Graphic presentation of the simplest statistical tests

Tsvetan B. Georgiev[1,2]
[1] New Bulgarian University, BG-1618 Sofia
[2] Institute of Astronomy and NAO, BG-1784 Sofia
tsgeorg@astro.bas.bg
(Submitted on 28.07.2013; Accepted on 03.09.2013)

Abstract. This paper presents graphically well known tests about change of population mean and standard deviation, about comparison of population means and standard deviations, as well as about significance of correlation and regression coefficients. The critical bounds and criteria for variability with statistical guaranty $P = 95\%$ and $P = 99\%$ are presented as dependences on the data number n. The graphs further give fast visual solutions of the direct problem (estimation of confidence interval for specified P and n), as well of the reverse problem (estimation of n, which is necessary for achieving a desired statistical guaranty of the result). The aim of the work is to present the simplest statistical tests in a comprehensible and convenient graphs, which will be always at hand. The graphs may be useful in the investigations of time series in astronomy, geophysics, ecology etc., as well as in the education.
Key words: Statistical inferences

Introduction

Any simple statistical test about variability checks the belonging of a suitable parameter, calculated from the data and called score, to confidence interval, corresponding to preliminary given (high) probability P (f.e. $P = 95\%$). The respective (low) probability $p = 1 - P$, called confidence level or error level, is used more often. If the score fails inside the interval, null hypothesis (H_0) is accepted, i.e. the check parameter (e.g. μ) may be considered constant. If the score fails outside the interval, alternative hypothesis (H_A) is accepted, i.e. the check parameter may be considered changed.

The theory and application of the statistical inferences based on confidence intervals is elaborated and described by Fisher (1925), Neyman & Pearson (1933), Tucker (1962), Zaks (1971), Cox & Hinkley (1974) etc.,as well as in the contemporary manuals.

In practice the user find the critical bounds of the confidence interval by preliminary calculated tables or by computer programs. Here we propose graphs of the critical bounds in dependence on the data number n. These graphs hold at least tree advantages. First, they are based on handy and easy calculated scores, specified here essentially for the practice. Second, they allow fast visual decision of the direct problem (check of null hypothesis by given n), as well as the reverse problem (estimation of necessary n for specified null hypothesis). Third, they provide decisions of both problems looking on a few graphs of critical bounds simultaneously. The disadvantage of these graphs is the low accuracy of the solution: only two significant digits.

Hereafter we suppose a normally distributed random variable X, presented by a sample of n its mutually independent realizations (data) $x_1, x_2, ..., x_n$. The main statistical parameters of the sample are the population mean μ, the population standard deviation σ, the sample mean m (average of the data) and the sample standard deviation s (mean square deviation of the data from m). The usual estimators are $m = (\varSigma x_j)/n$ and $s = [\varSigma \varDelta x_j/(n-1)]^{1/2}$, with

$\Delta x_j = x_j - m$. Hereafter the summing is executing always from 1 to n. Later we will suppose also a second such random variable, Y.

Two kinds of confidence intervals for the estimated parameter (e.g. μ) are of interest: two-sided, $(\delta_1(p/2), \delta_2(p/2))$ and one-sided, $(\infty, \delta(p))$ or $(delta(p), \infty)$. The first one accounts for possible increase or decrease and the second one concerns increase only (or, depending on the task, decrease only). Note that in all cases the level of error is $p = 1 - P$. Further, when the score parameter has symmetric random distribution (of Gauss-type or of Student-type), the two-sided interval is presented simply as $[-\delta(p/2), \delta(p/2))]$. Generally, the statistical test based on two-sided interval is harder.

In practice the dependence of the interval bounds (critical bounds) on the number of data n is the most essential and that's reason to build such kinds of graphs. The applying of the graphs includes calculation of the handy score value and compare it with the critical bounds over the data number n. When the score occurs under the bound, null hypothesis is accepted, i.e. the checked parameter is considered to be unchanged with $P\%$ statistical guaranty. Otherwise, alternative hypothesis is accepted, i.e. the checked parameter is considered to be changed with $P\%$ statistical guaranty.

This paper gives consecutively graphs for (1) intervals and variability of the population mean by Gauss and Student tests, (2) comparison of two sample means through Gauss and Student tests, (3) intervals and variability of the correlation and regression coefficients through Gauss and Student tests, (4) intervals and variability of the standard deviation through Pearson and Fisher tests.

1. Intervals and variability for the population mean μ through Gauss test and Student $T(n-1)$ test

Any sample value $x_j (j = 1, 2, \ldots, n)$ of mutually independent and normally distributed random data (results of measurements) may be scaled to the (standard) Gauss distribution by the substitution $z_j = (x_j - \mu)/\sigma$, as well as to the Student distribution $T(n-1)$ with degrees of freedom $f = n - 1$ by the substitution $t_j = (x_j - m)/s$.

That is why the scores (suitable random variables, calculated from the data), that are widely used in the statistical inferences about μ are:

(1.0) $z' = (m - \mu)/(\sigma/\sqrt{n})$ or $t' = (m - \mu)/(s/\sqrt{n})$.

Here μ, σ are (unknown) constants, but s and m are calculated from n data. According to the statistical theorems, the score z' follows Gauss distribution with standard deviation $\sigma/\sqrt{n}$ and t' follows Student T-distribution with degrees of freedom $f = n - 1$ with standard deviations $s/\sqrt{n}$.

Then any $P\%$ confidence intervals for z' or t' may be presented simply as $|z'| < z_c(p/2)$ or $|t'| < t_c(p/2; f)$ (two-sided) and $|z'| < z_c(p/2)$ or $|t'| < t_c(p/2; f)$ (as well as $|z'| > z_c(p/2)$ or $|t'| > t_c(p/2; f)$) (one-sided).

Here we propose more handy score values, (1.1) $z = z'/\sqrt{n} = (m - \mu)/\sigma$ or $t = t'/\sqrt{n} = (m - \mu)/s$,

where the difference $(m - \mu)$ again is expressed in the units of σ or s, but the factor $\sqrt{n}$ is applied as divisor of the modified interval bounds.

So, the modified confidence intervals for the scores z and t become

(1.2') $|z| < z_c(p/2)/\sqrt{n}$ or $|t| < t_c(p/2; f)/\sqrt{n}$,

two-sided, and

(1.2") $|z| < z_c(p)/\sqrt{n}$ or $|t| < t_c(p; f)/\sqrt{n}$,

one-sided.

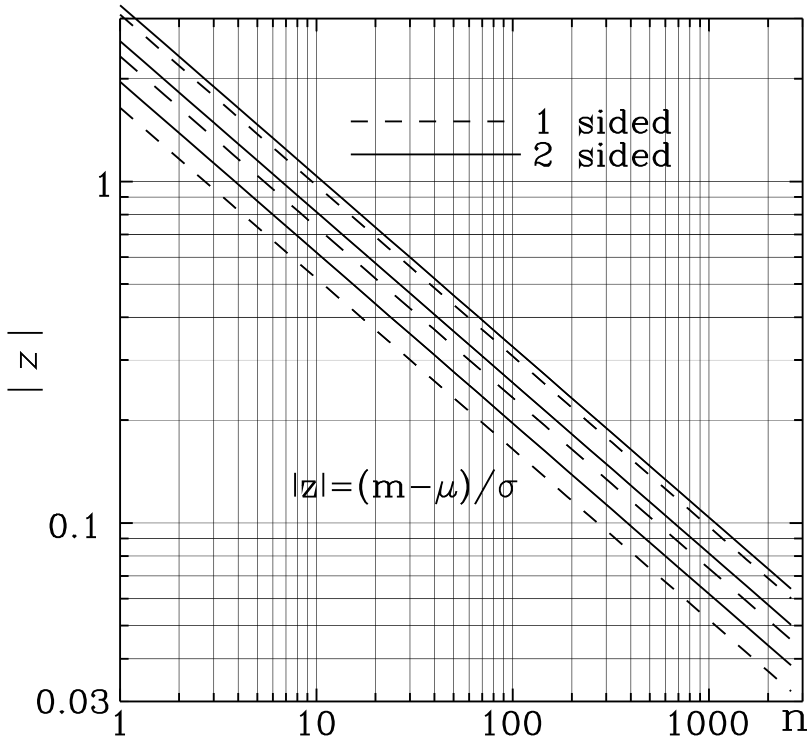

Fig. 1. Confidence intervals for Gauss tests, based on score $|z| = (m - \mu)/\sigma$, with bounds $z_c(p/2)/\sqrt{n}$ (two-sided, solid lines) and $z_c(p)/\sqrt{n}$ (one-sided, dashed lines),in dependence on data number n. The bounds correspond to statistical guaranty of 95, 99 and 99.9 %.

The distributions of Gauss and Student have symmetrical shapes and by this reason only the bounds of the intervals (1.2') and (1.2") are shown in the graphs. The bounds for the two-sided case are presented by solid curves and these for the on-sided case are given by dashed curves.

Figure 1 presents large scale critical bounds for the Gauss test with score z in respect to σ and in dependence on n. The bounds corresponds to cumulative percentages of 95, 99 and 99,9 % with respective quantile values of $z_c(p)$ equal to 1.645, 2.326 and 3.081 (one-sided case), as well as $z_c(p/2)$, equal to 1.96, 2.576 and 3.283 (two-sided case).

Figure 2 gives critical bounds for the Student $T(n-1)$ test in respect to s and in dependence on n. Tables of the quantile values for one-sided or two-sided intervals may be find in any manual after Fisher & Yates (1963). It is seen that at large number of data, f.e. $n > 50$, the Student bounds come close to the Gauss bounds (short dashed lines), but when the data number is low, the Student bounds must be used.

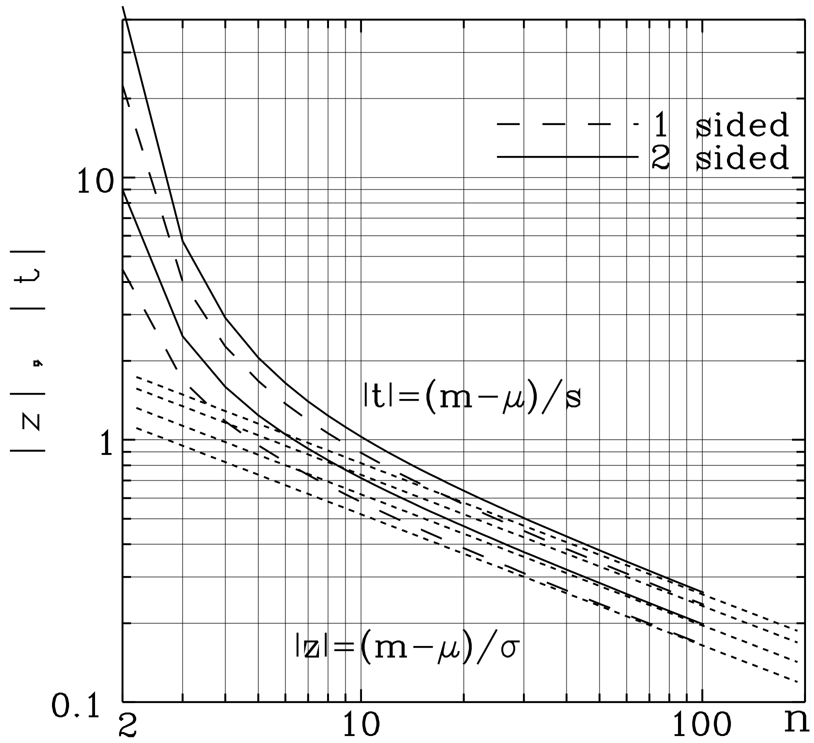

Fig. 2. Confidence intervals for Student $T(n-1)$ tests, based on the score $t = (m-\mu)/s$, with bounds $\pm t_c(p/2; f)/\sqrt{n}$ (two-sided, solid lines) and $t_c(p; f)/\sqrt{n}$ (one-sided, dashed lines), in dependence on the data number n. The bounds correspond to statistical guaranty of 95 and 99 %. The dotted lines show the bounds for the score $z = (m-\mu)/\sigma$, which is usable in in Fig.1.

The graphs in Fig.1 and Fig.2 may be used firstly for deriving of critical bounds of P% confidence intervals for unknown population mean μ through z or t scores (1.1).

The main purposes of the graphs in Fig.1 and Fig.2 are to ensure easy test about suspected change of a known population mean μ, as follows.

To check a possible change of μ, one calculates the value of $|z|$ or $|t|$. If the result exceeds the line of the chosen P% interval bound over n, the preliminary known μ may be announced as already changed and replaced by m within respective statistical guaranty P. Two-sided interval corresponds to principal change (increase or decrease) and one-sided interval corresponds to increase only or decrease only.

The graphs in Fig.1 and Fig. may be used also for fast solution of the reverse problem, i.e. deriving the number of the data n that are necessary for achieving of desired confidence interval for variability validation.

The graphs may be used also for determination of the P% confidence interval of the unknown μ using the interval bound z_l or t_l over n and the inequalities (1.1) and (1.2'). The bounds of the two-sided confidence intervals may be presented shortly as

(1.3') $m \pm \sigma.z_l$ or $m \pm s.t_l$,

and the bounds for one-sided confidence intervals are

(1.3") $m + \sigma.z_l$ or $m + s.t_l$.

2. Comparison of two sample means m_x and m_y through Gauss test and Student $T(2n-2)$ test

Suppose we have two samples of normal distributed and mutually independent random data: $x_j, j = 1, 2, \ldots, n_x$, with respective μ_x, σ_x, m_x and s_x, as well as $y_j, j = 1, 2, \ldots, n_y$, with respective μ_y, σ_y, m_y and s_y. In the theory the scores $z = (m_x - m_y)/(\sigma_{xy}/\sqrt{n})$ or $t = (m_x - m_y)/(s_{xy}/\sqrt{n})$ (with $m_x > m_y$) are used. The first of them follows Gauss distribution and the second one has Student distribution with degrees of freedom $f = n_x + n_y - 2$.

Here we concentrate on the simplest case when $n_x = n_y$ =n and $\sigma_x = \sigma_y = \sigma$ or $s_x = s_y = s$. Then following the statistical theorems we have $\sigma_{xy} = [(\sigma_x^2 + \sigma_y^2)/2]^{1/2} = \sigma/\sqrt{2}$, as well as $s_{xy} = s/\sqrt{2}$. Further the scores becomes

(2.0) $z' = (m_x - m_y)/(\sigma/\sqrt{n/2})$ or $t' = (m_x - m_y)/(s/\sqrt{n/2})$,

where the first of them follows Gauss distribution and the second one has Student distribution with degrees of freedom $f = 2n - 2$, ($T(2n-2)$), are usually used.

Here, like in Chapter 1, we introduce more handy scores:

(2.1) $z = z'/\sqrt{n/2} = (m_x - m_y)/\sigma$ or $t == t//\sqrt{n/2} = (m_x - m_y)/s$

with respective critical bounds of P% confidence intervals

(2.2') $|z| < z_c(p/2)/\sqrt{n/2}$ or $|t| < t_c(p/2; f)/\sqrt{n/2}$,

two-sided and

(2.2") $z < z_c(p)/\sqrt{n/2}$ or $t < t_c(p; f)/\sqrt{n/2}$

one-sided.

Figure 3 show the critical bounds of these intervals graphically. It provides visual solution of the problems like these accounted in the end of Chapter 1.

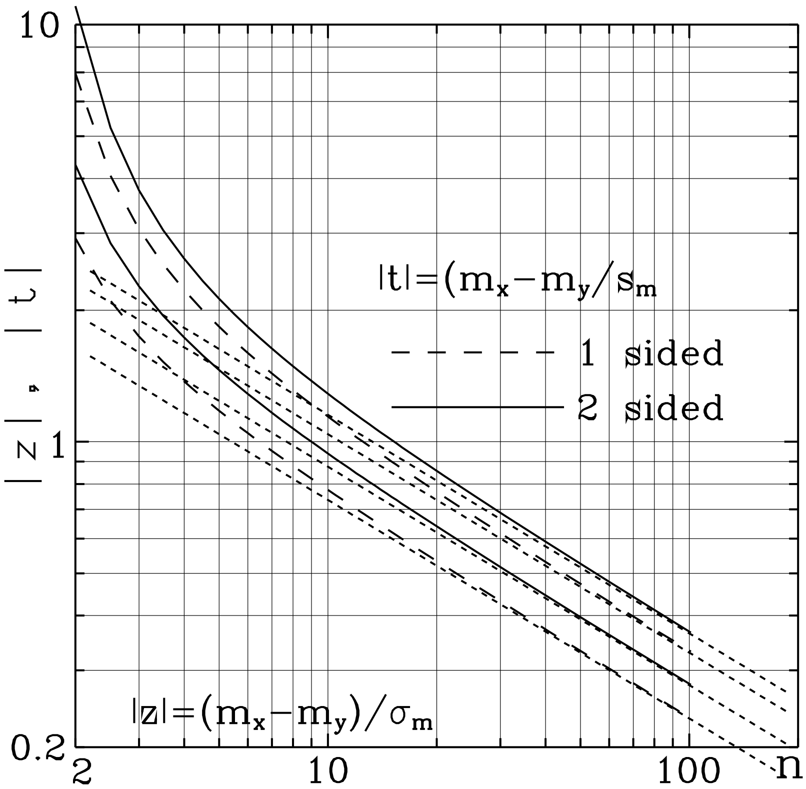

Fig. 3. Confidence intervals for the Student $T(2n-2)$ test based on the score $t = (m_x - m_y)/s$ with bounds $t_c(p/2; f)/\sqrt{n/2}$ (two-sided, solid lines) and $t_c(p; f)/\sqrt{n/2}$ (one-sided, dashed lines) for 95 and 99 % guaranty, in dependence on data number n. Dotted lines show the respective critical bounds for Gauss distributed score $z = (m_x - m_y)/\sigma$.

One important problem in the practice is the establishment of variability of the population mean μ by means of two samples of data with sample means m_x and m_y . Then if the value of z or t in (2.1) occurs above the specified curve above the data number n, the difference $(m_x - m_y)$, as well as the variability of μ may be considered to be significant with respective statistical guaranty.

Note that the confidence intervals for the difference $m_x - m_y$ are similar to these given in Chapter 1, however here they are $\sqrt{2}$ times higher. Again at large n the Student curves tend to Gauss lines and again the Gauss test is not recommendable, because the standard deviation σ is poorly known in the practice.

3. Intervals and variability of the correlation and regression coefficients through Gauss test and Student $T(n-2)$ test

Suppose n pairs of random data $(x_j, y_j), j = 1, 2, \ldots, n$, with statistical parameters $\mu_x, \sigma_x, m_x, s_x$, as well as $\mu_y, \sigma_y, m_y, s_y$. Suppose also the correlation coefficient ρ and regression model $y_j = \beta_0 + \beta_1.x_j + \epsilon$, where x_j are known, y_j are observed, but ρ, β_1 and β_0 are unknown. The deviations (Y-errors) ϵ_J are considered to be independent and normally distributed random variables with population mean 0 and (unknown) standard deviation σ_{yx}.

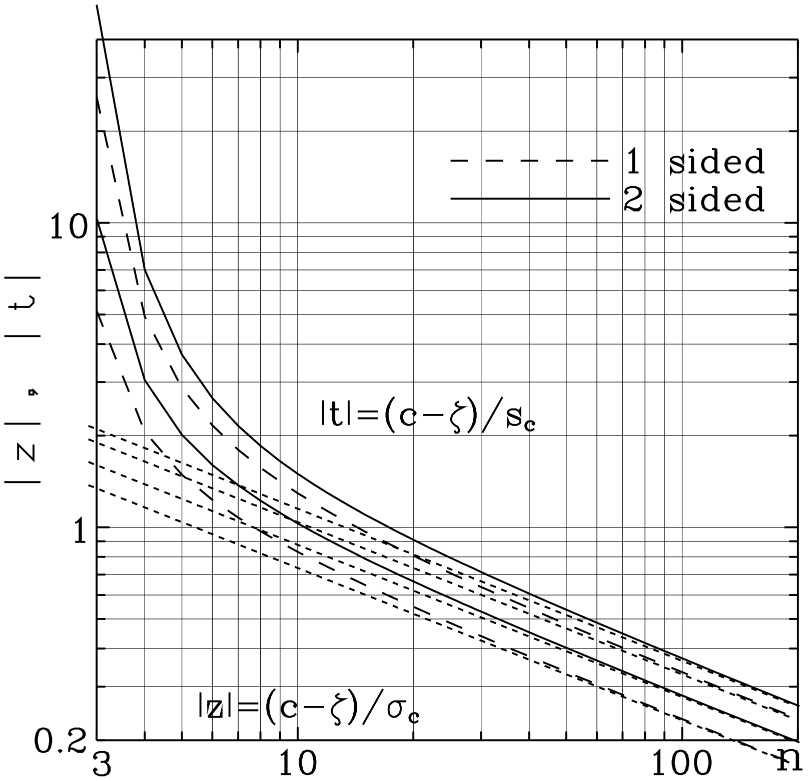

Fig. 4. Confidence intervals for the Student $T(n-2)$ tests based on t-scores (3.1), presented commonly as (3.5), with bounds (3.6), for 95 % and 99 % statistical guaranty, in dependence on the data number n. The solid curves correspond to two-sided tests and the dashed ones - to one-sided tests. Short-dashed lines show the respective bounds for the Gauss distributed z-scores, analogous of t-scores, but with σ instead s in every case (not included in the text).

Suppose also r is the sample correlation coefficient while b_0 and b_1 are the least-squares estimators of the regression coefficients with sample standard deviations s_r, s_{b0} and s_{b1}.

Hereafter the inequalities are presented only for scores for Student tests, because the value of the respective σ is unknown in principle. Though, the critical bounds with use of σ and Gauss distribution are shown for comparison in the graphs in Fig.4

In this Chapter 3 the tests have to answer whether the coefficients r, b_0 and b_1 are statistically equal to some specified (preliminary known) values, f.e. ρ, β_0 and β_1, respectively. The last values is often taken to be 0 and then the test show whether random variables X and Y may be consider to be related.

So, the commonly used score values are the relative differences

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

They follow $T(n-2)$ Student distributions with degrees of freedom $f = n-2$.

Further the correlation and regression coefficients are defined by use of the denotations $S_{xx} = \Sigma(x_j - m_x)^2$, $S_{yy} = \Sigma(y_j - m_y)^2$, as well as $S_{xy} = \Sigma(x_j - m_x)(y_j - m_y)$. Then the coefficients take the forms

(3.2') $r = S_{xy}/(S_{xx}.S_{yy})^{1/2}$,

(3.2") $b_1 = S_{xy}/S_{xx}$ and

(3.2"') $b_0 = m_y - b_1.m_x$.

The residual mean square deviation from the regression line is defined as

(3.3) $s_{yx} = [\Sigma(y_j - b_0 - b_1.x_j)^2/(n-2)]^{1/2}$.

The sample standard deviations of s_r, s_{b0} and s_{b1} are defined as

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

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

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

For simplification of the further presentation we well denote all scores, defined in (3.1) commonly as

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

where c is the derived (estimated from the data) coefficient, ζ is the checked preliminary known value of the coefficient (or 0) and s_c is the sample standard deviation of the coefficient.

Then the respective two-sided and one sided intervals as

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

Figure 4 shows bounds of the intervals (3.6), corresponding to the scores (3.1), presented commonly as (3.5). In this cases the bounds are just the Student curves of the type $T(n-2)$.

The testing procedure includes calculation of the score defined in (3.1) and comparing the result with the curves of the graph above the specified n. If the score occurs above the specified curve, the respective criterion for statistical significance of the coefficient (or its distinction from 0) is fulfilled.

The confidence intervals for the parameters ρ, β_0 and β_1 may be derived like in the end of Chapter 1.

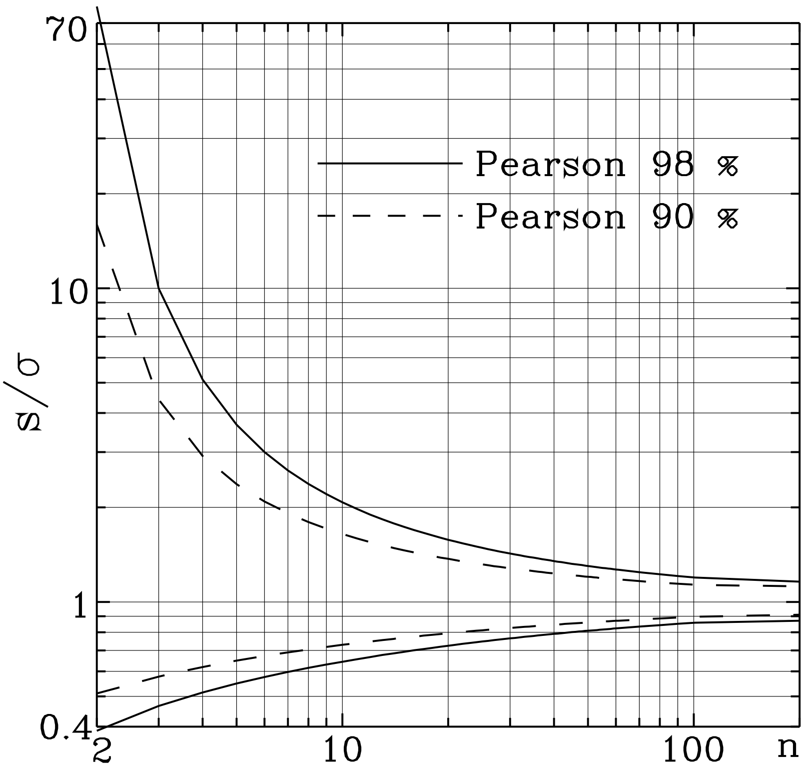

Fig. 5. Two-sided confidence intervals for the Pearson test based on the score s/σ with bounds given in (4.2') for $P = 98\%$ (solid curves) and $P = 90\%$ (dashed curves), in dependence on the data number n. Upper bounds of the one-sided intervals are presented in Fig.6.

4. Intervals and variability of the standard deviation through Pearson $\chi^2(n-1)$ and Fisher $F(n-1)$ tests

The $P\%$ confidence intervals for (unknown) σ is determining through the convenient score $u^2 = s^2/[\sigma^2/(n-1)]$, that follows Pearson χ^2-distribution with degrees of freedom $f = n-1$. This distribution has asymmetric shape.

The respective two-sided and one-sided confidence intervals for σ^2, given in the literature are

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

and

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

Here we again introduce more handy score $u = s/\sigma$, that expresses s in σ-units, and denote $\chi = (\chi^2)^{1/2}$. Then the two-sided and one-sided confidence intervals for u are

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

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

Figure 5 presents the respective two-sided critical bounds for 90 % and 98 % statistical guaranty in dependence on n. Note that ordinate axis in Fig.5 is presented after division by $(n-1)^{1/2}$. One-sided bounds for 90 % and 98 % statistical guaranty are presented by dashed curves in Fig.6.

The graphs in Fig.5 may be used for fast determination of the confidence interval of σ, as well as for estimation of n that is necessary for desired statistical guaranty of the result.

As an example, let us firstly consider σ to be preliminary well known. Suppose we have $n = 10$ new measurements with $s/\sigma = 2$. We may see in Fig.5 that this 2-fold increase of s in respect to σ is significant with 95 % guaranty, but it is not significant with 98 % guaranty. Secondly, if we hsve the same conditionsq but we derive $s/\sigma = 0.7$, we must conclude that the decrease of of s in respect to σ is significant with 95 % guaranty, but it is not significant with 98 % guaranty.

In practice the value of σ is usually poor known or in principle variable. Then the confidence intervals and tests about variability of σ must be based of two estimations of σ, s_0 and s, by application of Fisher distribution.

Suppose a preliminary independent estimation s_0 of σ, taken from $n_0 = n$ data. Then the P% confidence intervals for σ is determining through convenient score $v^2 = s^2/s_0^2$, for $s^2 > s_0^2$. This ratio follows Fisher $F(n-1)$ distribution with degrees of freedom $f = n-1$.

The Fisher distribution, like the Pearson distribution, has asymmetric shape with confidence intervals for v^2 given by the inequalities $1/F(p/2;f) < v^2 < F(p/2;f)$ (two-sided) and $v^2 < F(p;f)$ (one-sided). Because of the preliminary condition for $s^2 > s_0^2$ only the interval $v^2 < F(p/2;f)$, that is statistical criterion for change of σ, is used in the practice.

We introduce more handy score ratio $v = s/s_0$, for $s > s_0$, that expresses s in s_0-units. Then the upper limit of the confidence interval for s/s_0 is

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

Figure 6 shows the bounds of such intervals in dependence on n, accounting $f = n-1$, for 95 % and 99 % statistical guaranty.

The graphs in Fig.6 may be used for fast establishment of variability of σ like that of μ in Chapter 1, as well as like in the first example in this Chapter 4.

Conclusion

The graphs of the tests about the change of μ and σ has been already applied successfully for establishing of the stellar flickering (Georgiev, 2012), based on observations in the Rozhen NAO, published by Zamanov (2011) and Stoyanov (2012). The graphs occur very useful also in the lecturer practice of the author.

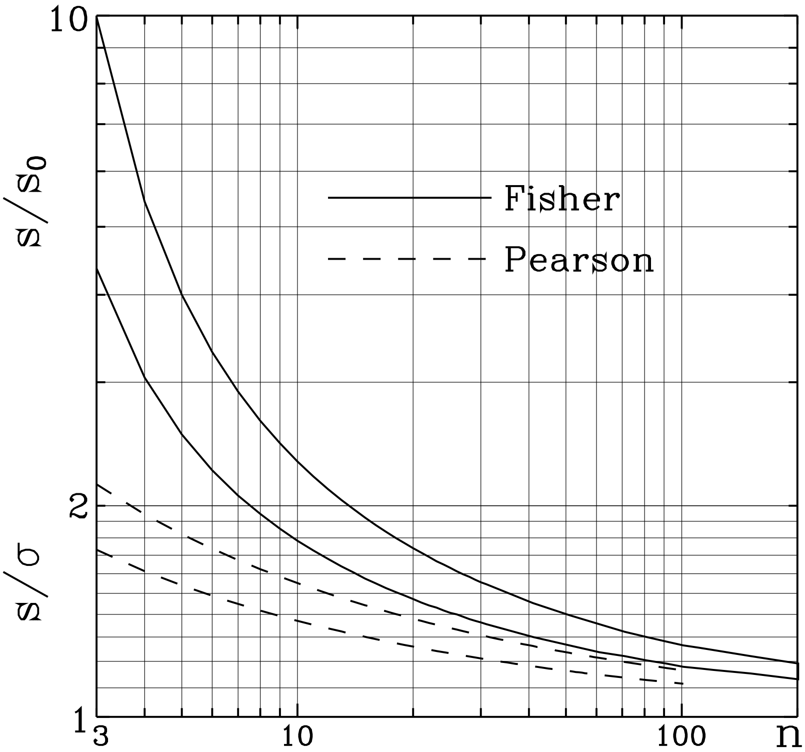

Fig. 6. Solid curves: upper bounds of the two-sided confidence intervals for the Fisher test, based on the score s/s_0 (for $s > s_0$) with bounds given in (4.3) for 95 % and 99 % statistical guaranty, in dependence on the data number n. Dashed curves: upper bounds of the one-sided confidence intervals for the Pearson test, based on the score s/σ (for $s > \sigma$) with bounds given in (4.2") for 95 % and 98 % statistical guaranty, in dependence on the data number n.

Acknowledgements

The author thanks Dr. Boris Komitov for the useful discussions, as well as the anonymous referee for the valuable recommendations about this paper.

References

Georgiev, T.B., 2012, *Bulg. Astron. J* 18/3, 77-105
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
Cox D.R, Hinkley D.V., 1974, *Theoretical Statistics*, Charman and Hall (in Russian: 1978)
Stoyanov, K., 2012,*Bulg. Astron. J* 18/2, 63-66
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)
Zamanov R.,2011, http://195.97.237.247