САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Mathematical Reduction of Experimental Data

Andrei Vlasov
a.vlasov@spbu.ru

Alexey Victorov
a.viktorov@spbu.ru

and

Anastasia Venediktova
a.venediktova@spbu.ru

САНКТ-ПЕТЕРБУРГ, 2025

ISBN 978-5-288-06514-9

UDC 53.088

Reviewers:
Prof., Dr. Sci. *Elena V. Aksenova*
St. Petersburg University | SPBU · Department of Statistical Physics
Prof., Dr. Sci. *Konstantin N. Mikhelson*
Institute of Chemistry, St. Petersburg State University, Russia · Department of Physical Chemistry

Vlasov Andrei, Victorov Alexey, and Venediktova Anastasia

(8,13 Мб) Mathematical Reduction of Experimental Data

The present course is aimed to provide systematic instructions for the choice and application of methods related to processing the data gotten through the physico-chemical experiments. A correct and conscientious estimation of the experimental results necessitates certain knowledge of mathematical statistics. So, in order not to convert mathematical routines into a set of formal recipes, whose content is “a black box”, the course explicates basic ideas of statistical methods in conjugation with concepts and notions of the calculus in probabilities. The authors were focused on giving statistical paradigm transparently, avoiding cumbersome mathematical machinery. As a rule, during the interpretation of experimental data several issues do arise. In first place, the experimental result inevitably contains methodological or/and random errors. Therefore, it is a must to handle a skill of accessing the reliability of experimental data, their error bars, to master getting an estimate of the true values of a measurable parameter in a series of sequel experiments. Another thing, in the course of experimental studying there comes about a problem of analytical representation of various experimental regularities. Quite often the mathematically analytical type of such dependencies is known a priori, however, formulae contain a number of constants (parameters), whose values are adjusted so that the experimental data are described in the best way. Owing to that, the experimental variables contain random errors, estimates of the said parameters also turn out to be random variables, containing errors, whose statistical analysis is necessary. The course presents typical examples of the methods of statistical estimation. Last, but not least, the course gives algorithms of hypotheses testing with regard to the true values of the measured quantities, their variances, coincidence of the statistical characteristics in separate and parallel samples.

Подписано к использованию 07.10.2025
Издательство СПбГУ. 199004, С.-Петербург, В.О., 6-я линия, д. 11.
Тел./факс +7(812) 328-44-22
E-mail: publishing@spbu.ru publishing.spbu.ru

Mathematical Statistics. By Bartel Leendert van der Waerden,
Virginia Thompson, and Ellen Sherman

An introduction to Probability Theory and Its Applications. By William Feller

Outlines of the course

- ❑ **Processing results of direct measurements, estimates of the purpose(experimentally gotten) quantities and their uncertainties, transfer of measurement uncertainties onto analytical functions of directly measured values**

- ❑ **Square-least regression: processing data on the response with a regard to the changing predictor**

- ❑ **Hypothesis testing**

Part 1

Processing results of direct measurements and transfer of uncertainties of direct measurements onto analytically defined functions of directly measured properties

Why is it necessary to process experimental data?

1. **There are uncertainties in the gotten values of directly measured quantities (there are errors, both random and systematic). So, in 1st place, one needs analyzing the reliability of the data**
2. **It is necessary to analytically present the data**
3. **Analytical representation includes certain sets of parameters, their estimation bases on the direct measurements, thereby parameters are also subject to some uncertainties, and yet again there is an issue of estimating these uncertainties**

Any experimental observation is done amidst action of many origins who can not be accounted for explicitly. From here comes the irreproducibility of the observation result. The latter can be formulated in terms of the error

Errors can be subdivided into:
systematic and random.
If systematic errors are eliminated, the analysis of random errors make appearance as the object of the calculus in probabilities and its digression in math statistics

We articulate three types of problems with regard to random variables:

- Given statistical law is known, one needs to estimate the parameters of the latter
- Hypothesis testing
- One needs establishing of the statistical law

So, a question arises: what is the statistical law?

Recalling some bits of the calculus in probabilities

We define inter-relation between the system we observe and outer conditions. We call the set of conditions determining the state of our system as the complex G of conditions. Now we switch to a particular property (event), whose appearance could be formalized in a list of outcomes or events A, B, C, D… etc. in the complex G. If any of them can take place or not in the G-complex we call them random events

1. If the event consisting in that either A or B takes place in the G-complex is termed their sum, A + B

2. An event is called persistent, if in the G-complex it takes place necessarily, U-event

3. An event is termed impossible, if it will never ever take place in the G-complex, V-event

4. If an event consists in that both, event A and event B, take place in the G-complex, it is referred to as their product, AB

5. Events A and B are called incompatible if AB = V

6. Events $\{A_1,\ A_n,\ \ldots.\}$ constitute a complete group of events, if at least one of them takes place necessarily in the G-complex, i.e. $\sum_i A_i = U$

Let us now consider a complete group of equally plausible events (intuitively asserted), who are pair-wise incompatible. Now we constitute a sample space of events = the said group, impossible event and all the events, who can be sub-divided into partial events, included into the group

Then the definition of the probability of an event from the sample space
would be the ratio =
the number of events into which the particular event could be subdivided
over
the number of events in the complete group

Obvious properties of the probability, stemming from the definition:

1. $w(A) \geq 0$

2. $w(U) \equiv 1$

3. $w(V) = 0$

4. the probability of an event, who is a sum of incompatible events A and B:

$$w(A+B) = w(A) + w(B)$$

e.g. Tossing a die

A – falling out of an even nominal

B – falling out a nominal “5”

Complete group contains 6 events, so, $w(A+B) = \frac{3}{6} + \frac{1}{6} = \frac{2}{3}$

Consider now so called conditional probabilities and regularities bound with them and the general definition of the probability

An example. What is the probability of that the sum of two tossed dice equals 6, if the fallen out sum is even

Define now a complete group for two distinguishable dice....Obviously it has 36 events. How many of them procure an even sum? 18. So, the complete group for defining the probability of getting sum 6 contains 18 events:

$$1_{green} + 5_{red};\ \ 2_{green} + 4_{red};\ 3_{green} + 3_{red};\ 4_{green} + 2_{red};\ 5_{green} + 1_{red}$$

So, $w(falling\ out\ 6, given\ the\ sum\ is\ even) = \frac{5}{18}$

Event A, given B occurred, $w(^{A}/_{B})$

What's the unconditional probability of having sum even?

Unconditional event B, $w(B)$

Obviously, $\frac{18}{36}$

What's the unconditional probability of an event that

Unconditional event A (6-sum), but B (even sum)

the fallen out sum simultaneously is even and equals 6?

occurred simultaneously $w(AB)$

Obviously, $\frac{5}{36}$

I take the liberty of the generalization without stringent prove:

$$w\left(A/_B\right) = \frac{w(AB)}{w(B)}$$

$$w(AB) = w\left(A/_B\right) \cdot w(B)$$

If $w\left(A/_B\right) = w(A)$, events A and B are independent:

$$w(AB) = w(A) \cdot w(B)$$

$$w(AB) = w(A) \cdot w(B)$$

Example of independent events.
What is the probability of that upon tossing two dice “5” nominal falls out on one die and an even nominal falls out on the other?

Probability of “5” on a dice is $\frac{1}{6}$.
Probability of even nominal on the other is $\frac{3}{6}$.
i.e. the r.h.s. is $\frac{3}{36} = \frac{1}{12}$.
Insofar as dice are distinguishable, the result must be multiplied by 2 => $\frac{1}{6}$.

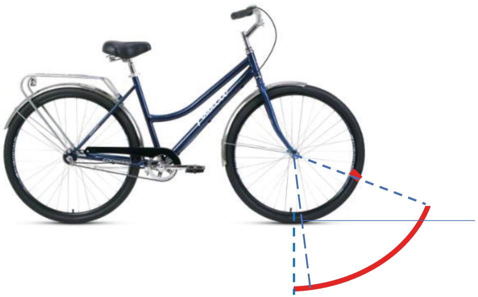
$$w(A) \cdot w(B) = \frac{1}{6}.$$

Now switch to the event AB, its complete group contains 36 events, and the very event splits into 6 results:

$$5_{green} + 2_{red};\ 5_{green} + 4_{red};\ 5_{green} + 6_{red};\ 2_{green} + 5_{red};\ 4_{green} + 5_{red};\ 6_{green} + 5_{red}$$

$$w(AB) = \frac{6}{36} = \frac{1}{6}.$$

Now switch to the continuous random events

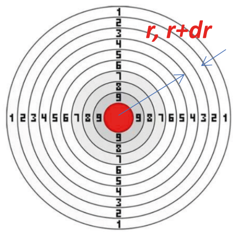


$$dw(x + dx) = \frac{dx}{2\pi R}$$

$$\int_0^{2\pi R} dw(x) = 1$$

$$w(x_1 \le x \le x_2) \int_{x_1}^{x_2} dw(x)$$

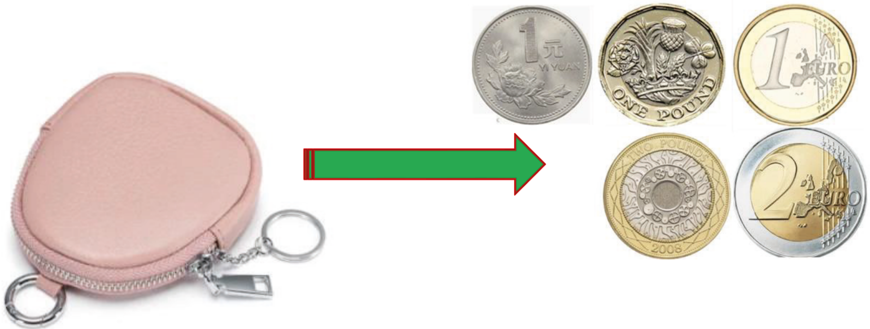
$$dw(x) = Cdx\ ;\ \ C = const$$



$$dw(r) = f(r)dr$$

$$f(r) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{r^2}{2\sigma^2}\right\}$$

Multidimensional random variables



Country's legal tender (i=1,2,3) / Coin value (j=1,2)	China	UK	EU
1	+	+	+
2		+	+

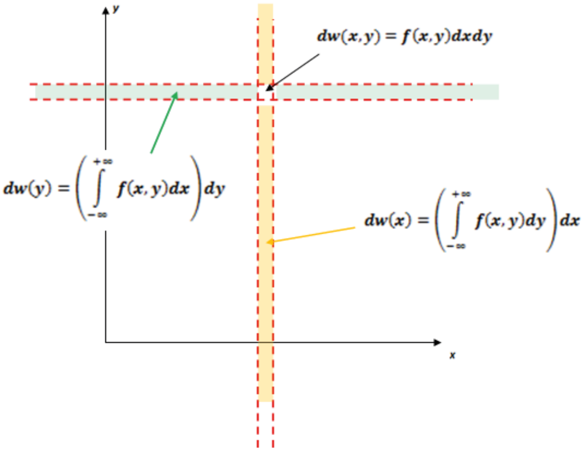
A unit event:

$$\left(x_i, y_j\right): \quad i = 1,2,3; j = 1,2$$

$$w\left(x_i, y_j\right) = \frac{1}{5}; \quad i = 1,2,3; j = 1,2$$

$$w(x_2) = \sum_{j=1}^{2} w\left(x_2, y_j\right) = \frac{2}{5}; \quad w(y_1) = \sum_{i=1}^{3} w(x_i, y_1) = \frac{3}{5}$$

Multidimensional continuous random variable



What is statistical law?
For a discrete random variable it is a bi-jection:

x_1	x_2	x_3	...	...	...	x_n	...
$w(x_1)$	$w(x_2)$	$w(x_3)$	...	...	...	$w(x_n)$	...

For a continuous random variable it is the function of the probability density:

$$f(x,y) \qquad\qquad f(x)$$

Apart from the dependence on the random variable, statistical law contain certain parameters, who's determination is extremely important for a judgment about true value of the studied quantity

We focus primarily on two of them:

- **Statistical expectation**

$$MX \equiv \sum_i x_i \cdot w(x_i)\,; \qquad MX \equiv \int_{-\infty}^{+\infty} x f(x) dx$$

- **Variance**

$$DX \equiv M(x - MX)^2$$

In first place, we discuss their formal properties, later we elucidate their statistical meaning

1. *Expectation of a sum of random variables equals the sum of expectations*:

$$M(x+y) = MX + MY$$

2. *Expectation of a constant (a persistent event) equals this constant :*

$$MC = C$$

3. *If variables are independent, then it holds:*

$$M(xy) = MX \cdot MY$$

3a. *Expectation of a linear function of a random variable holds*:

$$MCX = C \cdot MX$$

We prove properties of expectation in case of a discrete random variable

$$M(x+y)=\sum_{i,j}(x_i+y_j)w(x_iy_j)=\sum_{i,j}x_i\,w(x_iy_j)+\sum_{i,j}y_j\,w(x_iy_j)=$$
$$=\sum_i x_i\sum_j w(x_iy_j)+\sum_j y_j\sum_i w(x_iy_j)=\sum_i x_iw(x_i)+\sum_j y_j\,w(y_j)\equiv MX+MY$$

$$M(xy)=\sum_{i,j}x_iy_jw(x_iy_j)=\sum_{i,j}x_iy_jw(x_i)w(y_j)=\sum_i x_iw(x_i)\cdot\sum_j y_jw(y_j)\equiv MX{\cdot}MY$$

Now prove the properties of the variance:

1. If variables x and y are independent, then $D(x + y) = DX + DY$

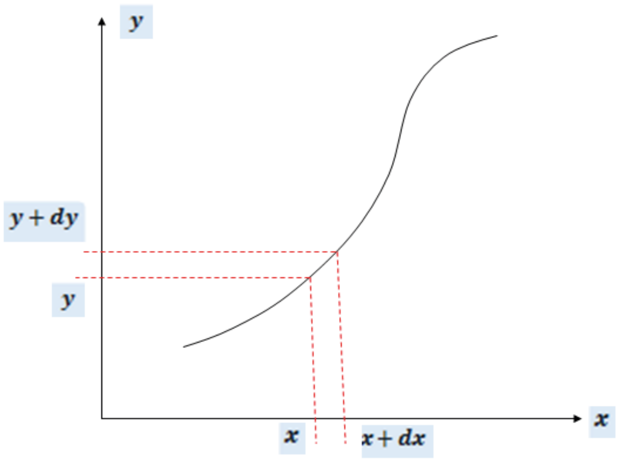
$$D(x+y) \equiv M[(x+y) - M(x+y)]^2 =$$

$$= M[(x - MX) + (y - MY)]^2 =$$

$$= M[(x - MX)^2 + (y - MY)^2 + 2(x - MX)(y - MY)] \equiv$$

$$\equiv DX + DY + M[2(x - MX)(y - MY)]$$

Statistical expectation of the function of random variable



$$dw(x) = dw(y)$$

$$MY \equiv \int_{-\infty}^{+\infty} y\,dw(y) = \int_{-\infty}^{+\infty} y(x)\,dw(x) = \int_{-\infty}^{+\infty} y(x) f(x)\,dx$$

Consider a random variable $\xi \in (-\infty, +\infty)$ with statistical law $f(\xi)$. Then for a function of this variable ξ^2 the expectation holds:

$$M\xi^2 = \int_{-\infty}^{+\infty} \xi^2 f(\xi) d\xi = \int_{-\infty}^{-\varepsilon} \xi^2 f(\xi) d\xi + \int_{-\varepsilon}^{+\varepsilon} \xi^2 f(\xi) d\xi + \int_{+\varepsilon}^{+\infty} \xi^2 f(\xi) d\xi >$$

$$> \int_{-\infty}^{-\varepsilon} \xi^2 f(\xi) d\xi + \int_{+\varepsilon}^{+\infty} \xi^2 f(\xi) d\xi$$

$$> \varepsilon^2 \left[\int_{-\infty}^{-\varepsilon} f(\xi) d\xi + \int_{+\varepsilon}^{+\infty} f(\xi) d\xi \right] \equiv w(|\xi| > \varepsilon)$$

$$\frac{M\xi^2}{\varepsilon^2} > w(|\xi| > \varepsilon)$$

Pafnutiy Chebyshev
1821–1894

specifically $\xi = MX - x \;\; => M\left[(MX - x)^2\right] \equiv DX$

$$\boxed{w(|x - MX| > \varepsilon) < \frac{DX}{\varepsilon^2}}$$

Chebyshev inequality

Aleksander Lyapunov
1857–1918

Central limit theorem

An experimental result (either per se or as an algebraic sum of true value and an error) is provided by operation of a number of random events. If a contribution of any of them is not overwhelming and each of them has finite expectation and variance, then consider the sum

$$\xi = \sum_i \xi_i$$

(ξ_i are presentation of random variables corresponding to contributing random events).

This sum complies with a statistical law close to the normal (a.k.a. Gaussian) law.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$$

$$\downarrow$$

$$x \in N(a, \sigma^2)$$

Each of ξ_i might comply with any statistical law, it should not provide an overwhelming contribution into ξ !!!!

Johann Carl Friedrich Gauss
1777–1855

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$$

$$MX \equiv \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-a+a) exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-a) exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx + \frac{a}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \frac{a}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} exp\left\{-\frac{y^2}{2\sigma^2}\right\} dy = 2\frac{1}{2}\frac{a}{\sqrt{2\pi\sigma^2}}\sqrt{2\pi\sigma^2} = a$$

Siméon Denis Poisson
1781–1840

$$\boxed{\int_{0}^{+\infty} e^{-\alpha x^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}}$$

$$DX = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-a)^2 exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y^2 exp\left\{-\frac{y^2}{2\sigma^2}\right\} dy = \frac{1}{2\sqrt{2\pi\sigma^2}}\sqrt{8\pi\sigma^6} = \sigma^2$$

$$\boxed{\int_{-\infty}^{+\infty} x^2\, e^{-\alpha x^2} dx = -\frac{\partial}{\partial\alpha} \int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = -\frac{\partial}{\partial\alpha}\sqrt{\frac{\pi}{\alpha}} = \frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}}$$

One more simple, though, important property of variance (independent of the statistical law) –
In our numeration it is the 2nd property:

If $C = const \Rightarrow\ D(Cx) = M[Cx - M(Cx)]^2 = M\{C^2(x - MX)^2\} = C^2 M(x - MX)^2 \equiv C^2 DX$

The central limit theorem asserts, that there are reasons to reckon that observation of a single random variable in a reproduceable complex of conditions comply with the normal law. If one knows analytical form of the statistical law, how can we determine its parameters?

We give a pinpointed formulation of the issue. All the perceivable results of the experiment constitute so called **entire assembly**, we by necessity have at most some number of outcomes of our test measurements, called a sample.

$\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_n}\}$ **– a sample, $\boldsymbol{n}$ – sample size.**

Following Lyapunov theorem, we presume that exact values of the sample are implementation of a random variable ξ, distributed normally. One experimental move we will consider independent of another.

An alternative standpoint, each of the values $\{x_1, x_2, \dots x_n\}$ are numeric values gotten by independent random variables $\{\xi_1, \xi_2, \dots \xi_n\}$ distributed in accord with a normal law having the same analytical form and the same parameters, although we do not know their values.

So, the question stands: how basing on the elements of the sample we can get estimates of the normal distribution parameters?

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{a}}(\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_n}\}) \qquad \widehat{\boldsymbol{\sigma}^2} = \widehat{\boldsymbol{\sigma}^2}(\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_n}\}).$$

Important features of these estimates: they are 1) random variables and 2) point estimates.

So, the question stands: how basing on the elements of the sample we can get estimates of the normal distribution parameters?

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{a}}(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n\}) \qquad \widehat{\boldsymbol{\sigma}^2} = \widehat{\boldsymbol{\sigma}^2}(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n\}).$$

Obvious thing is in that estimates are also random variables, and even for a fixed sample size they would alter from one sample to another. Therefore such estimates would hardly be rationalized from the standpoint of a question: how close are they to the true values of parameters a and σ^2?

The methodology of the answer is connected with the maximum likelihood estimation approach (**R. Fisher**, 1922).

Ronald Aylmer Fisher
1890–1962

Insofar as $\{\xi_1, \xi_2, \dots \xi_n\}$ are independent, the probability of an event $\{\xi_1 = x_1, \xi_2 = x_2, \dots \xi_n = x_n\}$ conforms to the theorem about independent events:

$$dw(\xi_1 = x_1 + dx_1, \xi_2 = x_2 + dx_2, \dots \xi_n = x_n + dx_n) = \prod_{i=1}^{n} dw(\xi_i = x_i) = \prod_{i=1}^{n} f(\xi_i, a, \sigma^2) d\xi_i$$

Likelihood function

$$\boldsymbol{L \equiv \left(\prod_{i=1}^{n} f(\xi_i, a, \sigma^2) \right)}$$

is the density of probabilities for an event consistent in that the random variables $\{\xi_1, \xi_2, \dots \xi_n\}$ simultaneously acquired a set of values:

$$\{\xi_1 = x_1 + dx_1, \xi_2 = x_2 + dx_2, \dots \xi_n = x_n + dx_1\}$$

Fisher principle asserts that estimation of the distribution parameters of L, and thereby, f, implement **$max\,L$**. It means that **what happened had the maximum probability**

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\};$$

$$\boldsymbol{lnL = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}\sum_i (x_i - a)^2}$$

$$\begin{cases} \dfrac{\partial lnL}{\partial \sigma^2} = -\dfrac{n}{2} \cdot \dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4}\sum_i (x_i - a)^2 = 0 \\ \dfrac{\partial lnL}{\partial a} = \dfrac{2}{2\sigma^2}\sum_i (x_i - a) = 0 \end{cases}$$

As far as aposteriori $\sigma < +\infty$ => $\widehat{\sigma^2} = \frac{1}{n}\sum_i (x_i - a)^2$ and hence $\sum_i x_i - n\hat{a} = 0.$

Eventually, we have:

$$\hat{\boldsymbol{a}} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}} \qquad \widehat{\boldsymbol{\sigma}^{\mathbf{2}}} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}} (\boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{a})^{\mathbf{2}}$$

$$\bar{x} = \hat{a}; \qquad D\xi_n = \widehat{\sigma^2}$$

Point estimates

$$\{x_i\}\ \epsilon \quad f(x)_{\theta_1,\theta_2,\dots\theta_m} \qquad 1 \le i \le n$$

1. Consistent estimate

Sample size

$$\lim_{n\to\infty} w\left(\left|\hat{\theta}_{\alpha}^{(n)} - \theta_{\alpha}\right| < \varepsilon\right) = 1$$

Number of numeric parameter of the law $f(x)$

2. Unbiased estimate

$$n = const \qquad M\hat{\theta}_{\alpha}^{(n)} = \theta$$

From the standpoint of these properties we analyze point estimates
In first place, the estimate of the expectation

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

single sample mean

1. $M\bar{x} = M\left(\frac{1}{n}\sum_i x_i\right) = \frac{1}{n}\sum_i Mx_i = M\xi = a$

   Therefore, single sample mean is unbiased estimate of the expectation

2. Recall the Chebyshev inequality:

   $w(\xi|-M\xi| > \varepsilon) < \frac{D}{\varepsilon^2}$

   We set $\xi = \bar{x}$, then

   $w(|\bar{x} - M\bar{x}| > \varepsilon) < \frac{D\bar{x}}{\varepsilon^2}$

   Consider the r.h.s.:

$$D\bar{x} = D\left(\frac{1}{n}\sum_i x_i\right) = \frac{1}{n^2}\sum_i Dx_i = \frac{n}{n^2}D\xi = \frac{1}{n}D\xi$$

Hence, the Chebyshev inequality takes the form:

$$w(|\bar{x} - M\bar{x}| > \varepsilon) < \frac{D}{n\varepsilon^2}$$

That means that with an increase of the sample size the probability of deviation of the sample mean from the expectation decreases!

There should be done a special note regarding the property of the estimate consistency.

All of the real measurements are performed with apparatus machinery, who has certain thresholds of precision δx. So, any single measurement of a sample has two terms $\boldsymbol{x_i^{(apparent)}} = \boldsymbol{x_i^{(random)}} \pm \boldsymbol{\delta x}$ and the definition of consistency involves the sample $\left\{\boldsymbol{x_i^{(apparent)}}\right\}$. Therefore, this definition deals with a double inequality:

$$\boldsymbol{\overline{x}^{(apparent)} - \varepsilon \leq a \leq \overline{x}^{(apparent)} + \varepsilon}$$

Its probability equals $1 - w\left(\left|(x^{(apparent)} - a\right| > \varepsilon\right) = 1 - \frac{Dx}{n\varepsilon}$

$$\boldsymbol{\overline{x}^{(apparent)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(random)} \pm \delta x}$$

Obviously, whatever big is the sample size, only the 1st term of the latter equation depends on it, apparatus threshold sensitivity staying constant. So, the 1st term can be stabilizing with, $n \to \infty$, whilst the 2nd one will never ever diminish

$$D\xi_n = \frac{1}{n}\sum_{i=1}^{n}(x - \bar{x})^2$$

sample variance

Now consider sample variance:

$$D\xi_n \equiv \frac{1}{n}\sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{n}\sum_i \left(x_i - a - (\bar{x} - a)\right)^2 = \frac{1}{n}\sum_i (x_i - a)^2 + \frac{1}{n}\sum_i (\bar{x} - a)^2 - \frac{2}{n}\sum_i (x_i - a)(\bar{x} - a)$$

$$= \frac{1}{n}\sum_i (x_i - a)^2 + (\bar{x} - a)^2 - 2(\bar{x} - a)\left[\frac{1}{n}\sum_i x_i - \frac{1}{n}\cdot a \cdot n\right] = \frac{1}{n}\sum_i (x_i - a)^2 - (\bar{x} - a)^2$$

This quantity is obviously random.

Calculate its expectation:

$$MD\xi_n$$

$$M\left[\frac{1}{n}\sum_i (x_i - a)^2\right] = \frac{1}{n}\sum_i M[(x_i - a)^2] = \frac{1}{n}\cdot n \cdot D$$

$$M[(\bar{x} - a)^2] \equiv D\bar{x} = \frac{D}{n}$$

$$MD\xi_n = D\xi - \frac{D\xi}{n} = \frac{(n-1)}{n} D\ \xi$$

The sample variance is a biased estimate of the entire assembly variance

Mean square deviation

$$S_n^2 \equiv \frac{n}{n-1} D\xi_n \qquad MS_n^2 = \frac{n}{n-1} M[D\xi_n] = \frac{n}{n-1} \cdot \frac{(n-1)}{n} \cdot D\xi = D\xi$$

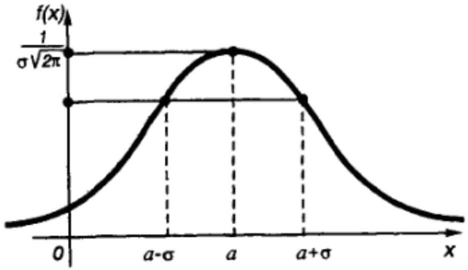
Mean square deviation, ***S***, is unbiased estimate of the assembly variance

Both estimates gotten are so called point estimates: although they procure estimation of the purpose quantities, they cater for no information as to how close we are regarding the estimated values.

What is needed is to establish statistical laws of their distribution.

Some more bits of the Gaussian function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$$

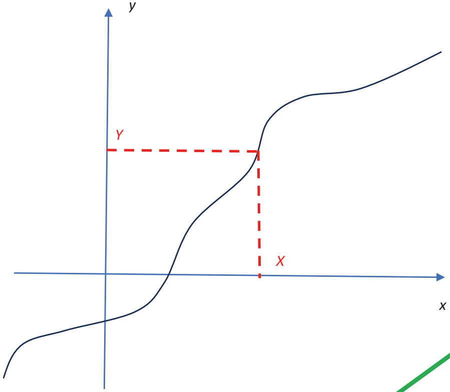


$$f'(x) = -\frac{(x-a)}{2\sigma^2\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$$

$$x_{max} = a$$

$$f''(x) = -\frac{1}{\sigma^2\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} + \frac{(x-a)}{\sigma^2\sqrt{2\pi\sigma^2}} \cdot \frac{(x-a)}{\sigma^2} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} = \frac{1}{\sigma^2\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \cdot \left[\frac{(x-a)^2}{\sigma^2} - 1\right]$$

$$x_{inflection} = a \pm \sigma$$



Statistical law for the function of the normally distributed random variable

$$\boldsymbol{w(y<Y)=w(x<X)}$$

Recap the differentiation of an integral with a varying upper limit:

$$\frac{d}{dx}\int_{const}^{x(y)} f(t)dt = f(x)\cdot\frac{dx}{dy}$$

$$\int_{-\infty}^{Y}\boldsymbol{\varphi(y)dy} = \int_{-\infty}^{X(Y)}\boldsymbol{f(x)dx} \quad => \quad \boldsymbol{\varphi(Y)} = \frac{\boldsymbol{dw(y<Y)}}{\boldsymbol{dY}} = \frac{\boldsymbol{d}}{\boldsymbol{dY}}\int_{-\infty}^{X(Y)}\boldsymbol{f(x)dx} = \boldsymbol{f(X)\cdot X'(Y)}$$

Since $\boldsymbol{X}$ is any value in a span of $\boldsymbol{x}$, the latter relation can be given as $\boldsymbol{\varphi(y)=f(x)\cdot\frac{dx}{dy}}$

If $\boldsymbol{x\in N(a,\sigma^2)}$, what would be the statistical law of the variable $\boldsymbol{y\equiv\frac{x-a}{\sigma}}$?

$$x = y\sigma + a \quad => \quad \varphi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \cdot \sigma = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{y^2}{2}\right\}$$

if $x \in N(a, \sigma^2) \Rightarrow y \equiv \frac{x-a}{\sigma}$: $y \in N(0, 1)$ **— standard normal distribution**

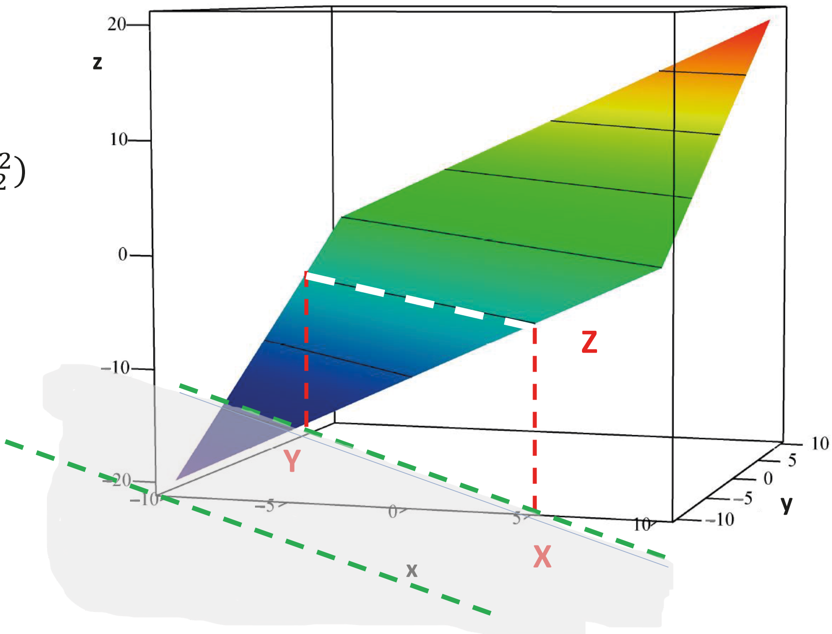
Composition of the normal laws

Independent random variables

$x \in N(a_1, \sigma_1^2)$ $y \in N(a_2, \sigma_2^2)$

What is the statistical law of

$$z = x + y\,?$$



$$z \le Z \Rightarrow$$
$$\Rightarrow \begin{cases} -\infty \le x \le +\infty \\ -\infty \le y \le Z - x \end{cases}$$

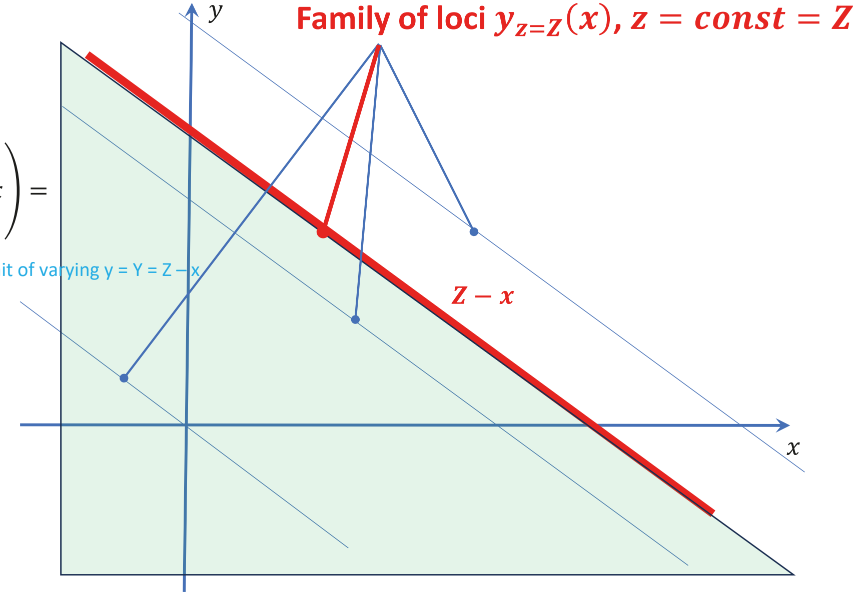
$$g(Z) = \frac{d}{dZ} \int_{-\infty}^{+\infty} \left(f(x) \left[\int_{-\infty}^{Z-x} f(y)\, dy \right] dx \right) =$$

NB: this is differentiation with regard to the upper limit of varying y = Y = Z – x

$$= \int_{-\infty}^{+\infty} f(x) f(Z - x) dx$$

Since we have $\forall Z \in z$ =>

$$g(z) = \int_{-\infty}^{+\infty} f(x) f(Z - x) dx$$



$$g(z) = \int\limits_{-\infty}^{+\infty} f(x)f(z-x)dx = \frac{1}{2\pi\sigma_1\sigma_2}\int\limits_{-\infty}^{+\infty} exp\left\{-\frac{(x-a_1)^2}{2\sigma_1^2}\right\} exp\left\{-\frac{(z-x-a_2)^2}{2\sigma_2^2}\right\} dx = $$

$$=$$

$$\frac{1}{2\pi\sigma_1\sigma_2}\int\limits_{-\infty}^{+\infty} exp\left\{-\frac{x^2-2xa_1+a_1^2}{2\sigma_1^2}-\frac{x^2-2x(z-a_2)+(z-a_2)^2}{2\sigma_2^2}\right\} dx$$

$$= \frac{1}{2\pi\sigma_1\sigma_2}\int\limits_{-\infty}^{+\infty} exp\left\{-x^2\boxed{\left(\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}\right)}+2x\boxed{\left(\frac{a_1}{2\sigma_1^2}+\frac{z-a_2}{2\sigma_2^2}\right)}-\boxed{\left(\frac{a_1^2}{2\sigma_1^2}+\frac{(z-a_2)^2}{2\sigma_2^2}\right)}\right\}dx =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2}\int\limits_{-\infty}^{+\infty} exp\{-\boxed{A}x^2+2\boxed{B}x-\boxed{C}\}dx =$$

$$\frac{1}{2\pi\sigma_1\sigma_2}\int\limits_{-\infty}^{+\infty} exp\left\{-\left(\sqrt{A}\cdot x\right)^2+2\sqrt{A}\cdot x\cdot\frac{B}{\sqrt{A}}+\frac{B^2}{A}-\frac{B^2}{A}-C\right\}dx =$$

$$\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} exp\left\{-\left(\sqrt{A}\cdot x-\frac{B}{\sqrt{A}}\right)^2\right\}\cdot exp\left\{\frac{B^2-AC}{A}\right\}dx=$$

$$=\frac{1}{2\pi\sigma_1\sigma_2}\cdot\ exp\left\{\frac{B^2-AC}{A}\right\}\cdot\frac{1}{\sqrt{A}}\int_{-\infty}^{+\infty} exp\left\{-\left(\sqrt{A}\cdot x-\frac{B}{\sqrt{A}}\right)^2\right\}d\left(\sqrt{A}\cdot x\right)$$

$$=\frac{1}{2\pi\sigma_1\sigma_2}\cdot\sqrt{\frac{\pi}{A}}\cdot exp\left\{\frac{B^2-AC}{A}\right\}$$

$$A=\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}$$

$$g(z)=\frac{1}{\sqrt{2\pi\boxed{(\sigma_1^2+\sigma_2^2)}}}exp\left\{-\frac{\left(z-\boxed{(a_1+a_2)}\right)^2}{2(\sigma_1^2+\sigma_2^2)}\right\}$$

$$\boldsymbol{z\in N\left(a_z,\sigma_z^2\right);\qquad a_z=a_x+a_y,\quad \sigma_z^2=\sigma_x^2+\sigma_y^2}$$

Let a variable comply with a normal law $x \in N(a, \sigma^2)$. And a function is defined $y(\text{x}) = C \cdot x$, $C = const$

What would be the statistical law of y?

Follow the proven theorem

$$\varphi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \cdot \frac{1}{C} = \frac{1}{\sqrt{2(\sigma \cdot C)^2}} exp\left\{-\frac{(Cx-Ca)^2}{2(\sigma \cdot C)^2}\right\} = \frac{1}{\sqrt{2\pi\sigma_y^2}} \mathrm{exp}\left\{-\frac{(y-a_y)^2}{2\sigma_y^2}\right\}$$

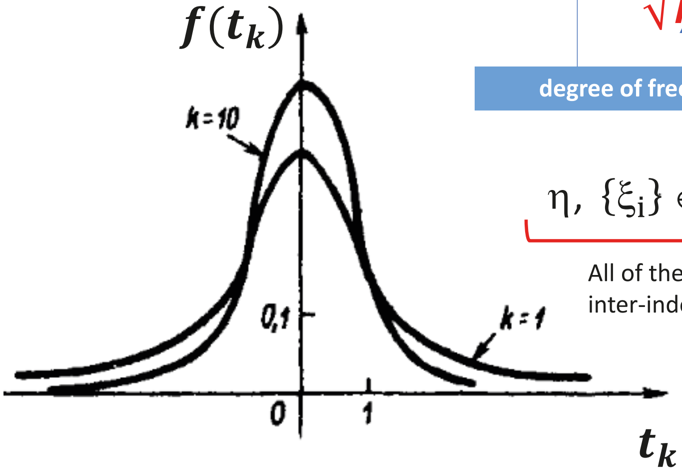
Obviously, $y \in N(a_y, \sigma_y^2)$, where $a_y = Ca; \sigma_y^2 = C^2\sigma^2$

Distributions connected with the standard normal law: Student distribution

$$t_k = \frac{\eta}{\sqrt{\frac{1}{k}\sum_{i=1}^{k}\xi_i^2}}$$

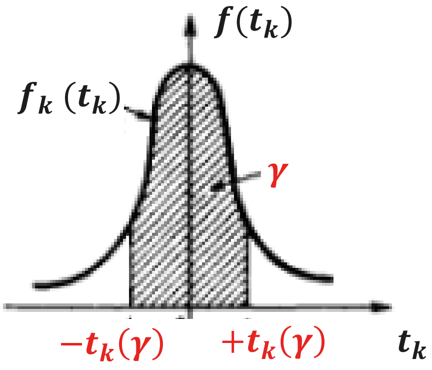
degree of freedom

$$\underbrace{\eta,\ \{\xi_i\} \in \ N(0,1)}_{\text{All of the random variables are inter-independent}}$$



William Sealy Gosset
*(a.k.a. **Student**)*
1876–1937

The Student distribution

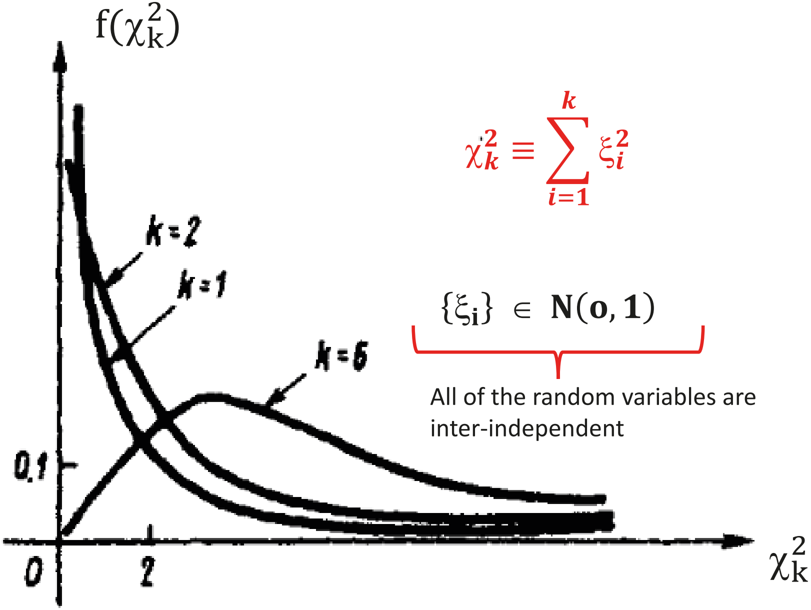


$\gamma = 0.95$ $\quad k = 6$

$$w(|t_k| < t_k(\gamma)) = \gamma$$

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Distributions connected with the standard normal law: Pearson distribution

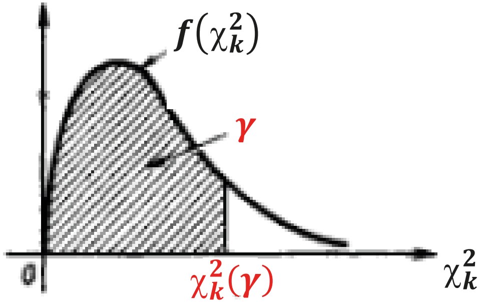


$$\chi^2_k \equiv \sum_{i=1}^{k} \xi_i^2$$

$$\underbrace{\{\xi_i\} \in \mathbf{N}(o,1)}_{\text{All of the random variables are inter-independent}}$$

Karl (Carl) Pearson
1857–1936

The Pearson χ^2 distribution



quantile

$\gamma = 0.95$ $k = 6$

k \ γ	0,50	0,70	0,80	0,90	0,95	0,98	0,99	0,999
1	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	9,34	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,5
14	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	15,34	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

$$w(\chi_k^2 < \chi_k^2(\gamma)) = \gamma$$

Fisher lemma

Let there be a sample of size n. Its elements are random variables $x_i \in N(a, \sigma^2)$. Consider variables $x_i^{'} = x_i - a$. By virtue of the proven theorems $x_i^{'} \in N(0, \sigma^2)$.

Calculate the sample mean $\bar{x}^{'} = \frac{1}{n}\sum_i x_i^{'} = \frac{1}{n}\sum_i (x_i - a) = \frac{1}{n}\sum_i x_i - a = \bar{x} - a$.

$$S^2(n-1) = \sum_i (x_i - \bar{x})^2 = \sum_i \left(x_i - a - (\bar{x} - a)\right)^2 = \sum_i \left(x_i^{'} - \bar{x}^{'}\right)^2 = \left(S^{'}\right)^2 (n-1)$$

Therefore, the mean average of the sample $x_i^{'} \in N(0, \sigma^2)$ is the same as of the original experimentally gotten sample $x_i \in N(a, \sigma^2)$.

Now consider a set of variables $\{y_i\}$,gotten by a linear transform of a sample $\{x_i^{'}\}$:

$y_j = \sum_{\alpha=1}^{n} C_{j\alpha} x_{\alpha}^{'}$.

Leopold Kronecker
1823–1891

As it follows from the previously proven theorems all of the $\{y_i\}$ are distributed normally. We calculate numeric values of their parameters:

$$My_j = M\left(\sum_{\alpha=1}^{n} C_{j\alpha} x_{\alpha}^{'}\right) = \sum_{\alpha=1}^{n} C_{j\alpha} M x_{\alpha}^{'} = 0.$$

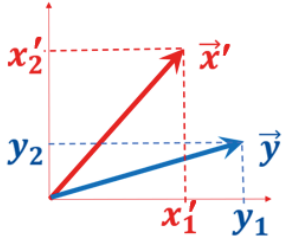
In a particular case of orthogonal matrix one should be mindful of this property:

$$\begin{pmatrix} \boxed{\begin{matrix} C_{11} & C_{12} \\ C_{21} & C_{22} \\ \cdots & \cdots \\ C_{n1} & C_{n2} \end{matrix}} & \begin{matrix} C_{13} \cdots \cdots \\ C_{23} \cdots \cdots \\ \cdots \\ C_{n3} \cdots \cdots \end{matrix} \end{pmatrix}$$

$$\sum_{\alpha} C_{i\alpha} C_{j\alpha} = \delta_{ij}$$

$$\sum_{\alpha} C_{\alpha i} C_{\alpha j} = \delta_{ij}$$

$$\delta_{ij} = \begin{cases} 1, & if\ i = j \\ 0 & ,if\ i \neq j \end{cases}$$



One more step further: calculate the expectation of a product $y_i \cdot y_j$

$$M\left(y_i \cdot y_j\right) = M[(y_i - 0) \cdot (y_j - 0)] = M\left[\sum_\alpha C_{i\alpha} x'_\alpha \cdot \sum_\beta C_{j\beta} x'_\beta\right] = \sum_{\alpha,\beta} C_{i\alpha} C_{j\beta} M\left(x'_\alpha x'_\beta\right)$$

Since x'_α and x'_β are independent, it holds:

$$\begin{cases} \alpha \neq \beta \quad \Rightarrow M x'_\alpha M x'_\beta = 0 \\ \alpha = \beta \quad \Rightarrow M\left(x'_\alpha - 0\right)\left(x'_\alpha - 0\right) = M\left(x'_\alpha - 0\right)^2 = \sigma^2 \end{cases}$$

Consequently, $M\left(y_i \cdot y_j\right) = \sum_{\alpha,\beta} C_{i\alpha} C_{j\beta} \delta_{\alpha\beta} \sigma^2 = \sigma^2 \sum_\alpha C_{i\alpha} C_{j\alpha} = \sigma^2\, \delta_{ij}$

Therefore all of the $\{y_i\}$ are non-correlated, and insofar as they are normally distributed, they are independent

This is in advance!!! The validation will be given in Part 2

Eventually, the “dry fallout” follows:

1. $M\left(y_i^2\right) = Dy_i^2 = \sigma^2$
2. If the C matrix is an orthogonal one, then $y_i \in N(0, \sigma^2)$
3. All the $\{y_i\}$ are statistically inter-independent

It is shown in statistics that there exists an orthogonal matrix:

$$\begin{pmatrix} \boxed{\frac{1}{\sqrt{n}} \quad \frac{1}{\sqrt{n}} \cdots \quad \frac{1}{\sqrt{n}}} \\ \vdots \qquad \ddots \qquad \vdots \\ \cdots \qquad \cdots \qquad \cdots \end{pmatrix}$$

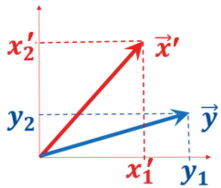
Then, the features of the transform are:

1. $y_1 \in N(0,\sigma^2)$, because $y_1 = \frac{x_1'}{\sqrt{n}} + \frac{x_2'}{\sqrt{n}} + \cdots \frac{x_n'}{\sqrt{n}} = \sqrt{n}\bar{x}'$ $\qquad \left(\bar{x}' \epsilon N\left(0, \frac{\sigma^2}{n}\right)\right)$
2. Due to the fact that orthogonal transform does not change the length of a vector, we have $\sum_i y_i^2 = \sum_i \left(x_i'\right)^2$

Then for a sample average square deviation one has:

$$S^2(n-1) \equiv \sum_i \left(x_i' - \bar{x}'\right)^2 = \sum_i \left(x_i'\right)^2 - 2\bar{x}' \sum_i x_i' + n\left(\bar{x}'\right)^2 = \boxed{\sum_i \left(x_i'\right)^2} - \boxed{n\left(\bar{x}'\right)^2} = \boxed{\sum_{i=1}^n y_i^2} - \boxed{y_1^2} = \sum_{i=2}^n y_i^2$$

So, the quantity $S^2(n-1)$ can be presented as a sum of $(\boldsymbol{n-1})$ statistically independent standard normal random variables $y_i \in N(0,\sigma^2)$



$\{y_i\}$ are inter-independent => $S^2(n-1)$ is independent from $\bar{x}'$, insofar as it equals $\frac{y_1}{\sqrt{n}}$

Now consider a quantity $\frac{S^2(n-1)}{\sigma^2} = \sum_{i=2}^{n} \left(\frac{y_i}{\sigma}\right)^2$, the sum on the r.h.s. embraces $(n-1)$ terms

$$\frac{y_i}{\sigma} \epsilon\, N(o,1)$$

Remember the definition of the Pearson variable:

$$\chi_k^2 \equiv \sum_{i=1}^{k} \xi_i^2$$

$$\frac{S^2(n-1)}{\sigma^2} = \chi_{n-1}^2$$

Now we turn back to the sample average: $\bar{x}n = \sum_{i=1}^{n} x_i$; $\quad x_i \in N(a, \sigma^2)$

$M\bar{x} = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = Mx_i = a$

$D\bar{x} = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D\,x_i = \frac{\sigma^2}{n}$

$\bar{x} \epsilon N(a, \sigma^2)$ => $\frac{(\bar{x}-a)}{\sigma/\sqrt{n}} \epsilon N(0,1)$

$$\boxed{\frac{(\bar{x}-a)}{\sigma/\sqrt{n}}} \times \frac{1}{\sqrt{\frac{1}{(n-1)} \cdot \boxed{\frac{S^2(n-1)}{\sigma^2}}}} = \frac{(\bar{\boldsymbol{x}}-\boldsymbol{a})\sqrt{\boldsymbol{n}}}{\boldsymbol{S}} = \boldsymbol{t}_{\boldsymbol{n}-\boldsymbol{1}}$$

Independent standard normal variables feature in the denominator and in the numerator

The Fisher lemma

Ronald Aylmer
Fisher
1890–1962

$$x_i \in N(a, \sigma^2), \qquad 1 \le i \le n$$

$$\Downarrow$$

$$\frac{(\bar{x} - a)\sqrt{n}}{S} = t_{n-1} \qquad \frac{S^2(n-1)}{\sigma^2} = \chi^2_{n-1}$$

Interval estimates

Now we know how to estimate true value of the measured quantity: we set it equal to the sample average; we estimate the variance by the value of the square root deviation. And so, we come to the issue of judging the quality of the said point estimates

If we have a point estimate (basing on sample size *n*) of some parameter of the distribution

$$\hat{\theta}_n \to \theta$$

Estimate

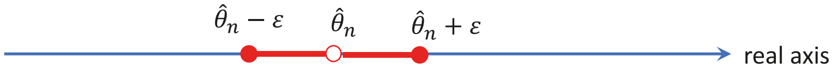
True value of the parameter

What we need is to establish quantitative answer to a question:
What is the link between the possible estimation **error** (accuracy) and **confidence probability** (reliability)?

$$w\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) \equiv \gamma$$

We need to interconnect the probability of the random event (the difference of the estimate does not surpass particular value of the deviation, in other words uncertainty) with boundaries of the interval, cast by this uncertainty

$$\left|\hat{\theta}_n - \theta\right| < \varepsilon \qquad \Rightarrow \qquad \hat{\theta}_n - \varepsilon < \theta < \hat{\theta}_n - \varepsilon$$



confidence interval

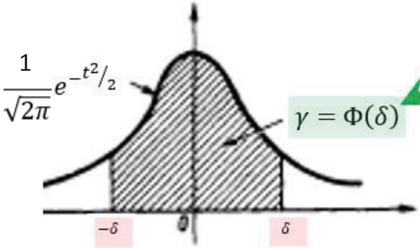
True value θ is somewhere inside confidence interval with confidence probability γ

Confidence interval for a single measurement with a known variance

The inequality $|x - a| < \varepsilon$ is equivalent to $\frac{|x-a|}{\sigma} < \frac{\varepsilon}{\sigma} \equiv \delta$. Therefore, their probabilities are equal: $w(|x - a| < \varepsilon) = w\left(\frac{|x-a|}{\sigma} < \frac{\varepsilon}{\sigma} \equiv \delta\right)$

$$w(|\delta_\gamma|) = \gamma$$

$$\Phi(\delta) = \frac{2}{\sqrt{2\pi}} \int_0^{\delta} e^{-t^2/2}\, dt = \gamma = w(|t| < \delta)$$



δ	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0238	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9189	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	8488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	99915	99917	99920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9938	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	—	—	—	—	—	—	—	—	—
5,0	0,99999994	—	—	—	—	—	—	—	—	—

Consider a task of judging about the confidence of a single measurement with known variance so that deviation from the true value will not surpass a value $k \cdot \sigma$, k is integer (1,2,3)

$$w(|x-a| \leq k\sigma) = w\left(\left|\frac{x-a}{\sigma}\right| \leq k = \Phi(k)\right)$$

$$w(|x - a| \leq 1 \cdot \sigma) = \Phi(1) = 0.6827$$

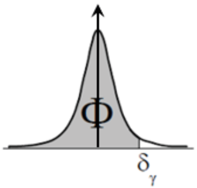
$$w(|x - a| \leq 2 \cdot \sigma) = \Phi(2) = 0.9545$$

$$w(|x - a| \leq 3 \cdot \sigma) = \Phi(3) = 0.9973$$

Empirical principle, "the rule of $3 \cdot \sigma$ ": if you perform a single measurement within a methodology and a setup with known general variance, then the probability to decline from the true value more than $\sqrt{DX}$ is less than 0.3 %

Alternatively, if the available data-base is only the Laplace function, then the procedure of getting mapping of γ onto δ_γ is as follows:

$$\Phi(\delta_\gamma) = \frac{1+\gamma}{2} = \int_{-\infty}^{\delta_\gamma} f(U)\,dU$$



γ	0.7500	0.9000	0.9500	0.9750	0.9900
$\Phi(\delta_\gamma)$	0.8750	0.9500	0.9750	0.9875	0.9950
δ_γ	1.1503	1.6449	1.9600	2.2414	2.5758

An interval estimate of a sample average for a known general assembly variance

$$\{x_i \,\epsilon\, N(a, \sigma_0^2)\} \Rightarrow \bar{x} \qquad \frac{\bar{x} - a}{\sigma_0 / \sqrt{n}} \,\epsilon\, N(0,1); \quad t = \frac{\bar{x} - a}{\sigma_0 / \sqrt{n}}; \quad w\big(|t| \le \delta_y\big) = \gamma$$

$$\gamma \rightarrow_{standard\ normal\ distribution} \rightarrow \pm\delta_\gamma$$

$$-\delta_y \le \frac{\bar{x} - a}{\sigma_0 / \sqrt{n}} \le \delta_y$$

Interval estimate for the expectation

$$\bar{x} - \boxed{\frac{\delta_\gamma \sigma_0}{\sqrt{n}}} \le a \le \bar{x} + \frac{\delta_{\boxed{\gamma}} \sigma_0}{\sqrt{n}}$$

ε, precision or error $\qquad$ γ, confidence probability

An exercise

A measurement of concentration in a probe is distributed in accord with a normal law with a known σ_0=2, the size of the probes' sample is 25. Find the precision of the average with a confidence 0.95

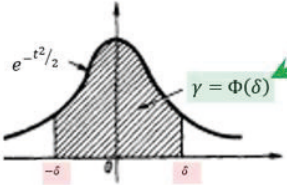
$$n = 25$$

$$\gamma = 0.95$$

$$\boldsymbol{\sigma_0 = 2}$$

$$\delta_\gamma = 1.96$$

$$\varepsilon = \frac{1.96 \cdot 2}{\sqrt{25}} = 0.784$$

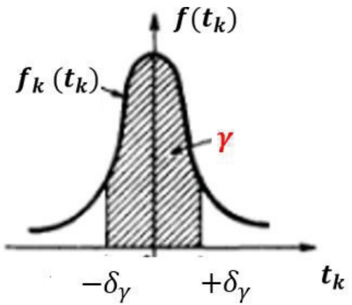


δ	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0238	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9189	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	8488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	99915	99917	99920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9938	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	—	—	—	—	—	—	—	—	—
5,0	0,99999994	—	—	—	—	—	—	—	—	—

An interval estimate of a sample average for an unknown general assembly variance

$$\frac{\bar{x}-a}{\sigma/\sqrt{n}} \epsilon\ N(0,1); \quad \frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1} \quad \Rightarrow \frac{\bar{x}-a}{S/\sqrt{n}} = t_{n-1}$$

$$w\left(\left|\frac{\sqrt{n}(\bar{x}-a)}{S}\right| \le \delta_\gamma\right) = \gamma$$



We set desired $\gamma; k = n-1 \Rightarrow_{Student\ distribution} \Rightarrow \ \delta_\gamma$

$$\bar{x} - \boxed{\frac{\delta_\gamma S}{\sqrt{n}}} \le a \le \bar{x} + \boxed{\frac{\delta_\gamma S}{\sqrt{n}}}$$

ε, precision (accuracy) or error

The Student distribution

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

The measurements of 16 probes provided mean average value of nanoobjects in the view-field of a microscope equals to 3000 species, square mean deviation is 20 units. Considering the number of nanoobjects conforming to the normal distribution, determine with confidence 0.9 the error bars for the expectation

The Student distribution

$$k = 15; \quad \gamma = 0.9$$

$$\Downarrow$$

$$\delta_\gamma = 1.753 \quad \Rightarrow$$

$$\Rightarrow error = \frac{1.753 \cdot 20}{4} = 8.765$$

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Measurements of the diameter of 20 pivots for a lab's set up gave an average $\overline{d} = 100$ mm; $S^2 = 256\ mm^2$. Diameters obey the normal law. Find the confidence probability that a true value of diameter belongs to the interval $a_d \epsilon \left(0.\ 9\overline{d}, 1.\ 1\overline{d}\right)$

$$\varepsilon = 0.1 \cdot \bar{d} = 10\,mm \quad \Rightarrow$$

Insofar as $\delta = \frac{\varepsilon\sqrt{n}}{S} = \frac{10\sqrt{20}}{16} = 3.125$ and $k = n - 1 = 24$,
one has $\gamma > 0.99$

An interval estimate of a variance

$$S - \varepsilon \le \sigma \le S + \varepsilon$$

$$1 - \varepsilon_1 \le \frac{\sigma}{S} \le 1 + \varepsilon_1, \qquad \varepsilon_1 = \frac{\varepsilon}{S}$$

Now, set confidence probability: γ If it is too big, it may well turn out that: $\varepsilon > 1$

$$max(0,1 - \varepsilon_1) \le \frac{\sigma}{S} \le (1 + \varepsilon_1) \qquad S \cdot (max[0; (1 - \varepsilon_1)]) \le \sigma \le S \cdot (1 + \varepsilon_1)$$

$$\frac{(n-1)}{(1+\varepsilon_1)^2} \le \frac{S^2(n-1)}{\sigma^2} \le \frac{(n-1)}{[max(0,1-\varepsilon_1)]^2}$$

$$\frac{S^2(n-1)}{\sigma^2} = \chi^2_{n-1}$$

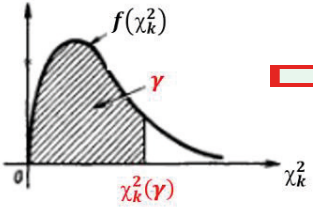
Recap: Newton – Raphson method

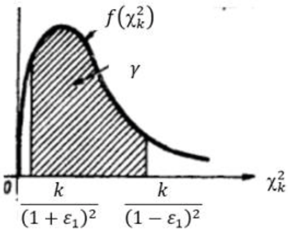
$$f(x) = 0$$

$$f\left(x^{(m+1)}\right) = f\left(x^{(m)}\right) + f'\left(x^{(m)}\right) \cdot \left(x^{(m+1)} - x^{(m)}\right)$$

$$f\left(x^{(m+1)}\right) = 0 \Rightarrow \boxed{x^{(m+1)} = x^{(m)} - \frac{f\left(x^{(m)}\right)}{f'\left(x^{(m)}\right)}}$$

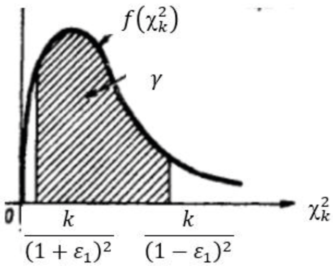
$$f'\left(x^{(m)}\right) \cong \frac{f\left(x^{(m)} + \Delta x\right)}{\Delta x}$$





$$\varepsilon_1^{(m)} \to \frac{k}{\left(1+\varepsilon_1^{(m)}\right)^2};\frac{k}{max^2\left[0,1-\varepsilon_1^{(m)}\right]} \to \Phi\left(\varepsilon_1^{(m)}\right) \to \left|\Phi\left(\varepsilon_1^{(m)}\right)-\gamma\right|$$

$$\varepsilon_1^{(m+1)} = \varepsilon_1^{(m)} - \frac{\Phi\left(\varepsilon_1^{(m)}\right)}{\Phi'\left(\varepsilon_1^{(m)}\right)}$$



$$w\left(\frac{k}{(1+\varepsilon_1)^2} \le \chi_k^2 \le \frac{k}{[max(0,1-\varepsilon_1)]^2}\right)$$

The algorithm goes as follows:

One sets $k = n - 1; \quad \gamma \quad \Rightarrow$

ε_1

$$S \cdot (max[0;(1-\varepsilon_1)]) \le \sigma \le S \cdot (1+\varepsilon_1)$$

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,568	906	1,602	2,946	6,923				
2	367	473	0,678	1,125	2,086	3,400	5,857	8,500	
3	290	370	482	0,730	1,270	1,932	3,000	4,200	9,00
4	248	306	398	563	0,941	1,382	2,056	2,700	5,00
5	221	277	348	475	738	1,104	1,594	2,00	3,80
6	200	251	308	416	623	0,918	1,306	1,650	3,00
7	185	232	290	380	576	800	1,143	1,393	2,50
8	173	216	269	354	516	713	0,986	1,225	2,05
9	162	202	252	329	476	650	889	1,094	1,75
10	153	192	239	304	442	596	814	0,980	1,50
12	140	176	218	276	388	527	700	840	1,30
14	130	162	290	252	357	468	620	740	1,14
16	122	150	188	236	325	422	564	671	1,02
18	115	143	177	223	297	390	500	600	0,92
20	108	136	168	210	282	370	480	567	85
25	096	122	148	187	247	317	408	485	70
30	0,088	111	137	172	226	281	369	425	60
35	085	101	127	156	207	261	347	400	56
40	076	095	119	146	193	342	312	375	52
50	068	084	105	133	174	212	270	311	45
60	062	077	095	122	155	193	242	283	40
70	057	072	088	112	145	180	222	250	37
80	054	067	082	103	138	167	200	236	35
90	051	063	078	096	131	151	192	220	32
100	048	060	074	092	125	146	184	200	30

With a confidence $\gamma = 0.98$ give an interval estimate of the general assembly variance $\sigma^2$for a sample of size 17 having an average square deviation $S^2 = 25$

$$\gamma = 0.98; \quad k = 17 - 1 = 16; \quad S = 5 \quad data - base\ source \Rightarrow \varepsilon_1 = 0.564$$

$$\boldsymbol{S \cdot (max[0; (1 - \varepsilon_1)]) \leq \sigma \leq S \cdot (1 + \varepsilon_1)}$$

$$S \cdot max(0; 1 - 0.564) = 5 \cdot (1 - 0.564) = 2.18; \qquad 5 \cdot (1 + 0.564) = 7.82$$

$$2.18 \leq \sigma \leq 7.82$$

$$4.7524 \leq \sigma^2 \leq 61.1524$$

With a confidence of 98 % this interval contains the true value of the general variance

This speculation may well be criticized: there are no grounds to suppose that the confidence interval is symmetrically allocated regarding the true value of the general-assembly square deviation (the square root of the general assembly variance).

So, we discuss an alternative approach to the issue:

$$1-\varepsilon_\gamma^{(1)} \le \frac{\sigma}{S} \le 1+\varepsilon_\gamma^{(2)}.$$

That is, we presume that probability, rather than precision, is distributed equally leftwards and rightwards with respect to the point estimate of the general variance. Yet again, we must exclude the situation when for some big confidence we get epsilon so big that the l.h.s. of the inequality is negative. Then, strictly speaking the condition reads:

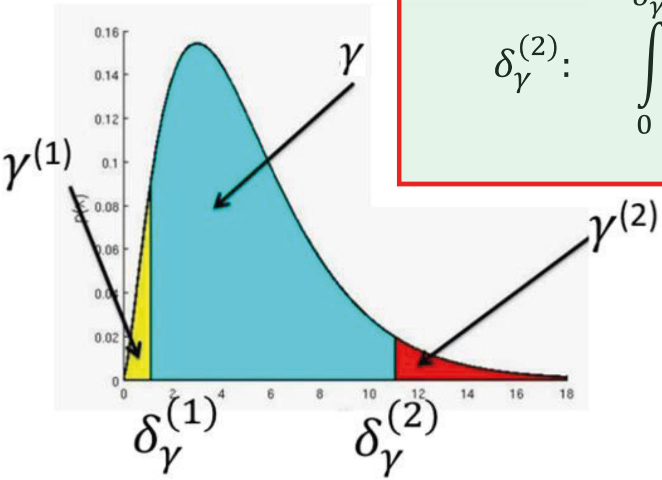
$max\left[0; 1-\varepsilon_\gamma^{(1)}\right] \le \frac{\sigma^2}{S^2} \le 1+\varepsilon_\gamma^{(2)}$ or equivalently:

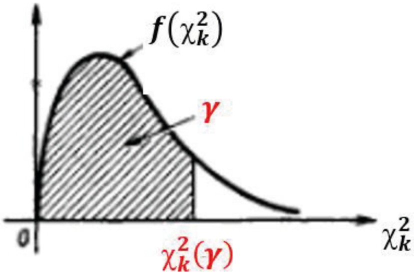
$$\underbrace{\boxed{\frac{n-1}{1+\varepsilon_\gamma^{(2)}}}}_{\delta_\gamma^{(1)}} \le \frac{S^2(n-1)}{\sigma^2} \le \underbrace{\boxed{\frac{n-1}{max\left[0; 1-\varepsilon_\gamma^{(1)}\right]}}}_{\delta_\gamma^{(2)}}$$

We set confidence γ, and along with it, a constraint for the probabilities of the χ^2_{n-1} to be left- and rightwards off the confidence interval, $\gamma^{(1)} = \gamma^{(2)}$.

$$\delta_\gamma^{(1)}: \quad \int_0^{\delta_\gamma^{(1)}} f(\chi_k^2)\,d(\chi_k^2) = \frac{1-\gamma}{2} = \gamma^{(1)}$$

$$\delta_\gamma^{(2)}: \quad \int_0^{\delta_\gamma^{(2)}} f(\chi_k^2)\,d(\chi_k^2) = \frac{1+\gamma}{2} = \gamma^{(2)}$$





$$\gamma = w(\chi_k^2 < \chi_k^2(\gamma))$$

k \ γ	0,50	0,70	0,80	0,90	0,95	0,98	0,99	0,999
1	0,455	1.074	1,642	2,71	3,84	5,41	6,64	10,83
2	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	5,35	7,23	8.56	10,64	12,59	15,03	16,81	22,5
7	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	9,34	11,78	13.44	15,99	18,31	21,2	23,2	29,6
11	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,5
14	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	15,34	18,42	20.5	23,5	26,3	29,6	32,0	39,3
17	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	26,3	30,3	32,9	36,7	40,1	44.1	47,0	55,5
28	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

Interval estimate of the variance, when the expectation is known

First, we show that an unbiased estimate in this case will be a quantity:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2$$

$$MS^2 = \frac{1}{n}M\sum_{i=1}^{n}(x_i - a)^2 = n \cdot \frac{1}{n}Dx = Dx = \sigma^2$$

$$S - \varepsilon \leq \sigma \leq S + \varepsilon$$

$$\varepsilon_1 = \frac{\varepsilon}{S}$$

$$max[0; 1 - \varepsilon_1] \leq \frac{\sigma}{S} \leq (1 + \varepsilon_1)$$

$$\frac{n}{(1 + \varepsilon_1)^2} \leq \frac{n \cdot S^2}{\sigma^2} \leq \frac{n}{max^2(0;\ 1 - \varepsilon_1)}$$

$$\Big\downarrow$$

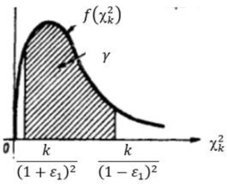
$$\chi_n^2$$

What's the distribution of the variable $\frac{nS^2}{\sigma^2}$?

$$\frac{x_i - a}{\sigma} \epsilon\ N(0,1) \Rightarrow \frac{n2}{\sigma^2} \equiv \sum_{i=1}^{n}\left(\frac{x_i - a}{\sigma}\right)^2 = \chi_n^2$$

$$\left.\begin{matrix}\gamma \\ k = n\end{matrix}\right\} \to \varepsilon_1$$

$$S(max[0; 1 - \varepsilon_1]) \leq \sigma \leq S(1 + \varepsilon_1)$$



Exercise

As a result of 50 analytical chemistry determinations of the known content of a gauging substance, a is thereby fixed and known, it was established that $S^2 = 2.56 \left(\frac{mmol}{L}\right)^2$. Assuming that determined concentration complies with a normal law, calculate the confidence of the confidence interval $(S - 0.16; S + 0.16)$

$S(1-\varepsilon_1) < \sigma < (1+\varepsilon_1)S \quad \varepsilon_1 = \frac{0.16}{\sqrt{2.56}} = 0.1 \qquad k = 50; \;\; \varepsilon_1 = 0.1 \;\; \rightarrow \;\; \gamma$

We need to find γ, consistent with $\varepsilon_1 = 0.1$. If we apply the Table, the only way to get through is linear interpolation:

$$\frac{\varepsilon_A - \varepsilon_B}{\gamma_A - \gamma_B} = \frac{\varepsilon_1 - \varepsilon_B}{\gamma_1 - \gamma_B} \Rightarrow \frac{0.021}{0.1} = \frac{0.016}{x}$$

$$x = 0.076 \Rightarrow \gamma = 0.076 + 0.6$$

With a confidence 67% the interval

$$S - 0.16 < \sigma < S + 0.16$$

will contain the true value of the variance

k \ γ	0,5	γ_B 0,6	γ_A 0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,568	906	1,602	2,946	6,923				
2	367	473	0,678	1,125	2,086	3,400	5,857	8,500	
3	290	370	482	0,730	1,270	1,932	3,000	4,200	9,00
4	248	306	398	563	0,941	1,382	2,056	2,700	5,00
5	221	277	348	475	738	1,104	1,594	2,00	3,80
6	200	251	308	416	623	0,918	1,306	1,650	3,00
7	185	232	290	380	576	800	1,143	1,393	2,50
8	173	216	269	354	516	713	0,986	1,225	2,05
9	162	202	252	329	476	650	889	1,094	1,75
10	153	192	239	304	442	596	814	0,980	1,50
12	140	176	218	276	388	527	700	840	1,30
14	130	162	290	252	357	468	620	740	1,14
16	122	150	188	236	325	422	564	671	1,02
18	115	143	177	223	297	390	500	600	0,92
20	108	136	168	210	282	370	480	567	85
25	096	122	148	187	247	317	408	485	70
30	0,088	111	137	172	226	281	369	425	60
35	085	101	127	156	207	261	347	400	56
40	076	095	119	146	193	342	312	375	52
50	068	084 ε_B	105 ε_A	133	174	212	270	311	45
60	062	077	095	122	155	193	242	283	40
70	057	072	088	112	145	180	222	250	37
80	054	067	082	103	138	167	200	236	35
90	051	063	078	096	131	151	192	220	32
100	048	060	074	092	125	146	184	200	30

Homework 1

An exercise on the maximum likelihood

Problem 1. The random variable t (days), time of the fail-safe work of the measuring sensor, has a distribution law:

$$f(t) = \vartheta exp\{-\vartheta t\}.$$

The table contains the data on the time of performance of 1000 sensors: first line gives varying average time of fail-safe time t_i and the 2nd line provides the information on the number of sensors, who showed this average time. By the maximum likelihood method give a point estimate of the unknown parameter ϑ

$t_i, days$	5	15	25	35	45	55	65
n_i	365	245	150	100	70	45	25

Problem 2. Find the minimal necessary number of measurements of the transition heat (the normal general assembly), which would ensure error not exceeding 0.3 kJ/mol with the confidence not less than 0.9, provided general assembly variance is 4 $(kJ/mol)^2$

Problem 3. Control test of 15 spectrometers established that the average functionality time (before the maintenance service) equals 3000 hours. There are reasons to reckon that the functionality time is distributed normally with a variance of 256 hours^2. Find the confidence of that the precision of the time estimate ε equals 10 hours

Problem 4. 10 measurements of a refractive index have been performed by one refractometer, the average square distribution S equaled 0.8. Find precision (an error bars) of the methodology and a confidence interval for the general variance, given the confidence for both distributions is taken equal to 0.95. Results of measurements are distributed in accord with a normal law

Interval estimates in the case, when there is no grounds to accept the normal distribution for the directly measured variable

$$w(|x-a| \geq \varepsilon) + w(|x-a| < \varepsilon) = 1$$

$$\underbrace{w(|x-a| < \varepsilon)}_{\gamma} = 1 - \underbrace{w(|x-a| \geq \varepsilon)}_{p} \geq 1 - \frac{DX}{\varepsilon^2}$$

$$p_{max} = w(|x-a| \geq \varepsilon) = w(|x-a| \geq k \cdot \sigma) = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

$$\boxed{\gamma_{min} = 1 - \frac{1}{k^2}}$$

$$k = 1 \quad \Rightarrow \quad x - \sigma \leq a \leq x + \sigma \quad \exists\, \gamma_{min} = 0.00$$

$$k = 2 \quad \Rightarrow \quad x - 2\sigma \leq a \leq x + 2\sigma \quad \exists\, \gamma_{min} = 0.75$$

$$k = 3 \quad \Rightarrow \quad x - 3\sigma \leq a \leq x + 3\sigma \quad \exists\, \gamma_{min} = 0.89$$

$$k = 5 \quad \Rightarrow \quad x - 5\sigma \leq a \leq x + 5\sigma \quad \exists\, \gamma_{min} = 0.96$$

What would be the interval estimate (given a certain confidence) for the expectation on the platform of the sample average, given the distribution law is unknown?

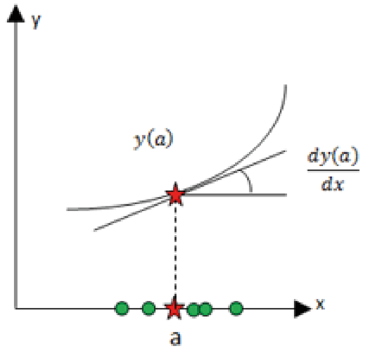
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$p_{max} = w(|\bar{x} - a| \geq \varepsilon) = w(|\bar{x} - a| \geq k\sigma) = \frac{\sigma_{\bar{x}}^2}{k^2\sigma^2} = \frac{\sigma^2}{nk^2\sigma^2} = \frac{1}{nk^2}$$

$$\boxed{\gamma = 1 - \frac{1}{nk^2}}$$

Interval estimate for the function of directly measured random variables

Practical cue: $\mu = \frac{K}{\theta} \cdot \frac{m_{subst.}}{m_{solvent}} \cdot 1000$



For a given set of conditions, $\{x_i\}$ is a sample providing point estimates of a sample

One-dimensional case: $y(x) = y(a) + \frac{dy(a)}{dx}(x-a) + \cdots;$

$$My(x) = My(a) + M\left[\frac{dy(a)}{dx}(x-a)\right] + \cdots$$

$$My(x) = y(a)$$

Multidimensional situation: $y = f(x_1, x_2, \dots, x_n).$

$$y = f(a_1, a_2, \dots, a_n) + \sum_i \frac{\partial f(\{a_i\})}{\partial x_i} \cdot (x_i - a_i) + \cdots \Rightarrow y \in N\left(a_{y,} \sigma_y^2\right)$$

$$a_y = My = Mf(a_1, a_2, \dots, a_n) + \sum_i M\left[\frac{\partial f(\{a_i\})}{\partial x_i} \cdot (x_i - a_i)\right] + \cdots$$

Brook Taylor
1685–1731

Obviously, if the Taylor expansion of $y = f(x_1, x_2, \dots, x_n)$ is linear truncated (the reason is in the predominant feasibility of small deviations from the expectations of directly measured quantities), then the expectation of y can be identified with function of expectations of directly measured variables

To the end of discussing σ_y^2 we recap some properties of the variance.

Let the variables ξ and η be random quantities, not necessarily independent.

Then one has: $D\xi(+\eta) = D\xi + D\eta + 2M[(\xi - M\xi)\cdot(\eta - M\eta)] = D\xi + D\eta + 2cov\xi(,\eta)$.

Generalizing to a case of linear combination one has:

$$D(C_1\xi + C_2\eta) =$$

$$= D(C_1\xi) + D(C_2\eta) + 2M\left[\left(C_1\xi - C_1a_\xi\right)\cdot\left(C_2\eta - C_2a_\eta\right)\right] = C_1^2D\xi + C_2^2D\eta + 2C_1C_2cov(\xi,\eta).$$

$$D\left(\sum_i C_i\xi_i\right) = \sum_i C_i^2D\xi_i + 2\sum_{i<j}\sum_j C_i\,C_j cov(\xi_i,\xi_j).$$

Now apply this formula to the Taylor expansion:

$$y(\{x_i\}) = f(a_1, a_2, \dots, a_n) + \sum_i \frac{\partial f(\{a_i\})}{\partial x_i} \cdot (x_i - a_i) + \cdots$$

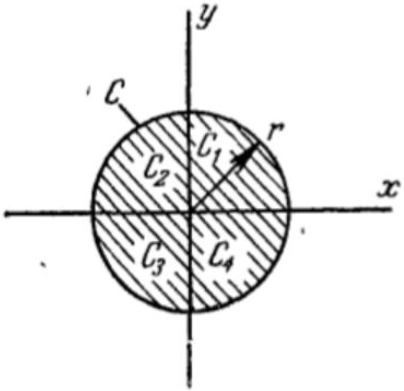
$$Dy = Df(a_1, a_{2,}, \dots., a_n) + D\left(\sum_i \frac{\partial f(\{a_i\})}{\partial x_i} \cdot (x_i - a_i)\right) = D\left(\sum_i \frac{\partial f(\{a_i\})}{\partial x_i} \cdot x_i\right) =$$

$$= \sum_i \left(\frac{\partial f(\{a_i\})}{\partial x_i}\right)^2 Dx_i + 2\sum_{i<j}\sum_j \frac{\partial f(\{a_i\})}{\partial x_i} \cdot \frac{\partial f(\{a_i\})}{\partial x_j} cov(x_i, x_j).$$

If the quantities of $\{x_i\}$, separately measured variables, are, by and large, inter-independent => $cov(x_i, x_j) = 0$.

And then

$$Dy = \sum_i \left(\frac{\partial f(\{a_i\})}{\partial x_i}\right)^2 Dx_i; \qquad \sigma_y^2 = \sum_i \left(\frac{\partial f(\{a_i\})}{\partial x_i}\right)^2 \sigma_{x_i}^2.$$



$$f(x,y)=\begin{cases} const, & x^2+y^2 \le r^2 \\ 0, & x^2+y^2 > r^2 \end{cases}$$

$$\iint\limits_{-\infty}^{+\infty} const \cdot dxdy = 1 \Rightarrow \quad const = \frac{1}{\pi r^2}$$

$$cov(x,y) = \iint xyf(x,y)dxdy = \frac{1}{\pi r^2}\iint xydxdy = 0$$

But, obviously x and y are dependent! If x=0, then $-r \le y \le r$; if $x = r$, then $y = 0$

No correlation does not necessarily signal statistical independence,
But in the case of normally distributed variables it does! (See next Chapter)

Both a_y and σ_y^2 are unknown. How can we estimate them?

In first place, we show, that $\bar{y} = f(\{\bar{x}_i\})$ is an unbiased estimate of a_y

$$M\bar{y} = M\left[f(\{a_i\}) + \sum_i \left(\frac{\partial f(\{a_i\})}{\partial x_i}\right)(\bar{x}_i - a_i) \cancel{+ \cdots}\right] \cong f(\{a_i\})$$

Neglection of the orders of magnitude above the linear one

Next, we show that $\left(\frac{\partial f(\{\bar{x}_i\})}{\partial x}\right)^2$ is an unbiased point estimate of $\left(\frac{\partial f(\{a_i\})}{\partial x}\right)^2$:

$$\frac{\partial f(\{\bar{x}_i\})}{\partial x} = \frac{\partial f(\{a_i\})}{\partial x} + \frac{\partial^2 f(\{a_i\})}{\partial x^2}(x_i - a_i) \cancel{+ \cdots}$$

$$\left(\frac{\partial f(\{\bar{x}_i\})}{\partial x}\right)^2 = \left(\frac{\partial f(\{a_i\})}{\partial x}\right)^2 + \cancel{\left[\frac{\partial^2 f(\{a_i\})}{\partial x^2}(\bar{x}_i - a_i)\right]^2} + 2\frac{\partial f(\{a_i\})}{\partial x}\frac{\partial^2 f(\{a_i\})}{\partial x^2}(\bar{x}_i - a_i)$$

$$M\left(\frac{\partial f(\{\bar{x}_i\})}{\partial x}\right)^2 = \left(\frac{\partial f(\{a_i\})}{\partial x}\right)^2$$

So, given the linear truncation of the Taylor expansion $\left(\frac{\partial f(\{\bar{x}_i\})}{\partial x}\right)^2$ is an unbiased point estimate of $\left(\frac{\partial f(\{a_i\})}{\partial x}\right)^2$

Now we introduce a quantity of a sample average deviation for y: $S_y^2 \equiv \sum_i \left(\frac{\partial f(\bar{x}_1, \bar{x}_2, \dots \bar{x}_n)}{\partial x_i}\right)^2 \cdot S_{x_i}^2$. It can be shown that this quantity renders an unbiased estimate of Dy.

Every $\bar{x}_i$ and $S_{x_i}^2$ are statistically independent by the virtue of Fisher lemma, hence $\frac{\partial f(\bar{x}_1, \bar{x}_2, \dots \bar{x}_n)}{\partial x_i}$ is independent of $S_{x_i}^2$. Then $MS_y^2 \equiv M\left[\sum_i \left(\frac{\partial f(\bar{x}_1, \bar{x}_2, \dots \bar{x}_n)}{\partial x_i}\right)^2 \cdot S_{x_i}^2\right] = \sum_i \left[M\left(\frac{\partial f(\bar{x}_1, \bar{x}_2, \dots \bar{x}_n)}{\partial x_i}\right)^2 \cdot MS_{x_i}^2\right] = \sum_i \left(\frac{\partial f(\{a_i\})}{\partial x_i}\right)^2 \cdot \sigma_i^2 = Dy = \sigma_y^2$.

Next, if the following assumption is applicable: $\sigma_y^2 \approx S_y^2$; in view of the fact: $y \in N(a_y, \sigma_y^2)$, the following random quantity holds the standard normal distribution:

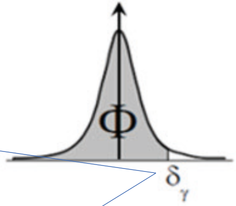
$$\frac{\bar{y} - a_y}{S_y} \in N(0, 1).$$

And so, setting γ, we pick up δ_γ from the data-base of the standard normal law.

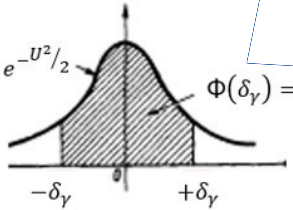
So, eventually the interval estimate of the function of directly measured random variables is:

$$\bar{y} - Sy \cdot \delta_\gamma \leq ay \leq \bar{y} + Sy \cdot \delta_\gamma.$$

$$\Phi(\delta_\gamma)=\frac{1+\gamma}{2}=\int_{-\infty}^{\delta_\gamma} f(U)\,dU$$



$$\bar{y}-S_y\cdot\delta_\gamma \le a_y \le \bar{y}+S_y\cdot\delta_\gamma$$



$$\Phi(\delta_\gamma)=\frac{2}{2\pi}\int_0^{\delta_\gamma} e^{-U^2/2}\,dU=\gamma=w(|U|<\delta_\gamma)$$

A simple flow-chart of giving an interval estimate of the indirectly measurable variable:

1. For each independent directly measured variable we get a sample $\left\{x_i^{(\alpha_i)}\right\}$, calculate $\overline{x_i}$ and $S_{x_i}^2$

   i — the number of directly measurable variable, α_i — sample size of the directly measured variable

2. Calculate numeric value $\overline{y} = f(\{\overline{x_i}\})$

3. Give explicit derivatives $\frac{\partial y}{\partial x_i}$ and calculate their numeric values in the point $\{\overline{x_i}\}$

4. Calculate numeric value of the quantity $S_y^2 \equiv \sum_i \left(\frac{\partial f(\overline{x}_1, \overline{x}_2, \ldots \overline{x}_n)}{\partial x_i}\right)^2 \cdot S_{x_i}^2$

5. Set desired value of the confidence γ and obtain the value of δ_γ

6. Eventually we get interval estimate for y:

$$\overline{y} - Sy \cdot \delta_\gamma \leq a_y \leq \overline{y} + S_y \cdot \delta_\gamma$$

Estimation of the limiting error

Another approach is to consider the absolute error as some small value, which is identified with the differential of functions against differentials of independent variables. Error transfer is determined by the rules of differentiation:

but it is taken into account that when adding and subtracting errors can always add up and you need to take the sum of the modules:

$$\boxed{dy} = \sum_i \left(\frac{\partial y}{\partial x_i}\right) \boxed{dx_i}$$

$$\boxed{\Delta y} = \sum_i \left|\left(\frac{\partial y}{\partial x_i}\right)\right| \cdot \boxed{|\Delta x_i|}$$

$$\boxed{\Delta y}$$

$$\bar{y} - S_y \cdot \delta_y \le a_y \le \bar{y} + S_y \cdot \delta_y$$

Rules of ob-truncation (rounding off)

1. The uncertainty (error) is to be rounded off to the first valid digit (significant figure) always rounding it up for a unity of the current digit order number:

$$\Delta x = \varepsilon = 0.237 \cong 0.3$$

2. Results of measurements, no matter direct or indirect, are rounded off to the precision of an uncertainty, i. e. the last significant figure should be in the same digit order number that the first significant figure in the “error”:

$$x \pm \Delta x = 243.871 \pm 0.026 = 243.87 \pm 0.03$$

Rounding off the result is provided by a simple truncation of figures, given the first cut off one does not surpass 5:

e. g. *rounding to the tenth of integer* $8.337 \cong 8.3$

3. If the first among the cut off figures is equal or greater than 5, than the last intact figures in increase by 1:

e. g. rounding to the hundredth $8.3351 \cong 8.34$

4. If the cut off figure is 5 and there no significant figures after it (or there are only zeros), then the last remaining figure is increased by unity if it is odd or leaved intact if it is even

e. g. *rounding to the hundredth digit order* $8.3350 \cong 8.34$ $8.3650 \cong 8.36$

Homework 1

Answers and discussion

An exercise on the maximum likelihood

Problem 1. The random variable t (days), time of the fail-safe work of the measuring sensor, has a distribution law:

$$f(t) = \vartheta exp\{-\vartheta t\}.$$

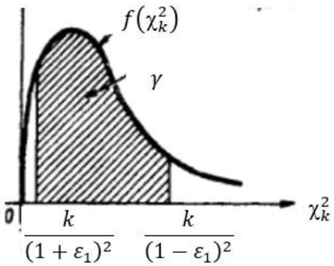
The table contains the data on the time of performance of 1000 sensors: first line gives varying average time of fail-safe time t_i and the 2[nd] line provides the information on the number of sensors, who showed this average time. By the maximum likelihood method give a point estimate of the unknown parameter ϑ

$t_i, days$	5	15	25	35	45	55	65
n_i	365	245	150	100	70	45	25

Problem 2. Find the minimal necessary number of measurements of the transition heat (the normal general assembly), which would ensure error not exceeding 0.3 kJ/mol with the confidence not less than 0.9, provided general assembly variance is 4 $(\mathrm{kJ/mol})^2$

Problem 3. Control test of 15 spectrometers established that the average functionality time (before the maintenance service) equals 3000 hours. There are reasons to reckon that the functionality time is distributed normally with a variance of 256 hours^2. Find the confidence of that the precision of the time estimate ε equals 10 hours

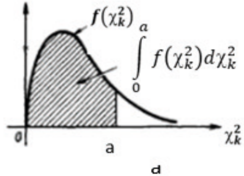
Problem 4. 10 measurements of refractive index have been performed by one refractometer, the average square distribution S equaled 0.8. Find precision (an error bars) of the instrument and confidence interval for the general variance, given the confidence for both distributions is taken equal to 0.95. Results of measurements are distributed in accord with a normal law

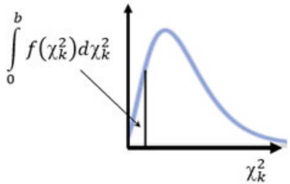


$$w\left(\frac{k}{(1+\varepsilon_1)^2} \le \chi_k^2 \le \frac{k}{[max(0, 1-\varepsilon_1)]^2}\right)$$

ε_1

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,568	906	1,602	2,946	6,923				
2	367	473	0,678	1,125	2,086	3,400	5,857	8,500	
3	290	370	482	0,730	1,270	1,932	3,000	4,200	9,00
4	248	306	398	563	0,941	1,382	2,056	2,700	5,00
5	221	277	348	475	738	1,104	1,594	2,00	3,80
6	200	251	308	416	623	0,918	1,306	1,650	3,00
7	185	232	290	380	576	800	1,143	1,393	2,50
8	173	216	269	354	516	713	0,986	1,225	2,05
9	162	202	252	329	476	650	889	1,094	1,75
10	153	192	239	304	442	596	814	0,980	1,50
12	140	176	218	276	388	527	700	840	1,30
14	130	162	290	252	357	468	620	740	1,14
16	122	150	188	236	325	422	564	671	1,02
18	115	143	177	223	297	390	500	600	0,92
20	108	136	168	210	282	370	480	567	85
25	096	122	148	187	247	317	408	485	70
30	0,088	111	137	172	226	281	369	425	60
35	085	101	127	156	207	261	347	400	56
40	076	095	119	146	193	342	312	375	52
50	068	084	105	133	174	212	270	311	45
60	062	077	095	122	155	193	242	283	40
70	057	072	088	112	145	180	222	250	37
80	054	067	082	103	138	167	200	236	35
90	051	063	078	096	131	151	192	220	32
100	048	060	074	092	125	146	184	200	30

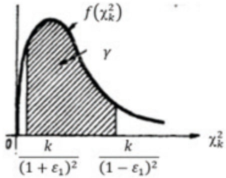




$$\Phi(\varepsilon_1^{(m)}) = \int_0^a f(\chi_k^2)d\chi_k^2 - \int_0^b f(\chi_k^2)d\chi_k^2$$

$$\varepsilon_1^{(m)} \rightarrow \overset{b}{\frac{k}{\left(1+\varepsilon_1^{(m)}\right)^2}};\overset{a}{\frac{k}{max^2\left[0,1-\varepsilon_1^{(m)}\right]}} \rightarrow \Phi\left(\varepsilon_1^{(m)}\right) \rightarrow \left|\Phi\left(\varepsilon_1^{(m)}\right)-\gamma\right|$$

$$\varepsilon_1^{(m+1)} = \varepsilon_1^{(m)} - \frac{\Phi\left(\varepsilon_1^{(m)}\right)}{\Phi'\left(\varepsilon_1^{(m)}\right)}$$



Next, determine the error-bars span **(precision) of the true value**

$$\varepsilon = \frac{S\delta_{\gamma}}{\sqrt{n}}, where\ \delta_{\gamma}$$

is the border value of the Student distribution

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	[illegible]	2,262	2,821	3,250	4,781
10	[illegible]	[illegible]	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699.	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

So, $\varepsilon = \frac{0.8}{\sqrt{10}} \cdot 2.262 = 0.5722$

Homework 2

Determining molecular weight of a substance by cryoscopy methodology

$$\mu = \frac{K}{\theta} \cdot \frac{m_{subst.}}{m_{solvent}} \cdot 1000; \qquad K = 1.853 \ \frac{g \cdot {}^{o}C}{mol}$$

θ Is the drop of of the melting temperature of the solution of a substance (admixture) against that in the pure solvent; $m_{subst.}$ — mass of a substance; $m_{solvent}$ — mass of a solvent (water in the particular example); K is the cryoscopy constant of a solvent (water)

Experimental data are:

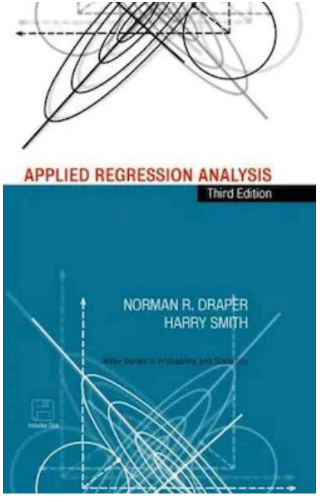
Number of a probe	θ (ºC)	$m_{subst.}$ (g)	$m_{solvent.}$ (g)
1	0.535	2.002	248.12
2	0.542	2.003	248.10
3	0.530	1.990	248.08
4	0.531	2.001	248.10
5	0.533	2.002	248.09

Estimate a confidence interval for the molecular mass of a dissolved substance with confidence probability 0.95

Part 2

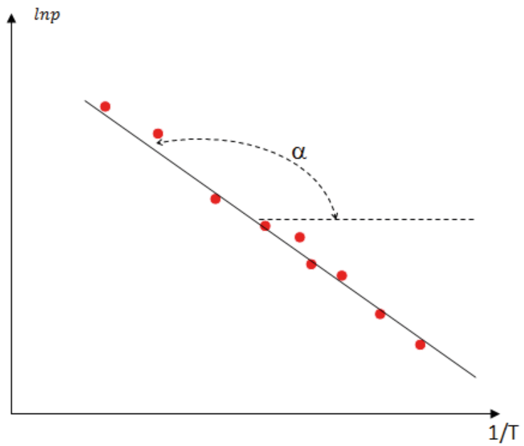
Square least regression

Norman R. Draper, Harry Smith. Applied Regression Analysis



Oftentimes an experiment is a study of the dependency of one quantity on another, both of them undergo variation in the same set of conditions. For example, we measure the electrical conductivity of a solution depending on the concentration of the electrolyte in it at fixed temperature, pressure and other external conditions.

$$\ln p = -B\frac{1}{T} + A$$



The value that we change at our discretion (in our example, concentration) is called the

predictor (a.k.a. factor)

and denotes specific elements of the array of its values

$$\{x_i\},\ i = 1, \dots n.$$

The quantities that take the values determined by the predictor (for us it is electrical conductivity), are called the

***response*,**

the elements of the array of its values are denoted

$$\{y_i\},$$

implying the correspondence of the elements of these two arrays when their numbers match. The response values are obviously inter-independent.

Experimentally established bijection predictor vs. response

$factor$	$response$
x_1	y_1
x_2	y_2
$\dots$	$\dots$
x_n	y_n

$$y(x_i) = \underbrace{\eta(x_i)}_{\text{regression}} + \varepsilon_i,$$

where $\eta\,(x)$ – functional dependence of the expectation of the response upon the current value of the predictor, and the measurement error with the number i is assumed to be a random variable obeying the normal distribution of

$$\varepsilon_i \in N\left(o, \sigma^2\right),$$

the variance being considered unknown, but having the same value for all response measurements. We look for the expectation as a predictor function in the form of a polynomial (let's call it a regression a.k.a. a statistical model):

$$\eta\,(x) = \underbrace{\theta_0 f_0(x) + \theta_1 f_1(x) + \cdots \theta_r f_r(x)}_{\text{regression}}$$

where analytical form of the function $\{f_\alpha\}$ is known, while coefficients $\{\theta_\alpha\}$ should be determined.

Taking into account the abovementioned, the results of the response measurements can be represented as a system of ratios:

$$\{y_i = \theta_0 f_0(x_i) + \theta_1 f_1(x_i) + \ \ldots + \theta_r f_r(x_i) + \varepsilon_{i\,,} \quad 1 \le i \le n$$

We consider this system as a basis for calculating unknown coefficients of the model. Let's rewrite it in a matrix form:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} f_0(x_1) & \cdots & f_r(x_1) \\ \vdots & \ddots & \vdots \\ f_0(x_n) & \cdots & f_r(x_n) \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ . \\ . \\ \theta_r \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

The matrix

$$A = \begin{pmatrix} f_0(x_1) & \cdots & f_r(x_1) \\ \vdots & \ddots & \vdots \\ f_0(x_n) & \cdots & f_r(x_n) \end{pmatrix}$$

is called the ***design matrix (a.k.a. plan matrix)***

We also introduce notation for vectors:

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ . \\ . \\ \theta_r \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix}; \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix}$$

We will mark the conjugation (a.k.a. transposition) of matrix objects with the upper index (superscript) *T*:

$$\varepsilon^{T} = (\boldsymbol{\varepsilon}_{1}, \boldsymbol{\varepsilon}_{2}, \dots, \boldsymbol{\varepsilon}_{n})$$

Then the system of the conditions can be rewritten:

$$\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\varepsilon}$$

We are up to a search for the coefficients (the entire vector of coefficients) from the condition of the minimum square of the length of the error vector, hence, the name of the method. The square of its length according to the Pythagorean theorem is:

The inner product

$$S^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - A\theta)^T (Y - A\theta) = \underbrace{Y^T Y}_{\varphi_1} - \underbrace{\theta^T A^T Y}_{\varphi_2} - \underbrace{Y^T A\theta}_{\varphi_3} + \underbrace{\theta^T A^T A\theta}_{\varphi_4}$$

Sine we are minimizing a scalar, whose algebraic sense is the sum of the squares of the difference between the regression and an experimental response, the method is termed the square least regression (= least squares regression)

S^2 is a scalar function of a vector argument θ. Before examining it for minimality, let's transform its expression somewhat

Consider the term φ_2 of the equation. First, focus at the product:

$$A^T Y = \begin{pmatrix} \sum_i f_0(x_i) y_i \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_i f_r(x_i) y_i \end{pmatrix}$$

Then the term φ_2 is entirely representable:

$$\theta^T A^T Y = \theta_0 \sum_{i=1}^{n} f_0(x_i) y_i + \theta_1 \sum_{i=1}^{n} f_1(x_i) y_i + \cdots \theta_r \sum_{i=1}^{n} f_r(x_i) y_i =$$

$$= y_1 \left[\theta_0 f_0(x_1) + \theta_1 f_1(x_1) + \cdots + \theta_r f_r(x_1)\right] + y_2\left[\theta_0 f_0(x_2) + \theta_1 f_1(x_2) + \cdots + \theta_r f_r(x_2)\right] +$$

$$+ \cdots y_n\left[\theta_0 f_0(x_n) + \theta_1 f_1(x_n) + \cdots + \theta_r f_r(x_n)\right] =$$

$$= y_1 \sum_{\alpha=0}^{r} \theta_\alpha f_\alpha(x_1) + y_2 \sum_{\alpha=0}^{r} \theta_\alpha f_\alpha(x_2) + \ldots y_n \sum_{\alpha=0}^{r} \theta_\alpha f_\alpha(x_n)$$

Consider the term φ_3

First, $A\theta = \begin{pmatrix} \sum_{\alpha=0}^{r} f_\alpha(x_1)\theta_\alpha \\ \sum_{\alpha=0}^{r} f_\alpha(x_2)\theta_\alpha \\ \cdot \\ \cdot \\ \sum_{\alpha=0}^{r} f_\alpha(x_n)\theta_\alpha \end{pmatrix}$

Then: $\varphi_3 = Y^T A\theta = y_1 \sum_{\alpha=0}^{r} f_\alpha(x_1)\theta_\alpha + y_2 \sum_{\alpha=0}^{r} f_\alpha(x_2)\theta_\alpha + \cdots + y_n \sum_{\alpha=0}^{r} f_\alpha(x_n)\theta_\alpha$

$$\varphi_2 = \varphi_3$$

Owing to that $\varphi_2 = \varphi_3$, the scalar purpose function takes the form:

$$S^2(\theta) = \underbrace{Y^T Y}_{\varphi_1} - \underbrace{2\theta^T A^T Y}_{2\varphi_2} + \underbrace{\theta^T A^T A \theta}_{\varphi_4}$$

One intermediary moment. Consider the product

$$A^T A$$

and designate it as

$$C \equiv A^T A$$

What's the size of the matrix?

Fisher information matrix

$$C_{ij} = \sum_{\alpha} a_{i\alpha}^T a_{\alpha j} = \sum_{\alpha} a_{\alpha i} a_{\alpha j} \qquad C_{ji} = \sum_{\alpha} a_{j\alpha}^T a_{\alpha i} = \sum_{\alpha} a_{\alpha j} a_{\alpha i}$$

$$C_{ij} = C_{ji}$$

Fisher matrix is a symmetric one

$S^2(\theta)$ is a scalar function of a vector argument

A necessary condition for its extremum is the zero value of its gradient in the coordinate space of the coefficient vector. The gradient is a linear operator; in order to calculate it, it is necessary to alternately find the gradients of the terms of the right side of equation

$$\varphi_2 = \theta^T A^T Y$$

$$\nabla\varphi_2 = \left(\frac{\partial\varphi_2}{\partial\theta_0}; \frac{\partial\varphi_2}{\partial\theta_1}; \dots \frac{\partial\varphi_2}{\partial\theta_r}\right)^T \qquad \frac{\partial\varphi_2}{\partial\theta_\alpha} = \sum_{i=1}^{n} f_\alpha(x_i) y_i$$

$$\nabla\varphi_2 = A^T Y$$

Now, turn to the term φ_4:

$$\varphi_4 = \theta^T A^T A\theta = \theta^T C\theta = (\theta_0, \theta_1, \dots, \theta_r)\begin{pmatrix} \sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha \\ \sum_{\alpha=0}^{r} C_{1\alpha}\theta_\alpha \\ \cdot \\ \cdot \\ \cdot \\ \sum_{\alpha=0}^{r} C_{r\alpha}\theta_\alpha \end{pmatrix} =$$

$$= \theta_0 \sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha + \theta_1 \sum_{\alpha=0}^{r} C_{1\alpha}\theta_\alpha + \cdots + \theta_r \sum_{\alpha=0}^{r} C_{r\alpha}\theta_\alpha$$

$$\nabla\varphi_4 = \left(\frac{\partial\varphi_4}{\partial\theta_0}; \frac{\partial\varphi_4}{\partial\theta_1}; \dots \frac{\partial\varphi_4}{\partial\theta_r}\right)^T$$

We'll consider the first line of this vector as an example:

$$\frac{\partial\varphi_4}{\partial\theta_0} = \sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha + \theta_0 C_{00} + \theta_\alpha C_{10} + \cdots + \theta_r C_{r0} = \sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha + \sum_{\alpha=0}^{r} \theta_\alpha C_{\alpha 0}$$

Then, using the symmetry of the Fisher matrix, we obtain:

$$\frac{\partial\varphi_4}{\partial\theta_0} = 2\sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha$$

For the entire gradient of φ_4 it holds:

$$\nabla\varphi_4 = 2\begin{pmatrix} \sum_{\alpha=0}^{r} C_{0\alpha}\theta_\alpha \\ \sum_{\alpha=0}^{r} C_{1\alpha}\theta_\alpha \\ \cdot \\ \cdot \\ \cdot \\ \sum_{\alpha=0}^{r} C_{r\alpha}\theta_\alpha \end{pmatrix} = C\theta = A^T A\theta$$

Eventually for the condition of the minimum length of the coefficient-dependent scalar we get:

$$\nabla S^2 = -2A^T Y + 2A^T A\theta = 0$$

And from here the minimizer of the scalar function is a vector of point estimates:

$$\boxed{\tilde{\theta} = \left(A^T A\right)^{-1} A^T Y}$$

NB: $\left(A^T A\right)^{-1} = (C)^{-1}$ – *reciprocal matrix a.k.a. matrix inverse*

Thus, the random vector $\widetilde{\theta}$ represents an estimate of an unknown vector θ, the set of components of which gives point estimates for the expectation of a random response at any value of the predictor. Let's see if this estimate is unbiased. The expectation of a matrix object is a matrix of the expectations of all its elements. Therefore:

$$MX = \begin{pmatrix} MX_1 \\ MX_2 \\ . \\ . \\ MX_k \end{pmatrix}$$

$$\begin{aligned} M\widetilde{\theta} &= M(A^TA)^{-1}A^TY = (A^TA)^{-1}A^TMY = \\ &= (A^TA)^{-1}A^TM(A\theta + \varepsilon) = \\ &= (A^TA)^{-1}A^TMA\theta + (A^TA)^{-1}A^TM\varepsilon = \\ &= (A^TA)^{-1}A^TMA\theta + 0 = (A^TA)^{-1}A^TA\theta = \theta \end{aligned}$$

Here familiar theorems about expectation and the fact that the expectation of the error vector is zero are used

Residual sum of squares

$$y = A\theta + \varepsilon$$

Vector of measured values

$$\tilde{y} = A\tilde{\theta}$$

Vector of predicted values

$$R^2 = \sum_{i=1}^{n}\left(y_i - \widetilde{\theta_0} f_0(x_i) - \widetilde{\theta_1} f_1(x_i) - \cdots \widetilde{\theta_r} f_r(x_i)\right)^2 = \sum_i \delta_i^2 = \delta^T \delta$$

1. It characterizes how the measured response values differ from the predicted ones
2. It is demonstrated in statistics, that: $M\left(\frac{R^2}{n-r-1}\right) = \sigma^2$, i. e. $\frac{R^2}{n-r-1}$ is an unbiased estimate of the measurement variance

From here stems an important inference. The elements of the predicted vector $\{\tilde{y}_i\}$ are within our assumption normally distributed random variables. However, they are bound with constraints $\tilde{\theta}_\alpha = \tilde{\theta}_\alpha(\{\tilde{y}_i\})$

Therefore, there are only $n - r - 1$ among them, who are statistically independent. Hence, following the Fisher lemma the unbiased estimate of the variance being resolved as a sum of $n - r - 1$ independent squares of the differences between the predicted and measured responses, features as the variable χ^2_{n-r-1}

$$\frac{R^2}{\sigma^2} = \chi^2_{n-r-1}$$

Worked example

Assignment: give point estimates for the parameters of the regression model

$$y_i = \theta_0 x_i + \theta_1 x_i^{-1} + \varepsilon_i$$

i	x_i	y_i	i	x_i	y_i
1	0,770	2,900	10	2,910	4,100
2	1,004	2,700	11	3,198	4,362
3	1,222	2,700	12	3,544	4,629
4	1,457	2,800	13	3,861	4,900
5	1,735	2,957	14	4,078	5,175
6	2,039	3,150	15	4,213	5,453
7	2,317	3,367	16	4,370	5,733
8	2,531	3,600	17	4,643	6,016
9	2,705	3,845	18	5,022	6,300
			19	5,394	6,586

$$A = \begin{pmatrix} f_0(x_1) & f_1(x_1) \\ \vdots & \vdots \\ f_0(x_{19}) & f_1(x_{19}) \end{pmatrix}$$

$$A = \begin{bmatrix} 0{,}770 & 1{,}299 \\ 1{,}004 & 0{,}996 \\ 1{,}222 & 0{,}818 \\ 1{,}457 & 0{,}686 \\ 1{,}735 & 0{,}576 \\ 2{,}039 & 0{,}490 \\ 2{,}317 & 0{,}432 \\ 2{,}531 & 0{,}395 \\ 2{,}705 & 0{,}370 \\ 2{,}910 & 0{,}344 \\ 3{,}198 & 0{,}313 \\ 3{,}544 & 0{,}282 \\ 3{,}861 & 0{,}259 \\ 4{,}078 & 0{,}245 \\ 4{,}213 & 0{,}237 \\ 4{,}370 & 0{,}229 \\ 4{,}643 & 0{,}215 \\ 5{,}022 & 0{,}199 \\ 5{,}394 & 0{,}185 \end{bmatrix}$$

$$A^T Y = \begin{pmatrix} 276{,}621 \\ 31{,}127 \end{pmatrix}$$

$$C = A^T A$$

$$C = \begin{pmatrix} 206{,}988 & 19{,}000 \\ 19{,}000 & 5{,}523 \end{pmatrix}$$

$$C = \begin{pmatrix} 206{,}988 & 19{,}000 \\ 19{,}000 & 5{,}523 \end{pmatrix}$$

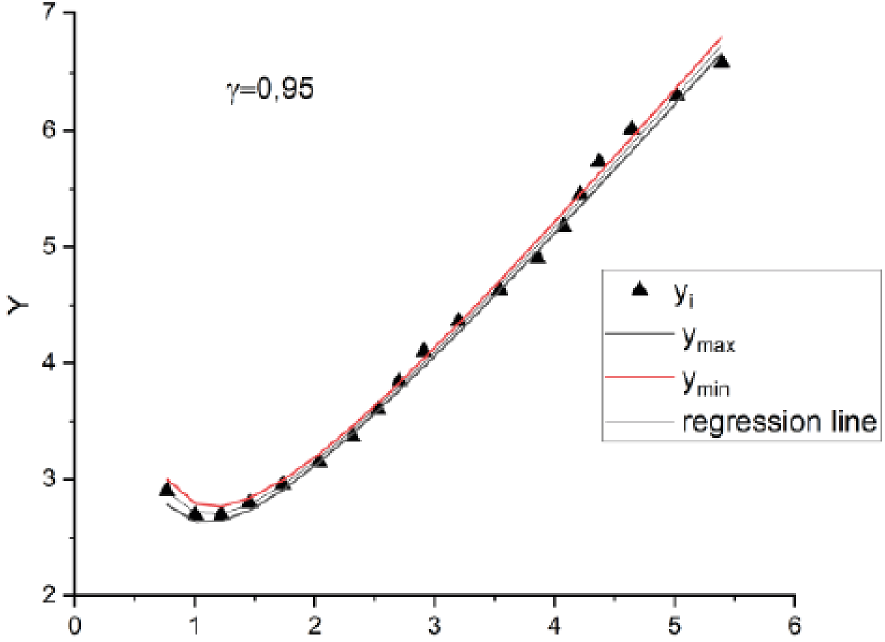
$$\|C^{-1}\|_{ij} = \frac{1}{det(C)} \cdot (-1)^{i+j} A_{iJ}(C)$$

$$A_{iJ}(C) \; - algebraic\ cofactor\ of\ matrix\ element\ \|C\|_{ij}$$

$$C^{-1} = \begin{pmatrix} 0{,}007 & -0{,}024 \\ -0{,}024 & 0{,}265 \end{pmatrix}$$

$$\hat{\theta} = \begin{pmatrix} 1{,}197 \\ 1{,}518 \end{pmatrix}$$

$$\boxed{y_i = 1{,}197 x_i + 1{,}518 x_i^{-1}}$$


γ=0,95
Y
y_i
y_{max}
y_{min}
regression line
7
6
5
4
3
2
0
1
2
3
4
5
6

$$\tilde{\boldsymbol{\theta}} = \left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{Y}$$

It is seen that the r.h.s. side is a linear transform of a random normal vector $\boldsymbol{Y}$. Therefore, according to the theorem on the composition of normal laws, parameter estimates are also normally distributed.

But!!!

Each of the coefficients $\widetilde{\boldsymbol{\theta}_{\alpha}}$ is expressed in terms of all of the measured response values, thus there may well be a statistical dependence of the predicted coefficients on each other, which is mediated by this dependence.

How can a description of the statistical inter-dependence of quantities be set up?

$$cov(X_1, X_2) = M[(X_1 - MX_1)(X_2 - MX_2)]$$

How can we generalize the definition of numerical characteristics for random vector quantities?
The vector expectation holds:

$$MX = \begin{pmatrix} MX_1 \\ MX_2 \\ . \\ . \\ MX_k \end{pmatrix}$$

The statistical dependence of the coordinates of a random vector is characterized by the so-called covariance matrix:

$$\begin{pmatrix} cov(X_1, X_1) & \cdots & cov(X_1, X_k) \\ \vdots & \ddots & \vdots \\ cov(X_k, X_1) & \cdots & cov(X_k, X_k) \end{pmatrix} = K(X)$$

In case $x_i \epsilon N\left(a_i, \sigma_i^2\right)$ $\qquad cov(X_i, X_i) = DX_i = \sigma_{x_i}^2$

To the end of simplicity of the rearrangements to come, we introduce a notion of the outer product of matrices:

$$X \otimes X^T = \begin{pmatrix} x_1 x_1 & \cdots & x_1 x_k \\ \vdots & \ddots & \vdots \\ x_k x_1 & \cdots & x_k x_k \end{pmatrix}$$

Henceforth, take care to distinguish it from the inner product, viz.

$$B^T B = \begin{pmatrix} \sum_{\alpha=1}^{k} B_{1\alpha} B_{\alpha 1} \\ \vdots \\ \sum_{\alpha=1}^{k} B_{k\alpha} B_{\alpha k} \end{pmatrix}$$

Via the introduced notion, the covariance matrix can be represented as an expectation of an outer product:

$$K(X) = M[(X - MX)\otimes(X - MX)^T]$$

$$M[(X - MX)\otimes(X - MX)^T] =$$

$$\begin{bmatrix} M(X_1 - MX_1)^2 & M(X_1 - MX_1)(X_2 - MX_2) & \ldots\ldots & M(X_1 - MX_1)(X_K - MX_K) \\ M(X_2 - MX_2)(X_1 - MX_1) & M(X_2 - MX_2)^2 & \ldots\ldots & M(X_2 - MX_2)(X_K - MX_K) \\ \ldots. & \ldots. & \ldots. & \ldots. \\ (X_K - MX_K)(X_1 - MX_1) & M(X_K - MX_K)(X_1 - MX_1) & \ldots. & M(X_K - MX_K)^2 \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma_1^2 & cov(x_1, x_2) & \ldots. & cov(x_1, x_k) \\ cov(x_2, x_1) & \sigma_2^2 & \ldots. & cov(x_2, x_k) \\ \ldots. & \ldots. & \ldots. & \ldots. \\ cov(x_k, x_1) & cov(x_k, x_2) & \ldots. & \sigma_k^2 \end{bmatrix}$$

Now in a Euclidean space, built on the axes corresponding to a set of components of a random vector, we define some linear transform:

$$Z = BX$$

Matrix B is not random (we set it at our discrete). Then its covariance matrix follows the above gotten presentation:

$$K(Z) = M\left[\left((BX - M(BX))\otimes(BX - M(BX))^{T}\right)\right] =$$

$$= M\left[(BX - BMX)\otimes\left((BX - BMX)^{T}\right)\right] = M[B(X - MX)\otimes(BX - MX)^{T}B^{T}] =$$

$$= BM[(X - MX)\otimes(BX - MX)^{T}]B^{T} = BK(X)B^{T}$$

Applying this scheme to the point estimate of the regression parameters, we get:

$$K(\widetilde{\theta}) = (A^T A)^{-1} A^T K(Y) \left[(A^T A)^{-1} A^T \right]^T = (A^T A)^{-1} A^T \sigma^2 I A \left[(A^T A)^{-1} \right]^T$$
$$= (A^T A)^{-1} \sigma^2 A^T A \left[(A^T A)^{-1} \right] = \sigma^2 C^{-1} I = \sigma^2 C^{-1}$$

$$I = \begin{pmatrix} 1 & 0 \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \ldots & 1 \end{pmatrix}$$

The circumstances employed up here are:

1. $A^T A$ is a Fisher matrix, it is a symmetric one, which means that the inverse matrix is also symmetric, and

2. the conjugation of a symmetric matrix keeps it unchanged

3. Next, the covariance matrix of the measured response values

$K(Y) = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}$ **— is a diagonal matrix (the elements of a random vector are independent of each other)**

4. Finally, the product of the matrix by its inverse is the unit matrix I

Note, that the vector of the predicted response values is a linear transform of the estimates' vector:

$$\widetilde{Y} = A\widetilde{\theta}$$

Hence, its covariance matrix is expressed as follows:

$$K(\widetilde{Y}) = AK(\widetilde{\theta})A^T = \sigma^2 AC^{-1}A^T$$

That is, the square root of the variance of the predicted parameters can be estimated from the diagonal elements of this covariance matrix:

$$\sigma_{\widetilde{\theta}_\alpha} = K_{\alpha\alpha}^{1/2}(\widetilde{\theta}) = \sigma\left\|C^{-1}\right\|_{\alpha\alpha}^{1/2}$$

From the linear transform $\widetilde{\theta} = \left(A^T A\right)^{-1} A^T Y$ and the fact that $\widetilde{\theta_\alpha}$ is the unbiased estimate of θ_α we see:

$$\widetilde{\theta_\alpha} - \theta_\alpha \in N\left(0, \sigma^2_{\widetilde{\theta}_\alpha}\right)$$

Thus, taking into account that:

$$\frac{R^2}{\sigma^2} = \chi^2_{n-r-1} \quad \overset{\text{Statistically independent}}{\Longleftrightarrow} \quad \frac{\widetilde{\theta_\alpha} - \theta_\alpha}{\sigma \| C^{-1} \|^{1/2}_{\alpha\alpha}} \in N(0,1),$$

a **Student variable** can be set up:

$$\frac{\widetilde{\theta_\alpha} - \theta_\alpha}{\sigma \| C^{-1} \|^{1/2}_{\alpha\alpha}} \cdot \frac{\sigma \sqrt{(n - r - 1)}}{R} = t_{n-r-1}$$

Then the confidence interval for the true value of each of the model parameters is:

$$\widetilde{\theta_\alpha} - \delta_{n-r-1}(\gamma) \cdot \frac{\left\|C^{-1}\right\|_{\alpha\alpha}^{1/2} \cdot R}{\sqrt{(n-r-1)}} \leq \theta_\alpha \leq \widetilde{\theta_\alpha} + \delta_{n-r-1}(\gamma) \cdot \frac{\left\|C^{-1}\right\|_{\alpha\alpha}^{1/2} \cdot R}{\sqrt{(n-r-1)}}$$

Here $\delta_{n-r-1}(\gamma)$ the percent points of the Student's distribution with the degree of freedom $n-r-1$, corresponding to the confidence probability γ

But! We should be mindful of the feasible interdependence between the elements of the vector $\widetilde{\theta_\alpha}$

Later we will give account of the joint confidence domain for the elements of $\widetilde{\theta_\alpha}$

Since the elements of the vector $\widetilde{\theta}$ are unbiased estimates of the corresponding elements of the vector θ, the variable $\widetilde{y}_i - My_i$ holds:

$$\widetilde{y}_i - My_i = \sum_{\alpha=0}^{r} [\widetilde{\theta_\alpha} - \theta_\alpha] f_\alpha(x_i) \ \epsilon \ N\left(0, \widetilde{\sigma}^2_{\widetilde{y}_i}\right),$$

where the variance can be determined *via* diagonal elements of the matrix $K(\widetilde{Y})$, viz., $\|AC^{-1}A^T\|_{ii}$:

$$\widetilde{\sigma}_{\widetilde{y}_i} = K_{ii}^{1/2}(\widetilde{y}) = \sigma\|AC^{-1}A^T\|_{ii}^{1/2}$$

Alas, we don't know σ^2

$$\frac{\widetilde{y}_i - My_i}{\sigma\|AC^{-1}A^T\|_{ii}^{1/2}} \in N(0,1)$$

Taking into consideration that

$$\frac{R^2}{\sigma^2} = \chi^2_{n-r-1}, \quad \xLeftrightarrow{\text{Statistically independent}} \quad \frac{\tilde{y}_i - My_i}{\sigma\|AC^{-1}A^T\|_{ii}^{1/2}} \in N(0,1)$$

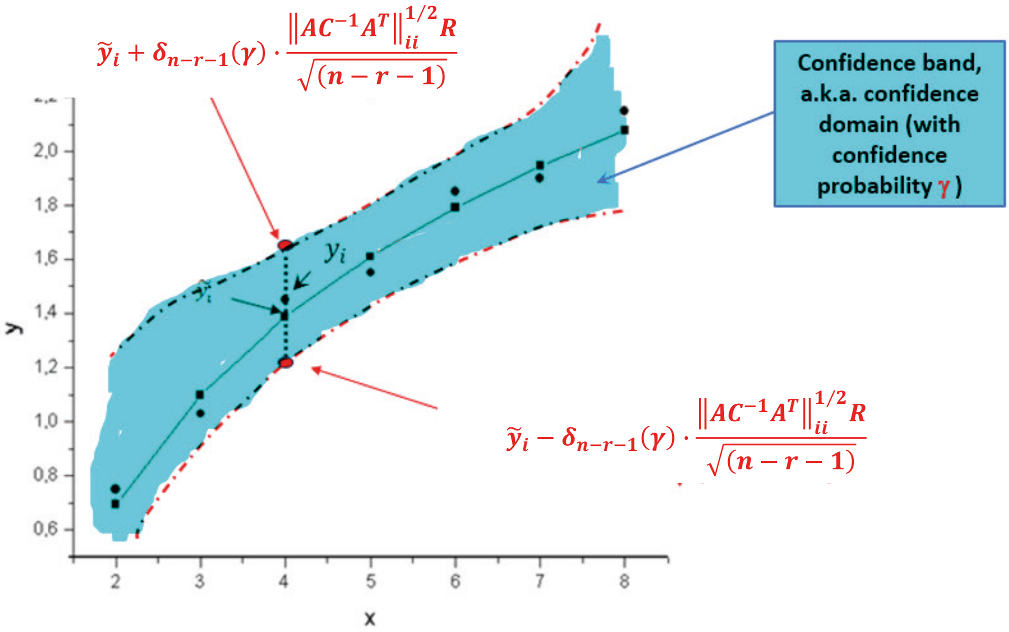
Yet another Student variable can be constructed:

$$\frac{\tilde{y}_i - My_i}{\cancel{\sigma}\|AC^{-1}A^T\|_{ii}^{1/2}} \cdot \frac{\cancel{\sigma}\sqrt{(n-r-1)}}{R} = t_{n-r-1}$$

The confidence interval for the true response value in each of the predictor-response pair is expressed by the following condition:

$$\tilde{y}_i - \delta_{n-r-1}(\gamma) \cdot \frac{\left\|AC^{-1}A^T\right\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}} \leq My_i \leq \tilde{y}_i + \delta_{n-r-1}(\gamma) \cdot \frac{\left\|AC^{-1}A^T\right\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}}$$

$\delta_{n-r-1}(\gamma)$ is gotten from the Student distribution law



$\tilde{y}_i + \delta_{n-r-1}(\gamma) \cdot \frac{\|AC^{-1}A^T\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}}$
$\tilde{y}_i - \delta_{n-r-1}(\gamma) \cdot \frac{\|AC^{-1}A^T\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}}$
Confidence band, a.k.a. confidence domain (with confidence probability γ)
y_i
$\tilde{y}_i$
y
x
2,0
1,8
1,6
1,4
1,2
1,0
0,8
0,6
2
3
4
5
6
7
8

Multi-dimensional normal law

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \cdot exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-a_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]\right\}$$

$$a_i = Mx_i \qquad \sigma_i^2 = Dx_i \qquad \rho = \frac{cov(x_1, x_2\,)}{\sqrt{Dx_1 \cdot Dx_2}} \qquad K(x) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

If ρ = 0, then variables are non-correlated. As for the probability, its distribution function gets the form:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot exp\left\{-\left[\frac{(x_1-a_1)^2}{2\sigma_1^2} + \frac{(x_2-a_2)^2}{2\sigma_2^2}\right]\right\} = \prod_{i=1,2}\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) exp\left\{-\frac{(x_i-a_i)^2}{2\sigma_i^2}\right\}$$

$$f(x_1, x_2) = f(x_1) \cdot f(x_2)$$

Non-correlated normal random variables are independent

Multi-dimensional normal law

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \cdot exp\left\{\underbrace{\boxed{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-a_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]}}_{F}\right\}$$

$$a_i = Mx_i \qquad \sigma_i^2 = Dx_i \qquad \rho = \frac{cov(x_1, x_2)}{\sqrt{Dx_1 \cdot Dx_2}} \qquad K(x) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$det\,K(x) = \sigma_1^2\sigma_2^2 - \rho\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2) \qquad K^{-1}(x) = \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \cdot \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$

Consider a product: $(x-a)^T K^{-1}(x)(x-a) =$

$$= \frac{1}{(1-\rho^2)}\begin{pmatrix} x_1 - a_1 & x_2 - a_2 \end{pmatrix}\begin{pmatrix} \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2} & -\rho\frac{\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2} \\ -\rho\frac{\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2} & \frac{\sigma_1^2}{\sigma_1^2\sigma_2^2} \end{pmatrix}\begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} =$$

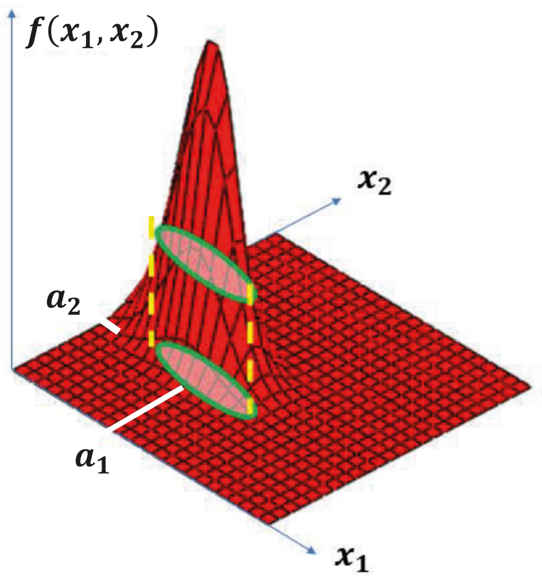
$$= \frac{1}{(1-\rho^2)}\begin{pmatrix} x_1 - a_1 & x_2 - a_2 \end{pmatrix}\begin{pmatrix} \frac{(x_1-a_1)}{\sigma_1^2} - \rho\frac{x_2-a_2}{\sigma_1\sigma_2} \\ -\rho\frac{x_1-a_1}{\sigma_1\sigma_2} + \frac{x_2-a_2}{\sigma_2^2} \end{pmatrix} =$$

$$= \frac{1}{(1-\rho^2)}\left[\frac{(x_1-a_1)^2}{\sigma_1^2} - \rho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} - \rho\frac{(x_2-a_2)(x_1-a_1)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]$$

So, $F = -\frac{1}{2}(x-a)^T K^{-1}(x)(x-a)$ $\quad$ $\det K(x) = \sigma_1^2\sigma_2^2 - \rho\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$

$$\Downarrow$$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\det K(x)}} exp\left\{-\frac{1}{2}(x-a)^T K^{-1}(x)(x-a)\right\}$$



The constant value of the distribution function and, thereby, the locus of small domains with equal probability is given by the constant value of the argument in the exponent:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{detK}} exp\left\{-\frac{1}{2}(x-a)^T K^{-1}(x-a)\right\}$$

$$\frac{1}{2}(x-a)^T K^{-1}(x)(x-a) = const$$

$$-ln\left[2\pi\sqrt{detK(x)} \cdot f(x_1, x_2)\right] = const$$

What is the probability, that a pair of random coordinates is within the ellipsoid cast with the above condition?

$\frac{1}{2}(x-a)^T K^{-1}(x)(x-a) = const$ The type of the expression on the l.f.s. is: $\varphi(x,x) = x^T A x$

An equivalent explicit algebraic expression holds:

$$\varphi(x,x) = \underbrace{a_{11}x_1x_1 + a_{21}x_1x_2 + \cdots + a_{21}x_2x_1 + \ldots + a_{nn}x_nx_n}_{\text{Quadratic form}}$$

Coefficients of this expression form a symmetric matrix: $a_{ij} = a_{ji}$

Set of numbers $\{x_i\}$ may be considered as coordinates of some vector in a certain frame

In a specially chosen coordinate frame this expression can be brough to the so-called canonical form, wherein the matrix of the quadratic form turns out to be a diagonal one:

$$\varphi(\tilde{x},\tilde{x}) = \sum_{i=1}^{n} \tilde{a}_{ii}\tilde{x}_i^2 = \tilde{x}^T \tilde{A} \tilde{x} \quad \Rightarrow \quad \tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{a}_{nn} \end{pmatrix}$$

The question stands: how to organize this special coordinate frame?

Imagine we have found it, then there must be an orthogonal linear transform inter-relating two coordinate frames, the one where the form has the original presentation, and a new one, where the form is canonical and the matrix of the form is diagonal

We denote the orthogonal linear transform from the new to the old frame as *B*: $x = B\tilde{x}$

Then in the new frame the quadratic form gets the presentation: $\varphi(\tilde{x}, \tilde{x}) = (B\tilde{x})^T AB\tilde{x} = \tilde{x}^T B^T AB\tilde{x}$

$$\tilde{A} = B^T AB$$

For any given vector, whose coordinates are presented in two different frames the numeric value of the quadratic form will stay constant… — Why?

It is established in the linear algebra, that a matrix $\widetilde{A}$ would be a diagonal one and the quadratic form would get a canonical representation, if the columns of the matrix $\boldsymbol{B}$ are the eigenvectors of the initial symmetric matrix $\boldsymbol{A}$ (in the initial coordinate frame), coordinates being those in a new frame build up so that its axes are co-linear with eigenvectors of the initial symmetric matrix. Therefore, the new coordinate frame is set on the eigenvalues of the initial matrix.

Worked example

Bring a quadratic form to a canonical representation and its matrix to a diagonal form

$$\varphi(x, x) = 2x_1x_2 \qquad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

***Df*: eigenvector x and eigenvalue λ are holding the following property of the linear transform presented by a matrix A:**

$$Ax = \lambda x$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

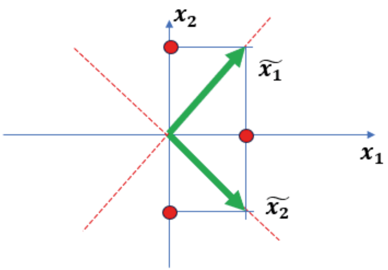
$$\begin{cases} -\lambda x_1 + x_2 = 0 \\ x_1 - \lambda x_2 = 0 \end{cases}$$

Nontrivial solutions for the homogeneous linear system exist, if the determinant of the system is zero: $\lambda^2 - 1 = 0$

1. λ **= +1**, then both of the system equations establish a locus $x_1 = x_2$,which is a bisector of the 1st and 3rd quadrants of the "old" frame. If we impose a constraint that its length is unity and at our discretion chose positive direction in the 1st quadrant. Then the vector is $\left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right\}$
2. λ **= –1**, similarly, it signifies a locus $x_1 = -x_2$, which is a bisector of the 2nd and 4th quadrants We chose a unit vector $\left\{\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right\}$

So far, two properties of the eigenvectors are exemplified, viz.:

1. Eigenvectors $\{\widetilde{x}_i\}$ are orthogonal;
2. Eigenvalues $\{\lambda_i\}$ are real



$$B = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \qquad AB = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$B^T AB = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Question en route: why the value of the quadratic form for any given vector will not alter with a change in the coordinate frame?

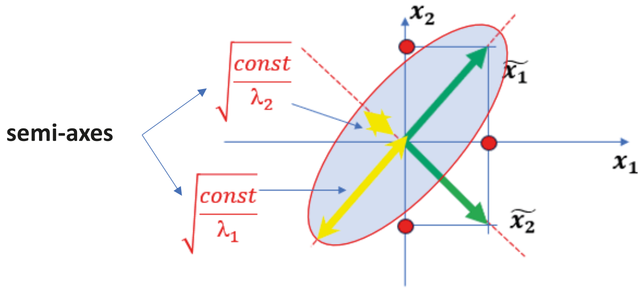
$$\lambda_1 = +1 \qquad \lambda_2 = -1$$

$$\varphi(x, x) = 2x_1x_2 = \tilde{x}_1^2 - \tilde{x}_2^2 = \varphi(\tilde{x}, \tilde{x})$$

Now, if we set $f(\widetilde{x}, \widetilde{x}) = const$,

it will cast a locus of vectors holding on this condition for $\lambda_i > 0$ in a coordinate frame, set up on eigenvectors of a linear transform **A**

$$\sum_{i=1}^{n} \widetilde{a}_{ii}(\widetilde{x}_i)^2 = \sum_{i=1}^{n} \lambda_i (\widetilde{x}_i)^2 = const \qquad \sum_{i=1}^{n} \left(\frac{\widetilde{x}_i}{\sqrt{\dfrac{const}{\lambda_i}}} \right)^2 = 1$$



If we set a condition:

$$\sum_{i=1}^{n} \lambda_i (\widetilde{x}_i)^2 \leq const \qquad \sum_{i=1}^{n} \left(\frac{\widetilde{x}_i}{\sqrt{\dfrac{const}{\lambda_i}}} \right)^2 \leq 1,$$

it will establish a set of vectors ending up in the interior of the elliptic domain

So, the issue stands: to establish the relationship between the spread a the random vector coordinate along a semi-axis (eigenvector of covariance inverse) and the confidence

Now let's return to the normal distribution law of a multidimensional vector $\tilde{\theta}$. The locus of vectors, who's ending points in the coordinate frame of the vector coordinates are constrained with equal probability of occurrence, is determined with a constant value of the argument in the exponent:

$$w\left[\left(\tilde{\theta}-\theta\right)^{T}K^{-1}\left(\tilde{\theta}\right)\left(\tilde{\theta}-\theta\right)=const\ (\gamma)\right]=\gamma$$

We pose a more loose question: how to find a domain in the space consistent with a constraint:

$$w\left[\left(\theta-\tilde{\theta}\right)^{T}K^{-1}\left(\tilde{\theta}\right)\left(\theta-\tilde{\theta}\right)\leq\delta_{\gamma}\right]=\gamma$$

The sequence of terms in the vector is changed… — why?

$$K(\tilde{\theta})=\sigma^{2}C^{-1} \quad \Longleftrightarrow \quad K^{-1}=\frac{1}{\sigma^{2}}C$$

$$\Downarrow$$

$$\frac{1}{\sigma^{2}}\left(\tilde{\theta}-\theta\right)^{T}C\left(\tilde{\theta}\right)\left(\tilde{\theta}-\theta\right)=const$$

Consider a partial case of a symmetric matrix 2×2

$$B = \frac{1}{\sigma^2} D = \frac{1}{\sigma^2}\begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} \qquad detB = \frac{1}{\sigma^4}\left(D_{11}D_{12} - D_{12}^2\right); \quad adB = \frac{1}{\sigma^2}\begin{pmatrix} D_{22} & -D_{12} \\ -D_{12} & D_{11} \end{pmatrix}$$

$$B^{-1} = \frac{\sigma^4}{\left(D_{11}D_{12} - D_{12}^2\right)} \cdot \frac{1}{\sigma^2}\begin{pmatrix} D_{22} & -D_{12} \\ -D_{12} & D_{11} \end{pmatrix} = \frac{\sigma^2}{detD}\begin{pmatrix} D_{22} & -D_{12} \\ -D_{12} & D_{11} \end{pmatrix} = \sigma^2 D^{-1}$$

$$detB^{-1} = \frac{\sigma^4}{(detD)^2}\left(D_{11}D_{12} - D_{12}^2\right) = \frac{\sigma^4}{detC}; \quad adB^{-1} = \frac{\sigma^2}{detD}\begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}$$

$$\left(B^{-1}\right)^{-1} = \frac{detD}{\sigma^4} \cdot \frac{\sigma^2}{detD}\begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} = \frac{1}{\sigma^2}\begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} = B$$

$$\left(\sigma^2 D^{-1}\right)^{-1} = \frac{1}{\sigma^2} D$$

$$K^{-1}(\tilde{\theta}) = \frac{1}{\sigma_0^2\sigma_1^2(1-\rho^2)}\begin{pmatrix} \sigma_1^2 & -\rho\sigma_0\sigma_1 \\ -\rho\sigma_0\sigma_1 & \sigma_0^2 \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma_0^2(1-\rho^2)} & -\dfrac{\rho}{(1-\rho^2)}\cdot\dfrac{1}{\sigma_0\sigma_1} \\ -\dfrac{\rho}{(1-\rho^2)}\cdot\dfrac{1}{\sigma_0\sigma_1} & \dfrac{1}{\sigma_1^2(1-\rho^2)} \end{pmatrix}$$

$$\widehat{K^{-1}} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \Rightarrow \quad \left(\widehat{K^{-1}}\right)^{-1} = \begin{pmatrix} \dfrac{1}{a} & 0 \\ 0 & \dfrac{1}{b} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_0^2 & 0 \\ 0 & \hat{\sigma}_1^2 \end{pmatrix}$$

$$K^{-1} = \frac{1}{\sigma^2}C \quad \Rightarrow \quad \widehat{K^{-1}} = \begin{pmatrix} \dfrac{\lambda_0}{\sigma^2} & 0 \\ 0 & \dfrac{\lambda_1}{\sigma^2} \end{pmatrix} \Rightarrow \quad \left\|\widehat{K^{-1}}\right\|_{\alpha\alpha} = \frac{1}{\hat{\sigma}_\alpha^2} = \frac{\lambda_\alpha}{\sigma^2}$$

$$\boxed{\hat{\sigma}_\alpha^2 = \frac{\sigma^2}{\lambda_\alpha}}$$

$$\theta - \tilde{\theta} \equiv \breve{\theta}$$

We shift the zero-point of the frame for random vectors to the point of the predicted values

$$D(x + const) = Dx; \quad cov(x_1 - const_1, x_2 - const_2) = cov(x_1, x_2)$$

Why? Prove it, PLS

It would leave the matrix intact

$$\frac{1}{\sigma^2}\breve{\theta}^T C(\tilde{\theta})\breve{\theta} = const$$

Eventually, we switch (as has been shown above in an example for a quadratic form) to a set of eigenvectors of the Fisher matrix $\hat{\theta}$. It gives a diagonal representation of the matrix $\hat{C}$

Eigenvalues of the Fisher matrix

$$\hat{\theta}^T \begin{pmatrix} \frac{\lambda_0}{\sigma^2} & 0 & 0 \dots 0 \\ 0 & \frac{\lambda_1}{\sigma^2} & 0 \dots 0 \\ \dots & \dots & \dots \\ 0 & 0 & 0 \quad \frac{\lambda_r}{\sigma^2} \end{pmatrix} \hat{\theta} = const$$

The meaning of a diagonal element of the diagonalized inverse of the covariance matrix is: $\frac{1}{\sigma^2_{\hat{\theta}_\alpha}}$

As we have established, $\sigma^2_{\hat{\theta}_\alpha} = \frac{\sigma^2}{\lambda_\alpha}$

$$\tilde{\theta}_\alpha - \theta_\alpha \equiv \breve{\theta}_\alpha \epsilon\, N(o, \sigma^2_\alpha) \;\;\rightarrow\;\; \hat{\theta}_\alpha \epsilon N\left(0, \sigma^2_{\hat{\theta}_\alpha}\right)$$

$$\frac{\hat{\theta}_\alpha}{\sigma / \sqrt{\lambda_\alpha}} \in N(0,1) \quad => \quad \sum_{\alpha=0}^{r} \left(\frac{\hat{\theta}_\alpha}{\sigma / \sqrt{\lambda_\alpha}} \right)^2 = \chi^2_{r+1}$$

All the variables are 1) standard normal and 2) inter-independent: all the elements off the diagonal of K and K^{-1} are zeros

But! The values σ^2 are unknown

So, we have: $$\sum_{\alpha=0}^{r}\frac{\hat{\theta}_\alpha^2}{\left(\sigma/\sqrt{\lambda_\alpha}\right)^2}=\chi^2_{r+1}$$ However, we are mindful of that: $$\frac{R^2}{\sigma^2}=\chi^2_{n-r-1}$$

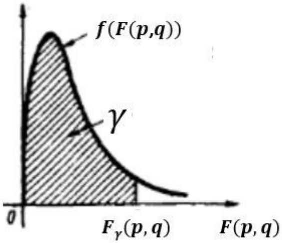
σ^2 is the same in all of the measurements, but is **unknown!**

Ronald Aylmer Fisher
1890–1962

And now we take a chance to praise the achievements of the already mentioned R. Fisher:

Fisher random variable: $$\frac{\chi^2_p/p}{\chi^2_q/q}\equiv F(p,q)$$

Fisher distribution for $\gamma = 0.95$



We set $\gamma, p = r + 1, q = n - r - 1$

$$w(F(p,q) < F_\gamma(p,q)) = \gamma$$

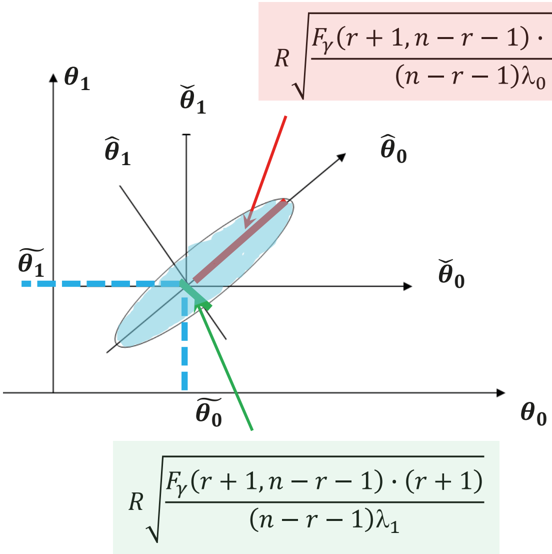
q	p = 1	2	3	4	5	6	8	12	24	∞
1	161,40	199,50	215,70	224,60	230,20	234,00	238,90	243,90	249,00	254,30
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,33	3,51	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

We set $\gamma, p = r+1, q = n-r-1$

$$\gamma \quad \Rightarrow \quad \frac{\left[\sum_{\alpha=0}^{r} \frac{\widehat{\theta_{\alpha}^{2}}}{\left(\frac{\cancel{\sigma}}{\sqrt{\lambda_{\alpha}}}\right)^{2}}\right](n-r-1)\cdot \cancel{\sigma^{2}}}{R^{2}(r+1)} \le F_{\gamma}(r+1, n-r-1)$$

$$\gamma \quad \Rightarrow \quad \frac{\left[\sum_{\alpha=0}^{r} \frac{\widehat{\theta_{\alpha}^{2}}}{\left(\frac{\sigma}{\sqrt{\lambda_{\alpha}}}\right)^{2}}\right](n-r-1)\cdot \sigma^{2}}{R^{2}(r+1)\cdot F_{\gamma}(r+1, n-r-1)} \quad \le \quad 1$$

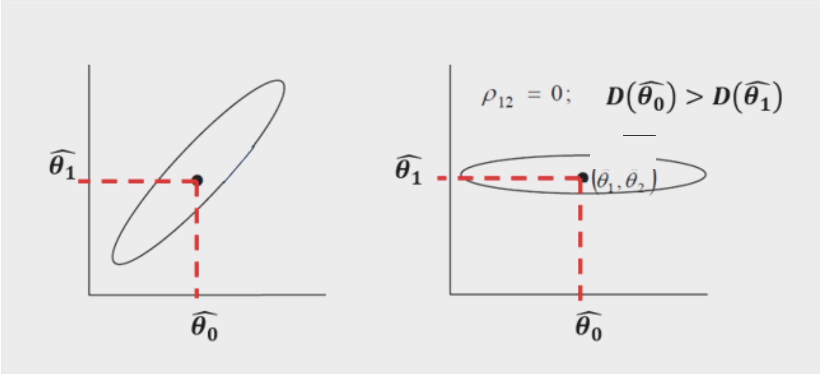
$$R\sqrt{\frac{F_{\gamma}(r+1, n-r-1)\cdot(r+1)}{(n-r-1)\lambda_{\alpha}}}$$



Semi-axes of the joint confidence ellipsoid are:

$$\boldsymbol{R\sqrt{\frac{F_\gamma(r+1,n-r-1)\cdot(r+1)}{(n-r-1)\lambda_\alpha}}}$$

Joint confidence domain of the regression parameters



parameters are non-correlated

$\widehat{\theta_0}$ is insignificant

Flow-chart of the linear regression of the experimental data

1. Setting the form of the regression, calculating the design matrix, Fisher matrix and point estimates of the regression parameters

$$\eta(x) = \theta_0 f_0(x) + \theta_1 f_1(x) + \cdots \theta_r f_r(x) \qquad A = \begin{pmatrix} f_0(x_1) & \cdots & f_r(x_1) \\ \vdots & \ddots & \vdots \\ f_0(x_n) & \cdots & f_r(x_n) \end{pmatrix} \qquad C \equiv A^T A \qquad \tilde{\theta} = \left(A^T A\right)^{-1} A^T Y$$

2. Calculation of the predicted regression, residual sum of squares and covariance matrices for the parameters and predicted regression values: $\tilde{y} = A\tilde{\theta}$; $R^2 = (\tilde{y} - y)^T(\tilde{y} - y)$; $K(\tilde{\theta}) = \sigma^2 C^{-1}$; $K(\tilde{Y}) = AK(\tilde{\theta})A^T = \sigma^2 A C^{-1} A^T$

3. Calculation of the individual interval estimates of the parameters and confidence band for the predicted regression values:

$$\widetilde{\theta_\alpha} - \delta_{n-r-1}(\gamma) \cdot \frac{\left\|C^{-1}\right\|_{\alpha\alpha}^{1/2} \cdot R}{\sqrt{(n-r-1)}} \le \theta_\alpha \le \widetilde{\theta_\alpha} + \delta_{n-r-1}(\gamma) \cdot \frac{\left\|C^{-1}\right\|_{\alpha\alpha}^{1/2} \cdot R}{\sqrt{(n-r-1)}},$$

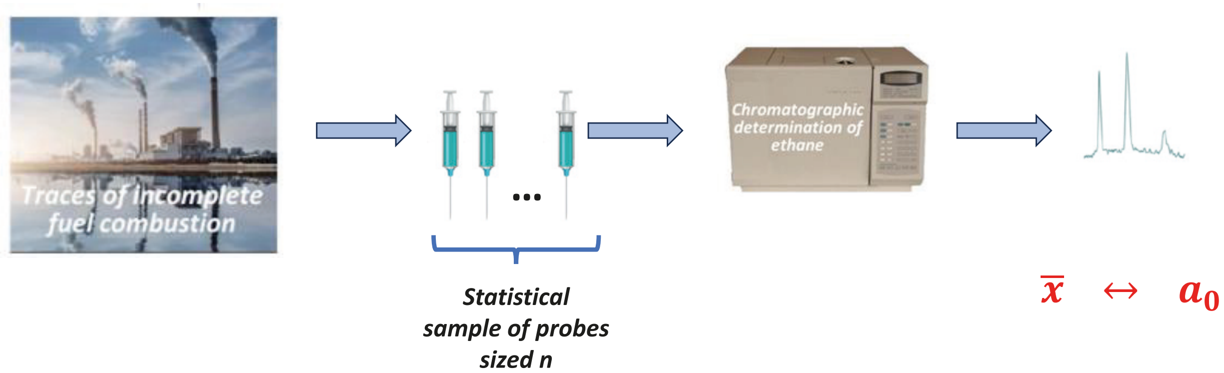
$$\tilde{y}_i - \delta_{n-r-1}(\gamma) \cdot \frac{\left\|AC^{-1}A^T\right\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}} \le My_i \le \tilde{y}_i + \delta_{n-r-1}(\gamma) \cdot \frac{\left\|AC^{-1}A^T\right\|_{ii}^{1/2} R}{\sqrt{(n-r-1)}}$$

4. Finding eigenvectors of the Fisher matrix, diagonalization of the matrix, finding semi-axes of the joint confidence ellipsoid for the regression parameters

$$semi-axis\ along\ \widetilde{\theta_\alpha} = R\sqrt{\frac{F_\gamma(r+1, n-r-1) \cdot (r+1)}{(n-r-1)\lambda_\alpha}}$$

Part 3
Hypotheses testing

Hypotheses about the average given the variance is known or unknown



Whether the expectation a_0 is lower against the maximum permissible concentration?

Any statement about the distribution law is a hypothesis

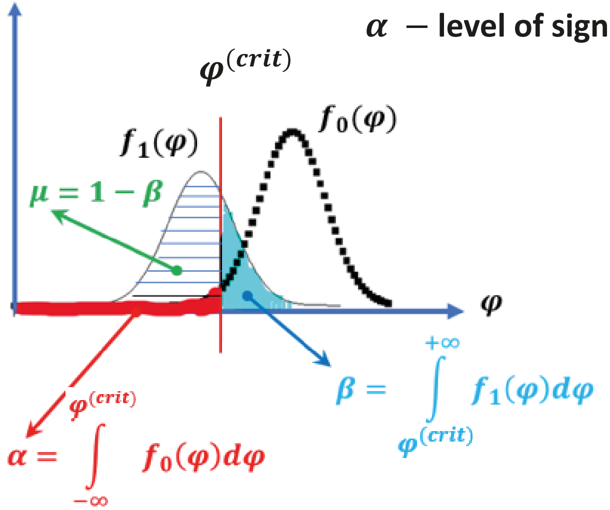
The methodology of statistical hypothesis testing consists in that a hypothesis is juxtaposed with some other statement about the distribution law: these two assertions are termed the basic and the alternative hypotheses

Both hypotheses are juxtaposed via analysis of the behavior of a quantity, whose value is determined by a sample values and whose distribution law is known

This random quantity (why is it random?) and its distribution law are termed a criterion

The distribution of the criterion within the basic and alternative hypotheses casts a **critical domain** for the basic hypothesis

α − level of significance $\qquad \mu$ − strength of the test



The probability to reject a correct hypothesis is
$\alpha = \int_{-\infty}^{\varphi^{(crit)}} f_0(\varphi)d\varphi$
The probability to accept a hypothesis, whilst it is incorrect $\beta = \int_{\varphi^{(crit)}}^{+\infty} f_1(\varphi)d\varphi$
The probability to reject an incorrect hypothesis is
$\mu = 1 - \beta = \int_{-\infty}^{\varphi^{(crit)}} f_1(\varphi)d\varphi$
The probability of accepting a basic hypothesis given it is a correct one =$1 - \alpha = \int_{\varphi^{(crit)}}^{+\infty} f_0(\varphi)d\varphi$

The better is α, the worse is μ. The practice is such that regularly α is set equal ***0.05***; the value of μ should be calculated from $f_1(\varphi^{(crit)})$

Types of hypotheses we will consider:

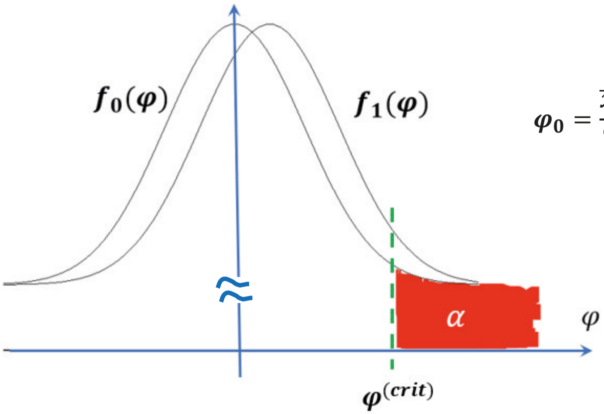
- **Hypothesis about the expectation from the sample given the variance is known**
- **Hypothesis about the expectation from the sample given the variance is unknown**
- **Hypothesis about variance from the sample**
- **Hypothesis about equality of average values from two sample having different variances**
- **Hypothesis about equality of averages from two samples having the same, but unknown variance**

Hypotheses about the average given the variance is known

Let there be a sample $\{x_1, x_1, \dots, x_1\}$;
the variance is known $\sigma^2 = \sigma_0^2$

Hypothesis H_0: $a = a_0$
Hypothesis H_1: $a = a_1 > a_0$

How to chose a criterion? $\varphi = \frac{\bar{x} - a}{\sigma_0 / \sqrt{n}}$



$$\varphi_0 = \frac{\bar{x} - a_0}{\sigma_0 / \sqrt{n}}; \quad \varphi_1 = \frac{\bar{x} - a_1}{\sigma_0 / \sqrt{n}} = \frac{\bar{x} - (a_0 + \Delta a)}{\sigma_0 / \sqrt{n}} = \varphi_0 - \frac{\Delta a}{\sigma_0 / \sqrt{n}}$$

So, an expedient thing is to set a critical domain on the r.h.s. of the real axis of the criterion

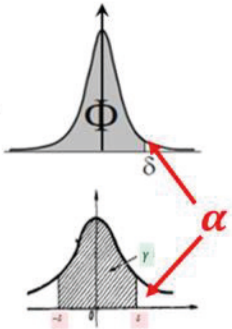
In order to get the boundary point of the one-sided critical domain one needs the database of the Laplace function. There are several types of presenting these data. Most often are cases **A** and **B**

A

$$\Phi(\delta_\gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\delta_\gamma} exp\left\{-\frac{x^2}{2}\right\} dx$$

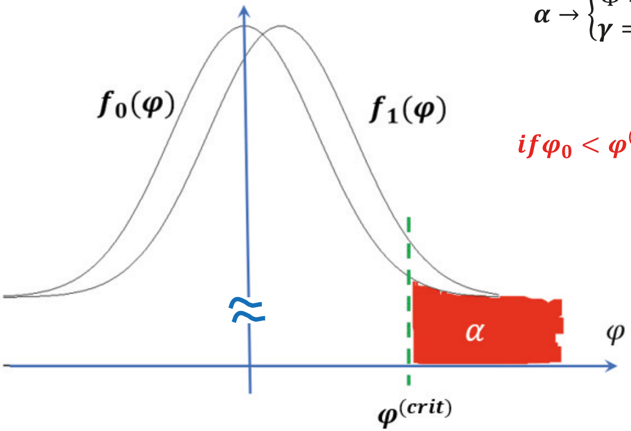
B

$$\gamma = \frac{2}{\sqrt{2\pi}} \int_0^{\delta_\gamma} exp\left\{-\frac{x^2}{2}\right\} dx$$



In case A: $\Phi = 1 - \alpha$*; in case B:* $\gamma = 1 - 2\alpha$*. The resulting value of* δ *will be the same*

$$\Phi(\delta_\gamma) = \frac{1+\gamma}{2}$$

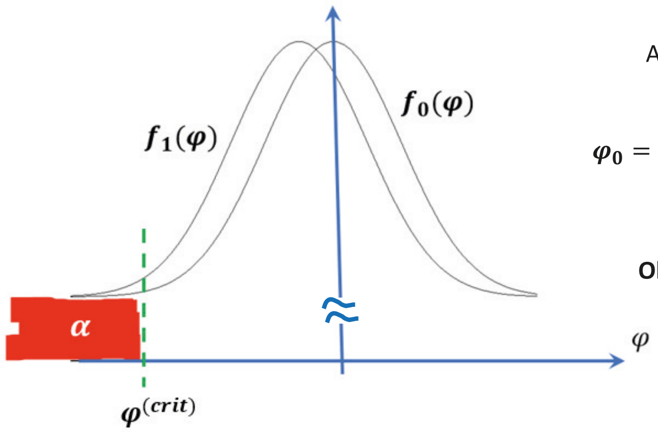


$$\alpha \to \begin{Bmatrix} \Phi = 1-\alpha \\ \gamma = 1-2\alpha \end{Bmatrix} \to \textit{any of Laplace function databases}$$

$$\to \varphi^{(crit)} = \delta_\gamma \to \textit{calculate } \varphi_0 = \frac{\bar{x} - a_0}{\sigma_0 / \sqrt{n}} \to$$

$$\textit{if } \varphi_0 < \varphi^{(crit)}, H_0 \textit{ is accepted; otherwise } H_0 \textit{ is rejected}$$

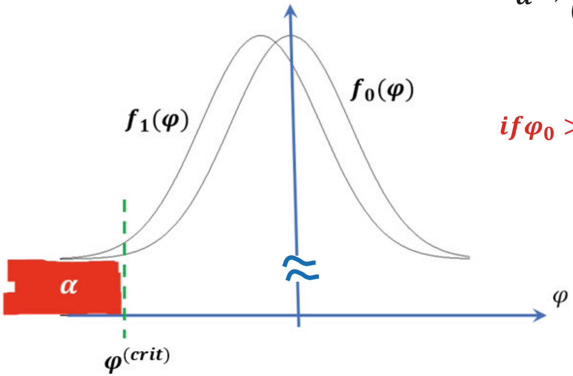
Hypothesis H$_0$: $a = a_0$
Hypothesis H$_1$: $a = a_1 < a_0$



A criterion: $\boldsymbol{\varphi = \frac{\bar{x}-a}{\sigma_0/\sqrt{n}} \epsilon N(0,1)}$

$$\varphi_0 = \frac{\bar{x}-a_0}{\sigma_0/\sqrt{n}}; \ \varphi_1 = \frac{\bar{x}-a_1}{\sigma_0/\sqrt{n}} = \frac{\bar{x}-(a_0-\Delta a)}{\sigma_0/\sqrt{n}} = \varphi_0 + \frac{\Delta a}{\sigma_0/\sqrt{n}}$$

Obviously, a critical domain is on the l.h.s. of the real axis of the criterion



$$\alpha \rightarrow \begin{Bmatrix} \Phi = 1 - \alpha \\ \gamma = 1 - 2\alpha \end{Bmatrix} \rightarrow any\ of\ Laplace\ function\ databases$$

$$\rightarrow \varphi^{(crit)} = -\,\delta_\gamma \rightarrow calculate\ \varphi_0 = \frac{\bar{x} - a_0}{\sigma_0 / \sqrt{n}} \rightarrow$$

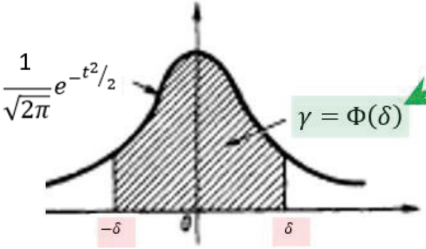
$$if\,\varphi_0 > \varphi^{(crit)}, H_0\ \ is\ accepted;\ otherwise\ \ H_0\ \ \ is\ rejected$$

Worked example

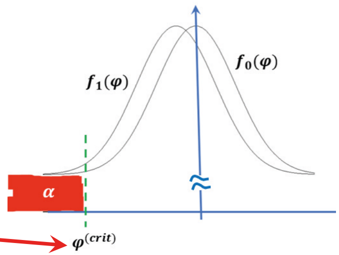
As a result of 9 measurements the average concentration of some substance in the solution was determined to be 48 mmol/L. The variance of the methodology is known to be σ_0^2= 9 (mmol/L)^2. Whether it is possible to accept 50 mmol/L as a true concentration at the level of significance α=0.05? How about 49 mmol/L? In the latter case determine the force of the criterion

A hint: in both cases formulate the alternative as $a_1 = \overline{x}$

$$\Phi(\delta) = \frac{2}{\sqrt{2\pi}} \int_0^{\delta} e^{-t^2/2}\, dt = \gamma = w(|t| < \delta)$$



δ	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0238	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9189	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	8488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	99915	99917	99920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9938	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	—	—	—	—	—	—	—	—	—
5,0	0,99999994	—	—	—	—	—	—	—	—	—

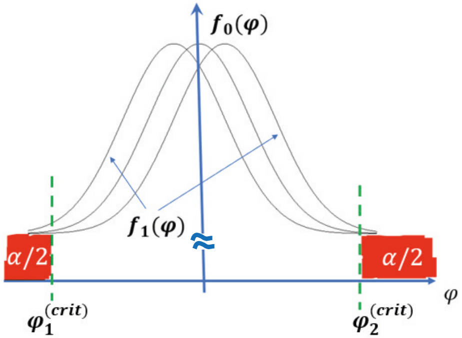


$$\gamma = 1 - 2\alpha = 0.90 \quad \Rightarrow \quad \delta_\gamma = 1.65, \quad -\delta_\gamma = -1.65$$

$$\varphi_0 = \frac{3(48-50)}{3} = -2\left(\frac{mmol}{L}\right) < -1.65 \quad \Rightarrow \quad H_0 \;\; in\ the\ 1^{st}\ case\ is\ rejected$$

$$\varphi_0 = \frac{3(48-49)}{3} = -1\left(\frac{mmol}{L}\right) > -1.65 \quad \Rightarrow \quad H_0 \;\; in\ the\ 2^{nd}\ case\ is\ accepted$$

$$\beta = \frac{1}{2} + \frac{\Phi\left(\left|\varphi^{(crit)}\right| - 1\right)}{2} = 0.5 + 0.24215 = 0.74215 \quad \Rightarrow \quad \mu = 1 - \beta \cong 0.26$$

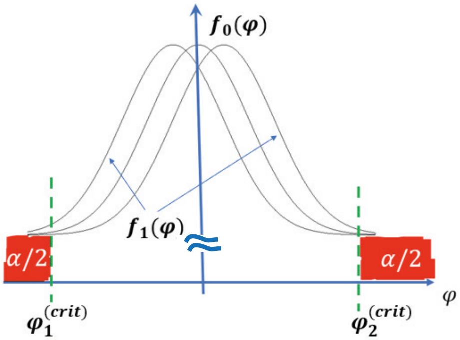


Hypothesis H_0: $a = a_0$
Hypothesis H_1: $a \neq a_0$

A criterion: $\boldsymbol{\varphi = \frac{\bar{x}-a}{\sigma_0/\sqrt{n}}}$

$$\varphi_0 = \frac{\bar{x} - a_0}{\sigma_0/\sqrt{n}}; \quad \varphi_1 = \frac{\bar{x} - a_1}{\sigma_0/\sqrt{n}} = \frac{\bar{x} - (a_0 \pm \Delta a)}{\sigma_0/\sqrt{n}} = \varphi_0 \pm \frac{\Delta a}{\sigma_0/\sqrt{n}}$$

It would make appearance as a rational thing to set a critical domain on the both wings of the Laplace function

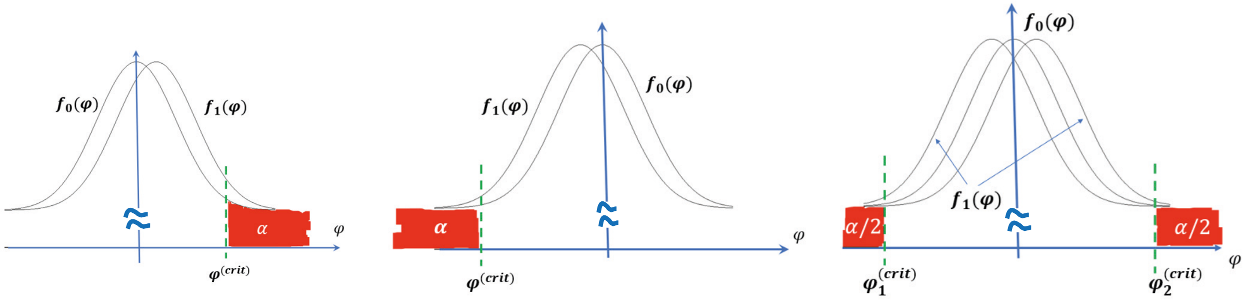


$$\alpha/_2 \rightarrow \begin{Bmatrix} \Phi = 1 - \alpha/_2 \\ \gamma = 1 - \alpha \end{Bmatrix} \rightarrow Laplace\ function\ database \rightarrow \varphi_2{}^{(crit)} = \delta_\gamma$$

$$\alpha/_2 \rightarrow \begin{Bmatrix} \Phi = \dfrac{1 - \alpha}{2} \\ \gamma = 1 - \alpha \end{Bmatrix} \rightarrow Laplace\ function\ database \rightarrow \varphi_1{}^{(crit)} = -\delta_\gamma$$

$$if\ \varphi_0 \epsilon \left[\varphi_1{}^{(crit)}, \varphi_2{}^{(crit)}\right], H_0\ \ is\ accepted; otherwise\ \ H_0\ \ \ is\ rejected$$

Hypotheses about the average with unknown variance



$$A\ sample\ \{x_i\}, 1 \le i \le n \ \Rightarrow\ \bar{x} == \frac{1}{n}\sum_{i=1}^{n} x_i\,;\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

A criterion: $\varphi = \dfrac{\bar{x}-a}{S/\sqrt{n}} = t_{n-1}$

A flow-chart for accepting an inference about hypothesis is basically the same as in the case of the known-variance hypotheses with the only difference in that the boundary of the critical domain is gotten from the database on the Student distribution

Worked example

Kinetic study of a certain reaction provided 20 observations giving the average time for it from the moment of mixing up components up to attaining chemical equilibrium $\bar{\tau}$ =77 min, the standard sample deviation $S_n^2 = 4$min^2. Can one reckon on the significance level 0.01 that $a_\tau = 80$ min is the reaction time in the given conditions

Hint: set an alternative hypothesis as the reaction time $\neq a_\tau$

$$\varphi = \frac{\sqrt{n}(\bar{\tau} - a_\tau)}{S_n} \qquad \begin{cases} H_0: & a_0 = 80 \\ H_1: & a_1 \neq a_0 \end{cases}$$

$$\begin{cases} \gamma = \alpha = 0.99 \\ k = n - 1 = 19 \end{cases} \Rightarrow \text{Student database} \Rightarrow \delta_\gamma = \pm 2.861$$

$$\varphi_0 = \frac{\sqrt{20}(77 - 80)}{2} = -6.708 < -2.861$$

$\Downarrow$

H_0 **will be rejected**

Hypotheses about the variance

Does the true variance exceed the permissible maximum value?

Hypothesis H_0: $\sigma^2 = \sigma_0^2$

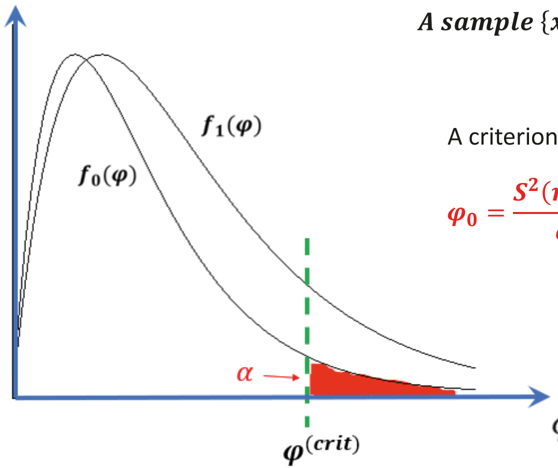
Hypothesis H_1: $\sigma^2 = \sigma_1^2 > \sigma_0^2$

$$A\ sample\ \{x_i\}, 1 \le i \le n \ \Rightarrow\ \bar{x} == \frac{1}{n}\sum_{i=1}^{n} x_i\,;\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

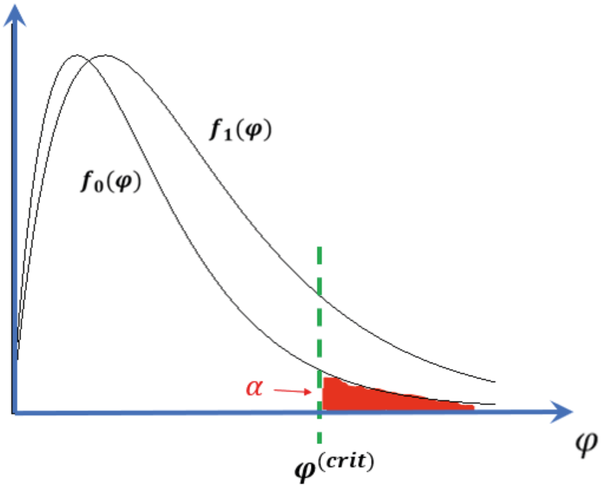
A criterion: $\varphi = \dfrac{S^2(n-1)}{\sigma^2} = \chi^2_{n-1}$

$$\varphi_0 = \frac{S^2(n-1)}{\sigma_0^2}\,;\ \varphi_1 = \frac{S^2(n-1)}{\sigma_1^2} = \frac{S^2(n-1)}{\sigma_1^2/\sigma_0^2}\cdot\frac{1}{\sigma_0^2} = \varphi_0\cdot\boxed{\frac{\sigma_0^2}{\sigma_1^2}}$$

< 1



R.h.s. critical domain

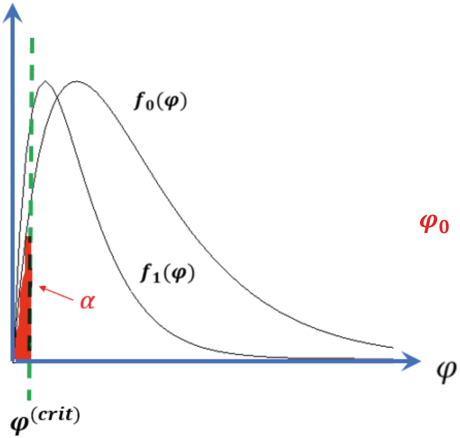


$$k = n - 1;\ \alpha \rightarrow \gamma = 1 - \alpha \rightarrow percent\ points\ of\ the\ Pearson\ distribution \rightarrow \varphi^{(crit)} = \delta_\gamma$$

$$if\ \varphi_0 \leq \varphi^{(crit)}, \qquad H_0\ is\ accepted;\ otherwise\ rejected$$

Another way of articulating alternative hypothesis

Hypothesis H_0: $\sigma^2 = \sigma_0^2$
Hypothesis H_1: $\sigma^2 = \sigma_1^2 < \sigma_0^2$

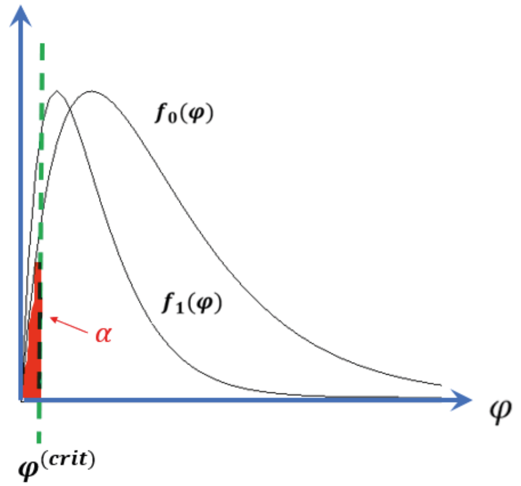


A criterion: $\varphi = \frac{S^2(n-1)}{\sigma^2} = \chi^2_{n-1}$

$$\varphi_0 = \frac{S^2(n-1)}{\sigma_0^2};\ \varphi_1 = \frac{S^2(n-1)}{\sigma_1^2} = \frac{S^2(n-1)}{\sigma_1^2 / \sigma_0^2} \cdot \frac{1}{\sigma_0^2} = \varphi_0 \cdot \boxed{\frac{\sigma_0^2}{\sigma_1^2}}$$

> 1

L.h.s. critical domain

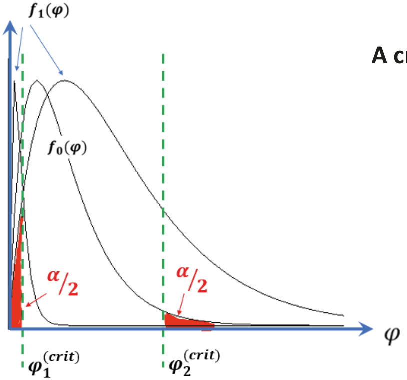


$$k = n - 1;\ \alpha \to \gamma = \alpha \to percent\ points\ of\ the\ Pearson\ distribution \to \varphi^{(crit)} = \delta_\gamma$$

$$if\ \varphi_0 \geq \varphi^{(crit)}, \qquad H_0\ is\ accepted;\ otherwise\ rejected$$

The third way of articulating alternative hypothesis

Hypothesis H_0: $\sigma^2 = \sigma_0^2$
Hypothesis H_1: $\sigma^2 = \sigma_1^2 \neq \sigma_0^2$

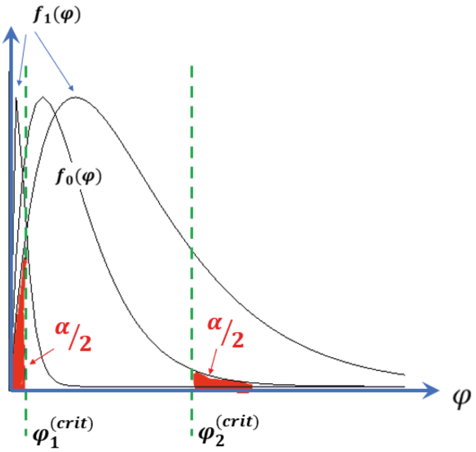


A criterion: $\varphi = \dfrac{S^2(n-1)}{\sigma^2} = \chi^2_{n-1}$

> 1
< 1

$$\varphi_0 = \frac{S^2(n-1)}{\sigma_0^2};\ \varphi_1 = \frac{S^2(n-1)}{\sigma_1^2} = \frac{S^2(n-1)}{\sigma_1^2/\sigma_0^2} \cdot \frac{1}{\sigma_0^2} = \varphi_0 \cdot \boxed{\frac{\sigma_0^2}{\sigma_1^2}}$$

Double-sided critical domain



$$k = n - 1;\ \alpha \to \gamma = {}^{\alpha}/_{2} \to percent\ points\ of\ the\ Pearson\ distribution \to \varphi_1^{(crit)} = \delta_\gamma$$

$$k = n - 1;\ \alpha \to \gamma = 1 - {}^{\alpha}/_{2} \to percent\ points\ of\ the\ Pearson\ distribution \to \varphi_2^{(crit)} = \delta_\gamma$$

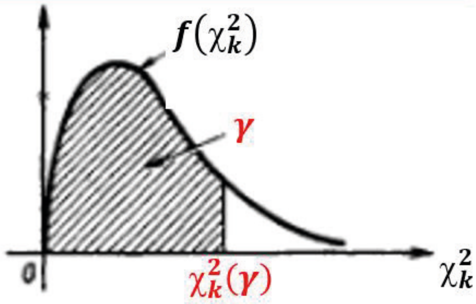
$$if\ \varphi_0 \epsilon \left[\varphi_1^{(crit)}, \varphi_2^{(crit)}\right], \qquad H_0\ is\ accepted; otherwise\ rejected$$

Worked example

The precision of a newly constructed detector of a certain substance was evaluated from the data on variance, which must not exceed 0.15 mmol/L. A statistical sample comprises measurements of 25 probes from one solution. The sample average deviation equals $\boldsymbol{S^2 = 0.25\,(mmol/L)^2}$. Whether it is possible to accept the precision of the setup relevant on the significance level $\boldsymbol{\alpha = 0.05}$?

Hint: set an alternative the statement that the setup precision is worse

The Pearson χ^2 distribution



$$w(\chi_k^2 < \chi_k^2(\gamma)) = \gamma$$

k \ γ	0,50	0,70	0,80	0,90	0,95	0,98	0,99	0,999
1	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	9,34	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,5
14	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	15,34	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

$$H_0: \quad \sigma^2 = \sigma_0^2$$
$$H_1: \quad \sigma^2 > \sigma_0^2$$

$$\varphi_0 = \frac{S^2(n-1)}{\sigma_0^2}$$

$$\varphi_1 = \frac{S^2(n-1)}{\sigma_0^2} \cdot \left(\frac{\sigma_0^2}{\sigma_1^2}\right) \quad \Rightarrow \quad critical\ domain\ is\ r.h.s.$$

$$\gamma = 1 - \alpha; \;\; k = n - 1 = 24 \quad \Rightarrow \quad \varphi^{(crit)} = 36.4$$

$$\varphi_0 = \frac{0.25 \cdot 24}{0.15} = 40 > 36.4 \quad \Rightarrow \quad the\ dasic\ hypothesis\ should\ be\ rejected$$

Hypothesis about equality of averages in two samples with different, but known variances

The situation of juxtaposing measurements of the same physico-chemical property with two methodologies

One sample $\{\boldsymbol{x_i}\}$ size $\boldsymbol{n_x}$ with variance $\boldsymbol{\sigma_x^2}$; another sample $\{\boldsymbol{y_i}\}$ size $\boldsymbol{n_y}$ with variance $\boldsymbol{\sigma_y^2}$

Do gotten averages from the two samples estimate the same expectation?

Variables $\dfrac{\bar{x}-a_x}{\sigma_x/\sqrt{n_x}}$ and $\dfrac{\bar{y}-a_y}{\sigma_y/\sqrt{n_y}} \in N(0,1)$ Hence, $\bar{x}-\bar{y} \in N\left(a_x-a_y;\ \dfrac{\sigma_x^2}{n_x}+\dfrac{\sigma_y^2}{n_y}\right)$

Criterion: $\varphi = \dfrac{\bar{x}-\bar{y}-(a_x-a_y)}{\sqrt{\dfrac{\sigma_x^2}{n_x}+\dfrac{\sigma_y^2}{n_y}}} \in N(0,1)$

$H_0:$ *the difference between* $\bar{x}$ *and* $\bar{y}$ *is accedental* $\Rightarrow M(\bar{x}-\bar{y})=0 \Rightarrow a_x = a_y$

$H_1:$ $a_x \neq a_y$

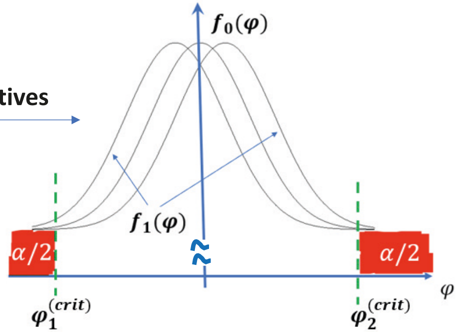
$$\varphi = \frac{\overline{x} - \overline{y} - (a_x - a_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \epsilon\, N(0,1)$$

H_0 : *the difference between* $\overline{x}$ *and* $\overline{y}$ *is accedental* $\Rightarrow M(\overline{x} - \overline{y}) = 0 \Rightarrow a_x = a_y$

H_1 : $a_x \neq a_y$

$$\varphi_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Positioning of alternatives



$\alpha \rightarrow \gamma = 1 - \alpha \rightarrow$
$\rightarrow$ *percent points of the Normal distribution*
$\rightarrow \; \rightarrow \varphi_1^{(crit)} = -\delta_\gamma; \;\; \varphi_2^{(crit)} = \delta_\gamma$

If $\varphi_0 \epsilon \left[\varphi_1^{(crit)}, \varphi_2^{(crit)}\right] \Rightarrow H_0$ *accepted*

Worked example

A statistical sample of 14 probes from the river was studied by a specific methodology and showed average concentration of detrimental contaminations $182\,mmol/L$, whilst another sample of 9 probes was studied by another methodology and showed the average concentration of the same substances $185\,mmol/L$. Variances of the employed methods were known and equaled: $\sigma_x^2 = 5\ (mmol/L)^2$ and $\sigma_x^2 = 7\ (mmol/L)^2$. Can one assert on the significance level 0.05 that both samples provide equal estimates of the same true values of the concentration?

A hint: an alternative consists in that the difference of the estimates is not a random case

$\boldsymbol{H_0}$: $\boldsymbol{a_x = a_y}$ $\quad \gamma = \mathbf{1} - 0.05 = 0.95 \quad \Rightarrow \quad \boldsymbol{Student\ database} \quad \Rightarrow \quad \boldsymbol{\delta_\gamma} = \pm 1.96$

$\boldsymbol{H_1}$: $\boldsymbol{a_x \neq a_y}$

$$\boldsymbol{\varphi_0} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{\mathbf{182 - 185}}{\sqrt{\frac{5}{14} + \frac{7}{9}}} = -2.82 < -1.96$$

Basic hypothesis should be rejected

Hypothesis about equality of averages in two samples with equal, but unknown variances

The situation of juxtaposing averages of the same physico-chemical property estimated from two statistical samples obtained by the same methodology with equal, yet unknown variances

One sample $\{x_i\}$ size n_x with variance S_x^2; another sample $\{y_i\}$ size n_y with variance S_y^2. $\sigma_x^2 = \sigma_y^2 = \sigma^2$

Do gotten averages from the two samples estimate the same expectation?

$$\bar{x} - \bar{y} \in N\left(a_x - a_y; \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right) \qquad \frac{\bar{x} - \bar{y} - (a_x - a_y)}{\sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}} \in N(0, 1) \qquad \frac{S_x^2(n_x - 1)}{\sigma^2} = \chi^2_{n_x - 1}; \frac{S_y^2(n_y - 1)}{\sigma^2} = \chi^2_{n_y - 1}$$

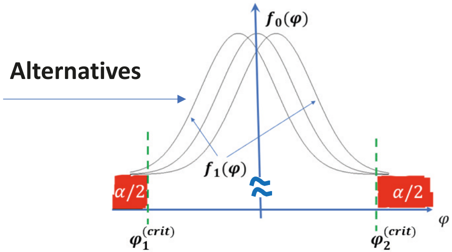
$$\varphi = \frac{\bar{x} - \bar{y} - (a_x - a_y)}{\cancel{\sigma}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \cdot \frac{\cancel{\sigma}}{\sqrt{\frac{1}{(n_x + n_y - 2)}\left(S_x^2(n_x - 1) + S_y^2(n_y - 1)\right)}} = t_{n_x + n_y - 2}$$

$$\varphi = \frac{\overline{x} - \overline{y} - (a_x - a_y)}{\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \cdot \frac{1}{\sqrt{\frac{1}{(n_x + n_y - 2)}\left(S_x^2(n_x - 1) + S_y^2(n_y - 1)\right)}} = t_{n_x+n_y-2}$$

$H_0:$ *the difference between* $\overline{x}$ *and* $\overline{y}$ *is accedental* $\Rightarrow M(\overline{x} - \overline{y}) = 0 \Rightarrow a_x = a_y$

$H_1:$ $a_x \neq a_y$

$$\varphi_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \cdot \frac{1}{\sqrt{\frac{\left(S_x^2(n_x - 1) + S_y^2(n_y - 1)\right)}{(n_x + n_y - 2)}}}$$



$k = n_x + n_y - 2; \alpha \rightarrow \gamma = 1 - \alpha \rightarrow$
percent points of the Student distribution
$\rightarrow \varphi_1^{(crit)} = -\delta_\gamma; \; \varphi_2^{(crit)} = \delta_\gamma$

$If\, \varphi_0 \epsilon \left[\varphi_1^{(crit)}, \varphi_2^{(crit)}\right] \Rightarrow H_0\, accepted$

Worked example

Two spots in the forest area were studied by a single analytical detector regarding the content of a particular contamination in the air. The sample from the spot X (number of probes $n_x = 9$) showed the following concentrations in the probes:

Number of probes	1	4	4
Registered concentration (mmol/L)	304	307	308

The other sample from spot Y (number of probes $n_y = 13$) gave the following data:

Number of probes	2	6	4	1
Registered concentration (mmol/L)	303	304	306	308

In the assumption of equal variances in both samples test on the significance level 0.1 a hypothesis H_0: $a_x = a_y$ with an alternative H_1: $a_x \neq a_y$

$$\gamma = 1 - \alpha$$
$$k = 9 + 13 - 2 = 20$$

$$\delta_\gamma = \pm 1.725$$

$$\overline{x} = \frac{304 \cdot 1 + 307 \cdot 4 + 308 \cdot 4}{9} = 307.11$$

$$\overline{y} = \frac{303 \cdot 2 + 304 \cdot 6 + 306 \cdot 4 + 308 \cdot 1}{13} = 304.77$$

$$S_x^2 = \frac{1}{8}\sum_1^9 (x_i - \overline{x})^2 = 2.378;$$

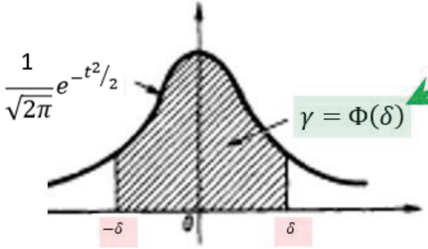
$$S_y^2 = \frac{1}{12}\sum_1^9 (y_i - \overline{y})^2 = 1.685;$$

$$\varphi_0 = \frac{307.11 - 304.77}{\sqrt{\frac{2.378 \cdot 8 + 1.685 \cdot 12}{20}\left(\frac{1}{9} + \frac{1}{13}\right)}} = 3.852\ \ (3.696) > \delta_\gamma \quad \Rightarrow \quad H_0 \qquad \textit{should be rejected}$$

Databases of statistical distributions

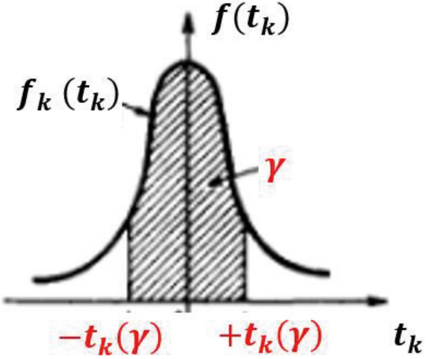
Percent points of the standard normal distribution

$$\Phi(\delta) = \frac{2}{\sqrt{2\pi}} \int_0^{\delta} e^{-t^2/2}\, dt = \gamma = w(|t| < \delta)$$



δ	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0238	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9189	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	8488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	99915	99917	99920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9938	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	—	—	—	—	—	—	—	—	—
5,0	0,99999994	—	—	—	—	—	—	—	—	—

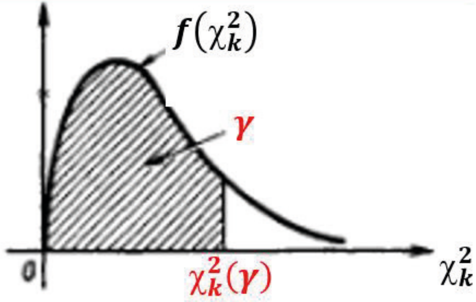
The Student distribution



$$w(|t_k| < t_k(\gamma)) = \gamma$$

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

The Pearson χ^2 distribution

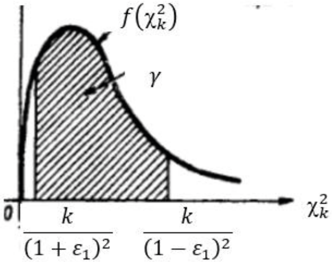


$$w(\chi_k^2 < \chi_k^2(\gamma)) = \gamma$$

k \ γ	0,50	0,70	0,80	0,90	0,95	0,98	0,99	0,999
1	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	9,34	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,5
14	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	15,34	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

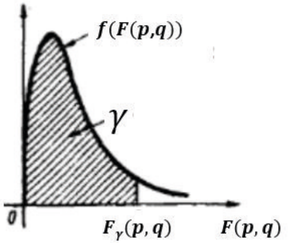
The values of ε_1 complying with a condition determined in accordance with the Pearson distribution:

$$w\left(\frac{k}{(1+\varepsilon_1)^2} \le \chi_k^2 \le \frac{k}{[max(0, 1-\varepsilon_1)]^2}\right)$$



k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,568	906	1,602	2,946	6,923				
2	367	473	0,678	1,125	2,086	3,400	5,857	8,500	
3	290	370	482	0,730	1,270	1,932	3,000	4,200	9,00
4	248	306	398	563	0,941	1,382	2,056	2,700	5,00
5	221	277	348	475	738	1,104	1,594	2,00	3,80
6	200	251	308	416	623	0,918	1,306	1,650	3,00
7	185	232	290	380	576	800	1,143	1,393	2,50
8	173	216	269	354	516	713	0,986	1,225	2,05
9	162	202	252	329	476	650	889	1,094	1,75
10	153	192	239	304	442	596	814	0,980	1,50
12	140	176	218	276	388	527	700	840	1,30
14	130	162	290	252	357	468	620	740	1,14
16	122	150	188	236	325	422	564	671	1,02
18	115	143	177	223	297	390	500	600	0,92
20	108	136	168	210	282	370	480	567	85
25	096	122	148	187	247	317	408	485	70
30	0,088	111	137	172	226	281	369	425	60
35	085	101	127	156	207	261	347	400	56
40	076	095	119	146	193	342	312	375	52
50	068	084	105	133	174	212	270	311	45
60	062	077	095	122	155	193	242	283	40
70	057	072	088	112	145	180	222	250	37
80	054	067	082	103	138	167	200	236	35
90	051	063	078	096	131	151	192	220	32
100	048	060	074	092	125	146	184	200	30

Fisher distribution for $\gamma = 0.95$



We set $\gamma, p = r + 1, q = n - r - 1$

$$w(F(p,q) < F_\gamma(p,q)) = \gamma$$

q	p = 1	2	3	4	5	6	8	12	24	∞
1	161,40	199,50	215.70	224,60	230,20	234,00	238,90	243,90	249,00	254,30
2	18,51	19.00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,33	3,51	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00