Statistical Intervals

It is 2 boundary numbers that indicate if the true parameter is captured within.

Imagine this as a net designed to capture the truth.

The Bounds do not give a range of values that the true parameter may be.

How effective the methodology is at capturing the truth.

When you construct a statistical interval you specify a confidence.

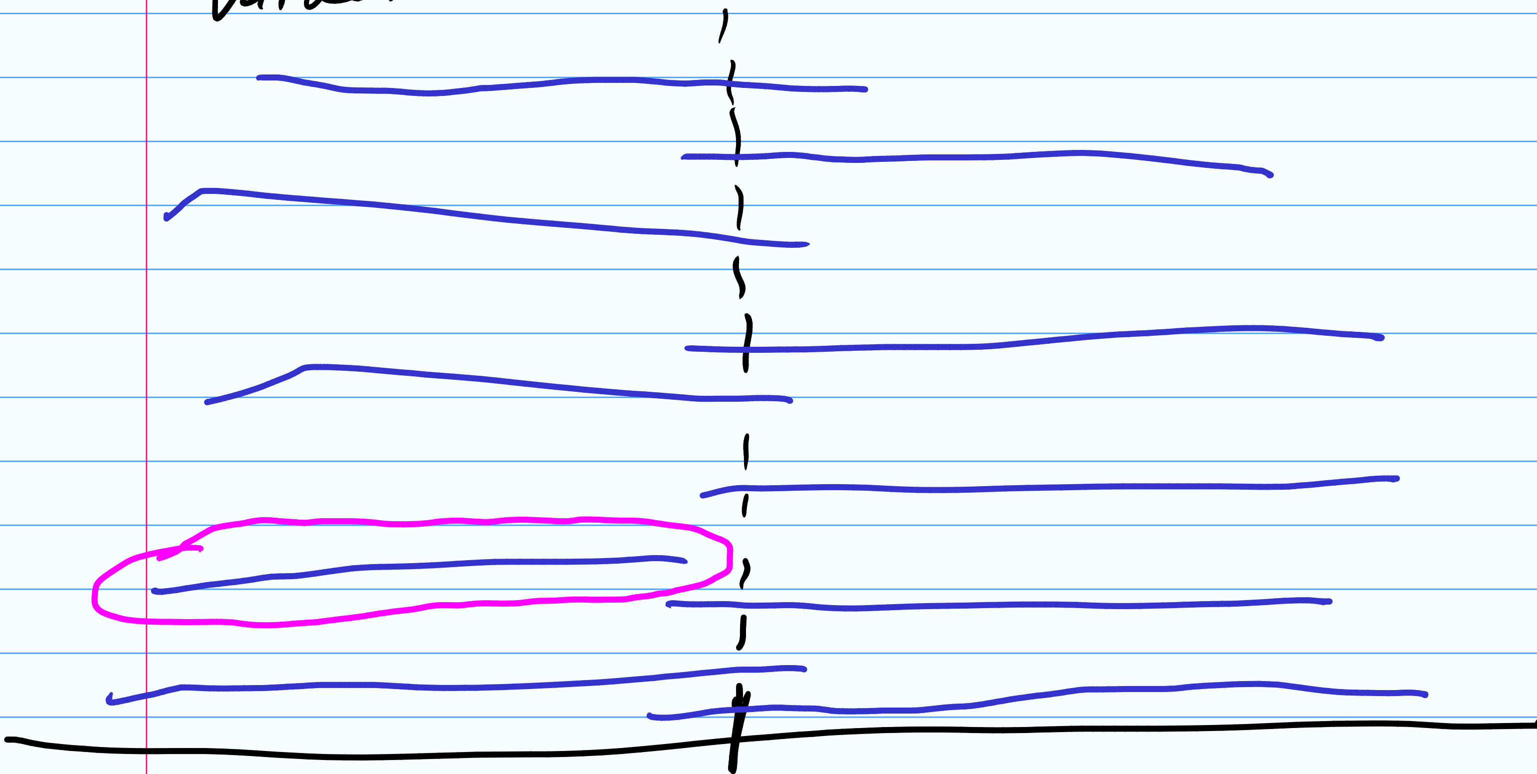
"Confidence" tells the probability that the interval will contain the truth.

90% CI

What does this mean?

If we repeat an experiment 1000, 900 of the intervals will capture the true parameter value.

μ ← true value



How do we construct a confidence Interval?

$(1-\alpha)100\%$ CI

α = The probability of being wrong.

$\alpha = 0.05$

$$P(L \le t.s. \le U) = .95$$

$$P(L' \le \mu \le U') = 0.95$$

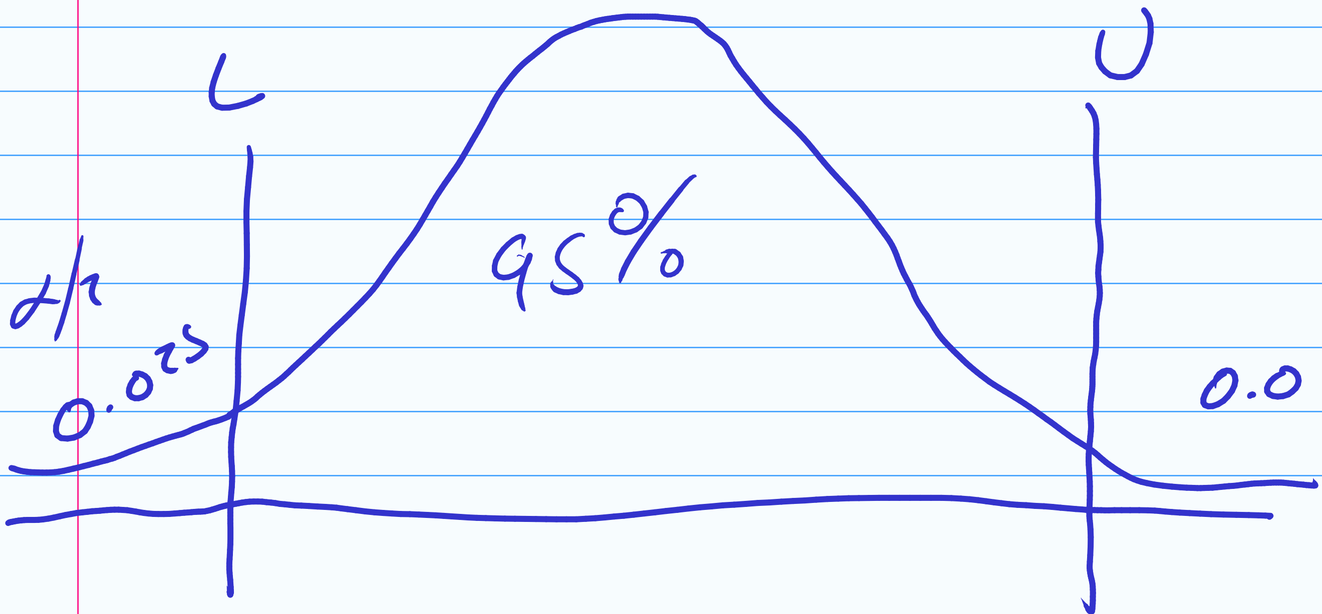
What is a test statistic?

$$t.s. = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

$$P(L \le t.s. \le U) = .95$$

$$P(ts \le U) - P(ts \le L) = 0.95$$

What is L and U?



$$P(t_S \le U) - P(t_S \le L) = 0.95$$

$$P(t_S \le L) = 0.025$$

$$P(t_S \le U) = 0.975$$

We can set U and L with a z-table or computer.

$$\int_{-\infty}^{U} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = 0.975$$

solve for U.

$$U = 1.96 \qquad L = -1.96$$

$$P(L \le t_S \le U) = .95$$

$$P(-1.96 \le t_S \le 1.96) = 0.95$$

$$P\left(-1.96 \le \frac{\bar{x} - \mu}{s/\sqrt{n}} \le 1.96\right) = 0.95$$

$$P(-1.96\, s/\sqrt{n} \le \bar{x} - \mu \le 1.96\, s/\sqrt{n}) = 0.95$$

$$P(-\bar{x} - 1.96\, s/\sqrt{n} \le -\mu \le -\bar{x} + 1.96\, s/\sqrt{n}) = 0.95$$

$$P(\bar{x} + 1.96\, s/\sqrt{n} \ge \mu \ge \bar{x} - 1.96\, s/\sqrt{n}) = 0.95$$

95% CI

$$\left(\bar{X} - 1.96\, s/\sqrt{n},\ \bar{X} + 1.96\, s/\sqrt{n}\right)$$

$(1-\alpha)100\%$ CI

$$\left(\bar{X} - z_{\alpha/2}\, s/\sqrt{n},\ \bar{X} + z_{\alpha/2}\, s/\sqrt{n}\right)$$

$$z_{\alpha/2} = \quad P(X \le z_{\alpha/2}) = 1 - \alpha/2$$

$$\bar{X} \pm z_{\alpha/2}\, s/\sqrt{n}$$

$$\bar{X} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$$

$$\bar{x} \pm t_{\alpha/2,\,DF}\, s/\sqrt{n}$$

small samples $\quad n < 30$

All intervals work for

r.s. that follow a normal.

$\bar{X} \overset{\circ}{\sim} N(\mu, s^2/n)$ by CLT

$\hat{\theta} \overset{\circ}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{nI(\theta)}}$$

If things follow a normal distribution.

Your generic formula is:

$$PE \pm CV \cdot SE$$

(point estimate) (critical value) (standard Error)

Z - scores

A z-score is a transformation of data that centered around the mean and scaled by the standard deviation.

Say X is a RV.

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}}$$

$$X \sim N(\mu, \sigma^2)$$

$$Z = \frac{X - \mu}{\sigma}$$

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad n \to \infty$$

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

σ - is unknown

$$Z \to t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$