UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Design and Analysis of Experiments

04 - Statistical Intervals

Version 2.11

Felipe Campelo
http://www.cpdee.ufmg.br/~fcampelo

Graduate Program in Electrical Engineering

Belo Horizonte
March 2015

“*Science is an integral part of culture. It’s not this foreign thing, done by an arcane priesthood. It’s one of the glories of the human intellectual tradition.*”

Stephen Jay Gould
1941-2002
American paleontologist

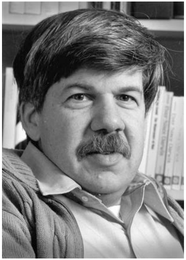


Image: Harvard University / AP

Statistical intervals are important in quantifying the uncertainty associated to a given estimate;

As an example, let's recap the coaxial cables example: *a coaxial cable manufacturing operation produces cables with a target resistance of* 50Ω *and a standard deviation of* 2Ω. *Assume that the resistance values can be well modeled by a normal distribution*.

Let us now suppose that a sample mean of $n = 25$ observations of resistance yields $\bar{x} = 48$. Given the sampling variability, it is very likely that this value is not exactly the true value of μ, but we are so far unable quantify how much uncertainty there is in this estimate.

Statistical intervals define regions that are likely to contain the true value of an estimated parameter.

More formally, it is generally possible to quantify the level of uncertainty associated with the estimation, thereby allowing the derivation of sound conclusions at predefined levels of certainty.

Three of the most common types of interval are:

- Confidence intervals;
- Tolerance intervals;
- Prediction intervals;

Confidence Intervals

Definition

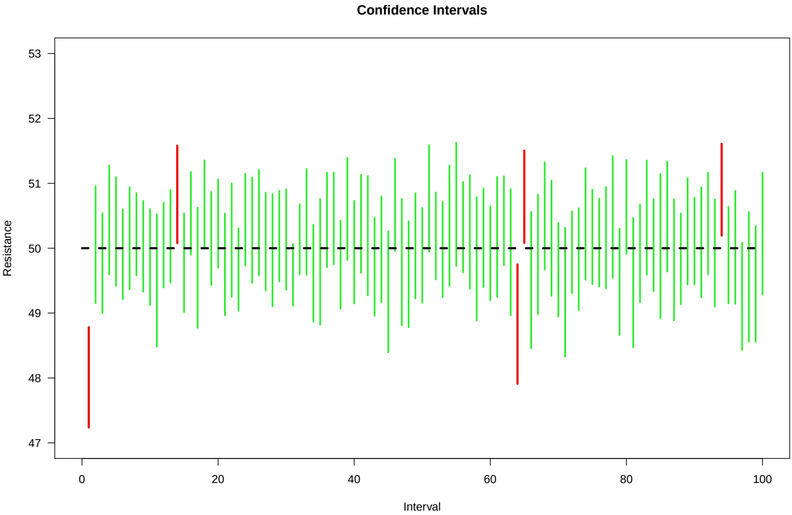
Confidence intervals quantify the degree of uncertainty associated with the estimation of population parameters such as the mean or the variance.

Can be defined as "*the interval that contains the true value of a given population parameter with a confidence level of* $100(1-\alpha)$";

Another useful definition is to think about confidence intervals in terms of confidence *in the method*: "The method used to derive the interval has a hit rate of 95%" - i.e., the interval generated has a 95% chance of 'capturing' the true population parameter."

Confidence Intervals

Example: 100 $CI_{.95}$ for a sample of 25 observations



For an interactive demonstration of the factors involved in the definition of a confidence interval, see `http://orcslab.cpdee.ufmg.br:3838/CI/`

The two-sided $CI_{(1-\alpha)}$ for the mean of a normal population with known variance σ^2 is given by:

$$\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $(1-\alpha)$ is the confidence level and z_x is the x-quantile of the standard normal distribution.

For the more usual case with an unknown variance,

$$\bar{x} + t_{\alpha/2}^{(n-1)}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\alpha/2}^{(n-1)}\frac{s}{\sqrt{n}}$$

where $t_x^{(n-1)}$ is the x-quantile of the t distribution with $n-1$ degrees of freedom.

A two-sided confidence interval on the variance of a normal variable can be easily calculated:

$$\frac{(n-1)s^2}{{\chi^2}_{1-\alpha/2}^{(n-1)}} \leq \sigma^2 \leq \frac{(n-1)s^2}{{\chi^2}_{\alpha/2}^{(n-1)}}$$

where ${\chi^2}_{x}^{(n-1)}$ represents the x-quantile of the χ^2 distribution with $n-1$ degrees of freedom. For the standard deviation one simply needs to take the squared root of the confidence limits.

Prediction Intervals

Definition

Prediction intervals quantify the uncertainty associated with forecasting the value of a future observation;

Essentially, one is interested in obtaining an interval within which he or she can declare that the next observation will fall with a given probability;

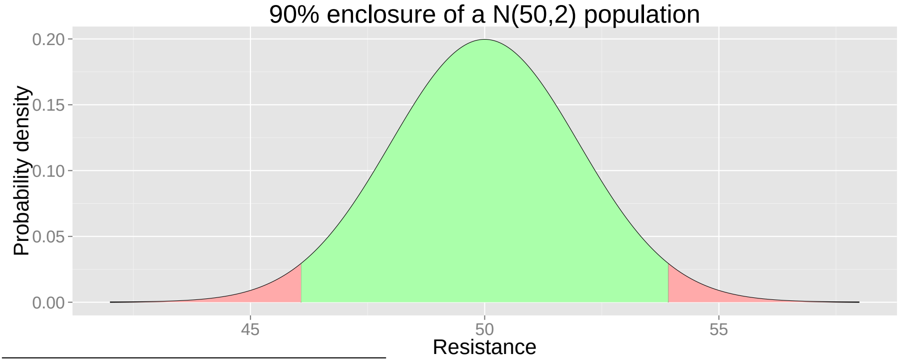
For a normal distribution, the tolerance interval for a single next observation (given an existing sample of size n) is:

$$\bar{x} + t_{\alpha/2}^{(n-1)} s\sqrt{1 + \frac{1}{n}} \leq x_{n+1} \leq \bar{x} + t_{1-\alpha/2}^{(n-1)} s\sqrt{1 + \frac{1}{n}}$$

Tolerance Intervals

Definition

"*A tolerance interval is an **enclosure** interval for a specified proportion of the sampled population, not its mean or standard deviation. For a specified confidence level, you may want to determine lower and upper bounds such that a given percent of the population is contained within them.*"[1].



[1] J.G. Ramírez: `https://git.io/v5ZFh`

The common practice in engineering of defining specification limits by adding $\pm 3\sigma$ to a given estimate of the mean arises from this definition - for a normal population $\approx 99.75\%$ of observations fall within $\mu \pm 3\sigma$.

However, as in most cases σ^2 is unknown, we have to use s^2 and compensate for the uncertainty in this estimation. The two-sided tolerance interval for a given population proportion γ is given as:[2]

$$\bar{x} \pm s\sqrt{\frac{(n-1)}{n}\frac{\left(n + z^2_{(\alpha/2)}\right)}{\chi^{2\,(n-1)}_{\gamma}}}$$

wherein γ is the proportion of the population to be enclosed, and $1-\alpha$ is the desired confidence level for the interval.

---

[2] NIST Engineering Statistics Handbook, `https://goo.gl/m6cxC6`

Statistical intervals quantify the uncertainty associated with different aspects of estimation;

Reporting intervals is always better than point estimates, as it provides the necessary information to quantify the location and uncertainty of your estimated values;

The correct interpretation is a little tricky (although not very difficult)[3], but it is essential in order to derive the correct conclusions based on the statistical interval of interest.

---

[3] See the table at the end of `https://git.io/v5ZFh`

Bibliography

Required reading

1. J.G. Ramírez, *Statistical Intervals: Confidence, Prediction, Enclosure*: `https://git.io/v5ZFh`
2. D.C. Montgomery and G.C. Runger, *Applied Statistics and Probability for Engineers*, Chapter 8. 3rd Ed., Wiley 2005.

Recommended reading

1. Simply Statistics (blog) - `http://simplystatistics.org`
2. R. Dawkins, *Climbing Mount Improbable*, W.W.Norton&Co.,1997.

About this material

Conditions of use and referencing

Please reference this work as:

Felipe Campelo (2015), *Lecture Notes on Design and Analysis of Experiments*. Online: https://github.com/fcampelo/Design-and-Analysis-of-Experiments
Version 2.11, Chapter 4; Creative Commons BY-NC-SA 4.0.

```
@Misc{Campelo2015-01,
  title={Lecture Notes on Design and Analysis of Experiments},
  author={Felipe Campelo},
  howPublished={\url{https://github.com/fcampelo/Design-and-Analysis-of-Experiments}},
  year={2015},
  note={Version 2.11, Chapter 4; Creative Commons BY-NC-SA 4.0.},
}
```