Probability and Statistics

Atar F. Babgei

Email: atar.babgei@its.ac.id

Statistical Inference
(**Estimation**)

Learning Outcome

By the end of this lecture you should be able to:

- Calculate a point estimate, interval estimate, margin of error, and level of confidence for single sample estimation problems
- Understand the distinction among confidence, prediction, and tolerance intervals

Point and Interval Estimation [1 / 2]

The sample mean $\bar{x}$ is an efficient **Point Estimate** of μ

An **interval estimate** of a population parameter θ is an interval of the form

$$\hat{\theta}_L < \mu < \hat{\theta}_U$$

Point and Interval Estimation [2 / 2]

Since different samples will generally yield different values of θ and, therefore, different values for $\hat{\theta}_L$ and $\hat{\theta}_U$,

$$P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha \text{, for } 0 < \alpha < 1$$

Definitions:

The interval computed from the selected is called

a $100(1 - \alpha)$ Confidence Interval

The fraction end points $\hat{\theta}_L$ and $\hat{\theta}_U$ are called

the lower and upper confidence limit

Single Sample: Estimating the Mean

Normal Distribution with Two sided bounds

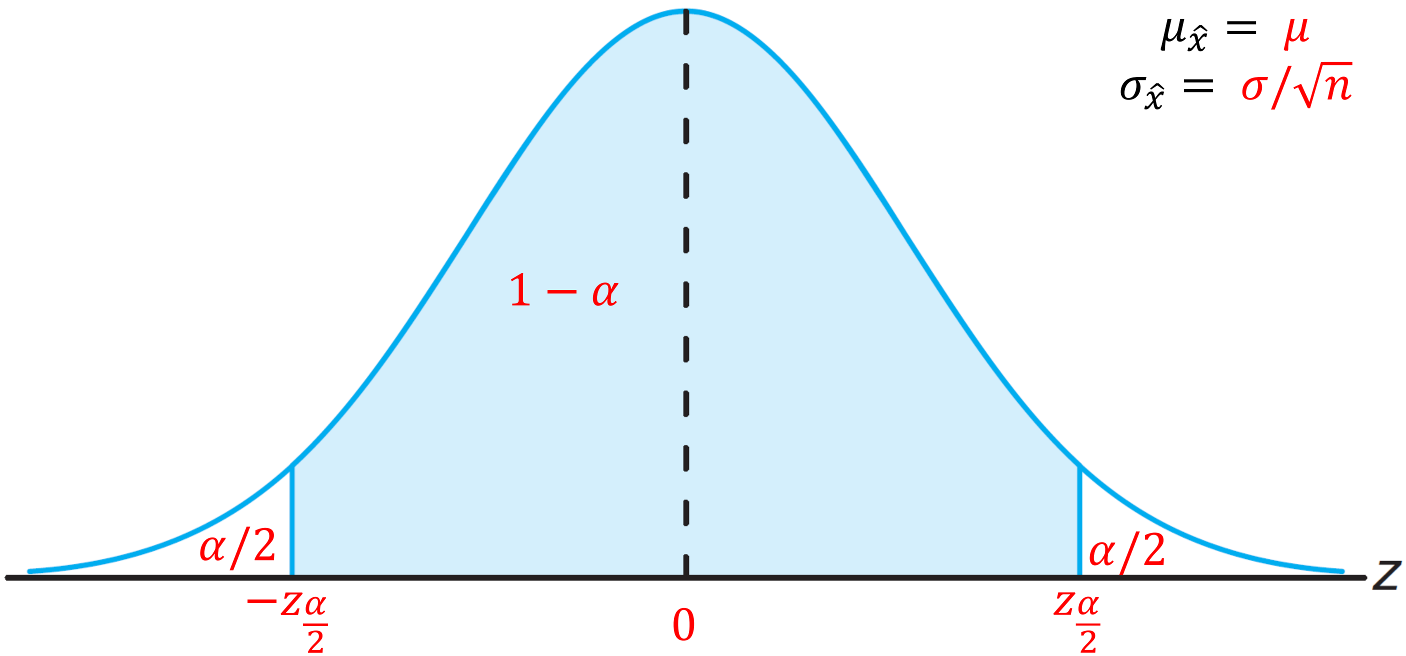

Figure 1. $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ [1]

Single Sample: Estimating the Mean

we can see from Figure that

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Hence,

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

By rearranging the equation, we obtain:

$$P\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Single Sample: Estimating the Mean

Confidence Interval:

$$\hat{\theta}_L < \mu < \hat{\theta}_U$$

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

Where $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = margin\ of\ error\ (e)$

Example 1.1

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per millilitre, assuming the population standard deviation is 0.3 gram per milliliter.

a) Find the 95% and 99% confidence intervals for the mean zinc concentration in the river.
b) How large a sample is required if we want to be 95% confident that our estimate of μ is of by less than 0.05?

Solution 1.1:

a) The point estimate of μ is $\bar{x} = 2.6$

$\sigma = 0.3$ $\quad\quad n = 36$

The 95% CI:

$\alpha = 0.05,\quad z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$$2.6 - (1.96)\frac{0.3}{\sqrt{36}} < \mu < 2.6 - (1.96)\frac{0.3}{\sqrt{36}}$$

Which reduces to:

$$2.5 < \mu < 2.7$$

Solution 1.1:

The 99% CI:

$$\alpha = 0.01, \qquad z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$$

$$2.6 - (2.575)\frac{0.3}{\sqrt{36}} < \mu < 2.6 - (2.575)\frac{0.3}{\sqrt{36}}$$

Which reduces to:

$$2.47 < \mu < 2.73$$

b) $n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{e}\right)^2 = \left(\frac{(1.96)(0.3)}{(0.05)}\right)^2 = 138.3$

One-Sided Confidence Bounds

What if there is one-sided bound only?

i.e. tensile strength

If $\bar{X}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1 - \alpha)\%$ confidence bounds for μ are given by

Upper one-sided bound: $\bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$

Lower one-sided bound: $\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}$

The Case of σ Unknown

When we only have less than 30 samples and the σ is unknown. The procedure is the same as that with σ known except that **σ** is replaced by **S** and the standard normal distribution is replaced by the **t-distribution.**

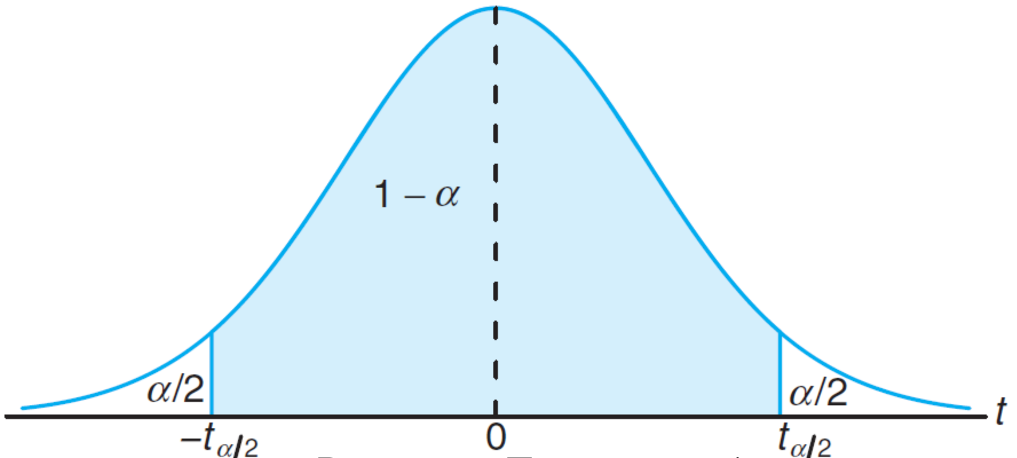

Figure 2. $P\left(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$

The Case of σ Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a 100 (1−α)% confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the t-value with $\vartheta = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right

Example 1.2

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters.

Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Solution 1.2:

$\bar{x} = 10$ $\qquad\qquad$ $s = 0.283$

$\alpha = 0.05,$ $\qquad$ $t_{\frac{\alpha}{2}} = t_{0.025} = 2.447$ for $\vartheta = 6$ DoF

The 95% CI is

$$10 - (2.447)\frac{0.283}{\sqrt{7}} < \mu < 10 - (2.447)\frac{0.283}{\sqrt{7}}$$

Which reduces to:

$$9.74 < \mu < 10.26$$

Prediction Interval

Sometimes, other than the population mean, we may also be interested in predicting the possible value of a future observation.

The development of a prediction interval is best illustrated by beginning with a normal random variable x_0 - $\bar{x}$, where x_0 is the new observation and $\bar{x}$ comes from the sample. Since x_0 and $\bar{x}$ are independent, we know that

$$Z = \frac{x_0 - \bar{x}}{\sqrt{\sigma^2 + \sigma^2/n}} = \frac{x_0 - \bar{x}}{\sigma\sqrt{1 + 1/n}}$$

Hence, the probability statement

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

The prediction interval is formalised as follows

$$\bar{x} - z_{\frac{\alpha}{2}}\,\sigma\sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\frac{\alpha}{2}}\sigma\sqrt{1 + 1/n}$$

Prediction Interval (unknown μ *and* σ)

For a normal distribution of measurements with unknown mean μ and variance σ^2, a 100 (1−α)% **prediction interval** of a future observation x_0 is

$$\bar{x} - t_{\alpha/2}\, s\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + t_{\alpha/2} s\sqrt{1+\frac{1}{n}}$$

where $t_{\alpha/2}$ is the t-value with $\vartheta = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

One-sided prediction interval can also be constructed

Lower bound: $\bar{x} - t_\alpha s\sqrt{1+\frac{1}{n}}$ Upper bound: $\bar{x} + t_\alpha s\sqrt{1+\frac{1}{n}}$

Example 1.3

Due to the decrease in interest rates, the First Citizens Bank received a lot of mortgage applications. A recent sample of 50 mortgage loans resulted in an average

loan amount of $257,300. Assume a population standard deviation of $25,000. For the next customer who fills out a mortgage application, find a 95% prediction interval for the loan amount

Solution 1.3

$\bar{x} = \$257{,}300$ $\sigma = \$25{,}000$ $n = 50$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

The 95% Prediction Interval is

$$257{,}300 - (1.96)(25{,}000)\sqrt{1 + \frac{1}{50}} < x_0 < 257{,}300 - (1.96)\,(25{,}000)\sqrt{1 + \frac{1}{50}}$$

Which reduces to:

$$\$207{,}812.43 < \mu < \$306{,}787.57$$

Tolerance Limit

Scientist or engineer may be interested in gaining a notion about where an individual notion about where an individual observation might fall. We can use tolerance interval which can be obtained by:

$$\bar{x} \pm ks$$

where k is determined such that one can assert with 100(1 − γ)% confidence that the given limits contain at least the proportion 1 − α of the measurements.

Usually it is fixed at **95%**

Example 1.4

A meat inspector has randomly selected 30 packs of 95% lean beef. The sample resulted in a mean of 96.2% with a sample standard deviation of 0.8%.

Find a tolerance interval that gives two-sided 95% bounds on 90% of the distribution of packages of 95% lean beef. Assume the data came from an approximately normal distribution.

Solution 1.4

a) The sample mean and standard deviation of given data is

$\bar{x} = 96.2$ $\quad$ $\mathrm{s} = 0.8$ $\quad$ $\mathrm{n} = 30$

$$\bar{x} \pm ks = 96.2 \pm (2.14)(0.8)$$

References

- “Probability & Statistics for Engineers & Scientists”, by Ronald E. Walpole, Raymond Myers, Sharon Myers, Keying Ye
- “Introduction to the Practice of Statistics”, Sixth Edition, by David S.Moore, George P. McCabe, and Bruce A. Craig
- “Probability and Statistics for Engineering and The Sciences” by Devore, Jay L.