Confidence intervals for linear regressions parameters:

A confidence interval (CI) at the $c = 100(1-\alpha)\%$ level refers to an interval around a stochastic quantity in which it will lie $c\%$ of times. A related quantity is the significance level which is denoted as α.

Confidence level c	Significance level α
99%	0.01
95%	0.05
90%	0.1

Hypothesis testing refers to the validation of a certain statement

statement called the alternate hypothesis (H_a), with respect to a given significance level.

Variable of interest $\longrightarrow$ Population mean of the linear regression parameters.

Hypothesis tests for the population mean

- Z-test (population σ is known)
- t-test (population σ is unknown)

Z-test

Let us consider an independent sample $(y_1, y_2, \cdots, y_n)$ collected from a population with a known population standard deviation

Sample mean $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

Sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

$$\rightarrow E[\bar{y}] = E\left[\frac{1}{n} \sum_{i=1}^{n} y_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[y_i] = \frac{1}{\cancel{n}} \times \cancel{n}\mu = \boxed{\mu}$$

$$\rightarrow Var[\bar{y}] = Var\left[\frac{1}{n} \sum_{i=1}^{n} y_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} Var[y_i] = \frac{1}{n^2} \times n\sigma^2 = \boxed{\frac{\sigma^3}{n}}$$

If we assume each y_i to be normally distributed, then:

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

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

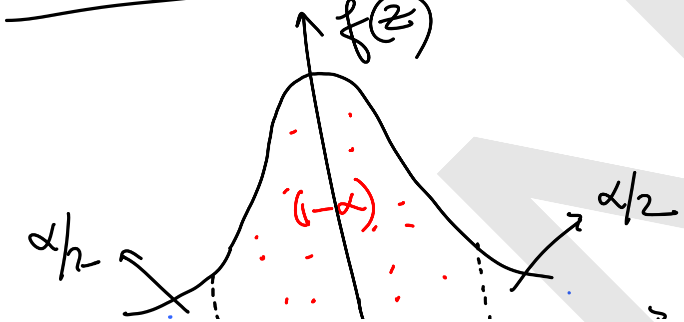
$$Z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \implies \boxed{Z \sim N(0, 1)}$$

Z-statistic

Hypothesis test → One-sided test

Hypothesis test → Two-sided test

Two-sided test

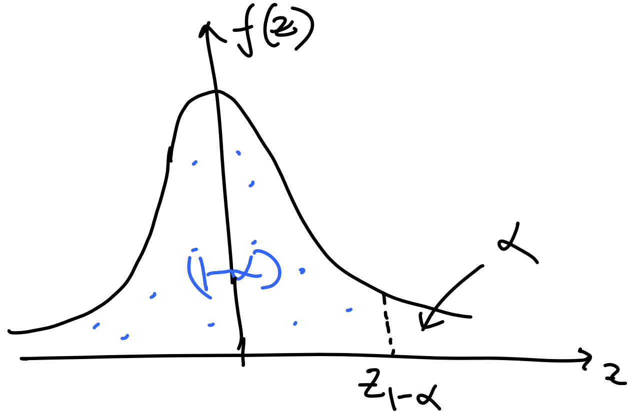


H_0 (null hypothesis) is rejected if

$$Z > Z_{1-\frac{\alpha}{2}} \quad \text{or} \quad Z < Z_{\frac{\alpha}{2}}$$

Otherwise the null hypothesis cannot be

One-sided test



H_0 (null hypothesis) is rejected if

$$z > z_{1-\alpha}$$

Otherwise the null hypothesis cannot be rejected.

Student's t test

Test for the population mean when the population standard deviation (σ) is unknown.

Random variable $T = \dfrac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

s: sample standard deviation

t_{n-1}: Student's t distribution with (n-1) degrees of freedom.

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

$$P\left(t_{\frac{\alpha}{2}} \leq \left(\frac{\bar{y}-\mu}{s/\sqrt{n}} \right) \leq t_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

$$P\left(t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \bar{y} - \mu \leq t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right) = 1-\alpha$$

$$P\left(-t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu - \bar{y} \leq -t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right) = 1-\alpha$$

$$\boxed{P\left(\bar{y} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{y} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right) = 1-\alpha}$$

Confidence intervals (CI) for linear regression parameters:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

To assign CI to $\hat{\beta}$, we need to know:

- $E(\hat{\beta})$
- $Var(\hat{\beta})$

We will assume that the errors (ϵ) in prediction are normally distributed.

$$E(\hat{\beta}) = E\left[(X^T X)^{-1} X^T Y\right]$$

$$Y = X\beta + \epsilon$$

actual parameters

error

$$E[\hat{\beta}] = E\left[(X^TX)^{-1}X^T(X\beta + \epsilon)\right]$$

$$= E\left[\underbrace{(X^TX)^{-1}X^TX}_{I}\beta + \underline{(X^TX)^{-1}X^T\epsilon}\right]$$

$$= E[I\beta] + (X^TX)^{-1}X^T \cancel{E(\epsilon)} \rightarrow 0$$

$$= E[\beta] + 0$$

$$E[\hat{\beta}] = E[\beta] = \beta$$

$\hat{\beta}$ is an <u>unbiased estimator</u> for β.

$$Var(\hat{\beta}) = ?$$

↳ To determine this, we will use the covariance matrix

$$Cov(\hat{\beta}) = E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T\right]$$

$$Y = X\beta + \epsilon$$

$$\hat{\beta} = (X^TX)^{-1}X^TY \implies X^TX\hat{\beta} = X^TY$$

$$X^TY = X^TX\beta + X^T\epsilon$$

$$X^TX\hat{\beta} = X^TX\beta + X^T\epsilon$$

$$\implies \hat{\beta} - \beta = (X^TX)^{-1}X^T\epsilon$$

$$Cov(\hat{\beta}) = E\left[(X^TX)^{-1}X^T\epsilon \cdot \left((X^TX)^{-1}X^T\epsilon\right)^T\right]$$

$$= (X^TX)^{-1}E[\epsilon\epsilon^T]$$

$$Var(\hat{\beta}) = diag\left[(X^TX)^{-1}E[\epsilon\epsilon^T]\right] = diag\left((X^TX)^{-1}\right)\cdot Var(\epsilon)$$

$$\sigma^2$$

$$\boxed{Var(\hat{\beta}) = \sigma^2 \, diag((X^T X)^{-1})}$$

$$\hat{\beta} \sim N(\beta, \sigma^2 \, diag((X^T X)^{-1}))$$

$$s^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\frac{\hat{\beta} - \beta}{\sqrt{s^2 \, diag((X^T X)^{-1})}} \sim t_{n-p}$$

# of parameters in the linear regression model

$$\boxed{P\left(\hat{\beta} - t_{n-p, 1-\frac{\alpha}{2}} \sqrt{s^2 \, diag((X^T X)^{-1})} \leq \beta \leq \hat{\beta} - t_{n-p, \frac{\alpha}{2}} \sqrt{s^2 \, diag((X^T X)^{-1})}\right) = 1-\alpha}$$

If we use $\alpha = 0.01$, we will obtain 99% CI