Determining the Confidence Interval for the Center and Width of a Structure in Fitting Measured Data by the Regression Line

1 G. Wimmer, 2 K. Karovič

1IMS and MI SAS, Bratislava, FNS UMB, Banská Bystrica, Slovakia and FNS MU Brno, Czech Republic,
2Institute of Measurement Science SAS, Bratislava, Slovakia
Email: wimmer@mat.savba.sk

Abstract. The two dimensional cross section of interest of a structure (e.g. a grating line) is schematically seen in Fig.2. Solid lines are structure’s edges. The solid bold arrow at the horizontal axis is the structure’s width w for given value F. Here the proper confidence intervals for width and center of the structure are derived.

Keywords: Confidence Interval, Measurement Uncertainty

1. Introduction

In metrology in case of assumed linear dependence between two quantities $x$, $y$ (e.g. length and electric signal, respectively) the functional dependence $y = f(x)$ is fitted using the regression line $y = a + bx$. It is assumed that measurement of the quantity $y$ in (exact) point $x$ is normally distributed, and the measurements are independent with equal standard deviations. The estimators of the regression coefficients $\hat{a}$, $\hat{b}$ and their standard deviations $s_a$, $s_b$ are determined from pairs of measured values $\{x_i, y_i\}_{i=1}^n$ using standard procedures (see e.g. [3]).

A common estimator of the $x$ value for assigned level $F$ of the quantity $y$ (e.g. electric signal) is

$$\hat{x} = \frac{\hat{F} - \hat{a}}{\hat{b}} ,$$

where $\hat{F}$ is a proper estimator of $F$, independent of $\hat{a}$, $\hat{b}$. Let $\hat{f}$ be the estimate of $F$ and $s_F$ be the estimate of the standard deviation of $\hat{f}$. By using the Law of Propagation of Uncertainties (see [2]) we obtain the estimate of the standard deviation of the estimator $\hat{x}$ as

$$s_x = \frac{1}{b} \sqrt{\frac{s^2}{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \left(\sum_{i=1}^n x_i^2 - 2 \frac{\hat{f} - \hat{a}}{\hat{b}} \sum_{i=1}^n x_i + n \left(\frac{\hat{f} - \hat{a}}{\hat{b}}\right)^2\right) + s_F^2} ,$$

where

$$s^2 = \frac{1}{n-2} \left(\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}\right) - \hat{b} \left(\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}\right) .$$

(1)
In deriving the (1-\(\alpha\))-confidence interval for \(x\) difficulties arise caused by the generally nonsymmetrical distribution of \(\hat{x}\) and so this confidence interval cannot be determined in usual way as \((\hat{x} - ks_x, \hat{x} + ks_x)\).

The desired (1-\(\alpha\))-confidence interval for \(x\) can be determined by the below described procedure.

2. Subject and Methods

The (1-\(\alpha\))-confidence interval for \(x\)

It is obvious (see e.g. [3]) that for an arbitrary \(x\) the (1-\(\alpha\))-confidence interval for (nonrandom value) \(a + bx (= y_x)\) is \((\hat{y}_x, \hat{y}_x)\), where

\[
\hat{y}_x = \hat{a} + \hat{b} x - sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

\[
\hat{y}_x = \hat{a} + \hat{b} x + sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

with

\[
d_x = \frac{1}{n} \left( 1 + \frac{\left( nx - \sum_{i=1}^{n} x_i \right)^2}{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \right),
\]

and \(t_{n-2} \left( 1 - \frac{\alpha}{2} \right)\) is the \(\left( 1 - \frac{\alpha}{2} \right)\)-quantile of the Student t-distribution with \(n-2\) degrees of freedom, where \(s\) is given by (1), for illustration see Fig. 1. According to [1], pp. 509-512, in case of sufficiently steep edges of the structure and small value of \(s\) (what is assumed here), the (1-\(\alpha\))-confidence interval for \(x\), denoted by \((\hat{x}, \hat{\mu}_x)\), can be constructed by the approach illustrated in Fig. 1. Given the errorless (nonrandom) value \(F\), the following relations hold true for the boundaries \(\hat{x}\) and \(\hat{\mu}_x\) of the (1-\(\alpha\))-confidence interval for \(x\)

\[
\hat{F} = \hat{a} + \hat{b} x - sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

\[
\hat{F} = \hat{a} + \hat{b} x + sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right).
\]

Solving both preceding equations the bounds \(\hat{x}\) and \(\hat{\mu}_x\) are given by

\[
\hat{x} = \frac{-B}{2A} - \frac{\sqrt{B^2 - 4AC}}{2A}, \quad \hat{\mu}_x = \frac{-B}{2A} + \frac{\sqrt{B^2 - 4AC}}{2A}
\]

for

\[
A = b^2 - d_x^2 \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

\[
b = \frac{1}{n} \left( \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 \right),
\]

with

\[
d_x = \frac{1}{n} \left( 1 + \frac{\left( nx - \sum_{i=1}^{n} x_i \right)^2}{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \right),
\]

and \(t_{n-2} \left( 1 - \frac{\alpha}{2} \right)\) is the \(\left( 1 - \frac{\alpha}{2} \right)\)-quantile of the Student t-distribution with \(n-2\) degrees of freedom, where \(s\) is given by (1), for illustration see Fig. 1. According to [1], pp. 509-512, in case of sufficiently steep edges of the structure and small value of \(s\) (what is assumed here), the (1-\(\alpha\))-confidence interval for \(x\), denoted by \((\hat{x}, \hat{\mu}_x)\), can be constructed by the approach illustrated in Fig. 1. Given the errorless (nonrandom) value \(F\), the following relations hold true for the boundaries \(\hat{x}\) and \(\hat{\mu}_x\) of the (1-\(\alpha\))-confidence interval for \(x\)

\[
\hat{F} = \hat{a} + \hat{b} x - sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

\[
\hat{F} = \hat{a} + \hat{b} x + sd_x \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right).
\]

Solving both preceding equations the bounds \(\hat{x}\) and \(\hat{\mu}_x\) are given by

\[
\hat{x} = \frac{-B}{2A} - \frac{\sqrt{B^2 - 4AC}}{2A}, \quad \hat{\mu}_x = \frac{-B}{2A} + \frac{\sqrt{B^2 - 4AC}}{2A}
\]

for

\[
A = b^2 - d_x^2 \left( t_{n-2} \left( 1 - \frac{\alpha}{2} \right) \right),
\]

\[
b = \frac{1}{n} \left( \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 \right),
\]

with

\[
d_x = \frac{1}{n} \left( 1 + \frac{\left( nx - \sum_{i=1}^{n} x_i \right)^2}{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \right),
\]
Determining some parameters of the structure

In order to determine the structure’s center and width, for a given (errorless, nonrandom) $F$ in analyzed cross section, it is necessary to use the $x$ values from two confidence intervals $(l_x, u_x)$ and $(l_x, u_x)$ that correspond to the structure’s edges (borders), see Fig. 2.

Using Bonferroni’s inequality (see e.g. in [3]) for the structure’s center $x_c$

$$P\left\{ x_{S} \in \left[ \frac{(1)x_c + (2)x_c}{2}, \frac{(1)x_c + (2)x_c}{2} \right] \right\} \geq 1 - 2\alpha,$$

i.e.

$$x_{S_1} = \frac{(1)x_c + (2)x_c}{2}, \quad x_{S_2} = \frac{(1)x_c + (2)x_c}{2}$$

is at least $(1-2\alpha)$-confidence interval for the structure's center.
Similarly, for the structure’s width $w$, 
\[ P\{w \in \left( \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}, \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \right) \} \geq 1 - 2\alpha, \]
and from this, 
\[ \left( w_1 = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}, w_2 = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \right) \] 
is the at least (1-2\alpha)-confidence interval for the structure’s width $w$.

![Fig. 2. Determining the sizes of the structure](image)

3. Conclusion

The above achieved assertions are applicable to measurement of the geometry of two-dimensional structures (or cross sections of three-dimensional structures) in the following manner:

For a chosen (errorless) value (level) $F$ of quantity $y$ (e.g. electric signal of a length gauge) and $\alpha \in (0,1)$, the (1-$\alpha$)-confidence interval for the structure’s border (bound) is $(s, u)$ where $s, u$ are given in (2), the (1-2\alpha)-confidence interval for the structure’s center is given in (3), and the (1-2\alpha)-confidence interval for the structure’s width is given in (4).

Acknowledgment

Contribution was supported by Alexander von Humboldt foundation and grants APVV RPEU-0008-06, VEGA 1/0077/09, VEGA 2/7082/27 and MŠMT ČR LC06024.

References

