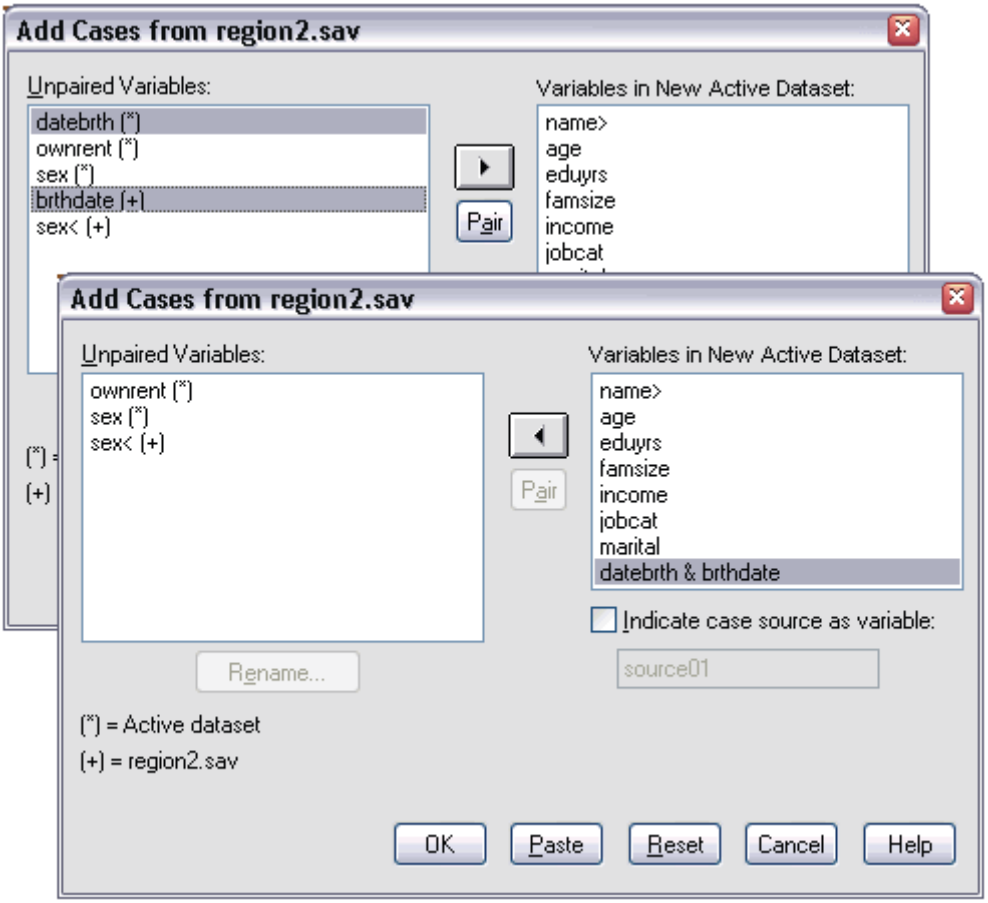
„*Indicate case source as variable*". Посочва източника на файлове с данни за всеки отделен случай. Тази променлива има стойността 0 за случаи от активния набор от данни и стойност 1 за случаи от външен файл с данни.

За да слеете файлове с данни със същите променливи и различни случаи

- Отворете един от файловете с данни, който искате да обедините. Ако имате няколко отворени набори от данни, „*направете*" един от наборите с данни, които искате да обедините с активния набор от данни. Случаят от този файл ще се появи първи в новата, обединена база данни.
- Изберете от менюто:
  „*Data*"
  „*Merge Files*"
  „*Add Cases…*"
- Изберете набор от данни или данни от SPSS формат, който да слеете с активния набор от данни.
- Отстранете всякакви променливи, които не желаете от „*Variables*" в „*New Active Dataset*" листа.
- Добавите всякакви двойки променлива от „Unpaired Variables" списъка, които представляват същата информация записана с различни имена на променливи в двата файла. За пример, дата на раждане може да има за име на променлива – „*brthdate*" в един файл и „*datevrth*" в друг файл.

За да изберете двойка от „*Unpaired*" променливи

- Кликнете върху една от променливите в „*Unpaired Variables*" листа.
- Ctrl кликнете върху другата променливи в списъка. (Натиснете клавиша Ctrl и натиснете левия бутон на мишката по същото време.)
- Кликнете „*Pair*", за да предвижите двойката променливи към „*Variables*" в „*New Active Dataset*" списъка. (Името на променливата от активния набор от данни се използва като име на променлива в обединения файл).



Фигура 9-5 Избирането на двойки променливи с Ctrl-клик

Добавяне на случаи: Преименуване

Можете да преименувате променливи, или от активния набор от данни или от друг набор от данни, преди да ги преместят от „*unpaired*" списъка към списъка на променливите, който ще бъде включен в обединения файл с данни. Преименуването на променливи ви позволява да:

- Използвайте името на променливата от друг набор от данни, а не от името на активния набор от данни за двойки променливи.

- Включвате две променливи със същото име, но с различни типове или с различна низова дължина. Например, за да включите и числовата променлива „*sex*" от активния набор от данни и низовата променливата „*sex*" от друг набор от данни, една от тях трябва да се преименува първа.

Добавяне на случаи: Информация за речника

Всичка съществуваща речникова информация (променливи и етикети със стойност, липсващи от потребителя стойности, показани формати) в активния набор от данни се прилага до обединения файл с данни.

- Ако някаква речникова информация за дадена променлива е недефинирана в активния набор от данни, се използва речникова информация от друг набор от данни.

- Ако активния набор от данни съдържа определени етикети със стойности или потребителски липсващи стойности за дадена променлива, всякакви допълнителни етикетни стойност или потребителски липсващи стойности за тази променлива в друг набор от данни се игнорират.

Обединяване на повече от две източника на данни

Използвайки командния синтаксис, можете да обединявате до 50 множества от данни и/или файлове с данни. За повече информация, вижте „*ADD FILES*" командата в „*Command Syntax Reference*" (достъпна от „*Help*" менюто).

„Add Variables" – Добавяне на променливи

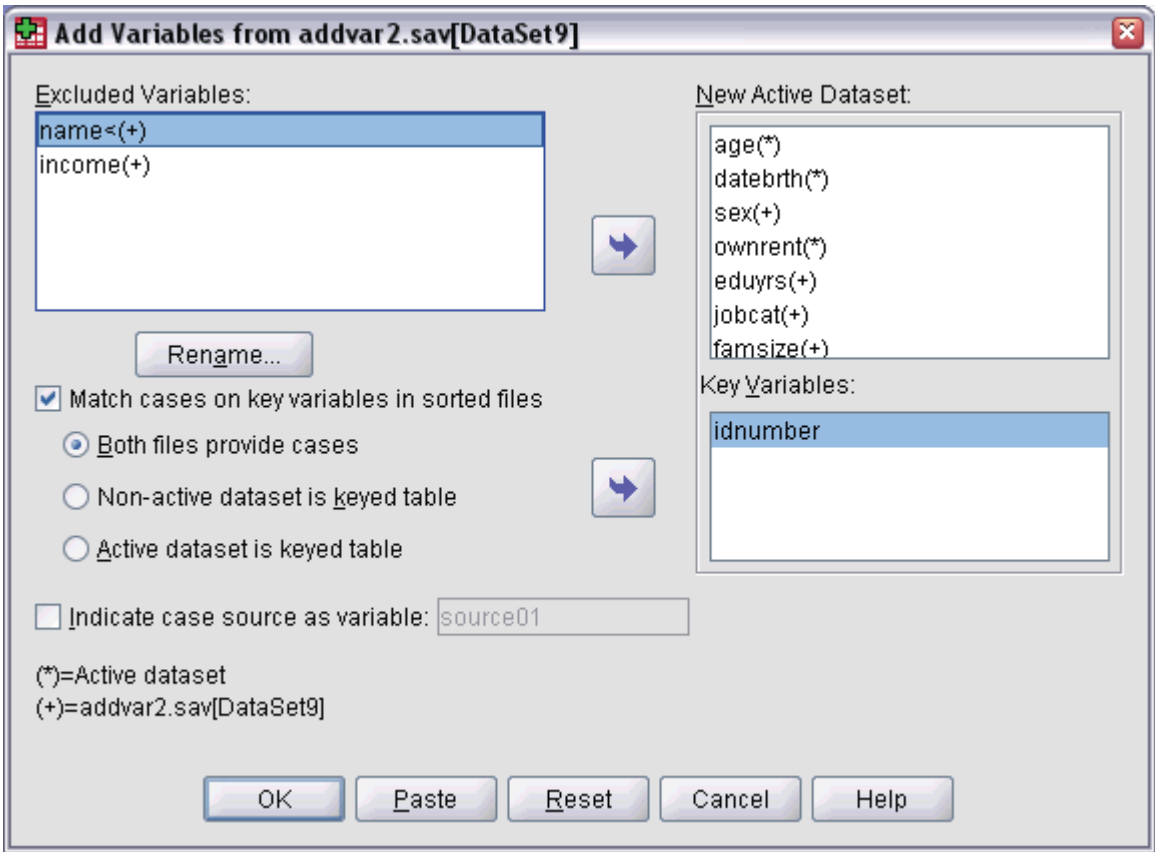
„*Add variables*" обединява активния набор от данни с друг отворен набор от данни или SPSS форматни данни, които съдържат същите случаи (редове), но различни променливи (колони). Например, можете да искате да обедините файл с данни, който съдържа резултатите преди теста с такъв, който съдържа резултатите след теста.

- Случаите трябва да бъдат подредени в същия ред и в двете поредици.

- Ако една или повече променливи ключ се използват за съвпадане на случаи, двете множества от данни трябва да бъдат сортирани във възходящ ред по променлив(ата/ите) ключ.

- Имената на променлива във втория файл с данни, който съвпадат с на имената на променливите в активното множество от данни, са изключени по подразбиране, защото „*Add Variables*" предполага, че тези променливи съдържат съвпадаща информация.

„Посочване на източника на случаи, като променлива". Посочва изходния файл с данни за всеки отделен случай. Тази променлива има стойност на 0 за случаите от активното множество от данни и стойност 1 за случаите от външния файл с данни.



Фигура 9-6 „Add Variable" диалогов прозорец

„Excluded Variables" - Външни променливи. Променливи, които биват изключени от новата, обединена база данни. По подразбиране този списък съдържа имена на всички променливи от другото множество от данни, на което имената на променливите се дублират с имената на променливите в активното множество от данни. Променливите от активното множество от данни се идентифицират със звездичка (*). Променливите от други множества от данни се идентифицират със знак плюс (+). Ако искате да включите външна променлива с дублиращо име в обединения файл, можете да я преименувате и да я добавите към списъка с променливите за включване.

„New Active Dataset" - Нови активни множества от данни. Променливите, които биват включени в ново, обединено множество от данни. По подразбиране, всички уникални имена на променливи и в двете множества от данни са включени в списъка.

„Key Variables" – Променливи ключ. Ако някои случаи в един набор от данни не съвпадат със случаите в друг набор от данни (което, значи че някои случаи липсват в едното множество от данни), използвайте променливи ключ, за да идентифицирате и правилно да съпоставите случаи от двете множества от данни. Можете също да използвате променливи ключ с файлове от тип „*table lookup*".

- Променливите ключ трябва да имат едни и същи имена и в двете множества от данни.

- И двете множества от данни трябва да са сортирани във възходящ ред по променливита ключ, и реда на променливите в списъка с променливи ключ трябва да бъдат същите като тяхната сортирана последователност.

- Случаите, които не се дублират с променливата ключ са включени в обединения файл, но не се обединяват със случаи от други файлове. Несъвпадащите случаи съдържат стойности само за променливите във файла, от които са взети; променливите от другия файл съдържат системно лишсващи стойности.

„Non-active or active dataset is keyed table". Ключова таблица или **„*table lookup file*",** е файл, в който данни за всеки отделен "случай" могат да бъдат добавени към множество случаи, в друг файл с данни. Например, ако един файл съдържа информация за отделните членове на семейството (като пол, възраст, образование) и другия файл съдържа цялата информация за семейството (като общият доход, размера на семейството, местоположение), можете да използвате файл с данните за семейството, като „*table lookup*" файл и да прилагате общите данни за семейстово за всеки индивидуален член от семейството в обединения файл с данни.

За да обедините файлове със същите случаи, но с различни променливи

- Отворете поне един от файлове с данни, които искате да обедините. Ако сте отворили няколко множество с данни направи едно от множествата с данни, които искате да обедините активното.
- Изберете от менюто:
  „*Data*"
  „*Merge Files*"
  „*Add Variables…*"
- Изберете множеството от данни или SPSS файл с данни, който да обедините с активното множество от данни.

За да изберете променливи ключ

- Изберете променливите от външния файл (+) от „*Exclude Variables*" списъка.
- Изберете съвпадащи на променливите ключ в сортираните файлове.
- Добавянете променливите при „*Key Variables*" списъка.

Променливите ключ трябва да съществуват едновременно в активното множество от данни и в другото множество от данни. И двете множества от данни трябва да бъдат сортирани във възходящ ред по променливита ключ и редът на променливите в „*Key Variables*" списъкът трябва да бъдат същите като тяхната сортирана редица.

Добавяне на променливи: преименуване

Можете да преименувате променливи, или от активното множество от данни или от друго множество от данни, преди да ги преместите в списъка с променливите, който ще бъде включен в обединения файл с данни. Това е преди всичко полезно, ако искате да включите две променливи със същите имена, които обаче съдържат различна информация.

Обединяване на повече от два файла

Използвайки командния синтакс, можете да обедините до 50 множества от данни и/или файлове с данни. За повече информация, използвайте „*MATCH FILES*" командата в „*Command Syntax Reference*" (достъпна от „*Help*" менюто).

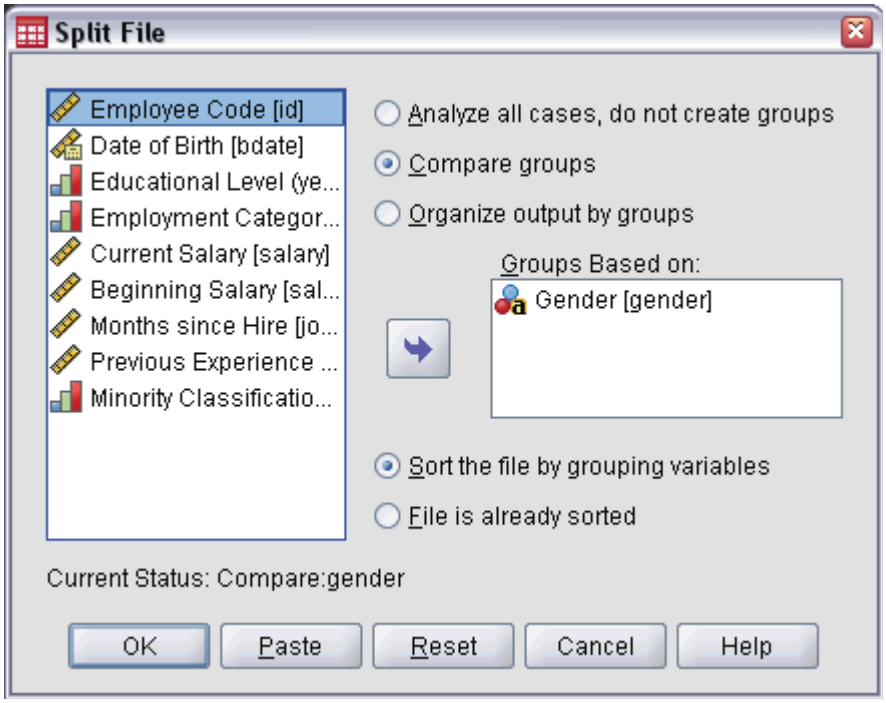
Разделяне на файла

„*Split File*" разделя файла с данни в отделни групи за анализ въз основа на стойностите на една или повече групиращи променливи. Ако изберете няколко групиращи променливи, случаите се групират по всяка променлива в категории от предходни променливи в „*Groups Based On*" списъка. Например, ако изберете „*gender*", като първата групираща променлива и „*minority*", като втората групираща променлива, случаите ще бъдат групирани по „minority" класификация във всяка „*gender*" категория.

- Можете да зададете до осем групиращи променливи.

„- Всяка низова променлива от осем байта (низови променливи по дълги от осем байта) се смятат за променлива близки до границата на осем групиращи променливи.”

- Случаите трябва да бъдат сортирани по стойностите на групиращите променливи и в същия ред, по който променливите са изброени в „*Groups Based On*” списъка. Ако файла с данни, не е вече сортиран, изберете „*Sort the file by grouping variables*”.



Фигура 9-7 „Split File” диалогов прозорец

„Compare groups” – Сравняване на групи. Групите с обединени файлове са представени заедно за сравнения. За главните таблици, една централна се създадева и всяка променлива на разделен файл може да бъде придвижвана по таблиците. За диаграми, отделна диаграма се създадена за всяка група разделени файлове и диаграмите са показват заедно в „*Viewer*”.

„Organize output by groups”. Всички резултати от всяка процедура, са показани отделно за всека група разделени файлове.

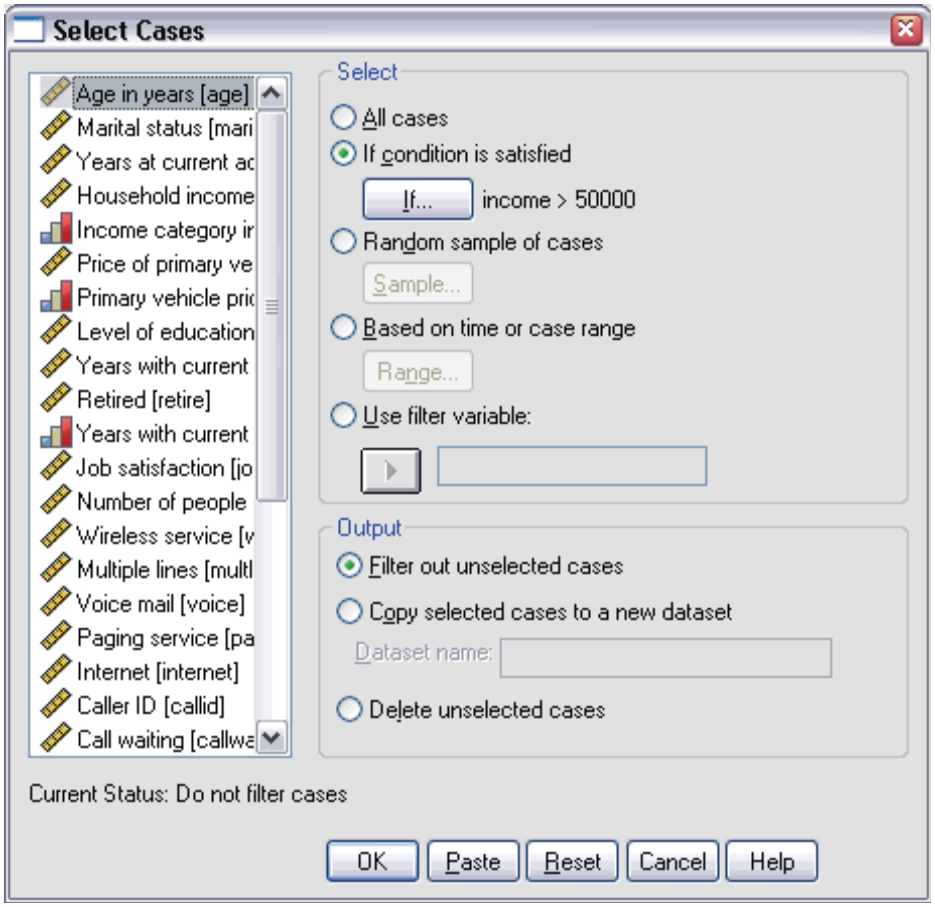
За да разделите файл с данни за анализ

- Изберете от менюто:
  „*Data*”
  „*Split File…*”
- Изберете „Compare groups” или „Organize output by groups”.
- Изберете една или повече групиращи променливи.

„Select Cases”

„Select Cases” предлага няколко метода за избор на подгрупа от случаи въз основа на критерии, които включват променливи и сложни изрази. Можете също така да изберете произволни извадки от случаи. Критериите, използвани за определяне на подгрупа могат да включват:

- Стойности на променливи и диапазони
- Дата и време диапазони
- номера на случаи (редове)
– Аритметични изрази
- Логически изрази
- Функции



Фигура 9-8 „Select Cases” диалогов прозорец

„All Cases” - Всички случаи. Изключва филтрирането на случаи и използва всички такива.

„If condition is satisfies” - **Ако условието е изпълнено.** Използва се условнен израз, за да се изберат случаи. Ако резултатът на условния израз е „*вярно*”, случаят се избира. Ако резултатът е „*невярно*” или е липсващ, случаят не се избира.

„Random sample of cases” - Случайна извадка от случаи. Избира случайна извадка въз основа на приблизителния процент или на точният брой на делата.

„Based on time or case range" - Въз основа на обхват от време променливи или случаи. Избира случаи въз основа на обхват от номера на случай или обхват от дати/време.

„Use filter variable" - Използвай филтрираща променлива. Използвайте избраната числова променлива от файла с данни, като филтрираща променлива. Случаи със стойност различна от 0 или липсваща такава за филтриращата променлива се избират също.

„Output" - **Резултат**

Този раздел контролира третирането на неизбраните случаи. Можете да изберете един от следните варианти за третиране на неизбран случаи:

„Filter out unselected cases" - Филтриране на неизбрани случаи. Неизбраните случаи, не са включени в анализите, но остават в Множеството от данни. Можете да използвате неизбраните случаи по-късно в сесията, ако изключите филтрирането. Ако изберете произволна проба или ако изберете случаи въз основа на условен израз, това създава променлива с име *filter_$* със стойност 1 за избраните случаи и стойност 0 за неизбраните.

„Copy selected cases to a new database" - Копиране на избраните случаи в ново множество от данни. Избраните случаи се копират в ново множество от данни, като оставят оригиналното множество от данни не променено. Неизбраните случаи не са включени в новото множество от данни и се оставят в първоначалното си състояние в оригиналното множество от данни.

„Delete unselected cases" - Изтриване на неизбраните случаи. Неизбраните случаи се заличени от множеството от данни. Изтритите случаи могат да бъдат възстановени само ако излезете от файла, без да запаметите промените и след това отворите отново файла. Заличаването на случаи е необратимо, ако запишете промените във файла с данни.

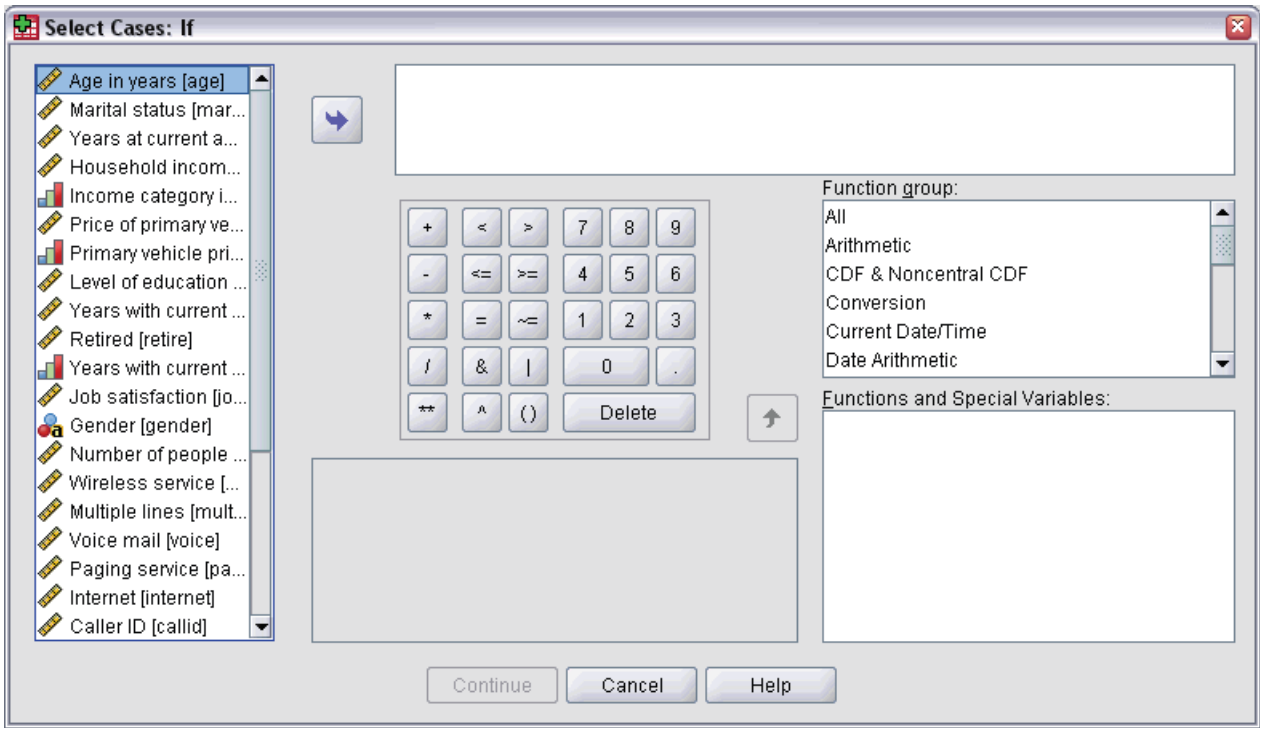
Забележка: Ако изтриете неизбран случаи и запишете файла, случаите не могат да бъдат възстановени.

За да изберете подмножество от случаи

- Изберете от менюто:
  „*Data*"
  „*Select Cases...*"
- Изберете един от методите за подбор на случаи.
- Посочете критериите за подбор на случаи.

„Select Cases: If” - Изберане на случаи: „*If*”

Този диалогов прозорец ви позволява да изберете подмножества от случаи, използвайки условни изрази. Условният израз връща стойност „*true*”, „*false*”, или липсваща стойност за всеки отделен случай.



Фигура 9-9 „Select Cases: If” диалогов прозорец

- Ако резултатът на условен израз е „*true*”, случаят се включва в избраното подмножество.

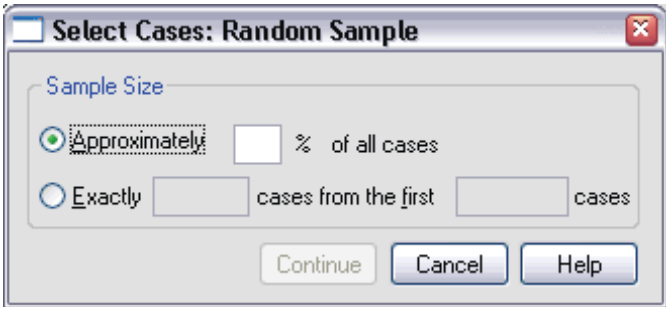
- Ако резултатът на условен израз е „*false*” или липсващ, случаят не се включва в избраното подмножесто.

- Повечето условни изрази използва един или няколко от шестте релационни оператора (<, >, <=, >=, =, ~=).

- Условните изрази могат да включват имена на променливи, константи, аритметични оператори, числови (и други) функции, логически променливи и релационни оператори.

„Select Cases: Random Sample” - Избиране на случаи: случайна извадка

Този диалогов прозорец ви позволява да изберете случайна извадка въз основа на приблизителния процент или на точният брой на случаите. Извадките се извършва без замяна; така същия случай не може да Бъде избиран повече от веднъж.



Фигура 9-10 „Select Cases: Random Sample” - диалогов прозорец

„Approximately” - Приблизително. Генерира случайна извадка от приблизително определен процент на случаите. Тъй като тази процедура създава независимо псевдо-случайно решение за всеки отделен случай, процентът на избраните случаи може да бъде само приблизително определен. Колкото повече случаи има във файла с данни, толкова по-близък е процентът от избраните случаи до определения процент.

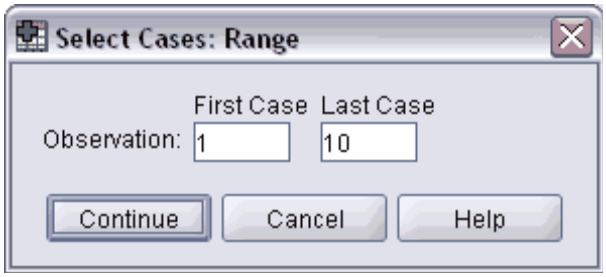
„Exactly” - Точно. Числото случаи, посочено от потребителя. Вие също трябва да посочите броят на случаите, от които, да се генерира пробата. Второто число трябва да бъде по-малко или равно на общия брой на случаите във файла с данни. Ако числото надхвърля общия брой на случаите във файла с данни, извадката ще съдържа пропорционално по-малко случаи, отколкото въведеното число.

„Select Cases: Range” – Избиране на случаи: обхват

Този диалогов прозорец избира случаи въз основа на обхват от номера със случай или обхват от дати или време.

- Обхватите със случаи се основават на редов номер, както е показано в „*Data Editor*” – Редактор за данни.

- Дата и час обхватите са валидни само за ***„time series”*** с определена променлива от тип дата („*Data menu*” – Дата меню, „*Define Dates*” – Определени дати).



Фигура 9-11 „Select Cases: Range” диалогов прозорец за обхвати от случаи (не са дефинирани променливите от тип дата)

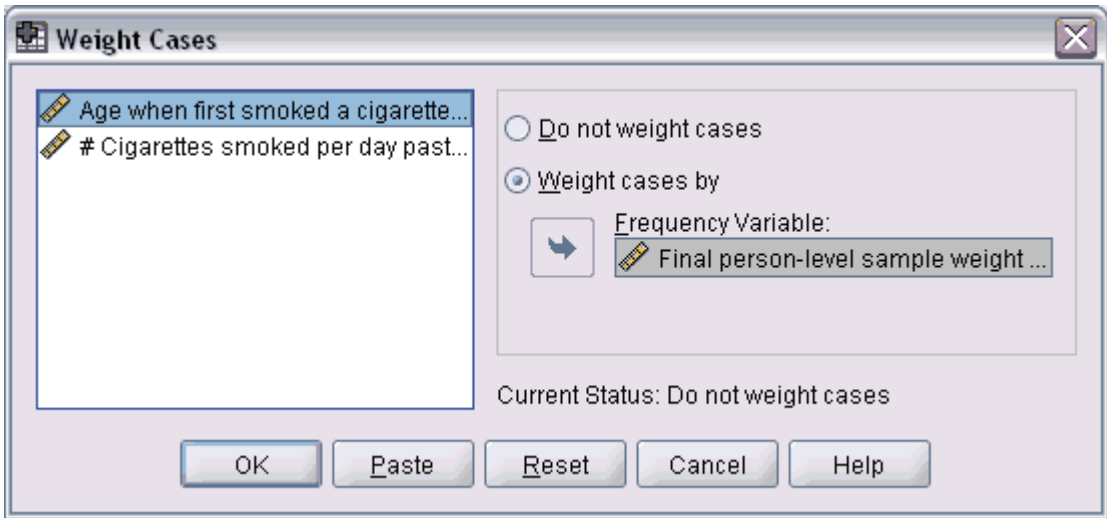
„Weight Cases” - Тегло на случаи

„*Weight Cases*” дава на случаите различни тегла (чрез симулирана репликация) за статистическите анализи.

- Стойностите на променливите-тегло трябва да посочват броя на наблюденията представени от единични случаи във файла с данни.

- Случаи с нула, отрицателни или липсващи стойности за променливата-тегло се изключват от анализа.

- Стойностите с плаваща запетая са валидни; те се използват там, където това е смислено и най-вероятно където случаите са таблични.



Фигура 9-12 „Weight Cases” диалогов прозорец

След като приложите променливата-тегло, тя остава в сила, докато не изберете друга променлива-тегло или изключите теглата. Ако запишете файл с данни - тегла, информацията за теглата се записва заеднос файла с данни. Можете да изключите теглата по всяко време, дори и след като файлът е запаметен във формат „*тегло*”.

Теглата в „*Crosstabs*”. Процедурата „*Crosstabs*” има няколко възможности за разглеждане на случаите-тегла.

Тегла в „*scatterplots*” и хистограми. „*Scatterplots*” и хистограмите имат опция за включване и изключване на случаите-тегла, но това не засяга случаите с нулева, отрицателна или липсваща стойност за променливата-тегло. Тези случаи остават изключени от графиката, дори и ако сте изключили теглата.

За да претеглите случаи

- Изберете от менюто:
  „*Data*”
  „*Weight Cases…*”
- Изберете „Weight cases by” – измерете случаите по.
- Изберете честотна променлива.

Стойностите на честотната променлива се използват като тегла на случаи. Например, случай със стойност 3 за честотната променлива ще представлява три случая във файла с данни - тегла.

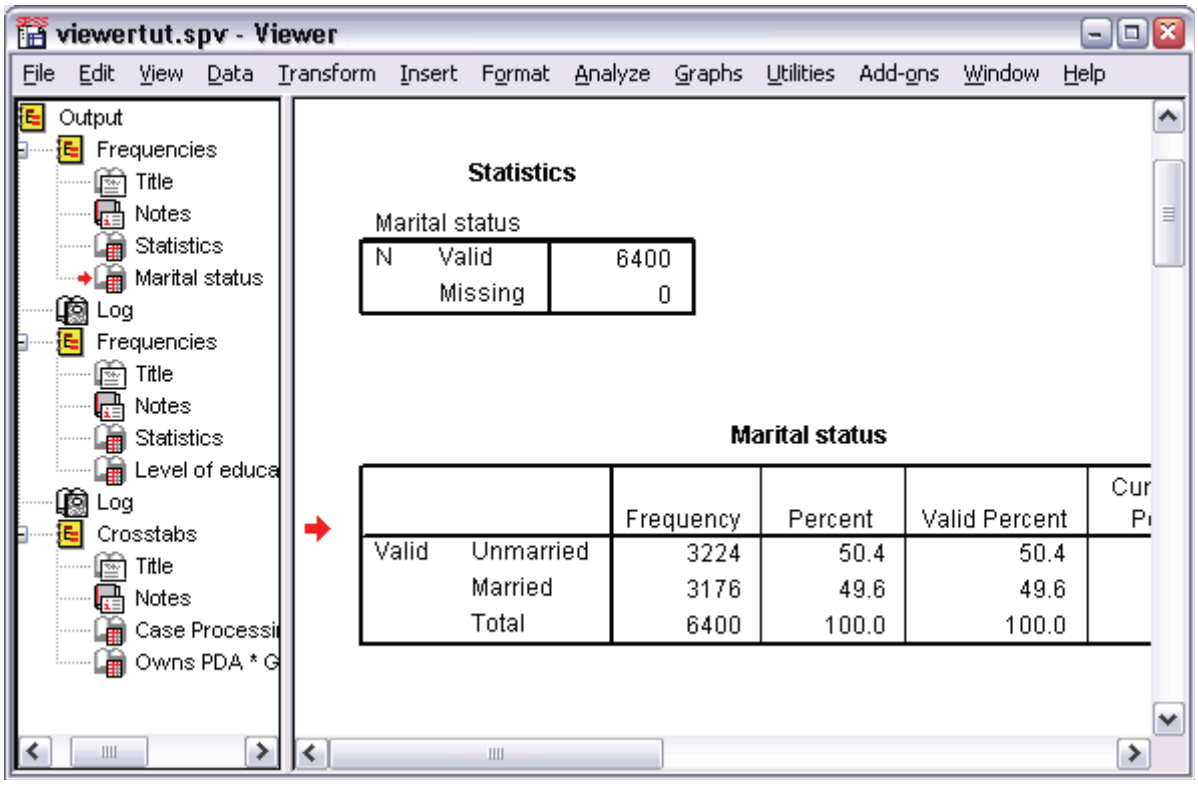
Глава 10. Работа с резултативния файл

Когато изпълните процедура, резултатите се показват във прозореца Изглед (Viewer). В този прозорец можете лесно да намерите резултатите, които искате да видите. Също така можете да модифицирате изходните резултати и да създадете документ, който да съдържа точно тези данни, които искате.

Изглед

Резултатите се показват във прозореца Изглед (Viewer). Вие можете да го използвате за да:

- Разглеждате данните
- Скривате или показвате таблици и диаграми
- Променяте реда на показване на данните, чрез преместване на определени елементи
- Премествате елементи между Изглед и други приложения



Фигура 10-1 Viewer

Изгледът е разделен на два панела:

- Левият панел съдържа компактен изглед на съдържанието.
- Десният панел съдържа статистически таблици, диаграми и текст.

Можете да кликнете върху елемент от компктнич изглед за да преминете директно към съответната таблица или диаграма. Можете да влачите дчсната страна на панела за да промените ширината му.

Показване и скриване на резултатите

В прозореца Изглед, можете по избор да показвате или скривате индивидуални таблици или резултати от цели процедури. Това действие е полезно когато искате да намалите видимата информация във панела с данните.

Скриване на Таблици и диаграми

- Кликнете два пъти иконата на елемента в панела с компактния изглед.

или

- Селектирайте елемента.
- От менюто изберете:

  Изглед

  Скриване

или

- Кликнете иконата затворена книга (Скриване) в лентата на компактнич изглед.

Отворената книга (Показване) става активната икона и указва, че елементът е скрит.

За да скриете резултатите от процедура:

- Щракнете кутията отляво на името не процедурата в компактния изглед.

Това действие скрива всички резултати от процедурата и скрива подструктурата на елемента в компактния изглед.

Преместване, изтриване и копиран на резултатите

Можете да пренаредите резултатите като копирате, преместите или изтриете елемент или група елементи.

Преместване на резултати в прозореца Изглед

- Изберете елементите в компактния изглед.
- Влачете до желаното място.

Скриване на резултати

- Изберете елементите в панела на компактния изглед или на данните.
- Натиснете клавиша **изтриване**.

или

- От менюто изберете :

  Редактиране

  Изтриване

Промяна на първоначалното подравняване

По подразбиране всички резултати са подравнени отляво. За да промените това направете следното:

- От менюто изберете:

  Редактиране

  Опции

  Изберете подпрозореца Изглед.

  В областта „Начално състочние на изходните данни“ (Initial Output State group), изберете типа елемент (например таблица,

диаграма, текстови данни).
Изберете подравняването, което желаете.

Промяна на пордравняването на изходни данни

- В панела с изглед з или панела с данните изберете елементите, които искате да подравните.
- От менто изберете:
  Формат
  Подравняване (ляво, дясно или центриране).

Панел Компактен изглед

Панелът компактен изглед показва съдържанието на прозореца Изглед. Можете да използвате този панел, за да се движите в панела с резултатите и да контролирате какво е показано. Повечето действия в този панел имат съответния резултат в панела с изходните данни.

- Избирането на елемент в панела за компактен изглед избира съответния елемент в панела с данните.
- Преместването на елемент в панела за компактен изглед премества елемент в панела с данните.
- Скриването на елемент в панела за комапктен изглед скрива съответните резултати в панела с данните.

Управление на панела за компактен изглед. Вие можете да:

- Показвате или скривате елемент в панела за компактен изглед
- Променяте индентацията на определени елементи
- Променяте размера на елементите в компактния изглед
- Променяте шрифта използван в компактния изглед.

За да скриете или покажете елемент в панела за за компактен изглед

щракнете върху кутийката отляво на елемента, който искате да скриете или да покажете.

Или

- Щракнете върху елемент в панела за компактен изглед.
- От менюто изберете:
  View
  Collapse (Скриване)

или

View
Expand (показване)

За да промените нивото на индентация

- Щракнете елемента в компактния изглед.
- Щракнете лявата стрелка на в лентата с инструменти за да изкачите в йерархията (да преместите елемента наляво).

Или

- Щракнете дясната стрелка в лентата с инструменти, за да понижите елемента в йерархията(да го преместите надясно).

Или

- Изберете в менюто:
  Edit
  Outline
  Promote *или* Demote

За да промените размера на елемент

- От менюто изберете:
  View
  Outline Size
  Изберете желания размер (Малък, Среден, Голям).

За да промените шрифта в компактния изглед

- От менюто изберете:
  View
  Outline Font...
  Изберете шрифт.

Добавяне на елементи към панела за компактен изглед

В панела за компактен изглед можете да добавяте елементи като заглавия, нов текст, диаграми или данни от други приложения.

За да добавите заглавие или текст

Текстови елементи, които не са свързани си таблица или диаграма могат да бъдат добавени към панела за компактен изглед.

- Щракнете върху таблицата, диаграмата или елемента, след който искате да добавите текстовия елемент.
- От менюто изберете:
  Insert
  New Title

или

  Insert
  New Text
- Щракнете два пъти върху новия обект.
- Въведете текста.

За да добавите текстов файл

- Щракнете върху таблицата, диаграмата или обекта, след който искате да добавите текст във панела с компактния изглед или панела с данните.
- От менюто изберете:
  Insert
  Text File...
  Изберете текстов файл.

Ако искате да редактирате текста, щракнете върху него два пъти с мишката.

Поставяне на обекти във прозореца Изглед

Обекти от други приложенич могат да се поставят в прозореца Изглед. Можете да използвате Paste After (Постави след) или Paste Special(Избор на поставяне). И двата вида поставяне позиционират новия елемент след текущия избран. Използвайте Paste Special, когато искате да контролирате формата на обекта, който поставяте.

Търсене и замяна на информация в прозореца Изглед

- За да намерите или замените информация в Изглед изберете от менюто:
  Edit
  Find (Търси)

или

  Edit
  Replace (Замени).

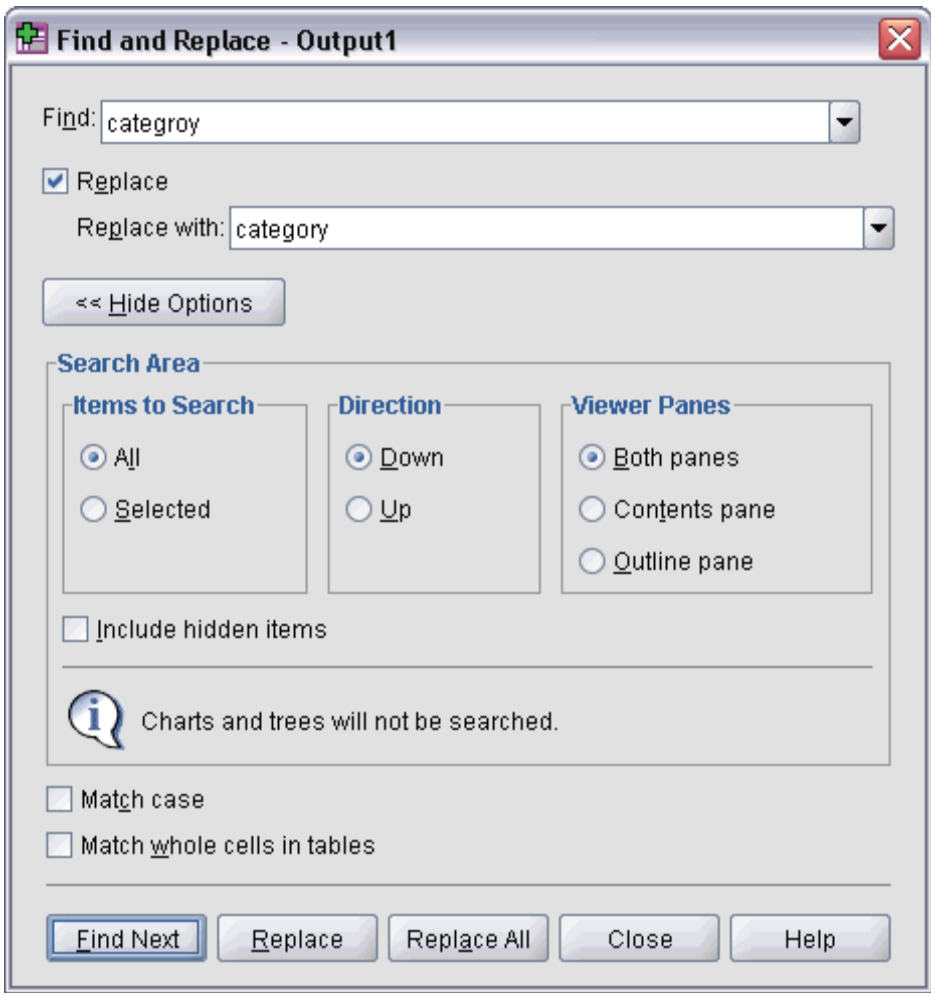


Figure 10-2 Find and Replace dialog box

Можете да използвате Търсене и замяна за да:

- Търсите в целия документ или в избрани елементи.

- Търсите преди или след текущата позиция.
- Търсите в двата панела заедно или поотделно панела за данни и за компактен изглед.
- Търсите скрити елементи. Това са елементи, скрити в панела за данни (например таблици от Notes, които са скрити по подразбиране) или скрити редове или колони в обобщаващи таблици.
- Задавате критерии за търсене със чувствителност за главни/малки букви.
- Задавате критерии за търсене в обобщаващи таблици по съдържанието на цели клетки.

Скрити елементи и слоеве от обобщаващи таблици

Слоевете под текущо видимия слой на многомерна обобщаваща таблица не се считат за скрити и се включват в областта на търсене дори когато скритите елементи са изключени от търсенето.

Скритите елементи включват скрити елементи в панела с данните (елементи с икона затворена книга в панела компактен изглед или елементи включени в скрити блокове на този панел) и редове/колони в обобщаващи таблици скрити по подразбиране (например празните редове и колони са скрити по подразбиране) или ръчно скрити чрез редактиране на таблицата. Скритите елементи се включват в търсенето само ако изберете ръчно Include hidden items. И в двата случая скритият елемент, който отговаря на критериите за търсене се показва, когато бъде намерен, но елементаъ след това се връща към първонаалното си състояние.

Копиране на резултати в други приложения

Обектите от резултативния файл може да се копират и поставят в други приложения, като например програми за текстообработка или електронни таблици. Можете да поставите обектите в различни формати. В зависимост от приложението, което получава данните, някои или всички от следните формати може да са достъпни:

Изображение (метафайл). Обобщаващите таблици, текстовата информация и диаграмите могат да се копират като метафайл изображения. Изображенията могат да се оразмеряват в приложението получател, като понякога са достъпни и ограничени възможности за редактиране. Обобщаващите таблици копирани като изображения запазват границите и характеристиките на шрифтовете си. Този формат е достъпен само под Windows операционни системи.

RTF (обогатен текстов формат). Обобщаващите таблици могат да се копират във обогатен текстов формат. В повечето случаи, това означава, че таблицата може да се редактира в приложението получател.

Забележка: Възможно е да има проблеми с показването на таблици с голяма ширина в Microsoft Word.

Растерни изображения (Bitmap). Обобщаващите таблици могат да се копират като растерни изображения.

Файлове за обмяна на Excel (BIFF). Съдържанието на таблицата може да се копира в електронна таблица и да запази числовата прецизност.

Текст. Съдържанието на таблицата може да се копира и като текст. Това е полезно при приложения като електронна поща, където може да се приема или предава само текст.

Ако приложението поддържа различни формати може да има опция Избор на Поставяне в менюто, която ви дава възможност да изберете формата или да покаже списък със възможни формати.

За да копирате и поставите Елементи от резултативния файл в друго приложение

- Изберете обектите в панела за компактен преглед или панела Изглед.
- От менюто Изглед изберете:
  Редактиране
    Копиране
- От менюто на приложението получател изберете:
  Редактиране
    Поставяне

или

  Редактиране
    Избор на поставяне...

Поставяне. Обектите се копират в системния буфер в няколко различни формата. Всяко приложение решава кой е предпочитания формат.
Избор на поставяне. Резултатите се копират в системния буфер в различни формати. Избор на поставяне ви да ва възможност да решите кой от достъпните в приложението получател формати да използвате.

Експорт на резултатите

Експорт на резултатите записва съдържанието на панела Изглед в HTML, текст, формати на Word/RTF, Excel или PowerPoint (изисква PowerPoint 97 или по късна версия) и PDF. Диаграмите може да се експортират в различни графични формати.
Забележка: Експортиране в PowerPoint формат е достъпно само под Windows операционни системи и не е активно във версията за студенти.
За да експортирате резултати

- Фокусирайте върху панела Изглед (щракнете с мишката в панела).
- От менюто изберете
  Файл
    Експортиране...
- Въведете име на файл (или префикс за диаграмите) и изберете желания формат.

Обекти за експортиране. Можете да експортирате всички обекти във Изглед, всички видими обекти или само избрани обекти.

Тип документ. Наличните опции са:

Word/RTF (*.doc). Обобщаващите таблици се експортират като Word таблици със запазени форматиращи атрибути, например граници на клетки, шрифтове и цветен фон. Текстът се експортира като форматиран RTF. Под операционни системи Windows диаграмите се включват в документа във формат EMF (разширен метафайл). Под други операцонни системи диаграмите се включват като PNG. Текстовите данни винаги се показват като шрифт с фиксирана ширина и се експортират със запазени форматиращи атрибути. Шрифтове с фиксирана ширина са необходими за правилно подравняване на текстовите данни разделени с интервал.
Забележка: Възможно е да има проблеми с показването на таблици с голяма ширина в Microsoft Word.
Excel (*.xls). Редовете, колоните и клетките на обобщаващите колони са експортират като редове, колони и клетки в Excel, със запазени форматиращи атрибути — например граници на клетки, шрифтове и едноцветен фон. Текстовите данни се експортират със запазени

атрибути на шрифта. Всеки ред в текстовите данни се представя като ред в Excel таблица, като цялото съдържание на реда е поместено в единична клетка. Диаграмите не се включват.

HTML (*.htm). Обобщаващите таблици се експортират като HTML таблици. Текстовите данни се експортират като предварително форматиран HTML. Връзка към диаграмите се включва във файла и е препоръчително да експортирате диаграмите във подходящ формат (например PNG или JPEG).

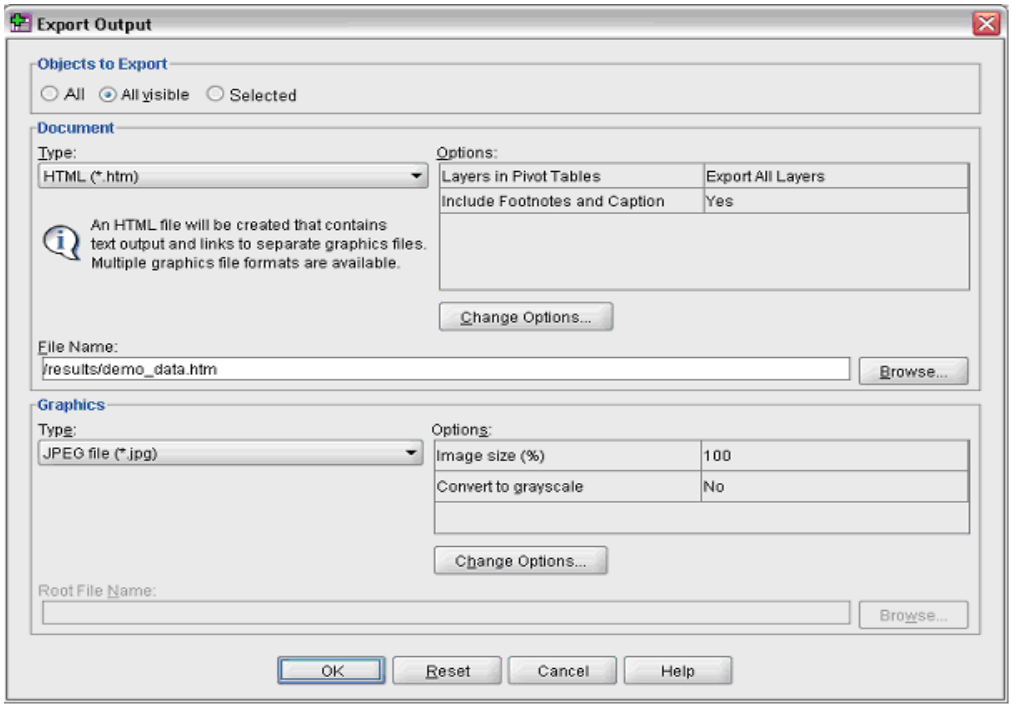
Portable Document Format (*.pdf). Резултатите се екпортират както се виждат във Прегледа за принтиране, със запазени форматиращи атрибути.

PowerPoint (*.ppt). Обобщаващите таблици се експортират като Word таблици и се запазват на отделни PowerPoint слайдове, със по една таблица на слайд.. Всички форматиращи атрибути се запазват — например граници на клетки, шрифтове и едноцветен фон. Текстовите данни се експортират като форматиран RTF. Диаграмите се еспортират в TIFF формат. За текстовите данни в SPSS винаги се използва шрифт с фиксирана ширина и се експортира със запазени форматиращи атрибути. Шрифтове с фиксирана ширина са необходими за правилно подравняване на текстовите данни разделени с интервал.

Забележка: Експортиране в PowerPoint формат е достъпно само под Windows операционни системи и не е активно във версията за студенти.

Текст (*.txt). Текстовите данни може да се представят като чист текст (plain text), UTF-8, and UTF-16. Обобщаващите таблици могат да се експортират с разделител интервал или табулация. Всички текстови данни се експортират с разделител интервал. За диаграмите се добавя ред с името на файла съдържащ еспортираната диаграма..

Без Формат (Само изображения). Наличните формати за експортиране включват: EPS, JPEG, TIFF, PNG, and BMP. Windows EMF (разширен метафайл) е също един от наличните формати.



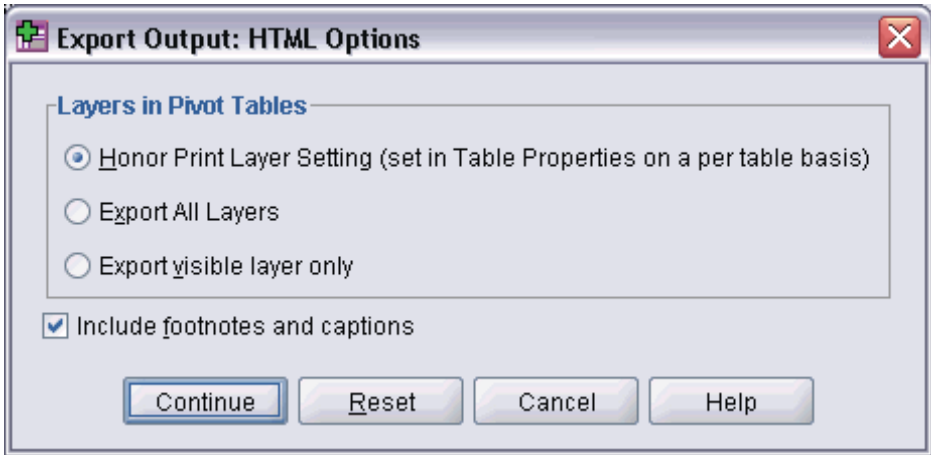
ФигураFigure 10-3 Export Output dialog box

Система за управляване на изходните резултати. Можете да експортирате всички данни или дефинирани от потребителя типове като текст, HTML, XML, or SPSS файлове с данни.

Опции за HTML, Word/RTF и Excel

Следните опции са достъпни при експортиране на обобщаващи таблици към Word/RTF, Excel и HTML:
Слоеве в обобщаващи таблици. По подразбиране добавяне или изклю1ване на слоеве на обобщаващи таблици се контролира от свойствата на съответната таблица. Можете да промените тази настройка и да добавите всички слоеве или да изключите всички слоеве, без текущо видимия слой.

Бележки под линия и обяснения. Включва или изключва бележките под линия и обясненията за обобщаващите таблици.
Note: За HTML, можете също така да контролирате и формата на изображенията за експортираните диаграми.
За да изберете опции за експортиране на HTML, Word/RTF, или Excel
Е Изберете HTML, Word/RTF или Excel като формат за експортиране.
Е Щракнете Промяна на Опции.



Фигура 10-4 HTML, Word/RTF, and Excel Export Output Options

Опции за PowerPoint

Следните опции са налични за PowerPoint:
Слоеве в обобщаващи таблици. По подразбиране включването или изключването на слоевете на обощаващите таблици се контролира от свойствата на съответната таблица. Можете да промените това и да включите всички слоеве или да изключите всички без текущия слой. . За повече информация моля прегледайте

Включване на бележки под линия и обяснения. Включва или изключва бележките под линия и обясненията за обобщаващите таблици.

Use Viewer outline entries as slide titles. Включва Заглавие на на всеки слайд създаден при експортирането. Всеки слайд съдържа единичен елемент експортиран от прозореца Изглед. Заглавието се състои от името на елемента в панела за компактен изглед.

За да изберете опции за експортиране в PowerPoint

- Изберете PowerPoint като формат за експортиране.
  Щракнете Промяна на Опции.

Забележка: Експортиране в PowerPoint формат се поддържа само под операционни системи Windows.

PDF Опции

Следните опции са налични за PDF:

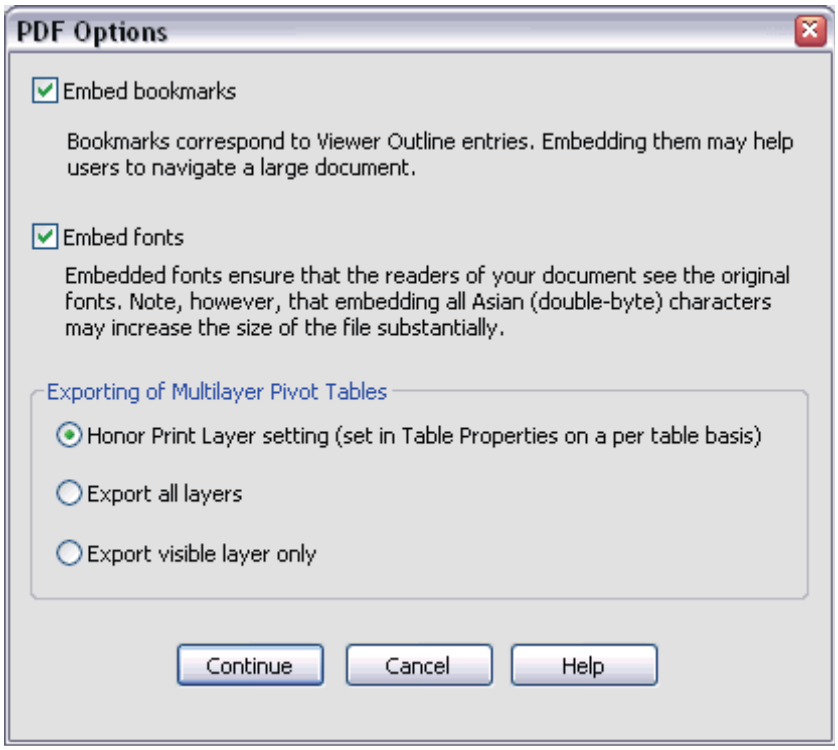
Embed bookmarks. Тази опция включва в PDF документа връзки, които отговарят на елементите в компактния изглед. Тези връзки улесняват придвижването в документи с голям брой елементи за експортиране.

Embed fonts. Включването на шрифтове осигурява еднакъв вид на документа на всички компютри. В противен случай, ако някои от използваните шрифтове не са налични на компютъра, който чете/отпечатва PDF документа, ще има заместване на шрифтове, което може да доведе до нежелани резултати.

Layers in pivot tables. По подразбиране включването или изключването на слоевете на обощаващите таблици се контролира от свойствата на съответната таблица. Можете да промените това и да включите всички слоеве или да изключите всички без текущия слой. .

За да изберете опции за експортиране в PDF

- Изберете Portable Document Format as като формат за експортиране.
  Щракнете Промяна на Опции.



Фигура 10-5 PDF Options dialog box

Други настройки за експортиране в PDF

Настройки за страници /Свойства на страници. Размерът на страниците, ориентацията, полетата, съдържанието, показването на заглавни и крайни части на страниците се контролира от настройките на страници и опциите за свойства на страници.
Свойства на таблици /Изглед на таблици. Мащабирането на широки или дълги таблици и отпечатването на слоеве се контролира от свойствата на съответната таблица. Тези свойства могат да бъдат запазени в Изглед на Таблици.
Текущ принтер/Принтер по подразбиране. Разделителната способност (DPI точки за инч) на PDF документа е настройка за текущия принтер или принтера по подразбиране (тази настройка може да се промени в Page Setup). Максималната разделителна способност е 1200 DPI. Ако настройката на принтера е над тази стойност, разделителната способност на PDF документа ще бъде 1200.
Забележка: Документите с висока разделителна способност могат да изглеждат неприемливо когато се отшечатват на принтери с ниска разделителна способност.

Опции за експорт на диаграми

При HTML, текст или експорт само на диаграми, можете да изберете графичен формат като за всеки формат може да има специфични настройки:
За да изберете графичен формат и настройки за експортираните диаграми:
Е Изберете HTML, Text или None (Graphics only) като тип документ.
Е Изберете графичен формат от падащото меню.
Е Щракнете Промяна Опции за да промените настройките за избрания формат.

Опции за експортиране в JPEG
Image size. Задаване на размера спрямо началното изображение – до 200%.
Convert to grayscale. Конвертира в черно-бяла графика.

Опции за експортиране в BMP
Image size. Задаване на размера спрямо началното изображение – до 200%.
Compress image to reduce file size. Използва компресия без загуба, за да намали размера на файла без това да се отрази на качеството.

Отпечатване на прозореца Изглед
Има две възможности за отпечатване на съдържанието на прозореца Изглед:
All visible output. Отпечатва само елементи, които са текущо показани в панела с данните. Скритите елементи (елементи, които имат икона затворена книга в панела за компактен изглед или елементи в скрити слоеве) не се отпечатват.
Selection. Отпечатва само елементи, които са избрани в панела за компактен изглед и/или панела с данни.

За да отпечатате Резултати и диаграми

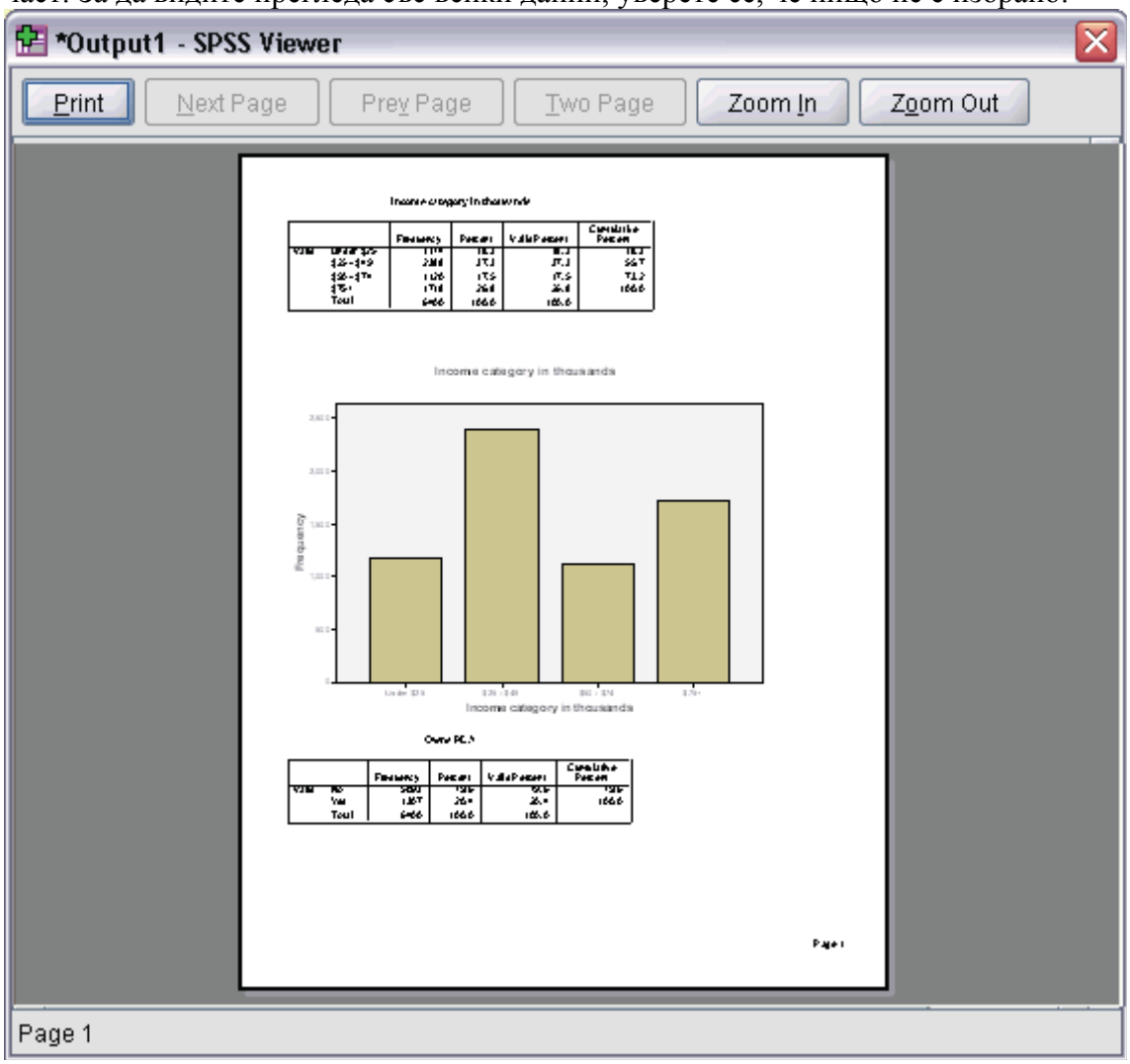
- Уверете се, че Изглед е активния прозорец (Щракнете някъде в прозореца)..
- От менюто изберете:
  File
  Print...
- Изберете опциите които желаете.
- Щракнете ОК за печат.

Преглед отпечатване

Прегледът на отпечатване ви показва какво ще се отпечата на всяка страница от документи с резултатите. Добре е да се види прегледа преди самото отпечатване, тъй като може да се видят елементи, които е възможно да не се виждат в панела с данните, например:

- Page break – нова страница.
- Скрити слоеве или обобщаващи таблици
- Прекъсвания в таблици с голяма ширина
- Заглавни и крайни части, които се отпечатват на всяка страница

Ако част от резултатите е избрана в панела с данните, прегледът ще покаже само тази част. За да видите прегледа със всики данни, уверете се, че нищо не е избрано.

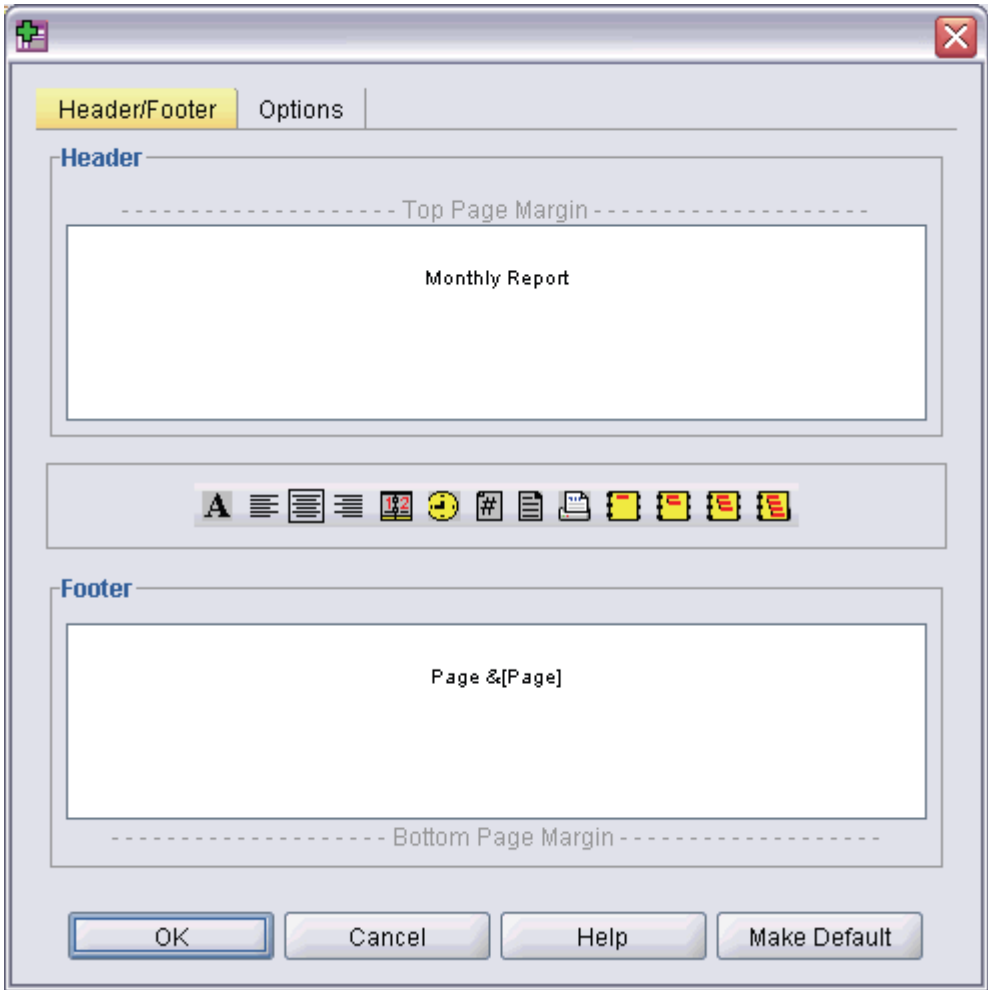


Фигура 10-7 Print Preview

Свойства на страница: Заглавни и крайни части

Заглавните и крайни части са информация, която се отпечатва в горното и долно поле на всяка страница. Можете да използвате произволно въведен текст като горно и долно поле. Можете също да използвате лентата в средата на прозореца за да добавите:

- Дата и час
- Номера на страници
- Име на файл
- Заглавни етикети
- Заглавия и подзаглавия на страници.



Фигура 10-8 Page Attributes dialog box, Header/Footer tab

Make Default задава текущите настройки като настройки по подразбиране за нови документи.

- Заглавните етикети показват първо, второ, трето и четвърто ниво на индентация за първия елемент на всяка страница.
- Page titles and subtitles отпечатва заглавия и подзаглавия на страници. Те могат да се създадат от New Page Title в менюто Insert или с командите TITLE и SUBTITLE. Ако не сте създали заглавич и подзаглавия, тази настройка нчма ефект.

Забележка: Настройките за шрифт за нови заглавия и подзаглавия се контролират от раздела Изглед във прозореца Опции(Опции в меню Редактиране). Настройките на шрифта за съществуващи заглавия и подзаглавия могат да се променят като се редактират заглавията в прозореца Изглед.

За да видите как изглеждат заглавните и крайните части изберете Print Preview от менюто Файл.

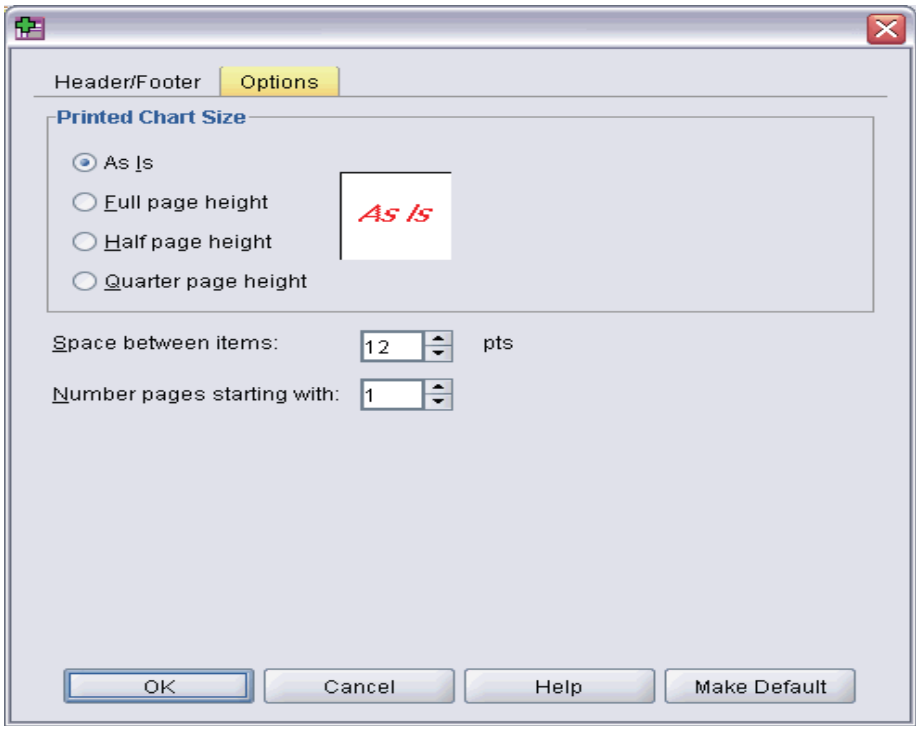
За да добавите Заглавни и крайни части

- Уверете се, че Изглед е активния прозорец (Щракнете някъде в прозореца).
- От менюто изберете:
  File
    Page Attributes...
- Щракнете раздела Header/Footer.
- Въведете заглавната и крайна части, които искате да се показват на всяка страница.

Свойства на страница: Опции

Тази диалогова кутия контролира размера на отпечатаните диаграми, разстоянието между отпечатаните елементи и номерата на страници.

- **Printed Chart Size.** Контролира размера на отпечатаната диаграма спрямо определения размер на страница. Съотношението ширина височина, не се влияе е от размера на диаграмата. Размерът на диаграмата се ограничава и от ширината и височината на страницата. Например когато диаграмата достигне лявото и дчсно поле на страницата не можете да увеличите повече размера, така че да запълните страницата и по височина.
- **Space between items.** Контролира разстоянието между отделните елементи. Всяка обобщаваща таблица, диаграма или текстово поле са отделен елемент. Тази настройка не променя показването на елементи в прозореца Изглед.
- **Number pages starting with.** Задава последователни номера на страници, като започва от въведеното число.
- **Make Default.** задава текущите настройки като настройки по подразбиране за нови документи



Фигура 10-9 Page Attributes dialog box, Options tab

За да промените Размера на отпечатаните диаграми, номера на страници и разстояние между отделните елементи:

- Уверете се, че Изглед е активния прозорец (Щракнете някъде в прозореца).
- От менюто изберете:
  File
  Page Attributes...
- Щракнете Опции.
- Променете настройките и настиснете OK.

Запазване на резултати

Съдържанието на прозореца Изглед може да бъде запазено в документ. Запазеният документ включва и двата панела (компактен изглед и данни).

За да запазите резултатите в документ

- От прозореца изглед изберете:
  File
  Save
- Въведете име на документ и натиснете Save.

За да запазите резултатите в различен формат (например HTML или текст), използвайте функята експорт в менюто File.

Глава 11. Работа с таблици

Обобщаващи таблици

Много от резултатите могат да бъдат представени в обобщени интеркативно таблици. Това означава, че можете да пренареждате редовете, колоните и слоевете.

Манипулиране на Обобщаващи таблици

Възможни са следните действия:

- Транспониране на редове и колони
- Преместване на редове и колони
- Създаване на многомерни слоеве
- Групиране и разделяне на редове и колони
- Показване и скриване на редове, колони и допълнителна информация
- Ротация на етикети ан редове и колони
- Търсене на дефиниции на термини

Активиране на обобщаваща таблица

Преди да можете да работите или да променяте обобщаваща таблица, първо трябва да я активирате. За да активирате таблицата:

- Щракнете два пъти в таблицата.

Или

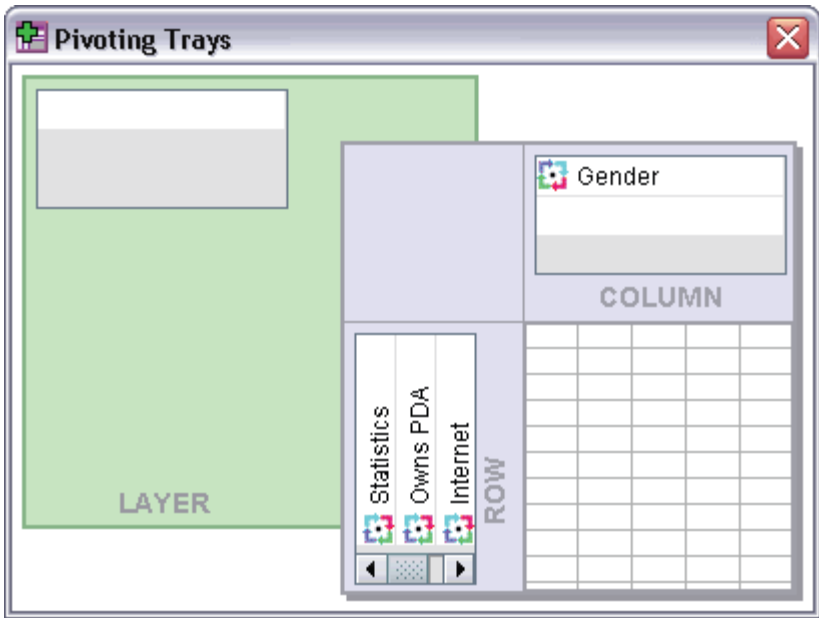
- Щракнете с десен бутон на мишката и от контекстното меню изберете Edit Content.
- От подменюто изберете In Viewer или In Separate Window.

По подразбиранем двойното щракване ще активира таблиците, освен ако не са много големи, в прозореца Изглед. Ако искате няколко таблици да са активирани по едно и също време, трябва да ги активирате в отделни прозорци.

Манипулациис таблицата

- Активирайте таблицата.
- От менюто изберете:
  Pivot
  Pivoting Trays

Таблицата има три измерения – редове, колони и слоеве. Измерението може да съдържа множество елементи (или да не съдържа нито един). Можете да промените организацичта на таблицата като местите елементи, между или вътр в отделните измерения. За да преместите елемент, просто го влачете с мишката до желаното място.



Фигура 11-1 Pivoting trays

Промяна на реда на показване на елементи в едно измерение

За да промените реда на показване на елементи в измерение на таблица (ред, колона, слой): Ако плотът не е активиран, от менюто Pivot Table изберете:

- Pivot
- Pivoting Trays

Влачете елементите от съответното измерение в плота.

Местене на редове и колони

- В таблицата (не в плота), щракнете етикета за реда или колоната, които искате да преместите.
- Влачете етикета до новата позиция.
- От контекстното меню изберете Insert Before (Вмъкни преди) или Swap (смени).

Забележка: Уверете се, че опцията Drag to Copy в меню Edit е изключена. Ако не е, изключете я.

Транспониране на редове и колони

Ако просто искате да обърнете редовете и колоните има проста алтернатива на използването на плота:

- Изберете от менюто

Pivot

Transpose Rows and Columns

Ефектът е същият като преместване на всички редове в измерението на колоните и всички колони в измерението на редовете.

Групиране на редове и колони

- Селектирайте етикетите на редовете или колоните, които искате да групирате(щракнете и влачете или използвайте Shift+ щракване за да селектирате няколко етикета).

- От менюто изберете:
  Edit
    Group

Автоматично се вмъква етикет за група. Щракнете два пъти върху етикета за да промените текста.

		Column Group Label		
		Female	Male	Total
Row Group Label	Clerical	206	157	363
	Custodial		27	27
	Manager	10	74	84

Фигура 11-2 Row and column groups and labels

Забележка: За да добавите редове или колони към съществуваща група, първо трябва да разделите елементите. Които са понастоящем в групата. Тогава можете да създадете нова група, която включва и допълнителните елементи.

Разделяне на редове и колони

- Щракнете където и да е в етикета на групата, съдържаща колоните и редовете които искате да разделите.
- От менюто изберете:
  Edit
    Ungroup

Разделянето автоматично премахва етикета на група.

Ротация на етикети на редове или колони

Можете да показвате етикетите на вътрешните колони или на редовете във външната колона изписани вертикално или хоризонтално.

- От менюто изберете:
  Format
    Rotate Inner Column Labels
  или
    Rotate Outer Row Labels

	Frequency	Percent	Valid Percent	Cumulative Percent
Did not complete high school	1390	21.7	21.7	21.7
High school degree	1936	30.3	30.3	52.0
Some college				
College degree				
Post-undergraduate deg				
Total				

	Frequency	Percent	Valid Percent	Cumulative Percent
Did not complete high school	1390	21.7	21.7	21.7
High school degree	1936	30.3	30.3	52.0
Some college	1360	21.3	21.3	73.2
College degree	1355	21.2	21.2	94.4
Post-undergraduate degree	359	5.6	5.6	100.0
Total	6400	100.0	100.0	

Фигура 11-3 Rotated column labels

Единствено етикетите на вътрешните колони и на редовете във външната колона могат да хоризонтално-вертикалната си ориентация.

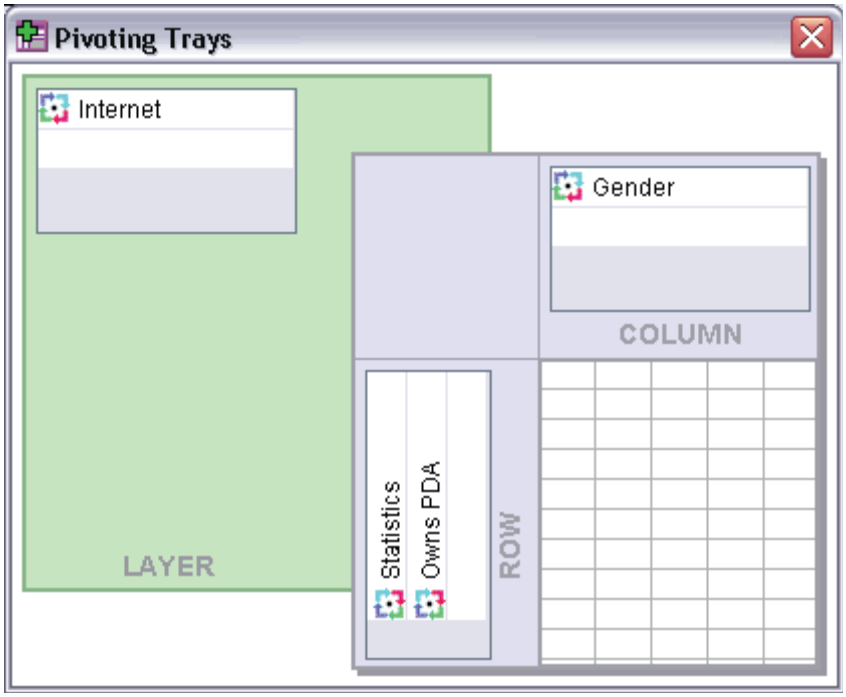
Работа със слоеве

Можете да покажете отделна двумерна таблица за всяка категория или комбинация от категории. Таблицата може да се представи като наредени един под друг слоеве, като само най-горния е видим.

Създаване и показване на слоеве

За да създадете слой:

- Активирайте таблицата.
- Ако плота не е показан, изберете от менюто Pivot Table:
  Pivot
  Pivoting Trays
- Влачете елементите от измерението на редовете или колоните в измерението на слоевете.



Фигура 11-4 Moving categories into layers

Преместването на елементи в измерението слой създава многомерна таблица, но се показва само двуизмерен отрязък. Видимата таблица е таблицата в наи-горния слой. Например, ако имате категорийна променлива да/не в измерението на слоевете, тогава многомерната таблица има два слоя: един за категорията да и един за категорията не.

Internet: Yes

			Gender		
			Female	Male	Total
Owns PDA	No	Count	513	512	1025
		Expected Count	512.5	512.5	1025.0
Total					

Internet: No

			Gender		
			Female	Male	Total
Owns PDA	No	Count	1897	1962	3859
		Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	49.2%	50.8%	100.0%
	Yes	Count	327	323	650
		Expected Count	320.6	329.4	650.0
		% within Owns PDA	50.3%	49.7%	100.0%
Total		Count	2224	2285	4509
		Expected Count	2224.0	2285.0	4509.0
		% within Owns PDA	49.3%	50.7%	100.0%

Figure 11-5 Categories in separate layers

Промяна на показвания слой

- Изберете категория от падащич списък със слоевете (в таблицата не в плота).

Internet | No

No
Yes

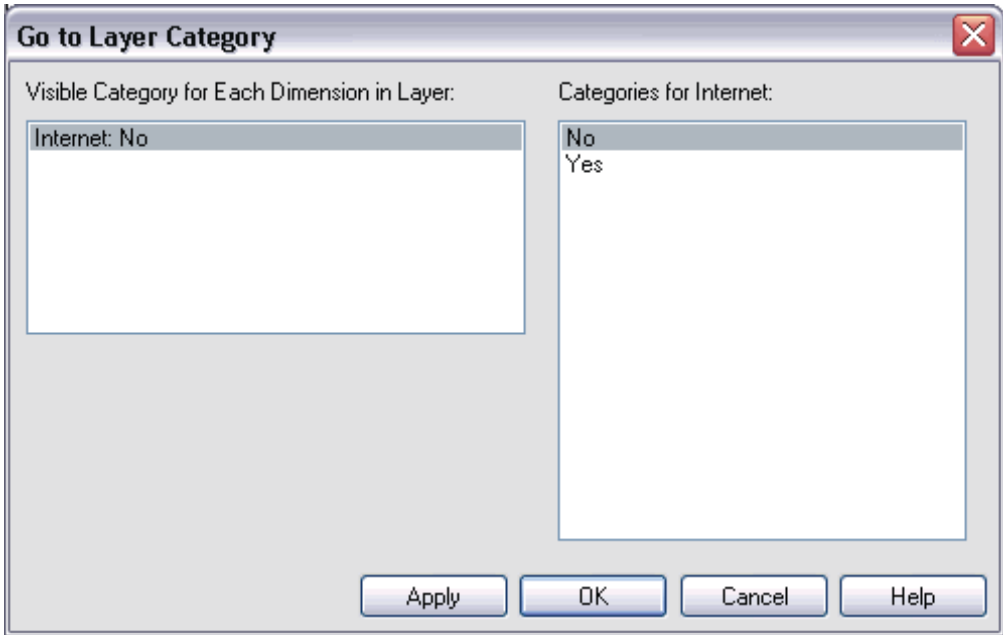
			Gender		
			Female	Male	Total
Owns PDA	No	Count	1897	1962	3859
		Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	49.2%	50.8%	100.0%
	Yes	Count	327	323	650
		Expected Count	320.6	329.4	650.0
		% within Owns PDA	50.3%	49.7%	100.0%
Total		Count	2224	2285	4509
		Expected Count	2224.0	2285.0	4509.0
		% within Owns PDA	49.3%	50.7%	100.0%

Категория Go to Layer (Премини към слой)

Категорията Go to Layer ви позволява да сменяте слоеве в таблицата. Тази диалогова кутия е особено полезна когато имате много слоеве или избраният слой има много категории.

- От менюто изберете:
  Pivot
  Go to Layer...

- В списъка Visible Category (Видима Категория), изберете измерение на слой. Списъкът с категориите ще покаже всички категории за избраното измерение.
- В списъка с категориите изберете желаната и натиснете OK или Apply.



Фигура 11-7 Go to Layer Category dialog box

Показване и скриване на елементи

Много видове клетки могат да се скрият, например:

- Етикети на измерение
- Категории, включително етикета на клетката и клетките с данни в ред или колона
- Етикети на категории(без да скривате летките с даннни)
- Бележки под линич, заглавич и обяснения

Скриване на редове и колони в таблица

- щракнете с десния бутон върху етикета на колонаа или реда, които искате да скриете.
- От контекстното меню изберете:
  Select
  Data and Label Cells
- Щракнете с десния бутон отново и от контекстното меню изберете Hide Category.

Или

- От менюто View изберете Hide.

Показване на скрити редове и колони в таблица

- Щракнете с десния бутон на мишката върху избрана колона или ред в същото измерение, като скрития ред или колона, които искате да покажете.
- От контекстното меню изберете:
  Select
  Data and Label Cells
- От менюто изберете:

View
Show All Categories in [име на измерението]

или

- За да покажете всички скрити редове и колони в активирана обобщаваща таблица, от менюто изберете:
  View
  Show All

Така ще покажете всички скрити редове и колони в таблицата. (Ако опцията Hide empty rows and columns е включена в свойства на таблицата, редове или колони, които са изцяло празни няма да се покажат.)

Скриване и показване на етикети на измерение

- Изберете етикета на измерение или който и да е етикет на категория вжеланото измерение.
- От менюто View или пт контекстното меню изберете Hide Dimension Label или Show Dimension Label.

Скриване и показване на заглавия на таблици

За да скриете заглавие:

- Изберете заглавието.
- От менюто View изберете Hide.

За да покажете скрити заглавия:

- От менюто View изберете Show All.

Изглед на таблица - TableLooks

Изгледът на таблица е набор от свойства, които определят как изглежда таблицата. Можете да изберете предварително дефиниран изглед или да създадате ваш собствен.

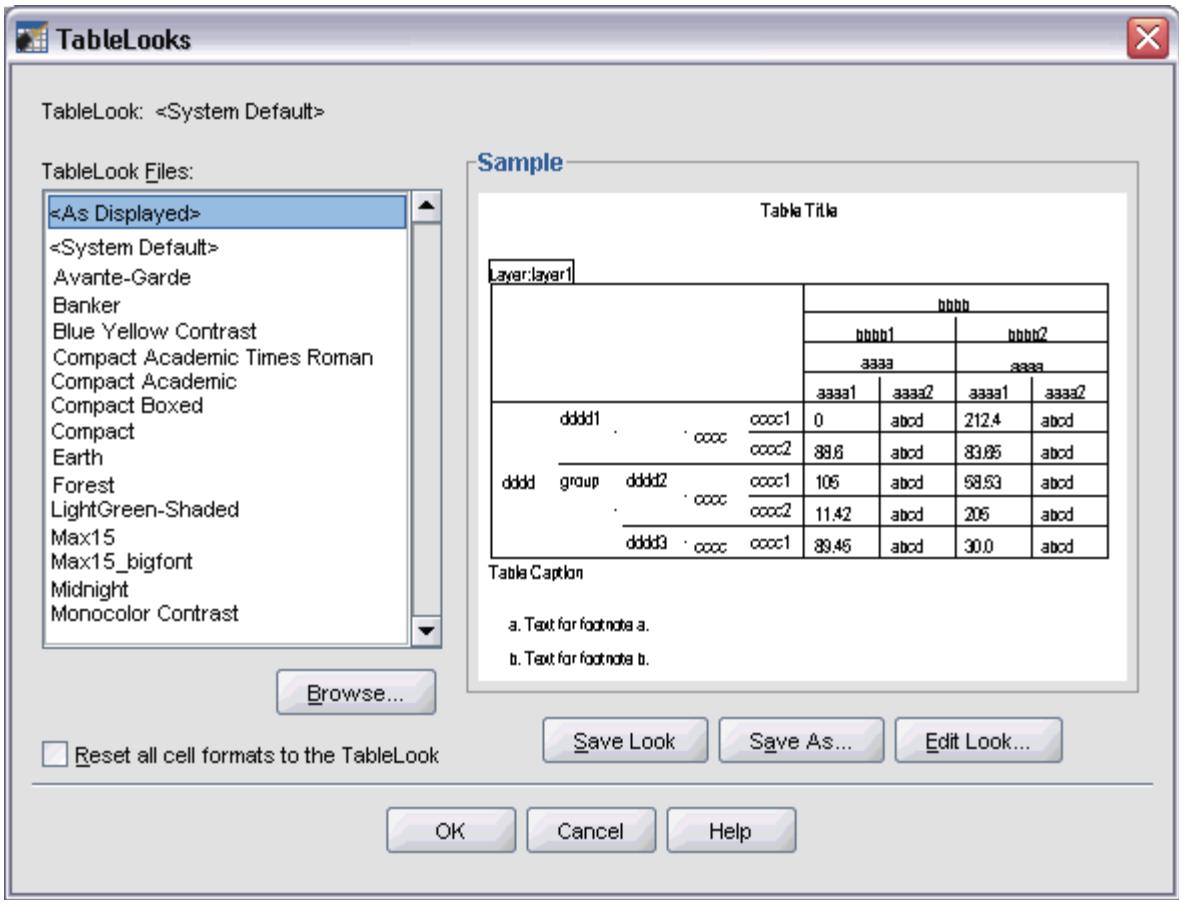
- Преди или след като приложите изглед на таблица, можете да променяте формата на отделните клетки или на групи от клетки, като използвате свойствата им. Променените клетки ще запазят формата си, дори и когато приложите нов изглед.
- Също така можете да върнете всички клетки към формата дефиниран в изгледа на таблица.

Това отменя редакциите на формата на отделните клетки. Ако изберете As Displayed в списъка TableLook Files, всички индивидуално редактирани клетки получават текущите свойства на таблицата.

Забележка: Изгледи на таблици създадени в ранни версии на SPSS не могат да се използват във версия 16.0 или по-късна.

За да приложите или запазите изглед на таблица - TableLook

- Активирайте обощаща таблица.
- От менюто изберете :
  Format
  TableLooks...
- Изберете изглед от списъка с файлове. За да изберете файл от друга директория натиснете Browse.
- Натиснете OK за да приложите изгледа към избраната таблица.



Фигура11-8 TableLooks dialog box

За да редактирате или създадете изглед на таблица

- В прозореца TableLooks, изберете т списъка желания изглед.
- Изберете Edit Look.
- Променете свойствата на атрибутите, които желаете и потвърдете с OK.
- Изберете Save Look за да запазите редактираният изглед, или Save As за да запазите нов изглед.

Редактирането на изглед на таблица се прилага само върху избраната таблица. Редактираният изглед няма да се приложи автоматично върху други таблици, които го използват, докато не изберете тези таблици и приложите отново съответният изглед.

Свойства на таблици

Свойствата на таблици ви позволяват да конфигурирате основните свойства на таблици, да приложите стилове за различни части от таблицата и да запишете тези конфигурации като изглед на таблица. Можете да:

- Контролирате основни свойства като скриване на празни редове и колони или опции за печат.
- Контролирате формата и положението на бележките под линия.
- Определяте конкретни форматиране за клетките в областта за данни, за етикети на редове и колони, както и за други области от таблицата.
- Контролирате ширината и цвета на линиите, които оформят границите на всяка област от таблица.

За да промените свойствата на таблица

- Активирайте таблицата.
- От менюто изберете:

  Format

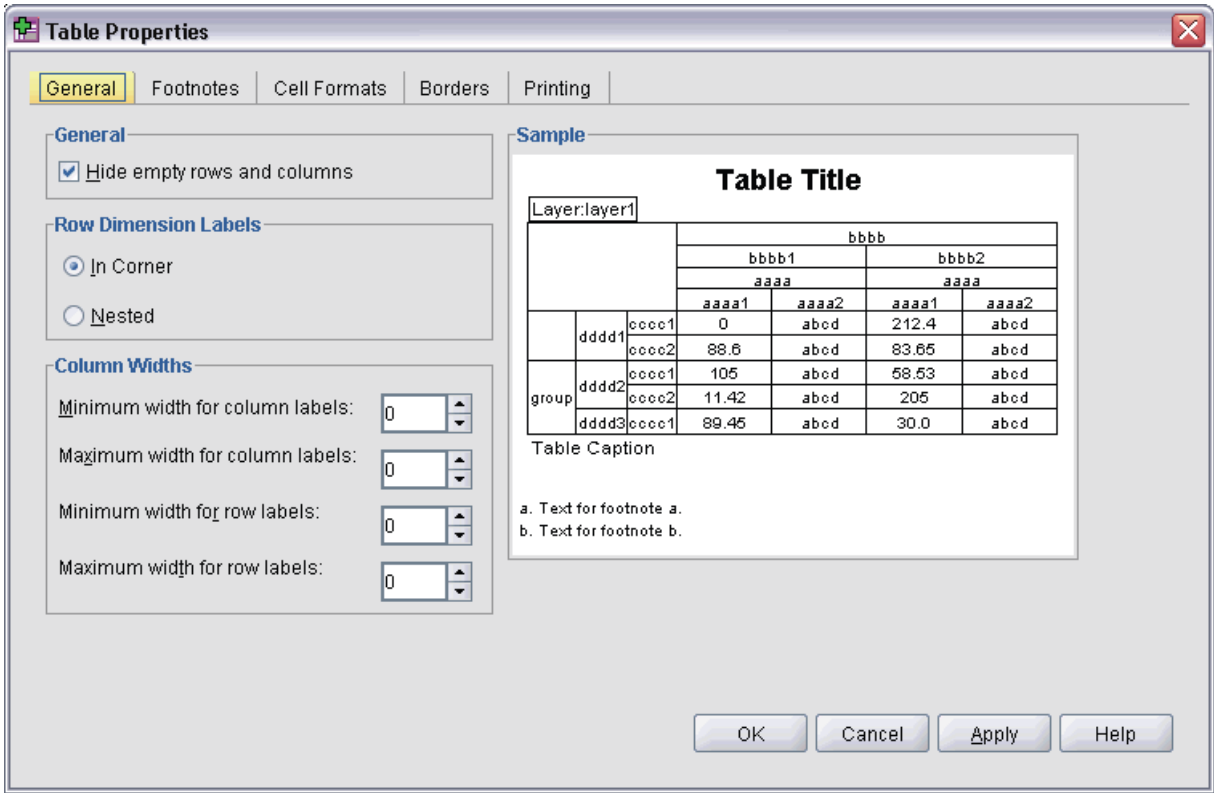
  Table Properties...
- Изберете подпрозорец(General, Footnotes, Cell Formats, Borders, or Printing).
- Изберете опциите, които желаете.
- Потвърдете с OK.

Променените свойства се прилагат към активираната обобщаваща таблица. За да приложите нови свойства към изглед на таблица, вместо само към таблицата, редактирайте изгледа (Меню Format, TableLooks).

Свойства на таблица: подпрозорец General (Основни свойства)

Някои свойства се отнасят за таблицата като цяло. Можете да:

- Покажете или скриете празните редове и колони. (Редове и колони, които нямат данни в клетките си.)
- Конфигурирате позицията на етикетите на редове: в горния ляв ъгъл или да са вложени.
- Конфигурирате максимална и минимална ширина на колона (в точки).

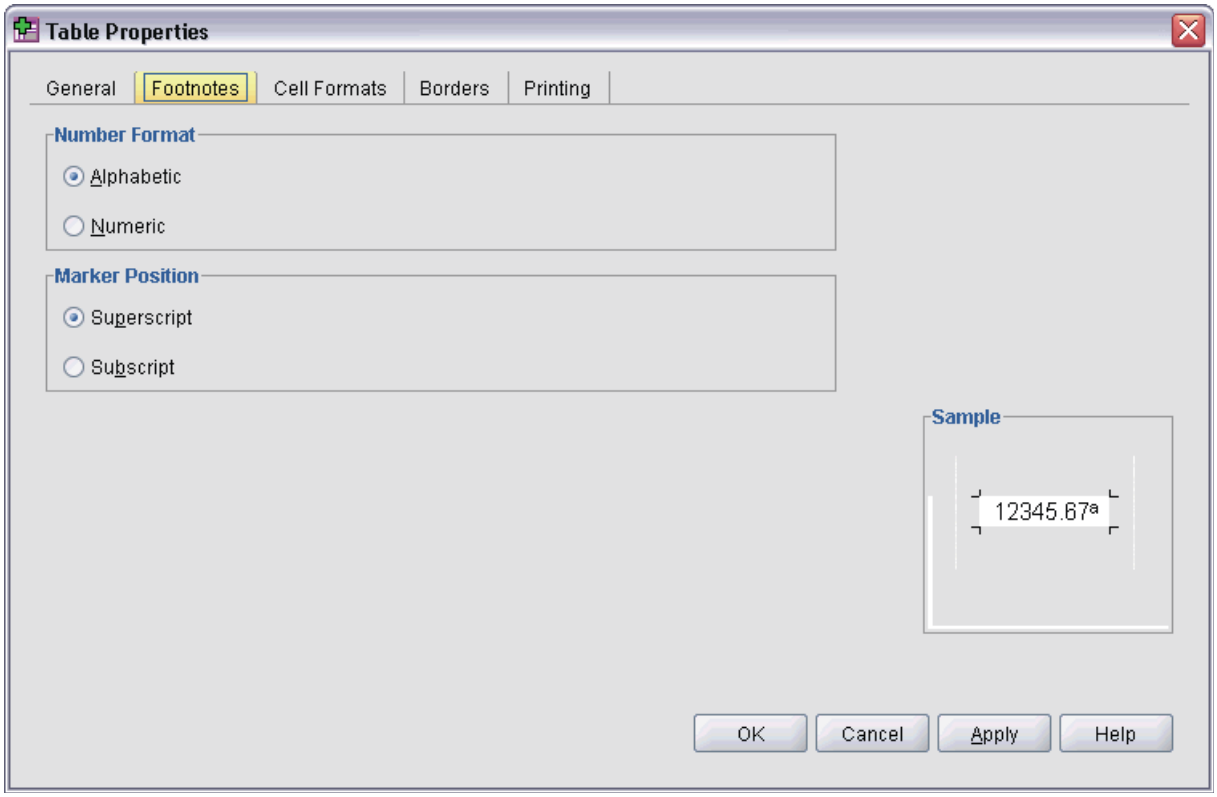


Фигура 11-9 Table Properties dialog box, General tab

За да промените свойствата на таблица:

- Изберете подпрозореца General.
- Изберете желаните опции.
- Потвърдете с OK или Apply.

Подпрозорец: Бележки под линия
Свойствата на бележките под линия включват стила и положението по отношение на текста.
- Стилът включва номера (1, 2, 3, ...) или букви(a, b, c, ...).
Маркерите за бележки под линия могат да са в ^{горно} или $_{\text{долно}}$ положение спрямо текста.



Фигура 11-10 Table Properties dialog box, Footnotes tab

За да промените свойствата на маркера за бележки под линия:
- Изберете подпрозорец Footnotes.
- Изберете подреждане по номера или букви.
- Изберете позиция.
- Потвърдете с OK или Apply.

Свойства на таблица: Подпрозорец Cell Formats(Формат на клетки)

По отношение на формата, клетките се делят на няколко основни елемента: заглавие, слоеве, ъглов етикет, данни, етикет на ред, етикет на колона, пояснения и бележки под линия. За всеки от тези елементи, можете да променяте съответните форматирания за клетка.
Форматиранията включват атрибути на текста (шрифт, големина, размер и цвят), хоризонтално и вертикално подравняване, фон и вътрешни полета.

Title

Layers

Corner labels	Column labels
Row labels	Data

Caption
Footnotes

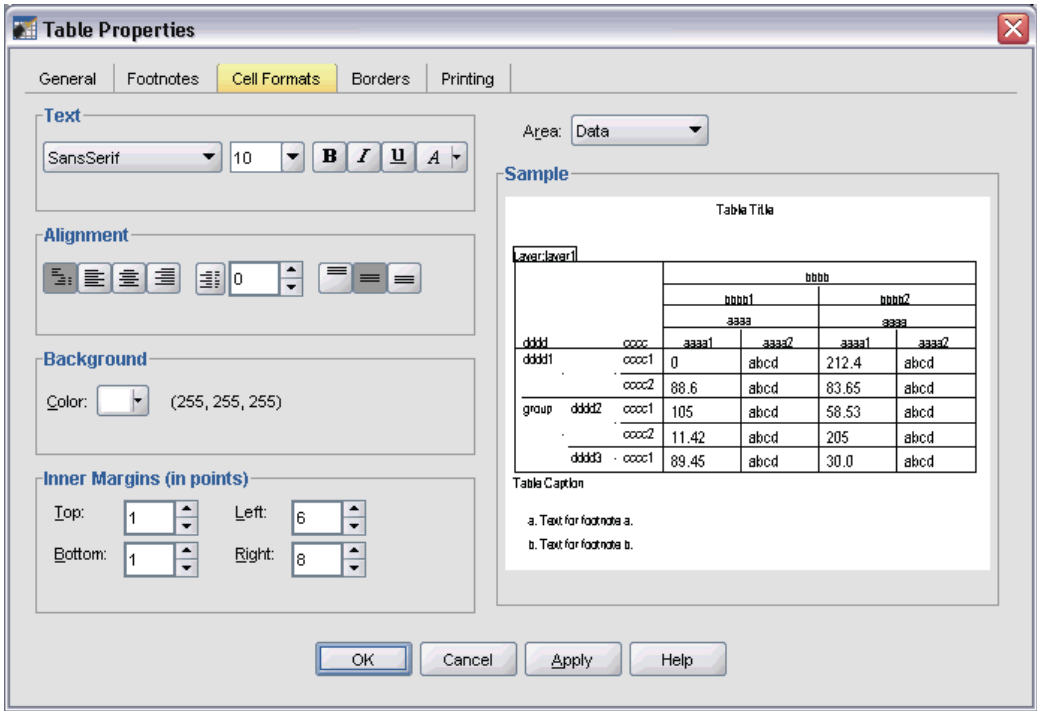
Фигура 11-11 Areas of a table

Форматирането на клетки се прилага към области (категории от информация). То не е атрибут на отделни клетки. Това е важна разлика, която трябва да се вземе под внимание когато обобщавате таблица.

Например:

- Ако зададете удебелен шрифт като форматиране за етикетите на колони, етикетите ще са удебелени, без значение какво е показано в измерението на колоните. Ако преместите елемент от измерението на колоните в друго измерение, той нчма да запази удебеления шрифт.

- Ако просто удебелите етикетите на колони като изберете клетките в активирана обобщаваща таблица и натиснете бутона Bold от лентата с инструменти, съдържанието на кетките ще остане удебелено, без значение в какво измерение ги преместите, а новите елементи преместенив измерението на колоните няма да са удебелени.



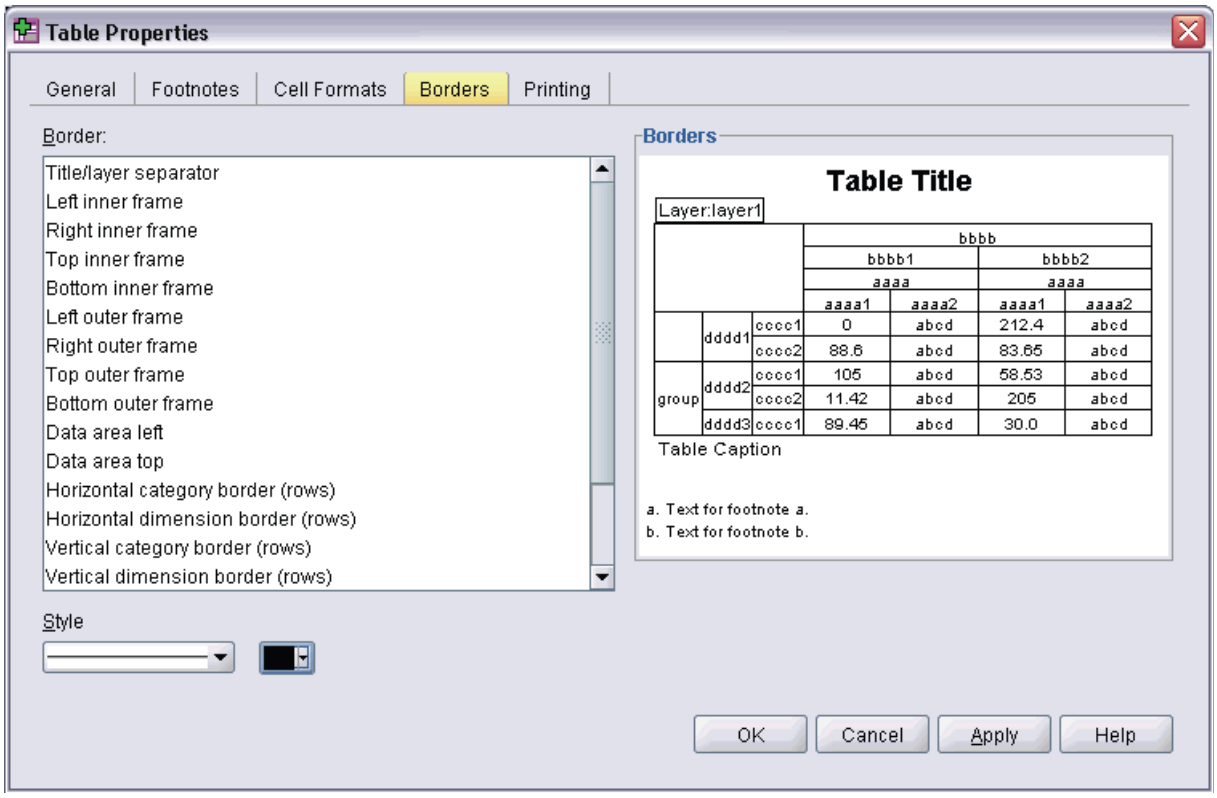
Фигура 11-12 Table Properties dialog box, Cell Formats tab

За да промените форматирането на клетки:

- Изберете подпрозорец Cell Formats.
- Изберете област от падащия списък или щракнете в област от показания модел(Sample).
- Изберете желаните атрибути на област. Изборът ви се прилага към модела.
- Потвърдете с OK или Apply.

Свойства на таблици: Подпрозорец Borders (граници на клетки)

Можете да изберете стил и цвят на линия за всяка граница. Ако изберете None за стил, няма да има ограничителна линия на съответната позиция.



Фигура11-13 Table Properties dialog box, Borders tab

За да промените границите на клетка:

- Изберете подпрозореца Borders.
- Изберете позиция, като щракнете върху името или като щракнете върху линия в показания модел.
- Изберете стил (или Без стил - None).
- Изберете цвят.
- Потвърдете с OK или Apply.

Свойства на таблици: Подпрозорец Printing (Печатане)

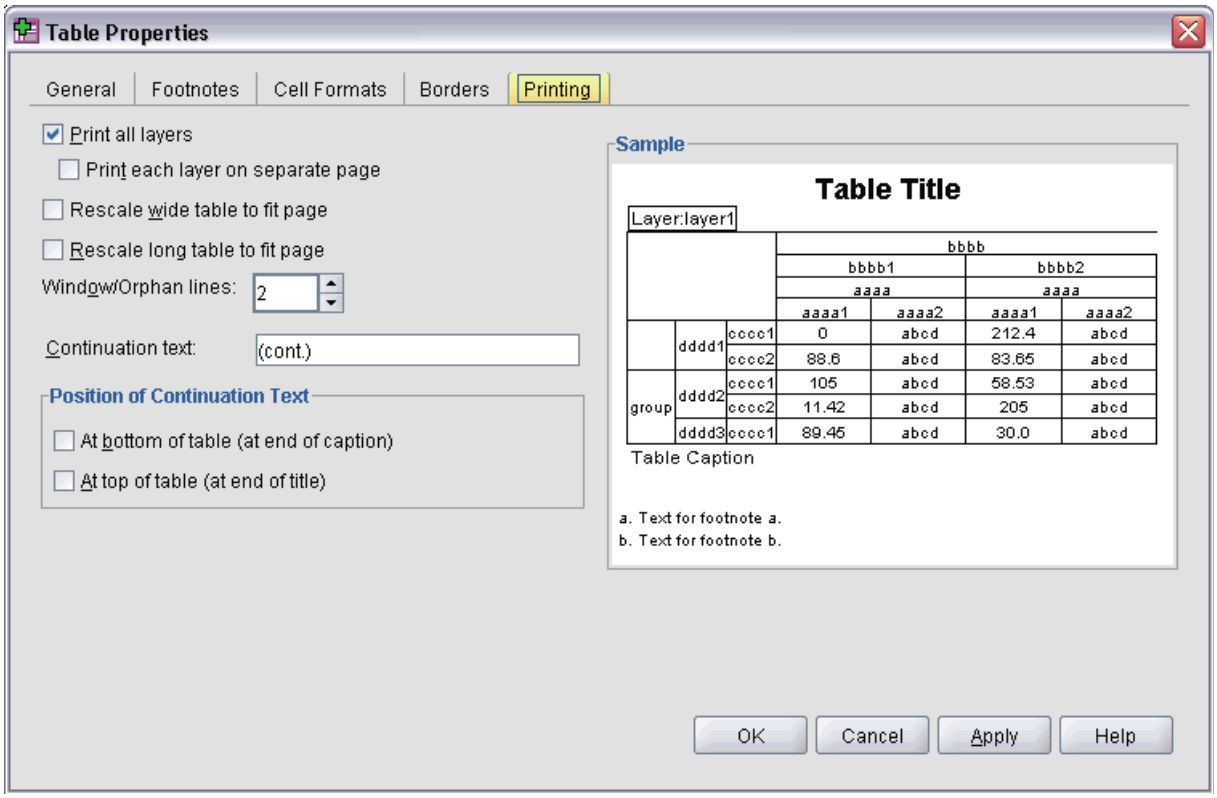
Можете да конфигурирате следните свойства за отпечатване на обощавщи таблици:

- Отпечатване на всички слоеве или само на наи-горния слой и отпечатване на всеки слой на отделна страница.
- Вертикално или хоризонтално смаляване на таблицата, за да се помести на страницата.

- Избягване на самотни линии, като конфигурирате минимален брой редове или колони, които ще се съдържат в определена област от таблицата, когато таблицата е прекалено широка или висока за дефинирания размер на страница.

***Забележка*:** Ако таблицата е прекалено висока за дадена страница, заради друга информация в началото на страницата, но самата таблица се помества в размера на страница, тогава цялата таблица се отпечатва на самостоятелна страница, без значение какви са настойките за самотни линии.

- Включване на текст за продължаване на таблици, които не се поместат на единична страница. Можете да изберете да показвате този текст в края и началото на всяка страница. Ако нито една от двете опции не е избрана, текстът няма да бъде показан.



Фигура 11-14 Table Properties dialog box, Printing tab

За да конфигурирате отпечатването на таблици:

- Изберете подпрозорец Printing.
- Изберете желаните опции.
- Потвърдете с OK или Apply.

Свойства на клетки

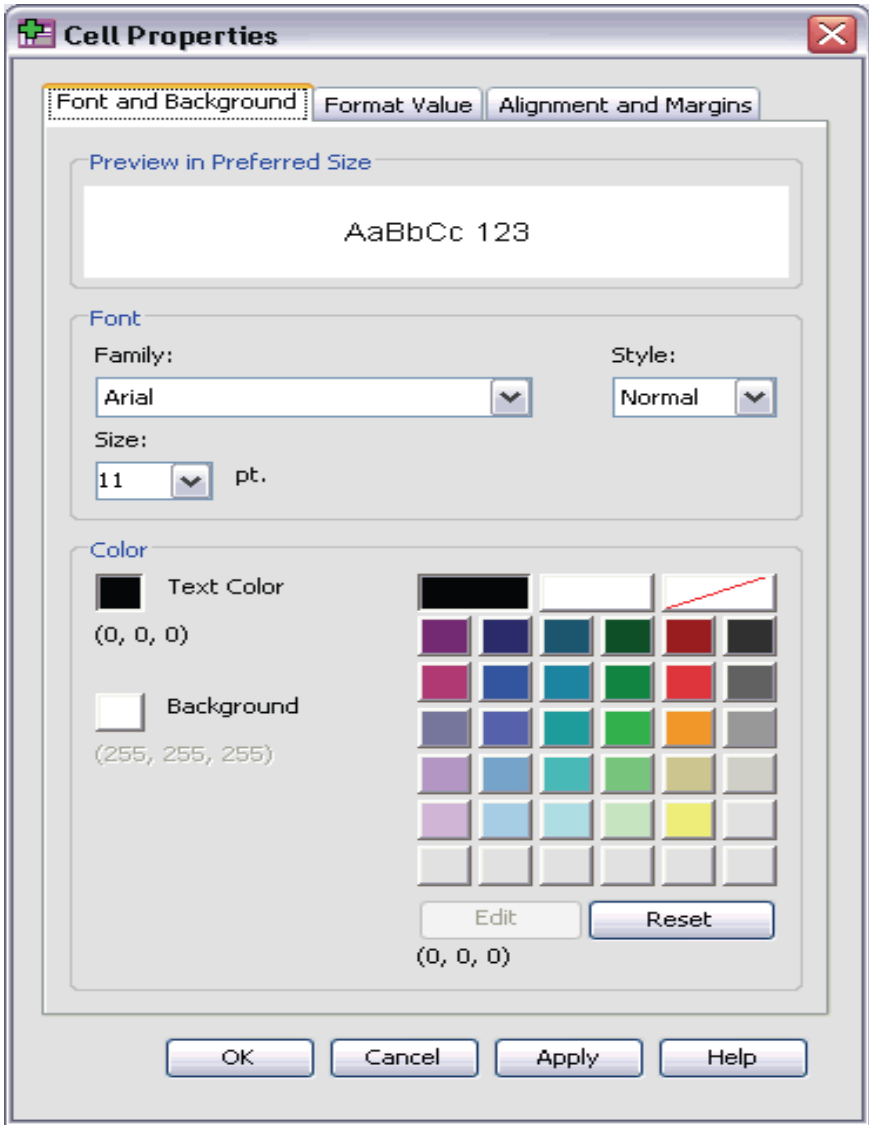
Свойствата на клетка се отнасят до единична клетка. Можете да променяте шрифта, формата на стойностм подравняването, полетата и цветовете. Свойствата на клетка предхождат свойствата на таблица: тоест, когато редактирате таблицата, това не се отразява на идивидуално форматирани клетки.

За да промените свойствата на клетка:

- Активирайте таблицата и изберете клетка(клетки) в активираната таблица.
- От менюто Format или от контекстното меню изберете Cell Properties.

Шрифт и Фон

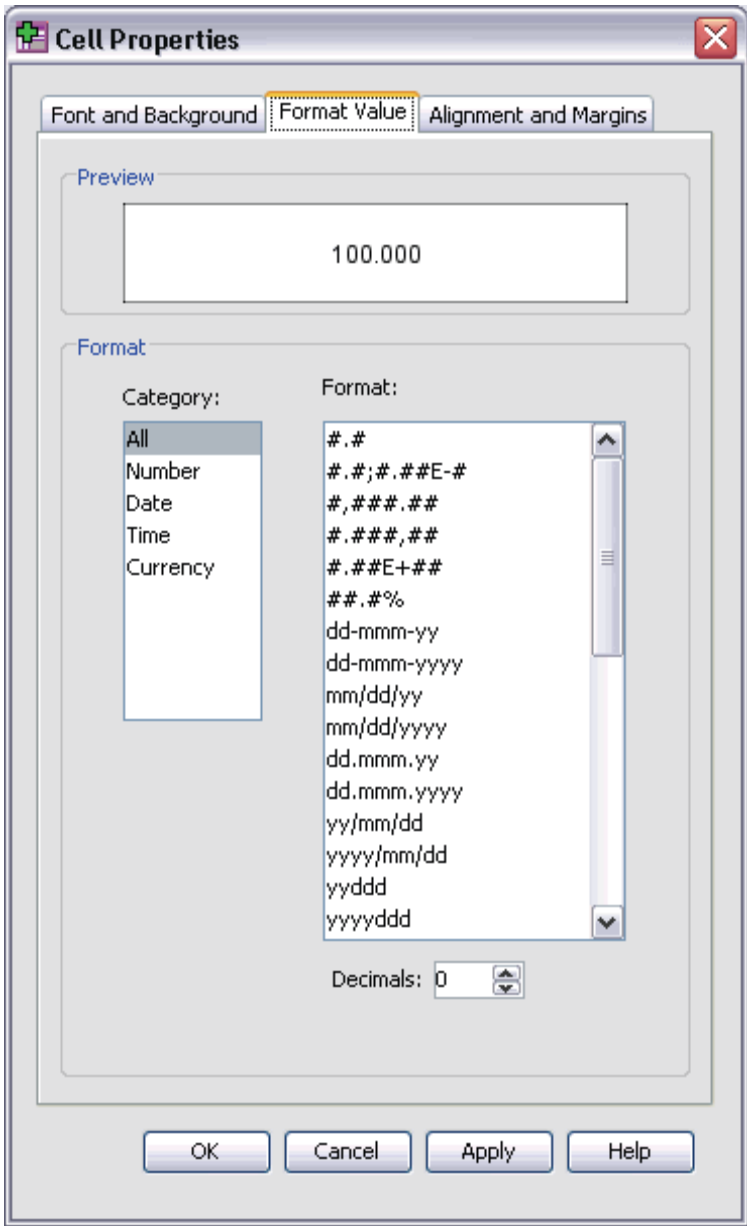
Подпрозорецът шрифт и фон (Font and Background) конфигурира цвета, стила и фона за избраните клекти в таблицата.



Фигура 11-15 Cell Properties dialog box, Font and Background tab

Формат на стойността

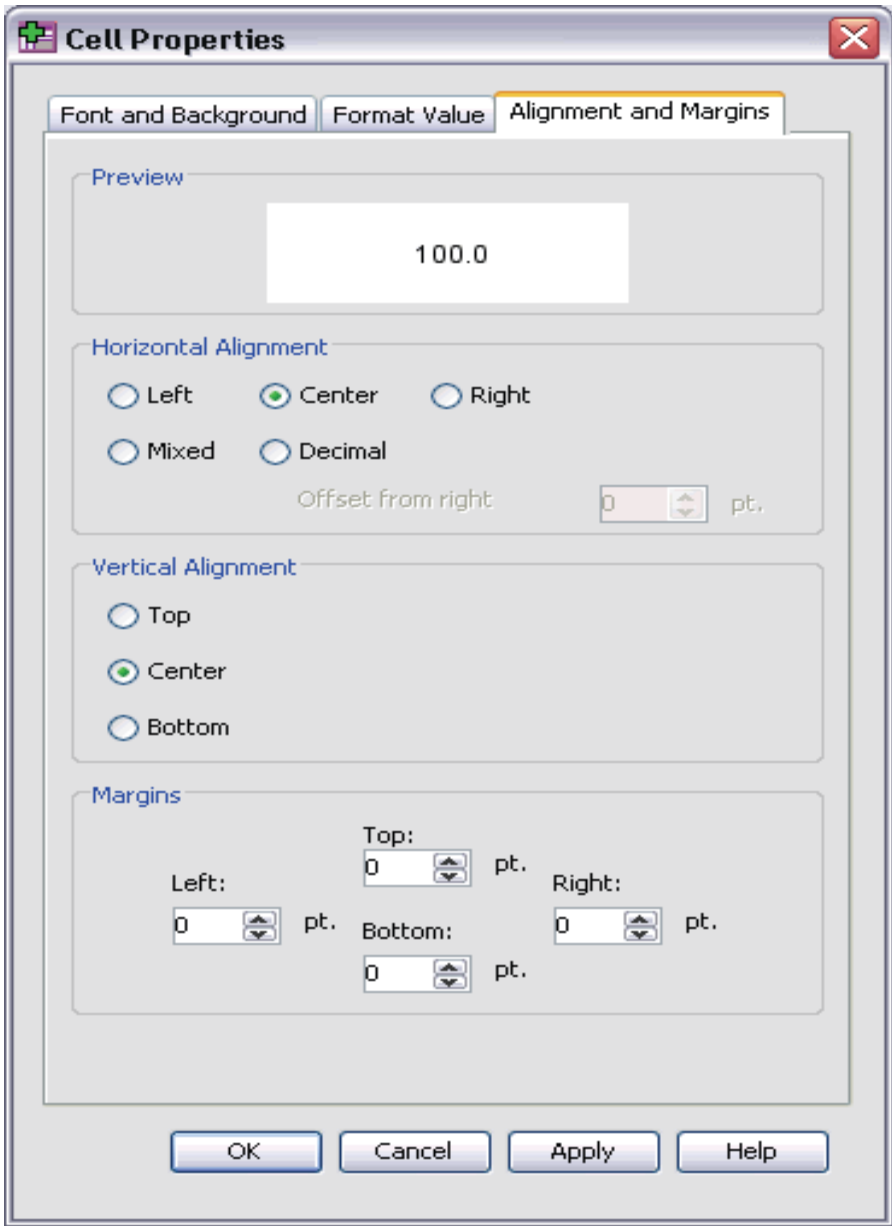
Подпрозорецът Format Value (Формат на стойност) конфигурира в каквъв формат да е стойността на дадената клетка. Можете да избирате от числа, дати, час и валути, като можете да задавате и колко цифри след десетичната запетая да се показват.



Фигура 11-16 Cell Properties dialog box, Format Value tab

Подравняване и полета

Подпрозорецър Alignment and Margins (подравняване и полете) конфигурира хоризонталнто и вертикално подравняване на стойностите, както и горните, долните, левите и десни полета за избраните клетки. Mixed (смесеното) хоризонтално подравняване действа според съдържанието на клетката, например датите се подравняват отдясно, а текстът отляво.



Фигура 11-17 Cell Properties dialog box, Alignment and Margins tab

Бележки под линия и пояснения

Можете да добавяте бележки под линия и пояснения към таблица. Можете също така да ги скривате, да променяте маркерите и да преномерирате бележките под линия. Някои от свойствата на бележките под линия се конфигурират заедно със свойствата на таблици.

Добавяне на бележки под линия и пояснения

За да добавите пояснения към таблица:

- От меню Insert изберете Caption.

Бележка под линия може да се прикачи към всякакво съдържание на клетка.

За да добавите бележка:

- Щракнете върху заглавие, клетка или пояснение в активирана обобщааяа таблица.
- От меню Insert изберете Footnote.

Показване/Скриване на пояснение

За да скриете пояснение:

- Изберете го.
- От менюто View изберете Hide.

За да покажете скритите пояснения:

- От менюто View изберете Show All.

Показване/Скриване на бележки под линия

за да скриете бележката:

- Изберете я.
- От менюто View изберете Hide (скриване) или изберете Hide Footnote от контекстното меню.

За да покажете скритите бележки под линия:

- Изберете Show All Footnotes от менюто View.
-

Маркер на бележка под линия

Маркерът конфигурира символите, които могат да отбелязват мястото на бележката под линия в текста.

Фигура 11-18 Footnote Marker dialog box

За да промените маркерите:

- Изберете бележка под линия.
- Изберете Footnote Marker от менюто Format.
- Въведете един или два символа.

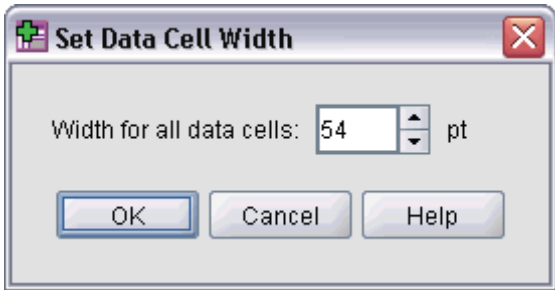
Преномериране на бележки под линия

Когато обобщавате таблица като промен[те местата на редове, колони или слоеве, може да се окаже, че бележките под линия са загубили последователноста си. За да ги преномерирате:

- Избрете Renumber Footnotes (Преномериране) от менюто Format.

Ширина на клетки с данни

Опцията Set Data Cell Width се използва за да се зададе еднаква ширина на всички клетки с данни.



Фигура 11-19 Set Data Cell Width dialog box

За да зададете ширина за клетките с данни:

- Изберете:
  Format
  Set Data Cell Widths...
- Въведете стойност за ширината.

Промяна на ширината на колона

- Щракнете и влачете с мишката границата на колона.

Показване на скритите граници в обобщаваща таблица

При таблици с множество скрити граници, може да се окаже удобно да покажете скритите линии. Това например ще ви улесни когато се опитвате да промените ширината на колона.

- От менюто View изберете Gridlines (Решетка).

Internet				Gender		
				Female	Male	Total
No	Owns PDA	No	Count	1897	1962	3859
			Expected Count	1903.4	1955.6	3859.0
			% within Owns PDA	49.2%	50.8%	100.0%
		Yes	Count	327	323	650
	Total					
Yes	Owns PDA					
	Total					

Internet				Gender		
				Female	Male	Total
No	Owns PDA	No	Count	1897	1962	3859
			Expected Count	1903.4	1955.6	3859.0
			% within Owns PDA	49.2%	50.8%	100.0%
		Yes	Count	327	323	650
			Expected Count	320.6	329.4	650.0
			% within Owns PDA	50.3%	49.7%	100.0%
	Total		Count	2224	2285	4509
			Expected Count	2224.0	2285.0	4509.0
			% within Owns PDA	49.3%	50.7%	100.0%
Yes	Owns PDA	No	Count	513	512	1025
			Expected Count	512.5	512.5	1025.0
			% within Owns PDA	50.0%	50.0%	100.0%
		Yes	Count	305	306	611
			Expected Count	305.5	305.5	611.0
			% within Owns PDA	49.9%	50.1%	100.0%
	Total		Count	818	818	1636
			Expected Count	818.0	818.0	1636.0
			% within Owns PDA	50.0%	50.0%	100.0%

Фигура 11-20 Gridlines displayed for hidden borders

Маркиране на Редове и колони в обобщаваща таблица

При обобщаващите таблици съществуват някои ограничения в това как да маркирате цял ред или колона, и маркировъчното оцветяване може да покрива прекъснати области то таблицата. За да маркирате цял ред или колона:

- Изберете етикет на ред или колона.
- От менюто изберете:

  Edit

  Select

  Data and Label Cells (данни и етикети на клетки)

или

- Щракнете с десен бутон етикета на категория за съответния ред или колона.
- От контекстното меню изберете:

  Select

  Data and Label Cells

или

- Щракнете етикета на ред или колона с натиснати Ctrl+Alt.

Отпечатване на обобщаващи таблици

Няколко фактора могат да повлияят на начина, по който отпечатаната таблица изглежда. Тези фактори могат да се конфигурират чрез промяна на атрибутите на таблицата.

- При многомерните обобщаващи таблици (таблици със слоеве) можете да отпечатате всички слоеве или само най-горния слой. При прекалено високи или широки таблици, можете автоматично да промените размера на таблицата, така че да се помести на страницата или да конфигурате къде прекъсват таблицата и страницата. При таблици, с височина или ширина, прекалено големи за единична страница , можете да конфигурирате къде се прекъсва таблицата между различните страници.

Използвайте опцията Print Preview (Предварителен преглед) в менюто File, за да проверите как ще изглеждат отпешатаните таблици.

Конфигуриране на прекъсванетона таблица при таблици с голяма ширина или височина

Таблиците, с височина или ширина, прекалено големи за единична страница автоматично се разделят и се отпечатватна няколко части. (При широките таблици, ако има достатъчно място, няколко части могат да се отпечатат на една страница) Можете да :

- Конфигурирате позициите в редовете или колоните, където таблицата да се раздели.
- да определите редове и колони, които трябва да останат заедно при разделяне на таблицата.
- да промените мащаба на големи таблици, така че дасе поместят на определената страница.

За да определите прекъсвания на редове и колони при отпечатване на таблица

- Щракнете върху етикета на колона, вляво от позицията, на която искате да вмъкнете прекъсване, или щракнете етикета на ред, над позицията където искате да се вмъкне прекъсването.
- От менюто изберете:
  Format
  Break Here (Прекъсни тук)

За да определите кои редове и колони да останат заедно при разделяне

- Маркирайте етикетите на редове или колони, които искате да останат заедно. (Щракнете и влачете или щракнете с натиснат Shift за да маркирате няколко реда/колони.)
- От менюто изберете:
  Format
  Keep Together (Запази заедно)

Създаване на Диаграма от обобщаваща таблица

- Щракнете два пъти върху таблицата, за да я активирате.
- маркирайте редовете, колоните или клетките, които искате да представите в диаграмата.
- Щракнете с десен бутон някъде в маркираната област.
- Изберете Create Graph (създай диаграма) от контекстното меню и изберете тип диаграма.

Глава 12. Работа в команден режим

Командният език ви дава възможност да запазвате и автоматизирате много задачи. Той също така предоставя една функционалност, която не може да се намери в менютата и диалоговите прозорци. Повечете команди можете да откриете в менютата и прозорците, но някои са достъпни само ако използвате командния език. Командният език ви позволява и да си запазите работата в команден файл за да можете да повторите анализа след известно време или да да стартирате автоматизирана задача с Production Facility.
Командният файл е просто един текстови файл, който съдържа команди. Докато е възможно да отворите команден прозорец и да въведете команди, често е по-лесно ако позволите на софтуера да ви помогне да си изградите команден файл като използвате един от тези методи.

- да копирате команден режим от диалоговите прозорци
- да копирате команден режим от дневника на резултатите (Output Log)
- да копирате команден режим от файл от журнала

Подробна информация за командния режим за справка съществува в две форми: като чсаст от цялостната система за помощ, и като отделен pdf файл, който се нарича *Command Syntax Reference* и който също може да бъде намерен в менюто за помощ.

Правилата на командния режим

Когато стартирате командите от прозорец на командния режим по време на сесия, това се нарича интерактивен режим.
Следните правила важат за командните спецификации по време на интерактивен режим:

- Всяка команда трябва да започне на нов ред. Командите могат да започват във всяка една колона от командни редове и да продължават колкото реда са необходими. Изключение прави командата END DATA, която трябва да започва в първата колона на първия ред след края на данните.
- Всяка команда трябва да завърши с точка като команден терминатор (завършек). Все пак е най-добре е да се изпусне командния терминатор на BEGIN DATA, за да не се третират данните на вътрешните редове като продължителни спецификации.
- Командният терминатор трябва да бъде последният символ различен от празно място в командата.
- При липсата на точка, която да изпълнява функцията на команден терминатор, се оставя един празен ред да служи за такъв.

Забележка: За да има съответствие с други комадни режими (включително и командни файлове, които стартират с командите INSERT или INCLUDE в интерактивна сесия, всеки един ред от команден режим не би трябвало да надхвърля 256 байта.

- Повечето подкоманди се разделят с наклонени черти (/). Наклонената черта преди първата подкоманда обикновено не е задължителна.
- Променливите имена трябва да бъдат изписани в пълния си вид.
- Текст, който се обгражда с апострофи или кавички трябва да бъде поместен на един ред.
- За да се посочат десетичните единици трябва да се използва период (.), без значение какво са регионалните или локалните настройки.
- Променливите имена, завършващи с точка могат да създатат грешки в командите, направени от диалоговите прозорци. Не можете да създавате такива променливи имена в диалоговите прозорци и като цяло трябва да ги избягвате.

В командния режим има значение дали използвате малки или главни букви и съкращения от по три или четири буквимогат да бъдат използвани за много командни спецификации. Можете да изпозлвате колкото си искате редове, за да дадете спецификация на една единствена команда. Също така можете да прекъсвате редовете или да поставяте празно място почти навсякъде, където се позволява едно празно място, като например от двете страни на наклонени черти, скоби, аритметични оператори или между променливи имена. Например:

FREQUENCIES
VARIABLES=JOBCAT GENDER
/PERCENTILES=25 50 75
/BARCHART.
и
freq var=jobcat gender /percent=25 50 75 /bar.
са приемливи алтернативи, които дават едни и същи резултати.

Файлове с командата INCLUDE

За командни файлове, които работят чрез командата INCLUDE важат правилата за режима на партидите. А те са следните:

- Всички команди в командния файл трябва да започнат в първата колона. Можете да използватесимволите за плюс и минус в първата колона ако искате да оеднаквите командната спецификация, за да направите командния файл по-четлив.
- Ако се използват повече редове за една команда, първата колона от всеки ред на едно продължение трябва да бъде празна.
- Комадните терминатори не са задължителни.
- Един ред не може да превишава 256 байта; всички допълнителни значи ще бъдат отрязани.

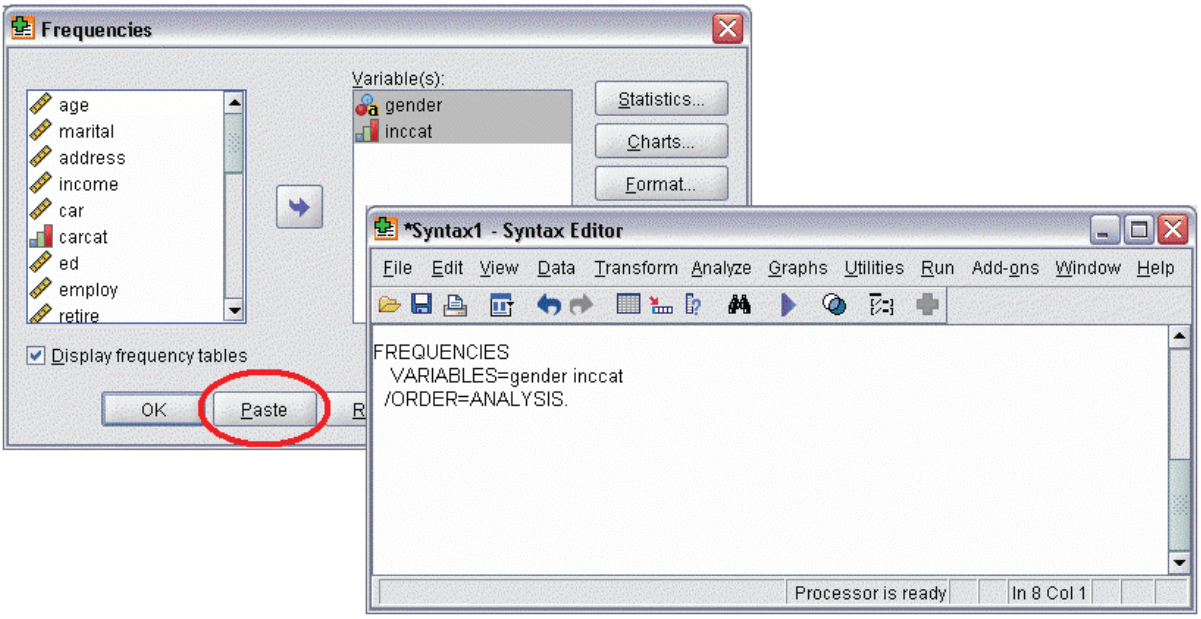
Ако нямате съществуващи командни файлове, които вече използват командата INCLUDE, може би вместо това трябва да използвате командата INSERT, след като тя може да служи за командните файлове, които се подчиняват и на двата вида правила. Ако създавате команден режим като копирате и прехвърляте изборите от диалоговите прозорци в команден прозорец, форматът на командите е подходящ за всеки един вид команден режим на работа. За повече информация вижте *Command Syntax Reference* (в PDF формат от менято за помощ).

Копиране на команден режим от диалогови прозорци. Най-лесният начин да създадеш файл в команден режим е да направиш нужните избори в диалоговите прозорци и да копираш и прехвърлиш тази информация в команден прозорец. Прехвърляйки командите във всяка една стъпка на дълъг анализ можете да изградите работен файл, който ви позволява да повторите анализа на по-късен етап или да стартирате автоматизирана задача с Production Facility.
В командния прозорец можете да стартирате копираните команди , да ги редактирате и да ги запазите като команден файл.

За да копирате команден режим от диалоговите прозорци:

- Отворете диалоговия прозорец и направете изборите, които желаете
- Изберете Paste

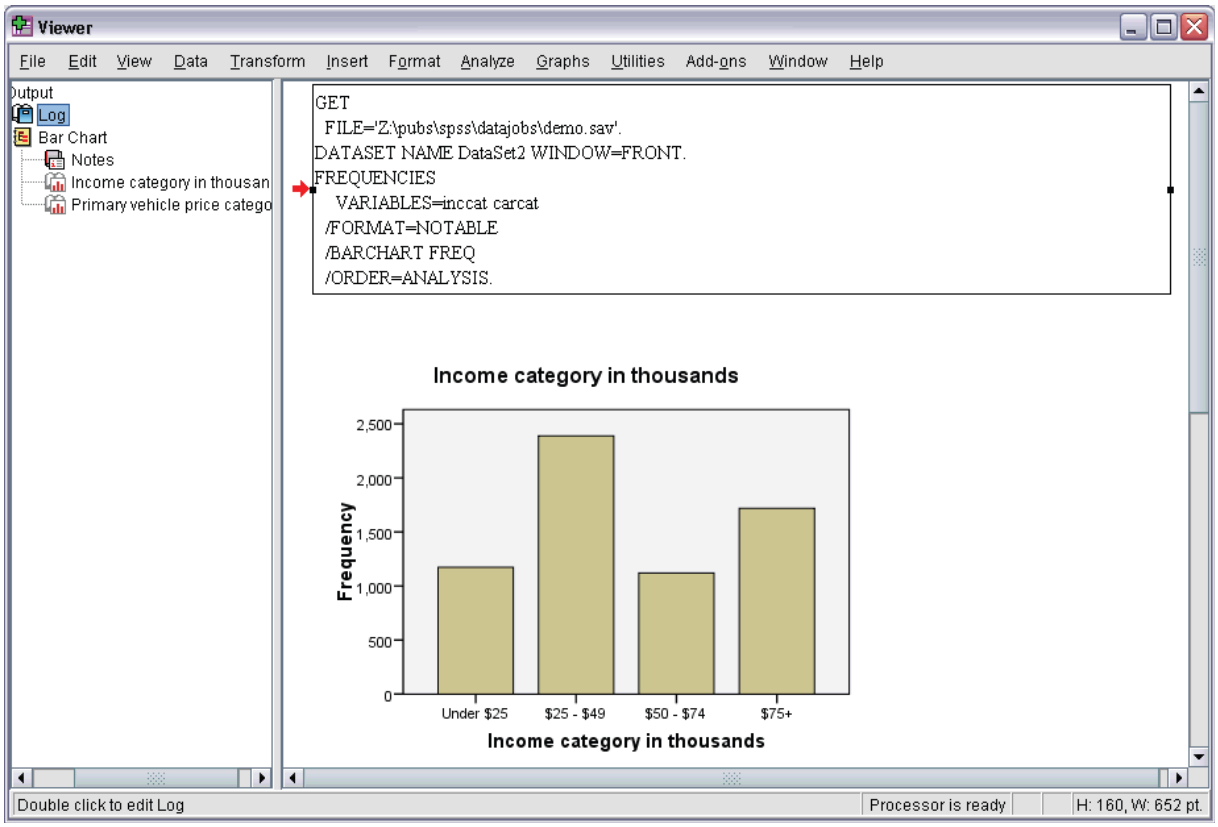
Комадният режим е прехвърлен в предназначения команден прозорец. Ако нямате отворен команден прозорец, нов такъв се отваря автоматично и там се прехвърлят командите.



Фигура12-1 *Прехвърляне на команди от диалогови прозорец*

Копиране на команди от дневника за резултатите (Output Log)

Можете да създадете команден файл като копирате командния режим от дневника, който се появява във Viewer. За да използвате този метод трябва да изберете Display commands (покажи командите) в дневкика от настройките на Viewer (Edit menu, Options, Viewer tab) before running the analysis. Тогава всяка команда ще се появи във Viewer заедно с резултатите от анализа. В командния прозорец можете да стартирате прехвърлените команди, да ги редактирате и да ги запазите като команден файл.



Фигура 12-2 *Команден режим в дневника*

За да копирате команди от дневника на резултатите (Output Log):

- преди да стартирате анализа изберете от менюто Edit / Options
- изберете Display commands от дневника, от Viewer; както стартирате анализите, командите за изборите ви в диалоговите прозорци се запазват в дневника отворете предишните командни файлове, които сте запазили или създайте нов такъв;
- за да създадете нов команден файл, изберете от менютата: File / New / Syntax
- във Viewer кликнете два пъти артикул от дневника, за да го активирате
- изберете текста, който желаете да копирате
- от менютата на Viewer изберете Edit / Copy
- от менютата в команден прозорец изберете Edit / Paste

Стартиране на командния режим

- маркирайте командите, които искате да стартирате в командния прозорец
- кликнете върху бутона Run (триъгълникът, който сочи надясно) от менюто Syntax Editor или
- изберете една от възможностите от менюто Run:
  - **All. Стартира всичките команди в командния прозорец.** Runs all commands in the syntax window.
  - **Selection. Стартира избраните в момента команди. Това всички команди, които са частично маркирани. Ако няма избрани команди, стартира командата, където в момента се намира курсорът.**
  - **Current. Стартира командата, където в момента се намира курсорът.**
  - **To End. Стартира всички команди от позицията на курсора до края на командите.**

Бутонът Run от менюто Syntax Editor стартира избраните команди или командите, където се намира курсорът ако няма избрани команди.

Фигура 12-3 *Бутонът Run*

Командни файлове в режим Unicode

В режим *Unicode* изходният формат за запазване на командни файлове създаден или променен по време на сесията е също

Unicode (UTF-8). Командни файлове във формат *Unicode* не могат да се четат от версии на SPSS преди 16.0. За да запазите команден файл във формат съвместим с по-ранни версии:

- от менютата на командния прозорец изберете File / Save As
- в диалоговия прозорец на Save As изберете Encoding / Local Encoding. Local encoding се определя от настоящата позиция.

Глава 13. Анализ на честоти (*Frequencies)*

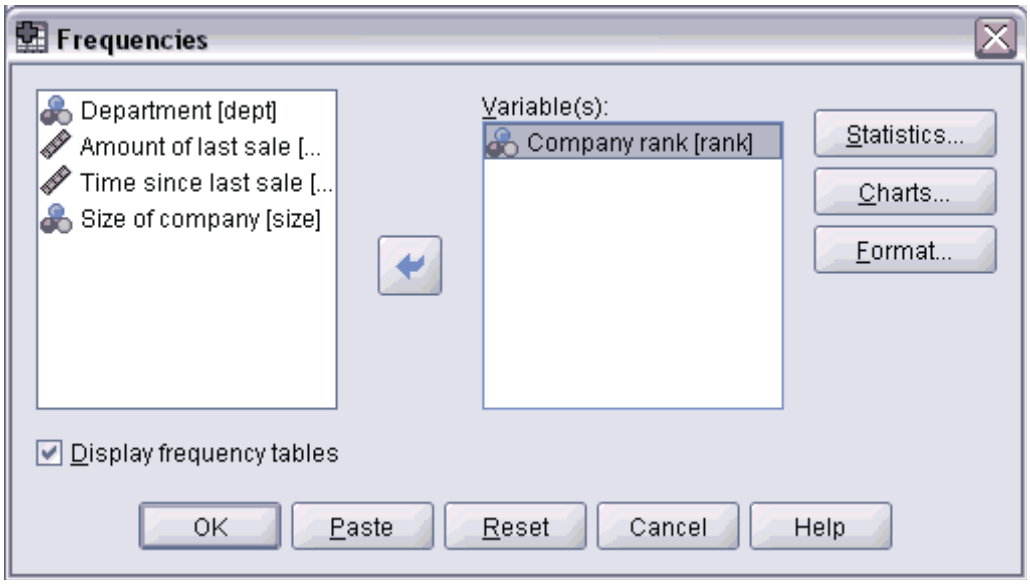
Процедурата за честоти осигурява статистика и графики, които помагат да се опишат различни видове променливи. Процедурата за честотите е добро място да започнете да разглеждате вашата информация. За доклад на честотите и графика можете да подредите стойностите или във възходящ, или в низходящ ред или можете да подредите категориите по техните честоти. Този доклад може да бъде съкратен когато енда променлива има различни стойности. Можете да назовете диаграми с фабричните честоти или с проценти.

Пример: Какво е разпределението на клиентите на една компания според типа на индустрията? От резултатите можете да научите, че 37,5% от клиентите ви са в правителствени агенции, 24,9% са в корпорации, 28,1% са в академични институции и 9,4% са в здравеопазването. За продължителна, колективна информация, като например стойността на продажбите, можете да разберете, че средната стойност е $3,576, със стандартно отклонение възлизащо на $1,078.

Статистика. Брой на честотите, проценти, кумулативни проценти, средно число, медиана, мода, сума, стандартно отклонение, дисперсии, размах, минимални и максимални стойности, стандардна грешка на средното число, коефициент на асиметрията, коефициент на ексцеса (и двата със стандартни грешки), квартили, персентили, определени от потребителя, стълбови диаграми, кръгови диаграми и хистограми.

Информация. Използвайте числени кодове или редици, за да кодирате категорийните променливи (редни или бройни).

Предположения. Табулациите и процентите дават полезно описание за информация от което и да е разпределение, особено за променливи с подредени или неподредени категории. Повечето от резюмираната статистика, която не е задължителна, като например средната стойност и стандартното отклонене се базират на обикновена теория и са подходящи за количествените променливи със симетрични разпределения. Твърдата статистика като медианата, квартилите и персентилите е подходяща за количествени променливи, които може или не да отговорят на презумпцията за нормалност.



Фигура 13-1 *Честоти и основен диалогов прозорец*

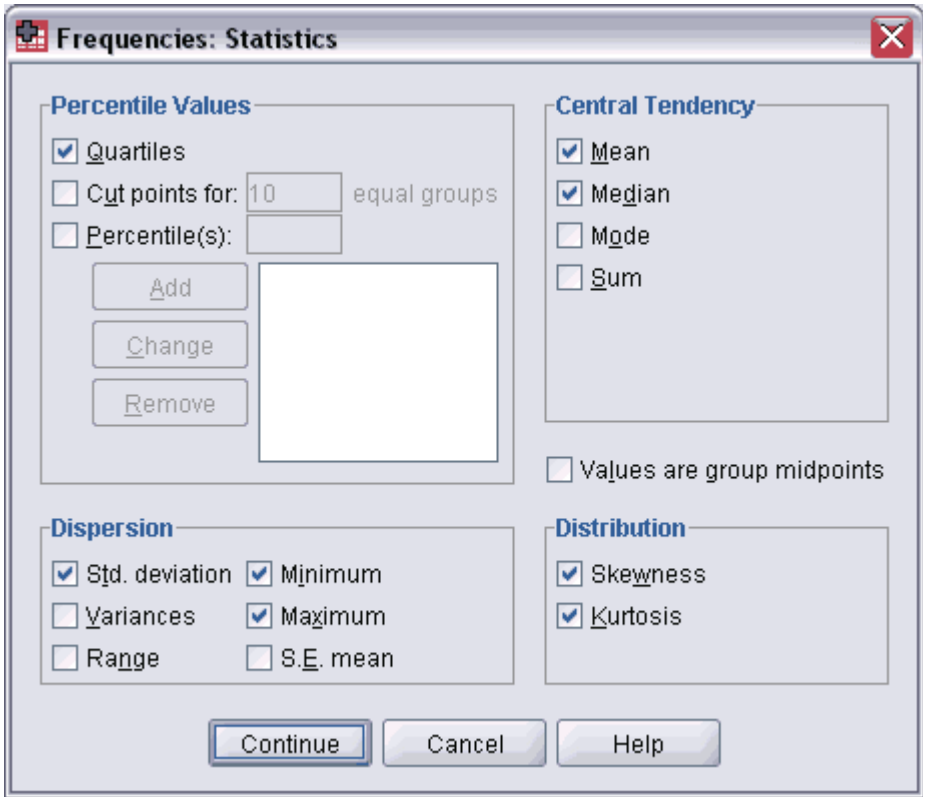
За да създадете честотни таблици:

- от менютата изберете Analyze / Descriptive / Statistics / Frequencies...

Изберете една или повече категорийни или количествени променливи
По желание можете да:

- изберете Statistics за описателна статистика за количествени променливи
- изберете Charts за кръгови, диаграми, стълбови диаграми и хистограми
- изберете Format за реда, в който да бъдат показани резултатите

Статистика на честотите



Фигура 13-2 *Диалогов прозорец на статистиката на честотите*

Стойности на персентилите. Стойности на количествена променлива, които разделят подредената информация по групи, така че определен процент е отгоре и друг процент е отдолу.
Квартили (25-ят, 50-ят и 75-ят перцентили) разделят резултатите в четири групи с еднакъв размер. Ако искате равен брой групи, различен от 4, изберете Cut points for n equal groups. Можете също така да посочите индивидуални перцентили (например: 95-ят персентил, под чиято стойност спадат 95% от резултатите.
Централна тенденция. Статистиката, която описва местоположението на разпределението включва средното число, медианата, модата и сумата на всички стойности.

- **Средно число.** Мерило за централна тенденция. Средно аритметично, сумата се разделя на броя случаи.
- **Медиана.** Стойността, над и под която попадат половината от случаите, 50-ят персентил. Ако има четен брой случаи, медианата е средно аритметично от двата случая в средата когато са подредени в низходящ или възходящ ред. Медианата е мерило за централна тенденция, но не е чуствителна към крайни стойности (както например средното число, което се влияе от много ниски или много високи стойности).

- **Мода.** Най-често срещаната стойност. Ако няколко стойности се срещат най-често, всяко едно от тях е модата. Процедурата на честотите показва само най-ниските стойности от тези многобройни моди.
- **Сума.** Сумата или сборът от стойностите на всички случаи.

Дисперсия. Статистика, която измерва вариациите в данните; включва стандартното отклонение, дисперсии, размах, минимални и максимални стойности и стандартна грешка при средното число.

- **Стандартно отклонение.** Измерва дисперсията около средното число. В нормално разпределение, 68% от случаите попадат в рамките на едно стандартно отклонение на средното число и 95% от случаите попадат в рамките на две стандартни отклонения. Например: ако средната възраст е 45 години със стандардно отклонение 10, 95% от случаите биха били между 25 и 65 в нормално разпределение.
- **Дисперсии.** Измерва дисперсията около средното число, равняваща се на отклоненията от средното число на квадрат, разделени на броя на случаите минус едно. The variance is measured in units that are the square of those of the variable itself.
- **Размах.** Разликата между най-големите и най-малките стойности на числена променлива, максималната стойност минус минималната.
- **Минимална стойност.** Най-малката стойност на числена променлива.
- **Максимална стойност.** Най-голямата стойност на числена променлива.
- **S. E. mean.** Измерва колко средното число може да варира от извадка до извадка, взети от едно и също разпределение. Може да се използва приблизително да се сравни средното число с хипотетична стойност (ако заключите, че двете стойности са различни ако пропорцията на разликата към стандартната грешка е по-малка от -2 или по-голяма от +2).

Разпределение. Коефициентите на асиметрията и на ексцеса описват формата и симетрията на разпределението. Те се представят със стандартните си грешки.

- **Коефициент на асиметрията.** Измерва асиметрията на разпределението. Нормалното разпределение симетрично и има коефициент на асиметрията 0. Разпределение със значителен положителен коефициент на асиметрията има дълга дясна опашка. Разпределение със значителен отрицателен коефициент на асиметрията има дълга лява опашка. Като насока, коефициент на асиметрията, който е два пъти повече от стандартната си грешка се счита за отклонение от симетрията.
- **Коефициент на ексцеса.** Измерва степента, до която резултатите се натрупват около централна точка. За нормално разпределение стойността на коефициента на ексцеса е 0. Положителен коефициент на ексцеса сочи, че резултатите се натрупват повече и имат по-дълги опашки от тези в нормално разпределение, а отрицателен коефициент на ексцеса сочи, че резултатите се натрупват по-малко и имат по-къси опашки.

Стойности и групови средни точки. Ако стойностите във вашите данни са средни точки на групи (например възрастта на всички хора в трийсетте са кодирани като 35), изберете тази опция, за да изчислите медианата и перцентилите за първоначалните, негрупирани данни.

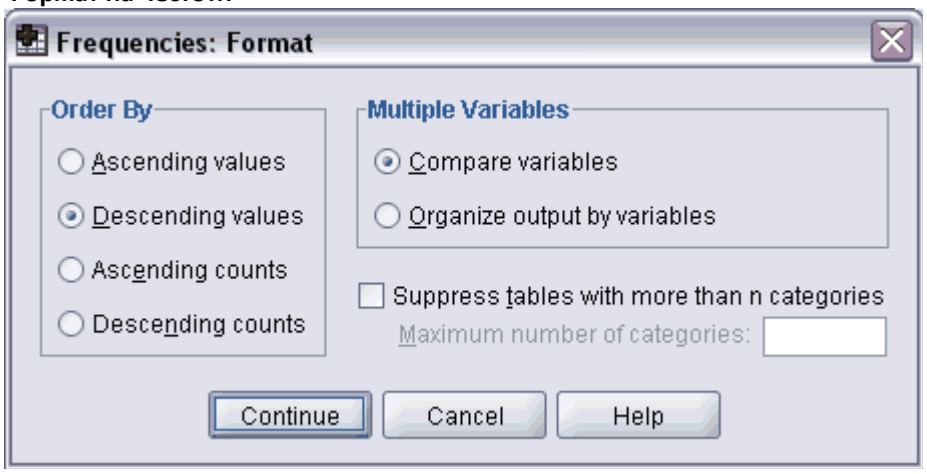
Честотни диаграми (*Frequencies Charts*)

Видове диаграми. Кръговата диаграма показва различни дялове от едно цяло. Всяко парче от кръговата диаграма съответства на група, която се определя от една единствена групираща променлива. Стълбовата диаграма показва броя за всяка отделна стойност или категория под формата на отделен стълб, като ви дава възможност да сравнявате категориите визуално. Хистограмата също има стълбове, но те са разположени по скала с еднакви интервали. Хистограмата показва формата, центъра и разпростирането на разпределението. Нормална крива наложена върху хистограмата ви помага да прецените дали данните са нормално разпределени.

Стойности на диаграмите. При стълбовите диаграми оста на скалата може да бъде обозначена или по честота или по проценти.

Chart Values. For bar charts, the scale axis can be labeled by frequency counts or percentages.

Формат на честоти



Фигура 13-4 *Диалогов прозорец на формат на честотите*

Подреждане. Честотната таблица може да бъде организирана според действителните стойности в данните или според честотата на тези стойности и това може да е или в низходящ, или във възходящ ред. Но все пак ако изискате хистограма или персентили, Frequencies решава, че променливата е количествена и представя стойностите и въз възходящ ред.
Множество променливи. Ако създадете таблици със статистика за множество променливи можете или да поставите всички променливи в една таблица, или за всяка променлива да има отделна таблица.

Съкращаване на таблици с повече от n категории. Тази опция предотвратява показването на таблици с по-голям брой променливи от определения.

Глава 14. Описание на метрични данни

Описателна статистика

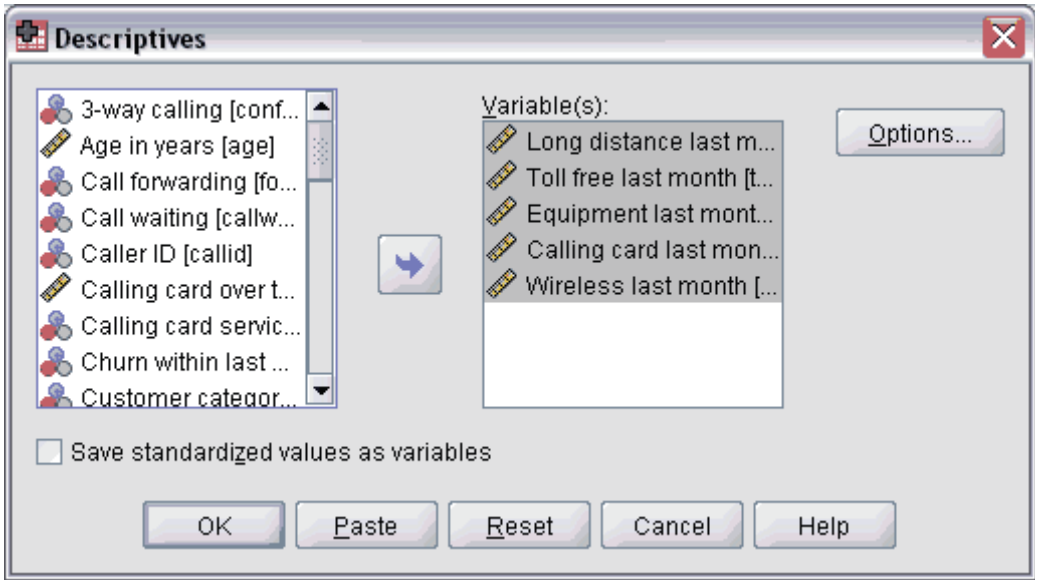
Процедурата за описателната статистика произвежда всеобхватна обобщена статистика за няколко променливи в една единствена таблица и изчислява стандартизирани стойности (*z* scores). Променливите могат да се подредят според големината на техните средни стойности (в низходящ или възходящ ред), по азбучен ред, или според основния ред, в който избирате променливите. Когато стандартизираните стойности са запазени, те се добавят към данните в Data Editor и са на разположение за диаграми, списъци с данни и анализи. Когато променливите се записват в различни единици (например брутен вътрешен продукт на глава от населението и процент literate, стандартизирана трансформация поставя променливите на обща скала за по-лесно визуално сравнение.

Пример: Ако всеки случай в данните ви съдържа дневните продажби за всеки член от търговския персонал (например продажбите по отделно на Боб, Ким и Брайън) направени всеки ден за няколко дена, процедурата Descriptives може да изчисли средните дневни продажби за всеки член на персонала и може да подреди резултатите от най-високата средна артитметична стойност до най-ниската.

Статистика. Размер на извадката, средно число, медиана, мода, минимална стойност, максимална стойност, сдандартно отклонение, коефициент на асиметрията, коефициент на ексцеса с техните стандартни грешки.

Данни. Използвайте числени променливи след като сте ги проверили гарфически за отсичане на грешки, отдалечени данни и аномалии при разпределението. Процедурата Descriptives е много ефикасна при голми файлове (с хиляди случая).

Предположения. Повечето от достъпната статистика (including *z* scores) е базирана на обикновена теория и е подходяща за числителни променливи (измервания на ниво интервал или пропорция) със симетрично разпределение. Избягвайте променливи с неподредени категории или с изкривено разпределение. Разпределението на *z* scores има същата форма като тази на първоначалните данни. Следователно изчисляването на *z* scores не е решение за проблемните данни.



Фигура 14-1 *Диалогов прозорец на описателната статистика*

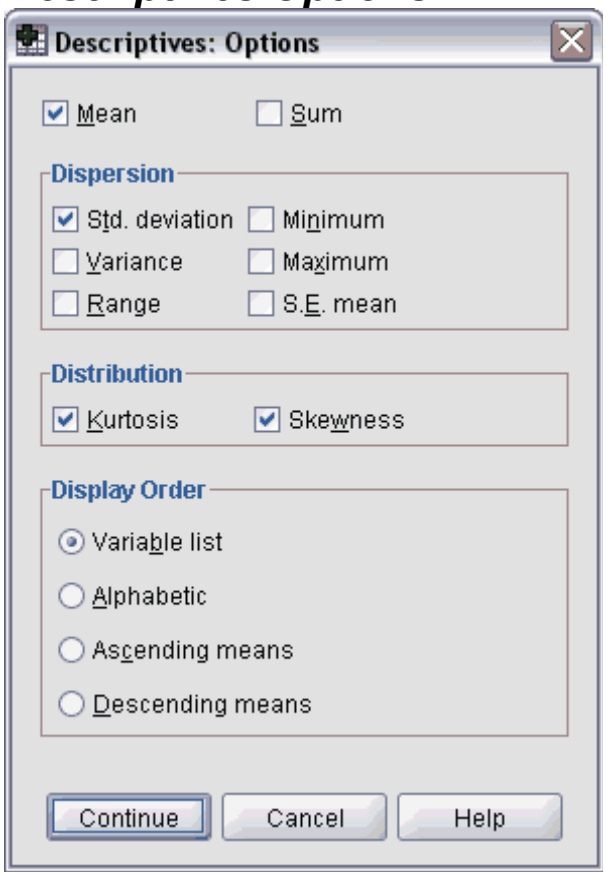
За да получите описателна статистика:

- От менютата изберете Analyze / Descriptive Statistics / Descriptives...
- Изберете една или повече променливи

По избор можете да:

- Изберете Save standardized values as variables, за да запазите *z* scores като нови променливи.
- Изберете Options за допълнителна статистика по избор и за ред на показването

Descriptives Options



Фигура 14-2 *Descriptives Options dialog box*

Средно число и сума. Средното число или средното аритметично се изчислява по подразбиране.
Дисперсия. Статистика, която измерва вариациите в данните; включва стандартното отклонение, дисперсии, размах, минимални и максимални стойности и стандартна грешка при средното число.

- **Стандартно отклонение.** Измерва дисперсията около средното число. В нормално разпределение, 68% от случаите попадат в рамките на едно стандартно отклонение на средното число и 95% от случаите попадат в рамките на две стандартни отклонения. Например: ако средната възраст е 45 години със стандардно отклонение 10, 95% от случаите биха били между 25 и 65 в нормално разпределение.
- **Дисперсии.** Измерва дисперсията около средното число, равняваща се на отклоненията от средното число на квадрат, разделени на броя на случаите минус едно. The variance is measured in units that are the square of those of the variable itself.
- **Размах.** Разликата между най-големите и най-малките стойности на числена променлива, максималната стойност минус минималната.
- **Минимална стойност.** Най-малката стойност на числена променлива.
- **Максимална стойност.** Най-голямата стойност на числена променлива.
- **S. E. mean.** Измерва колко средното число може да варира от извадка до извадка, взети от едно и също разпределение. Може да се използва приблизително да се сравни средното число с

хипотетична стойност (ако заключите, че двете стойности са различни ако пропорцията на разликата към стандартната грешка е по-малка от -2 или по-голяма от +2).

Разпределение. Коефициентите на асиметрията и на ексцеса описват формата и симетрията на разпределението. Те се представят със стандартните си грешки.

- **Коефициент на асиметрията.** Измерва асиметрията на разпределението. Нормалното разпределение симетрично и има коефициент на асиметрията 0. Разпределение със значителен положителен коефициент на асиметрията има дълга дясна опашка. Разпределение със значителен отрицателен коефициент на асиметрията има дълга лява опашка. Като насока, коефициент на асиметрията, който е два пъти повече от стандартната си грешка се счита за отклонение от симетрията.
- **Коефициент на ексцеса.** Измерва степента, до която резултатите се натрупват около централна точка. За нормално разпределение стойността на коефициента на ексцеса е 0. Положителен коефициент на ексцеса сочи, че резултатите се натрупват повече и имат по-дълги опашки от тези в нормално разпределение, а отрицателен коефициент на ексцеса сочи, че резултатите се натрупват по-малко и имат по-къси опашки. Като насока, коефициент на ексцеса, който е два пъти повече от стандартната си грешка се счита за отклонение от симетрията.

Ред. По подразбиране, променливите се подреждат в реда, в който сте ги избрали. По избор могат да се подредят по азбучен ред, във възходящ ред, или в низходящ ред.

Допълнителни командни характеристики на *DESCRIPTIVES*

Командният режим ви позволява да се:

- запазва стандартизирана статистика (*z* scores) за някои, но не всички променливи (с подкомандата VARIABLES)
- уточнят имена за новите променливи, които съдържат стандартизирана статистика (с подкомандата VARIABLES)
- изключват от анализа случаи с липсващи стойности за всяка една променлива(с подкомандата MISSING)
- подредят променливите според стойността на всеки един вид статистика, не само на средното число (с подкомандата SORT)

За пълна информация относно командния режим вижте *Command Syntax Reference.*

Глава 15. Предварително изследване на данните (*Explore)*

Процедурата Explore създава обобщена статистика и графично представяне за всички случаи или за отделни групи от случаи. Има много причини за използване на Explore процедурата - за проверка на данни, идентификация на откъсната част, описание, проверка на предположението и характеризиращи различия между **subpopulations** (групи от случаи). Проверката на данните може да покаже че вие имате необичайни стойностти, екстремни стойности, пропуски в данните, или други особености. **Exploring** на данните може да помогне да се определи дали статистическите, които се използват за анализ на данните са подходящи. Изследването може да покаже, че данните трябва да се трансформират ако техника изисква нормално разпределение. Или пък може да решите, че имате нужда от непараметрични тестове.

Пример: Погледнете разпределението на **maze-learning** за **rats** по четири различни **reinforcement** графици. За всяка от четирите групи, можете да видите дали разпределението е приблизително нормално и дали четирите разновидностти са равни. Можете също така да идентифицирате случаите с петте най-големи и петте най-малки. В **boxplots** и **stem-and-leaf** графично обобщават разпределението на време за обучение за всяка група.

Статистика и план: Средно, медиана, 5% махнати от средната, стандартна грешка, разлики, стандартно отклонение, минимум, максимум, обхват, среден квартил, обхват, **skewness** и **kurtosis** и тяхната стандартна грешка, доверителен интервал за средната (и определено ниво на доверие), персентили, **Huber's M-estimator, Andrews' wave estimator**, **Hampel's redescending M-estimator**, **Tukey's biweight estimator**, на петте най-големи и петте най-малки стойности на **Kolmogorov-Smirnov statistic** с ниво на значимост **Lilliefors** за тестване на нормалност и **Shapiro-Wilk statistic**. **Boxplots**, **stem-and-leaf plots**, хистограми, нормалност на план и **spread-versus-level plots** с тестове **Levene** и трансформации.

Данни: Процедурата **Explore** може да се използва за количествените променливи (измервания на интервал или съотношение на ниво). Факторна променлива (използвана за разбиване на данни в групи от случаи) трябва да има разумен брой на различни стойности (категории). Тези стойности могат да бъдат кратки низ или цифри. Променливата на етикета за случая, която се използва за етикетиране на **outliers** в **boxplots**, може да бъде кратък низ, дълъг низ(15 байта), или цифри.

Предположения: Разпределението на вашите данни не е задължително да бъде симетрично или нормално.

За да изследвате вашите данни

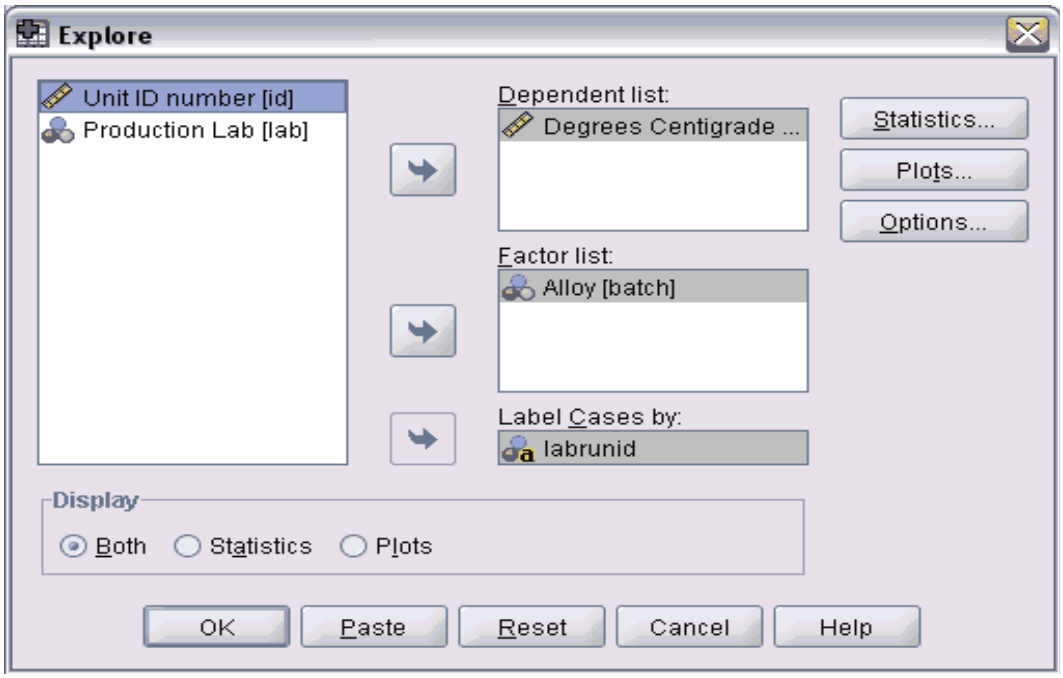
Изберете от менютата:

Analyze
Descriptive Statistics
Explore...

Изберете една или повече зависими променливи.

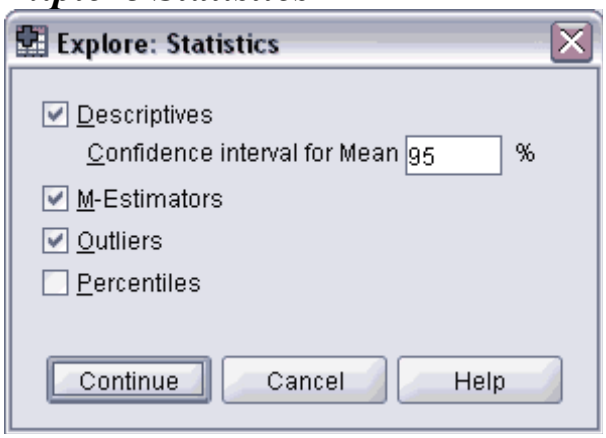
По желание можете да:

- Изберете една или повече факторни променливи, чиито стойности ще определят групите от случаите.
- Изберете идентификационна променлива която да поставя етикет случаи.
- Кликнете Statistics за устойчиви оценки, рязко отличаващи се измервания, персентили, и честотата на таблици.
- Кликнете **Plots** за хистограми, нормална вероятностна крива и тестове
- Щракнете върху Опции за контрол на липсващи стойности.



Фигура 15-1 Explore dialog Box

Explore Statistics



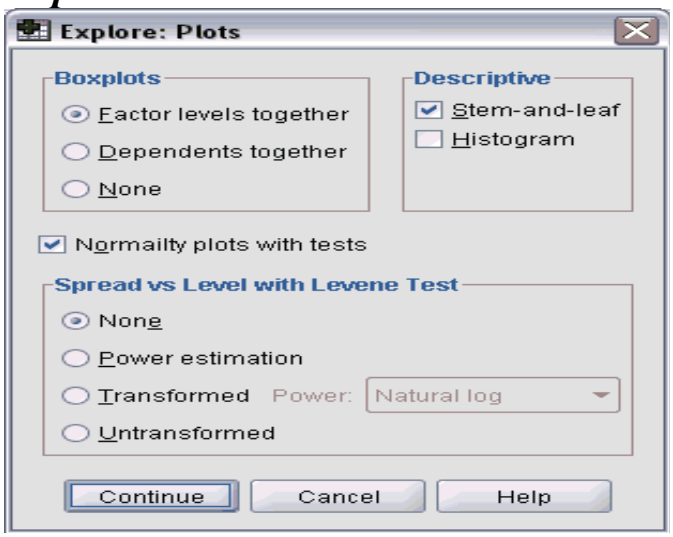
Фигура 15-2 Statistics диалогов прозорец

Descriptives. Тези измервания на централната тенденция и дисперсии се показват по подразбиране. Мерките на централната тенденция показва местоположението на разпространение; те включват средната, медианата, и 5% от средната. Мерки на дисперсия показват несходство на ценности; те включват стандартна грешка, вариация, стандартното отклонение, минимум, максимум, обхват, интерквартилен обсег. В описателна статистика също включва мерки за формата на разпределение; коефициентите на асиметрия и ексцес са показани с техните стандартни грешки. 95% доверителен интервал за средната е също така показан, можете да зададете различно ниво на доверие.

M-estimators. Устойчива алтернатива на извадковата средна. Изчислените оценки се различават в теглото, което те прилагат за случаите. **Huber's M-estimator**, **Andrews' wave estimator, Hampel's redescending M-estimator, and Tukey's biweight estimator** са показани.

Outliers. Показва петте най-големи и петте най-малки стойности със етикетите на случаи.
Percentiles. Показва стойности за 5-ти, 10, 25, 50-то, 75-ти, 90-та и 95-ти персентил.

Explore Plots



Фигура 15-3 *Explore Plots диалог*

Boxplots: Тези възможности контролират прозореца на **boxplots**, когато имате повече от една зависима променлива. Фактор равнища генерират отделни дисплеи за всяка зависими променлива. В рамките на прозореца, **boxplots** са показани за всяка от групите, определени с факторна променлива. Зависимите заедно генерират отделни прозорци за всяка група, определена с коефициент променлива. В рамките на прозореца, **boxplots** са показани едно до друго за всяка зависима променлива. Този прозорец е особено полезен, когато различните променливи представляват единна характеристика, измерена в различно време.

Descriptive: позволява да изберете **stem-and-leaf plots** и хистограми.

Normality plots with tests: Показва нормална вероятност и **detrended** нормални участъци. **Kolmogorov-Smirnov statistic**, с ниво на значимост **Lilliefors** за тестване на нормалност е показан. Ако нечислените тежести са определени, **Shapiro-Wilk statistic** се изчислява, когато среднопретегления размер на извадката е между 3 и 50. За липса на тежести или целочислени стойности, статистика, се изчислява, когато среднопретегления размер на извадката е между 3 и 5000.

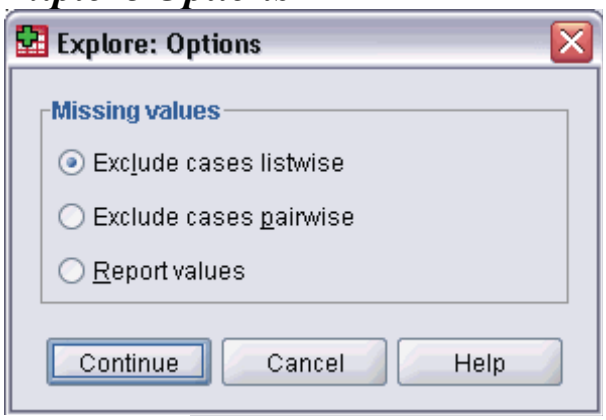
Spread vs. Level with Levene Test: Контролира данните за **spread-versus-level plots.** За всички **spread-versus-level plots**, на наклона на регресията и здравите тестове на Levene за хомогенност на дисперсията се показват. Ако изберете трансформация, тестовете на Levene се основават на промяната в данните. Ако не е избрана променлива за фактор, за **spread-versus-level plots** не се прави. **Power** оценяването създава графика на естествените логове на интерквартилните вариации срещу естествените логове от средните линии за всички клетки, както и оценка на трансформация на енергия за постигане на равни дисперсии в клетките. **Spread-versus-level plots**-помага да се определи нивото на енергия за трансформация, за да стабилизира (прави по-равни) разлики в групите. **Transformed** позволява да изберете една от алтернативите на **power**, може би след препоръка от **power** оценка, и произвежда графики на трансформираните данни. В интерквартилния обхват и средната стойност на трансформирани данни са начертани. **Untransformed** произвежда графики на суровите данни. Това е еквивалентно на трансформация с мощност от 1.

Explore Power Transformations

Има **power** трансформациите за **spread-versus-level plots**. За да се трансформират данни, трябва да се избере **power** за трансформация. Можете да изберете една от следните алтернативи:

- **Natural log.** Естествен лог на трансформация. Това е по подразбиране.
- **1/square root.** За всяка стойност на данните, реципрочно на корен квадратен е изчислено
- **Reciprocal.** Реципрочно на всяка стойност е изчислено
- **Square root.** Квадратния корен на всички данни се изчислява.
- **Square.** Всяка стойност на данните е на квадрат.
- **Cube.** Всяка стойност на данните е на куб

Explore Options



Фигура 15-4 *Explore Options диалог*

Missing Values. Контролира третирането на липсващите стойности.

- **Exclude cases listwise.** Случаи с липсващи стойности за всички зависими фактор променливи са изключени от всички анализи.
- **Exclude cases pairwise.**Случаи, без липсващи стойности на променливите в една група (клетка) са включени в анализа на тази група. В случая може да има липсващи данни за променливи, използвани в други групи.
- **Report values.** Липсващи стойности за фактор променливи се разглеждат като отделна категория. Целия изход е създаден за тази допълнителна категория. Честотните таблици включват категориите за изчезнали стойности. Липсващи стойности за фактор променлива са включени, но означени като липсващи.

EXAMINE Command Additional Features

Процедурата Explore използва **EXAMINE** синтаксиса. Синтаксиса на езика за команди също ви позволява да:

- . Извиква общия изход и графики в допълнение към изхода и графиките за групите дефинирани от потребителя(С командата **TOTAL**)
- Посочете обща скала за група от **boxplots** (С подкомандата **SCALE**)
- Посочете взаимодействия на фактор променливи (С подкомандата **VARIABLES**).
- Посочете персентили, различни от тези по подразбиране (С подкомандата **PERCENTILES**).
- Изчислете персентили според всеки от 5 методи (С подкомандата **PERCENTILES**).
- Посочете всяко правомощие, трансформация на **Spread-versus-level plots** (С подкомандата **PLOT**).
- Посочете броя на екстремни стойности да бъдат показани (С подкомандата **STATISTICS**).
- Задайте параметри за мрежата на М-оценителите, солидни оценки на място (С подкомандата **MESTIMATORS**).

Вижте ***Command Syntax Reference*** за пълна информация за синтаксиса.

Глава 16. Таблици на спрегнатост

Таблици на спрегнатост

Процедурата "Таблици на спрегнатост" формира двумерни и многомерни таблици и предлага разнообразие от тестове и начини на свързаност за двумерни таблици. Структурата на таблицата и дали категориите са подредени определят какъв тест или мярка да ползват.
Статистика на табове на спрегнатост и мерки на свързаност могат да бъдат изчислени само за двумерни таблици. Ако посочите ред, колона и фактор на слоя (контролна променлива), процедурата за табове на спрегнатост формира един панел за асоциативна статистика и мерки за всяка стойност от фактора на слоя(или комбинация на стойности за две или повече контролни променливи). Например, ако пола е фактор на слоя за таблица със стойности женен/омъжена(yes, no) съпоставена със живот(вълнуващ, рутинен, скучен) резултатите от двумерната таблица са пресметнати по различен начин – тези за жените са разделени от тези за мъжете и са визуализирани като панели един след друг.

Пример. Дали клиентите от малките компании са по печеливши в продажбите на услуги (например, обучение и консултиране) в сравнение с тези от по-големите фирми? От табове на спрегнатост, може да се научи, че по-голямата част от малките предприятия (по-малко от 500 служители), реализират високи печалби от услуги, докато по-голямата част от големи компании (повече от 2500 служители), реализират ниски печалби от услуги.

Статистики и мерки на асоциация (свързаност): Pearson chi-square, likelihood-ratio chi-square, linear-by-linear association test, Fisher's exact test, Yates' corrected chi-square, Pearson's *r*, Spearman's rho, contingency coefficient, phi, Cramér's *V*, symmetric and asymmetric lambdas, Goodman and Kruskal's tau, uncertainty coefficient, gamma, Somers' *d*, Kendall's tau-*b*, Kendall's tau-*c*, eta coefficient, Cohen's kappa, relative risk estimate, odds ratio, McNemar test, and Cochran's and Mantel-Haenszel statistics.

Данни. За да се определят категориите на всяка таблична променлива, използвайте стойности цифрова или низова (осем или по-малко байтове) променлива. Например, за пол, бихте могли да кодирате данните като 1 и 2, или като мъжки и женски.

Предположения. Някои статистически данни и мерки предполагат подредени категории (редни данни) или количествени стойности (интервал или съотношение на данни), както е описано в раздел за статистика. Други са валидни когато променливите от таблица съдържат неподредени категории (номинални данни). За **chi-square-based** статистика (**phi, Cramér's *V*** , и непредвиден коефициент), данните трябва да бъдат случайна извадка от **multinomial** дистрибуция.
Забележка: Поредните променливи могат да бъдат или цифрови кодове, които представляват категории (например, 1 = ниско, 2 = средно, 3 = високо) или стойности от тип низ. Въпреки това, азбучния ред на низови стойности се приема че отразява вярно реда на категориите. Например, за низ с стойности: ниски, средни и високи, реда на категориите се тълкува като високо, ниско, средно, което не е правилния ред. Като цяло е по-надеждно да се използват цифровите кодове за да се представят редни данни.

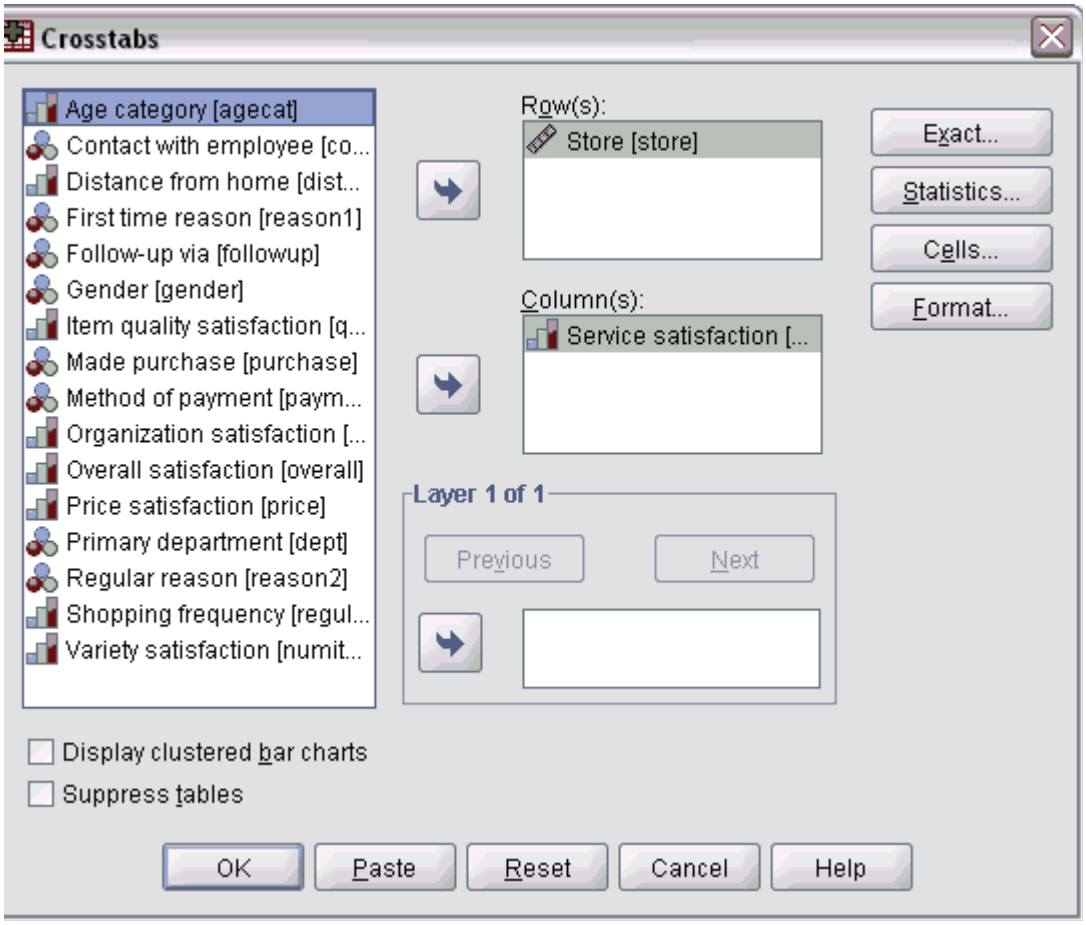
За да се направят табове на спрегнатост напреавете следното:
Изберете от меню:

- Analyze
  - Descriptive Statistics
    - Crosstabs...

Изберете един или повече редове и една или повече колонки.
По желание можете да:

- Изберете една или повече контролни променливи.

- Изберете Статистика за тестове и мерки на спрегнатост за двумерни таблици или подтаблици.
- Изберете Клетки за наблюдавани и очаквани стойности, проценти, и остатъците.
- Щракнете върху Формат за да контролирате реда на категориите.



Фигура 16-1 *Crosstabs dialog box*

Слоеве на табове за спрегнатост

Ако изберете една или повече променливи за слой, отделна **crosstabulation** се създава за всяка категория на всяка променлива за слой (контролна промелива). Например, ако имате една променлива за ред, една за колона и една променлива за слой с две категории, вие ще изкарате двумерна таблица за всяка категория от слоя. За да направите друг слой с контролни променливи, щракнете Напред. Подтаблиците се правят за всяка комбинация от категории за всяка променлива от първи слой, втора променлива от втори слой и т.н. Ако статистиката и мерки на спрегнатост се изискват, те се прилагат само за двумерни подтаблици.

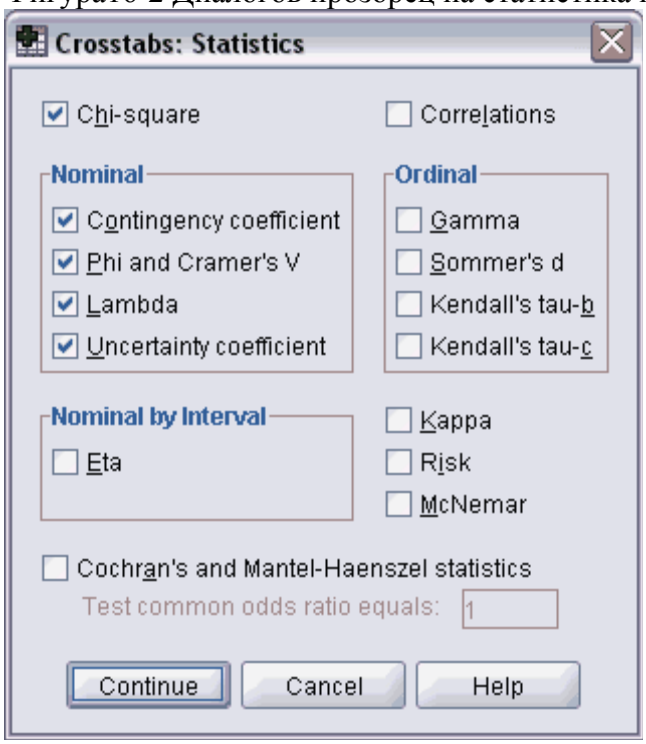
Crosstabs Клъстер Бар Графики

Показва клъстерните диаграми. Клъстерирана диаграма помага за обобщението на вашите данни във редица от случаи. Има по един клъстър от ленти за всяка стойност на променлива, която сте посочили в редове. Променливата, която определя лентите на всеки клъстер е променлива, която е зададена в меню колони. Има една редица от различен цвят

или шарени барове за всяка стойност на тази променлива. Ако зададете повече от една променлива в **Columns Or Rows**, диаграма се прави за всяка комбинация.

Статистика на слоеве на спрегнатост

Фигура16-2 Диалогов прозорец на статистика на слоеве на спрегнатост



Фигура16-2 Диалогов прозорец на статистика на слоеве на спрегнатост

За таблиците с два реда и две колони, изберете **Chi-square** за да изчислите **Chi-square** на **Pearson**, вероятността на съотношение **chi-square**, точен тест на **Fisher** и коригирана **Chi-square** на **Yate -** (приемствена корекция). За 2 × 2 таблици, точния тест на **Fisher** се изчислява, когато една таблица, която не е резултат от липсващи редове или колони в по-голяма таблица, притежава клетка с очаквана стойнот по-малко от 5. Коригираната **Chi-square Yate's** се изчислява за всички други 2 × 2 таблици. За таблиците с произволен брой на редовете и колоните, изберете **Chi-square** за изчисляване на **Chi-square** на **Pearson**, вероятността на съотношение **Chi-square**. Когато и двете променливи са количествени, **Chi-square** добива линейно-по-линейно асоциативен тест.

Корелация. За таблици, в които двата реда и колони съдържат предрени стойности, корелациите добиват коефициента на съответствие на **Spearman**, **rho**(само числови данни). **Rho** на **Spearman** е мярка за асоцииране между проъчки по ранг. Когато и двете променливи (фактори) са количествени, корелациите добиват коефициента на корелация на **Pearson**, който е мярка на линейни асоцииране между променливите.

Номинална. За номинални данни (не е **intrinsic** ред, като католици, протестанти, и евреи), можете да изберете непредвидени коефициент, **Phi** (коефициент) и **Cramer's V, Lambda** (симетрични и асиметрични **lambdas** и **Goodman and Kurskal's tau**), и коефициент на неопределимост.

◴ **Коефициент на неопределеност**: Мярка за асоциация, базирана на **Chi-square**. Стойността варира между 0 и 1, с 0, се показва липсата на връзка между променливите и стойностите на реда и колоната и стойности близки до 1 показват висока степен на асоцииране между променливите. Максималната възможна стойност зависи от броя на редовете и колоните в таблицата.

◴ **Phi and Cramer's V**. **Phi** е основана на **chi-square** мярката на асоциация, която включва разделяне на **Chi-square** статистика от обема на извадката, и като корен квадратен от резултата. **Cramer's V** е мярка за асоциация, базирана на **Chi-square**.

◴ **Lambda.** Мярка за асоциация, която да отразява пропорционално намаляване на грешка при определянето на стойности на зависима променлива като се използват стойностите на независимата променлива. Стойност 1 означава, че независимата променлива дава възможност да се определи зависимата променлива. Стойност 0 означава, че независимата променлива не дава възможност да се определи зависимата променлива.

◴ **Коефициент на несигурност.** Мярка за асоциация, която показва пропорционално намаление на грешка, когато стойностите на една променлива се използват за предвиждане на стойностите на други променливи. Например, стойност 0,83 показва, че познаването на една променлива намалява грешка при прогнозирането на стойностите на други променливи, с 83%. Програмата изчислява симетрични и асиметрични версии на Коефициент на несигурност.

◴ **Пореден.** За таблици, в които двете редове и колони съдържат подредени стойности, изберете Гама (**zero-order** за 2-мерни таблици и условно за 3 и 10-мерни таблици), **Kendall's tau-b**, и **Kendall's tau-c**. За прогнозиране на категории на колони от категории от ред, изберете **Somers'd**.

◴ **Гама.** Симетрична мярка на връзка между две редни променливи, с обхват между -1 и 1. Стойности, близки до абсолютната стойност 1 показват силна връзка между две променливи. Стойности, близки до 0, показват малко или никаква връзка. За 2-мерни таблици, нулево подредени гами са показани. За 3-мерни до N-мерни, таблици, условни гами са показани.

◴ **Somers'd**. Мярка за асоцииране между две редни променливи, с обхват между -1 и 1.Стойности, близки до абсолютната стойност 1 показват силна връзка между две променливи и стойности близки до 0, показват малко или никаква връзка между променливите. **Somers'd** е асиметрично разширение на гамата, която се отличава само във връзка с включването на броя на двойки вързани на независима променлива. Симетрична версия на тази статистика е изчислена.

◴ **Kendall's tau- b** Непараметрична мярка на корелация за редни или **ranked** променливи, които вземат под внимание **ties**. Знакът на коефициента показва посоката на връзката, а неговата абсолютна стойност показва силата, с по-големи абсолютни стойности показва по силни взаимоотношения. Възможните стойности варират от -1 до 1, а стойността на -1 или 1 може да бъде получена само от **square tables**.

◴ **Kendall's tau- c** Непараметрична мярка на асоциацията за редни променливи, която **take ties into account**. Знакът на коефициента показва посоката на връзката, и неговата абсолютна стойност показва силата, по-големи абсолютни стойности показват по силни

взаимоотношения. Възможни стойностите варират от -1 до 1, а стойността на -1 или 1 може да бъде получена само от **square tables**.

Номинална с интервал. Когато една променлива е категориина, а другата количествена, изберете ETA. Категориината променлива трябва да е кодирана числено.

ETA. Мярка за асоциация, която варира от 0 до 1, с 0, показваща липсата на връзка между променливи от ред и колона и стойности, близки до 1 показващи висока степен на асоциация. ETA е подходяща за зависима променлива, измерена по интервална скала (например, доход) и за независима променлива с ограничен брой на категории (например, пол). Две ETA стойности са изчислени: едната третира променлива от ред като интервална променлива и другата третира променлива от колона като интервална променлива.

Kappa. Cohen's kappa измерва съгласуването между оценката на двата **raters** когато и двата са класирани като един и същ обект. При стойност 1 се показва перфектно съгласуване. При стойност 0 показва, че съгласуването не е сигурно. **Kappa** е достъпна само за таблици, в които и двете променливи използват едни и същи стойностти: категория и брой на категории.

Risk. За 2 x 2 таблици, мярка на силата на връзката между наличието на фактор
и настъпването на дадено събитие. Ако доверителения интервал за статистиката включва стойност 1, не може да се предположи, че фактора е свързан със сибитието. Нечетнотото съотношение може да се използва като оценка или относителен риск когато възникването на фактора е рядко.

McNemar. Непараметричен тест за две свързани дихотомни променливи. Тества за промени в отговорите като използва **Chi-square** разпределение. Полезно за откриване на промени в отговорите поради експериментална интервенция в "пред-и-след" дизайни. За по-големи квадратни таблици, теста на **McNemar-Bowker** се отчита.

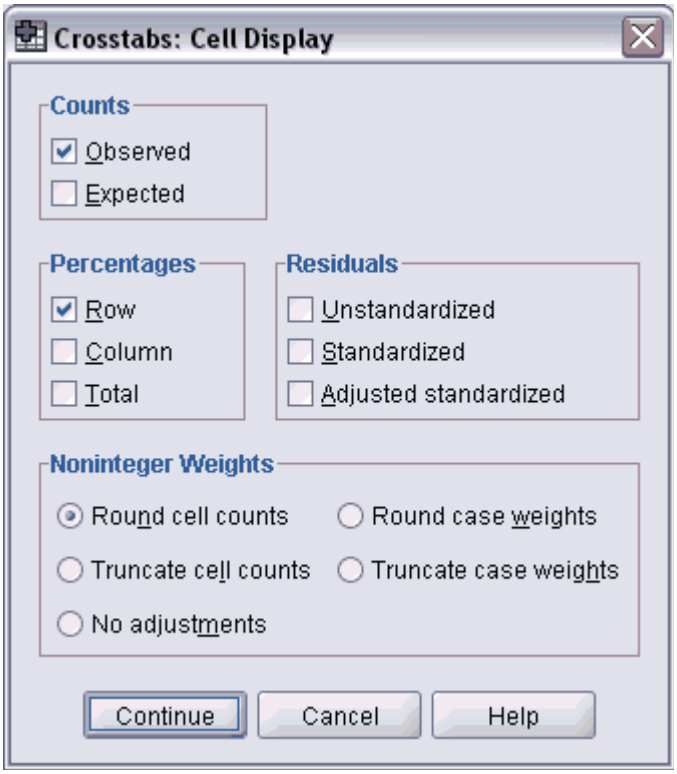
Cochran's and Mantel-Haenszel statistics. Статистиката на **Cochran and Mantel-Haenzel** може да се използва за да тества за независимост между дихотомната фактор променлива и дихотомната промелива за отговор, условно над **covariate** модели, дефинирани от един или повече промеливи от слой(контролни). Имайте предвид, че докато другите статистически данни са изчислени слой по слой, статистически данни на **Cochran and Mantel-Haenzel** се изчисляват веднъж за всички слоеве.

За да ви помогне да откриете модели в данните, които допринасят за значителния **Chi-square** тест, на **Crosstabs** процедурата показва очакваните честоти и три вида остатъци (отклонения), който измерват разликата между установените и очакваните честоти. Всяка клетка от таблицата може да съдържа комбинация от избрани брой, проценти, и остатъци.

Counts. Броят на наблюдаваните случаи и броя на очакваните случаи, ако промеливите по реди и колона са независими една от друга.

Percentages. Процентите, може да се добавят през редове и колони. Процентите
от общия брой на случаите, представени в таблицата (един слой) също са на разположение.

Residuals. Сурови нестандартизирани остатаци дават разликата между наблюдаваните и очакваните стойности.Стандартизираните и коригираните стандартизирани са също налични.



Фигура 16-3 *Crosstabs Cell Display диалог*

- **Unstandardized**. Разликата между наблюдавана стойност и очакваната стойност. В очакваната стойност е броят на случейте, които бихте очаквали в клетката, ако нямаше връзка между двете променливи. Положителния остатък показва, че има повече случаи в клетката, отколкото би имало, ако променливи на ред и колона са независими.

- **Standardized**. Остатъка разделен от оценката на неговото стандартно отклонение. Стандартизираните остатъци, които са известни като остатъците **Pearson**, имат стойност 0 и стандартно отклонение 1.

- **Adjusted standardized**. Остатъка за клетка (наблюдава минус очакваната стойност), разделена от оценка на неговата стандартна грешка. Крайния стандартизиран остатък е изразен в стандартно отклонени единици над или под средната.

- **Noninteger Weights**. Броят на клетките са обикновено целочислени стойности, тъй като те представляват броя на случаи във всяка клетка. Но ако файла с данни е претеглен с промелива за тегло с частична стойност (например, 1.25), броят на клетките може да бъде частична стойност. Можете да съкратите или приравните, преди или след изчисляване на броят на клетките или да използвате частична употреба клетки за показване двете таблица и статистически изчисления.

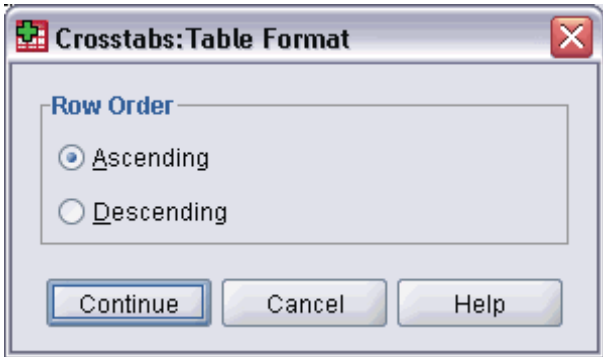
- **Round cell counts**. Теглата на случейте се използват както са, но тежестите в клетките се закръгляват преди изчисляването на статистиката.

- **Truncate cell counts**. Теглата на случейте се използват както са, но тежестите в клетките се трият преди пресмятането на статистическите данни.

- **Round case weights**. Теглата на случейте са закръглени преди употреба.
- **Truncate case weights**.Теглата на случейте са пресечен преди употреба.
- **No adjustments**. на случейте се използват както са и частичните клетки се използват.

Въпреки това, когато **Exact Statistics** (достъпна само с точния вариант тестове) се изисква, натрупаните тежести в клетки са или изтрити или закръглени преди изчисляване на **Exact Statistics**

Crosstabs Table Format



Фигура 16-4 *Crosstabs Table Format диалог*

Можете да подреждате таблицата по редове във възходящ или низходящ ред на стойностите на променливите.

Глава 17. Обобщено описание на метрични данни

Процедурата **Summarize** изчислява подгрупова статистика за променливи в категориите на една или повече групи променливи. Всички нива на групиране на променливи са **crosstabulated**. Можете да изберете реда, в който се показват статистическите данни. Също така е показана обобщена статистика за всяка отделна променлива във всички категории. Стойности от данни във всяка категория могат да бъдат изброени или **supressed**. С големи масиви от данни, можете да изберете само да изредите първите случаи N.

Пример. Какъв е средният размер продажби по регион и по индустрия на клиентите? Можете да откриете, че средната сума на продажбите с корпоративни клиенти е малко по-висока в западната област, отколкото в други региони, в региона на Западните се получава най-висок среден размер на продажбите.

Статистика.
Сума, броят на случаите, средна, медиана, групирана медиана, стандартна грешка на средната стойност, минимум, максимум, обхват, променлива стойност на първата категория на обединената променлива, променлива стойност на последната категория на обединената променлива, стандартно отклонение, различия, **kurtosis**, стандартна грешка на **kurtosis**, **skewness**, стандартна грешка на **skewness**, процент от общата сума, процент от общото N, процент от сумата, в процент от N, средна геометрична и средна хармонична.

Данни.
Групиращи променливи са категорични променливи, чиито стойности могат да бъдат цифрови или низови. Броя на категориите трябва да е достатъчно малък. Другите променливи трябва да може да бъдат класирани.

Предположения.
Някои от изборните подгрупи, като средната стойност и стандартното отклонение, се основават на нормализиращата теория и са подходящи за количествени променливи със симетрични дистрибуции. **Robust statistics, such as the median and the range**, са подходящи за количествени променливи, които могат или не могат да отговарят на допускането за нормалност.

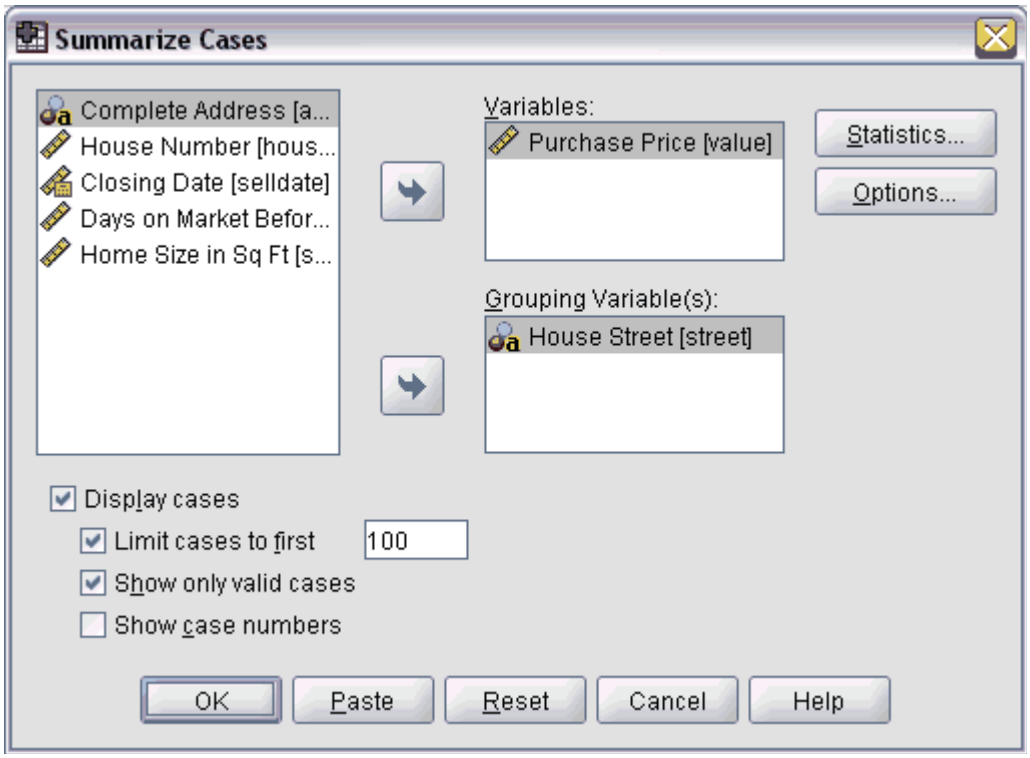
Получаване на Case Summaries

Изберерет от менюто:

- **Analyze**
  Reports
  Case Summaries...

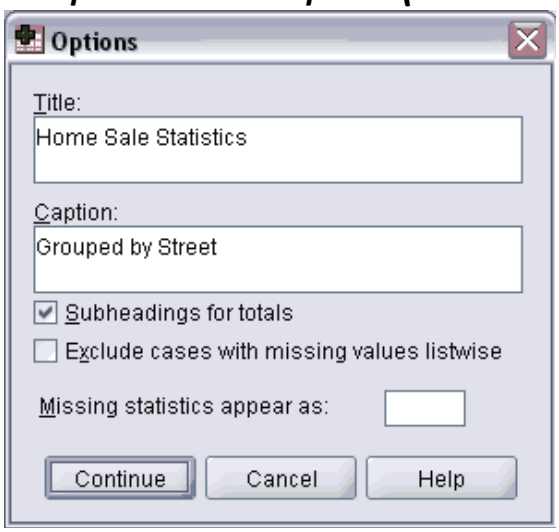
Изберете една или повече променливи. По желание можете да:

- Изберете една или повече групови променливи, които да разделят вашите данни на подгрупи.
- Щракнете върху Опции за промяна на заглавието, добавете надпис под изхода, или изключете случаите, с липсващи стойности.
- Кликнете Статистика за изборна статистика.
- Изберете **Display cases** за да се изброят случаите, във всяка подгрупа.
- По подразбиране, системата изкарва само първите 100 случаи на вашия файл.Можете да увеличите или намалите стойността до допустимите случаи за първи N(елемента) или отмяна да деселектирате този елемент, за да се изброят всички случаи.



Фигура 17-1 *Обобщен Cases диалог*

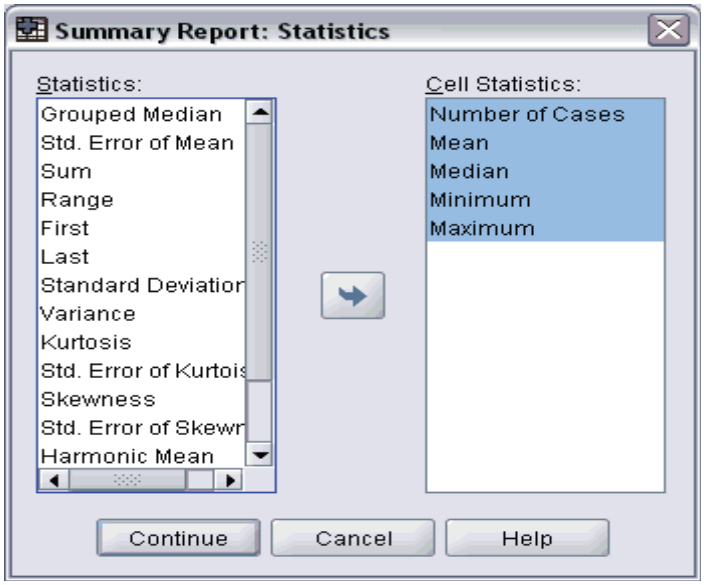
Опции на обобщение(Summarize Options)



Фигура17-2 *Диалогов прозорец с опции*

Summarize ви позволява да промените заглавието на вашия изход или да добавите надпис, който ще се появи под изходната таблица. Можете да контролирате затварянето на линиите в заглавия и надписи, като въведете \n, където и да искате да вмъкнете нов ред в текста. Можете също така да изберете да показвате или **suppress** подпозиции за суми и да включите или изключите случаи с липсващи данни за всяка от променливите, използвани в някои от анализите. Често това е желателно за да се обозначат липсващите случаи на продукцията, с период или звездичка. Въведете символ, фраза, или код, които бихте искали да се появи, когато липсва стойност, в противен случай, няма специален начин който се прилага към липсващите случаи в изхода.

Обобщена статистика



Фигура 17-3 *Диалог за обобщена статистика*

Можете да изберете една или повече от следните подгрупи статистика за променливите в рамките на всяка категория на всяка групова променлива: сумата, броя на случаите, средна, медиана, групирани медиана, стандартна грешка на средната стойност, минимум, максимум, обхват, променлива стойност на първата категория на обединената променлива, променлива стойност на последната категория на обединената променлива, стандартно отклонение, дисперсията, **kurtosis**, стандартна грешка на **kurtosis**, **skewness**, стандартна грешка на **skewness**, процент от общата сума, процент от общата N, процент от сумата, процент от N, средна геометрична, средна хармонична. Редът, в който изглежда статистиката в клетката Статистика е реда, в които те ще бъдат показани в продукцията. Обобщената статистика се показва за всяка променлива във всички категории.

First. Показва първата стойност данни, срещани в файла с данни.
Geometric Mean. В n-тия **root** на продукта от стойности на данни, където N представлява броят на случаи.
Grouped Median. Медиана, която е изчислена за данни, които са кодирани в групи. Например, с възрастови данни, ако всяка стойност в 30-те години е кодирана 35, всяка стойност в 40-те години е кодирана 45, и т.н., средната на групирана медиана се изчислява от кодирани данни.

Harmonic Mean. Използва се за оценка на групов среден размер, когато размерите на извадката в групите не са равни. Средната хармонична е общият брой на пробите, разделена на сумата от реципрочните на примерните големини.
Kurtosis. Мярка за степента, до която наблюденията се групират около една централна точка. За нормално разпределение, стойността на **kurtosis** статистика е нула. Положителен **kurtosis** показва, че наблюденията се групират повече и има повече опашки от тези в нормално разпределение, а отрицателни **kurtosis** показва, че наблюденията се групират по-малко и са с по малко опашки.
Last. Показва последно данните за стойността срещани във файла с данни.
Maximum. Най-голямата стойност на числова променлива.
Mean. Мярка на централната тенденция. Средно аритметично, сумата разделена на броя случаи
Median. Стойност над и под която спадат половината от случаите на 50-процентна. Ако е налице четен брой от случаи, медианата е средна стойност на двата средни случаи, когато те са сортирани във възходящ или низходящ ред. Средната мярка измерител на централната тенденция която не се влияе от отдалечените стойности (за разлика от средната, която може да бъде засегната от няколко изключително високи или ниски стойности).

Minimum. Най-малката стойност на числова променлива.
N.Брой на случаите (наблюдения или записи).
Percent of Total N. Процент от общия брой на делата във всяка категория.
Percent of Total Sum. Процент от общата сума за всяка категория
Range. Разликата между най-голямата и най-малката стойност на числова променлива, максималната минус минималната
Skewness. Мярка на асиметрия на разпределението. Нормалното разпределение е симетрично и има **skewness** стойност 0. Разпределение със значителен положителен skewness има дълга права опашка. А разпределение със значителен отрицателен skewness има дълга лява опашка. Като насока, една skewness стойност два пъти по голяма от нейната стандартна грешка е взета да покаже нейното отклонение от симетрията.
Standard Error of Kurtosis. Съотношението на kurtosis към стандартната си грешка може да се използва като тест за нормалност (това значи че може да отхвърли нормалността, ако съотношението е по-малко от -2 или по-голямо от 2). Голяма положителна стойност за kurtosis показва, че опашките на разпространението са по-дълги от тези на нормалното разпределение; отрицателна стойност за kurtosis показва по-къси опашки (стават като тези на **box-shaped uniform distribution**).
Standard Error of Skewness. Съотношението на skewness към стандартната си грешка може да се използва като тест за нормалност (това значи че може да отхвърли нормалността, ако съотношението е по-малко от -2 или по-голямо от 2). Голяма положителна стойност за skewness показва, дълга права опашка; крайна отрицателна стойност показва, дълга лява опашка.
Sum. Сумата, или общата сума от стойности, във всички случаи с нелипсващи стойности
Variance. Мярка за разсейване около средната стойност, равна на сумата на отклоненията от средната разделена с едно по-малко от броя на делата на квадрат. Различията се измерват в единици, които са квадрати на тези на променливата сама по себе си.

Глава 18. Погрупово описание на метрични данни

Средна аритметична

Процедурата ***Means*** изчислява средноаритметична на подгрупите и свързана едномерна статистика за зависими променливи в категориите на една или повече независими променливи. Възможно е да се намери еднофакторен анализ на дисперсята, ета и тестове за линейност.

Извадка. Измерете средното количество мазнина, абсорбирано при готвене с три различни вида олио за готвене и направете еднофакторен анализ на дисперсията, за да разберете, дали средноаритметичната е различна.

Статистика. Сума, брой случаи, средна, медиана, групирана медиана, стандартна грешка на средната, минимум, максимум, размах, променлива стойност на първата категория на групираната променлива, променлива стойност на последната категория на групираната променлива, стандартно отклонение, дисперсия, коефициент на ексцес /наклон, стандартна грешка на коефициента на ексцес, коефициент на асиметрия, стандартна грешка на коефициент на асиметрия, процент на крайната сума, процентно разпределение, геометрична средна, хармонична средна. Възможностите включват анализ на дисперсята, ета, ета-квадрат, тестове за линейност, R и R^2.

Данни. Зависимите променливи са количествени, а зависимите променливи са категорийни. Стойностите на категорийните променливи могат да бъдат числови или буквени.

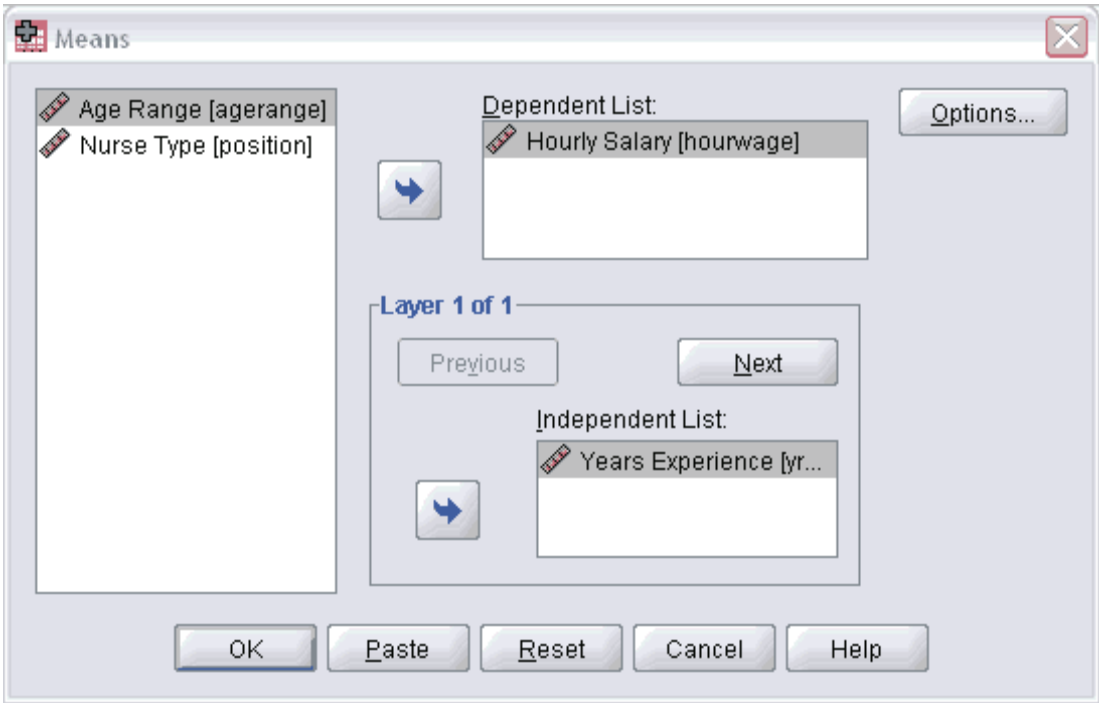
Предположения. Някои от възможните подгрупови статистики, като средноаритметичната и стандартното отклонение, се основават на нормална теория и са подходящи за количествени променливи със симетрични разпределения. Robust statistics, като медианата, са подходящи за количествени променливи , които могат или не могат да бъдат приети за нормалност. Анализът на дисперсията е строг към отклоненията от нормалността, но данните във всяка клетка трябва да са симетрични. Анализът на дисперсията също така предполага, че групите идват от популациите с равни дисперсии. За да се провери това твърдение се използва теста за хомогенност на дисперсията, който може да се намери в процедурата One-Way ANOVA.

За да се намери средноаритметичната на подгрупите:

От менюто се избира:

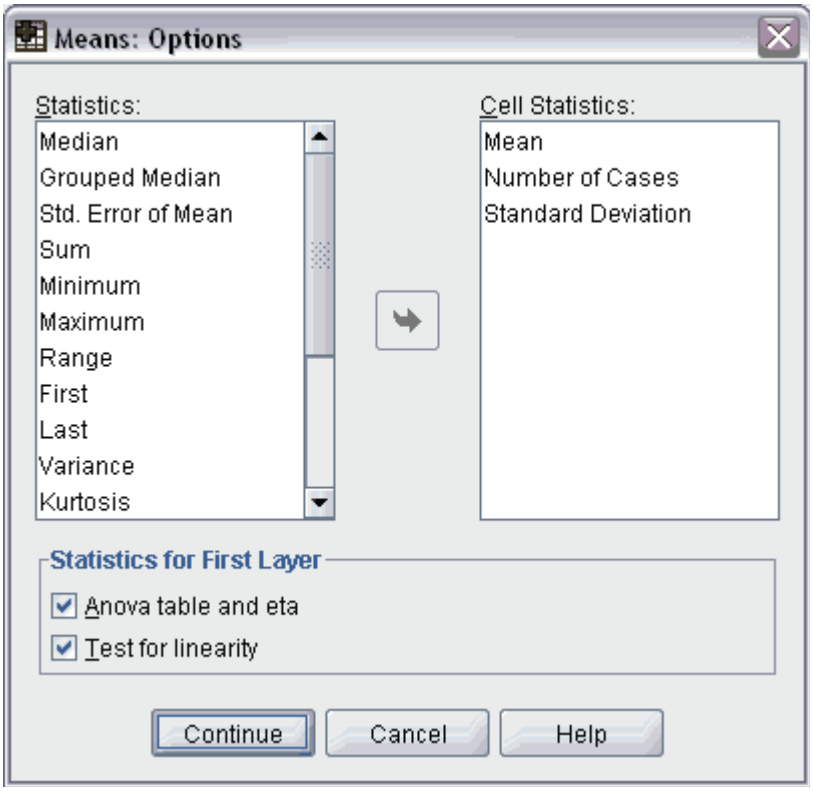
- Analyze
  Compare Means
  Means...

Изберете една или повече зависими променливи.
Използвайте един от следните методи, за да изберете категорийни независими променливи:
Изберете една или повече независими променливи. Различните резултати са показани отделно за всяка независима променлива.
Изберете един или повече слоя за независимите променливи. Всеки слой разделя извадкаа.
Ако имате една независима променлива в Слой 1 (Layer 1) и една независима променлива в Слой 2 (Layer 2), резултатите се появяват в една кръстосана таблица, като са едни срещу други в различни таблици за всяка независима променлива.
Има възможност, когато кликнете на Options за опции на статистиките, анализ на дисперсионната таблица, ета, ета на квадрат, R и R^2.



Фигура 18-1 *Прозорец на средноаритметичната*

Опции на MEANS



Фигура 18-2 *Прозорец на опции на MEANS*

Можете да изберете една или повече от следните статистики на подгрупите за променливите във всяка категория на всяка групираща променлива: сума, брой случаи, средна, медиана, групирана медиана, стандартна грешка на средната, минимум, максимум, размах, променлива стойност на първата категория на групираната променлива, променлива стойност на последната категория на групираната променлива, стандартно отклонение, дисперсия, коефициент на ексцес /наклон, стандартна грешка на коефициента на ексцес, коефициент на асиметрия, стандартна грешка на коефициент на асиметрия, процент на крайната сума, процент на крайният резултат на *N*, percentage of sum in, percentage of *N* in(?), геометрична средна, хармонична средна. Възможностите включват анализ на дисперсята, ета, ета на квадрат, тестове за линейност *R* и *R*2.

First. Показва първата стойност на данните, срещани във файла с данни.
Geometric Mean. (Геометрична средна) Корен от *N* на резултата от стойността на данните, където *N* е броят на случаите.
Grouped Median. (Групирана средна) Мредна, която се изчислява за данни, кодирани в група. Например за данни за възраст, ако всяка стойност в трийсетте е кодирана 35, всяка стойност в четирийсетте е кодирана като 45, и т. н., груповата средна е средната, изчислена от кодираните данни.
Harmonic Mean. (Хармонична средна) Използва се, за да се оцени средния размер на групата, когато размерите на извадката в групата не са равни. Хармоничната средна представлява крайният брой на извадките, разделен на сумата на реципрочните на размерите на извадката.
Kurtosis. (Коефициент на ексцес) Мярка за степента, до която наблюденията се групират около централна точка. За нормално разпределение стойността на статистиката на коефициента на ексцес е нула. Положителният коефициент на ексцес показва, че наблюденията групират повече и имат по-дълги опшки от тези в нормалното разпределение, а отрицаленият коефициент на ексцес показва, че наблюденията групират по-малко и имат по-къси опашки.
Last. Показва последните стойности на данните, които са във файла.
Maximum. (Максимум) Най-голямата стойност на числената променлива.
Mean. (Средноаритметична) Мярка за central tendency(?). Аритметичната средна, сумата, разделена на броя случаи.
Median. (Медиана) Стойността над и под която половината от случаите отпадат, петдесетия персентил. Ако броят на случаите е четно число, медианата е средната на двата средни случая, когато са подредени по възходящ или низходящ ред. Медианата е мярка на central tendency(?) не е чувствителна към outlying values (за разлика от средноаритметичната, която може да се влияе от няколко много високи или много ниски стойности).
Minimum. (Минимум) Най-малката стойност на числената променлива.
N. Броят случаи (наблюдения или записи).
Percent of total N. Процентът на крайния брой случаи във всяка категория.
Percent of total sum. Процентът на цялата сум във всяка категория.
Range. (Размах) Разликата между най-голямата и най-малката стойност на числова променлива, максимумът минус минимумът.
Skewness. (Коефициент на асиметрия) Мярката за асиметрия в разпределението. Нормалното разпределение е симетрично и има коефициент на ексцес равен на нула. Разпределението със забелевителен положителен коефициент на ексцес има дълга дясна опашка. Разпределение със голям отрицателен коефициент на ексцес има дълга лява опашка. Като насока, коефициент на ексцес със стойност два пъти по-голяма от неговата стандартна грешка се смята, че не е симетричен.
Standard Error of Kurtosis. (Стандартна грешка на коефициента на ексцес) Отношението на коефициента на ексцес към неговата стандартна грешка може да се използва като тест за нормалност (т. е. можете да сметнете нормалността за незначителна, ако отношението е по-малко от -2 или по-голямо от +2). Голяма положителна стойност на коефициента на ексцес показва, че опашките на разпределението са по-дълги от тези на нормалното разпределение; отрицателна стойност на коефициента на ексцес предполага по-къси опашки (които приличат на тези от box-shaped uniform distribution(?)).
Standard Error of Skewness. (Стандартна грешка на коефициента на асиметрия) Отношените между коефициента на асиметрия към неговата стандартна грешка може да се използва като тест за нормалност (т. е. можете да сметнете нормалността за незначителна, ако отношението е по-малко от -2 или по-голямо от +2). Голяма положителна стойност на коефициента на асиметрия ще покаже дълга дясна опашка; голямо отрицателно число предполага дълга лява опашка
Sum. (Сума) Сумата или крайната сума на стойностите при всички случаи с известни стойности.

Variance. (Дисперсия) Мярка на дисперсията около средноаритметичната, равна на сумата на отклоненията от средноаритметичната, повдигнати на квадрат, разделена на броя случаи минус едно. Дисперсията се измерва в единици, които са квадратът на самата променлива.

Статистики за първия слой (Layer)

Anova table and eta. Показва еднофакторна таблица на анализ на дисперсията и изчислява ета и ета квадрат (мярка за свързаност) за всяка независима променлива в първия пласт.

Test for linearity. (Тест за линейност) Изчислява сумата на квадратите, степените на свобода и средноаритметичната Изчислява сумата на квадратите, степените на свобода и квадрата на средната, свързан с линейните и нелинейните компоненти, също така и степента на F, R и R квадрат. Линейността не се изчислява, ако независимата променлива е къса дума.

Глава 19. Т-тестове

T Tests

Съществуват три вида *t* тестове:

Independent-samples t test (two-sample t test). Т-тестове за независими извадки (т тестове на две извадки). Сравнява средноаритметичните на една променлива за две групи от случаи. Предоставени са дескриптивна статистика за всяка група и теста на Левен за равностойност на дисперсиите, както и равна и различна дисперсия на т стойностите и 95% доверителен интервал за разликата в средноаритметичните.

Paired-samples t test (dependent t test). Свързани извадки за т тестове (Зависими т тестове) Сравнява средните на две променливи за една група. Този тест е също и за съвпадащи си двойки или модели на case-control study. Резултатът включва дескриптивни статистики за променливите в теста, корелацията между променливите, дескриптивната статистика за двойките разлики, т теста и 95% доверителен интервал.

One-sample t test. Едноизвадков т-тест. Сравнява средноаритметичната на една променлива с известна стойност или хипотеза за стойността. Дескриптивни статистики за променливите в теста се показват заедно с теста. 95% доверителен интервал за разликата между средноаритметичната и променливата на теста и хипотезата за стойността на теста е част от резултата по подразбиране.

Independent-Samples T Test. T тестове за независими извадки.

Процедурата за Т тестове на независими извадки сравнява средноаритметичните за две групи случаи. При идеален случай за този тест обектите трябва да са разделени на две групи на случаен принцип, така че всяка разлика в резултатите да се дължи на лечението (или липсата на такова), а не на други фактори. Това не се използва, ако сравнявате средния приход при мъжете и жените. Човек не е записан да е мъж или жена на случаен принцип. В такива ситуации трябва да се уверите, че разликите в другите фактори не маскират или увеличават забележителна разлика в средноаритметичните. Разликите в средния приход могат да бъдат повлияни от фактори като образованието (а, не от самия пол).

Пример. Пациенти с високо кръвно налягане са записани на случаен принцип в две групи – за плацебо и лечение. Тези с плацебо лечението получават неактивно хапче, а тези с истинското лечение получават ново лекарство, за което се предполага, че ще намали кръвното налягане. След като пациентите са лекувани два месеца, свързаните извадки на т теста се използват, за да се сравнят средните кръвни налягания за плацебо групата и за групата с истинското лечение. Всеки пациент е измерен веднъж и принадлежи на една група.

Statistics. Статистика. За всяка променлива: размер на извадката, средноаритметична, стандартно отклонение и стандартна грешка на средноаритметичната. За разликата в средноаритметичните: средноаритметична, стандартна грешка и доверителен интервал (вие можете да зададете нивото на доверие). Тестове: тест на Левен за равни дисперсии и за обобщени дисперсии и Т тестове за отделните дисперсии (separate-variances) за равенство на средните.

Data. Данни. Стойностите на количественената променлива, която ни интересува е на една колона на файла с данни. Процедурата използва групираща променлива с две стойности, за да раздели случаите на две групи. Групиращата променлива може да бъде числена или дума (като *да* или *не*). Като друга възможност можете да използвате количественена променлива – възраст – за да разделите случаите на две групи, специфицирайки точка на раделяне на групите (cutpoint) (точка на разделяне 21 разделя *възраст* на групи от *под 21 годишни* и *над 21 годишни*).

Assumptions. Предположения. За т тестове с равни дисперсии наблюденията трябва да са независими, случайните извадки от нормалните разпределения със същата дисперсия на популацията. За т тестовете с различни дисперсии наблюденията би трябвало да са независими, случайни извадки от нормалните разпределения. two-sample *t* test е строг към разминавания от нормалността. Когато разпределенията се проверяват графично, забележете, че са симетрични и нямат резултати, които са отделени от главната си част.

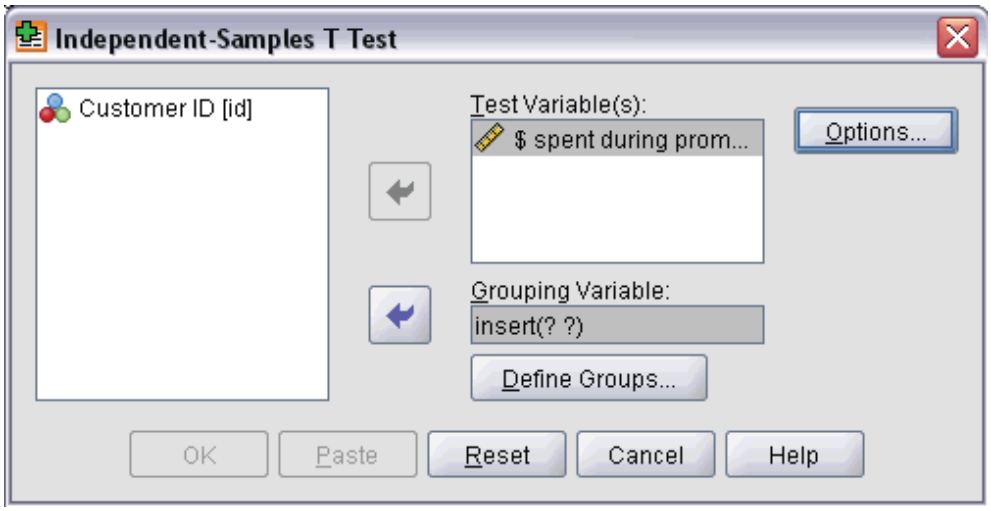
За да пролучите Т тест на независимите извадки

Изберете от менюто:

- Analyze
  - Compare Means
    - Independent-Samples T Test...

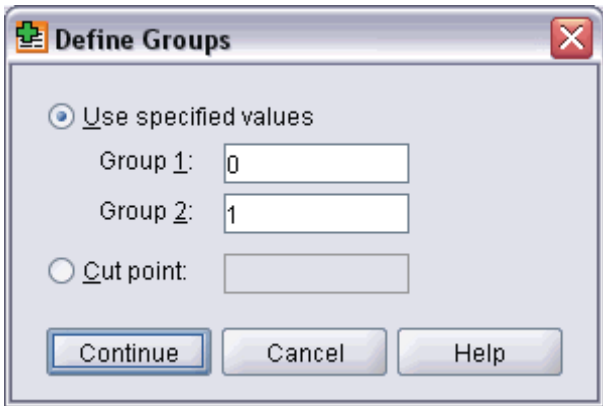
Изберете една или повече променливи за количествения тест. Отделен т тест е изчислен за всяка променлива. Изберете единствена групираща променлива и кликнете Define Groups, за да специфицирате два кода за групите, които искате да сравните.

Можете да кликнете на Options за да контролирате, това как да третирате липсващите данни и нивото на доверителния интервал.



Фигура 20-1 *Прозорец на Independent-Samples T Test*

Independent-Samples T Test Define Groups

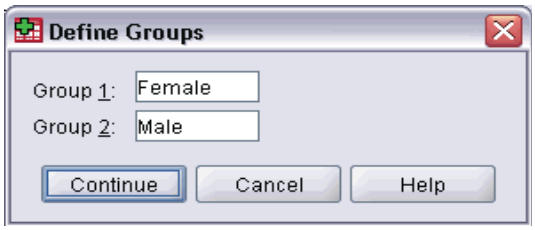


Фигура 20-2 *Прозорец на Define Groups числови променливи*

За чилсловите групи на променливите определете двете групи за т теста, като специфицирате двете стойности или точка на раделяне на групите:

Use specified values. Въведете стойност за Група 1 и още една стойност за Група 2. Случаи със всяка друга стойност са изключени от анализа. Цифрите не трябва да са записани като числа (например 6.25 и 12.5 са валидни).

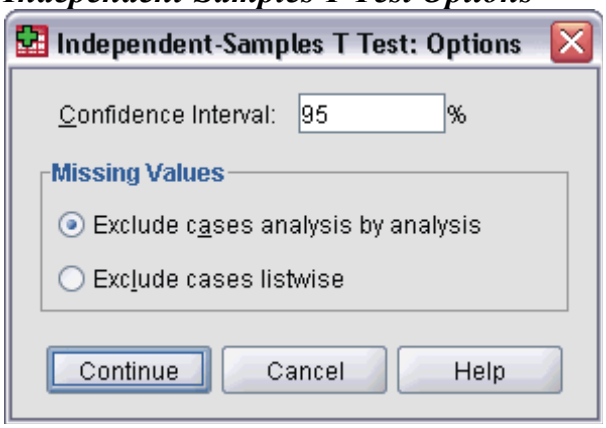
Cutpoint. Въведете цифра, която разделя стойностите на групиращата променлива на две. Всички случаи със стойности, които са по-малки от точката на разделяне на групата от едната група и по-големи или равни на точката на разделяне на групата от другата група.



Фигура 20-3 Прозорец на *Define Groups за номинални променливи*

За да групирате номинални променливи, въведете номинална за Група 1 и друга стойност за Група 2, като например *да* или *не*. Случаи с други номинали са изключени от анализа.

Independent-Samples T Test Options



Фигура 20-4 *Прозорец на Independent-Samples T Test Options*

Confidence Interval. (Доверителен интервал) По подразбиране 95% доверетелен интервал за разликата в средноаритметичните се появява. Въведете стойност от 1 до 99, за да изискате различно ниво на доверетелния интервал.
Missing Values. (Липсващи стойности) Когато проверявате различни променливи и липсват данни за една или повече променливи, можете да зададете на процедурата, кои случаи да включи и кои да изключи.
Exclude cases analysis by analysis. (Изключване на анализ на случаи по анализа) Всеки т тест използва всички случаи, които имат валидни данни за тестваните променливи. Размерите на извадката могат да са различни във всеки тест.
Exclude cases listwise. (Изключване на случаи по списък**)** Всеки т тест използва само случаи, които имат валидни данни за всички променливи, които се използват в поисканите т тестове. Размерът на извадката е постоянен при всички тестове.

Paired-Samples T Test. Свързани извадки в T тест.

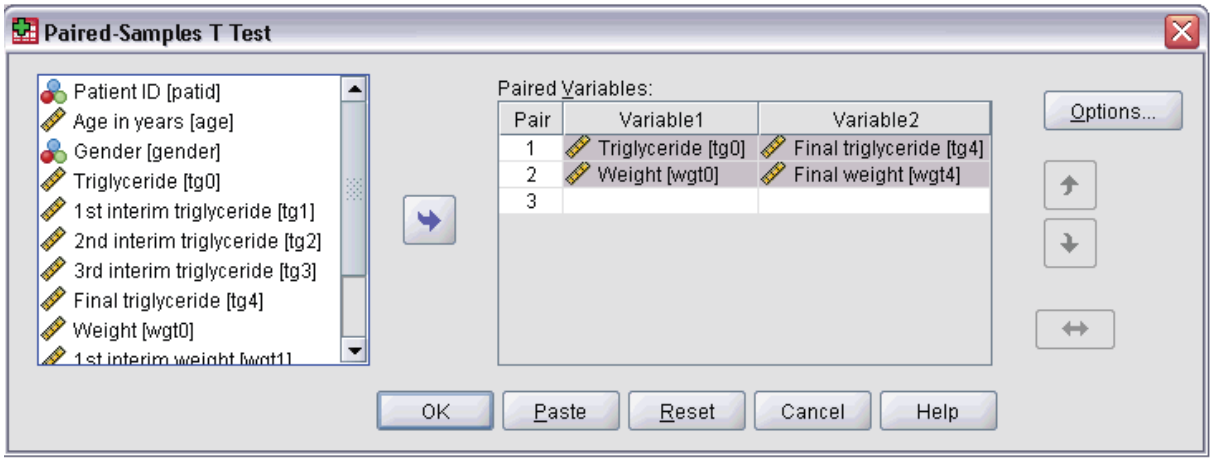
Процедурата ***Свързани извадки в T тест*** сравнява средноаритметичните на две променливи за една група. Процедуратата изчислява разликите между стойностите на двете променливи за всеки случай и проверява дали средната се различава от нула.
Example. Пример. В проучване на високо кръвно налягане всички пациенти са измерени в началото на проучването, назначено им е лечение и пак са измерени. Така всеки има две измервания често наричани измервания *преди* и *след.* Алтернатива е изследване случай - контрола, в което всеки запис в файла с данни съдържа резултатите за пациента и също така за неговата или нейната съответна контрола. В проучването за кръвното налягане пациентите и контролите могат да бъдат разделени по възраст (75 годишен пациент със 75 годишен от групата на контролите).
Statistics. Статистики. За всяка променлива: средна, размер на извадката, стандартно отклонение и стандартна грешка на средната. За всяка двойка променливи: корелация, средна разлика в средноаритметичните, T тест и доверетелен интервал за средната разлика (можете да зададете нивото на доверителност), стандартно отклонение и стандартна грешка на разликата на средната.

Data. Данни. За всеки свързан тест, задайте две количествени променливи (ниво на интервала на измерване или коефициента на измерването). За matched-pairs or case-control study, отговорът за всеки обект на тест и неговия съответен обект от контролите трябва да са в един и същи случай в файла с данни.
Assumptions. Предположения. Наблюденията за всяка двойка трябва да са направени при еднакви условия. Разликите на средната обикновено са разбъркани. Дисперсиите на всяка променлива могат да са равни или различни.

За да получите T тест на свързани извадки.
От менютата изберете:

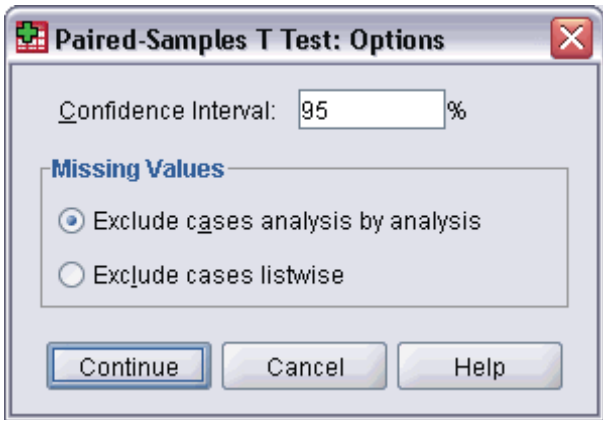
- Analyze
  Compare Means
  Paired-Samples T Test...

Изберете една или повече двойки променливи.
Можете да кликнете Options, за да изберете, как да контролирате липсващите данни и нивото на доверителния интервал.



Фигура20-5 *Прозорец на Paired-Samples T Test*

Paired-Samples T Test Options



Фигура 20-6 *Paired-Samples T Test Options dialog box*

Confidence Interval. Доверителен интервал. По подразбиране се показва 95% доверителен интервал за разликата в средните. Въведете стойност между 1 и 99, за да поискате различно ниво на доверителност.
Missing Values. Липсващи данни. Когато тествате няколко променливи и липсват данни за една или повече променливи, можете да зададете на процедурата, кои случаи да изключи и кои да включи.
Exclude cases analysis by analysis. (Изключване на анализ на случаи от анализа) Всеки т тест използва всички случаи, които имат валидни данни за тестваните променливи. Размерите на извадката могат да са различни във всеки тест.

Exclude cases listwise. (Изключване на случаи по списък**)** Всеки т тест използва само случаи, които имат валидни данни за всички променливи, които се използват в поисканите т тестове. Размерът на извадката е постоянен при всички тестове.

One-Sample T Test

Процедурата One-Sample T Test проверява дали средната на една променлива се различава от определена константа.

Примери. Изследователят може да иска да провери, дали средният коефициент на интелигентност за група студенти се различава от 100. Или производител на зърнени храни може да вземе произволни кутии от продукцията и да провери, дали средното тегло на извадките се различава от 1.3 паунда по 95% доверителен интервал.

Статистика. За всяка променлива на теста: средноаритметична, стандартно отклонение, стандартна грешка на средната. Средната разлика между всяка стойност на данните и хипотезата за стойността на теста, т тест, който проверява, дали тази разлика е нула и доверетелен интервал за тази разлика (можете да зададете нивото на доверие).

Данни. За да съпоставите стойностите на количествена променлива с хипотезата за стойността на теста, изберете количествена променлива и въведете предполаганата стойност на теста.

Assumptions. Предположения. Този тест предполага, че данните са разпределени нормално; обаче, този тест не приема отклонения от нормалността.

За да получите One-Sample T Test

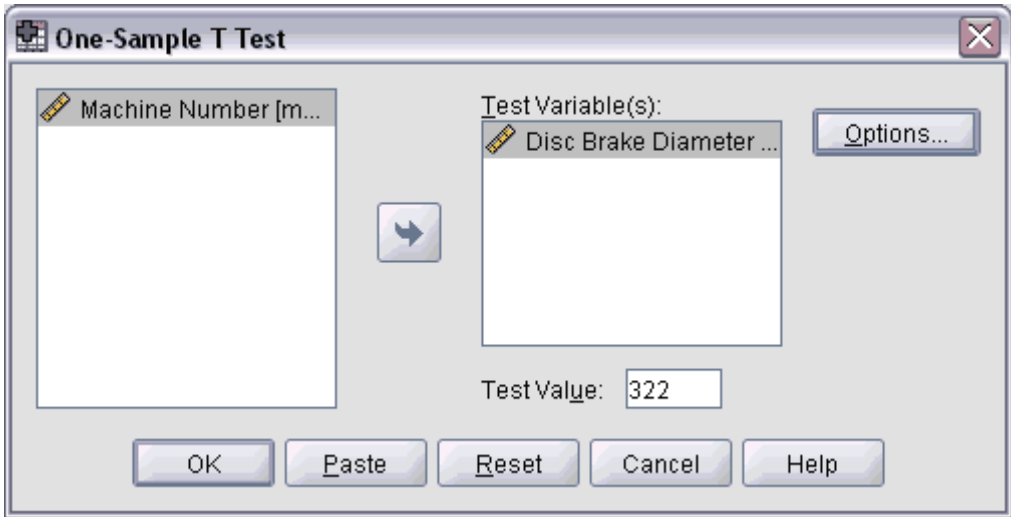
Изберете от менютата:

- Analyze
  - Compare Means
    - One-Sample T Test...

Изберете една или повече променливи, които да бъдат сравнени с предположената стойност.

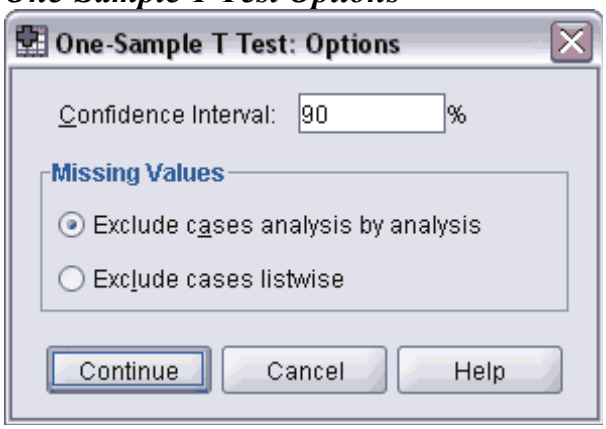
Въведете номинална стойност на теста, с която всяка средна на извадката ще бъде сравнена.

Можете да кликнете на Options, за да контролирате, как да третирате липсващите данни и нивото на доверетелния интервал.



Фигура 20-7 *Прозорец на One-Sample T Test*

One-Sample T Test Options



Фигура 20-8 *Прозорец на One-Sample T Test Options*

Confidence Interval. Доверителен интервал. По подразбиране се показва 95% доверителен интервал за разликата в средните. Въведете стойност между 1 и 99, за да поискате различно ниво на доверителност.
Missing Values. Липсващи данни. Когато тествате няколко променливи и липсват данни за една или повече променливи, можете да зададете на процедурата, кои случаи да изключи и кои да включи.
Exclude cases analysis by analysis. (Изключване на анализ на случаи от анализа) Всеки т тест използва всички случаи, които имат валидни данни за тестваните променливи. Размерите на извадката могат да са различни във всеки тест.
Exclude cases listwise. (Изключване на случаи по списък**)** Всеки т тест използва само случаи, които имат валидни данни за всички променливи, които се използват в поисканите т тестове. Размерът на извадката е постоянен при всички тестове.

T-TEST Command Additional Features

Командният синтактичен език също така Ви позволява да:
Правите т тестове от типовете one-sample и независими извадки, като използвате една и съща команда.
Сравнявате променлива с всяка друга променлива по списък с подвойкови т тестове (с подкомандата PAIRS).
Виж *Command Syntax Reference* (справка за синтакса на командите) за пълна информация за синтакса.

Глава 20. Едномерен еднофакторен дисперсионен анализ

Едномереният еднофакторен дисперсионен анализ представя еднофакторни анализи на количествени зависими променливи с един фактор (независима променлива. Дисперсионният анализ се използва за тестване хипотези, при които няколко средни са равни. Тази техника е едно разширение при множествените (двуизвадкови) t-тестове. В допълнение към определението - когато различията съществуват измежду средните, може да пожелаем да разберем коя средна е различна. Има два типа тестове за сравняване на средни: априори контрасти и post
hoc tests. Контрастите са тестове пуснати *преди* експеримента, а post
hoc tests са пуснати *след* провеждането на експеримента. Също можем да тестваме тенденциите чрез категориите.

Пример. Поничките поглъщат мазнина в различни количества когато те са сготвени. Един експеримент е зададен включващ три типа мазнина: фастъчено масло, царевично олио и мас. Фастъченото масло и царевичното олио са ненаситени мазнини, а маста е наситена мазнина. Може да се определи количеството мазнина, което е абсорбирана в зависимост от това каква мазнина е използвана, като се използва контраст за сравнение на разликите между наситените и ненаситените мазнини.

Статистики. За всяка група: номер на случаите, средна, стандартно отклонение, стандартна грешка на средната, минимум, максимум и 95% доверителен интервал за средната. Levene's test за хомогенността на дисперсията, графично изобразяване на дисперсионния анализ и тестове за силата на равенство на средните за всяка зависима променлива, user-specified a priori contrasts, post hoc range tests и множествено сравнение: Bonferroni, Sidak, Tukey's honestly significant difference, Hochberg's GT2,
Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch *F* test (R-E-G-W *F*), Ryan-Einot-Gabriel-Welsch range test (R-E-G-W *Q*), Tamhane's T2, Dunnett's T3, Games-Howell, Dunnett's *C*, Duncan's multiple range test, Student-Newman-Keuls (S-N-K), Tukey's *b*, Waller-Duncan, Scheffé, and least-significant difference.

Data. (Данни). Стойността на фастора (дискретна независима променлива) трябва да бъде integers, а зависимата променлива трябва да бъде количествена (ниво на измерване от интервала).

Предположение. Всяка група е от независима случайна извадка от популацията. Дисперсионния анализ е силно отдаличен от нормалността, въпреки че данните трябва да са симетрични. Групите трябва да идват от популации с равни дисперсии. За да тестваме това предположение ще използваме Levene's хомогенен дисперсионен тест.

За да получим едномерен еднофакторен дисперсионен анализ

От менюто избираме:

- Analyze
  Compare Means
    One-Way ANOVA...

избираме една или повече зависими променливи
избираме една независима променлива (фактор)

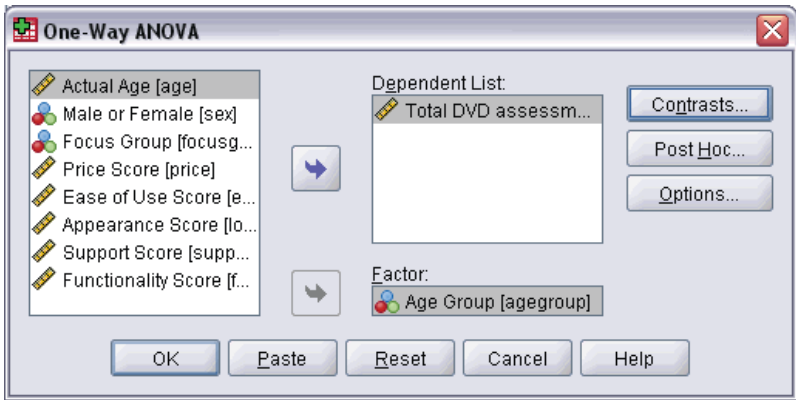


Figure 20-1 *One-Way ANOVA dialog box*

Еднофакторни ANOVA контрасти

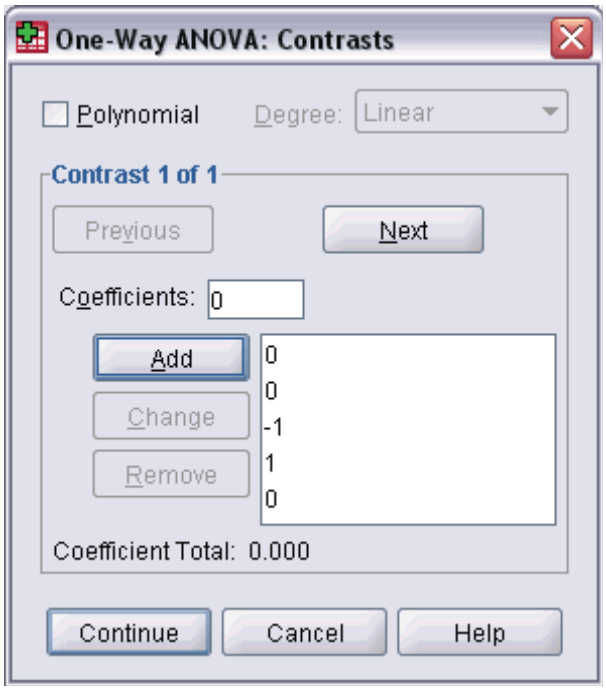


Figure 20-2 *One-Way ANOVA Contrasts dialog box*

Ние може да разделим групите – като сумираме квадратите на компонентите в тенденцията или да определим контрастите.

Полиноминал. Разделянието между групите сумирайки квадратите ни показва тенденционнине компоненти. Можем да тестваме тенденцията за зависима променлива чрез подредба на нивата на независимата променлива. Например, линейната тенденция (увеличава или намалява) при образуването на заплатата става чрез подредбата по нива на заплатата като се почне от най-високата спечелена.

- **Degree.** Може да изберем 1^{во}, 2^{ро}, 3^{то}, 4^{то} или 5^{то} ниво полиноминал.

Коефициенти. Потребителски специфичните контрасти мога да бъдат тествани чрез t статистиката. Напишете коефициент за всяка група (категория) за независима променлива и кликнете Add (добави) след всяко добавяне. Всяка нова стойност е добавена на върха на списъка с коефициенти. За да определите допълнителни набори контрасти, щракнете Next. Използвайте Next и Previous за да се премествате между различните набори контрасти.
Командата коефициенти е важна, защото тя съответства на нареждането при възкачване на стойностите при категориите на независимата променлива (фактора). Първия коефициент съответства на най-ниската групова стойност на независимата променлива, а последния коефициент съответства на най-високата стойност. Например, ако имаме 6 категории за независимата променлива, коефициенти – 1, 0, 0, 0, 0.5 и 0.5 контрасти в първата група с петата и шестата групи. За повечето заявки, коефициента трябва да се сумира до 0. Ако не се получи 0, ще се появи предопредително съобщение.

Еднофакторен ANOVA **Post Hoc Tests**

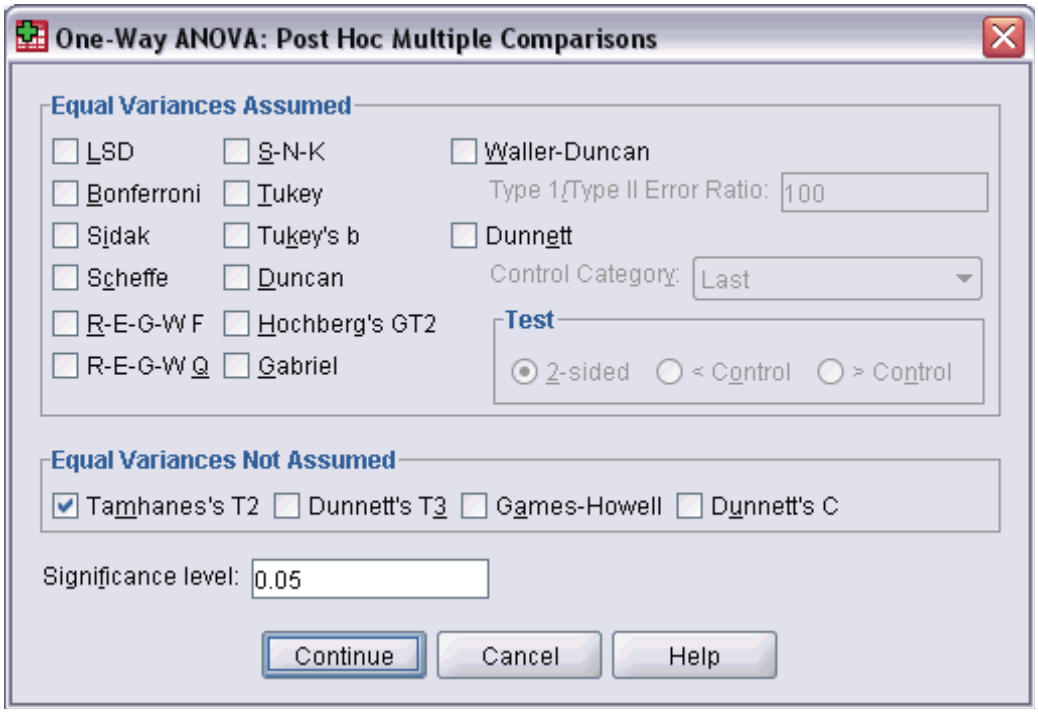


Figure 20-3 *One-way ANOVA Post Hoc Multiple Comparisons dialog box*

Веднъж вече определили, че съществуват разлики сред средните, post hoc range tests и pairwise multiple comparisons могат да определят как да ги различаваме. Хомогенният тест подрежда средните така че да се види че по своята същност те не са различни една от друга. Pairwise multiple comparisons test показва разликата между всяка двойка средни и дава матрица която индуцира различната група средни при ниво алфа 0.05.

Дисперсията приема равна стойност

Tukey's honestly significant difference test, Hochberg's GT2, Gabriel, и Scheffé са множествени сравнителни тестове и тестове с размах. Други достъпни тестове с размах са Tukey's *b*, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W *F* (Ryan-Einot-Gabriel-Welsch

F test), R-E-G-W *Q*(Ryan-Einot-Gabriel-Welsch range test), и Waller-Duncan. Достъпни множествени сравнителни тестове са Bonferroni, Tukey's honestly significant difference test, Sidak, Gabriel, Hochberg, Dunnett, Scheffé, and LSD (най-малка значителна разлика).

- **LSD.** Използва t тестове за да представи всички сравнения по двойки между групата средни. Нивото на грешката не се изменя независимо колко сравнения бихме направили.
- **Bonferroni.** Използва t тестове за да представи всички сравнения по двойки между групата средни, но контролира нивото на грешка, като има възможност да си го нагласиш за всеки тест по отделно.
- **Sidak.** Множествените сравнителни тестове са базирани на t статистиката (оценките). Сидак нагласява нивото на значимост за множествени сравнения и предоставя по стегнат вид от този на Bonferroni.
- **Scheffe.** Представя едновременни сравнения за всички възможни двойки комбинации от средните. Използва F извадково разпределение. Може да бъде използвано да провери всички възможни линейни комбинации на групата на средните, но просто ще ги сравнява по двойки.
- **R-E-G-W F.** Ryan-Einot-Gabriel-Welsch множествена процедура с долна стъпка базирана на F тест.
- **R-E-G-WQ.** Множествена процедура с долна стъпка базирана на the Studentized – овия размах.
- **S-N-K.** Прави всички сравнения между средните използвайки the Studentized размах. С еднакви извадкови размери, също така могат да се сравнят двойките средни използвайки стъпкова процедура (stepwise procedure). Средните са подредени от най-голямата до най-малката и се тества първо екстремните разлики.
- **Tukey.** Използва the Studentized размах за да сравни всички двойки по групи. Като поставя нивото грешката на експеримента за една колекция (сбор) от всички сравнения.
- **Tukey's b.** Използва the Studentized размах за да сравни всички двойки по групи. Критичната стойност е осреднена и символизира Tukey's honestly significant difference test and the Student-Newman-Keuls.
- **Duncan.** Прави сравнения използвайки стъпкова команда за индификационно сравнение често използван с the Student-Newman-Keuls test, но поставя защитно ниво за нивото на грешка при сбор от сравнения, което е по-добро отколкото нивото на грешка при отделните тестове. Използвайки the Studentized размахова оценка.
- **Hochberg's GT2.** Множественото сравнение и теста за размах използват модулите на the Studentized за максимум. Подобно на Tukey's honestly significant difference test.
- **Gabriel.** Тест за сравнение, който използва модула за максимум на Studentized и е предимно по-силен от Hochberg's GT 2 когато размерите на клетките са неравни. Тестът на Gabriel може да стане либерален, когато размерите на клетките се променят значително.
- **Waller-Duncan.** Множествения сравнителен тест се базира на t статистиката (оценката), използвайки a Bayesian подход.
- **Dunnett.** Множествения сравнителен t тест съчетава един набор от инструменти срещу единичната контролна средна. *2-странните* тестове тестват средната на всяко ниво (включително и контролната категория) на фактора който е различен от контролната група. < *Control* тества ако средната на което и да е ниво на фоктора е по-малка от контролната група. > *Control* тества ако средната на което и да е ниво на фоктора е по-голяма от контролната група.

Неравни дисперсии
Множествени сравнителни тестове който не сумират равни дисперсии са Tamhane's T2, Dunnett's T3,Games-Howell, and Dunnett's C.

- **Tamhane's T2.** Консервативен сравнителен тест базиран на t теста. Този тест е подходящ, когато дисперсиите са неравни.
- **Dunnett's T3.** Сравнителен базиран на максималните модули на the Studentized. Този тест е подходящ когато дисперсиите са различни.
- **Games-Howell.** Сравнителен тест, който понякога е либерален. Този тест е подходящ когато дисперсиите са различни.
- **Dunnett's C.** Сравнителен тест, който е базиран на размаха на the Studentized. Този тест е подходящ когато дисперсиите са различни.

Бележка: Може да ви се стори много лесно да интерпретирате output-а от post hoc тестовете ако не сте селектирали (избрали) скритите празни редове и колони в the Table Properties dialog box (в активираната таблица изберете Table Properties от менюто).

Опции на едномерен еднофакторен дисперсионен анализ

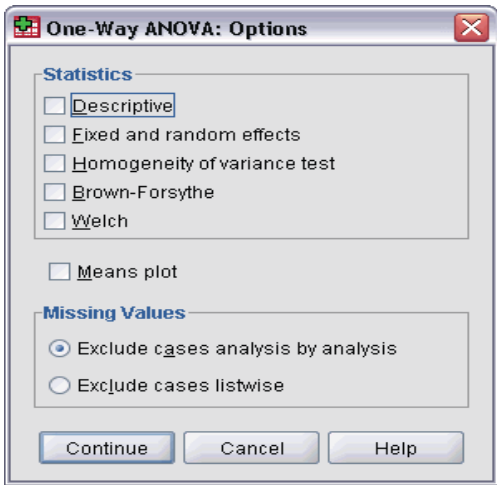


Figure 20-4 *One-Way ANOVA Options dialog box*

Statistics(Статистики). Изберете едно или повече от следващите:

- **Descriptive.** Пресмята номер от случаите, средната, стандартното отклонение, стандартната грешка на средната, минимум, максимум, и 95% доверителен интервал на всяка зависима променлива за всяка група.
- **Fixed and random effects.** Показва стандартното отклонение, стандартната грешка и 95% доверителен интервал за фиксиран модел, и стандартната грешка, 95% доверителен интервал и оценява дисперсията на компонентите по рандомизиран модел.
- **Homogeneity of variance test.** Пресмята статистиката на the Levene тествайки равенството в групата на дисперсиите. Този тест е независим от предположенията за нормалност.
- **Brown-Forsythe.** Пресмята тествайки равенството в групата на средните. Тази статистика (оценка) е препочитана за *F* статистиката когато преположението че дисперсиите са равни не се приема.

◷ **Welch.** Пресмята статистика на the Welch тествайки равенството на групата от средни. Тази статистика (оценка) е препочитана за *F* статистиката когато преположението че дисперсиите са равни не се приема.

Means plot. Показва чарт който изобразява субгрупата на средните(средните за всяка група са дифинирани със стойности на фактора променлива).

Missing Values. Контролира отношението на пропуснатите стойности.

◷ **Exclude cases analysis by analysis.** Случай с липсваща стойност низависимо дали се отнася за зависима или фактор променлива се изключва от анализа. Също и случай който е извън размаха, определен за фактора промелнива, не се използва.

◷ **Exclude cases listwise.** Случаи с липсващи стойности за фактор променлива или всяка зависима променлива се включват в лист на зависимите в главния диалогов бокс (кутия) и се изключват от всички анализи. Ако вие не се определили множествени зависими променливи, това няма да има ефект.

ONEWAY Command Additional Features

Командата syntax language също позволява да:

◷ Придобити фиксирани – и рандомизирани статистики (оценки). Стандартното отклонение, стандартна грешка на средната и 95% доверителен интервал за фискиран модел на ефекта. Стандартната грешка, 95% доверителен интервал и оценка на дисперсията на между компонентите (междинните компоненти) на за рандомизиран модел на ефекта (използвайки `STATISTICS=EFFECTS`).

◷ определи alpha нивата за най-малката значителна разлика, Bonferroni, Duncan, and Scheffé множествени сравнителни тестове (с `RANGES` субкоманда).

◷ Запише матрицата на средните, стандартното отклонение и честотите или да прочете матрицата на средните, честотите, обединените дисперсии и нивата на свобода за обединените дисперсии. Тези матрици могат да бъдат използвани на мястото на реда с данни където да придобият стойност с едномерения еднофакторен дисперсионен анализ (с `MATRIX` подкоманда).

Виж *Command Syntax Reference* за допълнителна информация.

Глава 21. Обобщен едномерен линеен модел

Обобщеният едномерен линеен модел предоставя регресивен анализ и анализ на дисперсията за една зависима промелива или няколко фактора (независими) промеливи. Фактора променлива дели популацията на групи. Използвайки този обобщен едномерен линеен модел можете да тествате нулевата хипотеза за ефекта на другите променливи върху средните на групирани единични зависими променливи. Може да проучите взаимодействията между факторите и ефектите на индивидуалните фактори, някои от които могат да са случайни. В допълнение ефектът на ковариантите и взаимодействието на ковариантите с факторите могат да бъдат включени. За регресивния анализ, независимите (предшестващи) праменливи са определени като коварианти.

И двата, балансирани и небалансирани, модела могат да бъдат тествани. В допълнение, за да тестваме хипотезите, обобщеният едномерен линеен модел създава оценки на параметрите.

Обикновено използваните априори контрасти са налични за да представят тестването на хипотезите. Допълнително, след като общия *F* тест е показал важността, че можем да използваме post hoc тестове за да оценим разликите между специфичните средни. Оценените маргинални средни дават оценки за предсказани средноаритметични стойности за клетките от модела, и профил плотс(взаимодействието в плотс) за тези средни ни позволяват лесно да визуализирате някои от връзките (отношенията). Останалите части, предсказани стойностите, разстоянието Cook's и стойностите за влияние могат да бъдат запазени като нови променливи в нашият файл с данни за последващи предположения. WLS тежестта ни позволява да определим променливата, която обикновено ни дава възможността за наблюдение върху различните тежести на анализа на квадратите с най-малка тежест(WLS), може би за да компенсира различната прецизност на измерване.

Пример. Данните са събрани за бегачи от Чикагския маратон за няколко години. Времето през което всеки бегач финишира е зависима променлива. Другите фактори включват времето (студено, приятно или горещо), броя на месеците тренировка, броя на предишни маратони и пол. Годините са обозначени като ковариант. Може би ще откриете, че полът има значителен ефект и че взаимодействието между пол и времето също е важно.

Методи. Тип I, Тип II, Тип III и Тип IV сумите от квадратите могат да бъдат използвани за оценка на различни хипотези. Тип III е по подразбиране.

Статистики. Post hoc размах тестът и множествените сравнения: най-малко значима разлика, Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch множествено *F*, Ryan-Einot-Gabriel-Welsch множествен размах, Student-Newman-Keuls, Tukey's honestly significant difference, Tukey's *b*, Duncan, Hochberg's GT2, Gabriel, Waller-Duncan *t* test, Dunnett (едностранни и двустранни), Tamhane'sT2, Dunnett's T3, Games-Howell, и Dunnett's *C*. Описателна статистика: наблюдавани средни, стандартни отклонения, и броя на всички зависимите променливи във всички клетки. Тестът на Levene за хомогенност на дисперсията.

Plots. Ниво сравнително разпространение, остатък и взаимодействие с профила.

Данни. Зависимата променлива е количествена. Факторите са категорични. Те могат да имат цифрови стойности или стрингови стойности до осем символа. Ковариантите са количествени променливи, които са свързани с зависимата променлива.

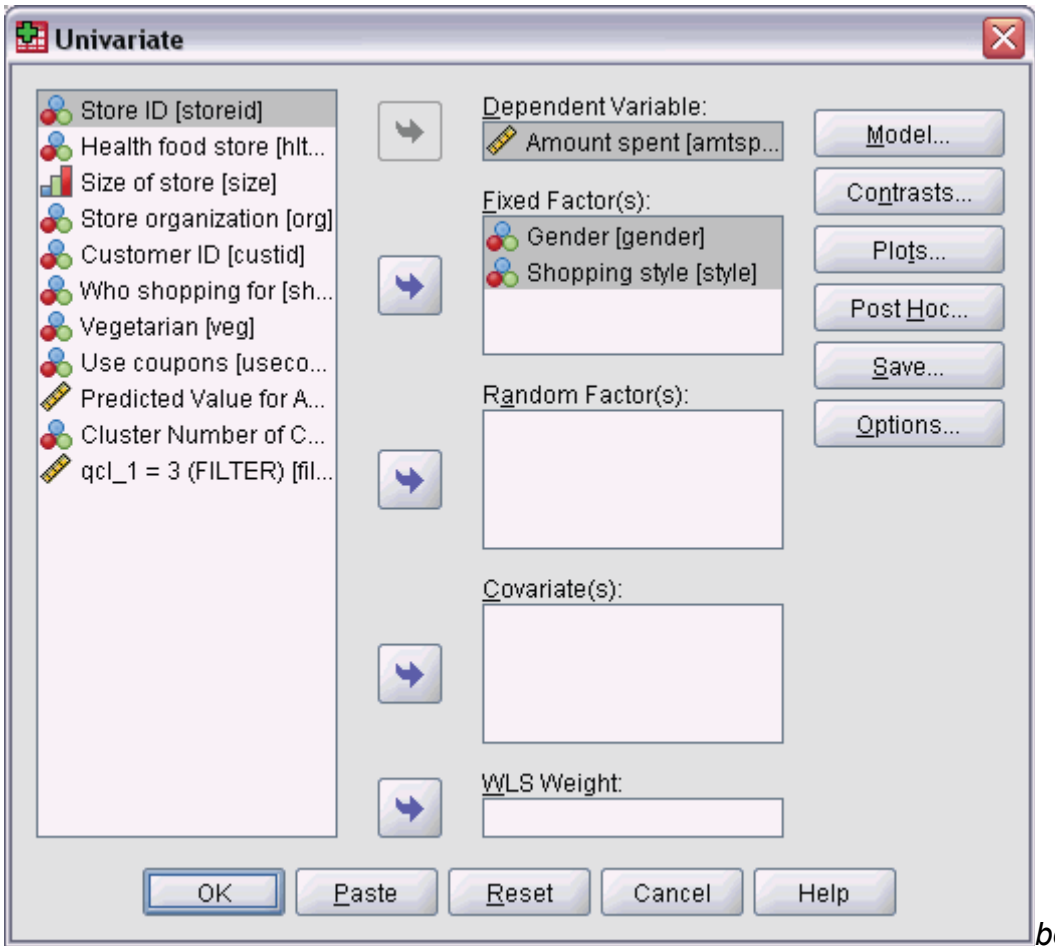
Предположения. Данните са от случайна извадка от нормална популация; а в популацията, всички отделни дисперсии са едни и същи. Дисперсионият анализ е устойчив към отклонения от нормалността, въпреки че данните трябва да бъдат симетрични. За да проверите предположенията, можете да използвате тестовете за хомогенност на

дисперсиите, тестове и фигурите за ниво на разпространие. Можем също така да разгледаме остатъците и остатъчните плотс.

За да се получи обобщен едномерен линеен модел в таблици

От менюто избираме:

- *Analyze*
  General Linear Model
  Univariate...



box

Figure 21-1 *GLM Univariate dialog*

Избираме зависима променлива.

Избираме променливи за Фиксиран(и) Фактор(и), Случайни Фактор(и) и Ковариант(и), като подходящи за нашите данни.

Като опция можем също така да използваме WLS Weight за да се уточни теглото на променливата чрез анализа теглото на най-малките квадрати. Ако стойността на теглото на променливата е нула, отрицателно или липсва, случаят е изключен от анализа. А вече променлива използвана в модела не може да се използва като тежест на променлива.

GLM Model

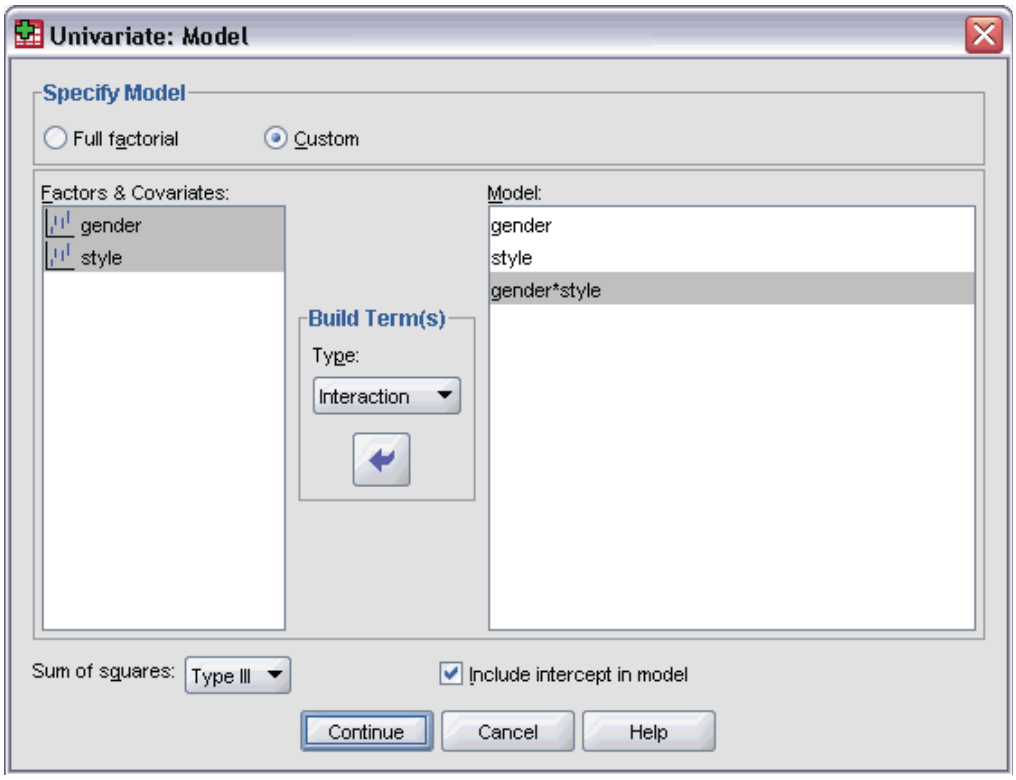


Figure 21-2 *Univariate Model dialog box*

Specify Model. Пълен факторен модел съдържа всички основни факторни ефекти, всички основни ковариантни ефекти и всички фактор- по-фактор взаимодействия. Той не съдържа ковариантните взаимодействия. Изберете опцията Custom за да посочете само един набор от взаимодействия или да посочате фактор-по-коварианти взаимодействия. Трябва да покажете всички времена за да бъдат включени в модела.

Factors and Covariates. Факторите и ковариантите са в списъка.

Model. Моделът зависи от естеството на нашите данни. След като изберем опцията Custom можем да изберете Основните Ефекти и взаимодействия, които са от интерес за нашия анализ.

Sum of squares. Метод за изчисляване на сумите на квадратите. За балансирани или небалансирани модели с не липсващи клетки, Тип III сбора на квадратите е методът, който е най-често използван.

Include intercept in model. Прекъсването обикновено е включено в модела. Ако може да се предположи, че данните преминават през началото, може да се изключи прекъсването.

Build Terms

За избраните фактори и коварианти:

Interaction. Създава най-високото ниво на взаимодействието на всички избрани променливи. Това е по подразбиране.

Main effects. Създава основните ефекти за всяка отделна избрана променлива.

All 2-way. Създава всички възможни два начина на взаимодействие на избраните променливи.

All 3-way. Създава всички възможни три начина взаимодействия на избраните променливи.
All 4-way. Създава всички възможни четири начина взаимодействия на избраните променливи.
All 5-way. Създава всички възможни пет начина взаимодействия на избраните променливи.

***Sum of Squares** (Сумата на квадратите)*
За този модел, можем да изберем тип на сумирането на квадратите. Тип III е най-често използван и е по подразбиране.
Type I. Този метод е известен като метод на йерархично разлагане на сумата-на-квадратите. Всеки период се коригира само от период, който го предхожда в този модел. Тип I сумира квадратите, които обикновено се използват за:
- За балансиран модел на дисперсионния анализ, в който всички основни ефекти са определени преди всяко взаимодействие на ефектите от първи ред, всяко взаимодействие на ефектите от първи ред са определени преди всяко взаимодействие на ефектите от втори ред и така нататък.
- Модел на полиноминална регресия, в който всяко по-ниско ниво на команда е определено преди всяко по-високо ниво на команда.
- Чисто вложения модел, в който на първо определения ефект е вложен в рамките на втория определен ефект, вторият определен ефект е вложен в рамките на третия, и така нататък. (Тази форма на гнездене може да бъде определен само чрез използване на синтаксиса (syntax).)

Type II. Този метод изчислява сумите на квадратите на ефект в модела коригиран за всички други "подходящи" ефекти. Подходящият ефект е този, който отговаря на всички ефекти и които не съдържа разглеждания ефект. Метод тип II често се използва за:
- За балансиран модел на дисперсионния анализ.
- Всеки модел, в който притежава само основните ефекти.
- Всеки регресионен модел.
- Чист гнездящ дизайн. (Тази форама на гнездене може да се определи с използването на синтаксис)

Type III. По подразбиране. Този метод изчислява сумите на квадратите на един ефект в дизайна като сумите на квадратите са приспособени за всякакви други ефекти, които не ги съдържат и са перпендикулярни на всички ефекти (ако има такива), които ги съдържа. Тип III сума на квадратите има едно основно предимство в това, че те са инвариантни по отношение на честотата в клетката, толкова дълго колкото общата форма на оценката остава константа. Ето защо, този тип на сумиране от квадратите често се счита за полезен за един небалансиран модел с не липсващи клетки. Във факторения модел с не липсват клетки, този метод е еквивалентен на техниката на the Yates' за средно-претеглени квадрати на средните. Тип III-сбора на- квадратите, метод който е често се използван за:
- Всички модели, изброени в тип I и тип II.
- Всеки балансиран или небалансиран модел, без празни клетки.
- Всеки балансиран модел или небалансиран модел с празни клетки.

Type IV. Този метод е предназначен за ситуация, в която има липсващи клетки. За всеки ефект F в дизайна, ако F не се съдържа в друг ефект, тогава Тип IV = тип III = тип II. Когато F се съдържа в други ефекти, Тип IV разпределя контрасти направени сред

параметрите в F за всички справедливо по-високо ниво ефекти. Тип IV метод сума -на-квадратите често се използва за:

- Всички модели, изброени в тип I и тип II.
- Всеки балансиран модел или небалансиран модел с празни клетки.

GLM Contrasts

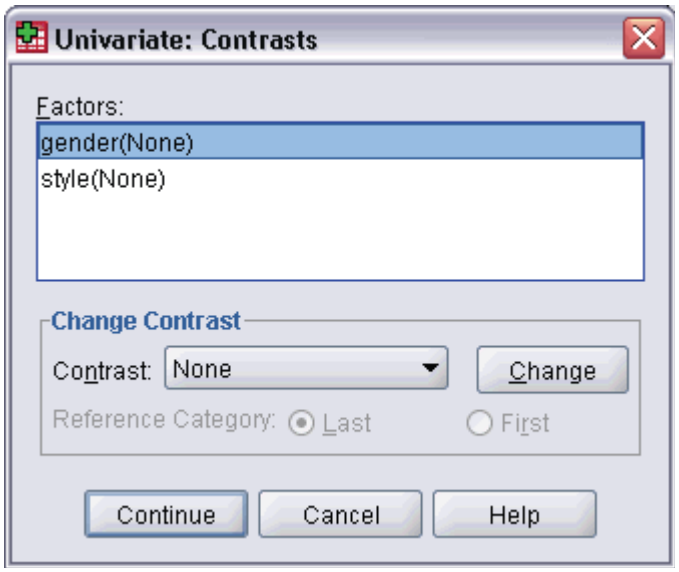


Figure 21-3 *Univariate Contrasts dialog box*

Контрастите се използват за изпитване на различията между нивата на един фактор. Можем да зададем контраст за всеки фактор в модела (в повтарящ се модел с мерки, за всеки от междусубектен фактор).

Контрасти представляват линейни комбинации на параметрите. Хипотезата за изпитване се основава на нулевата хипотеза **LB** = 0, където **L** е контрастът коефициенти на матрицата и **B** е параметърен вектор. Когато контрастът е уточнен, **L** матрицата е създадена. В колоните на матрицата **L**, съответстваща на фактор съответстващ на контраста. Останалите колони се регулират така, че матрицата **L** е измерена. Аутпута (Решението) включва F статистика за всеки набор от контрасти. Също така показва разликите на контраста, които са Bonferroni тип едновременно доверителен интервал на базата на Student's t разпределение.

Available Contrasts

Налични контрасти са отклонение, извадка, разликата, Helmert, повтарящи се и полиноминална. За отклонение контрастите и извадковите контрасти, можем да изберем дали референтната категория е последната или първата категория.

Contrast Types

Deviation. Сравнява средната на всяко ниво (с изключение при референтната категория) спрямо средната на всички нива (голямата средна). Нивата на фактора могат да бъдат в произволен ред.

Simple. Сравнява средната на всяко ниво на средната на определено ниво. Този тип на контраст е полезен, когато е налице в контролна група. Можем да изберем първата или последната категория, като препратка.

Difference. Сравнява средната на всяко ниво (с изключение на първото) до средната на предишните нива. (Понякога се наричан обратни Helmert контрасти.)

Helmert. Сравнява средната на всяко ниво на фактора (с изключение на последния) до средната от последващите нива.
Repeated. Сравнява средната на всяко ниво (с изключение на последния) до средната от следващите ниво.
Polynomial. Сравнява линейния ефект, квадратичния ефект, кубичния ефект, и така нататък. Първа степен на свобода съдържа линейния ефект във всички категории; а втора степен на свобода, квадратичения ефект, и така нататък. Тези контрасти често се използват за оценка на полиноминалните тенденции.

GLM Profile Plots

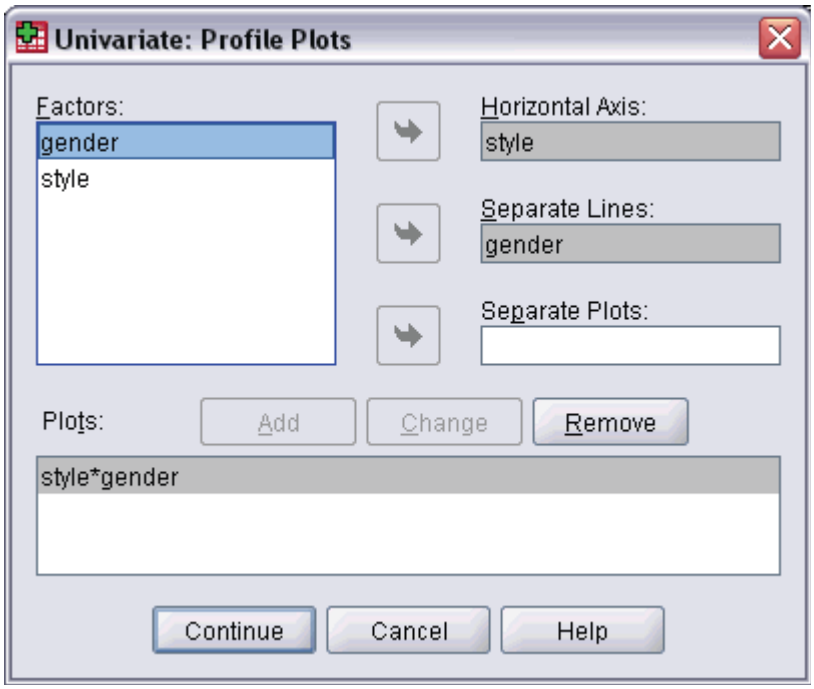


Figure 21-4 *Univariate Profile Plots dialog box*

Профил плотс (взаимодействие на плотс) са полезни за сравняване на маргинални средни в своя модел. Един профил плот е линейния плот, в който всяка точка показва оценената маргинална средна на зависимата променлива(коригирани за всички променливи) на едно ниво на фактора. Нивата на втория фактор могат да бъдат използвани за създаването на отделни редове. Всяко ниво в третия фактор може да се използва за създаване на отделен плот. Всички фиксирани и случайни фактори, ако има такива, са на разположение на плота. За многомерни анализи, профил плотове са създадени за всяка зависима променлива. При повторни анализ на мерките, както и двата междусубектни фактори и по индивидулни фактори могат да бъдат използвани в профил плотс. GLM Многовариантен и GLM Повтарящи се Мерки са налични само ако имате инсталирана the Advanced Models option. Профилът плот на един фактор показва дали оценените маргинални средни се увеличават или намалява през нивата. За два или повече фактори, паралелните линии сочат, че не съществува взаимодействие между факторите, което означава, че можете да се проучи нивото на само един от факторите. Непаралелните линии показват взаимодействие.

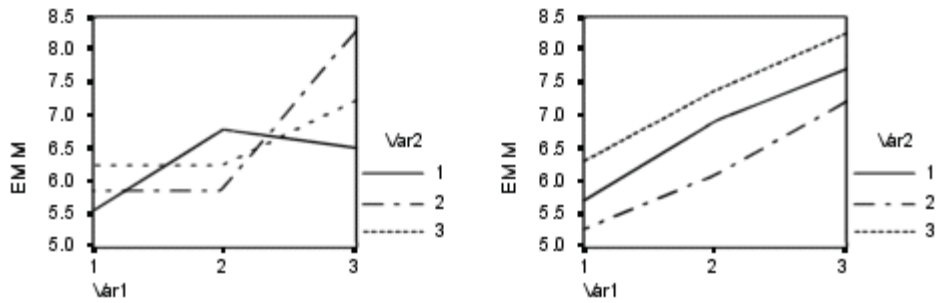


Figure 21-5 ***Nonparallel plot (left) and parallel plot (right)***

След като плота е определен като изберем факторите за хоризонталната ос и, евентуално, факторите за отделните линии и отделните плотс, плотът трябва да се добави към the Plots list.

GLM Post Hoc Comparisons

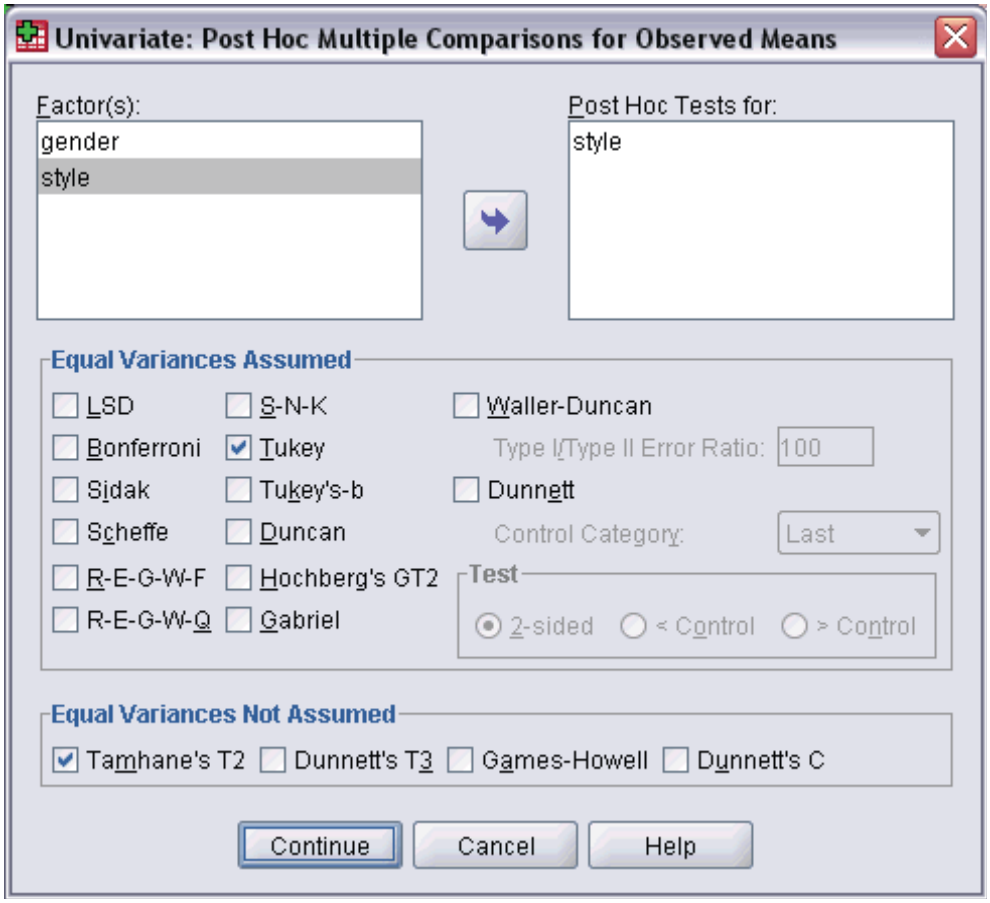


Figure 21-6 *Post Hoc dialog box*

Post hoc multiple comparison tests. След като сме установили, че съществуват различия между средните, post hoc range теста и множествените сравнения на няколко двойки може да определи кои средни се различават. Сравненията са направени на некоригирани стойности. Тези тестове се използват само за фиксирани междусубектни фактори. В GLM Повтарящи се Мерки, тези тестове не са налични, ако има не междусубектни фактори, и the post hoc множествените сревненителни тестове се

извършват средно за нивата на по индивидуалните фактори. За GLM Многовариантен, the post hoc тестовете се извършват за всяка зависима променлива поотделно. GLM Многовариантен и GLM Повтарящи се Мерки са налични само ако имате инсталирана the Advanced Models option. Тестовете на The Bonferroni и Tukey's за честна значителна разлика са обикновено използваникато множествен сравнителен тест. **Тестът на Bonferroni**, е базиран на Student's *t* статистика, настройва наблюдаваното ниво на значимост за факта, че множествените сравнения са направени. **Тестът *t* на Sidak** също настройва на ниво на значимост и предвижда по-строги граници, отколкото границите на теста Bonferroni. **Тестът на Tukey's** за честно значима разлика използва Studentized статистически размах, за да направи всички двойки сравнения между групите и да определи нивото на експериментална грешка до нивото на грешка за събираната информация за всички двойки сравнения. При тестване на голям брой двойки на средната, тестът на Tukey's за честно значима разлика е по-силен от тестът на Bonferroni. При малък брой на двойки, Bonferroni е по-силен.

Hochberg's GT2 е подобен на тестът на Tukey's за честно значима разлика, но се използва модула за максимум на Studentized. Обикновено, тестът на Tukey е по-силен. **Тестът на Gabriel's за двойни сравнения** също използва модула за максимум на Studentized и като цяло е по-силен от GT2 Hochberg, когато размерите на клетката са неравни. Тестът на Gabriel's може да се превърне в либерален, когато размеритеклетката се различават значително.

***T* Тестът на Dunnett's за множествено сравнение** сравнява набор от обработки срещу една контролна средна. Последната категория е категория по подразбиране за контрол. Освен това, можем да изберем първата категория. Можем също така да изберем двустранен или едностранен тест. За да проверим дали средната на всяко ниво (с изключение на контролната категория) на фактора не е равна на тази на контролната категория, използвайка един двустранен тест. За да тествате дали средната на всяко равнище на фактора е по-малка от тази на контролната категория, избереме < Control. По същия начин, за да тестваме дали средната на всяко ниво на фактора е по-голяма от тази на контролната категория, избереме > Control. Ryan, Einot, Gabriel, and Welsch (R-E-G-W),разработват два множествено стъпкови долни размахови теста. Множествено стъпковата процедура първо тества дали всички средни са равни. Ако всички средни не са равни, подгрупи на средната се тестват за равенство. **R-E-G-W F** се основава на *F* тест и **R-E-G-W Q** е въз основа на Studentized размах. Тези тестове са по-силни, отколкото теста на Duncan's за множествен размах и Student-Newman-Keuls (които са също многостъпкови процедури), но те не се препоръчват за неравноразмерни клетки. Когато дисперсиите са неравни, използвайте **Tamhane's T2** T2 Tamhane's (консервативен тест базиран на а *t* тест), **Dunnett's T3** (тест за сравнение базиран на Studentized модул замаксимум), **Games-Howell тест за сравнение** (понякога либерален), или Dunnett's C (тест за сравнение на базата на Studentized размах). **Duncan's multiple range test**, Student-Newman-Keuls (**S-N-K**), and **Tukey's b** са тестове за размах, които подреждат средните и изчисляват стойността на размаха. Тези тестове не са използвани по-често отколкото тестовете, които предварително обсъдихме. The **Waller-Duncan t test** използва подхода на Bayesian. Този размахов тест използва хармоничната средна на обема на извадката, когато извадките са неравни. Теста на the **Scheffé** за нивото на значимост има за цел да даде възможност на всички възможни линейни комбинации на групата на средните да бъдат тествани, а не само по двойки сравнения са на разположение в тази функция. Резултат на теста на Scheffé е често по-консервативен от други тестове, което означава, че по-голямата разлика между средната се изисква за значимост.

Най-значимата разлика (**LSD**)за множествения тест за сравнение е еквивалентна на няколко отделни *t* теста между всички двойки в групите. Недостатък на този тест е, че не е направен опит да се адаптират към наблюдаваното ниво на значимост за множество сравнения.

Тестовете показват. Двойките сравнения са предвидени за LSD, Sidak, Bonferroni, Games-Howell, Tamhane's T2 и T3, Dunnett's *C*, и Dunnett's T3. Хомогенните подмножества за тестовете за размах са условие за S-N-K, Tukey's *b*, Duncan, R-E-G-W*F*, R-E-G-W *Q*, и Waller. Тестът на Tukey's за честно значима разлика, Hochberg's GT2, Gabriel's test, и Scheffé's test са едновременно няколко сравнителни теста и тестове за размах.

GLM Save

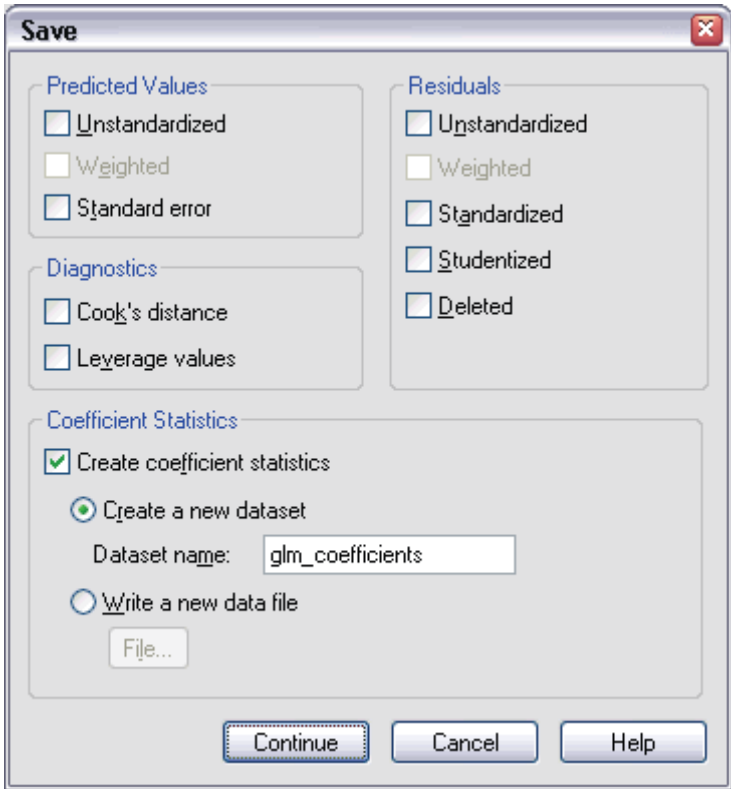


Figure 21-7 *Save dialog box*

Можете да запишете стойностите предсказани от модела, остатъчните премервания, както и мерки, свързани с нови променливи в the Data Editor. Много от тези променливи могат да бъдат използвани за разглеждане на предположения за данните. За да запишете стойностите за употреба при друга сесия на SPSS, трябва да запишете текущия файл с данни.

Предсказани стойности. Стойностите се предсказват от модела за всеки случай.

- **Unstandardized.** Стойността на модела прогнозира зависимата променлива.
- **Weighted.** Претеглена нестандартизирана прогноза. Само ако WLS променливата е предварително избрана.
- **Standard error.** Приблизителната оценка на стандартното отклонение от средната стойност на зависимите променливи за случаи, които имат едни и същи стойности на независимите променливи.

Diagnostics. Мерки за идентифициране на случаи с необичайни комбинации от стойности за независими променливи и случаи, които могат да окажат голямо влияние върху модела.

- **Cook's distance.** Една мярка за това колко от остатъците от всички случаи ще се променят, ако определен случай е бил изключен от изчисляването на коефициентите на регресия. Голям Cook's D показва, че с изключването на един случай от изчисляването може да промени статистиката на коефициента на регресия значително.

- **Leverage values.** Не центриран лостов механизъм приемащ определени стойности. Относителното влияние на всяко наблюдение на модела е удобно.

Residuals. Един нестандартизиран остатък е действителната стойност на зависимата променлива минус стойността предсказана от модела. Стандартизиран, Studentized, и унищожаващ остатъците са също на разположение. Ако дадена променлива WLS бъде избрана, то тя ще бъде претеглена, ще нестандартизира остатъците.

- **Unstandardized.** Разликата между наблюдаваната стойност и стойността предсказана от модела.
- **Weighted.** Претегля нестандартизираните остатъци. Налична само ако WLS променлива бъде предварително избрана.
- **Standardized.** Остатъка е разделен от оценката на неговото стандартно отклонение. Стандартизираните остатъци, които са известни като остатъците на Пиърсън, имат средна равна на 0 и стандартно отклонение равно на 1.
- **Studentized.** Остатъка е разделен от оценката на неговото стандартно отклонение, което варира от случай до случай, в зависимост от разстоянието на стойностите за всеки отделен случай на независимите променливи от средните на независимите променливи.
- **Deleted.** Остатък за един случай, когато този случай се изключва от изчисляването на регресивните коефициенти. Това е разликата между стойността на зависимата променлива и предвижданата коригирана стойност.

Coefficient Statistics. Записва дисперсионна – обратнодисперсионна матрица от параметърните оценки в модела до нов набор от данни в текущата сесия или външен SPSS-формат на файла с данни. Също така, за всяка зависима променлива, ще има ред на оценките на параметъра, един ред за стойностите на значимост за *t* статистиката, съответстваща на параметърните оценки, и един ред на остатъчните степени на свобода. За многовариантния модел, има подобни редове за всяка зависима променлива. Можете да се използва тази матричен файл и в други процедури, които четат матрични файлове.

GLM Options

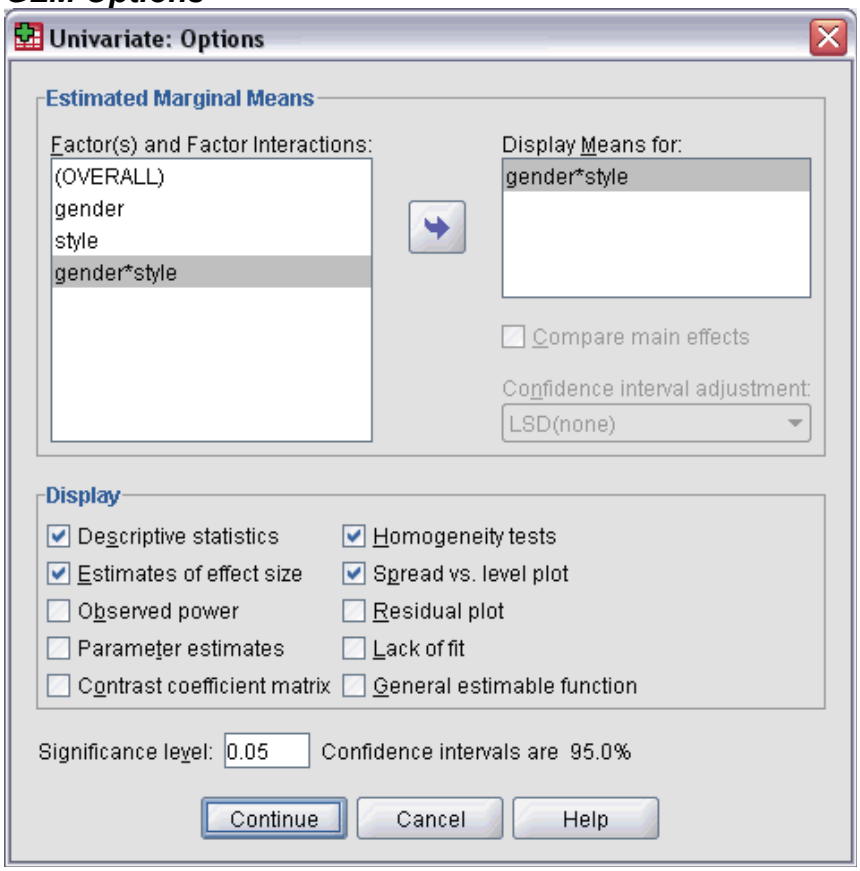


Figure 21-8 *Options dialog box*

Избираемите статистики са налични от този диалогов прозорец. Статистиките са изчислени, като се използва фиксирани-ефекти модел.

Estimated Marginal Means. Изберете факторите и взаимодействията, за които искате оценки от популационните маргинални средни в клетките. Тези средни са коригирани за ковариантите, ако има такива.

Compare main effects. Осигурява некоригирани по двойки сравнения между оценяваните маргинални средни пределното за всеки основен ефект в *the model*, както и в рамките на между- и вътресубектните фактори. Тази позиция е налична само ако основните ефекти са избрани в рамките на the *Display Means For list*.

Confidence interval adjustment. Изберете най-малко значимата разлика (LSD), Bonferroni, или Sidak корекция на доверителните интервали и значимост. Тази позиция е достъпна само ако *Compare main effects* е избрана.

Display. Изберете *Descriptive statistics* за производството на наблюдаваните средни, стандартни отклонения, и броя на всички зависимите променливи във всички клетки. *Оценките на ефекта размер* дава частичен ETA-квадратна стойност за всеки ефект и всяка параметърна оценка. Ета-квадратна статистика описва част от общите колебания дължащи се на фактора. Изберете *Observed power* за получаване на сила на теста, когато алтернативните хипотези се определят въз основа на наблюдаваните стойности. Изберете *Parameter estimates* за да произведете параметърните оценки, стандартните грешки, *t* тестовете, доверителните интервали, и наблюдаваната сила за всеки тест. Изберете *Contrast coefficient matrix*, за да получите **L** матрица.

Тестовете за хомогенност произвеждат теста на Levene за хомогенност на дисперсията за всяка зависима променлива във всички нива на комбинации в междусубектните фактори,

за междусубектните фактори само. Разпространението-срещу-нивота и възможностите за остатъчни плотс са полезни за проверка на предположенията за данни. Тази позиция е забранена, ако няма фактори. Изберете *Residual plot* за да се получи наблюдавани-по-прогнозирани-по-стандартизирани остатъчен плот за всяка зависима променлива. Тези плотс са полезни за разследване на поемането на равна дисперсия.

Изберете Lack of fit за да проверите дали връзката между зависимата променлива и независимите променливи могат да бъдат адекватно описани от модела. *General estimable function* ни позволява да си изградим случайни тестове за хипотеза на базата на обща оценъчна функция. Редове във всяка контрастна коефициентна матрица са линейни комбинации от общата оценъчна функция.

Significance level. Може би ще пожелаете да зададете ниво на значимост, използвайки post hoc tests и нивото на сигурност, използвани често за построяването на доверителни интервали. Посочената стойност се използва също и за изчисляване на наблюдаваната сила на теста. Когато зададете ниво на значимост, свързаното с тях ниво на доверителни интервали се показва в диалоговия прозорец.

UNIANOVA Command Additional Features

Езикът команда синтаксис също позволява да:

- Посочите ефектите за гнездене вложени в дизайна (с помощта на *подкомандата DESIGN*).
- Посочите тестовете на сравнителните ефекти с линейна комбинация от ефекти или една стойност (с помощта на подкомандата *TEST*).
- Посочите няколко контраста (с помощта на подкомандата *CONTRAST*).
- Включите потребителските липсващи стойности (с помощта на подкомандата *MISSING*).
- Посочите EPS критериите(с помощта на подкомандата *CRITERIA*).
- Конструирате custom **L** matrix, **M** matrix, или **K** matrix (с помощта на подкомандите *LMATRIX, MMATRIX* и *KMATRIX*).
- За отклонение или опростени контрасти, посочите междинна категория (с помощта на подкомандата *CONTRAST*).
- Посочите показателите за полиноминалните контрасти (с помощта на подкомандата *CONTRAST*).
- Посочите условната грешка за post hoc сравнения (с помощта на подкомандата *POSTHOC*).
- Изчислите оценките за маргиналните средни за всеки фактор или факторно взаимодействието между факторите в списъка с фактори(с помощта на подкомандата *EMMEANS*).
- Посочите имената на временние променливи (с помощта на подкомандата *SAVE*).
- Конструирате корелацията в матричния файл с данни (с помощта на подкомандата *OUTFILE*).
- Да построим матричен файл с данни, който да съдържа статистики от междусубектна ANOVA таблица (с помощта на подкомандата *OUTFILE*).
- Запазим матрицата за дизайна като нов файл с данни (като се използва подкомандата OUTFILE).

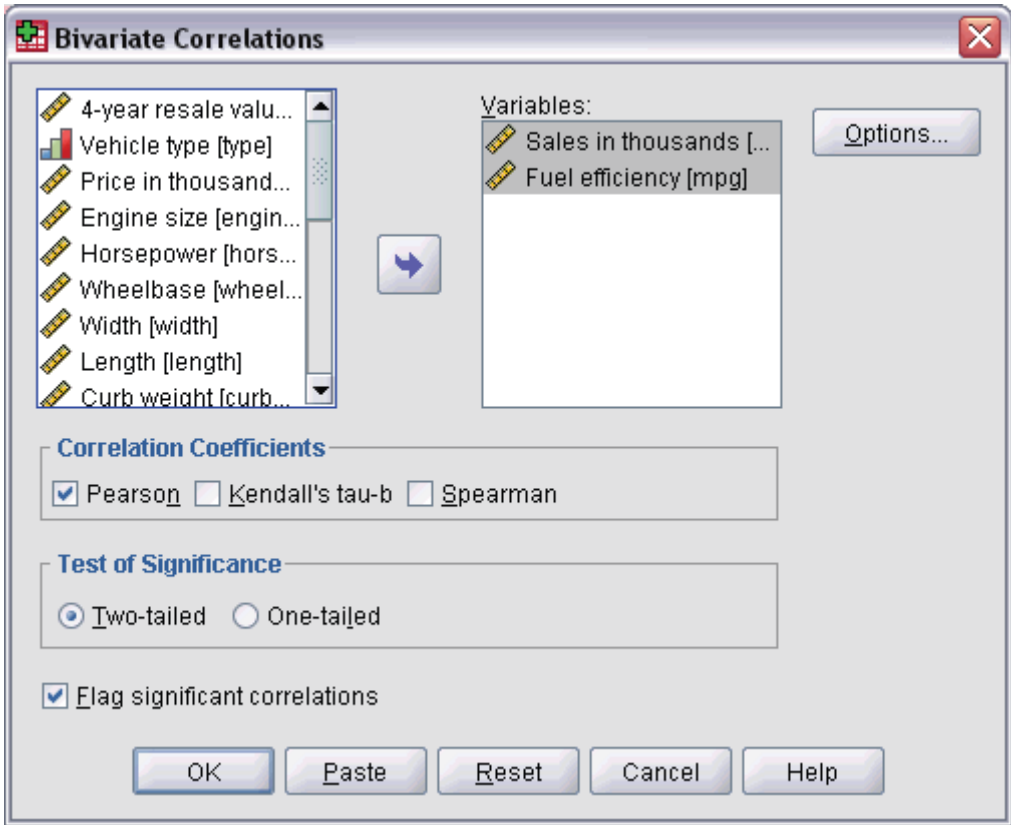
Вижте *Command Syntax Reference* за пълна информация за синтаксиса.

Глава 22. Корелация

Процедурата на корелация изчислява коефициентът на Пирсъновата корелация, Спиърмъновата Rho и Кендъловата rho-b в техните нива на значимост. Корелациите измерват как променливи или редовете са свързани. Преди изчисляването на коефициента на корелация, сортирайте вашите данни за очертания (което може да доведе до подвеждащи резултати) и доказателства за наличието на линейна връзка. Коефициентът на Пирсъновата корелация е мярка за линейна свързаност. Две променливи могат да бъдат напълно свързани, но ако връзката не е линейна, коефициентът на Пирсъновата корелация не е подходящ за статистическо измерване на тяхната свързаност.

Пример. Броят на игри, спечели от баскетболния отбор корелира ли се със средния брой на отбелязани точки на мач? Диаграмата показва, че има линейна връзка. Анализирайки данните от сезон 1994-1995 на НБА потвърждава, че коефициентът на Пирсъновата корелация (0,581) е значителен на ниво 0,01. Може да се предположи, че колкото са повече игрите спечелени на сезон, толкова по-малко точки са отбелязани от опонентите. Тези променливи са отрицателно корелирани (-0,401), и съотношението е значително на ниво 0,05.
Статистика. За всяка променлива: брой на случаи с не-липсващи стойности, средни (аритметични) и стандартни отклонения. За всяка двойка променливи: коефициентът на Пирсъновата корелация, Rho на Спиърмън, tau-*b* на Кендъл, крос-продукт на отклонения, и ковариация.
Данни. Използвайте симетрични количествени променливи за коефициентът на Пирсъновата корелация и количествени променливи или променливи с подредени категориите за Rho на Спиърмън и tau-*b* на Кендъл.
Предположения. Коефициентът на Пирсъновата корелация предполага, че всяка двойка променливи е нормална.

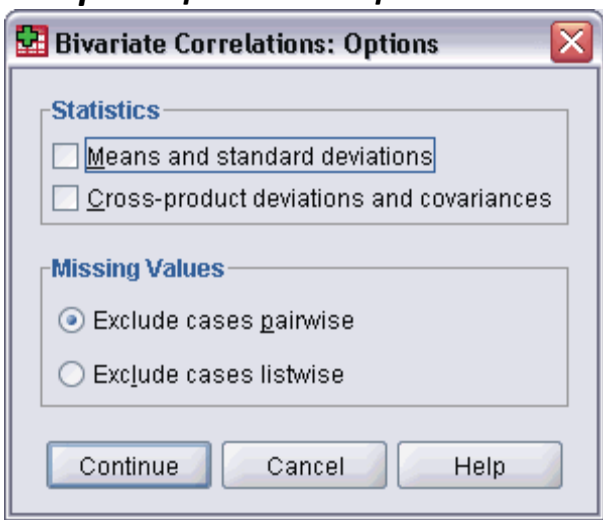
От менюто изберете:
- Analysis
  - Correlate
    - Bivariate ...



Фигура 22-1 Диалогов прозорец на корелацията

Изберете две или повече цифрови променливи.
Следните опции също са на разположение:
 Коефициенти на корелацията. За количествен, нормално разпределени променливи, изберете коефициент на Пирсъновата корелация . Ако данните не са разпространени или са подредени в категориите, изберете Rho на Спиърмън и tau-*b* на Кендъл, които измерват връзката между редивете.
Коефициентите на корелация се измершат в стойност от -1 (перфектна отрицателна връзка) до 1 (добра положителна връзка). Стойност 0 показва, че връзката не е линейна. При тълкуване на резултатите си, внимавайте да не си извадите никакви причинно-следствени изводи, дължащи се на значителна корелация.
 Тест за значимост. Можете да изберете двустранни или едностранни вероятности. Ако посоката на връзката е известна предварително, изберете едностранна вероятност. В противен случай, изберете двустранна.
 Обозначаване със знаме на значима корелация. Коефициентите на корелация значими на ниво 0,05, са идентифицирани с една звездичка, а тези значими на ниво 0,01 се идентифицират с две звездички.

Корелационни опции



Фигура 22-2 *Диалогов прозорец на корелационните опции*

Статистика. За Пирсъновата корелация можете да изберете един или и двете от следните неща:

 Средни и стандартни отклонения. Показан за всяка отделна променлива. Броят на случаите с не-липсващи стойности е също така показана. Липсващи стойности се обработват по променлива-към-променлива основа, независимо от настройката на вашите липсващи стойности.
Отклонения и ковариация на крос-продукта. Показан за всяка двойка променливи. Кръстосания продукт на отклонения е равен на сумата на продуктите на средно-коригираните променливи.
Това е числителя на коефициента на Пирсъновата корелация . Ковариацията е нестандартизираната мярка на връзката между две променливи, равна на отклонението на крос-продукта разделен от N-1.
Липсващи стойности. Можете да изберете един от следните:
 Изключи случаи по двойки. Случаи с липсващи стойности за един или и двете от един чифт променливи за един коефициент на корелация са изключени от анализа. Тъй като всеки коефициент се основава на всички случаи, които са валидни кодове на тази конкретна двойка променливи, цялата налична информация се използва във всички изчисления. Това може да доведе до серия от коефициенти, които се основават на различен брой на случаите.
 Изключи случаи по списък . Случаи с липсващи стойности за някоя от променливите са изключени от всички корелации.

Корелации и NONPAR CORR и Допълнителни функции за управление
Синтактичният език за управление също ви позволява да:
 Записване на корелационна матрица за Пирсъновата корелация , която може да се използва на мястото на сурови данни за получаване на други анализи, като анализ на факторите (с субкомандна матрица).
 Получаване на корелация на всяка променлива на списъка с всяка променлива на втори списък (с

помощта на думата WITH от субкомандата VARIABLES).
Вижте синтаксис на командите за справка и за пълна информация на синтаксиса.

Частични корелации

Процедурата на частичната корелация изчислява коефициентите на частичните корелации, които описват линейни връзка между две променливи, като същевременно контролират последиците от едно или повече допълнителни променливи. Корелациите са мерки от линейна свързаност. Две променливи могат да бъдат напълно свързани, но ако връзката не е линейна, корелационният коефициент не е подходящ за статистическо измерване на тяхната свързаност.

Пример. Има ли връзка между здравното финансиране и процента на заболяване? Въпреки че може да Очакваме връзката да е отрицателна, проучване отчита значителна положителна корелация:
С нарастването на финансирането на здравеопазването, процента на заболяваемост се увеличи. Контролът за степента на посещенията на осигуряващите здравни услуги, обаче, на практика премахва наблюдаваната положителна корелация. Здравното финансиране и процента на заболяване само изглеждат положително свързани, тъй като повече хора имат достъп до здравеопазването поради неговото финансиране, което води до повече болести докладвани от лекари и болници.

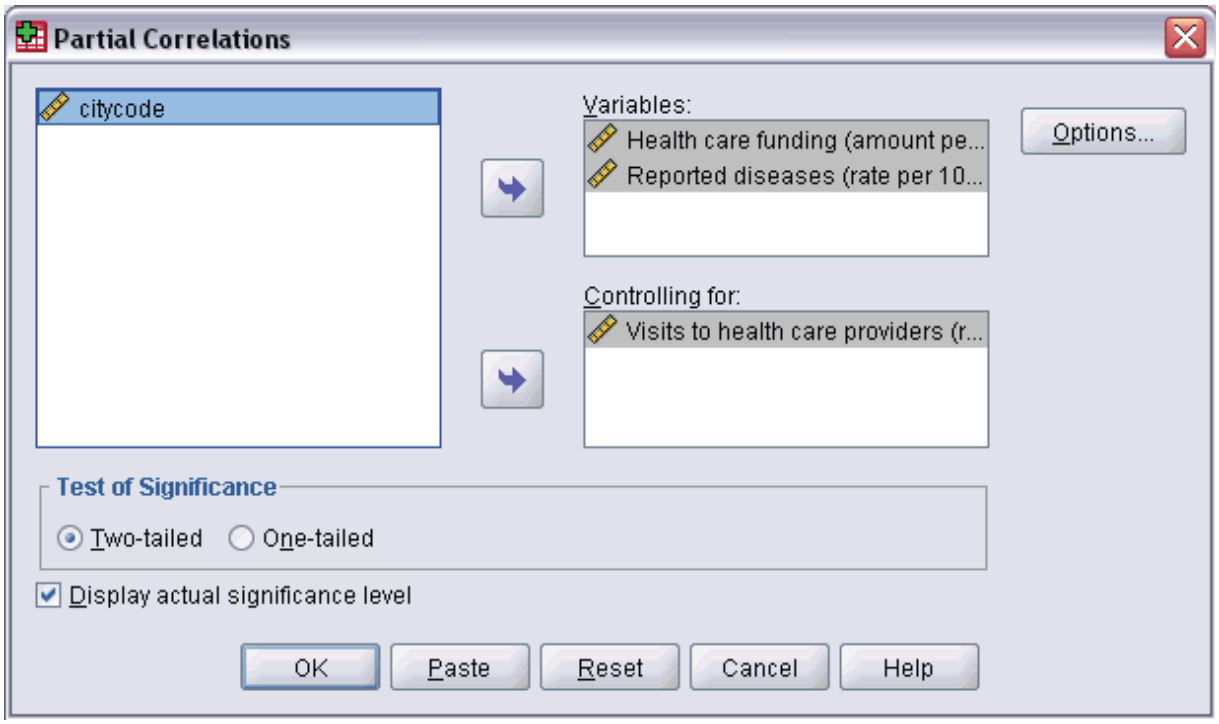
Статистика. За всяка променлива: брой на случаи с не-липсващи стойности, средни и стандарти отклонения. Матрици за частична и нулева корелация , със степени на свобода и
нива на значимост.

Данни. Използвайте симетрични, количествени променливи.

Предположения. Процедурата на частичната корелация предполага, че всяка двойка променливи е нормално.

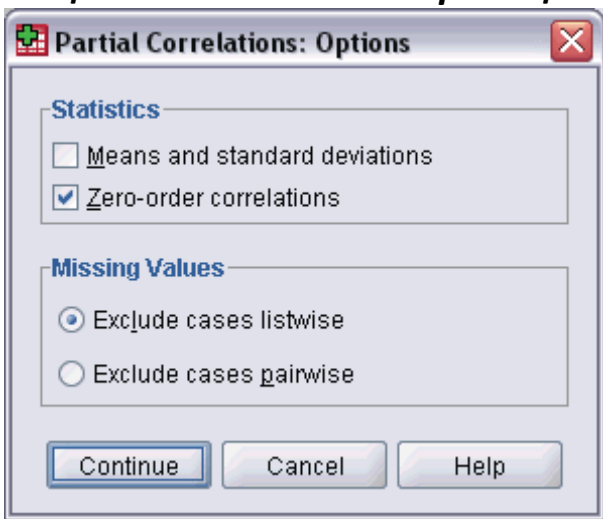
За да се получи частична корелация
От менюто изберете:

- Analysis
  - Correlate
    - Partial....



Фигура 22-3 *Диалогов прозорец на частичните корелации*

Опции на частната корелация



Фигура 22-4 *Диалогов прозорец на опциите на частна корелация*

Глава 23. Линейна регресия

Линейната регресия изчислява коефициентите на линейно уравнение, включващи един или повече независими променливи, които най-добре предскаже стойността на зависима променлива. Например, можете да опитате да предвидите годишните продажби на продавач (зависима променлива) от независими променливи като възраст, образование и дългогодишен опит.

Пример. Броят на игри, спечели баскетболен отбор в един сезон свързани ли са със средния броя на точките на отбора отбелязани по време на играта? Диаграмата показва, че тези променливи са линейно свързани. Броят на спечелени игри и средния брой точки, отбелязани от противника са също линейно свързани. Тези променливи имат негативно отношение. Ако броят спечелени игри се увеличава, средният брой спечелени точки от опонента намалява. С линейна регресия, можете да моделирате връзката на тези променливи. Един добър модел може да се използва, за да се предскаже колко игри ще спечели даден отбор.

Статистика. За всяка променлива: брой на валидни случаи, средно и стандартно отклонение. За всеки Модел: коефициент на регресията, матрицата на корелацията, частна и частична корелация, множество R, R2, коригирана R2, промени в R2, стандартна грешка на оценката, таблица на анализ-на-противоречие, прогнозни стойнисти и остатъците. Също така, 95% доверителни интервали за всеки коефициент на регресията, матрица на вариация-ковариация, коефициент на дисперсията на инфлацията, толерантност, Дърбин-Уотсън тест, мерки за разстояние (Mahalanobis, готвач), DfBeta, DfFit, интервали за прогнозиране, и диагностика на случаите. Диаграми: scatterplots, частични диаграми, хистограми, и нормална диаграми на вероятностите.

Данни. Зависимите и независимите променливи трябва да бъдат количествени. Категорични променливи величини, като религия, важна сфера на изучаване, или региона на пребиваване, трябва да се запише в двоични (сляпо) променливи или други видове контрастни променливи.

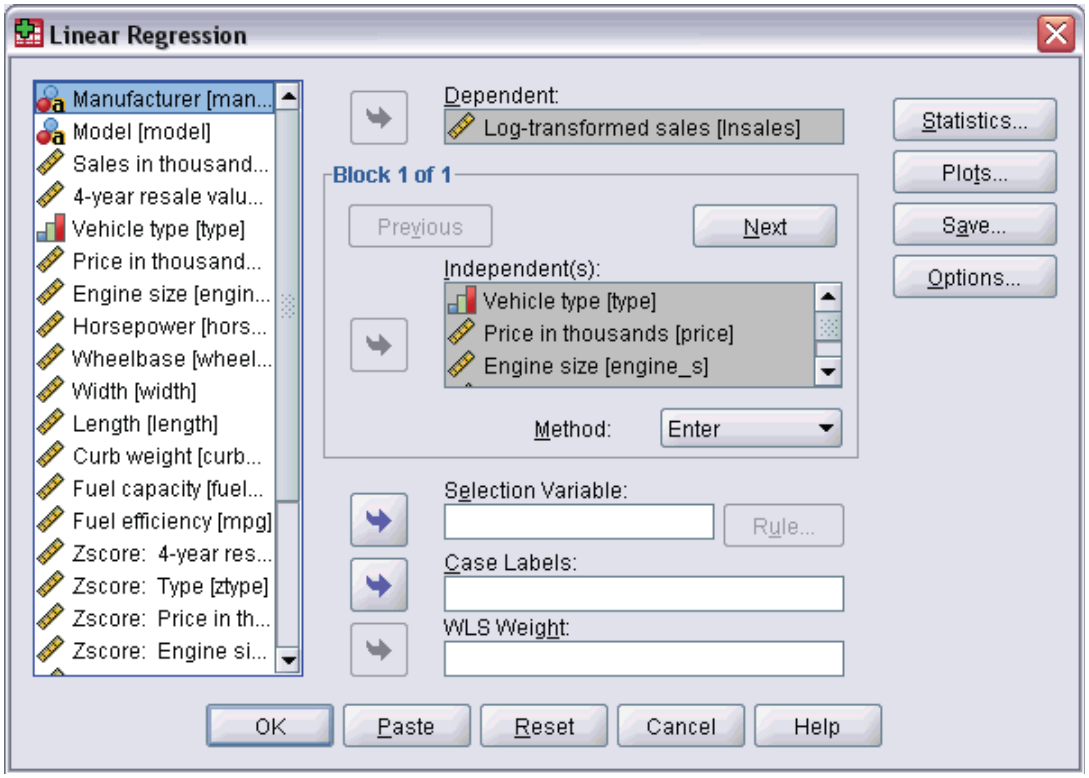
Предположения. За всяка стойност на независима променлива, разпределението на зависима променлива трябва да бъде нормално. Различията на разпределението на зависима променлива трябва да бъде постоянна за всички стойности на независимата променлива. Връзката между зависима променлива и всяка независима променлива трябва да бъде линейна, и всички наблюдения трябва да бъдат независими. За да се получи анализ на линейната регресия.

От менюто изберете:

- Analysis
  - Regression
    - Linear …

В диалоговия прозорец на линейната регресия изберете цифрова зависима променлива.
Изберете един или няколко числови независими променливи.
По желание можете да:

1. групирате независими променливи в блокове и определите различни методи за влизане за различните подгрупи на променливи.
2. Изберете селектирени променливи за ограничаване на анализа до определена подгрупа от случаите, които имат специална стойност (и) за тази променлива.
3. Изберете променлива за идентифициране на определен случай за определяне на точки върху диаграмите.
4. Изберете цифров WLS тегло на променливата при средно претеглена най-малко анализ на квадрат. WLS. Позволява ви да получите средно претеглена модел на квадрат. Основни информационни данни се претеглят с реципрочни на техните вариации. Това означава, че наблюденията с големи разлики имат по-малко значение за анализа от наблюдения, свързани с малки разлики. Ако стойността на теглото на променливата е нула, отрицателно или липсва, случаят се изключват от анализа.



Фигура 23-1 Диалогов прозорец на линейна регресия

Методите на избор на променлива линейна регресия

Методът за избор ви позволява да зададете колко независими променливи да влязът в анализа. Чрез използването на различни методи можете да създавате различни регресионни модели от един и същи набор от променливи.

- Натиснете (Регресия). Процедура за подбор на променливи, в която всички променливи в даден блок се вписват в една единствена стъпка.
- Поетапно. На всяка стъпка независима променлива, която не е в уравнението с най-малката вероятността от F се вписва, но само ако тази вероятност е достатъчно малък. Променливите, които са вече в уравнението на регресия, са премахнати, ако вероятността им на F става достатъчно голяма. Методът се прекратява, когато няма повече променливи, които са допустими за включване или отстраняване.
- Премахни. Процедура за подбор на променливите, в която са отстранени всички променливи в блок в една единствена стъпка.
- Задно премахване. Процедура за подбор на променливата, в която са вписани всички променливи в уравнението и след това последователно се отстраняват. Променливата с най-малката частична корелация със зависима променлива се счита за първа за премахване. Ако тя отговаря на критерия за отстраняване, тя се отстранява. След като първата променлива е отстранена, променливата, която остава в уравнението с най-малкото частично корелация се счита за следващата. Процедурата спира, когато не са останали променливи в уравнението, които отговарят на критериите за отстраняване.
- Селекция напред. Процедура по поетапен подбор на променливи, които последователно се вписват в модела. Първата променлива, която влиза в уравнението е тази с най-положителна или отрицателна корелация с зависима променлива. Тази променлива се вписва в уравнението, само ако тя отговаря на критерия за влизане. Ако първата променлива влезе в уравнението, независимата променлива, която не е в уравнението и която е с най-голямата частна корелация, се счита за следваща. Процедурата спира, когато не са останали променливи, които отговарят на критериите за влизане. Стойността на значимост в уравнението се базира на напасване на единен модел. Ето защо, стойности на занчимост обикновено са недействителни, когато е използван прогресивен метод (поетапно, напред, или назад). Всички променливи трябва да минат критерия за толерантност, за да се впишат в уравнението, независимо от метода. Нивото на толерантност е 0,0001. Също така, променливата не се вписва, ако ще доведе до намаляване на толерантността на друга променлива, която вече е в модела. Всички избрани независими променливи се добавят към единен модел на регресия. Въпреки това, можете да определите

различни методи за влизане за различните подгрупи от променливи. Например, можете да въведете един блок от променливи в регресионния модел използвайки поетапния подбор и втори блок чрез използване на избор „напред”. За да добавите втори блок от променливи към регресионен модел, щракнете върху Напред.

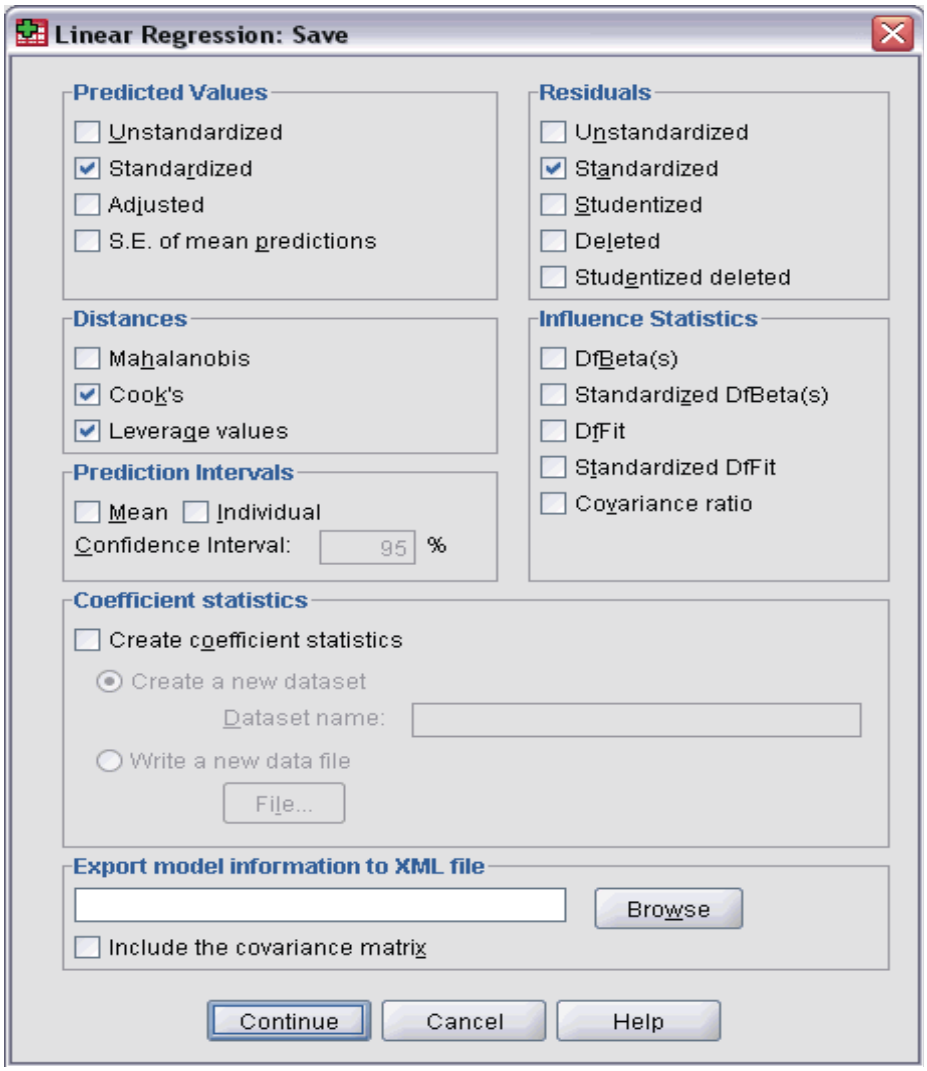
Комплект правила за линейната регресия

Фигура 23-2 Диалогов прозорец за правилата за линейна регресия

Случаите, определени от селективното правило, са включени в анализа. Така например, ако изберете променлива, изберете равени, и тип 5 за стойност, тогава само случаи, за които избраната променлива има стойност, равна на 5, са включени в анализа. Стойността на редовете също е разрешена.

Таблиците могат да помогнат в утвърждаването на допусканията за нормалност, линейност и равенство на дисперсиите. Таблиците са полезни и за откриване на необичайни наблюдения и значими случаи. След като ги запазите като нови променливи, прогнозирани стойности, остатъци, и други диагностики са на разположение в Редактор на данни за изграждането на таблици от независимите променливи. Следните таблици са на разположение:
Scatterplots. Можете да използвате две от следните: за зависима променлива, стандартизирани прогнозни стойности, стандартизирани остатъците, заличени остатъците, коригирани прогнозни стойности. Сравнете стандартизираните остатъци и стандартизираните прогнозни стойности, за да проверите за линейност и равенство на дисперсиите.
Източник-списък на променливи . Изброява зависимите променливи (DEPENDNT) и следните прогнозни и
Остатъчни променливи: Стандартизирани прогнозни стойности (* ZPRED), стандартизираните остатъци (* ZRESID),
Изтрити остатъци (* DRESID), коригирани прогнозни стойности (* ADJPRED), Studentized остатъците (* SRESID), Studentized заличени остатъците (* SDRESID).
Изработване на всички частни участъци. Показва диаграми на остатъците на всяка независима променлива и остатъците на зависима променлива, когато двете променливи намаляват отделно от останалите независими променливи. Най-малко две независими променливи трябва да бъдат в уравнението за да бъда представена частична диаграма.
Стандартизирана Остатъчна диаграма. Можете да получите хистограма на стандартизираните остатъци и таблица на нормалната Вероятност като сравнявате разпространението на стандартизирани остатъци и нормалното разпределение.
Ако всички таблици са поискани, значи ви е показана обобщена статистика за стандартизираните прогнозни стойности и стандартизираните остатъци (* ZPRED и * ZRESID).

Линейна регресия: запазване на нови променливи



Фигура 23-4 диалоговия прозорец за записване на линейна регресия

Можете да запазите прогнозните стойности, остатъци, както и други полезни статистически данни за диагностика. Всеки избор добавя една или повече нови променливи, които да активен файл с данни. Прогнозираната стойност. Стойности, които регресният модел предвижда за всеки отделен случай.

- **Нестандартизирани**. Стойността на модела прогнозира за зависима променлива.
- **Стандартизирани**. Трансформация на всяка прогнозирана стойност в стандартизирана форма. Това означава, че прогнозираната средна стойност се изважда от прогнозната стойност, а разликата е разделена на стандартното отклонение от очакваните стойности. Стандартизираната прогнозирана стойност е със средна стойност от 0 и стандартно отклонение от 1.
- **Коригирани**. Приблизителната стойност за един случай, когато този случай се изключва от изчисляването на коефициентите на регресия.
- **S.E. на средната прогнози.** Стандартни грешки от очакваното стойности. Прогнозата на стандарта отклонение от средната стойност на зависима променлива за случаи, които имат едни и същи стойности на независимите променливи.

Разстояния. Мерки за идентифициране на случаи с необичайни комбинации от стойности за независимите променливи може да спаси прогнозира ценности, остатъци, както и други полезни статистически данни за диагностика. Всеки избор добавя една или повече нови променливи, които да си активен файл с данни.

Прогнозираната ценности. Ценности, които регресионен модел предвижда за всеки отделен случай.

- **Unstandardized.** Стойността на модела прогнозира за зависима променлива.
- **Стандартизирани.** А трансформацията на всеки прогнозира стойност в стандартизирана форма. Това означава, че прогнозира средна стойност се изважда от прогнозната стойност, а разликата е разделено стандартното отклонение от очакваното стойности. Стандартизиран прогнозира стойности са средно от 0 и стандартно отклонение на 1.

 Коригирани. Приблизителната стойност за един случай, когато този случай се изключват от изчисляването на
Коефициентите на регресия.
 S.E. на средните прогнози. Стандартни грешки от очакваните стойности. Прогнозата на стандартно отклонение от средната стойност на зависима променлива за случаи, които имат едни и същи стойности на независимите променливи.
Разстояния. Мерки за идентифициране на случаи с необичайни комбинации от стойности за независим променливи и случаи, които могат да окажат голямо влияние на регресионния модел.

 Mahalanobis. Мярка за това колко стойностите на определен случай на независими променливи се различават от средната стойност на всички случаи. Голямо разстояние на Mahalanobis идентифицира случай, който има много високи стойности на една или повече от независимите променливи.
 Cook's. Една мярка за това колко остатъците от всички случаи ще се променят, ако всеки конкретен случай бъде изключен от изчисляването на коефициентите на регресия. Голяма Д на Кук показва, че с изключение на един случай от изчисляването на регресията променя коефициентите
значително.
 Лостови стойности. Мерки за въздействие на част от регресията. В центъра
лост варира от 0 (няма влияние) до (N-1) / N. Прогнозни интервали от време. Горните и долните граници за средните и индивидуалните прогнозирани интервали.
 Средна. Долна и горна граници (две променливи) за предсказване интервал от средната прогнозира отговор.
 Лице. Долна и горна граници (две променливи) за прогнозирани интервали на зависимите променливи за един случай.
 Доверителен интервал. Въведете стойност между 1 и 99,99, за да се определи нивото на достоверност за два прогнозирани интервала. Средна или индивидуална стойност трябва да бъдат избрани, преди да бъде въведена тази стойност. Типични стойности на доверителен интервал са 90, 95 и 99.
Остатъци. Действителната стойност на зависима променлива минус стойността предсказана от регресия уравнение.

 Нестандартизирани. Разликата между наблюдаваната стойност и стойността предсказана от модела.

 Стандартизирани. Остатъкът разделен на оценката на неговото стандартно отклонение.

Стандартизираните остатъци, които са известни като Пирсънови остатъци, имат стойност от 0 и стандарто отклонение от 1.

Studentized. Остатъкът разделен на оценката на неговото стандартно отклонение, което варира

за всеки отделен случай, в зависимост от разстоянието на стойностите на всеки отделен случай независимите променливи
от средните стойности на независимите променливи.

 Заличени. Остатъкът за случаите, когато този случай се изключва от изчисляването на

регресните коефициенти. Това е разликата между стойността на зависима променлива и предвижданата коригирана стойност.

 Заличена Studentized Заличените остатъци за случай разделен на стандартната грешка.

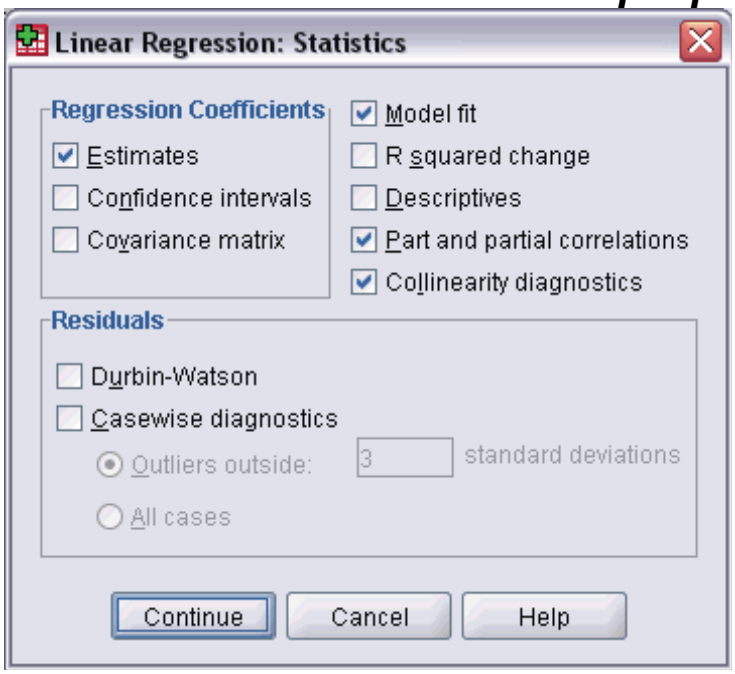
Разликата между заличените Studentized остатъцу и свързаните с нея Studentized остатъци показва каква разлика елиминирането на случая прави при своята собствена прогноза.
статистика на влиянието. Промяната в коефициентите на регресия (DfBeta [и]) и прогнозираните стойности (DfFit), които са резултат от изключването на един конкретен случай. Стандартизираните DfBetas и DfFit стойности също са на разположение заедно с съотношението на ковариация.

 DfBeta (a). Разликата в бета стойността е промяната в коефициента на регресия, което води до

изключването на един конкретен случай. Стойността се изчислява за всеки термин в модела, включително константата.
 Стандартизиран DfBeta. Стандартизирана разлика в бета стойност. Промяната в коефициента на регресия води до изключването на един конкретен случай. Вие може да искате да разгледате случаи с абсолютни стойности, по-големи от 2, разделени на корен квадратен от N, където N е броят на случаите. Стойността се изчислява за всеки термин в модела, включително константата.
 DfFit. Разликата в годни стойност е промяната в прогнозната стойност, която е резултат от

изключване на един конкретен случай.
 Стандартизиран DfFit. Стандартизиран разликата в стойността. Промяната в прогнозната стойност, която води до изключването на един конкретен случай. Вие може да искате разгледате стандартни стойности, които в абсолютна стойност са повече от 2 пъти на квадратния корен на P / N, където п е броят на параметрите в модела и N е броят на случаите.
 Съотношение на ковариация. Съотношението на детерминанта на ковариационната матрица в конкретен случай изключен от изчисляването на коефициентите на регресия към детерминантата на ковариационната матрица с включени всички случаи. Ако съотношението е близо до 1, случаят не променя значително матрицата на ковариация.
Статистика на коефициента. Спестява регресионни коефициенти за набор от данни или файл с данни. Базите от данни са налични за последващо използване, но не са записани като файлове, освен ако изрично да са записани като такива преди края на сесията. DataSet имена трябва да съответстват на правилата за именуване на променливе. За повече информация, виж имената на променливите в глава 5 на стр. 74.
Износ на информация за XML файл. Параметрите на оценки и (по желание) техните ковариации са изнесени за даден файл в XML (PMML) формат. SmartScore и SPSS Server (отделен продукт) може да използва този модел файл, за да се прилага информационният модел за други файлове с данни за точкувани цели.

Статистика на линейната регресия



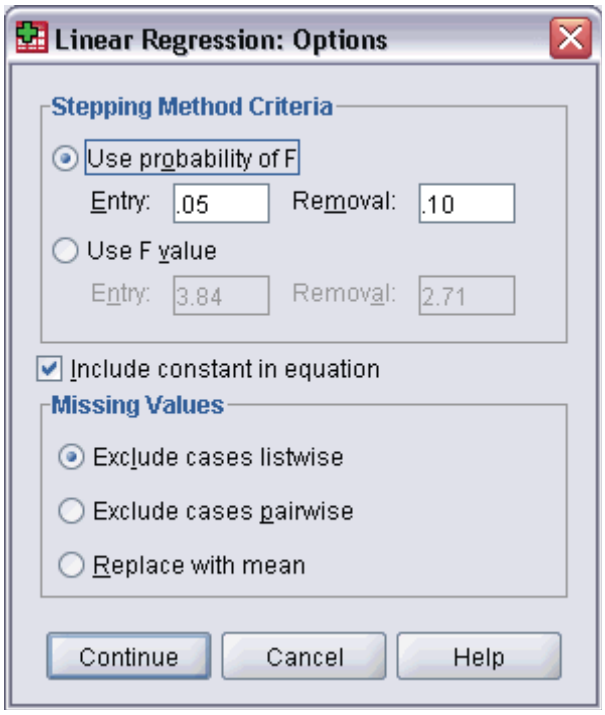
Фигура 23-5 *Диалогов прозорец на статистиката*

Следните статистически данни са достъпни:

Коефициенти на регресията. Оценката показва Регресионен коефициент Б, стандартна грешка на Б,стандартизиран коефициент бета, т стойност за Б, и двестранно ниво на значимост на Т. интервали на Доверие, които показват 95% доверителни интервали за всеки коефициент на регресията или ковариационната матрица. Ковариационната матрица показва варионна- ковариационна матрица на регресните коефициенти с ковариациите извън диагонала, а вариациите на диагонала. Корелационната матрица също е показана.
Прилагане на Модела . Променливите вписани и отстранени от модела, са включени и следните goodness-of-fit статистики са показани: многобройни R, R2 и коригирани R2, стандартна грешка на оценката и таблица на анализ-на-противоречие. R квадрат промята. Промяната в R2 статистиката, която се получава чрез добавяне или изтриване на независима променлива. Ако R2 промяна свързана с променливата е голяма, това означава, че променливата е добра прогноза за зависима променлива.
Описателни. Дава брой валидни случаи, средната стойност и стандартното отклонение за всяка променлива в анализа. Установена е корелационна матрица с едностранно ниво на значимост и броят

от случаите за всяка връзка са показани.
Частична корелация. Съотношението, което остава между две променливи, след премахване на корелация, че се дължи на тяхното взаимно свързване с други променливи. Съотношението между зависима променлива и независима променлива, когато линейния ефекти на другите независими променливи в модела са били отстранени от двете.
Частна корелация. Съотношението между зависима променлива и независима променлива, когато линейни ефекти на други независими променливи в модела са били отстранени от независимата променлива. Тя е свързана с промяната в R-квадрат, когато дадена променлива се прибави към уравнение. Понякога наричана получастна корелация.
Collinearity диагностика. Collinearity (или multicollinearity) е нежелана ситуация, когато една независима променлива е линейна функция на други независими променливи. Eigenvalues на мащабирана кръстосана продуктува матрица, индекси на състояние, и пропорции на вариации-разлагане са показани заедно с факторите, дисперсията на инфлацията (VIF) и отклоненията за отделните променливи.
Остатъци. Показва Дърбин-Уотсън тест за серийната корелация на остатъците и методи за диагностика на случаите, които отговарят на критерия за избор (outliers горе н стандартни отклонения).

Опции на линейната регресия



Фигура 23-6 *Диалогов прозорец на* ***Опции на линейната регресия***

Следните опции са възможни:

Засилващ метод на критерии. Тези опции се прилагат, когато нито напред, назад, или поетапния подборен метод на променливите е определен. Променливите могат да бъдат вписани или премахнати от модела в зависимост или от значението (вероятността) на стойността на F или самата стойност на F.
 Ползване на вероятностите на F. Променливата се вписва в модела, ако нивото на значимост на стойността на F е по-малка от стойността за влизане и се отстранява, ако нивото на значимост е по-голямо от стойността за премахване.Стойността за въвеждане трябва да бъде по-малка от тази за отстраняване, както и двете стойности трябва да бъдат положителни. За да въведете
повече променливи в модела, се увеличи стойността влизане. За да премахнете по-променливи от модел, по-ниска стойност на премахване.
 Използване на F стойност. А променлива се вписва в модела си Е, ако стойността е по-голяма от стойността Влизане и се отстранява, ако F стойност е по-малко от стойността отстраняване. Въведеното трябва да бъде по-голяма от изваждането и двете стойности трябва да бъде положителен. За да въведете повече променливи в модела, по-ниски от стойността за влизане. За да премахнете повече променливи от този модел, увеличете стойността за отстраняване.

Включване на константа в уравнението. По подразбиране, регресионен модел включва константа. Премахването на тази опция води до регресия чрез произход, която е рядкокс. Някои от резултатите на регресия чрез произход не са сравними с резултатите на регресия, която включва константа. За пример, R2, не може да се тълкува по обичайния начин.

Липсващи стойности. Можете да изберете един от следните:

- **Изключи случаи по списъци.** Само в случаи с валидни стойности за всички променливи са включени в анализите.
- **Изключи случаи по двойки.** Случаи с пълни данни за чифт променливи са взаимно свързани и са използвани за изчисляване на коефициента на корелация, на който се основава регресионен анализ. Степените на свободата се основават на минималните двойки N.
- **Заменете със средна стойност**. Всички случаи са използвани за изчисления, със средна стойност на променливите, коитозаменят липсващите наблюдения.

Допълнителни функции за управление на регресията

Командата синтаксис на езика също ви позволява да:

- Напишете корелационната матрица или да прочетете една матрица на мястото на необработените данни, за да получите вашият регресивен анализ (със подкомандата „матрица”).
- Посочете нивата на толеранс (с подкомандата „критерии”).
- Получаване на множество модели за една и съща или различни зависими променливи (с подкомандите „метода” и „зависим”).
- получите допълнителна статистика (с подкомандите „DESCRIPTIVES” и „СТАТИСТИКА”).

Вижте синтаксис на командите за справка и пълна информация на синтаксиса.

Глава 24. Непараметрични тестове

Nonparametric Tests

Непараметричните тестове предвиждат редица процедури, които не се нуждаят от предположенията относно разпределенията.

Хи-квадрат тест. Представя променливата в категории и изчислява хи-квадрат базирайки се на разликите между наблюдаваните и очаквани честоти.

Биномен Тест. Сравнява наблюдаваната честота във всяка категория на дихотомните променливи с очакваните честоти от биномното разпределение.

Runs Test. Проверява дали появата на две стойности на дадена променлива е случайна.

One-Sample Kolmogorov-Smirnov Test. Сравнява наблюдаваната кумулативно разпределителна функция за променлива със специфична теоретична дистрибуция, която може да бъде нормална, единна, показателна или Поасон.

Two-Independent-Samples Tests. Сравнява две групи от случаи на една променлива по Mann-Whitney U тест, two-sample Kolmogorov-Smirnov тест, тест на Moses за изпитване на крайни реакции и Wald- Wolfowitz runs тест.

Two-Related-Samples Tests. Сравнява разпределенията на две променливи, като наличните възможности са Wilcoxon signed-rank тест, тест на знака, както и тест McNemar.

Tests for Several Independent Samples. Сравнява две или повече групи от случаи на една променлива. Наличните тестове са The Kruskal-Wallis тест, проверка на медианата и Jonckheere-Terpstra тест.

Tests for Several Related Samples. Сравнява разпределенията на две или повече променливи.Тест на Friedman , Kendall's *W*, и Cochran's *Q* са наличните тестове за сравнение.

Квартили и средна стойност, стандартно отклонение, минимум, максимум, и броя на не липсващите случаи са достъпни за всички по-горе споменати изследвания.

Хи-квадрат Тест

Процедурата представя променливата в категории и изчислява статистиката на хи-квадрат. Този тест сравнява наблюдаваните и очаквани честоти във всяка категория за да провери дали всички категории съдържат еднакви пропорции на стойностите и проверява дали зададената от потребителя пропорция е на лице.

Примери. Хи-квадрат тест може да се използва за да се провери дали един пакет желирани бонбони съдържа еднакво количество сини, кафяви, зелени, оранжеви, червени и жълти бонбони. Също така може да се провери дали пакетчето съдържа 5% сини, 30% кафяви, 10% зелени, 20% оранжеви, 15% червени и 15% жълти бонбони.

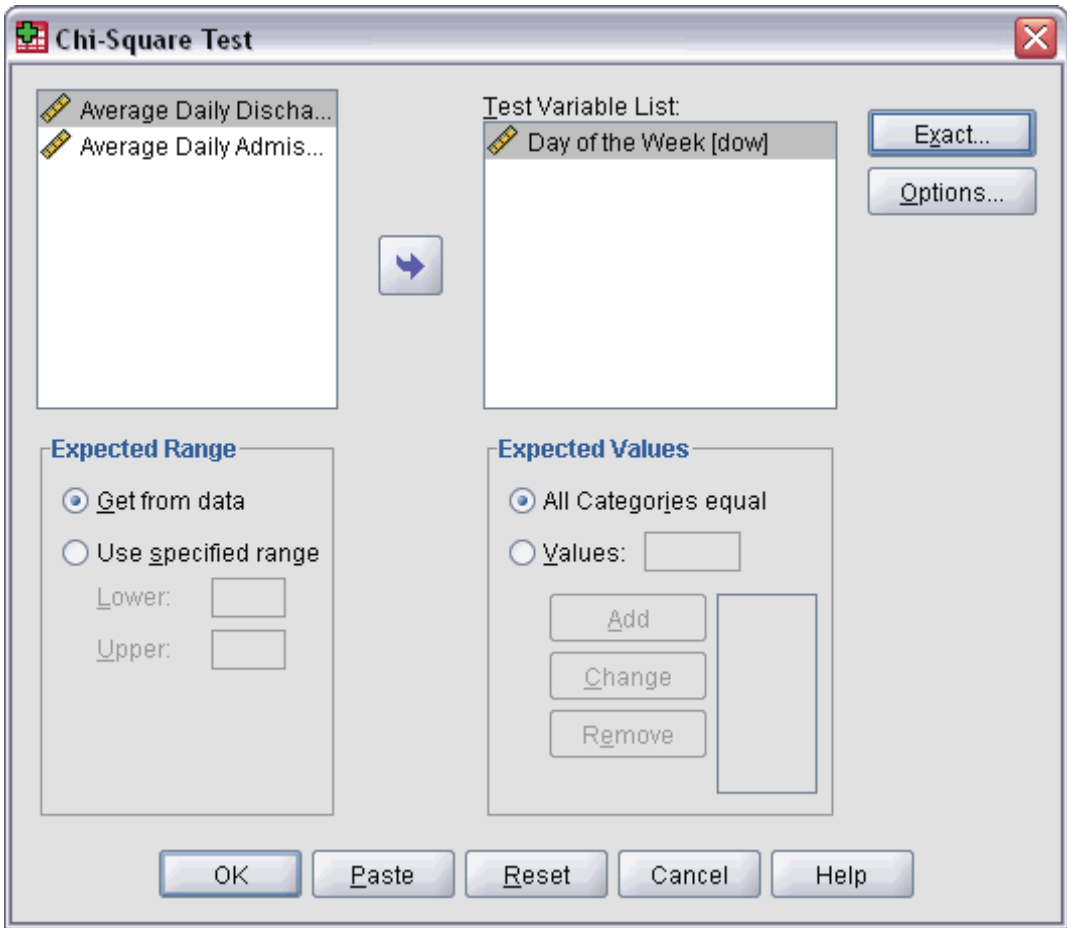
Статистики. Средна стойност, стандартно отклонение, минимум, максимум и квартили. Процентите на липсващи и налични случаи, броя на очакваните и наблюдаваните случаите за всяка категория, остатъци и хи-квадрат статистика.

Данни. Използвайте подреден или неподреден цифрови променливи (редни или номинални равнища на измерване). За да се превърне string променливи в числови променливи, използвайте функцията Automatic Recode, която е на разположение в меню Transform.

Предположения. Непараметричните тестовете не изисква предположения относно формата на извършените разпределение. Данните се приемат за случайна извадка. Очакваните честоти за всеки категория трябва да бъдат най-малко 1. Не повече от 20% от категориите, трябва да очаква честотата на по-малко от 5.

За да направите Хи-квадрат тест от менюто изберете:

- Analyze
  Nonparametric Tests
  Chi-Square...



Фигура 24-1 *Диалогов прозорец на Хи-квадрат тест*

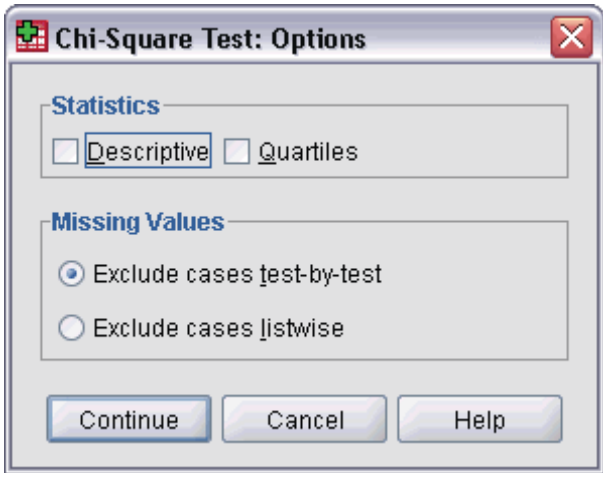
Изберете една или повече променливи. Всяка променлива се представя като отделен тест.
По желание, щракнете върху Options за описателна статистика, квартили, както и контрол на липсващите данни.

Хи-квадрат Тест очаквани граници и очаквани стойности

Очаквани граници. По подразбиране, всяка отделна стойността на променливата е определено като категория. За да се създадат категории, в рамките на определен период, изберете Use specified range и въведете целочислени стойности за долна и горна граници. Категориите са определени за всяка целочислена стойност в рамките на включената гама, като случаи с показатели извън границите, са изключени. Например, ако зададете стойност 1 за Долна и на стойност от 4 за горната, само целочислени стойности от 1 до 4 се използват за хи-квадрат тест.

Очаквани стойности. По подразбиране, всички категории имат равни очакваните стойности. Категории може да има и потребителя с посочени очакваните пропорции. Изберете Values (стойности), въведете стойността, която е по-голям от 0, за всяка категория на тестовата променлива, и след това щракнете върху Add (Добави). Всеки път, когато добавите стойност, тя се появява в дъното на списъка със стойност. Редът на стойностите е важен, той отговаря на възходящия ред на категорийните стойности на тестовата променлива. Първата стойност в списъка, отговаря на най-ниската група стойности на тестовата променлива, а последната стойност съответства на най-високата стойност. Елементите на стойността в списъка са обобщени и след това всяка стойност е разделена на тази сума, за да се изчисли съотношението на случаите в съответната категория. Например, за стойности 3, 4, 5, 4 очакваните пропорции са 3 / 16, 4 / 16, 5 / 16 и 4 / 16.

Хи-квадрат тест Опции



Фигура 36-2 *Диалогов прозорец с опции за Хи-квадрат тест*

Statistics. Можете да изберете една или и две обобщени статистики.

- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил.

Missing Values. Контролира работата с липсващи стойности.

- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.

- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Command допълнителни функции (Хи-квадрат тест)

Командния синтаксис също ви позволява да:

- Посочете различни минимални и максимални стойности и очакваните честоти за различни променливи (с подкоманда CHISQUARE).

- Тестване на една и съща променлива срещу различни очаквани честоти или да използва различни диапазони (с подкоманда EXPECTED).

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Биномен Тест

Биномен тест е процедура за сравняване на наблюдаваните честоти на две категории на дихотомна променлива до честотите, които се очакват в рамките на биномно разпределение с вероятност определен параметър. По подразбиране, вероятностния параметър за двете групи е 0.5. За промяна на вероятностите, можете да въведете тестова пропорция за първата група. Вероятността за Втората група ще бъде 1 минус специфичната вероятност за първата група.

Пример. Когато хвърляне с монета, вероятността да се падне ези е равен на 1 / 2. Въз основа на тази хипотеза, монета е хвърлена 40 пъти, и резултатите се записват (ези-тура. От Бином теста, може да откриете, че 3 / 4 от хвърлянията са били ези и че се наблюдава ниско ниво на значимост (0.0027). Тези резултати показват, че вероятността да се падне ези не е равена на 1 / 2; монетата е вероятно предубедена.

Статистики. Средна стойност, стандартно отклонение, минимум, максимум, брой на не липсващи случаи и квартили.

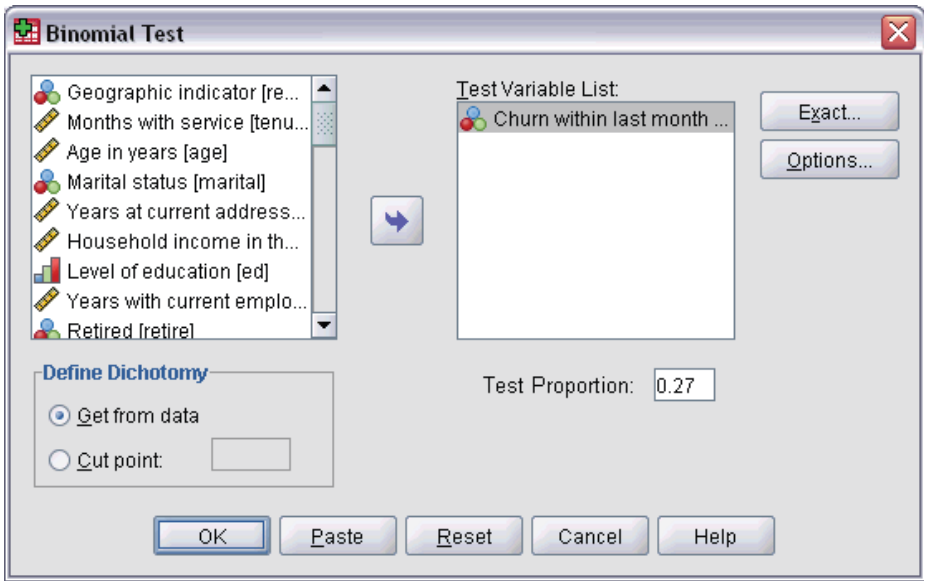
Данни. Променливите трябва да бъдат числови и дихитимни. За да се превърне string променливи в числови променливи, използвайте функцията Automatic Recode, която е на разположение в меню Transform. Дихотомна променлива означава че променливата може да приеме само две възможни стойности: ДА или НЕ, ВЯРНО или НЕВЯРНО, 0 или 1 и т.н. Ако променливите не са дихотомни то трябва да определите т.нар. точка на срез (cut point). Измерванията над тази точка се назначават в една група, а тези под точката на срез - в друга група.

Предположения. Непараметричните тестовете не изисква предположения относно формата на извършените разпределение Данните се приемат за случайна извадка.

ЗА да направите Биномен Тест

От менюто изберете:

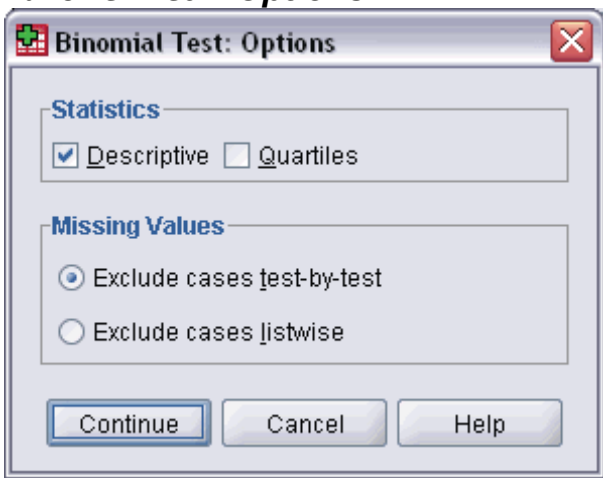
- Analyze
  - Nonparametric Tests
    - Binomial...



Фигура 24-3 *Диалогов прозорец Биномен Тест*

Изберете една или повече числови променливи.
За по-детайлна настройка, кликнете Options за descriptive statistics, квартили, и настройка на липсващи данни.

Биномен Тест Options



Фигура 24-4 *Диалогов прозорец за опции на Биномен Тест*

Statistics. Можете да изберете една или и двете сумарни статистики.
- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил

Missing Values. Контролира работата с липсващи стойности.
- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Command Additional Features (Binomial Test)

Командния синтаксис също ви позволява да:
- Избирате определени групи (и да изключвате други групи) когато дадена променлива притежава повече от две категории (със субкоманда BINOMIAL).
- Определете различни cut points или вероятности за различни променливи (със субкоманда BINOMIAL).
- Тествайте една и съща променлива с различни cut points или вероятности (със субкоманда EXPECTED).

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Сериен тест (Runs Test)

Процедурата Run Тест проверява дали появата на две стойности на дадена променлива е на случаен принцип. Run е последователност на подобни наблюдения. Проба с твърде много или твърде малко Runs предполага, че извадката не е на случаен принцип.

Пример. Да предположим, че 20 души са анкетирани, за да разберете дали те ще купят даден продукт. Предполагаемата случайност на пробата ще бъде сериозно компрометирана, ако всички 20 души са от един пол. Runs тест може да се използва за да се определи дали пробата е изготвен на случаен принцип или не.

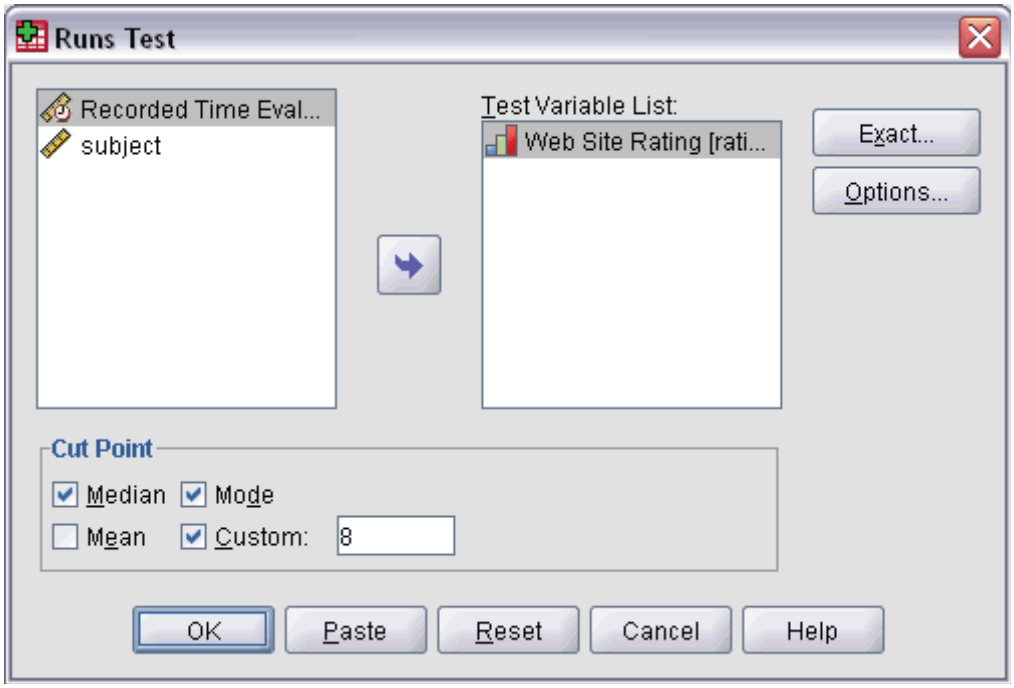
Статистики. Средна стойност, стандартно отклонение, минимум, максимум, брой на не липсващи случаи и квартили.

Данни. Променливите трябва да бъдат числови. За да се превърне string променливи в числови променливи, използвайте функцията Automatic Recode, която е на разположение в меню Transform.
Предположения. Непараметричните тестове не изискват предположения относно формата на извършените разпределения. Използвайте проби от непрекъснати вероятностни разпределения.

За да стартирате СерийнияТест

От менюто изберете:

- Analyze
  - Nonparametric Tests
    - Runs...



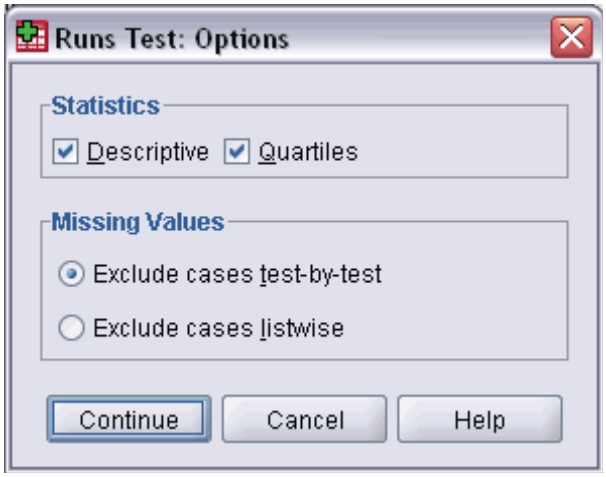
Фигура 34-5 *Добавяне на собствена cut point*

Изберете една или повече числови тестови променливи
По желание, натиснете Options за descriptive statistics, квартили, и настройки за липсващи данни.

Runs Test Cut Point

Cut Точка. Задава точка на срез за дихотомизация на променливите, които сте избрали. Можете да използвате наблюдаваната средна стойност, медиана, или мода, или можете да използвате определена стойност като краен пункт. Случаи със стойности, които са по-малко от cut точката са възложени в една група, както и случаи със стойности, които са по-голям или равен на cut точката, са възложени на друга група. Един тест се извършва за всяка избрана cut точка.

Runs Test Опции



Фигура 24-6 *Диалогов прозорец с опции за Runs тест*

Statistics. Можете да изберете една или и двете сумарни статистики.
- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил

Missing Values. Контролира работата с липсващи стойности.
- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Допълнителни функции (Runs Test)

Командния синтаксис също ви позволява да:
- Определяне на различни cut точки за различни променливи (Със субкоманда RUNS).
- Test Да тествате едни и същи променливи с различни cut точки (Със субкоманда RUNS).

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Едноизвадков тест на Колмоголов-Смирнов

One-Sample Kolmogorov-Smirnov Тест сравнява наблюдаваната кумулативна разпределителна функция на променлива с определена теоретична дистрибуция, която може да бъде нормална, еднородна, Poisson, или експоненциална. Kolmogorov-Smirnov *Z* се изчислява от най-голямата разлика (в абсолютна стойност) между наблюдаваните и теоретични кумулативно разпределени функции. Този goodness-of-fit тест проверява дали наблюденията са могли да произлезнат от определена дистрибуция.

Пример. Много параметрични тестове изискват нормално разпределени променливи. One-sample Kolmogorov-Smirnov тест може да бъде използван за да определи дали дадена променлива (например *доход*) е нормално разпределена.

Статистики. Средна стойност, стандартно отклонение, минимум, максимум, брой на не липсващи случаи и квартили.

Данни. Използвайте количествени променливи.

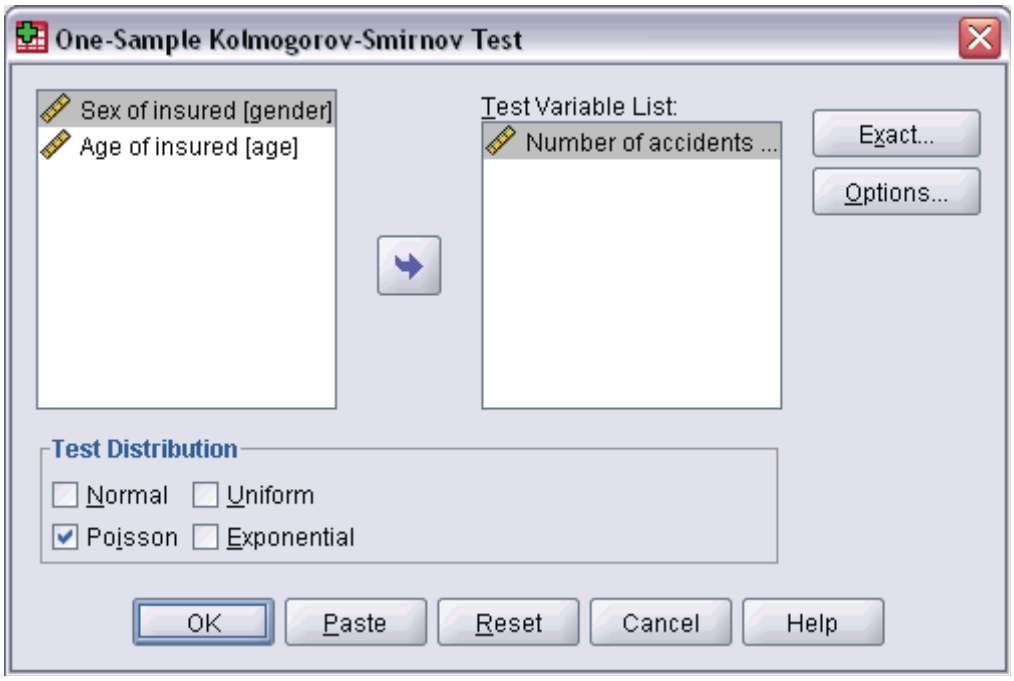
Предположения. Тестът предполага, че параметрите на тестовото разпределение са предварително определени. Тази процедура оценява параметрите на извадката. Средната стойност и стандартното отклонение на извадката са параметрите за нормално разпределение,минимални и максимални стойности определят обхвата на равномерно

разпределение, самата извадка е параметъра за разпределение по Poisson, а средната стойност е параметъра за експоненциално разпределение. Силата на теста да открива отклонения от хипотетичните разпределения мрежа може да бъде сериозно намален. За тестване с нормално разпределение с определени параметри, коригирайте K-S Lilliefors тест (на разположение в Explore).

За да стартирате One-Sample Kolmogorov-Smirnov Тест

От менюто изберете:

- Analyze
  - Nonparametric Tests
    - 1-Sample K-S...

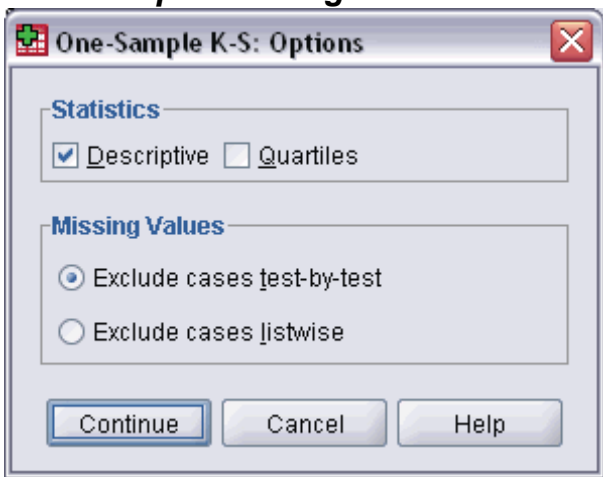


Фигура 24-7 *Диалогов прозорец на One-Sample Kolmogorov-Smirnov Тест*

Изберете един или няколко числови променливи тест. За всяка променлива се произвежда отделен тест.

За по-детайлна настройка, кликнете Options за descriptive statistics, квартили, и настройка на липсващи данни.

One-Sample Kolmogorov-Smirnov Тест Опции



Фигура 24-8 *Диалогов прозорец с опции за One-Sample K-S Options*

Statistics. Можете да изберете една или и двете сумарни статистики.
- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил.

Missing Values. Контролира работата с липсващи стойности.
- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Допълнителни функции (One-Sample Kolmogorov-Smirnov Тест)

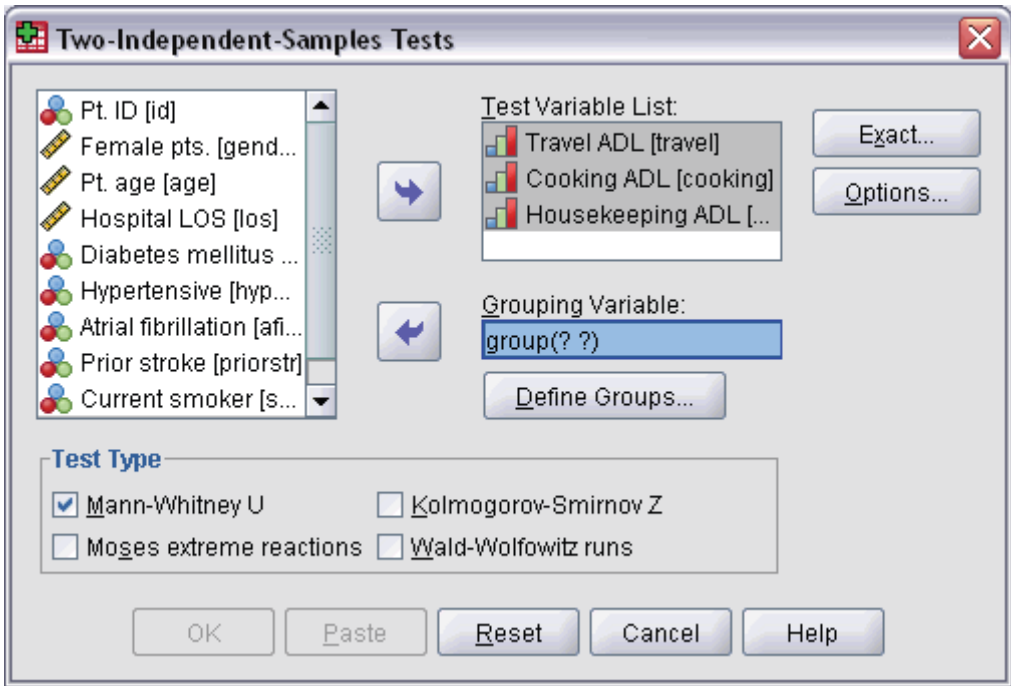
Конадния синтаксис също Ви позволява да определите параметрите на test distribution (Със субкоманда K-S).
Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Тестове за две независими извадки

Two-Independent-Samples Тест сравнява две групи от случаи по една променлива.
Пример. Разработени са нови дентални скоби, които са предназначени да бъдат по-удобни, да изглеждат по-добре и да предоставят по-бърз напредък в пренареждането на зъбите. За да разберете дали новите скоби трябва да се носят толкова дълго, колкото старите скоби, 10 деца са случайно избрани да носят старите скоби, както и още 10 деца са избрали да носят нови скоби. От Mann-Whitney U тест, може да откриете, че средно, деца с нови скоби не трябва да носят скоби колкото децата със старите скоби.
Статистика. Средна стойност, стандартно отклонение, минимум, максимум, брой на не липсващи случаи и квартили. Тества: Mann-Whitney *U*, Moses extreme reactions, Kolmogorov-Smirnov *Z*, Wald-Wolfowitz runs.
Данни. Използвайте числови променливи които могат да бъдат подредени.
Предположения. Използвайте случайни и независими извадки. Mann-Whitney *U* тест изисква двете извадки да са сходни по форма.

За да стартирате Two-Independent-Samples Тест
От менюто изберете:

- Analyze
  Nonparametric Tests
    2 Independent Samples...



Фигура 24-9 *Диалогов прозорец на Two-Independent-Samples Тест*

Изберете една или повече числови променливи.
Изберете групираща променлива и натиснете Define Groups (определи групи) за да разделите на две групи или извадки.

Типове Two-Independent-Samples Тест

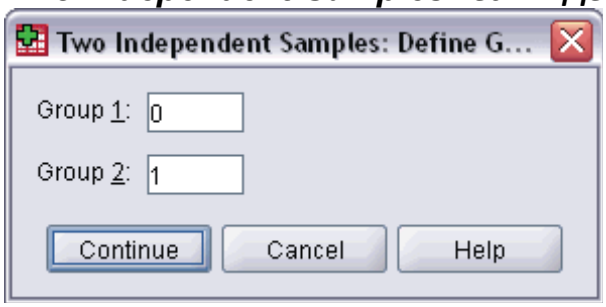
Типове тестове. Четири тестове са на разположение за проверка дали две независими извадки (групи) произлизат от една и съща популация.

Mann-Whitney U тест е най-популярният за проверка на две независими извадки. Еквивалентен е на тестовете на Wilcoxon и Kruskal-Wallis. Mann-Whitney проверява дали две извадки от популации да идентични по локализация. Наблюденията и от двете групи са обединени и подредени, като средният ранг възложени в случай на връзки. броя на връзките трябва да бъдат малки в сравнение с общия брой на наблюдения. Ако популациите са идентични по местоположение то ранговете трябва да са случайно разпределени между двете извадки. Теста изчислява колко пъти резултата на група 1 е по-голям от този на група 2 и обратното. Сметнато по Mann-Whitney *U* се получава по-малкото от двете числа. Wilcoxon rank sum *W* , също показва, по-малката от двете рангови суми. Ако и двете извадки имат еднакъв брой наблюдения, *W* е ранговата сума на дефинираната като първа група в диалоговия прозорец Two-Independent-Samples Define Groups.

Kolmogorov-Smirnov Z тест и **Wald-Wolfowitz runs тест** засичат разликите в локалозацията и разпределението more general. Kolmogorov-Smirnov е базиран на абсолютната разлика между наблюдаваните кумулативни разпределителни функции и на двете извадки. Когато тази разлика е значително голяма, двете разпределения се считат

за различни. Wald-Wolfowitz комбинита и рангира наблюденията и от двете групи. Ако двете извадки са от една популация то двете групи би трябвало да рандомизирани по ранговата скала.
Moses extreme reactions тест предполага, че експерименталната променлива ще влияе на някои индивиди по един начин, а на други по друг. Теста проверява за екстремни резултати сравнени с контролна група. Този тест се съсредоточава върху размаха на контролната група като измерител на колко екстремните стойности в експерименталната група повлияват на размаха сравнен с размаха на контролната група. Контролната група се дефинира като първа в диалоговия прозорец Two-Independent-Samples Define Groups. Наблюденията и от двете групи се комбинират и ранжират.

Two-Independent-Samples Тест – Дефиниране на групи.



Фигура 24-10 Диалогов прозорец за дефиниране на групи при *Two-Independent-Samples*

Statistics. Можете да изберете една или и двете сумарни статистики.
- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил.

Missing Values. Контролира работата с липсващи стойности.
- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Command Additional Features (Two-Independent-Samples Tests)

Командния синтаксис също Ви позволява да определите боря на случаите които да бъдат съкратени при Moses теста (Със субкоманда MOSES).
Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Тестове за две свързани извадки

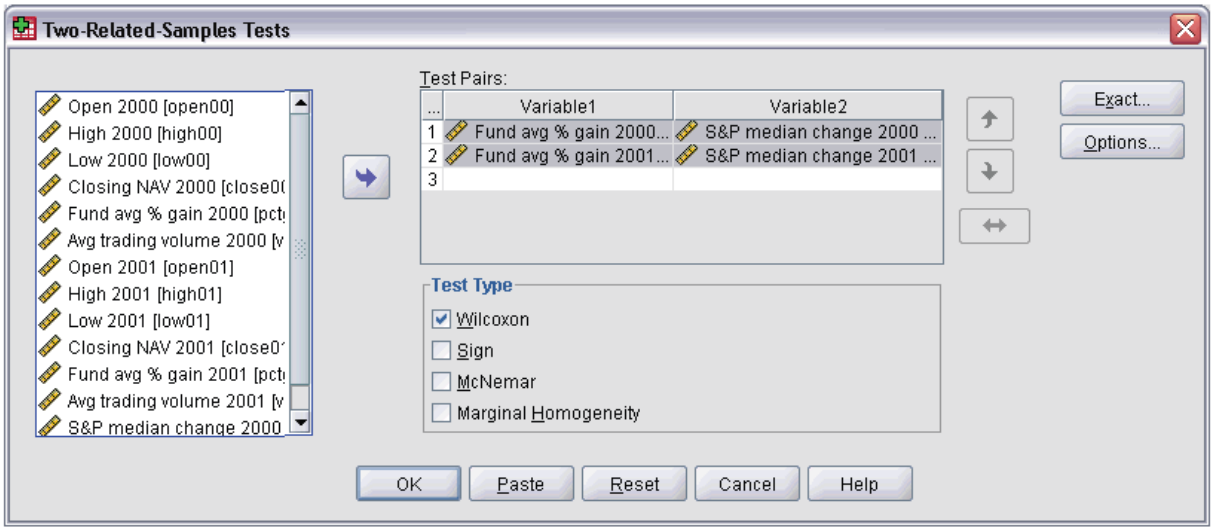
Two-Related-Samples Тест, сравнява разпределението на две променливи.
Пример. Като цяло, семействата получават ли цената която искат, когато продават домовете си? Чрез прилагането на Wilcoxon signed-rank тест за данни, за 10 къщи, можете да научите, че седем семейства получават по-малко от исканата цена, едно семейство получава повече пари, както и две семейства получават желаната сума.
Статистика. Средна стойност, средна стойност, минимум, максимум, брой на не липсващи случаи, и квартили. Тестове: Wilcoxon signed-rank, sign, McNemar. Ако е инсталирана и опцията Exact Tests,
(достъпна само за операционните системи Windows), marginal homogeneity тест също е достъпен.
Данни. Използвайте числови променливи които могат да бъдат подредени.

Предположения. Въпреки че не са въведени определени разпределения за двете променливи, разпределението на двойката разлики в популацията се счита за симетрично.

За да стартирате Two-Related-Samples Тест

От менюто изберете:

- Analyze
  - Nonparametric Tests
    - 2 Related Samples...



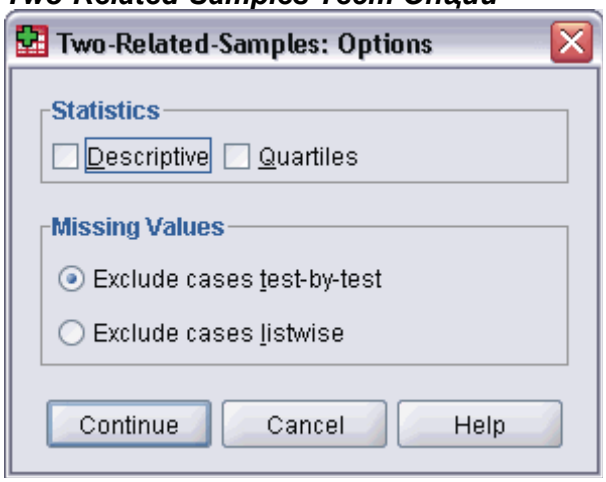
Фигура 24-12 *Диалогов прозорец на Two-Related-Samples Тест*

Изберете една или повече двойки променливи.

Two-Related-Samples Test Types

Тестовете в тази секция сравняват разпределението на две свързани променливи. Използването на различните тестове зависи от типа на данните. Ако данните са последователни, използвайте sign test или Wilcoxon signed-rank test. **Sign test** изчислява разликите между две променливи за всички случаи и класифицира разликите като позитивни, негативни или неутрални. Ако двете порменливи са разпределени сходно, боря на позитивните и негативните разлики няма да се различават значително. **Wilcoxon signed-rank test** considers включва информация за знака на разликата и магнитута на разликите между двойките. Понеже Wilcoxon signed-rank теста включва повече информация относно данните, е по-прецизен от sign test. Ако данните са двоични, използвайте **McNemar тест**. Този тест се използва обикновено в повтарящи се измервания, в които реакцията на всеки обект се отчита два пъти, веднъж преди и втори след определена ситуация. Тестът определя дали първоначалното отчитане (преди ситуацията) е равно на вторичното отчитане (след събитието). Този тест е полезен за засичне на промени в реакциите след експериментална намеса в before-and-after designs. Ако данните са категориини използвайте **marginal homogeneity тест.** Този тест е разширение на теста на McNemar за преминаване ор двоични отзиви към multinomial отзиви. Проверява промените в отзивите (използвайте разпределяне по хи-квадрат) и е полезен за отчитане на промени в отговорите предизвикани от експериментални намеси в before-and-after дизайна. Теста marginal homogeneity е достъпен само след инсталирането на Exact Tests.

Two-Related-Samples Тест Опции



Фигура 24-13 *Диалогов прозорец с Опции за Two-Related-Samples*

Statistics. Можете да изберете една или и двете сумарни статистики.

- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Квартили.** Показва стойностите на 25тия, 50тия, и 75тия персентил

Missing Values. Контролира работата с липсващи стойности.

- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Command Additional Features (Two Related Samples)

Командния синтаксис също ви позволява да тествате дадена променлива със всяка променлива от списъка.

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Тестове за няколко независими извадки

Теста за Several Independent Samples сравнява две или повече групи от случаи по една променлива.

Пример. Дали три марки 100-ватови крушки се различават по живот? С Kruskal-Wallis one-way analysis of variance, може и да научите, че се различават.

Статистика. Средна стойност, standard deviation, минимум, максимум, брой на не липсващи случаи и квартили . Tests: Kruskal-Wallis *H*, median.

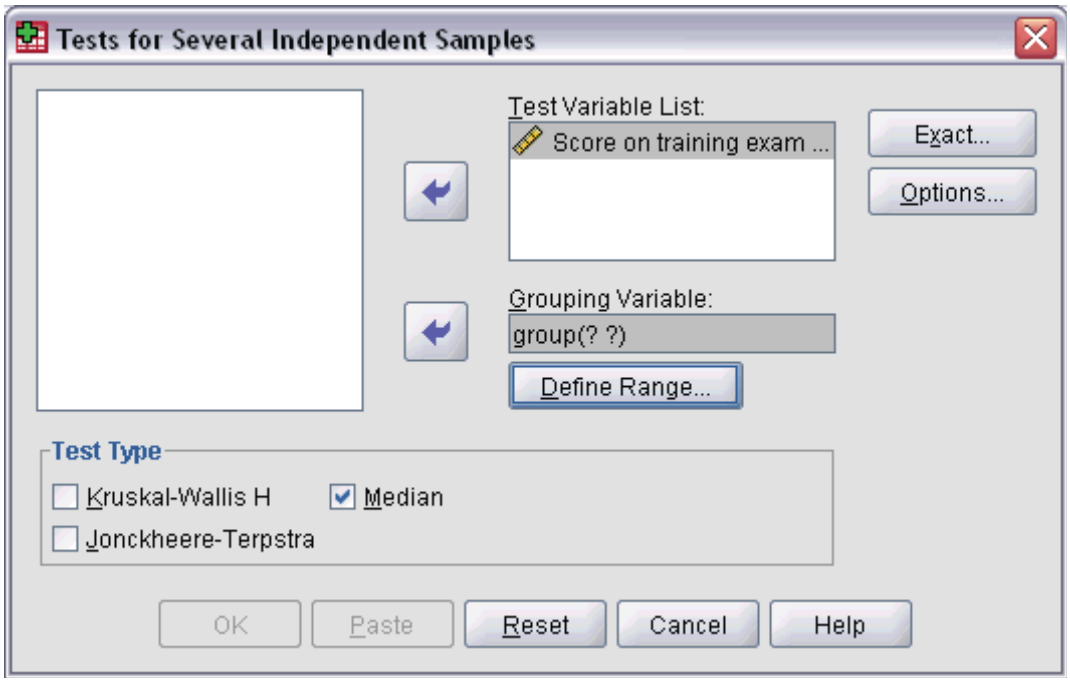
Данни. Използвайте числови променливи които могат да бъдат подредени.

Предположиния. Използвайте независими случайни проби. Kruskal-Wallis *H* тест изисква тестовите проби да са сходни по форма.

За да стартирате тестове за Several Independent Samples

От менюто изберете:

- Analyze
  - Nonparametric Tests
    - K Independent Samples...



Фигура 24-14 *Defining the median тест (тест за дефиниране на медианата)*

Изберете една или повече числови променливи.
Изберете групираща променлива и натиснете Define Range за да определите минимума и максимума на стойността на групиращата променлива.

Tests for Several Independent Samples Test Types

Три теста са наличниза определяне дали няколко независими извадки произлизат от една популация. The Kruskal-Wallis *H* тест, тест на медианата (median test) и Jonckheere-Terpstra тест, всички те проверяват дали няколко независими извадки са от една популация.

Kruskal-Wallis H тест, представлява допълнение към Mann-Whitney *U* тест, и представялва непараметричния аналог на еднозначния анализ на вариациите и засича разлики в разпределението на локациите. Тестът на медианата

(median test), които е по-общ тест (не и толкова категоричен) засича разлики в разпределението и формата. Kruskal-Wallis *H* тест и теста на медианата предполагат че няма *a priori* при задаването *k* популацията от която е извлечена извадката. При наличието на нормално *a priori* подреждане (възходящо или низходящо) на *k* популацията, теста **Jonckheere-Terpstra** е по-подходящ. Например, *k* популацията може да представи *k* повишаваща температура. The hypothesis that different temperatures produce the same response

distribution is tested against the alternative that as the temperature increases, the magnitude of the

response increases. Here, the alternative hypothesis is ordered; therefore, Jonckheere-Terpstra is the most appropriate test to use. Jonckheere-Terpstra тест е достъпен само ако сте инсталирали допълнителният модул Exact Tests.

Tests for Several Independent Samples Define Range

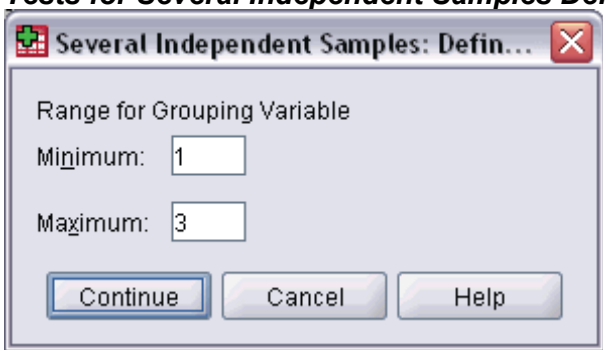
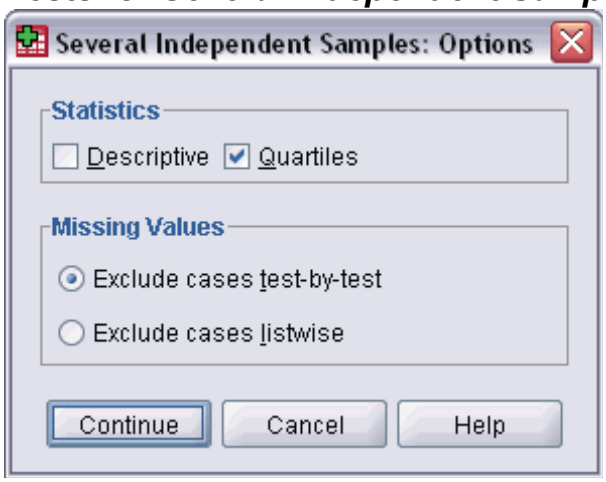


Figure 24-15 *Several Independent Samples Define Range dialog box*

За да дефинирате размаха, въведете стойности за Minimum и Maximum които отговарят на най-ниските и на най-високите стойности на групиращата променлива . Случаи с показатели извън тези граници се изключват. Например, ако определите минимална стойност 1 и аксимална 3, само стойностите равни и между 1 и 3 ще се използват. Минимума трябва да е по-малак от максимума и двете величини трябва да се дефинират задължително.

Tests for Several Independent Samples Опции



Фигура 24-16 *Диалогов прозорец с опции за Several Independent Samples*

Statistics. Можете да изберете една или и двете сумарни статистики.
- **Descriptive.** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Показва стойностите на 25тия, 50тия, и 75тия персентил

Missing Values. Контролира работата с липсващи стойности.
- **Exclude cases test-by-test.** Когато няколко теста са определени, всеки тест се проверява отделно за липсващи стойности.
- **Exclude cases listwise.** Случаи с липсващи стойности за която и да е променлива, са изключени от всички анализи.

NPAR TESTS Command Additional Features (K Independent Samples)

Комадния синтаксис също ви позволява да определите стойност различна от медианата за теста на медианата - median test (Със субкоманда MEDIAN - медиана).
Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Тестове за нясолко свързани извадки

Теста за Several Related Samples сравнява разпределението на две или повече променливи.

Пример. Счита ли обществото, че професиите като лекар, адвокат, полицай и учител носят различни нива на престиж? Десет човека са запитани да подредят по престиж тези четири професии.

Теста на Friedman показва, че наистина обществото асоциира различни нива на престиж за четрите професии.

Статистика. Средна стойност, стандартно отклонение, минимум, максимум, брой на не липсващи случаи и квартили. Тестове: Friedman, Kendall's *W*, and Cochran's *Q*.

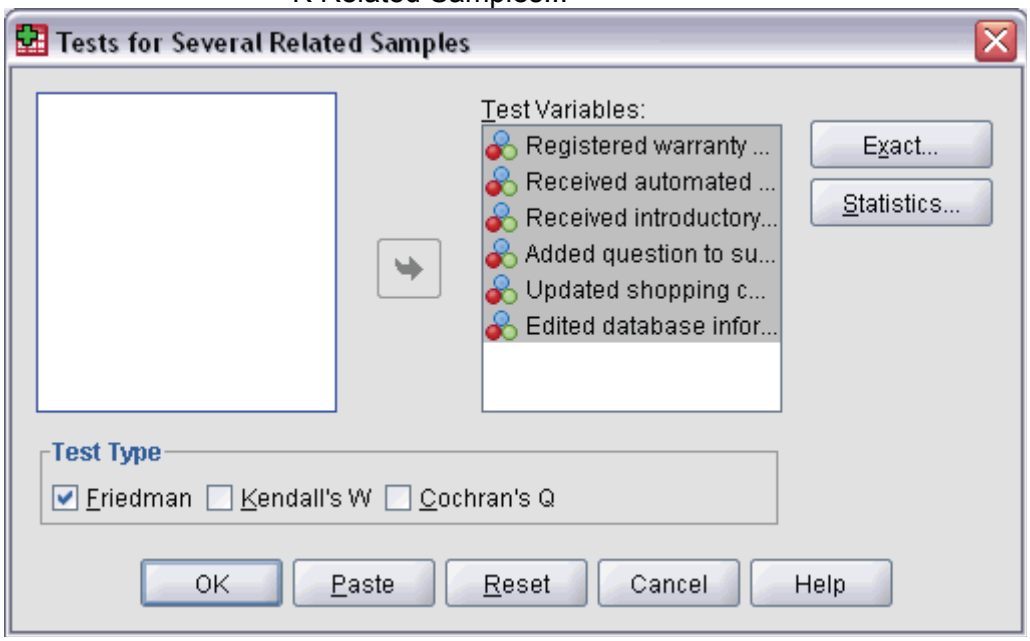
Данни. Използвайте числови променливи които могат да бъдат подредени

Предположения. Непараметричните тестове не изискват предположения за формата на първоначалното разпределение. Използвайте зависими случайни проби.

За да стартирате тестове за Several Related Samples

От менюто изберете:

- Analyze
  - Nonparametric Tests
    - K Related Samples...



Фигура 24-17 *Selecting Cochran as the test type*

Изберете две или повече числови тестови променливи.

Tests for Several Related Samples Test Types

Три теста са налични за сравняване на разпределението на няколко свързани променливи.

Теста на **Friedman** е непараметричния аналог на едноизвадковия тест с повторни измервания или на двуфакторен дисперсионен анализ с едно измерване в клетка.

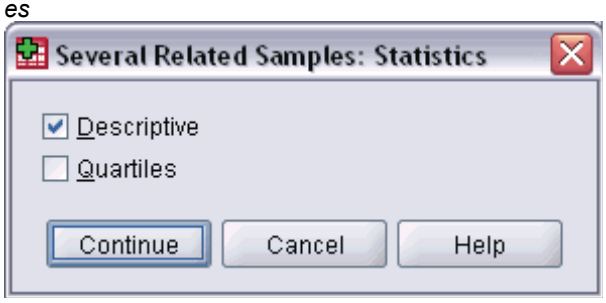
Тестът на Friedman проверява нулевата хипотеза, че *k* свързани променливи принадлежат към едзна и съща популация. Всичките *k* променливи са подредени от 1 до *k*. Статистиката на теста се базира на тази подредба.

Kendall's W е производна на статистиката на Friedman. Kendall's *W* се интерпретира като коефициент на съгласуваност, който е мярка за съгласие между оценителите. Всеки случай е съдя или оценител, и всяка променлива е предмет или човек който бива

оценяван. За всяка променлива сумата от оценките се изчислява. Kendall's *W* влиза в границите между 0 (без съгласие) и 1 (пълно съгласие).
Cochran's Q е идентичен с теста на Friedman, но се използва когато всички отговори са двойни. Този тест е надграждане на теста на McNemar за ситуацията на *k свързани извадки*. Cochran's *Q* проверява хипотезата, че няколко свързани дихотомни променливи имат еднаква средна стойност. Променливите се измерват върху един индивид или съвпадащи индивиди.

Tests for Several Related Samples Statistics

es



Фигура 24-18 Диалогов прозорец със статистики за *Several Related Sampl*

Можете да изберете.

- **Descriptive. .** Показва средната стойност, стандартно отклонение, минимум, максимум и броя на не липсващите случаи.
- **Quartiles.** Displays Показва стойностите на 25тия, 50тия, и 75тия персентил

NPAR TESTS Command Additional Features (K Related Samples)

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Глава 25. Анализ на надежността

Reliability Analysis

Reliability Analysis изчислява броя на често използваните измервания за скала на надеждността и също така предоставя информация за взаимоотношенията между различните елементи в скалата. Intraclass корекциите на коефициентите могат да бъдат използвани за да се изчисли inter-rater reliability estimates.

Пример. Дали въпросника ми измерва по полезен начин удовлетвореността на клиентите? С използването на анализ на надеждността, можете да определите степента с която елементите във вашия въпросник са свързани помежду си, можете да получите общия индекс на повторяемостта или вътрешната скала на последователността като цяло, и можете да определите проблемните елементи които трябва да бъдат премахнати от скалата.

Статистики. Descriptives за всяка променлива и скала, обобщена статистика на всички елементи, inter-item correlations and covariances, оценка на надеждността, ANOVA таблица, intraclass корекция на коефициента, Hotelling's *T*2, и Tukey's test of additivity.

Модели. Следните модели за надеждност са налични:

- **Alpha (Cronbach).** Този модел е модел на вътрешната съгласуваност, въз основа на средната интер-точка на съответствието.
- **Split-half.** Този модел разделя скалата на две части и разглежда връзките между тях.
- **Guttman.** Този модел изчислява ниските граници на надеждността на Guttman.
- **Parallel.** Този модел предполага, че всички обекти имат равни дисперсии, равни дисперсии на грешките в репликациите.
- **Strict parallel.** Този модел се припокрива с Parallel като добавя предположението, че средните стойности на всички елементи са равни.

Данни. Данните могат да бъдат дихотомни, ординарни или интервали, но всички трябва да бъдат числово кодирани.

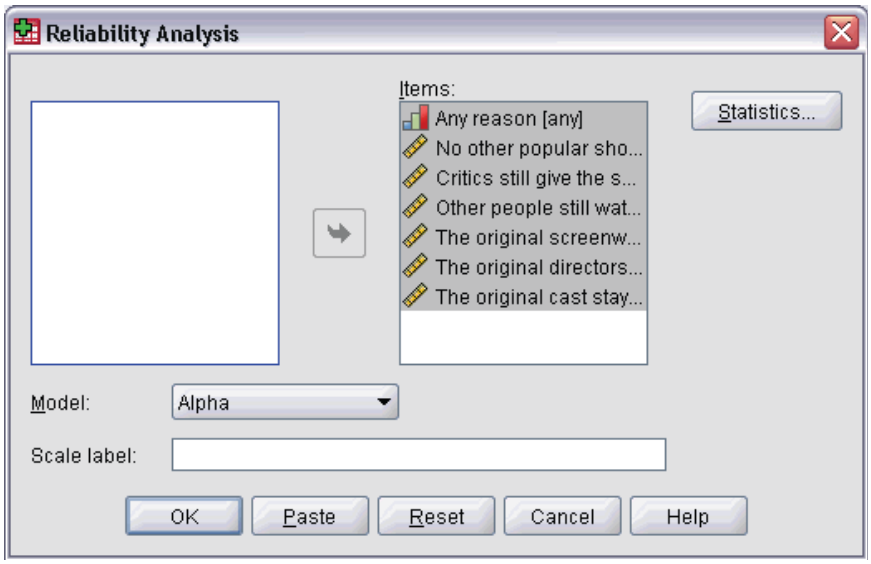
Предположения. Наблюденията трябва да бъдат независими и грешките трябва да бъдат uncorrelated между елементите. Всяка двойка елементи трябва да бъдат bivariate нормално разпределени. Скалите трябва да бъдат така построени, че всеки елемент да бъде линейно свързан с общата оценка.

Свързани процедури. Ако желаете да разгледате dimensionality на елементитте на скалата си (за да видите дали повече от един конструкт е необходим да се вземе на предвид в условията за оценяване на елементите), използвайте factor analysis или multidimensional scaling. За да определите хомогенните групи от променливи, използвайте йерархичен клъстер анализ за групиране на променливите.

За да стартирате Reliability Analysis

От менюто изберете:

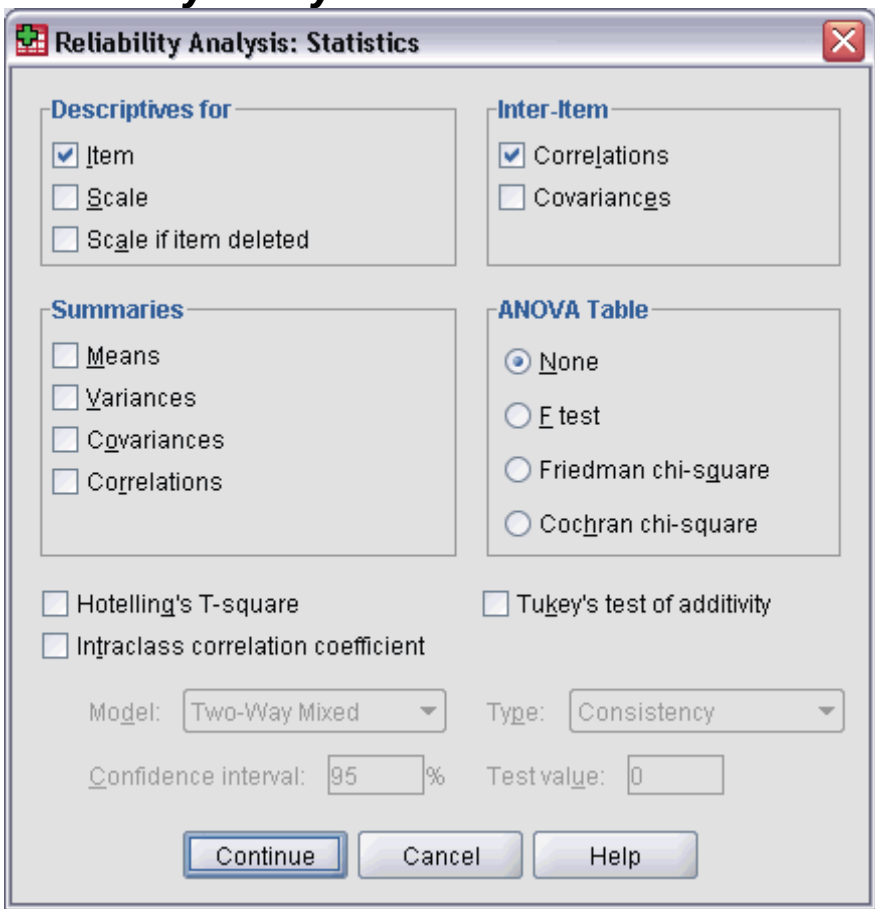
- Analyze
  - Scale
    - Reliability Analysis...



Фигура 25-1 Диалогов прозорец на *Reliability Analysis*

Изберете две или повече променливи като потенциални компоненти на additive скала.
Изберете модел от падащото меню

Reliability Analysis Statistics



Фигура 25-2 Диалогов прозорец със статистики за *Reliability Analysis*

Можете да изберете различни статистически данни, които описват Вашия мащаб и позиции. Статистически данни, които се отчитат от по подразбиране включва броя на делата, броят на предметите, оценките и надеждност, както следва:

- **Alpha models.** Коефициент alpha; за дихотомни данни, равно е на коефициента Kuder-Richardson 20 (KR20).
- **Split-half models.** Корелация между Guttman split-half надеждност, Spearman-Brown надеждност (за равни и неравни дължини) и коефициент alpha за всяка половина.
- **Guttman models.** Коефициенти на надеждността от lambda 1 до lambda 6.
- **Parallel and Strict parallel models.** Тест за goodness of fit of модела; определя вероятността от грешки, общо противоречие и същински вариации; определя общата inter-item корелация; оценява надеждността и дава непредубедената оценка на надеждността.

Descriptives for. Представя описателна статистика за скалите или елементи в скалите.

- **Item.** Представя описателна статистика за продуктите във всички случаи.
- **Scale.** Представя описателна статистика за скалите.
- **Scale if item deleted.** Показва обобщена статистика сравнявайки всяка точка от скалата която е съставена от други елементи. Статистиката включва средна стойност и вероятност елемента да бъде премахнат от скалата, корелация между елемент и скала изградена от други елементи и Cronbach's alpha ако елемента ще бъде изтрит от скалата.

Summaries. Показва описателна статистика за разпределението на елементите във всички точки в скалата.

- **Means.** Сумарна статистика за средната стойност на елементите. Най-малката, най-голямата и средната стойност на елементите, съотношението между най-високата и най-ниската стойност и обхвата и отклонението на средната на елементите се показват.
- **Variances.** Сумарна статистика за разликите в елементите. Най-голямото, най-малкото и средното отклонение на елементите, обхвата и отклонението в разликите на елементите и съотношението между най-голямото и най-малкото отклонение в елементите се показват.
- **Covariances.** Сумарна статистика за inter-item covariances. Най-малката, най-голямата и средната inter-item covariances, обхвата и отклонението в inter-item covariances, и съотношението между най-голямата и най-малката inter-item covariances се показват.
- **Correlations.** Сумарна статистика за inter-item корелации. Най-малката, най-голямата и средна inter-item корелация, обхвата и варирането на inter-item корелациите, и съотношението между най-голямата и най-малката inter-item корелация се показват.

Inter-Item. Показва матрица или covariances на корелацията между елементите.

ANOVA Table. Показва тестове за равни средни стойности.

- **F test.** Displays a repeated measures analysis-of-variance table.
- **Friedman хи-квадрат.** Показва хи-квадрата на Friedman и коефициента на съгласуване на Kendall. Тази опция е подходяща за рангирани данни. Хи-квадрат теста заменя F в таблицата ANOVA.
- **Cochran хи-квадрат.** Показва Cochran's Q. Тази опция е подходяща при дихотомни данни. Q статистиката заменя обикновената F статистика в таблицата ANOVA.

Hotelling's T-square. Прави многовариантни тестове на нулева хипотеза, че всички елементи на скалата имат еднаква средна стойност.

Tukey's test of additivity. Провежда тест за да провери предположението, че няма мултиплицирано взаимодействие между елементите.

Intraclass correlation coefficient. Провежда измервания за съгласуваност или съгласие на стойностите в случаите.

- **Model.** Изберете модела за изчисляване на корелационния коефициент. Наличните модели са Two-Way Mixed, Two-Way Random, and One-Way Random. Select Two-Way Mixed,

Когато ефектите от хората са случайни, а ефектите от елементите фиксирани използвайте Two-Way Mixed, а когато и ефектите от хората и елементите са случайни изберете Two-Way Random или изберете One-Way Random когато ефектето от хора са случайни.

- **Type.** Изберете тип на индекса. Възможностите са съгласуваност или пълно съгласие.
- **Confidence interval.** Определете нивото на интервала на доверие. По подразбиране е 95%.
- **Test value.** Посочете хипотетичната стойност за коефициента на теста за хипотези. Тази стойност е стойността с която наблюдаваната стойност се сравнява. По подразбиране е 0.

RELIABILITY Command Допълнителни функции

Командния синтаксис също ви позволява да:

- Четете и анализирате корелационна матрица.
- Напишете корелационна матрица за по-късен анализ.
- Определите разделящи точки различно от равните половини предоставени от метода split-half.

Виж *Command Syntax Reference* за пълна информация относно синтаксиса.

Глава 26. Анализ на въпроси с повече от един отговор

Multiple Response Analysis

Two procedures are available for analyzing multiple dichotomy and multiple category sets. The Multiple Response Frequencies procedure displays frequency tables. The Multiple Response Crosstabs procedure displays two- and three-dimensional crosstabulations. Before using either procedure, you must define multiple response sets.

Example. This example illustrates the use of multiple response items in a market research survey. The data are fictitious and should not be interpreted as real. An airline might survey passengers flying a particular route to evaluate competing carriers. In this example, American Airlines wants to know about its passengers' use of other airlines on the Chicago-New York route and the relative importance of schedule and service in selecting an airline. The flight attendant hands each passenger a brief questionnaire upon boarding. The first question reads: Circle all airlines you have flown at least once in the last six months on this route—American, United, TWA, USAir, Other. This is a multiple response question, since the passenger can circle more than one response. However, this question cannot be coded directly because a variable can have only one value for each case. You must use several variables to map responses to each question. There are two ways to do this. One is to define a variable corresponding to each of the choices (for example, American, United, TWA, USAir, and Other). If the passenger circles United, the variable *united* is assigned a code of 1, otherwise 0. This is a **multiple dichotomy method** of mapping variables. The other way to map responses is the **multiple category method**, in which you estimate the maximum number of possible responses to the question and set up the same number of variables, with codes used to specify the airline flown. By perusing a sample of questionnaires, you might discover that no user has flown more than three different airlines on this route in the last six months. Further, you find that due to the deregulation of airlines, 10 other airlines are named in the Other category. Using the multiple response method, you would define three variables, each coded as 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta*, and so on. If a given passenger circles American and TWA, the first variable has a code of 1, the second has a code of 3, and the third has a missing-value code. Another passenger might have circled American and entered Delta. Thus, the first variable has a code of 1, the second has a code of 5, and the third a missing-value code. If you use the multiple dichotomy method, on the other hand, you end up with 14 separate variables. Although either method of mapping is feasible for this survey, the method you choose depends on the distribution of responses.

Multiple Response Define Sets

The Define Multiple Response Sets procedure groups elementary variables into multiple dichotomy and multiple category sets, for which you can obtain frequency tables and crosstabulations. You can define up to 20 multiple response sets. Each set must have a unique name. To remove a set, highlight it on the list of multiple response sets and click Remove. To change a set, highlight it on the list, modify any set definition characteristics, and click Change.

You can code your elementary variables as dichotomies or categories. To use dichotomous variables, select Dichotomies to create a multiple dichotomy set. Enter an integer value for Counted value. Each variable having at least one occurrence of the counted value becomes a category of the multiple dichotomy set. Select Categories to create a multiple category set having the same range of values as the component variables. Enter integer values for the minimum and maximum values of the range for categories of the multiple category set. The procedure totals each distinct integer value in the inclusive range across all component variables. Empty categories are not tabulated.

Each multiple response set must be assigned a unique name of up to seven characters. The procedure prefixes a dollar sign ($) to the name you assign. You cannot use the following reserved names: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length*, and *width*. The name of the multiple response set exists only for use in multiple response procedures. You cannot refer to multiple response set names in other procedures. Optionally, you can enter a descriptive variable label for the multiple response set. The label can be up to 40 characters long.

To Define Multiple Response Sets

E From the menus choose:
Analyze
Multiple Response
Define Sets...

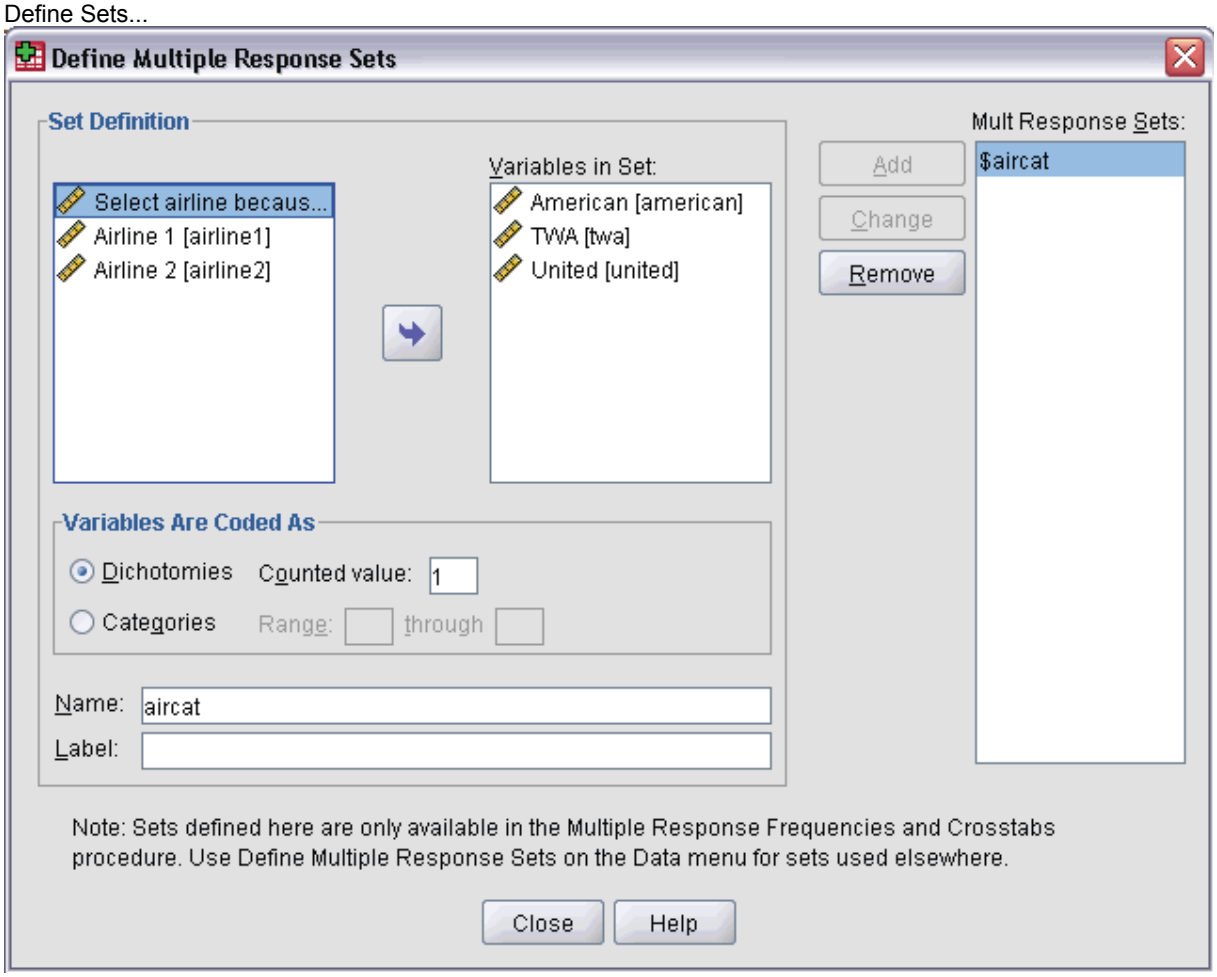


Figure 26 -1 *Define Multiple Response Sets dialog box*

E Select two or more variables.
E If your variables are coded as dichotomies, indicate which value you want to have counted. If your variables are coded as categories, define the range of the categories.
E Enter a unique name for each multiple response set.
E Click Add to add the multiple response set to the list of defined sets.

Multiple Response Frequencies

The Multiple Response Frequencies procedure produces frequency tables for multiple response sets. You must first define one or more multiple response sets (see "Multiple Response Define Sets").

For multiple dichotomy sets, category names shown in the output come from variable labels defined for elementary variables in the group. If the variable labels are not defined, variable names are used as labels. For multiple category sets, category labels come from the value labels of the first variable in the group. If categories missing for the first variable are present for other variables in the group, define a value label for the missing categories.

Missing Values. Cases with missing values are excluded on a table-by-table basis. Alternatively, you can choose one or both of the following:

- **Exclude cases listwise within dichotomies.** Excludes cases with missing values for any variable from the tabulation of the multiple dichotomy set. This applies only to multiple response sets defined as dichotomy sets. By default, a case is considered missing for a multiple dichotomy set if none of its component variables contains the counted value. Cases with missing values for some (but not all variables) are included in the tabulations of the group if at least one

variable contains the counted value.

◷ **Exclude cases listwise within categories.** Excludes cases with missing values for any variable from tabulation of the multiple category set. This applies only to multiple response sets defined as category sets. By default, a case is considered missing for a multiple category set only if none of its components has valid values within the defined range.

Example. Each variable created from a survey question is an elementary variable. To analyze a multiple response item, you must combine the variables into one of two types of multiple response sets: a multiple dichotomy set or a multiple category set. For example, if an airline survey asked which of three airlines (American, United, TWA) you have flown in the last six months and you used dichotomous variables and defined a **multiple dichotomy set**, each of the three variables in the set would become a category of the group variable. The counts and percentages for the three airlines are displayed in one frequency table. If you discover that no respondent mentioned more than two airlines, you could create two variables, each having three codes, one for each airline. If you define a **multiple category set**, the values are tabulated by adding the same codes in the elementary variables together. The resulting set of values is the same as those for each of the elementary variables. For example, 30 responses for United are the sum of the five United responses for airline 1 and the 25 United responses for airline 2. The counts and percentages for the three airlines are displayed in one frequency table.

Statistics. Frequency tables displaying counts, percentages of responses, percentages of cases, number of valid cases, and number of missing cases.

Data. Use multiple response sets.

Assumptions. The counts and percentages provide a useful description for data from any distribution.

Related procedures. The Multiple Response Define Sets procedure allows you to define multiple response sets.

To Obtain Multiple Response Frequencies

E From the menus choose:

Analyze
Multiple Response
Frequencies...

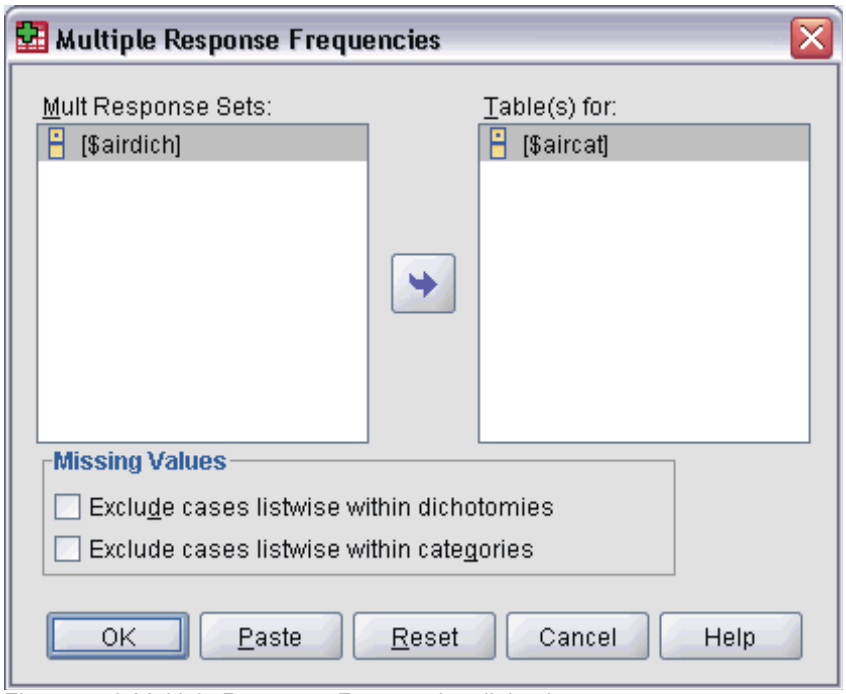


Figure 26-2 *Multiple Response Frequencies dialog box*

E Select one or more multiple response sets.

Multiple Response Crosstabs

The Multiple Response Crosstabs procedure crosstabulates defined multiple response sets, elementary variables, or a combination. You can also obtain cell percentages based on cases or responses, modify the handling of missing values, or get paired crosstabulations. You must first define one or more multiple response sets (see "To Define Multiple Response Sets").

For multiple dichotomy sets, category names shown in the output come from variable labels defined for elementary variables in the group. If the variable labels are not defined, variable names are used as labels. For multiple category sets, category labels come from the value labels of the first variable in the group. If categories missing for the first variable are present for other variables in the group, define a value label for the missing categories. The procedure displays category labels for columns on three lines, with up to eight characters per line. To avoid splitting words, you can reverse row and column items or redefine labels.

Example. Both multiple dichotomy and multiple category sets can be crosstabulated with other variables in this procedure. An airline passenger survey asks passengers for the following information: Circle all of the following airlines you have flown at least once in the last six months (American, United, TWA). Which is more important in selecting a flight—schedule or service? Select only one. After entering the data as dichotomies or multiple categories and combining them into a set, you can crosstabulate the airline choices with the question involving service or schedule.

Statistics. Crosstabulation with cell, row, column, and total counts, and cell, row, column, and total percentages. The cell percentages can be based on cases or responses.

Data. Use multiple response sets or numeric categorical variables.

Assumptions. The counts and percentages provide a useful description of data from any distribution.

Related procedures. The Multiple Response Define Sets procedure allows you to define multiple response sets.

To Obtain Multiple Response Crosstabs

E From the menus choose:

Analyze
Multiple Response
Crosstabs...

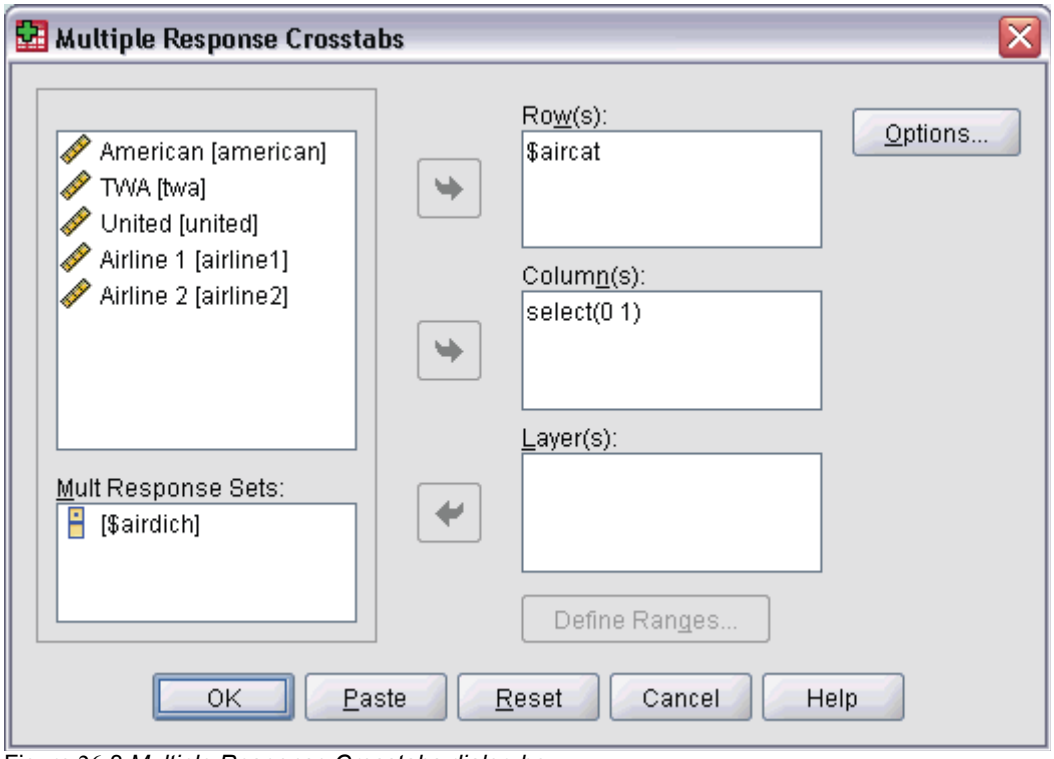


Figure 26-3 *Multiple Response Crosstabs dialog box*

E Select one or more numeric variables or multiple response sets for each dimension of the crosstabulation.
E Define the range of each elementary variable.
Optionally, you can obtain a two-way crosstabulation for each category of a control variable or multiple response set. Select one or more items for the Layer(s) list.

Multiple Response Crosstabs Define Ranges

x

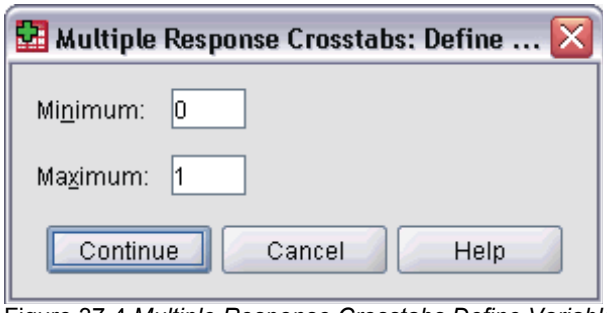


Figure 37-4 *Multiple Response Crosstabs Define Variable Range dialog bo*

Value ranges must be defined for any elementary variable in the crosstabulation. Enter the integer minimum and maximum category values that you want to tabulate. Categories outside the range are excluded from analysis. Values within the inclusive range are assumed to be integers (non-integers are truncated).

Multiple Response Crosstabs Options

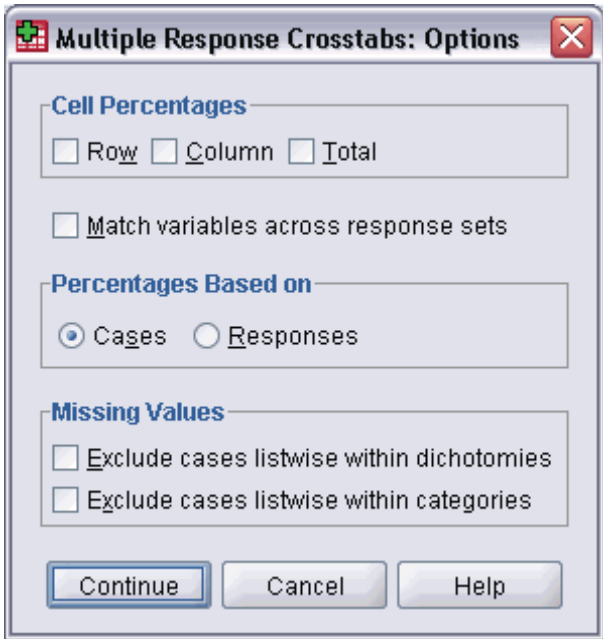


Figure 37-5 *Multiple Response Crosstabs Options dialog box*

Cell Percentages. Cell counts are always displayed. You can choose to display row percentages, column percentages, and two-way table (total) percentages.
Percentages Based on. You can base cell percentages on cases (or respondents). This is not available if you select matching of variables across multiple category sets. You can also base cell percentages on responses. For multiple dichotomy sets, the number of responses is equal to the number of counted values across cases. For multiple category sets, the number of responses is the number of values in the defined range.
Missing Values. You can choose one or both of the following:

- **Exclude cases listwise within dichotomies.** Excludes cases with missing values for any variable from the tabulation of the multiple dichotomy set. This applies only to multiple response sets defined as dichotomy sets. By default, a case is considered missing for a multiple dichotomy set if none of its component variables contains the counted value. Cases with missing values for some, but not all, variables are included in the tabulations of the group if at least one variable contains the counted value.
- **Exclude cases listwise within categories.** Excludes cases with missing values for any variable from tabulation of the multiple category set. This applies only to multiple response sets defined as category sets. By default, a case is considered missing for a multiple category set only if none of its components has valid values within the defined range.

By default, when crosstabulating two multiple category sets, the procedure tabulates each variable in the first group with each variable in the second group and sums the counts for each cell; therefore, some responses can appear more than once in a table. You can choose the following option:

Match variables across response sets. Pairs the first variable in the first group with the first variable in the second group, and so on. If you select this option, the procedure bases cell percentages on responses rather than respondents. Pairing is not available for multiple dichotomy sets or elementary variables.

MULT RESPONSE Command Additional Features

The command syntax language also allows you to:

- Obtain crosstabulation tables with up to five dimensions (with the `BY` subcommand).
- Change output formatting options, including suppression of value labels (with the `FORMAT` subcommand).

See the *Command Syntax Reference* for complete syntax information.

Глава 27. Как да оформим резултатите

Reporting Results

Case listings and descriptive statistics are basic tools for studying and presenting data. You can obtain case listings with the Data Editor or the Summarize procedure, frequency counts and descriptive statistics with the Frequencies procedure, and subpopulation statistics with the Means procedure. Each of these uses a format designed to make information clear. If you want to display the information in a different format, Report Summaries in Rows and Report Summaries in Columns give you the control you need over data presentation.

Report Summaries in Rows

Report Summaries in Rows produces reports in which different summary statistics are laid out in rows. Case listings are also available, with or without summary statistics.

Example. A company with a chain of retail stores keeps records of employee information, including salary, job tenure, and the store and division in which each employee works. You could generate a report that provides individual employee information (listing) broken down by store and division (break variables), with summary statistics (for example, mean salary) for each store, division, and division within each store.

Data Columns. Lists the report variables for which you want case listings or summary statistics and controls the display format of data columns.

Break Columns. Lists optional break variables that divide the report into groups and controls the summary statistics and display formats of break columns. For multiple break variables, there will be a separate group for each category of each break variable within categories of the preceding break variable in the list. Break variables should be discrete categorical variables that divide cases into a limited number of meaningful categories. Individual values of each break variable appear, sorted, in a separate column to the left of all data columns.

Report. Controls overall report characteristics, including overall summary statistics, display of missing values, page numbering, and titles.

Display cases. Displays the actual values (or value labels) of the data-column variables for every case. This produces a listing report, which can be much longer than a summary report.

Preview. Displays only the first page of the report. This option is useful for previewing the format of your report without processing the whole report.

Data are already sorted. For reports with break variables, the data file must be sorted by break variable values before generating the report. If your data file is already sorted by values of the break variables, you can save processing time by selecting this option. This option is particularly useful after running a preview report.

To Obtain a Summary Report: Summaries in Rows

E From the menus choose:
Analyze
Reports
Report Summaries in Rows...

E Select one or more variables for Data Columns. One column in the report is generated for each variable selected.

E For reports sorted and displayed by subgroups, select one or more variables for Break Columns.

E For reports with summary statistics for subgroups defined by break variables, select the break variable in the Break Column Variables list and click Summary in the Break Columns group to specify the summary measure(s).

E For reports with overall summary statistics, click Summary to specify the summary measure(s).

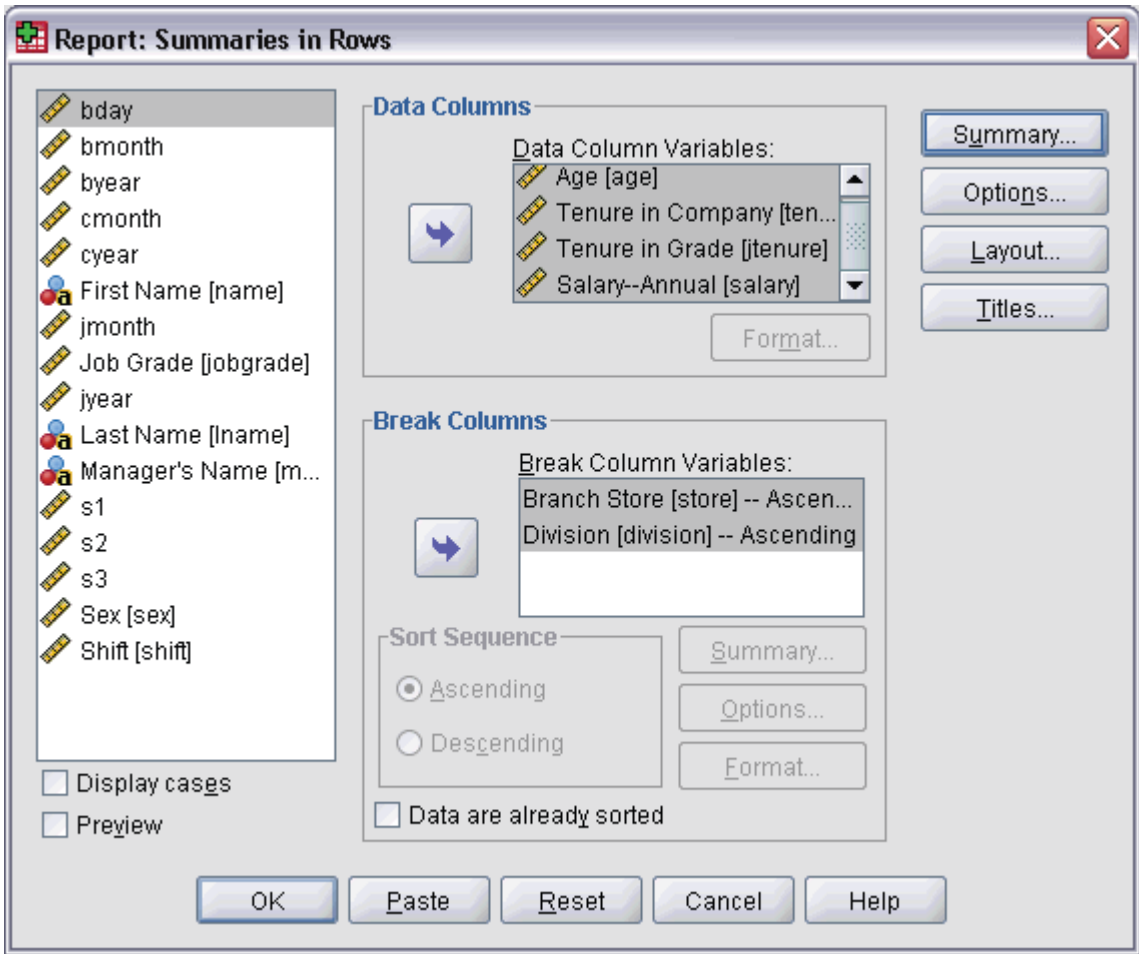


Figure 27-1 *Report Summaries in Rows dialog box*

Report Data Column/Break Format

The Format dialog boxes control column titles, column width, text alignment, and the display of data values or value labels. Data Column Format controls the format of data columns on the right side of the report page. Break Format controls the format of break columns on the left side.

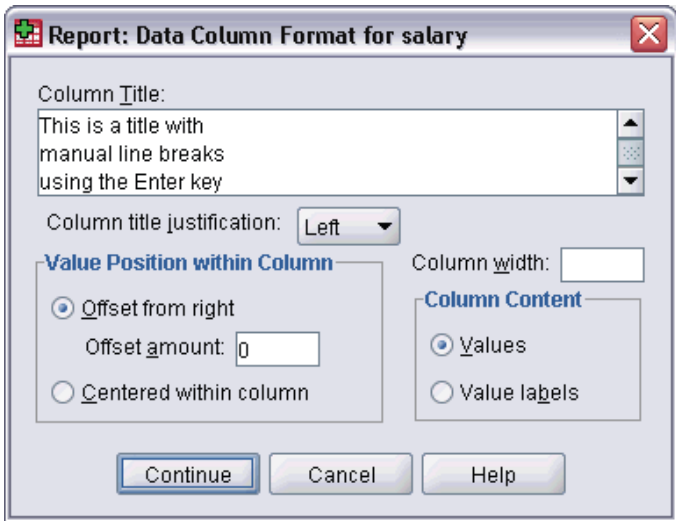


Figure 27-2 *Report Data Column Format dialog box*

Column Title. For the selected variable, controls the column title. Long titles are automatically wrapped within the column. Use the Enter key to manually insert line breaks where you want titles to wrap.

Value Position within Column. For the selected variable, controls the alignment of data values or value labels within the column. Alignment of values or labels does not affect alignment of column headings. You can either indent the column contents by a specified number of characters or center the contents.

Column Content. For the selected variable, controls the display of either data values or defined value labels. Data values are always displayed for any values that do not have defined value labels. (Not available for data columns in column summary reports.)

Report Summary Lines for/Final Summary Lines

The two Summary Lines dialog boxes control the display of summary statistics for break groups and for the entire report. Summary Lines controls subgroup statistics for each category defined by the break variable(s). Final Summary Lines controls overall statistics, displayed at the end of the report.

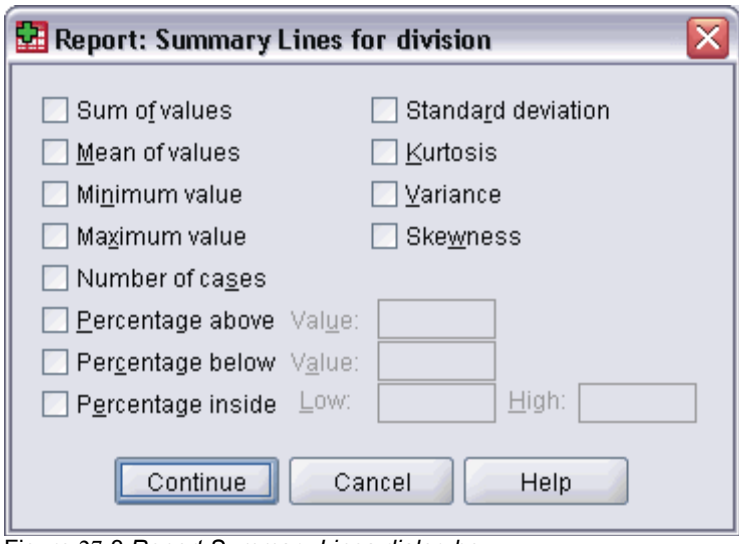


Figure 27-3 *Report Summary Lines dialog box*

Available summary statistics are sum, mean, minimum, maximum, number of cases, percentage of cases above or below a specified value, percentage of cases within a specified range of values, standard deviation, kurtosis, variance, and skewness.

Report Break Options

Break Options controls spacing and pagination of break category information.

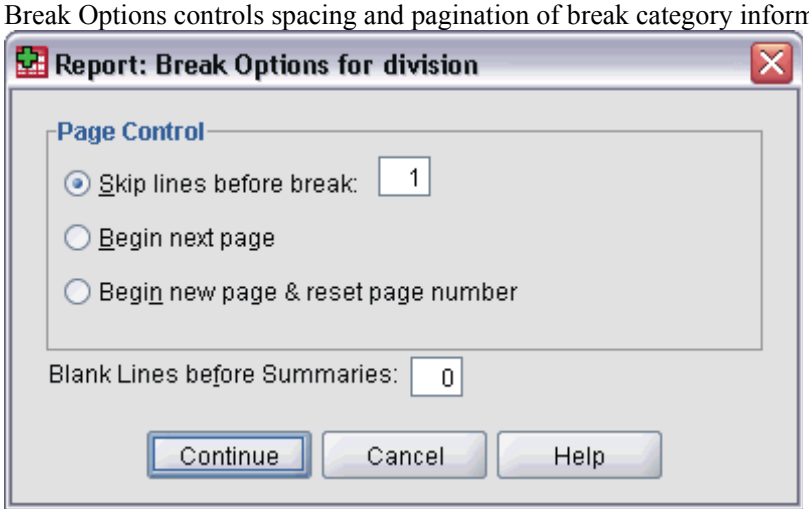


Figure 27-4 *Report Break Options dialog box*

Page Control. Controls spacing and pagination for categories of the selected break variable. You can specify a number of blank lines between break categories or start each break category on a new page.

Blank Lines before Summaries. Controls the number of blank lines between break category labels or data and summary statistics. This is particularly useful for combined reports that include both individual case listings and summary statistics for break categories; in these reports, you can insert space between the case listings and the summary statistics.

Report Options

Report Options controls the treatment and display of missing values and report page numbering.

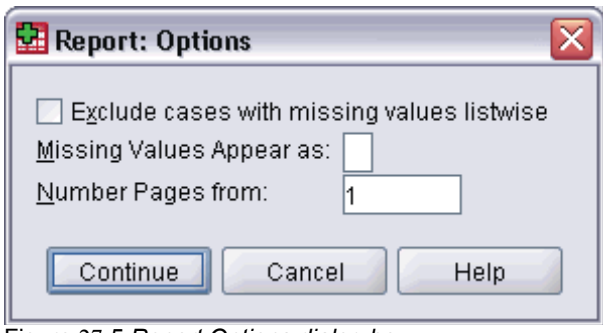


Figure 27-5 *Report Options dialog box*

Exclude cases with missing values listwise. Eliminates (from the report) any case with missing values for any of the report variables.

Missing Values Appear as. Allows you to specify the symbol that represents missing values in the data file. The symbol can be only one character and is used to represent both **system-missing** and **user-missing** values.

Number Pages from. Allows you to specify a page number for the first page of the report.

Report Layout

Report Layout controls the width and length of each report page, placement of the report on the page, and the insertion of blank lines and labels.

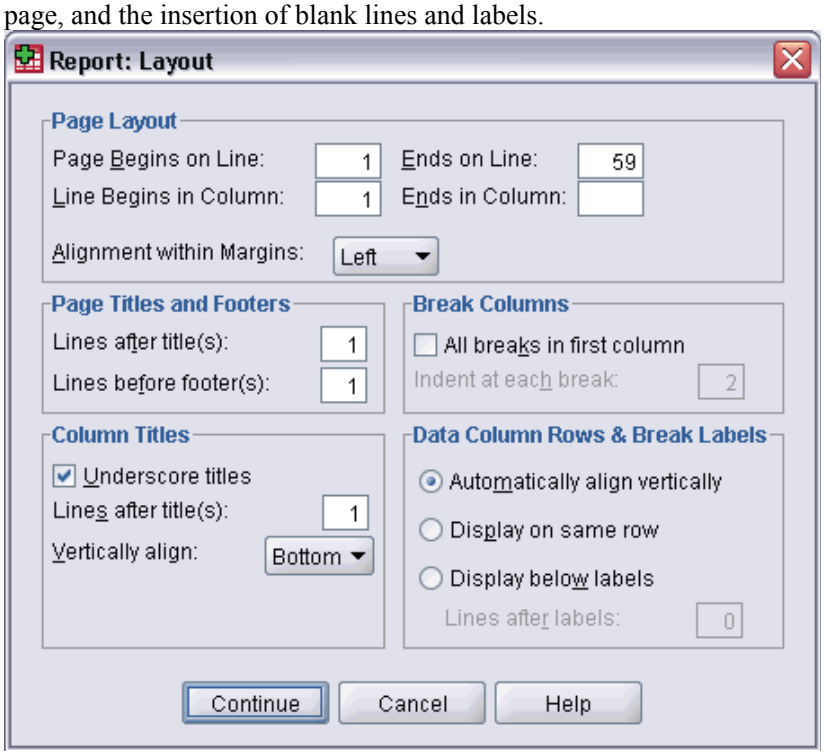


Figure 27-6 *Report Layout dialog box*

Page Layout. Controls the page margins expressed in lines (top and bottom) and characters (left and right) and reports alignment within the margins.
Page Titles and Footers. Controls the number of lines that separate page titles and footers from the body of the report.
Break Columns. Controls the display of break columns. If multiple break variables are specified, they can be in separate columns or in the first column. Placing all break variables in the first column produces a narrower report.
Column Titles. Controls the display of column titles, including title underlining, space between titles and the body of the report, and vertical alignment of column titles.
Data Column Rows and Break Labels. Controls the placement of data column information (data values and/or summary statistics) in relation to the break labels at the start of each break category. The first row of data column information can start either on the same line as the break category label or on a specified number of lines after the break category label. (Not available for column summary reports.)

Report Titles

Report Titles controls the content and placement of report titles and footers. You can specify up to 10 lines of page titles and up to 10 lines of page footers, with left-justified, centered, and right-justified components on each line.

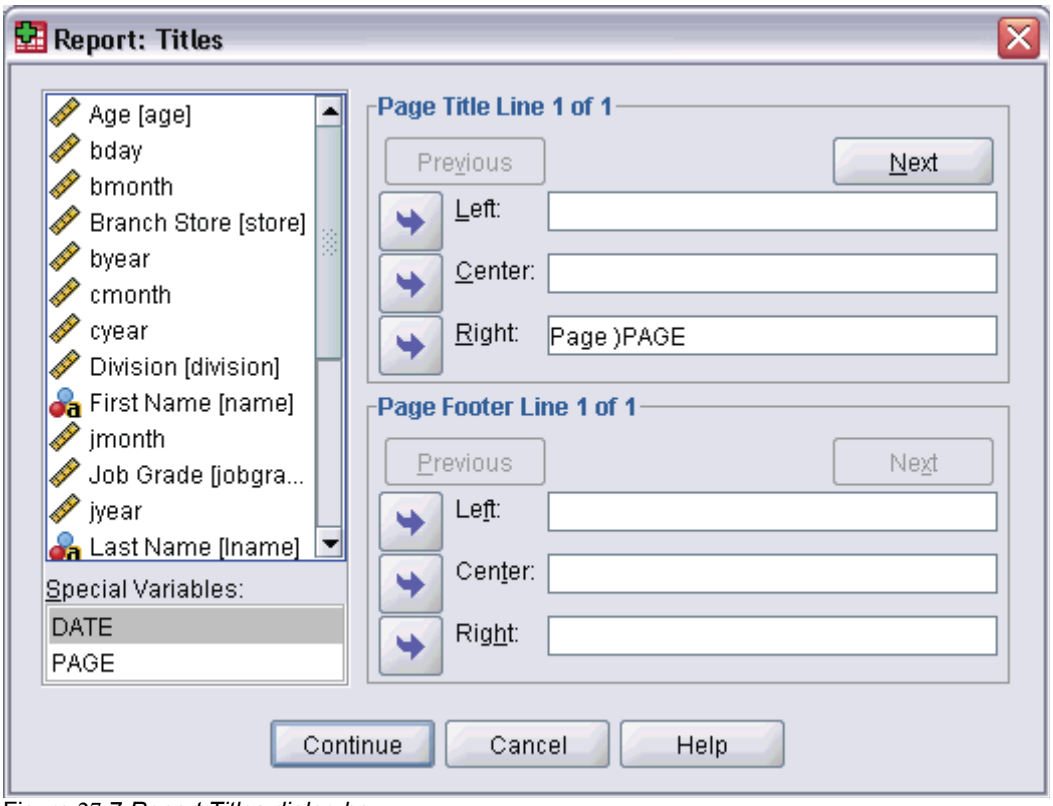


Figure 27-7 *Report Titles dialog box*

If you insert variables into titles or footers, the current value label or value of the variable is displayed in the title or footer. In titles, the value label corresponding to the value of the variable at the beginning of the page is displayed. In footers, the value label corresponding to the value of the variable at the end of the page is displayed. If there is no value label, the actual value is displayed.
Special Variables. The special variables *DATE* and *PAGE* allow you to insert the current date or the page number into any line of a report header or footer. If your data file contains variables named *DATE* or *PAGE*, you cannot use these variables in report titles or footers.

Report Summaries in Columns

Report Summaries in Columns produces summary reports in which different summary statistics appear in separate columns.

Example. A company with a chain of retail stores keeps records of employee information, including salary, job tenure, and the division in which each employee works. You could generate a report that provides summary salary statistics (for example, mean, minimum, and maximum) for each division.

Data Columns. Lists the report variables for which you want summary statistics and controls the display format and summary statistics displayed for each variable.

Break Columns. Lists optional break variables that divide the report into groups and controls the display formats of break columns. For multiple break variables, there will be a separate group for each category of each break variable within categories of the preceding break variable in the list. Break variables should be discrete categorical variables that divide cases into a limited number of meaningful categories.

Report. Controls overall report characteristics, including display of missing values, page numbering, and titles.

Preview. Displays only the first page of the report. This option is useful for previewing the format of your report without processing the whole report.

Data are already sorted. For reports with break variables, the data file must be sorted by break variable values before generating the report. If your data file is already sorted by values of the break variables, you can save processing time by selecting this option. This option is particularly useful after running a preview report.

To Obtain a Summary Report: Summaries in Columns

E From the menus choose:

Analyze
Reports
Report Summaries in Columns...

E Select one or more variables for Data Columns. One column in the report is generated for each variable selected.

E To change the summary measure for a variable, select the variable in the Data Column Variables list and click Summary.

E To obtain more than one summary measure for a variable, select the variable in the source list and move it into the Data Column Variables list multiple times, one for each summary measure you want.

E To display a column containing the sum, mean, ratio, or other function of existing columns, click Insert Total. This places a variable called *total* into the Data Columns list.

E For reports sorted and displayed by subgroups, select one or more variables for Break Columns.

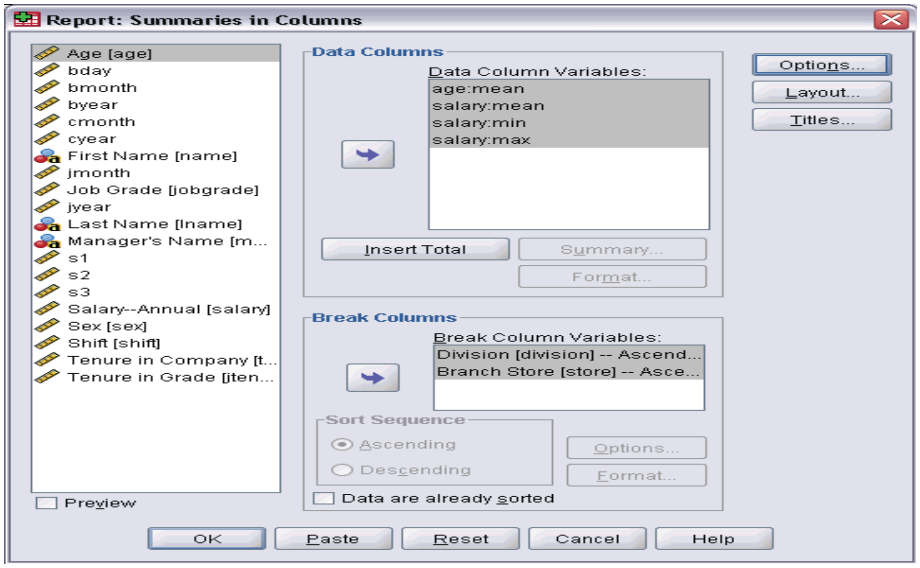


Figure 27-8 *Report Summaries in Columns dialog box*

Data Columns Summary Function

Summary Lines controls the summary statistic displayed for the selected data column variable.

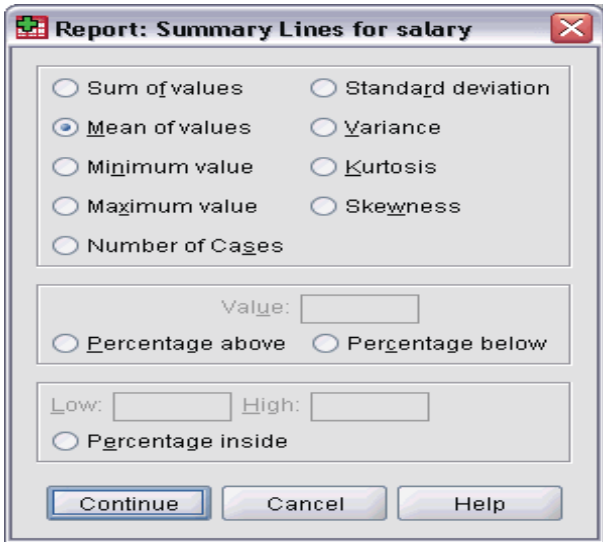


Figure 27-9 *Report Summary Lines dialog box*

Available summary statistics are sum, mean, minimum, maximum, number of cases, percentage of cases above or below a specified value, percentage of cases within a specified range of values, standard deviation, variance, kurtosis, and skewness.

Data Columns Summary for Total Column

Summary Column controls the total summary statistics that summarize two or more data columns. Available total summary statistics are sum of columns, mean of columns, minimum, maximum, difference between values in two columns, quotient of values in one column divided by values in another column, and product of columns values multiplied together.

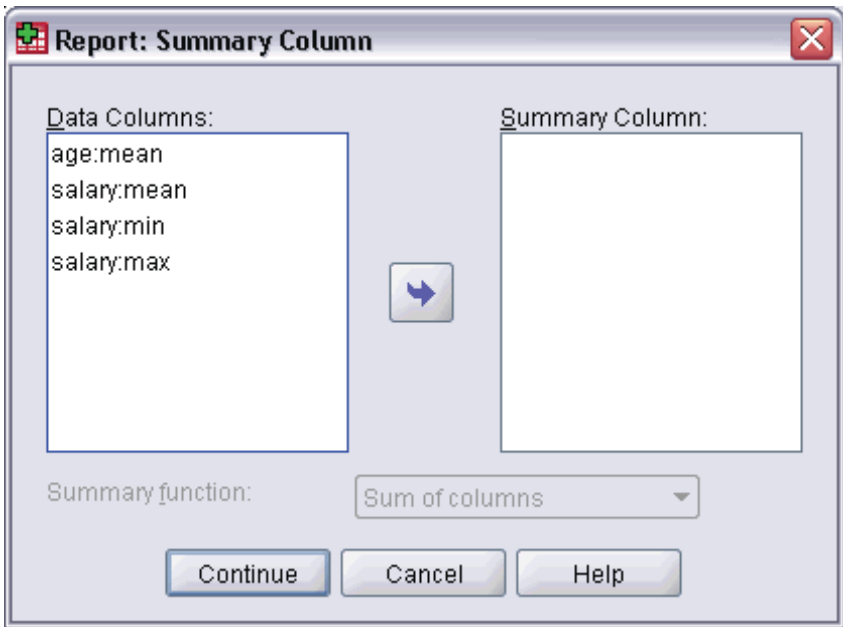


Figure 27 -10 *Report Summary Column dialog box*

Sum of columns. The *total* column is the sum of the columns in the Summary Column list.
Mean of columns. The *total* column is the average of the columns in the Summary Column list.
Minimum of columns. The *total* column is the minimum of the columns in the Summary Column list.
Maximum of columns. The *total* column is the maximum of the columns in the Summary Column

list.
1st column – 2nd column. The *total* column is the difference of the columns in the Summary Column list. The Summary Column list must contain exactly two columns.
1st column / 2nd column. The *total* column is the quotient of the columns in the Summary Column list. The Summary Column list must contain exactly two columns.
% 1st column / 2nd column. The *total* column is the first column's percentage of the second column in the Summary Column list. The Summary Column list must contain exactly two columns.
Product of columns. The *total* column is the product of the columns in the Summary Column list.

Report Column Format

Data and break column formatting options for Report Summaries in Columns are the same as those described for Report Summaries in Rows.

Report Summaries in Columns Break Options

Break Options controls subtotal display, spacing, and pagination for break categories.

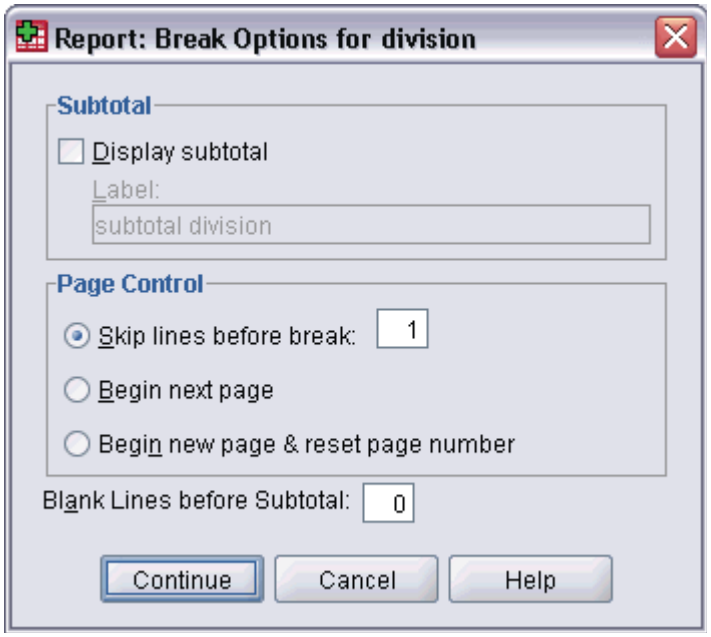


Figure 27 -11 *Report Break Options dialog box*

Subtotal. Controls the display subtotals for break categories.
Page Control. Controls spacing and pagination for categories of the selected break variable. You can specify a number of blank lines between break categories or start each break category on a new page.
Blank Lines before Subtotal. Controls the number of blank lines between break category data and subtotals.

Report Summaries in Columns Options

Options controls the display of grand totals, the display of missing values, and pagination in column summary reports.

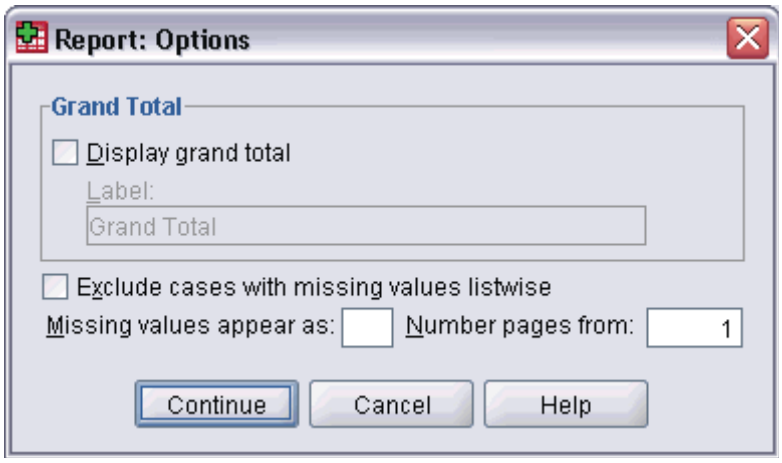


Figure 27-12 *Report Options dialog box*

Grand Total. Displays and labels a grand total for each column; displayed at the bottom of the column.
Missing values. You can exclude missing values from the report or select a single character to indicate missing values in the report.

Report Layout for Summaries in Columns

Report layout options for Report Summaries in Columns are the same as those described for Report Summaries in Rows.

REPORT Command Additional Features

The command syntax language also allows you to:

- Display different summary functions in the columns of a single summary line.
- Insert summary lines into data columns for variables other than the data column variable or for various combinations (composite functions) of summary functions.
- Use Median, Mode, Frequency, and Percent as summary functions.
- Control more precisely the display format of summary statistics.
- Insert blank lines at various points in reports.
- Insert blank lines after every *n*th case in listing reports.

Because of the complexity of the `REPORT` syntax, you may find it useful, when building a new report with syntax, to approximate the report generated from the dialog boxes, copy and paste the corresponding syntax, and refine that syntax to yield the exact report that you want.
See the *Command Syntax Reference* for complete syntax information.

Глава 28. Графики

Общо за графичната структура (Overview of the Chart Facility)

Високо–резолюционни графики и диаграми се създават като използваме менюто за графики и от функциите на менюто за анализ. Тази глава ни дава общ преглед върху графичната структура.

Създаване и редакция на графика

Преди да започнем да изграждаме графика, трябва да имаме данните си в Редактора за данни. Можете директно да въведете данните в Редактора, да отворите предварително запаметени данни; да прочетете нов работен документ; tab-delimited data file или файл бази данни.
Упътването на Помощното меню има онлайн примери за създаване и моделиране на графики и онлайн Помощна система даваща информация за създаване и моделиране на всички видове графики.

Създавене на графики

Изграждача на Графики ни позволява да построим графика по предварително начертана, галерияс графики (палитра с графики) или чрез сглобяване с индивидуални отделни части (като например, оси и пръчки). Графика може да се построи чрез придърпване и поставяне на готова графика от галерията или на отделни елементи върху работния плот, който се намира в дясно от Листа с Променливи и заема голямо пространство от Кутията на Изграждача на Графика(Chart Builder dialog box)

Когато създаваме графика работния лист ни дава възможност за предварителен изглед на графиката. Въпреки, че се използват дадени променливи етикети и измерване на нивата, не се изобразяват действителните данни. Вместо това се използват произволно генерирани данни за представяне на груб чертеж на това как ще изглежда графиката. За нови потребители използването на галерията е препоръчително.

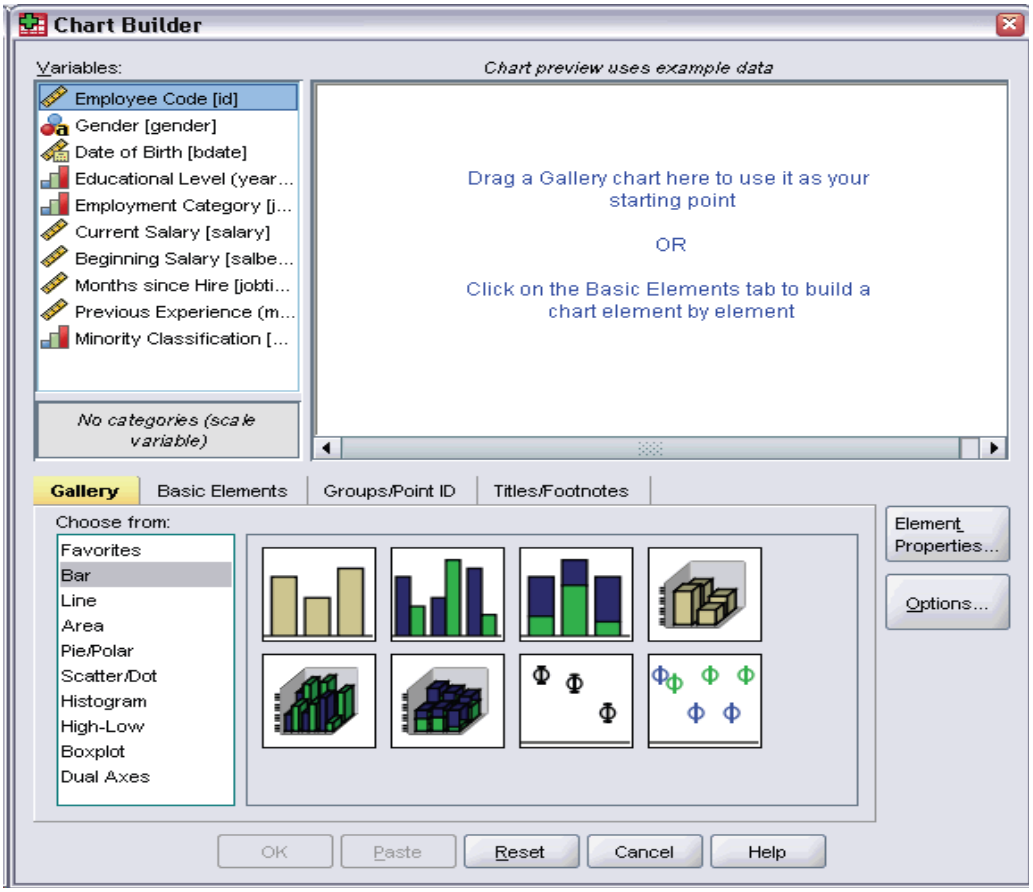
Как да стартираме Изграждача на Графика

От менюто избираме :

- Graphs

  Chart Builder ....

Това води до отварянето на прозореца Изграждане на Графика



Фиг. 28-1 *Chart Builder dialog box*

Изграждане на графика използвайки Галерията

Най-лесният начин за изграждане на графикае чрез използване на Галерията. Следват се основни стъпки за изграждане на графика от галерия.

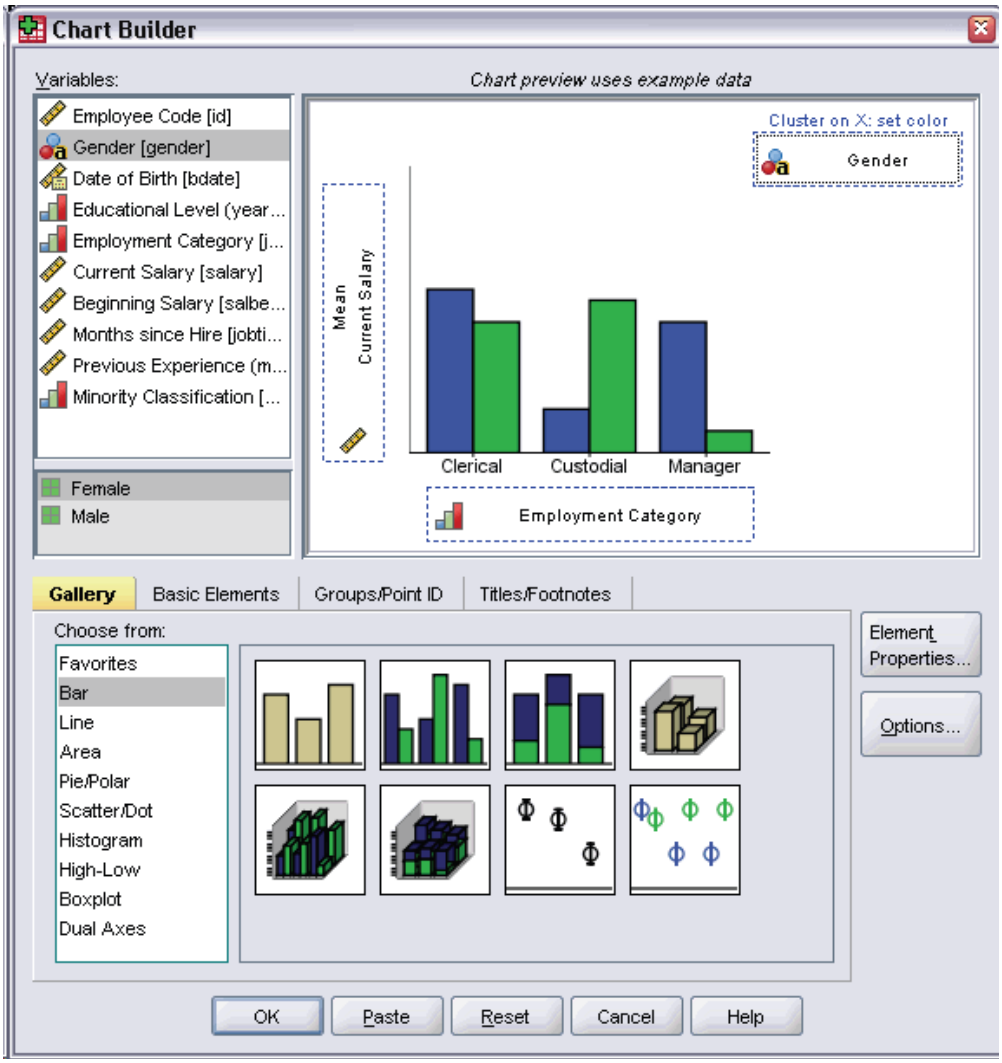
– клипнете върху Галерния таб, ако все още не се е появил

- В листа на Избор на графика , изберете категорията на графиката. Всяка категория предлага няколко вида.

- Плъзнете картинката на избраната графика върху работния лист (canvas). Можете да го направите и като клипнете 2 пъти върху картинката. Ако работния лист има вече графика, то галерната графиката заменя съответната ос и графичните елементи на графиката.

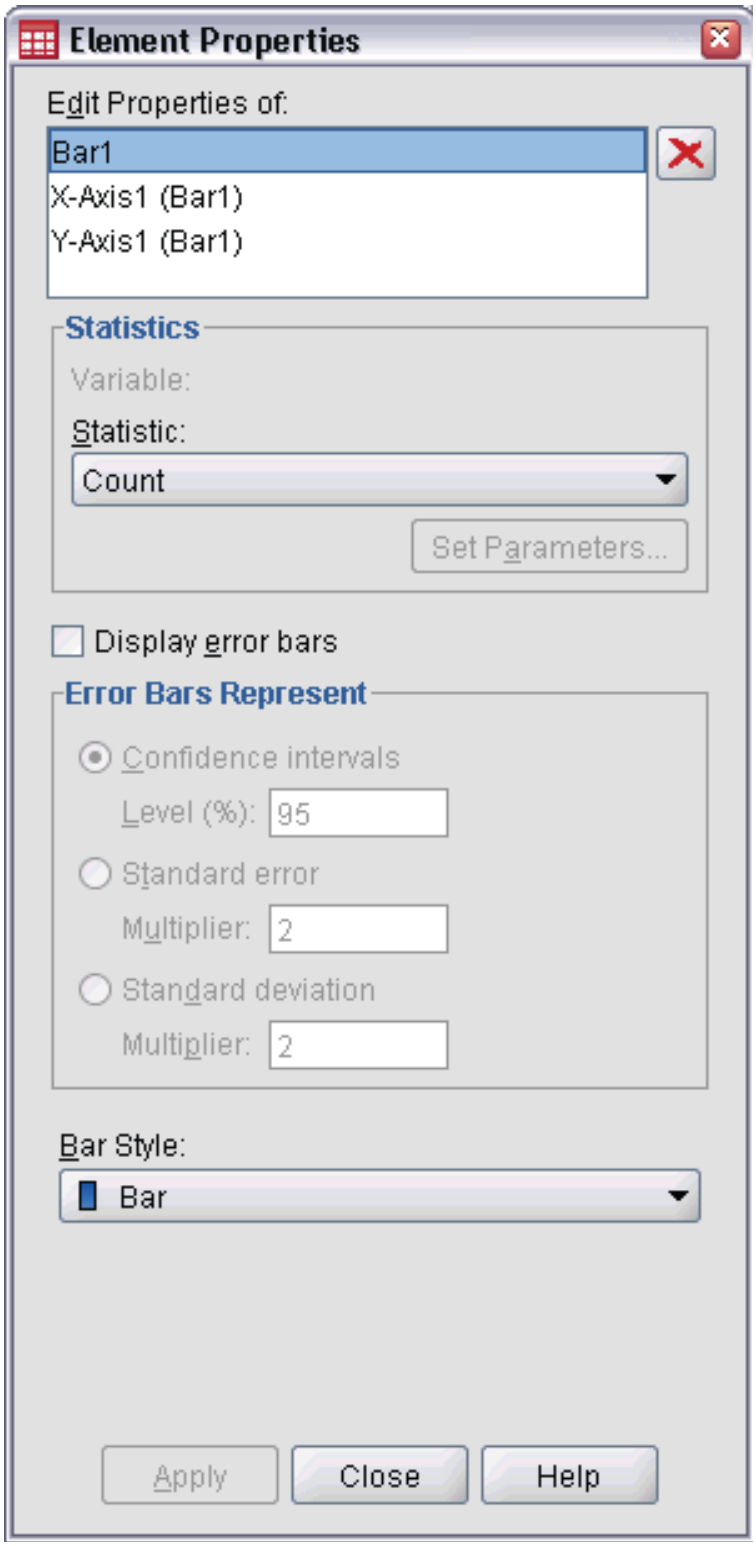
- Вмъкнете променлива от списъка с Променливи и я сложете в съответните осни зони, ако има такива. Ако таиава зони вече съществуват и показват някаква статистика и искате да използвате тази статистика, не е необходимо да вкарвате променлива в тази зона. Необходимо е да добавите променлива само, тогава когато текстът в зоната е син. Ако текстът е черен, зоната вече съдържа променлива или статистическа данна.

Забележка: Измерването на нивото на променливите е от изключително значение. Изграждача на Графика прави графиката въз основа на измерено ниво. Освен това, изградената графика може да изглежда различно за различните нива на измерването. Вие можете временно да смените нивото на измерените променливи чрез клипване с десния бутон върху променливата и изберете съответната опция.



Фиг. 43-2 *Изграждне на Графика със запълнени зони.*

Е- Ако трябва да смените статистическа данна или видоизмените някое свойство на осите или на легендите(като размер на размаха (scale range), изберете Елемент Свойства



Фиг. 28-3 Свойства на Елемента диалогова кутия

- От списъка Редакция на Свойства изберете елемента, който искате да промените (За информация относно специфичните свойства клипнете на Помощ)
- След като направите промяната, клипнете Приложи
- Ако трябва да добавите още променливи в графиката (например за групиране или облицовка) клипнете на Групи/Покажи ID tab в диалоговата кутия за Изграждане на

Графика и изберете една или повече опция. След това придърпайте окончателната променливи до новите зони на спускане в работния лист(canvas)
- Ако искате да разместите графика(като например да направите решетките хоризонтални) клипнете на таба с Основни Елементи и след това Размести
- Клипнете ОК за да създадете нова графика. Графиката се изобразява в диаскоп/Viewer

Фиг. 28-4 Графика изобразена в Прозореца за резултати / *Bar chart displayed in Viewer window*

Редакция на Графика

Редактора за Графики ни предоставя мощна, лесна за използване среда, в която можете да изграждате свой графики и да изследвате вашите данни. Характеристиките на Редактора на Графика са:

- ✓ **Прост, интуитивен интерфейс –** имате бърз достъп за редакция на части от графиката чрез менюта, подменюта и тулбарове . Освен това можете да въведете текст директно върху графиката.
- ✓ **Широка гама от опции за форматиране и статистически цели.** Можете да избирате от цяла гама от стилове и статистически опции.
- ✓ **Мощни проучвателни инструменти –** Можете да разглеждате вашите данни по различни начини, например чрез етикиране, пренареждане и ротация. Можете да промените вида

на графиката и ролята на променливите в нея. Можете да добавите криви на разпределение, на променливата (fit), на интерплотацията и линии за справка (and reference lines)

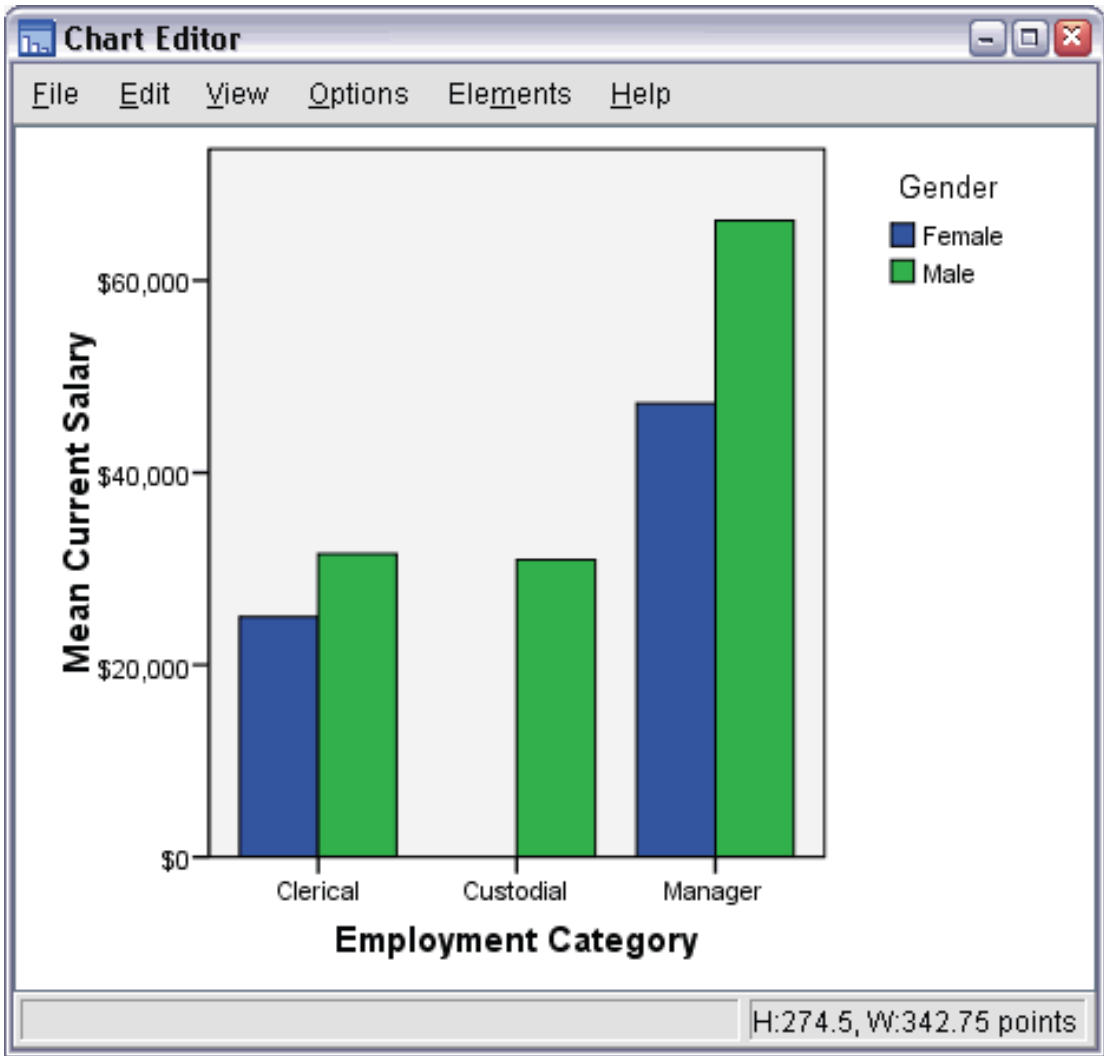
- ✓ **Гъвкави шаблони за последователен вид и поведение-** Можете да създавате шаблони и да ги използвате за създаване на графики с облик и фукнции, които вие желаете**.** Например, ако винаги искате конкретна ориентация за ос етикети, можете да определите ориентацията в шаблона и да приложите шаблона в други графики.

Как да Изглежда Редактора на Графика

Е- Създайте графика в SPSS, или отворете Диаскопов файл с графика

Е- Клипнете два пъти върху графиката в Диаскопа

Графиката се изобразява в Редактора на Графика



Фиг. 28-5 -Графика изобразена в Редактора на Графика

Съществени правила за Редактора на Графика

Предоставя редица методи за манипулиране на графики

Менюта

Много от действята, които можете да изпълнявате в Редактора на Графики стават чрез менюта, особено след добавянето на елемент към графиката. Например, използвате менюта за добавянето на крива в разпръснати точки. След добавянето на елемент към графиката често използвама свойствата на диалоговата кутия за да определим свойствата на добавения елемент.

Свойства на Диалоговата кутия

Свойствата на една графика, както и свойствата на елемент от графиката могат да бъдат намерени в Свойства на Диалоговата кутия. За да видите тези свойства трябва да:

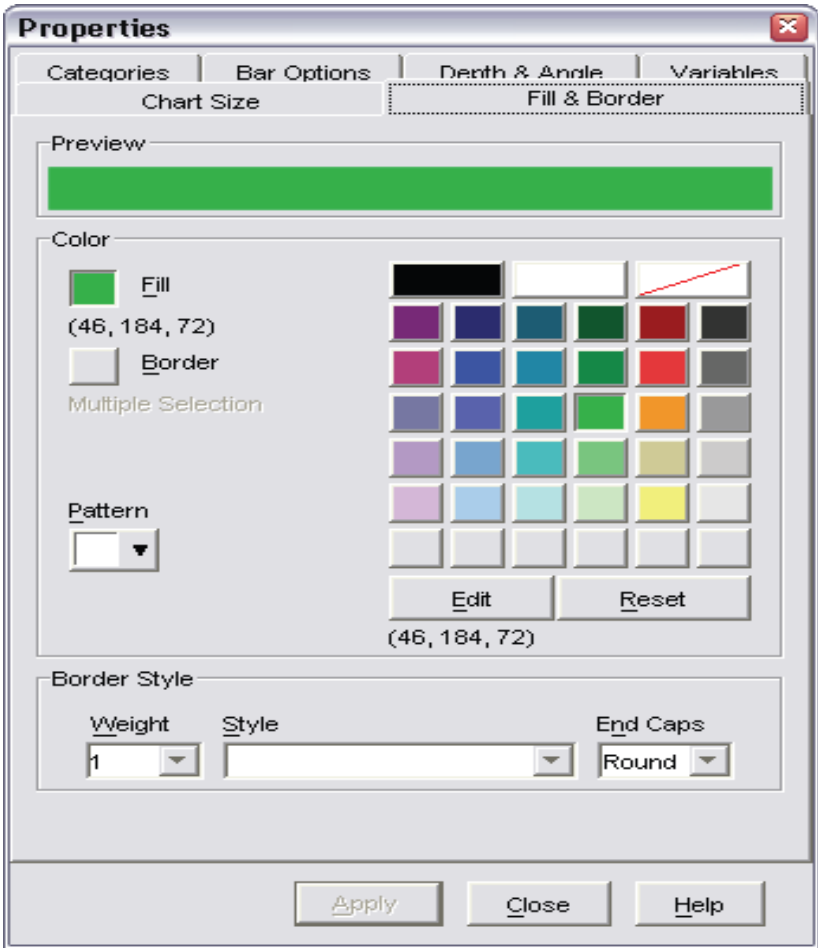
- Клипнете два пъти върху елемент от графиката

или

- изберете елемент от графиката и от менюто да изберете:

Редакция

Свойства, Допълнителен /Additionally , Свойствата на Диалоговата кутия автоматично се появяват след като добавите елемент към графиката.



Фиг. 28-6 Свойства на диалоговата кутия, Properties dialog box, Fill & Border tab

The Properties dialog box has tabs that allow you to set the options and make other changes to a chart. The tabs that you see in the Properties dialog box are based on your current selection. Some tabs include a preview to give you an idea of how the changes will affect the selection when you apply them. However, the chart itself does not reflect your changes until you click Apply. You can make changes on more than one tab before you click Apply. If you have to change the selection to modify a different element on the chart, click Apply before changing the selection. If you do not click Apply before changing the selection, clicking Apply at a later point will apply changes only to the element or elements currently selected.
Depending on your selection, only certain settings will be available. The help for the individual tabs specifies what you need to select to view the tabs. If multiple elements are selected, you can change only those settings that are common to all the elements.

Toolbars

The toolbars provide a shortcut for some of the functionality in the Properties dialog box. For example, instead of using the Text tab in the Properties dialog box, you can use the Edit toolbar to change the font and style of the text.

Saving the Changes

Chart modifications are saved when you close the Chart Editor. The modified chart is subsequently displayed in the Viewer.

Chart Definition Options

When you are defining a chart in the Chart Builder, you can add titles and change options for the chart creation.

Adding and Editing Titles and Footnotes

You can add titles and footnotes to the chart to help a viewer interpret it. The Chart Builder also automatically displays error bar information in the footnotes.

How to Add Titles and Footnotes

E Click the Titles/Footnotes tab.
E Select one or more titles and footnotes. The canvas displays some text to indicate that these were added to the chart.
E Use the Element Properties dialog box to edit the title/footnote text.

How to Remove a Title or Footnote

E Click the Titles/Footnotes tab.
E Deselect the title or footnote that you want to remove.

How to Edit the Title or Footnote Text

When you add titles and footnotes, you cannot edit their associated text directly on the chart. As with other items in the Chart Builder, you edit them using the Element Properties dialog box.
E Click Element Properties if the Element Properties dialog box is not displayed.
E In the Edit Properties Of list, select a title, subtitle, or footnote (for example, Title 1).
E In the content box, type the text associated with the title, subtitle, or footnote.
E Click Apply.

Setting General Options

The Chart Builder offers general options for the chart. These are options that apply to the overall chart, rather than a specific item on the chart. General options include missing value handling, templates, chart size, and panel wrapping.
E Click Options.
Modify the general options. Details about these follow.
E Click Apply.

User-Missing Values

Break Variables. If there are missing values for the variables used to define categories or subgroups, select Include so that the category or categories of user-missing values (values identified as missing by the user) are included in the chart. These “missing” categories also act as break variables in calculating the statistic. The “missing” category or categories are displayed on the category axis or in the legend, adding, for example, an extra bar or a slice to a pie chart. If there are no missing values, the “missing” categories are not displayed.
If you select this option and want to suppress display after the chart is drawn, open the chart in the Chart Editor and choose Properties from the Edit menu. Use the Categories tab to move the categories that you want to suppress to the Excluded list. Note, however, that the statistics are not

recalculated if you hide the “missing” categories. Therefore, something like a percent statistic will still take the “missing” categories into account.
Note: This control does not affect system-missing values. These are always excluded from the chart.
Summary Statistics and Case Values. You can choose one of the following alternatives for exclusion of cases having missing values:
- **Exclude listwise to obtain a consistent base for the chart.** If any of the variables in the chart has a missing value for a given case, the whole case is excluded from the chart.
- **Exclude variable-by-variable to maximize the use of the data.** If a selected variable has any missing values, the cases having those missing values are excluded when the variable is analyzed.

To see the difference between listwise and variable-by-variable exclusion of missing values, consider the following figures, which show a bar chart for each of the two options.

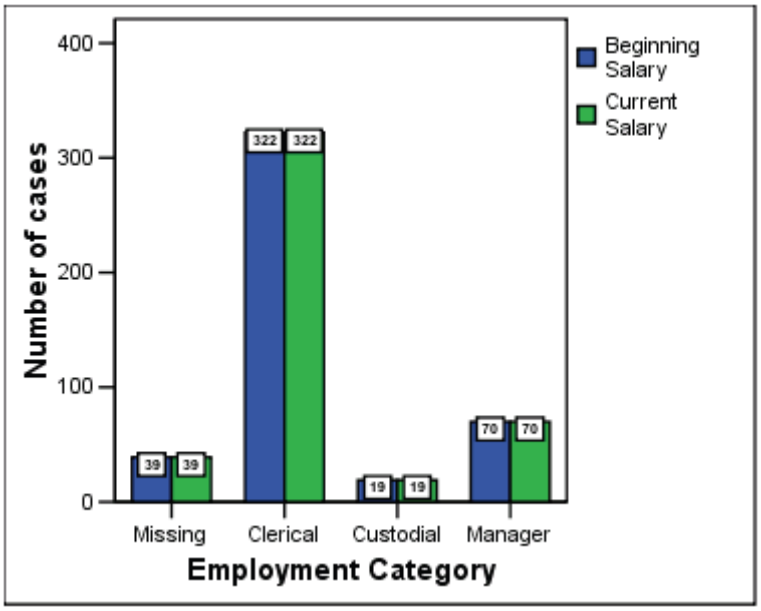


Figure 28-7 Listwise exclusion of missing values

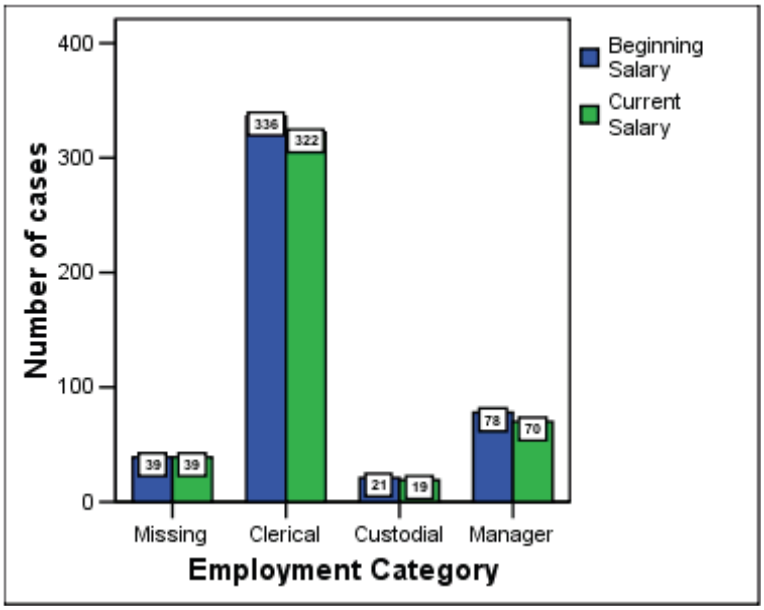


Figure 28-8 Variable-by-variable exclusion of missing values

The charts were created from a version of the *Employee data.sav* file that was edited to have some system-missing (blank) values in the variables for current salary and job category. In some other cases, the value 0 was entered and defined as missing. For both charts, the option Display groups defined by missing values is selected, which adds the category *Missing* to the other displayed job categories. In each chart, the values of the summary function, *Number of cases*, are displayed in the bar labels.

In both charts, 26 cases have a system-missing value for the job category, and 13 cases have the user-missing value (0). In the listwise chart, the number of cases is the same for both variables in each bar cluster because whenever a value was missing, the case was excluded for all variables. In the variable-by-variable chart, the number of nonmissing cases for each variable in a category is plotted without regard to missing values in other variables.

Templates

A chart template allows you to apply the attributes of one chart to another. When you open a chart in the Chart Editor, you can save it as a template. You can then apply that template by specifying it at creation or later by applying it in the Chart Editor.

Default Template. This is the template specified by the Options. You can access these by choosing Options from the Edit menu in the Data Editor and then clicking the Charts tab. The default template is applied first, which means that the other templates can override it.

Template Files. Click Add to specify one or more templates with the standard file selection dialog box. These are applied in the order in which they appear. Therefore, templates at the end of the list can override templates at the beginning of the list.

Chart Size and Panels

Chart Size. Specify a percentage greater than 100 to enlarge the chart or less than 100 to shrink it. The percentage is relative to the default chart size.

Panels. When there are many panel columns, select Wrap Panels to allow panels to wrap across rows rather than being forced to fit in a specific row. Unless this option is selected, the panels are shrunk to force them to fit in a row.