STATS 2107
Statistical Modelling and Inference II
Lecture notes
Chapter 1: Estimation

Jono Tuke

School of Mathematical Sciences, University of Adelaide

Semester 2 2017

Welcome

Contact

Jono Tuke:

`simon.tuke@adelaide.edu.au`

Lectures

- 5 a fortnight
- Wed 3; Thu 4; Fri 1 (odd weeks)

Tutorials

Monday 2pm or 3pm (even weeks)

Practicals

Tuesday 12pm (odd weeks)

Assignments

Due at 5pm on Friday in

- Week 3
- Week 5
- Week 7
- Week 9
- Week 11

Project

Due at 5pm on Friday in Week 12.

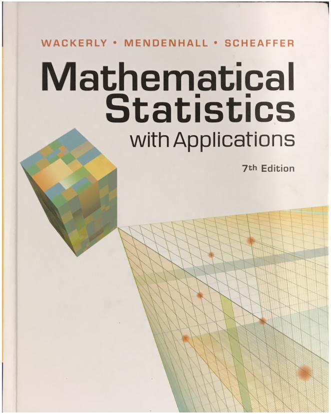
Assessment

- Assignments 20%
- Project 10%
- Exam 70%

Consulting

Room EMG13 Mondays 9am.

Reference text



Statistics

Setup

Suppose $Y_1, Y_2, \ldots, Y_n$ are random variables with cdf, F_θ for some $\theta \in \Theta$, where Θ denotes the set of legitimate parameter values, called the *parameter space*.

We would like to estimate the unknown parameter θ from the data $y_1, y_2, \ldots, y_n$.

Definition

A function $T(Y_1, Y_2, \ldots, Y_n)$ is called a **statistic**. A statistic T that takes values in Θ is called an **estimator** for θ.

MSE

Definition

Let T be an estimator for θ, then the **mean squared error** of T is defined by

$$MSE_T(\theta) = E[(T - \theta)^2].$$

Example

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent identically distributed (i.i.d.) $N(\mu, \sigma^2)$ random variables and let

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

be an estimator for μ. Calculate

$$MSE_{\bar{Y}}(\mu).$$

Example

Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. Bernoulli random variables with probability of success θ and let

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

be an estimator for θ. Prove

$$MSE_{\bar{Y}}(\theta) = \frac{\theta(1-\theta)}{n}.$$

Theorem

For any $k > 0$,

$$P(|T - \theta| \geq k\sqrt{MSE}) \leq \frac{1}{k^2}.$$

Proof

Bias

Definition

Let T be an estimator for θ, then the **bias** of T is defined by

$$b_T(\theta) = E[T] - \theta.$$

If $b_T(\theta) = 0$ for all θ, then T is said to be an **unbiased estimator** for θ.

Theorem

$$MSE_T(\theta) = var(T) + b_T(\theta)^2.$$

Proof

BLUE

Definition

An estimator of the form:

$$T = a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n = \sum_{i=1}^{n} a_i Y_i,$$

for some constants $a_1, a_2, \ldots, a_n$ is called a **linear estimator**.

Example

Is the sample mean $\bar{Y}$ a linear estimator?

Lemma

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random variables with

$$E[Y_i] = \mu_i \text{ and } var(Y_i) = \sigma_i^2.$$

Let

$$T = \sum_{i=1}^{n} a_i Y_i,$$

then

$$E[T] = \sum_{i=1}^{n} a_i \mu_i \text{ and } var(T) = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Futhermore, if

$$Y_i \sim N(\mu_i, \sigma_i^2) \text{ independently, then}$$

$$T \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Best Linear Unbiased Estimator (BLUE)

The best linear unbiased estimator for a parameter θ is the linear, unbiased estimator for θ that has minimum variance.

Lemma

Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with

$$E[Y_i] = \mu \text{ and } var(Y_i) = \sigma.$$

The linear estimator

$$T = \sum_{i=1}^{n} a_i Y_i$$

is unbiased for μ if and only if

$$\sum_{i=1}^{n} a_i = 1.$$

Theorem

Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with

$$E[Y_i] = \mu \text{ and } var(Y_i) = \sigma,$$

then the BLUE for μ is given by

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Proof

Standard error of an estimator

If T is an unbiased estimator for θ, then the standard deviation of the estimator is called the **standard error**:

$$SE(T) = \sqrt{var(T)}.$$

Confidence intervals

Definition

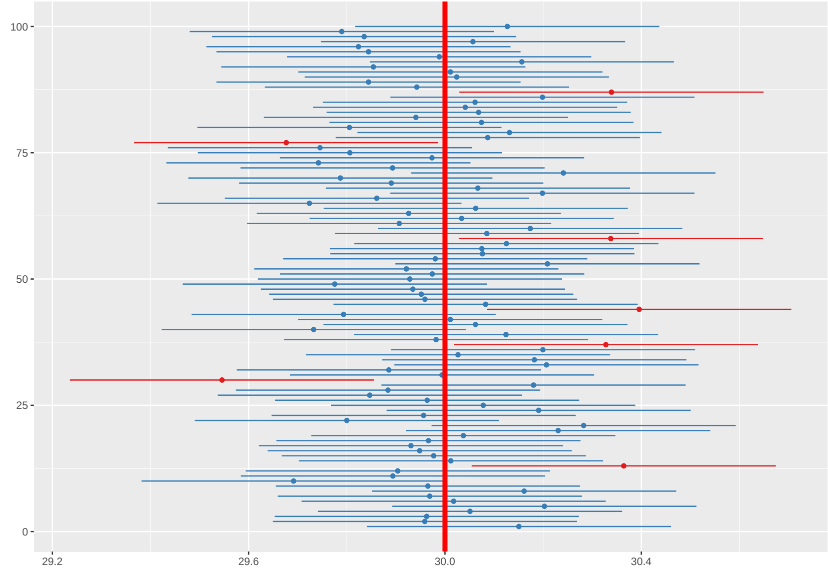
A random interval (L, U) is called a $100(1-\alpha)\%$ confidence interval for the parameter θ if it satisfies

$$P(L < \theta < U) = 1 - \alpha.$$

Remarks

- Note that the endpoints L and U are random: not θ.
- The quality $(1 - \alpha)$ is called the **coverage probability**. It lies between 0 and 1.

What do we mean by confidence?



Example

Suppose that $Y_1, Y_2, \dots, Y_n$ are i.i.d. $N(\mu, \sigma^2)$ with σ^2 known, then

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

where

$$P(Z > z_{\alpha/2}) = \alpha/2, \text{ for } Z \sim N(0,1),$$

is a $100(1-\alpha)\%$ confidence interval for μ.

Proof

Hypotheses tests

Statistical hypothesis

A statistical hypothesis is a statement about the unknown parameter θ. The most general formulation is

$$H_0 : \theta \in \Theta_0$$

where

$$\Theta_0 \subset \Theta,$$

and Θ is the parameter space.

Scalar parameter

Usually of form

$$H_0 : \theta = \theta_0.$$

Hypothesis test

A test of the hypothesis H_0 is a rule that tells us, for a given set of data $y_1, y_2, \ldots, y_n$ whether we should **retain (accept)** or **reject** H_0.

Hypothesis test

Usually a test is constructed from a **test statistic**, T, and a **critical region**, C, with the rule

- Reject H_0 if $T \in C$
- Retain H_0 if $T \notin C$

Example

If $y_1, y_2, \ldots, y_n$ are i.i.d. $N(\mu, \sigma^2)$ observations with σ^2 known.

Write down the appropriate test statistic and critical region to test

$$H_0 : \mu = \mu_0.$$

Type I and type II errors

	Retain H_0	Reject H_0
H_0 True	Correct conclusion	Type I error
H_0 False	Type II error	Correct conclusion

Significance level:

$$\alpha = P(\text{reject } H_0 | H_0 \text{ true}).$$

Power:

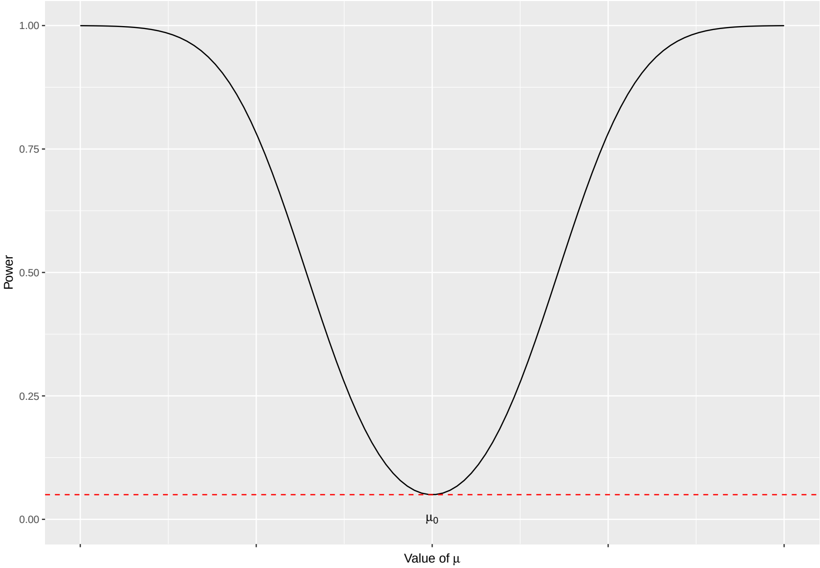
$$\beta = P(\text{reject } H_0 | H_0 \text{ false}) = 1 - P(\text{Type II error}).$$

Example

Show that in previous example of normal observations with σ^2 known, the test has significance level α.

Power function

Suppose that H_0 is false in the previous example, *i.e.*$\mu \neq \mu_0$



One-sided tests

Consider $y_1, y_2, \ldots, y_n$ are i.i.d. $N(\mu, \sigma^2)$ observations with σ^2 known.

We can test the one-sided hypothesis

$$H_0 : \mu \leq \mu_0.$$

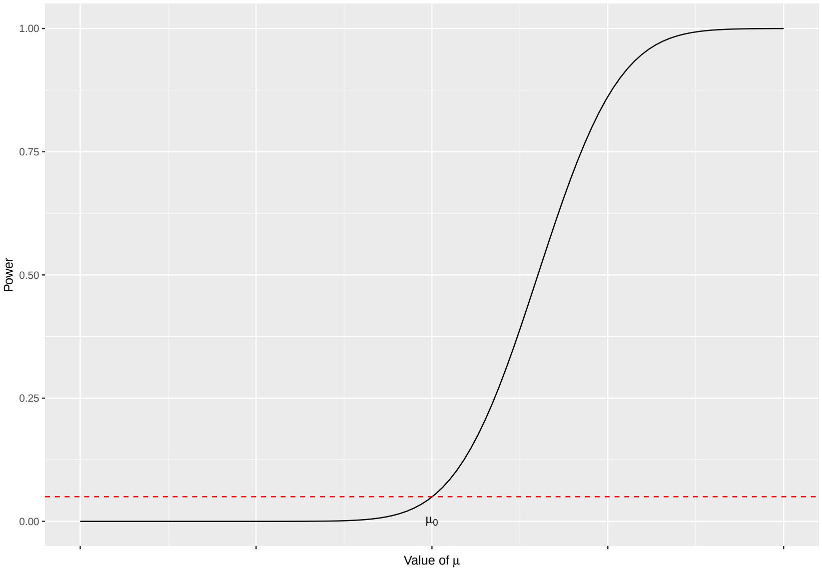
The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$

The rule is

$$\text{Reject } H_0 \text{ if } z \geq z_\alpha.$$

One-sided power function

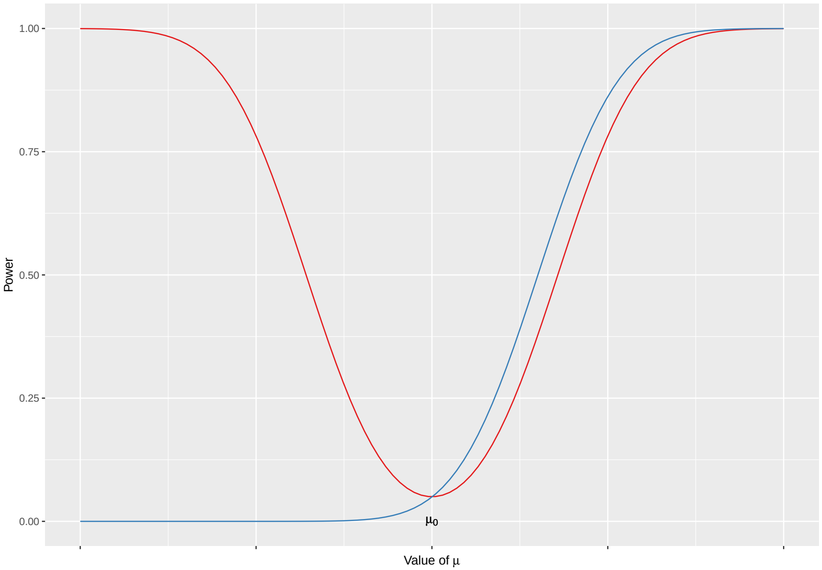


One-sided versus two-sided

- For one-sided test the type I error probability is not a single number, i.e.,

$$P(\text{reject} H_0 | H_0 \text{ true}) \leq \alpha$$

One-sided versus two-sided



One-sided versus two-sided warning

Assume that observed value of test statistic is

$$z_\alpha < z_{obs} < z_{\alpha/2}$$

Do not decide the number of sides after seeing the results

P-values

P-value

The P-value is the probability of observing data as extreme as that observed, if the null hypothesis is true.

Example

Suppose $y_1, y_2, \ldots, y_n$ are i.i.d. $N(\mu, \sigma^2)$ observations with σ^2 known. Consider the null hypothesis

$$H_0 : \mu = \mu_0,$$

with test statistic

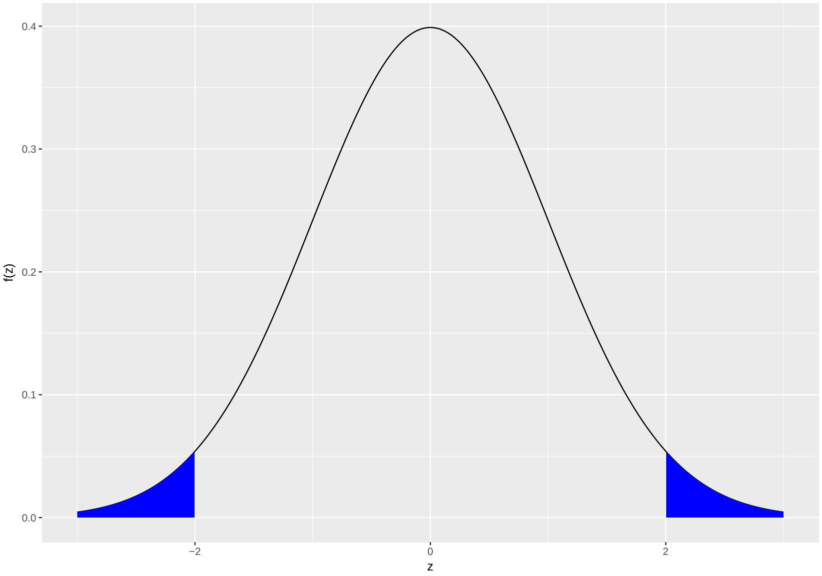
$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}},$$

then

$$\text{P-value} = P(|Z| \geq |z|)$$

for $Z \sim N(0, 1)$.

Example



Two-sided test using P-values

- Reject H_0 if P-value $\leq \alpha$
- Retain H_0 if P-value $> \alpha$.

Alternative interpretation of P-value

The P-value is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

Guidelines

P-value	conclusion
P-value > 0.1	no evidence against H_0
$0.05 <$ P-value ≤ 0.1	weak evidence against H_0
$0.01 <$ P-value ≤ 0.05	strong evidence against H_0
P-value ≤ 0.01	very strong evidence against H_0

P-values

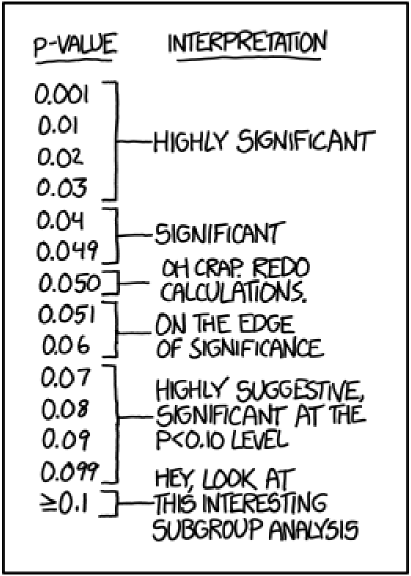


Figure 1: https://xkcd.com/1478/

Confidence intervals and hypothesis tests

If (L, U) is a $100(1-\alpha)\%$ confidence interval for θ, then the test

- Reject $H_0 : \theta = \theta_0$ if $\theta_0 \notin (L, U)$
- Retain $H_0 : \theta = \theta_0$ if $\theta_0 \in (L, U)$

is a test of $H_0 : \theta = \theta_0$ with significance level α.