Graphs of confidence interval limits for the basic statistical tests in dependence on the data number

Tsvetan B. Georgiev
New Bulgarian University, Department of Natural Science
21 Montevideo Str., BG-1618, Sofia

ABSTRACT

Graphs of the confidence interval limits for statistical guaranties of 95 % and 99 % are presented in dependence on the data number *n*. The tests comprehend (i) change of the population mean or standard deviation, (ii) compare of averages or mean quadratic deviations, (iii) significance of th regression and correlation coefficients, as well as (iv) confidence intervals and standard deviation changes through Student's and Fisher's tests. The graphs give fast solution of the direct problem – estimation of the confidence interval limits in dependence on the available data number *n*, as well as the reverse problem – estimation of the necessary data number *n* for achieve the desired statistical guaranty. The graphs may be useful in the practice of the scientific investigations. They are an efficient tool in the education about the basics of the statistical inferences too.

Keywords: *data – analysis; statistical – methods; statistical - inferences*

Графики на доверителните интервални граници за основните статитстически тестове в зависимост от броя данни

Цветан Б. Георгиев
Нов български университет, Департамент „Природни науки"

РЕЗЮМЕ

Представени са графики на границите на доверителните интервали за статистически гаранции 95% и 99% в зависимост от броя данни *n*. Обхванати са тестовете относно (i) изменение на средно-аритметична стойност или средно-квадратично отклонение, (ii) сравнение на две средно-аритметични стойности или две средно-квадратични отклонения, (iii) значимост на коефициенти на корелация или регресия, както и (iv) доверителни интервали и променливост на стандартно отклонение чрез пирсънов и фишеров тестове. Графиките дават бързо визуално решение на правата задача – оценка на граница на доверителен интервал в зависимост от наличния брой данни *n* , както и на обратната задача – оценка на необходимия брой данни *n* за постигане на желаната статистическа гаранция. Графиките могат да бъдат полезни за бързи статистически изводи в практиката на научните изследвания. Те се оказват ефективни и като образователно средство относно основите на теорията на статистическите изводи.

Ключови думи: *анализ на данни, статистически методи, статистически изводи*

1. Увод

Всеки прост статистически тест оценява принадлежността на неизвестната стойност на статистически параметър, напр. средната стойност (математическото очакване) μ или стандартното отклонение σ, към някакъв интервал на стойностите на параметъра, с предварително зададена (висока) вероятност, напр. P=95%. Границите на интервала се оценяват чрез някакъв изчислим от данните целеви параметър (ЦП) , напр. подходяща функция на средно-аритметична стойност m или подходяща функция средно-квадратично отклонение s. Зададената вероятност и съответният на нея интервал се наричат доверители или гаранционни. Доверителната вероятност, напр. P = 95%, се нарича и ниво на доверие. На практика вместо P по-често се използва съответната допълнителна вероятност $p=1-P$, в случая p = 5%, която се нарича ниво на грешка.

Смисълът на статистическия тест е най-нагледен в случая на проверка за променливост на някакъв параметър, напр. средната стойност μ: Ако съответният ЦП принадлежи на доверителния интервал, смята се, че възможното изменение на параметъра е статистически незначимо, а ако не принадлежи – статистически значимо. В първия случай се приема т.н. нулева хипотеза (H_0, няма значимо изменение на параметъра, нищо интересно), а във втория случай се приема т.н. алтернативна хипотеза (H_A, има значимо изменение, става интересно). Тези специални статистически хипотези не се „доказват“ или „опровергават“, както обикновено в математиката, а се „приемат“ или „отхвърлят“, като окончателно решение се взема от потребителя.

При неголям брой данни оценката на статистическия параметър, например μ или σ, чрез, съответно, m или s, е несигурна. Точно тогава се препоръчва да се оцени и доверителния интервал на статистическия параметър. Този подход, наречен интервално оценяване на статистически параметри, е въведен от английския генетик и статистик Роналд Фишер (Fisher, 1925). Съответната методика е изложена системно в множество монографии (Neyman & Pearson 1933, Tucker 1962, Zaks 1971, Cox & Hinkley 1974) и др.

На практика потребителят намира границите на доверителния интервал чрез предварително съставени таблици или чрез компютърни програми. Обаче, графиките на граници на интервали в зависимост от броя данни n имат поне три предимства пред таблиците и компютъра. Първо, за тях се използват удобни и лесно изчислими (на ум) ЦП, явяващи се функция напр. на m и s. Второ, те дават бързо визуално решение на правата задача (проверка на алтернативната хипотеза по дадено n) и на обратната задача (оценка на необходимото n за евентуално приемане на алтернативната хипотеза). Трето, те дават визуални решения на двете задачи едновременно, а тук и за две характерни доверителни вероятности – 95 % и 99 %. Недостатъкът на графиките е ниската точност на резултата, само две значещи цифри, но практика тази точност е почти винаги достатъчна.

Тук и нататък се предполага нормално разпределена случайна величина *X,* представена чрез извадка от n нейни независими реализации (данни) x_1, x_2, …, x_n. Главните статистически параметри на *X* са неизвестната средна стойност (population mean; математично очакване) μ, неизвестното стандартно отклонение (population standard deviation, стандартно отклонение) σ. Смятат се за изчислени средно-аритметичната стойност на извадката (sample mean, ean arithmetic value) m и средно- квадратичното отклонение спрямо средната стойност на извадката (sample standard deviation, mean quadratic deviation) s. Обичайните оценки на m и s се получават по формулите

(0.0) $m = (\Sigma x_j)/n$ и $s = [\Sigma\Delta x_j)/(n\text{-}1)]^{1/2}$,

където $\Delta x_j = x_j - m$. Тук и навсякъде нататък сумирането е винаги от 1 до n.

На практика се използват два вида доверителни интервали: едностранно-ограничени (едностранни), $(-\infty, \delta(p)]$ или $[\delta(p), \infty)$ и двустранно ограничени (двустранни), $[\delta_1(p/2), \delta_2(p/2)]$. При търсене на променливост на вече известна величина видът на интервала се определя в зависимост от задачата. Едностранният интервал се отнася за параметри, които предположително могат *само да намаляват* или *само да се увеличават*. Примери са нарастването на броя на хората и намаляването на броя на горилите с течение на времето, в наши дни. Двустранният интервал се отнася за параметри, които могат *както да намаляват, така и да се да се увеличават*. Пример е максималната дневна температура на дадено място, която флуктуира в рамките на няколко дни. И в двата случая статистическата гаранция се отнася за предварително зададено на ниво на гаранция P и съответно ниво на грешка p=1-P. И в двата случая изчисленият доверителни интервал расте с нарастването на стандартното отклонение σ, ако то се смята за известно, или пък, на практика – с нараствването на сасредно-квадратичното отклонение s (т.е. с намаляването на броя на измерванията n. Последното твърдение отразява интуитивното схващане, че при по-точни и/или повече на брой измервания, резултатът е по-близо до истинската, неизвестна стойност на статистическия параметър.

Границите на доверителните интервали се определят за изчислим от данните ЦП. Статистическият тест се състои в сравнение на конкретния ЦП с границите на съответния доверителен интервал, определен за дадено P и зададен брой измервания n. Когато ЦП има симетрично случайно разпределение (от гаусов или стюдънтов тип) двустранният интервал има симетричния вид $[\delta(p/2), \delta(p/2)]$, а иначе – не. Статистическият извод за променливост (на отхвърляне на нулевата хипотеза) чрез двустранен интервал, е по-строг отколкото чрез едностранен интервал.

По-нататък се дават зависимости на интервалните граници $\delta(\cdot)$ от броя данни n, защото така те са най-удобни за практиката. Границата на доверителния интервал е ордината $\delta(\cdot)$ над абсцисата на съответния брой данни n. Когато ЦП попада в доверителния интервал се приема нулевата хипотеза, т.е. проверяваният статистически параметър се смята за непроменен с P % гаранция. Иначе се приема алтернативната хипотеза, т.е., изменението на параметъра се смята статистически значимо с P % гаранция.

В тази статия се дават последователно графики за (1) интервали и променливост на средната стойност чрез гаусов и стюдънтов тест, (2) сравнение на две средни стойности чрез гаусов и стюдънтов тест, (3) интервали и променливост на корелационни и регресионни коефициенти чрез стюдънтов тест и (4) интервали и променливост на стандартното отклонение чрез пирсънов и фишеров тест. Изложението следва основните пунктове на публикацията на Georgiev (2014).

1. Доверителни интервали и променливост на средната стойност μ чрез гаусов тест или стюдънтов $T(n\text{-}1)$ тест

Тук и нататък винаги се предполага, че са добити n данни (независими един от друг резултати от измервания) x_j , j=1, 2, …, n), явяващи се реализации на случайна величина с нормално разпределение. За теоретичното разглеждане тук се предполага още, че тези данни са стандартизирани, т. е мащабирани към разпределение с нулева средна стойност и единично стандартно отклонения по два начина: (i) към гаусово разпределение чрез субституцията $z_j = (x_j - \mu)/\sigma$ и (ii) към стюдънтово разпределение $T(n\text{-}1)$, с брой степени на свобода $f = n - 1$, чрез субституцията $t = (x_j - m)/s$.

Тогава обичайно дефинираните стандартни или абсолютни ЦП, които са случайни величини (защото са функции на случайни величини) са

(1.0') $|z'| = (m - \mu)/(\sigma/n^{1/2})$ или $|t'|= (m - \mu)/(s//n^{1/2})$.

Тези ЦП се използват широко за определуне на доверителни интервали и статистически изводи относно средната стойност μ. Според съответните теореми, z' следва гаусово разпределение, със стандартно отклонение $\sigma/n^{1/2}$, а t' следва стюдънтово $T(n\text{-}1)$ разпределение с f = n-1 степени на свобода, със стандартно отклонение $s/n^{1/2}$.

При тези означения всеки P % доверителен интервал за z' и t' може да бъде представен чрез съответните му критични граници $-z_c(p/2) < z' < z_c(p/2)$ или $-t_c(p/2;f) < t' < t_c(p/2;f)$ (двустранен), както и $-\infty < z' < z_c(p)$ и $-\infty < t' < t\ (p;f)$ (едностранен, за проверка дали средната стойност е само нараснала) или $z_c(p) < z' < \infty$ и $t_c(p;f) < t' < \infty$ (едностранен, за проверка дали средната стойност е само намаляла).

В този раздел се въвеждат по-естествени и по-удобни за пресмятане на ум, *относителни* ЦП (които също са случайни величини):

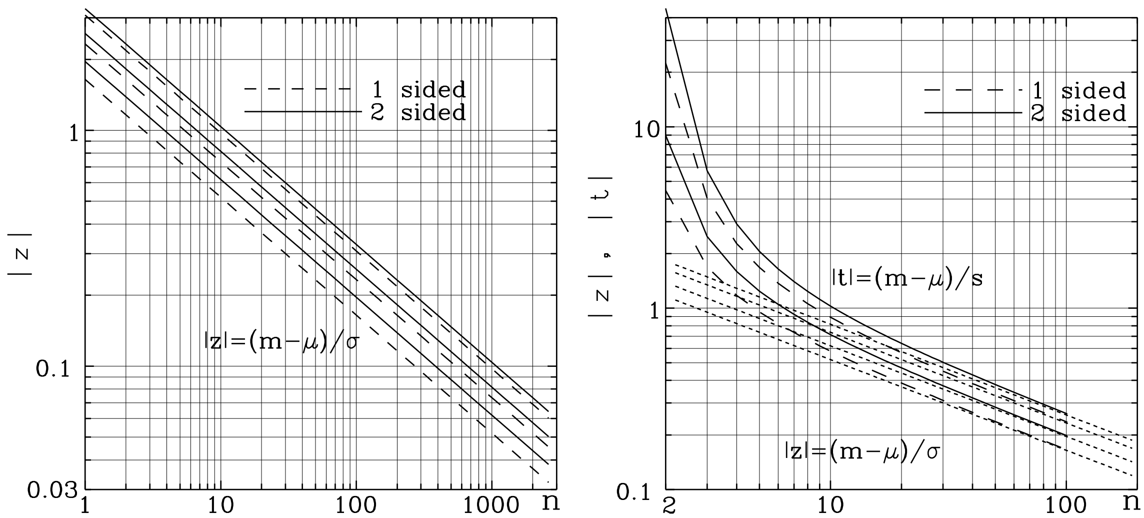
(1.0) $|z| = (m - \mu)/\sigma$ или $|t| = (m - \mu)/s$,

при които разликата $(m - \mu)$ се изразява в единици σ или s. Тогава едностранните доверителни интервали стават

(1.1) $|z| < z_c(p)/n^{1/2}$ или $|t| < t_c(p;f)/n^{1/2}$,

а двустранните доверителни интервали стават

(1.2) $|z| < z_c(p/2)/n^{1/2}$ или $|t| < t_c(p/2;f)/ n^{1/2}$,



Фиг.1. *a*: Доверителни граници за гаранции 95, 99 и 99.9 % (отдолу нагоре) в зависимост от броя данни n за гаусовия ЦП $|z| = (m - \mu)/\sigma$. Едностранните граници по (1.1) са представени чрез плътни прави, а двустранните по (1.2) – чрез прекъснати. *b*: Доверителни граници за гаранции P равни на 95 и 99 % в зависимост от броя данни n за стюдънтовия ЦП $|t| = (m - \mu)/\sigma$. Едностранните граници по (1.1) са представени чрез плътни криви, а двустранните по (1.2) – чрез прекъснати. Пунктирните прави показват за сравнение хода на доверителните граници за гаранции P равни на 95 и 99 % за $|z| = (m - \mu)/\sigma$.

В тази статия графиките на фиг.1. показват доверителните граници в зависимост от броя данни. Понеже разпределенията на Гаус и Стюдънт са симетрични, на фиг.1*a* и фиг.1*b* са дадени само горните, положителни, граници на интервалите (1.1) и (1.2).

Фигwра 1*a* представя критичните граници z_c за z от (1.1) и (1.2), при гаусово разпределение на z, т.е. при предварително известно стандартно отклонение σ, в зависимост от n. Границите са за кумулативни вероятности P равни на 95, 99 и 99.9%, със съответни квантили (процентили) равни на1.96, 2.576 и 3.283 (двустранен случай, плътни прави), или равни на 1.645, 2.326 и 3.181 (едностранен случай, прекъснати прави)

Фигура 1*b* представя критичните граници t_c за t от (1.1) и (1.2), при стюдънтово разпределение $T(n\text{-}1)$ на t в зависимост от n. Тук σ се смята за неизвестна величина и затова се използва s. Границите са за кумулативни вероятности 95 и 99%. Квантилите, използвани за построяването на кривите на Фиг.1*b* могат да се намерят в таблици в цитираните учебници и таблици в Интернет, като всички те произхождат от Fisher & Yates (1963). Границите за двустранния случай са плътни криви, а за едностранния прекъснати. На Фиг.1*b* се вижда, че при голям брой данни кривите на стюдънтовите граници клонят към прави (пунктирни линии), съответни на гаусовите граници, дадени в по-широк диапазон на Фиг.1*a*, т.е. когато данните са малко, трябва да се използват стюдънтовите граници.

Правата задача на графиките на фиг.1 е да способстват лесен и бърз статистически извод относно евентуалното изменение на някаква предварително известна средна стойност μ по n нови измервания, със средно аритметична стойност m и стандартно отклонение σ (фиг.1a) или средно квадратично отклонение s (фиг.1b). Обратната задача на графиките на фиг.1 е да дават лесен и бърз отговор на въпроса: „Колко най-малко данни n трябват, щото заподозряната в променливост средна стойност μ да бъде обявена за променена с висока статистическа гаранция." Ако стандартната грешка на измерванията σ е предварително известна, може да се прилага гаусов тест по фиг.1a,. а иначе, което винаги е статистически по-обосновано – стюдънтов тест по фиг.1b. На практика обикновено се работи с относителни стандартни отклонения, вкл. в проценти. Тогава и разликите в (1.1) и (1.2) следва да се изразяват в проценти.

Ето една *примерна задача* относно възможна промяна на интензитет на светлинен източник. Същото се отнася за параметър или източник на замърсяване на околната среда и др.п. Светлинен източник има предварително известен интензитет μ, а фотометър има предварително известна относителна стандартна грешка σ/μ=1%. Направени са n=10 измервания и е получена средно аритметична стойност m, която е с 2% по-висока (или с 20 % по-ниска) от стойността на предварително известния интензитет μ. Относителното средно квадратично отклонение на данните (относителната грешка на измерванията) е s/m=2%. Дали може да се смята с висока гаранция, (1) че източникът се е променил или (2) че само е увеличил (или само намалил) интензитета си? Колко измервания трябват за да се твърди с гаранция 99 %, че източникът е изменил интензитета си с 1%?

Правата задача в случая следва да отговори на въпроса „Дали измереното увеличение на интензитета е съществено, щото източникът да бъде обявен (1) за променлив, чрез двустранен тест или (2) за само усилен (или само отслабен), чрез едностранен тест?" За прилагане на гаусов тест трябва да смятаме, че стойността σ е предварително известна и постоянна. Тогава имаме ЦП $|z|$ = 2%/1% = 2. На фиг.1a над абсциса n=10 стойността z = 2 стои над всички прави (плътни и прекъснати). Отговорите са: Със статистическа гаранция над 99.9% източникът е (1) „променен", а също така е (2) „само усилен" (или „само отслабен"). За прилагане на стюдънтов тест, което се налага при неизвестна стойност на σ и/или при малко данни, както тук, следва да се използва s вместо σ. Тогава имаме ЦП $|t|$ = 2%/2% = 1. На фиг.1b над абсциса n=10 стойността t = 1 стои малко над долната плътна крива. Отговорите са: (1) Източникът е променлив с над 95% гаранция, но не е променлив с над 99% гаранция; (2) Ако се проверява дали източникът само се усилва (или само отслабва) такъв по-ограничен тип промяна следва да се приеме за значим с над 99% гаранция. Трябва да се отбележи, че за случая „само намаляване" се подхожда както при случая „само увеличаване", пак чрез прекъснатите линии на двете фигури. Също така, ако се очаква увеличаване, а се наблюдава намаляване, или обратно, подозираното увеличаван (или намаляване) следва да се обявят за статистически абсолютно непотвърдени. Винаги крайното решение взема потребителят.

Обратната задача в случая следва да отговори на въпроса „При стойностите σ/μ=1% и s/m=2% колко най-малко наблюдения n_{mim} са необходими за да се установи със 99 % гаранция, че източникът е променил интензитета си с амплитуда 1%?" При гаусовия тест (фиг.1*a*), за z=1%/1%=1, по средната плътна линия намираме $n_{\mathrm{mim}} \geq 7$, но този неголям брой измервания приемлив само ако стандартното отклонение е добре известно. По стюдънтовия тест (фиг.1*b*) за t=1%/2%=0.5 по горната плътна линия намираме значително по-голямата и статистически по-обоснована стойност $n_{\mathrm{mim}} \geq 30$. Трябва да се изтъкне, че на фиг.1, в log-log координати, гаусовите граници на интервали са прави линии. Обаче, стюдънтовите граници на интервали, които са статистически по-обосновани и по-строги, са криви линии. При много данни, напр. при n>50, стюдънтовите граници клонят асимптотично към гаусовите. Към малък брой данни, напр. n<10, стюдъънтовите криви вървят нагоре с увеличаваща се стръмност.

2. Сравняване на две средно аритметични стойности m_x и m_y чрез гаусов или стюдънтов $T(2n-2)$ тест

Нека имаме две извадки от нормално разпределени, а също така вътрешно и взаимно независими серии от случайни данни x_j, j=1, 2, … , n_x и y_j, j=1, 2, … , n_y, със съответни μ_x, μ_y, σ_x, σ_y, m_x, m_y, s_x и s_y. Дали (неизвестните) средни стойности μ_x и μ_y следва да се смятат за съществено различни? В този случай, за разлика от примерната задача в Раздел 1, стойностите μ_x, μ_y са неизвестни, което е обичайно в практиката. Подхождаме както в предния раздел.

В теорията и практиката при $m_x > m_y$, се използват следните целеви параметри

$$|z'| = (m_x - m_y)/(\sigma_{xy}/n^{1/2}) \text{ или } |t'| = (m_x - m_y)/(s_{xy}/n^{1/2}),$$

следващи съответно теоретично гаусово разпределение или стюдънтово разпределение със брой степени на свобода $f = n_x + n_y - 2$.

Тук се разглежда най-простият случай когато имаме $n_x = n_y = n$ и $\sigma_x = \sigma_y = \sigma$ или $s_x = s_y = s$. Тогава според статистическите теореми имаме $\sigma_{xy} = [(\sigma_x^2 + \sigma_y^2)/2]^{1/2} = \sigma/2^{1/2}$, както и, аналогично, $s_{xy} = s/2^{1/2}$. Следователно, съответните на така опростената задача ЦП ще имат „стандартния" вид

$$(2.0') \quad |z'| = (m_x - m_y)/(\sigma/(n/2)^{1/2}) \text{ или } |t'| = (m_x - m_y)/(s/(n/2)^{1/2}).$$

Първият от тях следва гаусово разпределение, а вторият – стюдънтово, $T(2n-2)$, с брой степени на свобода $f = 2n-2$.

И тук, както в Раздел 1, се въвеждат по-удобните за пресмятане на ум параметри

$$(2.0) \quad |z| = (m_x - m_y)/\sigma \text{ или } |t| = (m_x - m_y)/s.$$

Чрез тях наблюдаваната разлика (m_x - m_y)се изразява като относителна величина, в единици σ или s. Тогава съответните P % доверителни интервали са

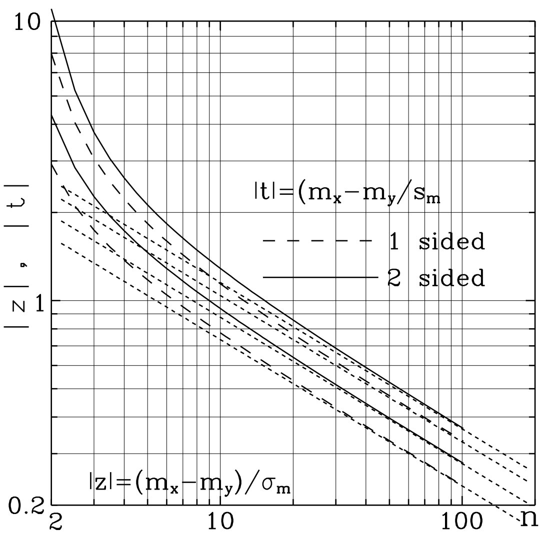
$$(2.1) \quad |z| < z_c(p)/(n/2)^{1/2} \text{ или } |t| < t_c(p;f)/(n/2)^{1/2}$$

за двата едностранни случаи и

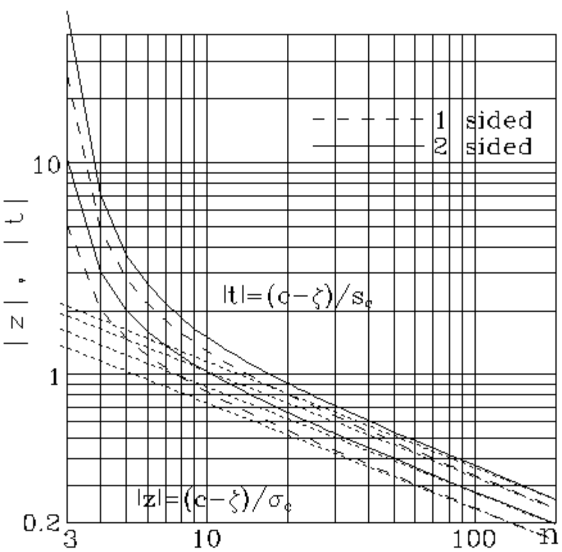
$$(2.2) \quad |z| < z_c(p/2)/(n/2)^{1/2} \text{ или } |t| < t_c(p/2;f)/(n/2)^{1/2}$$

за двустранния случай. Фигура 2 показва границите на интервалите (2.1) и (2.2) графично, чрез които се решават лесно права и обратна задача, както в Раздел 1.

Най-често срещаната задача е при заподозряна променливост на средната стойност μ на някаква величина да се установи евентуалната й промяна по две серии от измервания. Ако стойността на $|z|$ или $|t|$, изчислена според (2.0) се окаже над съответната линия на графиката за дадено n, то промяната (m_x - m_y) може да бъде смятана за статистически значима, с доверителна вероятност P.Доверителните интервали за разликата μ_x - μ_y са подобни на тези за m - μ в Раздел 1, но тук те са $2^{1/2}$ пъти по-широки ($2^{1/2}$ пъти по-високи на графиките).



Фиг.2. Доверителни граници за гаранции 95 и 99 % в зависимост от броя данни n за гаусово разпределения ЦП $|z| = (m_x - m_y)/s$ (2.0) за едностранен интервал (2.1) (прекъснати линии) и двустранен интервал (2.2) (плътни линии).



Фиг.3. Доверителни граници за гаранции 95 и 99 % в зависимост от броя данни n за стюдънтово разпределения ЦП $|t| = (c - \zeta)/s_c$, (3.5) за едностранен интервал (2.1) (прекъснати криви) и двустранен интер- вал (непрекъснати криви), по (3.6).

На двете диаграми пунктирните прави показват критичните граници за аналогичните гаусово разпределени целеви параметри, в които вместо *s* се използва *σ*. Гаусово разпределението ЦП не са включени в текста, понеже нямат практическо приложение. При търсенето на промяна на средната стойност неголям брой данни и по принцип следва да се използва средно-квадратичното отклонение *s*. Гаусовият тест може да се използва като крайно изключение само ако стандартната грешка на измерването *σ* е предварително добре известна.

3. Доверителни интервали и променливости на коефициенти на корелация и регресия чрез гаусов и стюдънтов *T*(*n*-2) тест

Нека имаме две извадки от нормално разпределени и вътрешно независими случайни данни x_j,и y_j, j=1, 2, … , n, със съответни μ_x, μ_y, σ_x, σ_y, m_x, m_y, s_x и s_y.

За разлика от Раздел 2, тук се предполага, че x_j и y_j данните може да са свързани чрез ненулев коефициент на корелация ρ и линейна регресия на Y спрямо X от вида $y_j = \beta_0 + \beta_1.x_j + \varepsilon_j$, с ненулев коефициент β_1. Данните x_j и y_j, се смятат за известни (измерени), а ρ, β_1 и β_0 подлежат на тестване за значимост (т.е. за отличие от 0) и/или за променливост (т.е. за сравняване с предварително известни оценки). Отклоненията от регресионната права ε_j се смятат за реализации на нормално разпределена случайна величини със средна стойност 0 и (неизвестно) стандартно отклонение σ_{yx}.

Нека r да бъде оценката на коефициента на корелация, b_0 и b_1 да бъдат оценките на регресионните коефициенти, получени по метода на най-малките квадрати, а s_r, s_{b0} и s_{b1} да бъдат съответните изчислени извадкови средно квадратични отклонения. По-нататък неравенствата касаят само стюдънтови разпределения, понеже стойностите на съответните стандартни отклонения σ_r, σ_{b0} и σ_{b1} на практика не могат да се смятат предварително известни и константни. Все пак, границите на интервалите с използване на гаусови разпределения са показани за сравнение на фиг.3 с пунктирни линии.

В този Раздел 3 тестовете за променливост трябва да отговарят по принцип на въпроса дали коефициентите r, b_0 и b_1 са съществено различни от съответни, предварително известни стойности ρ, β_0 и β_1. При това, предварителните стойности могат да се приемат за нули и тогава тестовете ще показват дали извадковите коефициенти r, b_0 и b_1 са статистически значими (гарантирано различни от нула), т.е. дали случайните величини X и Y са статистически зависими. Подхожда се както в предишните раздели.

Обичайно използваните ЦП, които тук не се налага да се подменят с по-прости, са

(3.1) $t_r = (\rho - r)/s_r$, $t_{b0} = (b_0 - \beta_0)/s_{b0}$ и $t_{b1} = (b_1 - \beta_1)/s_{b1}$.

Тези ЦП следват стюдънтово разпределение $T(n\text{-}2)$ с брой степени на свобода $f = n\text{-}2$.

За компактнист на формулите, нека да използваме означенията

$S_{xx} = \Sigma(x_j - m_x)^2$, $S_{yy} = \Sigma(y_j - m_y)^2$ и $S_{xy} = \Sigma(x_j - m_x).(y_j - m_y)$.

Тогава според метода на най-малките квадрати формулите за коефициентите са

(3.2) $r = S_{xy}/(S_{xx}.S_{yy})^{1/2}$, $b_1 = S_{xy}/S_{xx}$ и $b_0 = m_y - b_1.m_x$,

а остатъчното средно квадратично отклонение от регресионна права, съответно на неизвестното стандартно отклонение σ_{yx}, е

(3.3) $s_{yx} = [\Sigma(y_j - b_0 - b_1.x_j)^2/(n\text{-}2)]^{1/2} =$
$= [n.(S_{yy} - 2b_1S_{xy} + b_1^2S_{xx})/(n\text{-}2)]^{1/2}$.

Според теорията средно квадратичните отклонения на коефициентите са

(3.4') $s_r = [(1 - r_{xy}^2) / (n\text{-}2)]^{1/2}$,

(3.4.") $s_{b1} = s_{yx}/S_{xx}^{1/2}$ и

(3.4."') $s_{b0} = s_{yx}\ [(1/n) + (m_x^2/S_{xx})]^{1/2}$.

За простота ще означим трите ЦП, въведени в (3.1), обобщено като

(3.5) $|t| = (c - \zeta)/s_c$,

където c е оцененият по данните коефициент, ζ е проверяваната стойност на коефициента (или 0), а s_c е средно-квадратичното отклонение на коефициента от вида (3.4), оценено по данните. Тогава съответните двустранен и едностранен интервали се записват общо като

(3.6) $|t| = < t_c(p/2;f)$ и $|t| < t_c(p;f)$.

Съответната тестова процедура включва изчисляване на необходимия ЦП по (3.1) и сравняване на резултата с височината на съответната крива над съответното n по фиг.3.

4. Доверителни нтервали и променливост на стандартното отклонение чрез пирсънов $\chi^2(n\text{-}1)$ и фишеров $F(n\text{-}1)$ тестове

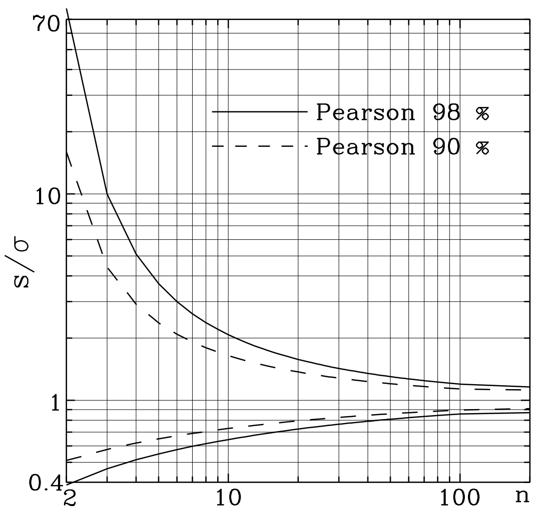
Доверителният интервал за P % статистическа гаранция за (неизвестно) стандартно отклонение σ се определя конвенционално чрез ЦП

(4.0) $u^2 = s^2/[(\sigma^2/(n\text{-}1)]$,

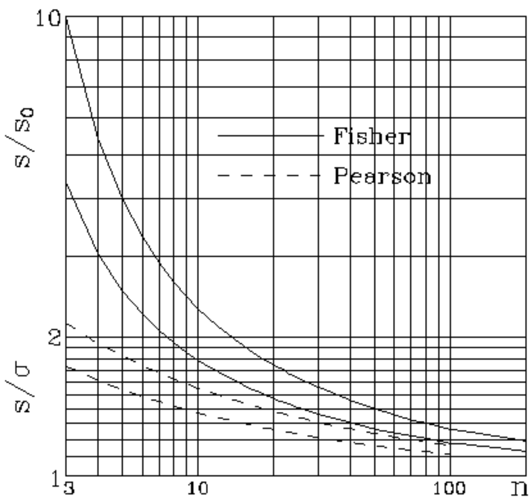
който има пирсъново $\chi^2(n\text{-}1)$ разпределение с $f = n - 1$ степени на свобода. За разлика от предните раздели, пирсъновото разпределение има асиметрична форма. Съответните двустранни и едностранни доверителни интервали за σ^2, са

(4.1') $(n\text{-}1).s^2/\chi^2(1\text{-}p/2;f) < \sigma^2 < (n\text{-}1).s^2/\ \chi^2(p/2;f)$ и

(4.1") $\sigma^2 < (n\text{-}1).s^2/\ \chi^2(p;f)$.



Фиг.4. Граници на двустранните (асиме трични) доверителни интервали за целе вия параметър u = s/σ (4.2), с пирсъново разпределение, при P = 90 % (прекъснати криви) и P = 98 % (плътни криви) в зави симост от броя данни n.



Фиг.5. Граници на 95% и 98 % едно странни интервали за u = s/σ (4.2), с пирсъново разпределение (прекъсна ти криви) и съответни 95% и 99 % ед нотранни интервали за u = s/s_0 (4.3), с фишерово разпределение (плътни криви).

Тук отново въвеждаме по-удобен ЦП, u=s/σ, който изразява s в единици $\underline{\sigma}$, както и означението $\chi = (\chi^2)^{1/2}$. Тогава двустранните и едностранни интервали за ЦП s/σ стават

(4.2') $(n\text{-}1)^{1/2}/\chi(p/2;f) < s/\sigma < (n\text{-}1)^{1/2}/\ \chi^2(1\text{-}p/2;f)$ и

(4.2") $\sigma/s < (n\text{-}1)/\ \chi(1\text{-}p;f)$.

Фигура 4 показва двустранните критични граници за 95% и 98% статистическа гаранция в зависимост от n. Съответните едностранни граници са представени с прекъс-нати криви на Фиг.5.

Както и в Раздел 1 и Раздел 3, графиките на фиг.4 могат да се използват на първо място за определяне на доверителния интервал на σ според s, както и за оценяване на необходимия брой данни n за достигане на желана статистическа гаранция на резултата. На второ място те могат да се използват като тест за променливост.

Като пример, нека първо смятаме, че стандартното отклонение σ е добре известно и да имаме 10 данни, за които s/σ = 2. Това е случай на заподозряно увеличаване на предварително известно стандартно отклонение. На фиг.4, над n=10, се вижда, че това двукратно нарастване на s спрямо σ е статистически значимо с 95 % гаранция, но е статистически незначимо с 98% гаранция. По аналогия, ако сме получили s/σ = 0.7, това е случай на заподозряно намаляване на предварително известно стандартно отклонение. Тогава чрез долните криви линии на фиг.4 следва да отбележим, че изменението на s спрямо σ е статистически значимо с 95 % гаранция, но е статистически незначимо с 98% гаранция. В крайна сметка решава потребителят.

На практика стандартното отклонение σ е променливо или известно с ниска точност. Тогава тестът за променливост на σ може да се базира на две оценки на средно квадратичното отклонение, s_0 и s, чрез конвенционалния ЦП $v^2 = s^2/[(s_0^2/(n\text{-}1)]$, за $s^2>s_0^2$. Това отношение следва фишерово разпределение с f=n-1 степени на свобода, означавано като $F(n\text{-}1)$. Съответните доверителни интервали за v^2 са $1/F(p/2;f) < v^2 < F(p/2;f)$ (двустранен, асиметричен) и $v^2 < F(p;f)$ (едностранен). Поради предварителното условие $s^2>s_0^2$ дясната граница на двустранния интервал е достатъчна за статистическия тест относно променливост на σ.

И тук ние въвеждаме по-удобен ЦП $v=s/s_0$, за $s>s_0$, който изразява s в единици s_0. Тогава горната граница на двустранния доверителен интервал става

(4.3) $s/s_0 < F(p/2;f)^{1/2}$.

Фигура 5 показва границите на този интервал в зависимост от n, за 95% и 99% статистическа гаранция. Графиките на фиг.5 могат да бъдат използвани за установяване на променливост на σ така както на μ в Раздел 1 или както в предишния пример тук.

5. Заключение

Графиките на тестовете за променливост на μ и σ бяха използвани успешно за установяване на съществени краткосрочни изменения на блясъка на една катаклизмична звезда (Georgiev, 2012). Заедно с това нагледното представяне на тестовете се оказа удобно в работата на автора и при преподаването на основите на статистиката.

ЛИТЕРАТУРА

Cox D.R, Hinkley D.V., 1974, Theoretical Statistics, Charman and Hall (in Russian: 1978)

Georgiev, T.B., 2012. Bulg. Astron. J. 18/3, 77-105

Georgiev, T.B., 2014. Bulg. Astron. J. 20, 14-25

Fisher R.A., 1925, Statistical Methods for Research Workers, Edinburgh, Oliver and Boyd

Fisher R.A., Yates F., 1963. Statistical Tables for Biological, Agricultural and Medical Research, 6th ed., Edinburgh

Neyman, J., Pearson, E.S., 1933, On the Problem of the Most Efficient, Tests of Statistical Hypotheses, Phil. Trans. R. Soc., Series A 231

Tucker, H., 1962, An introduction to Probability and Mathematical Statistics, New Yurk, Acasemic Press

Zacks, S., 1974, The Theory of the Statistical Inference, John Wiesly (in Russian: 1975)