Course 02403 Introduction to Mathematical Statistics

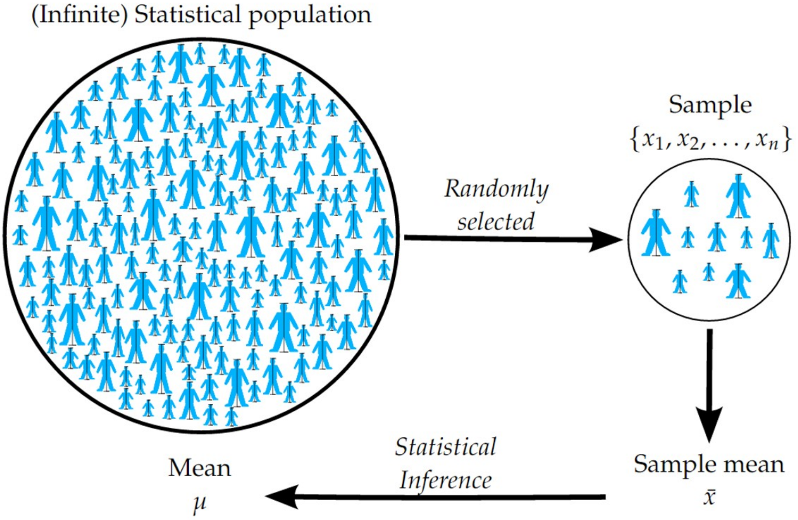
Lecture 13: An overview of course content

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Overview

Overview

Statistics

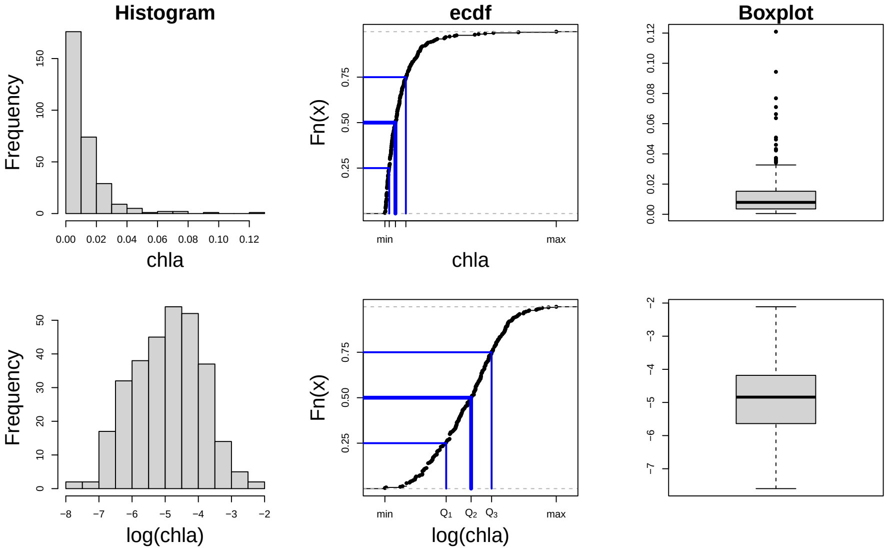

(Infinite) Statistical population
Randomly selected
Sample
$\{x_1, x_2, \ldots, x_n\}$
Sample mean
$\bar{x}$
Statistical Inference
Mean
μ

Summary statistics (Nøgletal)

We use *summary statistics* to summarize and describe data (random variables)

- **Average:** Measure of center / location
- **Median:** Measure of center / location
- **Variance:** Variation
- **Standard deviation:** Variation (same unit as data)
- **Coefficient of variation:** Variation in data (unit less)
- **Covariance:** (linear) interdependence
- **Correlation:** (linear) interdependence (unit less)
- **Quantiles:** For making statements about the data distribution
- **Second order moment repesentation:** Mean and variance-covariance

Graphical summaries: Plots



Overview

Random variable

Before the experiment is carried out, we have a random variable

$$Y \text{ (or } Y_1, \ldots, Y_n)$$

indicated with capital letters.
Formally Y is a function that assign probabilities to subsets of possible outcomes, e.g. if Y is the number rolled with a fair dice then $P(Y = 1) = \frac{1}{6}$ and $P(Y \in \{1, 2\}) = \frac{2}{6}$.

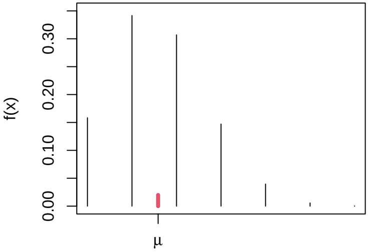
After the experiment is carried out, we have a *realization* or *observation*

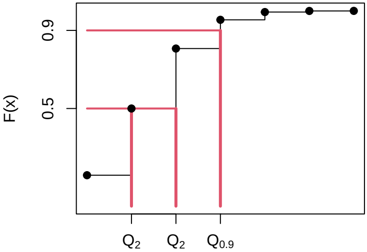
$$y \text{ (or } y_1, \ldots, y_n)$$

indicated with lowercase letters. y is a number (i.e. NOT a random variable), e.g. we roll 2 with a fair dice.

Summary statistics (Empirical and model based)

	Empirical	Discrete random variable
Mean	$\bar{y} = \sum y_i \frac{1}{n}$	$\mu = \sum y_i f(y_i)$
Variance	$s^2 = \sum (y_i - \bar{y})^2 \frac{1}{n-1}$	$\sigma^2 = \sum (y_i - \mu)^2 f(y_i)$
Median	$y_{(\lceil n/2 \rceil)}$ [1]	"$F^{-1}(0.5)$" [2]
Quantile	Q_τ [1]	"$F^{-1}(\tau)$"





[1] see lect01 for precise definition

[2] More precisely: x s.t. $P(Y \leq y) \geq 0.5$ and $P(Y \geq y) \geq 0.5$

Discrete distributions used in this course

Distribution	f(y)	μ	σ^2	Typical application
$Y \sim B(n,p)$	$\binom{n}{y} p^y (1-p)^{n-y}$	np	$np(1-p)$	Flip a coin n-times (succes prob p).
$Y \sim H(n,a,N)$	$\frac{\binom{a}{y}\binom{N-a}{n-y}}{\binom{N}{n}}$	$n\frac{a}{N}$	$n\frac{a}{N}\frac{(N-a)}{N}\frac{N-n}{N-1}$	Number of white balls drawn from an urn with N balls and a white balls.
$Y \sim P(\lambda)$	$\frac{\lambda^y}{y!}e^{-\lambda}$	λ	λ	Number of arivals per hour when average number of arivals per hour is λ.

- There exist a number of other discrete distributions, fit for different porpuses.

Continuous distributions: overview

Distribution	pdf	μ	σ^2	Typical application
$Y \sim U(\alpha, \beta)$	$\frac{1}{\beta-\alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$	Constant density in the interval (α, β), zero outside.
$Y \sim Exp(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Time between arrivals when mean time equal λ.
$Y \sim N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	Distribution of measurement errors.
$Y \sim LN(\alpha, \beta^2)$	$\frac{1}{x\beta\sqrt{2\pi}} e^{-\frac{(\ln(x)-\alpha)^2}{2\beta^2}}$	$e^{\alpha+\beta^2/2}$	$\mu^2(e^{\beta^2}-1)$	Distribution of concentrations.

Rules for random variables (*both discrete and continuous*)

Let Y be a random variable, while a and b are constants, then

$$\mathrm{E}[aY+b] = a\mathrm{E}[Y]+b$$
$$\mathrm{V}[aY+b] = a^2\mathrm{V}[Y]$$

Let $Y_1,\ldots,Y_n$ be random variables, then

$$\mathrm{E}\left[\sum_{i}^{n} a_i Y_i\right] = \sum_{i=1}^{n} a_i \mathrm{E}[Y_i]$$

If $Y_1,\ldots,Y_n$ are *independent*, then

$$\mathrm{V}\left[\sum_{i=1}^{n} a_i Y_i\right] = \sum_{i=1}^{n} a_i^2 \mathrm{V}[Y_i]$$

Matrix calculation rules for random variable

Let $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ be random vectors with

$$V \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{bmatrix},$$

then

Theorem

Let the variance-covariance matrix of $[\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T]^T$ be as above and let $\boldsymbol{b}$ be a vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of appropriate dimensions, then

$$E[\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{b}] = \boldsymbol{A}E[\boldsymbol{Y}_1] + \boldsymbol{b}$$
$$Cov[\boldsymbol{A}\boldsymbol{Y}_1, \boldsymbol{B}\boldsymbol{Y}_2] = \boldsymbol{A}Cov[\boldsymbol{Y}_1, \boldsymbol{Y}_2]\boldsymbol{B}^T = \boldsymbol{A}\boldsymbol{\Sigma}^{12}\boldsymbol{B}^T$$

and as a special case

$$V[\boldsymbol{A}\boldsymbol{Y}_1] = Cov[\boldsymbol{A}\boldsymbol{Y}_1, \boldsymbol{A}\boldsymbol{Y}_1] = \boldsymbol{A}\boldsymbol{\Sigma}^{11}\boldsymbol{A}^T.$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be such that $\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{B}\boldsymbol{Y}_2$ can be formed, then

$$V[\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{B}\boldsymbol{Y}_2] = \boldsymbol{A}\boldsymbol{\Sigma}^{11}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{\Sigma}^{22}\boldsymbol{B}^T + \boldsymbol{A}\boldsymbol{\Sigma}^{12}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Sigma}^{21}\boldsymbol{A}^T.$$

Error propagation

Assume that Y_i are random variables with $E(Y_i) = \mu_i$ og $V(Y_i) = \sigma_i^2$ og $Cov(Y_i, Y_j) = \sigma_{ij}$

We need to find:

$\sigma^2_{f(Y_1,\ldots,Y_n)} = \mathrm{Var}(f(Y_1,\ldots,Y_n))$

(Generalization of) Method 4.3: for non-linear functions:

$$\sigma^2_{f(Y_1,\ldots,Y_n)} \approx \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_i}\right)^2 \sigma_i^2 + 2\sum_{i}\sum_{j>i} \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\sigma_{ij}$$

Where the derivatives of f are evaluated at $\mu_1, ..., \mu_n$.
Or the more general matrix formulation

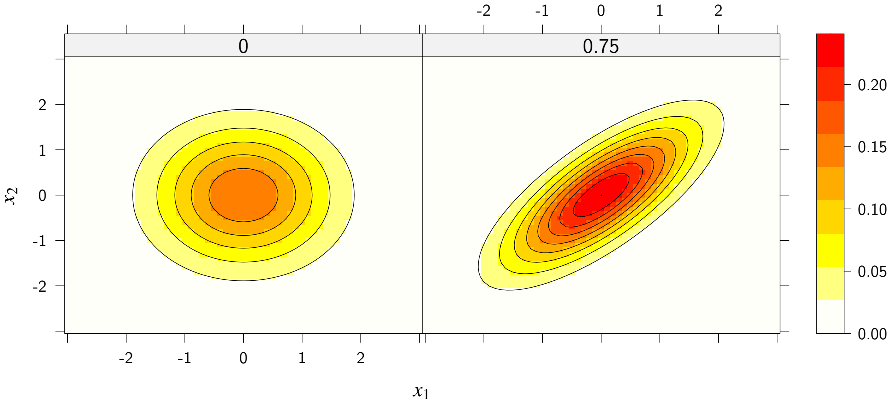
$$\begin{aligned} E[f(\boldsymbol{Y})] &\approx f(\boldsymbol{\mu}) \\ V[f(\boldsymbol{Y})] &\approx \boldsymbol{J}_f(\boldsymbol{\mu})\boldsymbol{\Sigma}\boldsymbol{J}_f(\boldsymbol{\mu})^T. \end{aligned}$$

Simulation is also a possibility.

Overview

Bivariate Normal density

Normal density for different values of the correlation.



Multivariate normal: general definition

Definition (Multivariate normal distribution)

Let Z_i, $i = 1, ..., n$, be iid. standard normal random variables, s.t. $(\boldsymbol{Z} = [Z_1, ..., Z_n]^T)$

$$\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}).$$

Then the random vector $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}$, with $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, follow an m-dimensional multivariate normal distribution with

$$\begin{aligned} E[\boldsymbol{Y}] &= \boldsymbol{b} \\ V[\boldsymbol{Y}] &= \boldsymbol{A}\boldsymbol{A}^T, \end{aligned}$$

this holds also when $\boldsymbol{A}\boldsymbol{A}^T$ is not positive definite.

Overview

Some general concepts

Central Estimator:
An estimator, $\hat{\theta}$, is central (or non-biased), if and only if, the mean value of the estimator equals θ

Consistent Estimator
A central estimator, $\hat{\theta}$, that converge in probability is called a consistent estimator (you can think of this as $V(\theta_n) \to 0$).

Efficient Estimator
An estimator $\hat{\theta}_1$ is a more efficient estimator for θ than $\hat{\theta}_2$ if:

1. $\hat{\theta}_1$ and $\hat{\theta}_2$ both are central estimators of θ
2. The variance of $\hat{\theta}_1$ is less than the variance of $\hat{\theta}_2$

Estimate
When we have the actual sample and have calculated the summary statistic, we have an estimate (this is not a random variabele)

Sampling distributions

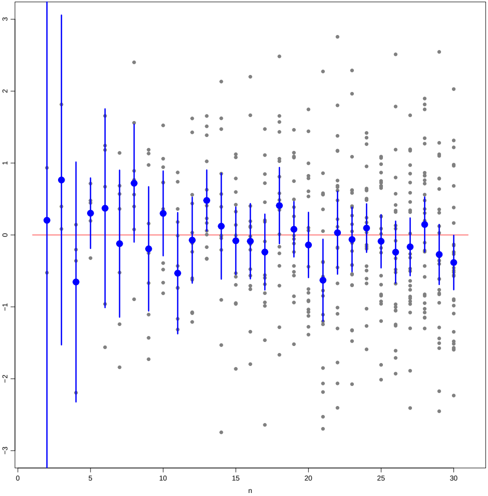
In lecture 1 we saw a number of summary statistics, we now assume that

$$Y_i \sim N(\mu, \sigma^2), \quad \text{and iid.}$$

In this and the next lecture we will answer the following questions

- What is the distribution of $\bar{Y}$? $\bar{Y} \sim N(\mu, \sigma^2/n)$
- What is the distribution of S^2? $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
- What is the distrubution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$? $\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$
- If we calculated observed variances from two different groups, what is then the distribution of $\frac{S_1^2}{S_2^2}$? $\frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$, if variance equal in the two groups.

Confidence intervals for increasing sample size



Overview

The general linear model

The general linear model is a statistical model that can be written in the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

or $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$.

- $\boldsymbol{Y}$ are the observations
- $\boldsymbol{X}$ is the design matrix
- $\boldsymbol{\beta}$ is a vector of parameters
- $\boldsymbol{\epsilon}$ are the residuals

The general linear model as a projection

The fitted values in a general linear model can be ritten as

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

and the observed residuals can be written as

$$\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y},$$

where

- $\boldsymbol{H}$ is an orthogonal projection matrix
- $\boldsymbol{r}$ and $\hat{\boldsymbol{Y}}$ are independent.
- The dimension of the model is $Trace(\boldsymbol{H}) = Rank(\boldsymbol{X}) = p$
- If two design matrices have the same projection matrix then the models are equivalent.

Cochrans theorem

Theorem (Cochran's theorem)

Let $\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, and let $\boldsymbol{H}_i$ be orthogonal projection matrices such that

$$\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{Y} = \frac{1}{\sigma^2}\sum_{i=1}^{K}\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y}$$

i.e. $\sum_{i=1}^{K}\boldsymbol{H}_i = \boldsymbol{I}_n$, *with* $Rank(\boldsymbol{H}_i) = p_i$, *and* $\sum_i p_i = n$ *then*

1. $\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y} \sim \chi^2(p_i)$
2. $\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y}$ *and* $\boldsymbol{Y}^T\boldsymbol{H}_j\boldsymbol{Y}$ *are independent for* $i \neq j$.

Type I partioning of variation

Consider a series of nested hypothesis

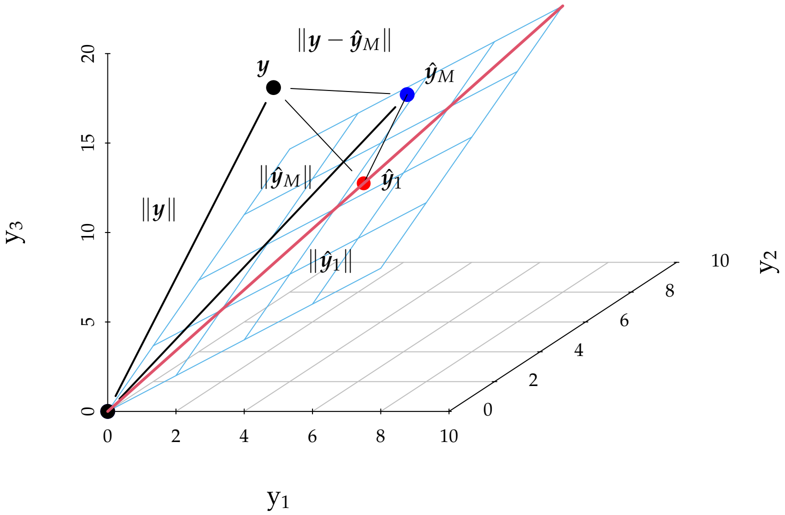
$$H_0 \subset H_1 \subset \cdots \subset H_M \subset \mathbb{R}^n$$

corresponding to the design matrices

$$\begin{aligned}
\boldsymbol{X}_0 =& \mathbf{1} \\
\boldsymbol{X}_1 =& [\mathbf{1} \quad \tilde{\boldsymbol{X}}_1] \\
&\vdots \\
\boldsymbol{X}_i =& [\boldsymbol{X}_{i-1} \quad \tilde{\boldsymbol{X}}_i] \\
&\vdots \\
\boldsymbol{X}_M =& [\boldsymbol{X}_{M-1} \quad \tilde{\boldsymbol{X}}_M],
\end{aligned}$$

and corresponding projection matrices $\boldsymbol{H}_i = \boldsymbol{X}_i(\boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i^T$.

Example: Items on a scale, projections



Overview

Interval Estimation: Confidence Intervals

Consider the $1-\alpha$ random interval $I(\boldsymbol{Y},\alpha)$ for the variable θ (e.g. μ or σ^2) such that

$$P(\theta \in I(\boldsymbol{Y},\alpha)) = 1-\alpha.$$

Notice that the interval (not θ) is random. The realization $I(\boldsymbol{y},\alpha)$ is referred to as thhe **confidence interval** for θ.

Repeated sampling interpretation

If the experiment is repeated K times $(K >> 1)$, then we expect that the true parameter is included in $(1-\alpha)K$ intervals.

Notice that in practice we have one interval and make statements based on that one interval! For n large formulas can be applied for even for no-normal data (CLT).

Confidence interval for the mean

$$I(\boldsymbol{y}, \alpha) = \bar{y} \pm t_{1-\alpha/2} \frac{s}{\sqrt{n}} = \bar{y} \pm ME$$

The Margin of Error is (for σ unknown)

$$ME = t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

The margin of error (and hence the confidence interval width)

- increase when α decrease towards zero
- decrease with number of observations
- vary from sample to sample (through s)

Further the location varies (through $\bar{y}$) from sample to sample.

Interpretation

The onfidence interval is NOT about single observations. It is about the location of the true (unknown) mean value.

Method 3.19: Confidence intervals for variance and standard deviation

Let $Y_i \sim N(\mu, \sigma^2)$ for $i = 1, \dots, n$ be iid.

Variance:

A $100(1-\alpha)\%$ confidence interval for the variance σ^2 is given by:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}};\ \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right],$$

where the quantiles come from a χ^2 distribution with $n-1$ degrees of freedom.

Standard deviation:

A $100(1-\alpha)\%$ confidence interval for the standard deviation σ is:

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}};\ \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right].$$

Hypothesis test

A *null hypothesis* is rejected if the outcome (i.e. the data) is *unusual* under the null hypothesis (and the assumptions).

Definition

An observation is unusual if the probability of the observation or something more extreme is small (i.e. less than α). More extreme is understod in terms of distance to the null hypethesis.[a]

[a] This is the no-directional p-value and unidirectional tests also exist.

Example: Assume that $Y \sim N(0,1)$, we observe $y = 3$ in order to determine if y is unusual we calculate

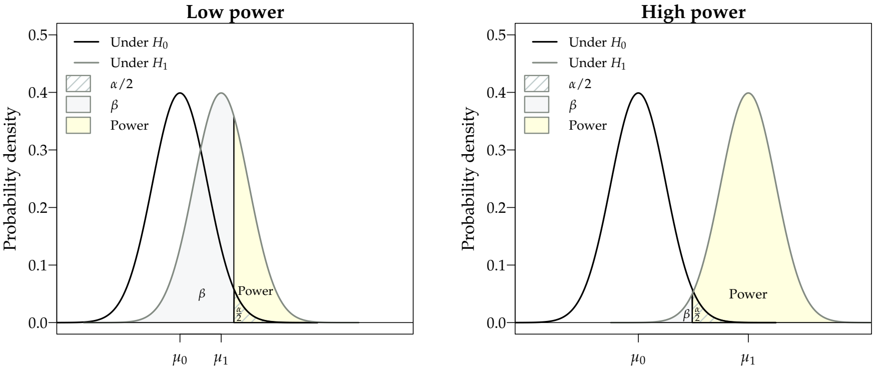
$$P(Y > |y|) + P(Y < -|y|) = 2(1 - P(Y < |y|) = 2(1 - F(y)) = 0.0027$$

which means that it is unusual if α is chosen as 0.05.

Power

Two possible truths against two possible conclusions:

	Rejecting H_0	Not rejecting H_0
H_0 is true	Type I error (α)	Correct acceptance of H_0
H_0 is false	Correct rejection of H_0	Type II error (β)



Overview

Pooled variance set up

Let $Y_{1,1},\ldots,Y_{1,n_1}$ be iid. with $Y_{1,i} \sim N(\mu_1,\sigma^2)$ and $Y_{2,1},\ldots,Y_{2,n_2}$ be iid. with $Y_{2,i} \sim N(\mu_2,\sigma^2)$, test the hypothesis $\mu_1-\mu_2=\delta_0$. We will consider the test statistics

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - \delta_0}{S_{\bar{Y}_1-\bar{Y}_2}}$$

In an informal way we can write this as

$$T = \frac{\text{observation-hypothesis}}{\text{standard deviation under assumptions}}$$

We need to establish the distribution of T.

Pooled variance, test statistics and confidence interval

Based on the distribution of T we find the test statistics

Test statistics and p-value

$$t_{\text{obs}} = \frac{(\bar{y}_1 - \bar{y}_2) - \delta_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}; \quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2},$$

and

$$p-value = 2P(T > |t_{obs}|) = 2(1 - P(T < |t_{obs}|); \quad T \sim t(n_1+n_2-2)$$

$(1-\alpha)$-confidence interval for δ

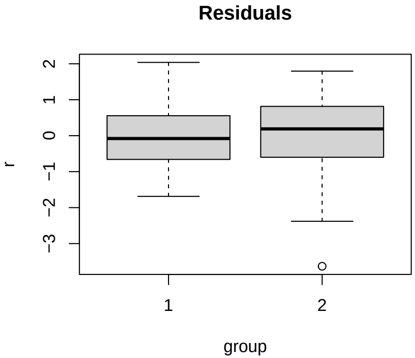
$$CI(\alpha) = \bar{y}_1 - \bar{y}_2 \pm t_{1-\alpha/2} s_p \sqrt{\frac{1}{n_2} + \frac{1}{n_2}}$$

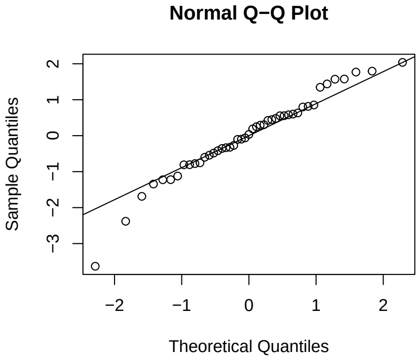
where $t_{1-\alpha/2}$ is based on the t-distribution with n_1+n_2-2 degrees of freedom.

Checking assumptions

Pooled variance

If the pooled variance two sample test is used then in addition the equal variance assumption should also be checked (eg. box-plots or compare observed variances). (see LM formulation for another way)





Two sample t-test as an LM

The two sample t-test, assumening equal variance in the two groups, can be written as an LM

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

where $\boldsymbol{X} \in \mathbb{R}^{(n_1+n_2)\times 2}$, while the projection matrix $(\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)$ is unique the design matrix $\boldsymbol{X}$ is not. Some parametrizations are

$$\boldsymbol{X}_1 = \begin{bmatrix} \boldsymbol{1}_{n_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n_2} \end{bmatrix}; \quad \boldsymbol{X}_2 = \begin{bmatrix} \boldsymbol{1}_{n_1} & \boldsymbol{0} \\ \boldsymbol{1}_{n_2} & \boldsymbol{1}_{n_2} \end{bmatrix}; \quad \boldsymbol{X}_3 = \begin{bmatrix} \boldsymbol{1}_{n_1} & -\frac{1}{2}\boldsymbol{1}_{n_1} \\ \boldsymbol{1}_{n_2} & \frac{1}{2}\boldsymbol{1}_{n_2} \end{bmatrix}$$

- Different parmarametrization result in different parameter interpretations, while the fitted values are unaffected.

Overview

Multiple linear regression

We can of course imagine more than one explanatory variable, corresponding to the model

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i} + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2) \text{ og i.i.d.}$$

or

$$\begin{aligned}\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} &= \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \varepsilon_i \sim N(0, \sigma^2) \\ &= \boldsymbol{X}\beta + \varepsilon \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}),\end{aligned}$$

Parameter estimates

All results from the simple linear regression carry over (with minor adjustment). The estimators of the parameters in the simple multiple regression model are given by

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

and the covariance matrix of the estimates is

$$V[\hat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

and central estimate for the residual variance is

$$\hat{\sigma}^2 = \frac{RSS}{n-(p+1)}$$

Residuals and model reduction

Further

- Confidence intervals for parameters
- Confidence and prediction intervals for the line
- Standardized residuals and leverage
- Multicollinarity
- Polynomial regression

Overview

One-Way ANOVA - Decomposition and ANOVA Table

- With the model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

  the total variation in data can be decomposed:

$$SST = SS(Tr) + SSE\,.$$

- 'One-way' implies that there is only one factor in the experiment (with k levels).
- The method is called analysis of variance because testing is done by comparing variances.

Two-Way Analysis of Variance - Model

- The model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

  where the errors are independent and identically distributed with

$$\varepsilon_{ij} \sim N(0, \sigma^2)\,.$$

  - μ is the overall mean
  - α_i represents the effect of treatment $i \in \{1, ..., k\}$
  - β_j represents the effect of block $j \in \{1, ..., l\}$
  - There are k treatments and l blocks

- Note: In this course, we only have one observation in each cell (i.e., with the same α and the same β) when performing two-way ANOVA.

ANOVA-table

Source of variation	Deg. of freedom	Sums of squares	Mean sum of squares	Test-statistic F	p-value
Treatment	$k-1$	$SS(Tr)$	$MS(Tr) = \frac{SS(Tr)}{k-1}$	$F_{\text{Tr}} = \frac{MS(Tr)}{MSE}$	$P(F > F_{\text{Tr}})$
Block	$l-1$	$SS(Bl)$	$MS(Bl) = \frac{SS(Bl)}{l-1}$	$F_{\text{Bl}} = \frac{MS(Bl)}{MSE}$	$P(F > F_{\text{Bl}})$
Residual	$(k-1)(l-1)$	SSE	$MSE = \frac{SSE}{(k-1)(l-1)}$		
Total	$n-1$	SST			

Generalizations, its is all about $\boldsymbol{X}$

We have studied the model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0},^2 \boldsymbol{I}).$$

The (very) general theory is given in Chapter 9, and specific examples

1. The columns of $\boldsymbol{X}$ given as zero and ones (t-test and ANOVA, Chapter 3 and 8)
2. The columns of $\boldsymbol{X}$ given as real numbers (regression, Chapter 5 and 6)

Further we assumed balanced design in the two-way ANOVA. The set up is fairly easily generalized (meaning that all general formulas for the LM transfer directly), to

- Non-balanced design, i.e. different number of observations in each group (e.g. missing observations, or multiple observations in some cells).
- Multiple (more than 2) factors.
- Mix of regression and factor analysis (different slopes in different groups)
- Interaction effects, i.e. "effect of factor A and B" $\neq$ "effect of factor A $+$ factor B".

Overview

Bootstrapping: An overview

We have seen 4 not so different method boxes

1. With or without distribution assumptions (parametric or non-parametric)
2. Analyses with one or two samples (one or two groups)

Note:

Means are also included in *random sample functions*. That is, these methods can also be applied for analyses beyond means!

Inferences for Proportions

- Specific methods, one, two and $k > 2$ samples
  - Binary/categorical response
- Estimation and confidence interval of proportions
  - Large sample vs. small sample methods
- Hypotheses for one proportion
- Hypotheses for two proportions
- Analysis of contingency tables (χ^2-test) (All expected > 5)

Overview

Some further perspectives

- Likelihood theory (estimations teknik)
- General Linear Models (GLM) - generalisering af multiple linear regression og variansanalyse
- Generalized Linear Models - LM but non-Gaussian data
- Correlations struktures
- Stochastic dynamical systems
- and much more....

Some next level cources

- 02405 Probability
- 02417 Time Series Analysis
- 02418 Statistical modelling: Theory and practice
- 02411 Design of experiments
- 02413 Statistical Quality Control
- 02441 Applied Statistics and Statistical Software

And explore for cources that are not in this list....

Some advanced cources

There will be a “Statistical modelling” specialization from next semester.

- 02426 Non-linear random effect models: time-independent and dynamic models
- 02427 Advanced Time Series Analysis
- 02407 Stochastic Processes - Probability 2
- 02807 Computational Tools for Data Science
- 02429 Analysis of correlated data: Mixed linear models
- 02582 Computational Data Analysis
- 02443 Stochastic Simulation
- 02586 Statistical genetics
- 02409 Multivariate Statistics

And explore for cources that are not in this list....

Some advanced cources

And of course combined with machine learning, control and other topics.

Overview