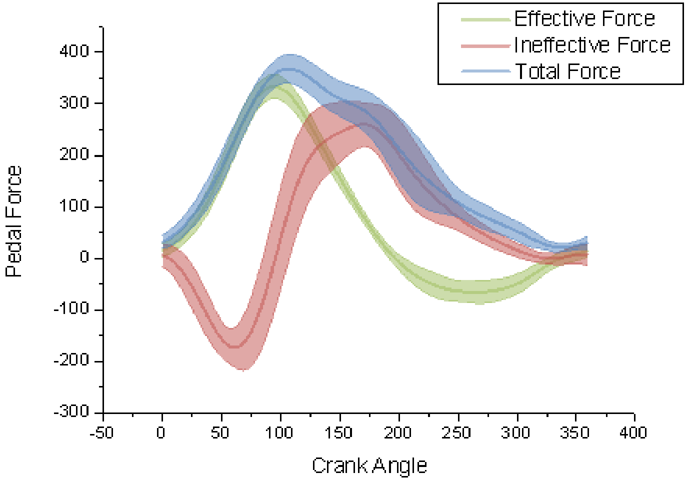
MATH 205: Statistical methods

Lab 7: Confidence intervals

A good prediction comes with a range



Confidence

- Assume that you have been using an AI to predict the stock price of Microsoft every day in the last few years
- The prediction comes as a range, e.g., [295, 305]
- The algorithm, on average, is correct 95 out of 100 days
- Then we say that a prediction from this AI has a confidence of 95%

Confidence interval

Suppose we are studying a distribution with with mean μ (unknown) and standard deviation $\sigma = 0.85$. A random sample of $n = 50$ specimens is selected with sample average $\bar{X}$. We know that $\bar{X}$ can be used to approximate μ.

Question: Find a general rule of c (that depends on $\bar{X}$) such that

$$P\left[\bar{X} - c \leq \mu \leq \bar{X} + c\right] = 0.95$$

$\rightarrow$ The interval $[\bar{X} - c, \bar{X} + c]$ is a confidence interval with confidence level 95%.

The Central Limit Theorem

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit when $n \to \infty$, the standardized version of $\bar{X}$ have the standard normal distribution

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right) = \mathbb{P}[Z \leq z] = \Phi(z)$$

Rule of Thumb:

If $n > 30$, the Central Limit Theorem can be used for computation.

Confidence interval

We know that

$$P\left[\bar{X} - c \leq \mu \leq \bar{X} + c\right] = 0.95$$

is equivalent to

$$P\left[\bar{X} - \mu - c \leq 0 \leq \bar{X} - \mu + c\right] = 0.95$$

is equivalent to

$$P\left[\frac{-c}{\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{c}{\sqrt{n}}\right] = 0.95$$

Confidence interval

- We have

$$P\left[-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right] = 0.95$$

- Rearranging the inequalities gave

$$P\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

- This means that if you use

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

as a range to estimate μ, then you are correct 95% of the time.

Normal distribution with know σ

- Using
$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$
as a range to estimate μ is correct 95% of the time.
- If after observing $X_1 = x_1$, $X_2 = x_2, \ldots$, $X_n = x_n$, we compute the observed sample mean $\bar{x}$. Then
$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$
is a 95% confidence interval of μ

z-critical value

NOTATION

z_α will denote the value on the measurement axis for which α of the area under the z curve lies to the right of z_α. (See Figure 4.19.)

For example, $z_{.10}$ captures upper-tail area .10 and $z_{.01}$ captures upper-tail area .01.

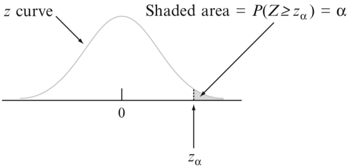


Figure 4.19 z_α **notation illustrated**

Since α of the area under the standard normal curve lies to the right of z_α, $1-\alpha$ of the area lies to the left of z_α. Thus z_α *is the* $100(1-\alpha)$*th percentile of the standard normal distribution*. By symmetry the area under the standard normal curve to the left of $-z_\alpha$ is also α. The z_α's are usually referred to as ***z* critical values**. Table 4.1 lists the most useful standard normal percentiles and z_α values.

$100(1-\alpha)\%$ confidence interval

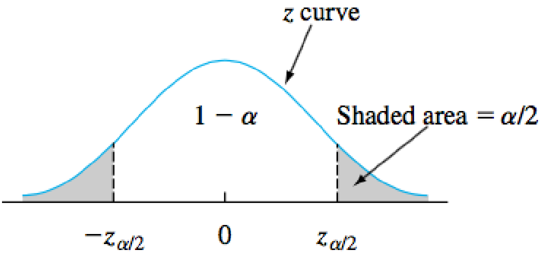


Figure 8.4 $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$

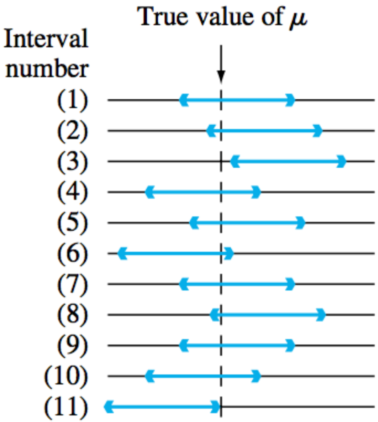
$100(1-\alpha)\%$ confidence interval

A **$100(1-\alpha)\%$ confidence interval** for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{8.5}$$

or, equivalently, by $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Interpreting confidence intervals



95% confidence interval: If we repeat the experiment many times, the interval contains μ about 95% of the time