Two elementary statistical coverage problems

Autor(en):   **Pestman, Wiebe R.**

Objekttyp: **Article**

❙Elemente der Mathematik

Two elementary statistical coverage problems

Wiebe R. Pestman

Wiebe R. Pestman obtained his doctor's degree in mathematics at the University of Groningen. He is currently lecturer and researcher at the University Medical Center in Utrecht. His mathematical interests are in functional analysis, probability and statistics, operator algebras, and harmonic analysis.

1 Formulating the problems and setting the notations

Suppose a researcher draws a sample $X_1, X_2, \ldots, X_n$, from some population, and computes the corresponding sample mean. Assume that the population has a Gaussian probability distribution. A second researcher draws, independently, a sample $Y_1, Y_2, \ldots, Y_n$ from the same population. Then what is the probability that the 95% confidence interval for the population mean, generated by $Y_1, Y_2, \ldots, Y_n$, covers the mean of $X_1, X_2, \ldots, X_n$? Below this probability will be denoted by P_n. As a second coverage problem, what is the probability that the 95% confidence interval generated by the $X_1, X_2, \ldots, X_n$ and that by the $Y_1, Y_2, \ldots, Y_n$ are disjoint? Below this probability will be denoted by Q_n. The aim of this paper is to get an expression for the values that can be taken on by P_n and Q_n. The sample means of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ will be denoted by $\bar{X}$ and $\bar{Y}$, respectively. The sample variances will be denoted as S_X^2 and S_Y^2. The Greek capital Φ will stand for the cumulative distribution function of a standard Gaussian distribution, that is to say, a Gaussian distribution with mean 0 and standard deviation 1.

Liegen zwei voneinander unabhängige Stichproben gleicher Größe aus derselben Grundgesamtheit mit dem Erwartungswert μ vor, so kann man nach der Wahrscheinlichkeit dafür fragen, dass ein aus der zweiten Stichprobe erzeugtes Konfidenzintervall für μ das arithmetische Mittel der ersten Stichprobe enthält. Ebenso kann man fragen, mit welcher Wahrscheinlichkeit zwei Konfidenzintervalle aus den beiden Stichproben disjunkt sind. In der vorliegenden Arbeit werden diese Wahrscheinlichkeiten für normalverteilte Grundgesamtheiten nicht nur asymptotisch, sondern auch für kleine Stichproben berechnet. Die Ergebnisse werden im Kontext von statistischen Hypothesentests diskutiert, bei denen die Überlappung von Konfidenzintervallen als Kriterium zur Entscheidung herangezogen wird.

2 The family of t-distributions

To define general t-distributions, suppose that Z is a random variable with a standard Gaussian distribution and that V is a random variable with a χ^2-distribution with n degrees of freedom. Assume that Z and V are statistically independent. Then, by definition, the t-distribution with n degrees of freedom and non-centrality parameter α is the probability distribution of the random variable

$$\frac{Z + \alpha}{\sqrt{V/n}}.$$

The cumulative distribution function of this distribution will be denoted as F_n^α. In cases where $\alpha = 0$ one talks about *central* t-distributions or simply about t-distributions. Now, to connect the above to sampling, let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be two independent samples from a population with a standard Gaussian distribution. Then the random variables $\sqrt{n}\,\bar{X}$ and $\sqrt{n}\,\bar{Y}$ are independent and they both have a standard Gaussian distribution. It follows from this that $\sqrt{n/2}\,(\bar{X} - \bar{Y})$ also has a standard Gaussian distribution. Moreover $(n-1)S_Y^2$ has a χ^2-distribution with $n-1$ degrees of freedom (see [5]). Besides this (see [5], [6]), the variables $(n-1)S_Y^2$ and $\sqrt{n/2}\,(\bar{X} - \bar{Y})$ are statistically independent. It follows that the quantity

$$\frac{\sqrt{n/2}\,(\bar{X} - \bar{Y}) + \alpha}{S_Y} = \frac{\sqrt{n/2}\,(\bar{X} - \bar{Y}) + \alpha}{\sqrt{(n-1)S_Y^2/(n-1)}} \tag{1}$$

has a t-distribution with $n - 1$ degrees of freedom and non-centrality parameter α. The following sections will heavily lean on this observation.

3 A solution to the first coverage problem

To determine the probabilities P_n, first note that, without loss of generality, it may be assumed that the samples are drawn from a standard Gaussian distribution. Namely, there is coverage under $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ if and only if one has coverage with respect to the standardized samples. For this reason it will be systematically assumed that the two samples are drawn from a standard Gaussian distribution. To solve the problem, recall that the endpoints of the confidence interval for the population mean generated by the sample $Y_1, Y_2, \ldots, Y_n$ are given by

$$\bar{Y} \pm t_{n-1}\frac{S_Y}{\sqrt{n}}$$

where t_{n-1} stands for the quantile belonging to the fraction 0.975 in a t-distribution with $n - 1$ degrees of freedom. It follows from this formula that an arbitrary number $v \in \mathbb{R}$ is covered by the interval if and only if the distance of v to $\bar{Y}$ is less than $t_{n-1} \times S_Y/\sqrt{n}$. More precisely, the number v is covered if and only if

$$|\bar{Y} - v| < t_{n-1}\frac{S_Y}{\sqrt{n}}.$$

In particular, the sample mean $\bar{X}$ is covered if and only if

$$|\bar{Y} - \bar{X}| < t_{n-1}\frac{S_Y}{\sqrt{n}}.$$

The above, however, is the same as saying that

$$\left|\frac{\sqrt{n/2}\,(\bar{X} - \bar{Y})}{S_Y}\right| < \frac{t_{n-1}}{\sqrt{2}}.$$

In virtue of (1) the probability P_n that the above occurs can be captured in terms of (central) cumulative t-distribution functions as

$$P_n = F_{n-1}\left(t_{n-1}/\sqrt{2}\right) - F_{n-1}\left(-t_{n-1}/\sqrt{2}\right).$$

Equivalently, one has

$$P_{n+1} = F_n\left(t_n/\sqrt{2}\right) - F_n\left(-t_n/\sqrt{2}\right). \tag{2}$$

By the formula above the values of the coverage probabilities can easily be computed. In the table below some values are given:

n	2	3	4	5	10	50	100	500
P_n	0.929	0.907	0.890	0.879	0.856	0.838	0.836	0.835

The table suggests that the coverage probabilities converge to a limit. In the following section this will be confirmed indeed. In an accuracy to 3 decimals the limit value will turn out to be 0.834.

The formula (2) can easily be visualized as area under the curve of a central t-density with n degrees of freedom. In the figure below the area of the grey region presents the probability P_{n+1}.
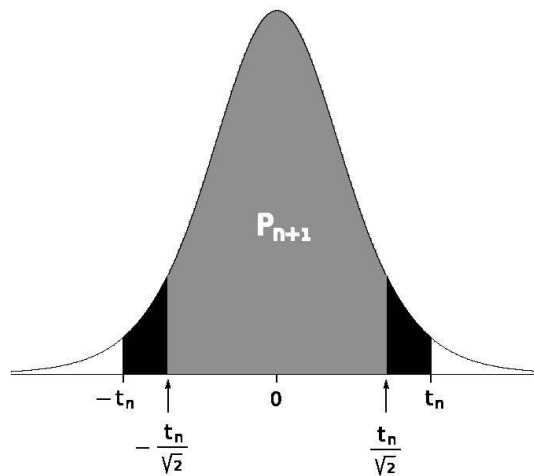
The area of the grey and the black regions together is 0.95.

# 4 The asymptotic value P of the probabilities P_n

The asymptotic behavior of the probabilities P_n is easily revealed. Namely, the cumulative central t-distribution functions F_n converge pointwise to the cumulative distribution Φ of the standard Gaussian distribution. So for all numbers $v \in \mathbb{R}$ one has

$$\lim_{n \to \infty} F_n(v) = \Phi(v).$$

The limit function Φ being continuous, the convergence is actually uniform on $\mathbb{R}$ (see for example [5, §VII.5]). In virtue of (2) it follows from this (see [7]) that

$$P = \lim_{n \to \infty} P_n = \Phi\left(t/\sqrt{2}\right) - \Phi\left(-t/\sqrt{2}\right) = 1 - 2\,\Phi\left(-t/\sqrt{2}\right) \tag{3}$$

where t stands for the quantile belonging to 0.975 in the standard Gaussian distribution. In a precision of 7 decimals the above equals 0.8342315. In the figure below the convergence of the coverage probabilities P_n is sketched.
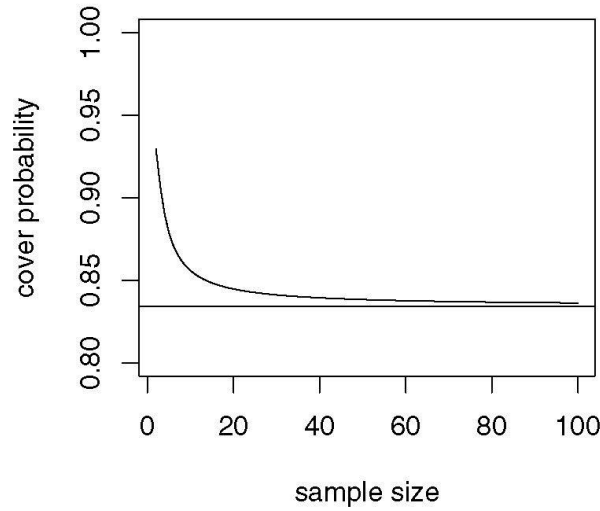
In the figure the fraction 0.8342315 is depicted as a horizontal asymptote. The values of the coverage probabilities can easily be checked through simulations.

5 A solution to the second coverage problem

For reasons that will become apparent in the last section, the second coverage problem will be solved in a somewhat more general setting. Assume that the two independent samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are drawn from an $N(\mu_X, \sigma)$-distributed and an $N(\mu_Y, \sigma)$-distributed population, respectively. Let Q_n be the probability that the two corresponding 95% confidence intervals are disjoint. Now define δ as

$$\delta = \frac{\mu_Y - \mu_X}{\sigma}.$$

The value of Q_n remains unchanged when replacing μ_X by 0, σ by 1 and μ_Y by δ. For this reason it is from now on assumed that the $X_1, X_2, \ldots, X_n$ emanate from a $N(0, 1)$-distributed and the $Y_1, Y_2, \ldots, Y_n$ from a $N(\delta, 1)$-distributed population. The corresponding value of Q_n will be denoted as $Q_n(\delta)$. To compute the values of $Q_n(\delta)$, recall that

the endpoints of the 95 % confidence interval for the population mean generated by the samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are given by

$$\bar{X} \pm t_{n-1} \frac{S_X}{\sqrt{n}} \quad \text{and} \quad \bar{Y} \pm t_{n-1} \frac{S_Y}{\sqrt{n}}$$

where t_{n-1} stands for the quantile belonging to the fraction 0.975 in a central t-distribution with $n - 1$ degrees of freedom. It follows from these formulas that there is non-overlap if and only if

$$\bar{X} + t_{n-1} \frac{S_X}{\sqrt{n}} < \bar{Y} - t_{n-1} \frac{S_Y}{\sqrt{n}} \quad \text{or} \quad \bar{Y} + t_{n-1} \frac{S_Y}{\sqrt{n}} < \bar{X} - t_{n-1} \frac{S_X}{\sqrt{n}} .$$

The two events above mutually exclude each other. For this reason one may write

$$Q_n(\delta) = \mathbb{P}\left(\bar{X} + t_{n-1} \frac{S_X}{\sqrt{n}} < \bar{Y} - t_{n-1} \frac{S_Y}{\sqrt{n}} \right)$$
$$+ \mathbb{P}\left(\bar{Y} + t_{n-1} \frac{S_Y}{\sqrt{n}} < \bar{X} - t_{n-1} \frac{S_X}{\sqrt{n}} \right) . \tag{4}$$

Now denote the first and second summand in the above by $Q_n^1(\delta)$ and $Q_n^2(\delta)$, respectively. By symmetry these two probabilities are related as

$$Q_n^2(\delta) = Q_n^1(-\delta) . \tag{5}$$

For this reason one has

$$Q_n(\delta) = Q_n^1(\delta) + Q_n^1(-\delta) . \tag{6}$$

Hence it suffices to focus on the probability $Q_n^1(\delta)$, that is, the probability that the event

$$\bar{X} + t_{n-1} \frac{S_X}{\sqrt{n}} < \bar{Y} - t_{n-1} \frac{S_Y}{\sqrt{n}}$$

occurs. This happens if and only if

$$\bar{X} - \bar{Y} + t_{n-1} \frac{S_X}{\sqrt{n}} < -t_{n-1} \frac{S_Y}{\sqrt{n}} .$$

This, in turn, occurs if and only if

$$\frac{\bar{X} - \bar{Y} + t_{n-1} S_X / \sqrt{n}}{S_Y / \sqrt{n}} < -t_{n-1} . \tag{7}$$

The inequality above can be rewritten as

$$\frac{\sqrt{n/2}\left(\bar{X} - (\bar{Y} - \delta) \right) + t_{n-1} S_X / \sqrt{2} - \delta\sqrt{n/2}}{S_Y} < \frac{-t_{n-1}}{\sqrt{2}} . \tag{8}$$

The *conditional* probability that the above occurs, given $S_X = s$, can, by (1), be captured in terms of non-central cumulative t-distribution functions as

$$F_{n-1}^{t_{n-1}S_X/\sqrt{2}-\delta\sqrt{n/2}}\left(-t_{n-1}/\sqrt{2}\right) = F_{n-1}^{t_{n-1}\sqrt{V/2(n-1)}-\delta\sqrt{n/2}}\left(-t_{n-1}/\sqrt{2}\right) \qquad (9)$$

where V stands for $(n-1)S_X^2$. The latter variable, however, has a χ^2-distribution with $n-1$ degrees of freedom. Denoting the density of this distribution by g_{n-1} one obtains, via the law of total probability, that

$$Q_n^1(\delta) = \int_0^\infty F_{n-1}^{t_{n-1}\sqrt{v/2(n-1)}-\delta\sqrt{n/2}}\left(-t_{n-1}/\sqrt{2}\right) g_{n-1}(v)\, dv. \qquad (10)$$

The integrals on the right side can be evaluated numerically for arbitrary values of n and δ. Thus, via (6), the values of $Q_n(\delta)$ can be computed. In the table below some of such computed values are given for $\delta = 0$.

n	2	3	4	5	10	100
Q_n	$0.378\,\%$	$0.476\,\%$	$0.514\,\%$	$0.532\,\%$	$0.552\,\%$	$0.557\,\%$

The table suggests that the probabilities Q_n converge to a limit. In the following section this will be confirmed indeed.

6   The asymptotic value Q of the probabilities Q_n

The asymptotic behavior of the probabilities Q_n is easily revealed by having a closer look at (8). The left side of (8) namely, for $\delta = 0$, can be rewritten as

$$\frac{Z_n + t_{n-1}S_X/\sqrt{2}}{S_Y} \qquad (11)$$

where Z_n stands for a variable that has a standard Gaussian distribution. If $n \to \infty$, then both S_X and S_Y converge to the standard deviation of the population, that is, to 1. Moreover the t_{n-1} converge to the number t that presents the quantile belonging to the fraction 0.975 in a standard Gaussian distribution. Exploiting the independence of the Z_n, S_X, S_Y it follows that (11) converges in distribution to the Gaussian distribution with mean $t/\sqrt{2}$ and standard deviation 1. So, denoting the cumulative distribution function of (11) by G_n, one has

$$\lim_{n\to\infty} G_n(v) = \Phi(v - t/\sqrt{2}) \qquad \text{for all} \quad v \in \mathbb{R}.$$

The limit function Φ being continuous, the convergence is actually uniform on $\mathbb{R}$ (see [5, §VII.5]). Exploiting this (see for example [7]) it follows from (6) that

$$Q = \lim_{n\to\infty} Q_n = \lim_{n\to\infty} 2\,G_n\left(-t_{n-1}/\sqrt{2}\right) = 2\,\Phi\left(-2t/\sqrt{2}\right) = 2\,\Phi\left(-t\sqrt{2}\right). \qquad (12)$$

In a precision of 7 decimals the above equals $0.5574597\,\%$. From the above it follows that the asymptotic probability to see two intersecting confidence intervals is given by

$$1 - \lim_{n\to\infty} Q_n = 1 - 2\,\Phi\left(-t\sqrt{2}\right). \qquad (13)$$

Compare this probability with (3).

7 Using interval overlap as a decision criterion

Given an $N(\mu_X, \sigma^2)$-distributed and an $N(\mu_Y, \sigma^2)$-distributed population, the hypothesis

$$H_0 : \mu_X = \mu_Y$$

is sometimes tested in the following way: Two samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are drawn from both populations and their corresponding 95 % confidence intervals are computed. Conclusive in the decision procedure is then whether the intervals intersect or not. If they intersect, then the hypothesis H_0 is maintained and if they are disjoint, then H_0 is rejected. If the confidence intervals are computed at a level of 95 % confidence, then, by (12), one arrives in the hypothesis test at an asymptotic significance level of $2 \, \Phi(-t\sqrt{2})$, where t stands for the quantile belonging to 0.975 in a standard Gaussian distribution. This leads to an asymptotic significance level of 0.5574597 % (see also [2], [4], [9]). In order to arrive at an asymptotic significance level of 5 % the value of t must be adapted into a value $\tilde{t}$ for which

$$2 \, \Phi\left(-\tilde{t}\sqrt{2}\right) = 0.05 \,.$$

This amounts to $\tilde{t} = t/\sqrt{2} = 1.385904$. The corresponding confidence level of this value for $\tilde{t}$ is 0.8342315. Note that this is precisely the asymptotic coverage probability P that arose in the first coverage problem! When the confidence level is set to this particular percentage the asymptotic significance level of the decision procedure is equal to 5 %. For finite sample sizes the confidence level P leads to the following significance levels, below denoted as α, in the decision procedure:

n	2	3	4	5	10	20	500
α	4.22 %	4.73 %	4.87 %	4.93 %	4.99 %	5.00 %	5.00 %

How does the decision procedure sketched above perform relative to a 2-sample t-test when testing at the same significance level, that is, at the level listed in the second line of the table above? It seems natural to compare the two decision procedures then as to their power. By the Neyman-Pearson theory the first method cannot possibly have a higher power than the 2-sample t-test, which, given an arbitrary fixed significance level, maximizes the power (see [5]). The discrepancy in power between the two procedures seems, however, to be minor. The power of the method of disjoint intervals is presented by the probability $Q_n(\delta)$, where, as before

$$\delta = \frac{\mu_Y - \mu_X}{\sigma} \,.$$

Thus the power can be computed via the formulas (6) and (10). Computations of $Q_n(\delta)$ for various values of n and δ suggest that in all possible cases the difference in power, relative to a 2-sample t-test, is less than 0.005.

References

[1] Barr, D.R.: Using confidence intervals to test hypotheses. *J. Quality Technology* 1 (1969), 256–258.

[2] Goldstein, H.; Healy, M.J.R.: The Graphical Presentation of a Collection of Means. *J. Royal Statistical Association, Series A* 158 (1995) 1, 175–177.

[3] Nelson, L.S.: Evaluating overlapping confidence intervals. *J. Quality Technology* 21 (1989), 140–141.

[4] Payton, M.E.; Greenstone, M.H.; Schenker, N.: Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *J. Insect Science* 3 (2003), 34–39. Available at www.insectscience.org/3.34

[5] Pestman, W.R.: *Mathematical Statistics*. Second edition, Walter de Gruyter Verlag, Berlin 2009.

[6] Pestman, W.R.: Teaching the independence of $\bar{X}$ and S^2. *Elem. Math.* 53 (1998), 107–111.

[7] Rudin, W.: *Principles of Mathematical Analysis*. McGraw-Hill, New York 1976.

[8] Ryan, G.W.; Leadbetter, S.D.: On the misuse of confidence intervals for two means in testing for the significance of the difference between the means. *J. Modern Applied Statistical Methods* 1 (2002), 473–478.

[9] Schenker, N.; Gentleman, J.F.: On judging the significance of differences by examining the overlap between confidence intervals. *Amer. Statist.* 55 (2001), 182–186.

Wiebe R. Pestman
University Medical Center Utrecht
Department: Julius Center
Section: Biostatistics
Heidelberglaan 100
NL-3584CX Utrecht, The Netherlands
e-mail: wpestman@umcutrecht.nl