Dept. of Probability and Mathematical Statistics

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

doc. RNDr. Arnošt Komárek, Ph.D.

NMSA407 Linear Regression

Winter term 2021–22

Cast

Lectures (Tuesday 11:30 – 14:40 in K1)

break of about 10 minutes at some point around the middle

doc. RNDr. Arnošt Komárek, Ph.D.

`komarek@karlin.mff.cuni.cz`

`http://msekce.karlin.mff.cuni.cz/~komarek`

2nd floor next to the stairs

Exercise class (Thursday 15:40 in K4 and 17:20 in K11)

RNDr. Matúš Maciak, Ph.D.

`maciak@karlin.mff.cuni.cz`

`http://www.karlin.mff.cuni.cz/~maciak`

1st floor between the stairs and the library

Exercise class (Tuesday 17:20 in K4)

Mgr. Stanislav Nagy, Ph.D.

`nagy@karlin.mff.cuni.cz`

`http://msekce.karlin.mff.cuni.cz/~nagy`

4th floor

Study materials

Webpage of the course

http://msekce.karlin.mff.cuni.cz/~komarek/vyuka/nmsa407.html

Central webpage of the exercise classes

http://msekce.karlin.mff.cuni.cz/~maciak/nmsa407_2022.php

Study materials

1. Self-written notes made during the lecture.

2. Course notes
   - Should be used selectively as a supplement to self-written notes.
   - They contain (much) more than what's required to pass the exam.
   - Some parts of the lecture will be presented a bit differently as compared to the course notes.

3. Slides
   - Pure complement to information being provided orally and "on the blackboard" (irrespective of what "blackboard" means during the COVID-19 pan(dem)ic).

Past experience suggests that individual reading of the notes only is in most cases insufficient to be prepared for exam. The course notes are intended as a **supplement** of the lecture, **not its replacement**.

Basic supplementary

Khuri, A. I. (2010). *Linear Model Methodology*.
Boca Raton: Chapman & Hall/CRC. ISBN 978-1-58488-481-1.

Zvára, K. (2008). *Regrese*.
Praha: Matfyzpress. ISBN 978-80-7378-041-8.

Extended supplementary

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis, Second Edition*.
New York: John Wiley & Sons. ISBN 978-0-471-41540-4.

Draper, N. R., Smith, H. (1998). *Applied Regression Analysis, Third Edition*.
New York: John Wiley & Sons. ISBN 0-471-17082-8.

Sun, J. (2003). *Mathematical Statistics, Second Edition*.
New York: Springer Science+Business Media. ISBN 0-387-95382-5.

Weisberg, S. (2005). *Applied Linear Regression, Third Edition*.
Hoboken: John Wiley & Sons. ISBN 0-471-66379-4.

Anděl, J. (2007). *Základy matematické statistiky*.
Praha: Matfyzpress. ISBN 80-7378-001-1.

Cipra, T. (2008). *Finanční ekonometrie*.
Praha: Ekopress. ISBN 978-80-86929-43-9.

Zvára, K. (1989). *Regresní analýza*.
Praha: Academia. ISBN 80-200-0125-5.

The lectures shall not follow closely any of the books.

Exercise classes

During semester

- Practical analyses of various types of datasets.
- Theoretical assignments.

Principal computational environment

- System R (http://www.R-project.org).
- Possibly (but not necessarily) combined with RStudio (http://www.rstudio.org).
- Exercise classes are not a course in R programming!
- Emphasis on interpretation of results.

"Technical" materials (how to do calculations in R):

- R tutorials at http://msekce.karlin.mff.cuni.cz/~komarek/vyuka/nmsa407.html
- Just supplementary.

Course credit (*Zápočet*)

- Details have been (will be) provided on the web and during the first "exercise classes".

Exam

1. **Written part** composed of theoretical and semi-practical assignments (no computer analysis).

2. **Oral part** (extent depending on results of the written part).

- The exam dates for the written part will be communicated in due time via SIS. All ($\pm$five) exam dates will be in a period

  January 10 – February 11, 2022.

- There will be no exam dates later on!

Unavoidable prerequisites

- NMSA331 and 332: Mathematical Statistics 1 and 2;
- NMSA333: Probability Theory 1;
- NMSA336: Introduction to Optimisation;
- NMAG101 and 102: Linear Algebra and Geometry 1 and 2.

Other prerequisites

- All other compulsory (optional) subjects of Bachelor study branch General mathematics, direction Stochastics.

Prerequisite knowledge

The most important areas of general mathematics and mathematical statistics which are unavoidable to be able to follow this course include:

- **Vector spaces, matrix calculus**;
- **Probability space, conditional probability, conditional distribution, conditional expectation**;
- **Elementary asymptotic results** (laws of large numbers, central limit theorem for i.i.d. random variables and vectors, Cramér-Wold theorem, Cramér-Slutsky theorem);
- **Foundations of statistical inference** (statistical test, confidence interval, standard error, consistency);
- **Basic procedures of statistical inference** (asymptotic tests on expected value, one- and two-sample t-test, one-way analysis of variance, chi-square test of independence);
- **Maximum-likelihood theory** including asymptotic results and the delta method;
- **Working knowledge of R**.

1

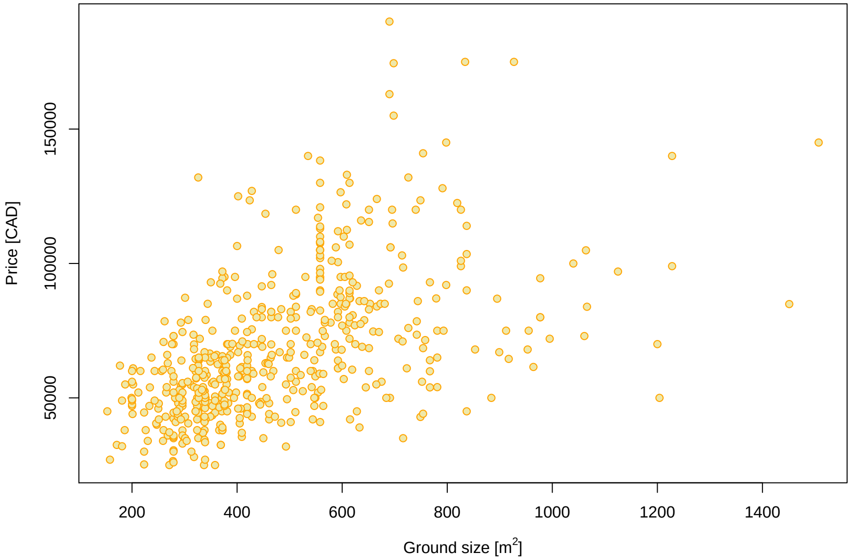
Linear Model

Section **1.1**

Regression analysis

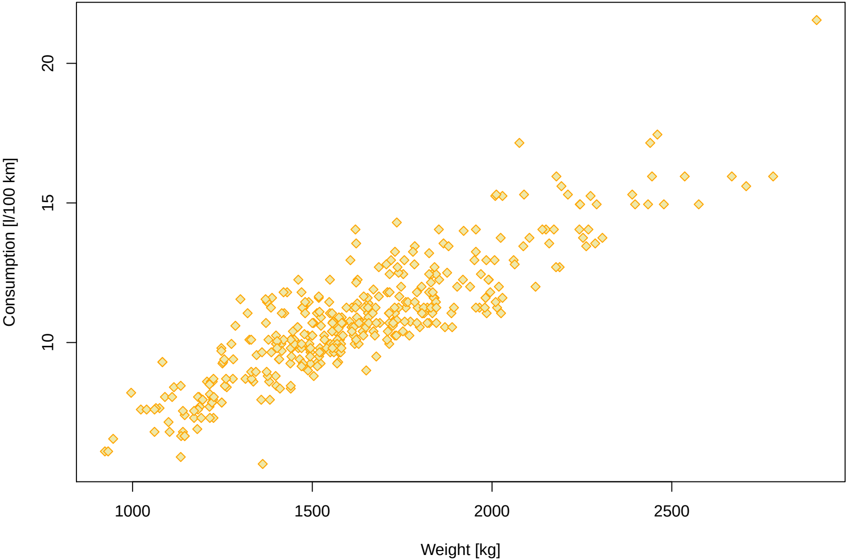
Houses1987 ($n = 546$)

price ~ ground



Cars2004nh (subset, $n = 409$)

consumption ∼ weight

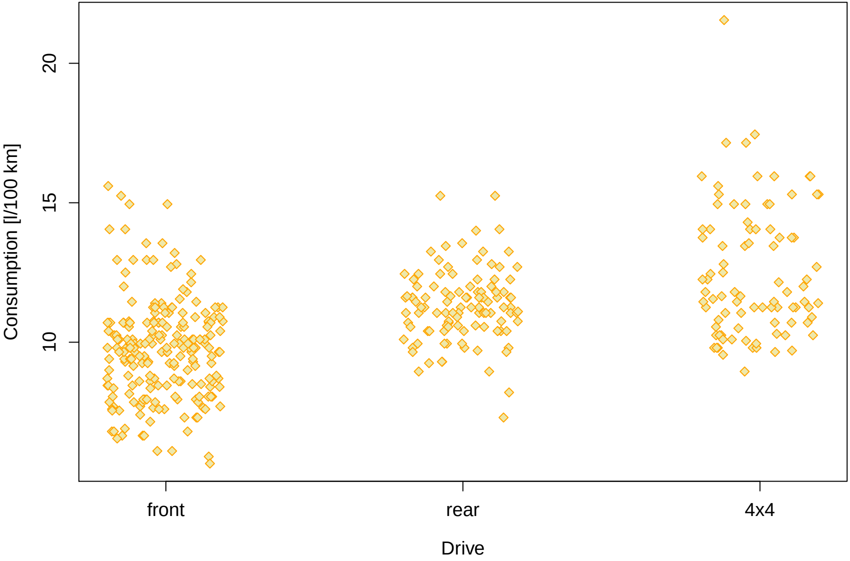


Cars2004nh (subset, $n = 409$)

consumption ∼ drive

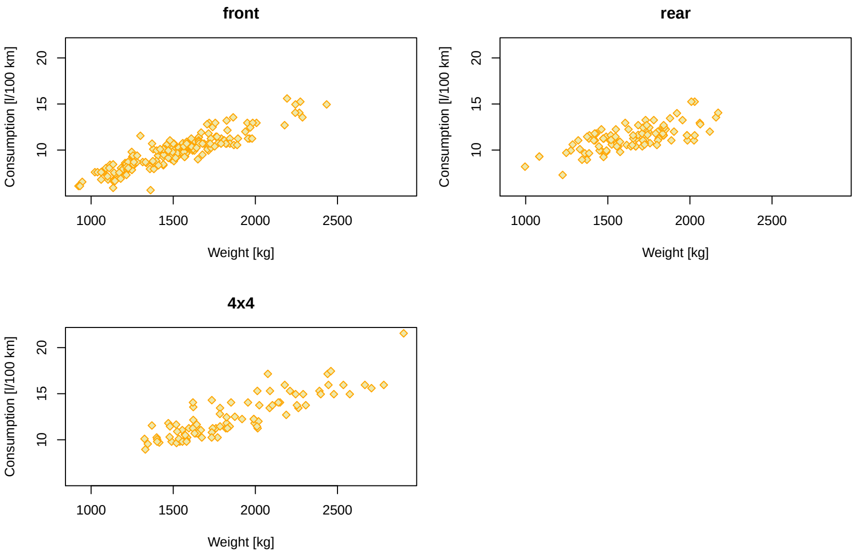
Cars2004nh (subset, $n = 409$)

consumption ~ drive



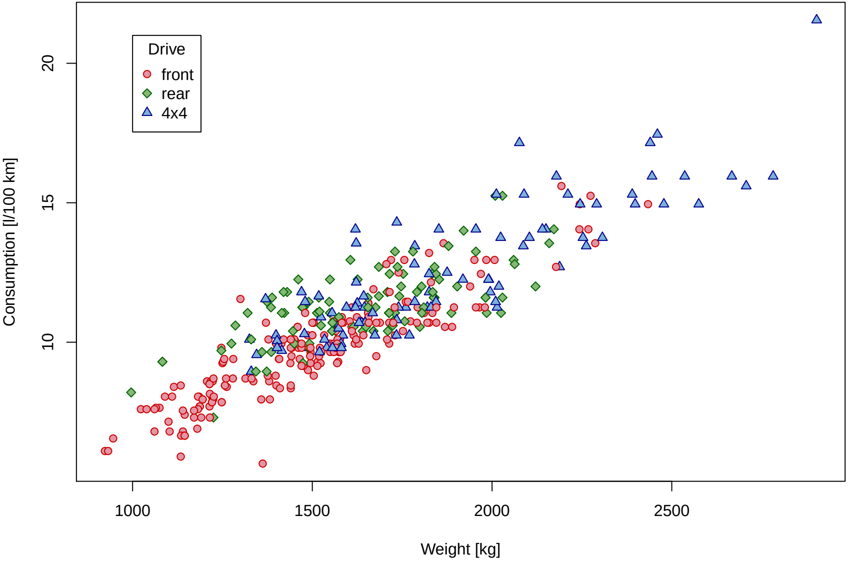
Cars2004nh (subset, $n = 409$)

consumption ∼ weight, drive



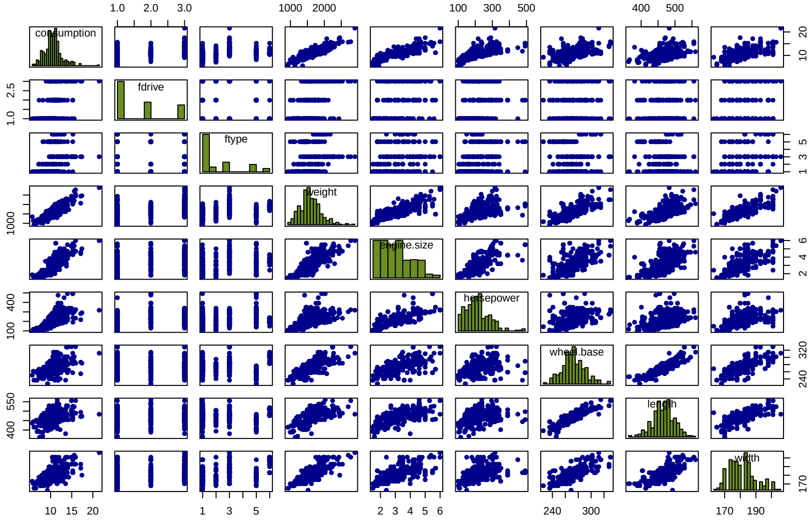
Cars2004nh (subset, $n = 409$)

`consumption ~ weight, drive`



Cars2004nh (subset, $n = 384$)

consumption ~ drive, type, weight, engine.size, horsepower, wheel.base, length, width



Section **1.2**

Linear model: Basics

Definition 1.1 Linear model with i.i.d. data.

The data $\left(Y_i, \boldsymbol{X}_i^\top\right)^\top \overset{\text{i.i.d.}}{\sim} \left(Y, \boldsymbol{X}^\top\right)^\top$, $i = 1, \ldots, n$, satisfy a *linear model* if

$$\mathbb{E}\big(Y \,\big|\, \boldsymbol{X}\big) = \boldsymbol{X}^\top \boldsymbol{\beta}, \qquad \text{var}\big(Y \,\big|\, \boldsymbol{X}\big) = \sigma^2,$$

where $\boldsymbol{\beta} = \left(\beta_0, \ldots, \beta_{k-1}\right)^\top \in \mathbb{R}^k$ and $0 < \sigma^2 < \infty$ are unknown parameters.

1.2.2 Interpretation of regression coefficients

$$\boldsymbol{x} = \big(x_0, \ldots, x_j \ldots, x_{k-1}\big)^\top \in \mathcal{X},$$

$$\boldsymbol{x}^{j(+1)} := \big(x_0, \ldots, x_j + 1 \ldots, x_{k-1}\big)^\top \in \mathcal{X},$$

$$\boldsymbol{x}^{j(+\delta)} := \big(x_0, \ldots, x_j + \delta \ldots, x_{k-1}\big)^\top \in \mathcal{X}$$

1.2.3 Linear model with general data

$$\mathbb{X} = \begin{pmatrix} X_{1,0} & \dots & X_{1,k-1} \\ \vdots & \vdots & \vdots \\ X_{n,0} & \dots & X_{n,k-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix} = \left(\boldsymbol{X}^0, \ \dots, \ \boldsymbol{X}^{k-1}\right).$$

Lemma 1.1 Conditional mean and covariance matrix of the response vector.

Let the data $\left(Y_i, \boldsymbol{X}_i^\top\right)^\top \overset{\text{i.i.d.}}{\sim} \left(Y, \boldsymbol{X}^\top\right)^\top$, $i = 1, \dots, n$ *satisfy a linear model. Then*

$$\mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{X}\big) = \mathbb{X}\boldsymbol{\beta}, \qquad \operatorname{var}\big(\boldsymbol{Y} \,\big|\, \mathbb{X}\big) = \sigma^2 \mathbf{I}_n.$$

Definition 1.2 Linear model with general data.

The data $(\boldsymbol{Y}, \mathbb{X})$, satisfy a *linear model* if

$$\mathbb{E}(\boldsymbol{Y} \mid \mathbb{X}) = \mathbb{X}\boldsymbol{\beta}, \qquad \mathrm{var}(\boldsymbol{Y} \mid \mathbb{X}) = \sigma^2 \mathbf{I}_n,$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{k-1})^\top \in \mathbb{R}^k$ and $0 < \sigma^2 < \infty$ are unknown parameters.

1.2.4 Rank of the model

Assumptions

- $n > k$;
- $\mathsf{P}\big(\text{rank}(\mathbb{X}) = r\big) = 1$ for some $r \leq k$.

Definition 1.3 Full-rank linear model.

A *full-rank linear model* is such a linear model where $r = k$.

1.2.5 Error terms

$$\boldsymbol{\varepsilon} = \big(\varepsilon_1, \dots, \varepsilon_n\big)^\top = \big(Y_1 - \boldsymbol{X}_1^\top \boldsymbol{\beta}, \dots, Y_n - \boldsymbol{X}_n^\top \boldsymbol{\beta}\big)^\top = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}$$

Lemma 1.2 Moments of the error terms.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n\big)$. *Then*

$$\begin{aligned} \mathbb{E}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{X}\big) &= \mathbf{0}_n, & \mathbb{E}\big(\boldsymbol{\varepsilon}\big) &= \mathbf{0}_n, \\ \mathrm{var}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{X}\big) &= \sigma^2\, \mathbf{I}_n, & \mathrm{var}\big(\boldsymbol{\varepsilon}\big) &= \sigma^2\, \mathbf{I}_n. \end{aligned}$$

1.2.6 Distributional assumptions

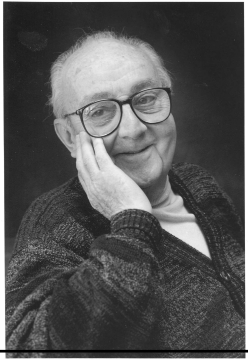
1.2.7 Fixed or random covariates

Essentially, all models are wrong, but some are useful. The practical question is how wrong do they have to be to not be useful.

George E. P. Box

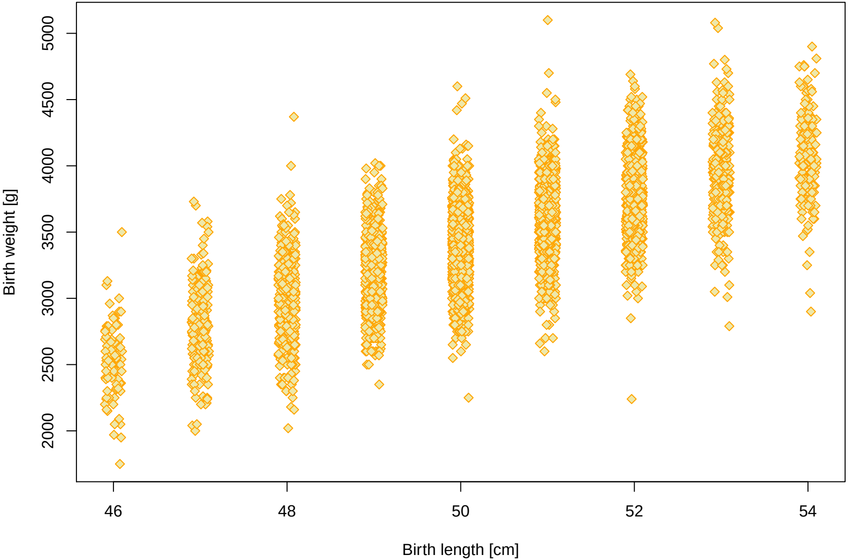
October 18, 1919 in Gravesend, Kent, England

– March 28, 2013 in Madison, Wisconsin, USA.



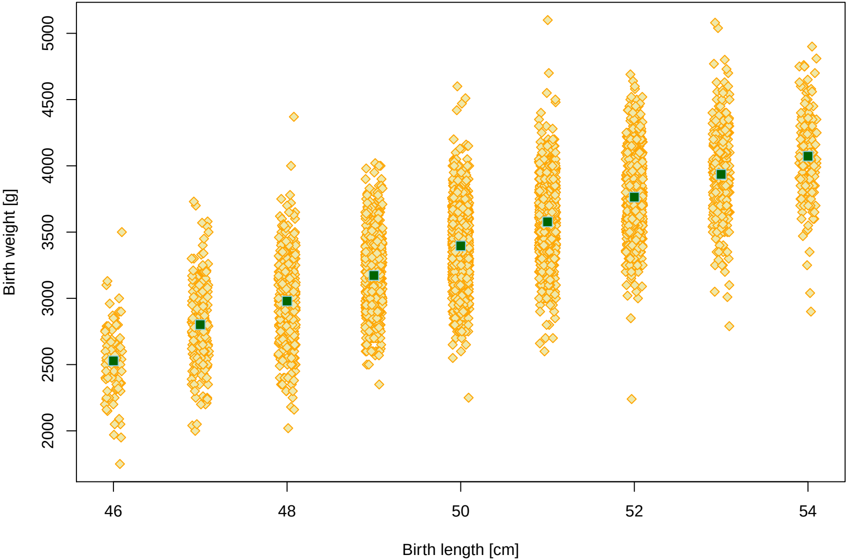
HosiO ($n = 4838$)

bweight ∼ blength



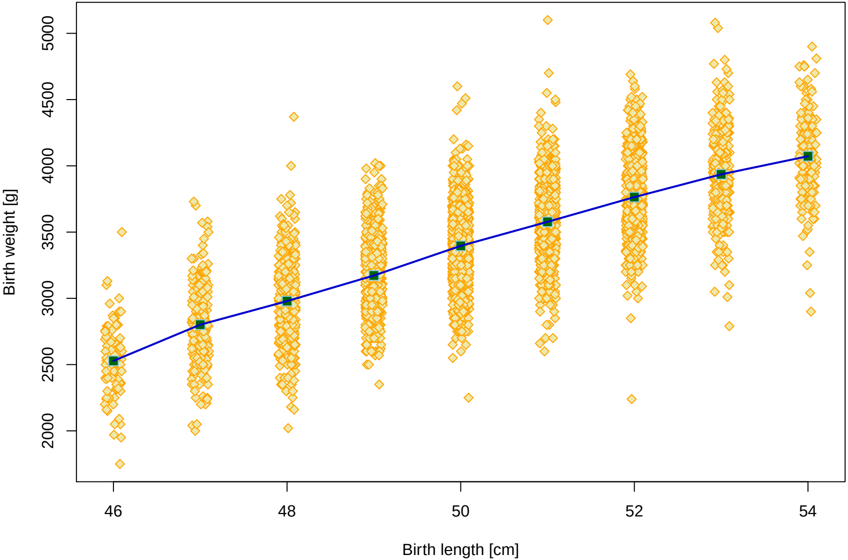
HosiO ($n = 4838$)

bweight ~ blength



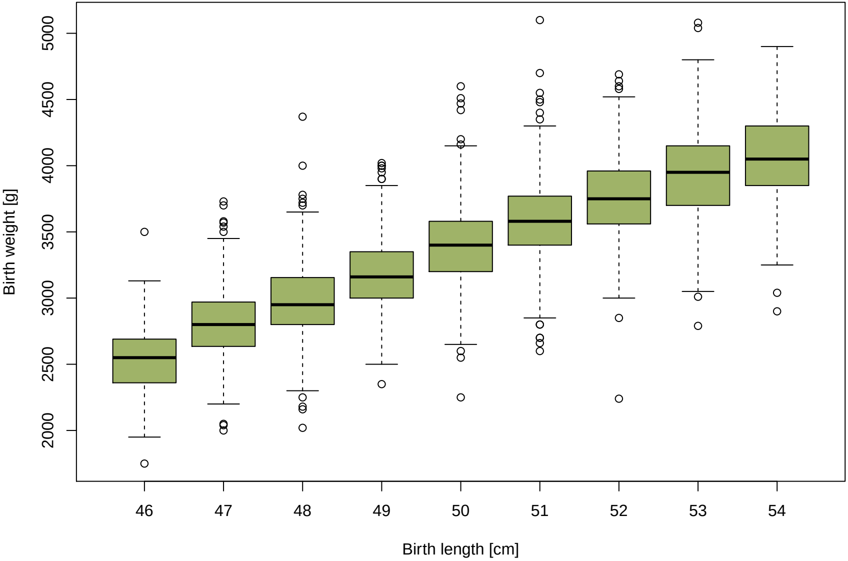
`Hosi0` ($n = 4838$)

`bweight` $\sim$ `blength`



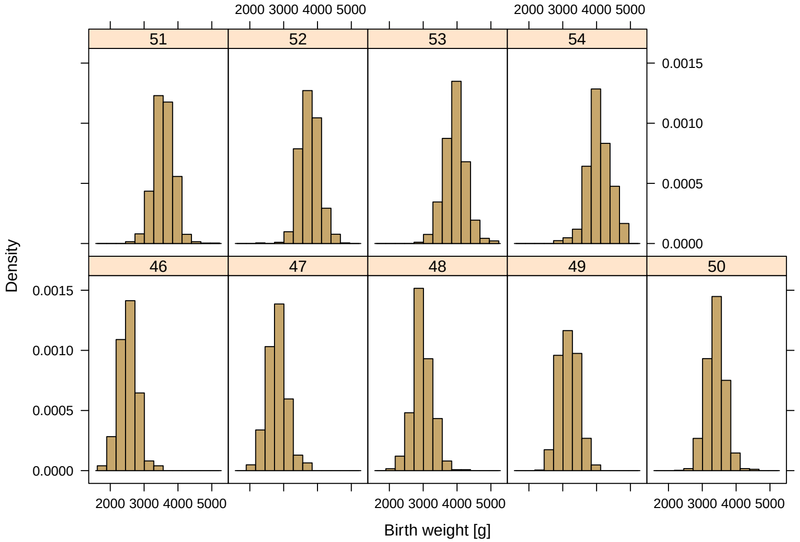
HosiO ($n = 4838$)

bweight ∼ blength



Hosi0 ($n = 4838$)

bweight ∼ blength



2

Least Squares Estimation

Section **2.1**

Sum of squares, least squares estimator and normal equations

2.1 Sum of squares, least squares estimator and normal equations

Definition 2.1 Sum of squares.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$. The function $\mathsf{SS}: \mathbb{R}^k \longrightarrow \mathbb{R}$ given as follows

$$\mathsf{SS}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2 = \big\|\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\big\|^2 = \big(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\big)^\top \big(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\big), \qquad \boldsymbol{\beta} \in \mathbb{R}^k$$

will be called the *sum of squares* of the model.

2.1 Sum of squares, least squares estimator and normal equations

Lemma 2.1 Least squares estimator.

Assume a full-rank linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. *There exist a unique minimizer to* $\text{SS}(\boldsymbol{\beta})$ *given as*

$$\widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top \mathbb{X}\big)^{-1} \mathbb{X}^\top \boldsymbol{Y}.$$

2.1 Sum of squares, least squares estimator and normal equations

Definition 2.2 Least squares estimator, normal equations.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. The quantity $\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top\boldsymbol{Y}$ will be called the *least squares estimator (LSE)* of the vector of regression coefficients $\boldsymbol{\beta}$. The linear system $\mathbb{X}^\top\mathbb{X}\,\boldsymbol{\beta} = \mathbb{X}^\top\boldsymbol{Y}$ will be called the system of *normal equations*.

2.1 Sum of squares, least squares estimator and normal equations

Lemma 2.2 Moments of the least squares estimator.

Let $\boldsymbol{Y} \,|\, \mathbb{X} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$. *Then*

$$\mathbb{E}\big(\widehat{\boldsymbol{\beta}} \,\big|\, \mathbb{X}\big) = \boldsymbol{\beta}, \qquad \mathbb{E}\big(\widehat{\boldsymbol{\beta}}\big) = \boldsymbol{\beta},$$

$$\mathrm{var}\big(\widehat{\boldsymbol{\beta}} \,\big|\, \mathbb{X}\big) = \sigma^2 \left(\mathbb{X}^\top \mathbb{X}\right)^{-1}.$$

Section **2.2**

Fitted values, residuals, projections

2.2 Fitted values, residuals, projections

Definition 2.3 Regression and residual space of a linear model.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim (\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, $\text{rank}(\mathbb{X}_{n\times k}) = r \leq k$. The *regression space* of the model is a vector space $\mathcal{M}(\mathbb{X})$. The *residual space* of the model is the orthogonal complement of the regression space, i.e., a vector space $\mathcal{M}(\mathbb{X})^{\perp}$.

2.2 Fitted values, residuals, projections

Definition 2.4 Fitted values, residuals.

Consider a full-rank linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. The vector

$$\widehat{\boldsymbol{Y}} := \mathbb{X}\widehat{\boldsymbol{\beta}} = \mathbb{X}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top\boldsymbol{Y}$$

will be called the vector of *fitted values* of the model. The vector

$$\boldsymbol{U} := \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$$

will be called the vector of *residuals* of the model.

2.2 Fitted values, residuals, projections

Notation. $\mathbb{H} := \mathbb{X}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top, \ \mathbb{M} := \mathbf{I}_n - \mathbb{H}$.

Lemma 2.3 Algebraic properties of fitted values, residuals and related projection matrices.

(i) $\widehat{\boldsymbol{Y}} = \mathbb{H}\boldsymbol{Y}$ *and* $\boldsymbol{U} = \mathbb{M}\boldsymbol{Y}$ *are projections of* $\boldsymbol{Y}$ *into* $\mathcal{M}(\mathbb{X})$ *and* $\mathcal{M}(\mathbb{X})^\perp$*, respectively;*

(ii) $\widehat{\boldsymbol{Y}} \perp \boldsymbol{U}$;

(iii) $\mathbb{H}$ *and* $\mathbb{M}$ *are projection matrices into* $\mathcal{M}(\mathbb{X})$ *and* $\mathcal{M}(\mathbb{X})^\perp$*, respectively;*

(iv) $\mathbb{H}^\top = \mathbb{H}, \ \mathbb{M}^\top = \mathbb{M}$;

(v) $\mathbb{H}\,\mathbb{H} = \mathbb{H}, \ \mathbb{M}\,\mathbb{M} = \mathbb{M}$;

(vi) $\mathbb{H}\,\mathbb{X} = \mathbb{X}, \ \mathbb{M}\,\mathbb{X} = \mathbf{0}_{n\times k}$.

2.2 Fitted values, residuals, projections

Terminology *(Hat matrix, residual projection matrix).*
For a linear model of (not necessarily full-rank)
$\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big),$ rank$(\mathbb{X}_{n\times k}) = r \leq k$.

- $\mathbb{H} = \mathbb{Q}\,\mathbb{Q}^\top = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-}\mathbb{X}^\top$: *hat matrix* ,
  where $\mathbb{Q}_{n\times r} = \big(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r\big)$ is an orthonormal vector basis of the regression space $\mathcal{M}\big(\mathbb{X}\big)$;
- $\mathbb{M} = \mathbb{N}\,\mathbb{N}^\top = \mathbf{I}_n - \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-}\mathbb{X}^\top$: *residual projection matrix* ,
  where $\mathbb{N}_{n\times r} = \big(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_{n-r}\big)$ is an orthonormal vector basis of the residual space $\mathcal{M}\big(\mathbb{X}\big)^\perp$.

Section **2.3**

Gauss-Markov theorem

2.3 Gauss-Markov theorem

Theorem 2.4 Gauss–Markov.

Assume a linear model $\boldsymbol{Y} \,|\, \mathbb{X} \sim (\mathbb{X}\beta, \sigma^2 \mathbf{I}_n)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = r \leq k$. *Then the vector of fitted values* $\widehat{\boldsymbol{Y}}$ *is, conditionally given* $\mathbb{X}$, *the best linear unbiased estimator (BLUE) of a vector parameter* $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{Y} \,|\, \mathbb{X})$. *Further,*

$$\mathrm{var}(\widehat{\boldsymbol{Y}} \,|\, \mathbb{X}) = \sigma^2 \,\mathbb{H} = \sigma^2 \,\mathbb{X}(\mathbb{X}^\top \mathbb{X})^- \mathbb{X}^\top.$$

2.3 Gauss-Markov theorem

Historical remarks

- The method of least squares was used in astronomy and geodesy already at the beginning of the 19th century.
- 1805: First documented publication of least squares.

  Adrien-Marie Legendre. Appendix "*Sur le méthode des moindres quarrés*" ("*On the method of least squares*") in the book *Nouvelles Méthodes Pour la Détermination des Orbites des Comètes* (*New Methods for the Determination of the Orbits of the Comets*).
- 1809: Another (supposedly independent) publication of least squares.

  Carl Friedrich Gauss. In Volume 2 of the book *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* (*The Theory of the Motion of Heavenly Bodies Moving Around the Sun in Conic Sections*).
  - C. F. Gauss claimed he had been using the method of least squares since 1795 (which is probably true).
- The Gauss–Markov theorem was first proved by C. F. Gauss in 1821 – 1823.
- In 1912, A. A. Markov provided another version of the proof.
- In 1934, J. Neyman described the Markov's proof as being "elegant" and stated that Markov's contribution (written in Russian) had been overlooked in the West.

  ⇒ The name Gauss–Markov theorem.

Theorem 2.5 Gauss–Markov for linear combinations.

Assume a full-rank linear model $\boldsymbol{Y} \mid \mathbb{X} \sim (\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. *Then*

(i) *For a vector* $\mathbf{l} = (l_0, \dots, l_{k-1})^\top \in \mathbb{R}^k$, $\mathbf{l} \neq \mathbf{0}$, *the statistic* $\widehat{\theta} = \mathbf{l}^\top \widehat{\boldsymbol{\beta}}$ *is the best linear unbiased estimator (BLUE) of the parameter* $\theta = \mathbf{l}^\top \boldsymbol{\beta}$ *with*

$$\text{var}(\widehat{\theta} \mid \mathbb{X}) = \sigma^2 \mathbf{l}^\top (\mathbb{X}^\top \mathbb{X})^{-1} \mathbf{l} > 0.$$

(ii) *For a given matrix*

$$\mathbb{L} = \begin{pmatrix} \mathbf{l}_1^\top \\ \vdots \\ \mathbf{l}_m^\top \end{pmatrix}, \quad \mathbf{l}_j \in \mathbb{R}^k,\ \mathbf{l}_j \neq \mathbf{0}, \quad j = 1, \dots, m, \quad m \leq k$$

with linearly independent rows ($\text{rank}(\mathbb{L}_{m\times k}) = m$), *the statistic* $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ *is the best linear unbiased estimator (BLUE) of the vector parameter* $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$ *with*

$$\text{var}(\widehat{\boldsymbol{\theta}} \mid \mathbb{X}) = \sigma^2 \mathbb{L} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{L}^\top,$$

which is a positive definite matrix.

Section **2.4**

Residuals, properties

2.4 Residuals, properties

Definition 2.5 Residual sum of squares.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\text{rank}(\mathbb{X}_{n\times k}) = r \leq k$. The quantity $\text{SS}_e = \big\|\boldsymbol{U}\big\|^2 = \sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n}\big(Y_i - \widehat{Y}_i\big)^2 = \big\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\big\|^2$ will be called the *residual sum of squares* of the model.

2.4 Residuals, properties

Lemma 2.6 Altenative expressions of residuals and residual sum of squares.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = r \leq k$. *The following then holds.*

(i) $\boldsymbol{U} = \mathbb{M}\boldsymbol{\varepsilon}$, *where* $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}$;

(ii) $\mathrm{SS}_e = \boldsymbol{Y}^\top\mathbb{M}\boldsymbol{Y} = \boldsymbol{\varepsilon}^\top\mathbb{M}\boldsymbol{\varepsilon}$.

2.4 Residuals, properties

Lemma 2.7 Moments of residuals and residual sum of squares.

Let $\boldsymbol{Y} \,|\, \mathbb{X} \sim \left(\mathbb{X}\beta,\, \sigma^2 \mathbf{I}_n\right)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = r \leq k$. *Then*

(i) $\mathbb{E}\left(\boldsymbol{U} \,\middle|\, \mathbb{X}\right) = \mathbf{0}_n$, $\quad \mathrm{var}\left(\boldsymbol{U} \,\middle|\, \mathbb{X}\right) = \sigma^2\, \mathbb{M}$;

(ii) $\mathbb{E}\left(\mathrm{SS}_e \,\middle|\, \mathbb{X}\right) = \mathbb{E}(\mathrm{SS}_e) = (n-r)\sigma^2$.

2.4 Residuals, properties

Definition 2.6 Residual mean square and residual degrees of freedom.

Consider a linear model $\boldsymbol{Y} \,|\, \mathbb{X} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\operatorname{rank}(\mathbb{X}_{n\times k}) = r \leq k$.

(i) The *residual mean square* of the model is the quantity $\mathsf{SS}_e/(n-r)$ and will be denoted as MS_e. That is,

$$\mathsf{MS}_e = \frac{\mathsf{SS}_e}{n-r}.$$

(ii) The *residual degrees of freedom* of the model is the vector dimension of the residual space $\mathcal{M}(\mathbb{X})^{\perp}$ and will be denotes as ν_e. That is,

$$\nu_e = n - r.$$

Section **2.5**

Parameterizations of a linear model

2.5 Parameterizations of a linear model

Definition 2.7 Equivalent linear models.

Assume two linear models: M_1: $\boldsymbol{Y} \,\big|\, \mathbb{X}_1 \sim \big(\mathbb{X}_1\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big)$, where $\mathbb{X}_1$ is an $n \times k$ matrix with $\mathrm{rank}\big(\mathbb{X}_1\big) = r$ and M_2: $\boldsymbol{Y} \,\big|\, \mathbb{X}_2 \sim \big(\mathbb{X}_2\boldsymbol{\gamma},\, \sigma^2\mathbf{I}_n\big)$, where $\mathbb{X}_2$ is an $n \times l$ matrix with $\mathrm{rank}\big(\mathbb{X}_2\big) = r$. We say that models M_1 and M_2 are *equivalent* if their regression spaces are the same. That is, if

$$\mathcal{M}\big(\mathbb{X}_1\big) = \mathcal{M}\big(\mathbb{X}_2\big).$$

Section **2.6**

Matrix algebra and a method of least squares

2.6 Matrix algebra and a method of least squares

- Quantities to calculate for the LSE in a full-rank model ($\mathrm{rank}(\mathbb{X}_{n\times k}) = k$):

$$\mathbb{H} = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top, \qquad \mathbb{M} = \mathbf{I}_n - \mathbb{H} = \mathbf{I}_n - \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top,$$

$$\widehat{\boldsymbol{Y}} = \mathbb{H}\boldsymbol{Y} = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}, \quad \mathrm{var}\big(\widehat{\boldsymbol{Y}}\,\big|\,\mathbb{X}\big) = \sigma^2\mathbb{H} = \sigma^2\,\mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top,$$

$$\boldsymbol{U} = \mathbb{M}\boldsymbol{Y} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}, \qquad \mathrm{var}\big(\boldsymbol{U}\,\big|\,\mathbb{X}\big) = \sigma^2\mathbb{M} = \sigma^2\Big\{\mathbf{I}_n - \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\Big\},$$

$$\widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}, \qquad \mathrm{var}\big(\widehat{\boldsymbol{\beta}}\,\big|\,\mathbb{X}\big) = \sigma^2\,\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}.$$

2.6.1 QR decomposition

2.6.2 SVD decomposition

See the *Fundamentals of Numerical Mathematics (NMNM201)* course.

3

Basic Regression Diagnostics

Section **3.1**

(Normal) linear model assumptions

3.1 (Normal) linear model assumptions

1. $\mathbb{E}\big(Y_i \,\big|\, \boldsymbol{X}_i = \boldsymbol{x}\big) = \boldsymbol{x}^\top \boldsymbol{\beta}$ for some $\boldsymbol{\beta} \in \mathbb{R}^k$ and (almost all) $\boldsymbol{x} \in \mathcal{X}$.
   $\equiv$ Correct regression function

2. $\mathrm{var}\big(Y_i \,\big|\, \boldsymbol{X}_i = \boldsymbol{x}\big) = \sigma^2$ for some σ^2 irrespective of (almost all) values of $\boldsymbol{x} \in \mathcal{X}$.
   $\equiv$ *homoscedasticity*

3. $\mathrm{cov}\big(Y_i,\; Y_l \,\big|\, \mathbb{X} = \mathbf{x}\big) = 0$, $i \neq l$, for (almost all) $\mathbf{x} \in \mathcal{X}^n$.
   $\equiv$ The responses are conditionally uncorrelated.

4. $Y_i \,|\, \boldsymbol{X}_i = \boldsymbol{x} \sim \mathcal{N}\big(\boldsymbol{x}^\top \boldsymbol{\beta},\, \sigma^2\big)$, for (almost all) $\boldsymbol{x} \in \mathcal{X}$.
   $\equiv$ Normality

3.1 (Normal) linear model assumptions

Assumptions in terms of the errors ε:

1. $\mathbb{E}\big(\varepsilon_i \,\big|\, \boldsymbol{X}_i = \boldsymbol{x}\big) = 0$ for (almost all) $\boldsymbol{x} \in \mathcal{X}$,
   and consequently also $\mathbb{E}\big(\varepsilon_i\big) = 0$, $i = 1, \ldots, n$.
   $\equiv$ the regression function of the model is correctly specified.

2. $\mathsf{var}\big(\varepsilon_i \,\big|\, \boldsymbol{X}_i = \boldsymbol{x}\big) = \sigma^2$ for some σ^2 which is constant irrespective of (almost all) values of $\boldsymbol{x} \in \mathcal{X}$.

   Consequently also $\mathsf{var}\big(\varepsilon_i\big) = \sigma^2$, $i = 1, \ldots, n$.
   $\equiv$ *homoscedasticity* of the errors.

3. $\mathsf{cov}\big(\varepsilon_i,\, \varepsilon_l \,\big|\, \mathbb{X} = \mathbf{x}\big) = 0$, $i \neq l$, for (almost all) $\mathbf{x} \in \mathcal{X}^n$. Consequently also $\mathsf{cov}\big(\varepsilon_i,\, \varepsilon_l\big) = 0$, $i \neq l$.
   $\equiv$ The errors are uncorrelated.

4. $\varepsilon_i \,\big|\, \boldsymbol{X}_i = \boldsymbol{x} \sim \mathcal{N}\big(0,\, \sigma^2\big)$ for (almost all) $\boldsymbol{x} \in \mathcal{X}$ and consequently also $\varepsilon_i \sim \mathcal{N}\big(0,\, \sigma^2\big)$, $i = 1, \ldots, n$.
   $\equiv$ The errors are normally distributed and owing to previous assumptions,
   $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, \sigma^2)$.

3.1 (Normal) linear model assumptions

Assumptions and residual properties

1. (A1) $\Longrightarrow \mathbb{E}\big(\boldsymbol{U} \,\big|\, \mathbb{X}\big) = \mathbf{0}_n$.
2. (A1) & (A2) & (A3) $\Longrightarrow \mathrm{var}\big(\boldsymbol{U} \,\big|\, \mathbb{X}\big) = \sigma^2\, \mathbb{M}$.
3. (A1) & (A2) & (A3) & (A4) $\Longrightarrow \boldsymbol{U} \,|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\, \mathbb{M}\big)$.

Section **3.2**

Standardized residuals

3.2 Standardized residuals

Definition 3.1 Standardized residuals.

The *standardized residuals* or the vector of standardized residuals of the model is the vector $\boldsymbol{U}^{std} = \big(U_1^{std}, \dots, U_n^{std}\big)$, where

$$U_i^{std} = \begin{cases} \dfrac{U_i}{\sqrt{\mathrm{MS}_e\, m_{i,i}}}, & m_{i,i} > 0, \\ \text{undefined}, & m_{i,i} = 0, \end{cases} \qquad i = 1, \dots, n.$$

3.2 Standardized residuals

Lemma 3.1 Moments of standardized residuals under normality.

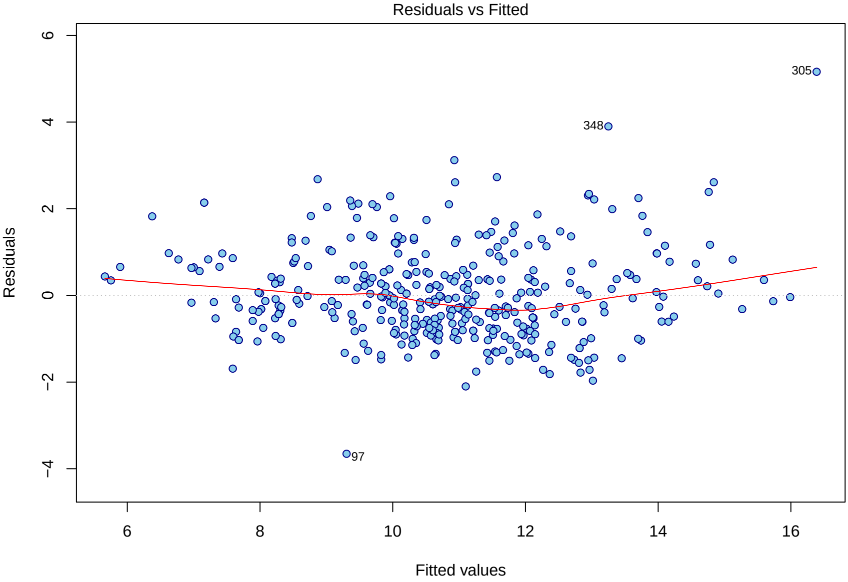
Let $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$ *and let for chosen* $i \in \{1, \dots, n\}$, $m_{i,i} > 0$. *Then*

$$\mathbb{E}\big(U_i^{std} \,\big|\, \mathbb{X}\big) = 0, \qquad \mathrm{var}\big(U_i^{std} \,\big|\, \mathbb{X}\big) = 1.$$

Section **3.3**

Graphical tools of regression diagnostics

3.3.1 (A1) Correctness of the regression function



3.3.1 (A1) Correctness of the regression function

Overall inappropriateness of the regression function

➠ scatterplot $(\widehat{\boldsymbol{Y}}, \boldsymbol{U})$ of residuals versus fitted values.

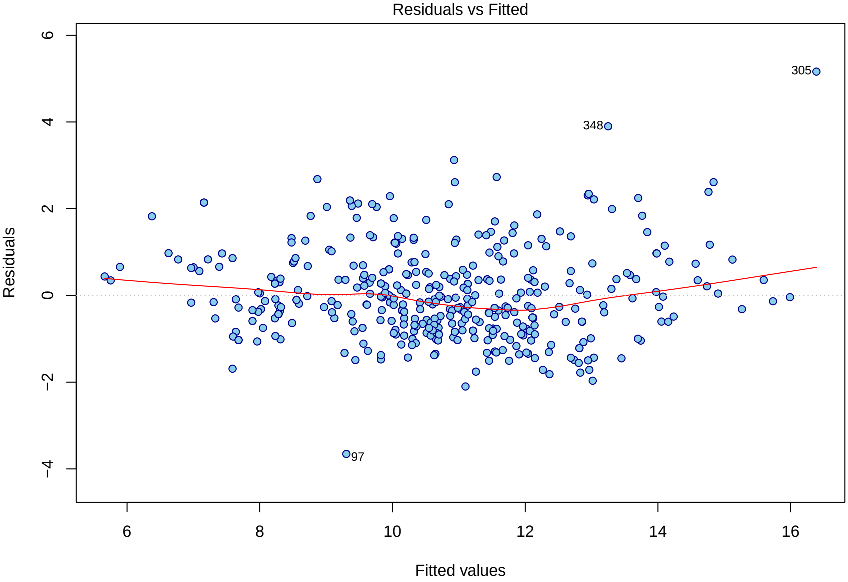
Nonlinearity of the regression function with respect to a particular regressor $\boldsymbol{X}^j$

➠ scatterplot $(\boldsymbol{X}^j, \boldsymbol{U})$ of residuals versus that regressor.

Possibly omitted regressor $\boldsymbol{V}$

➠ scatterplot $(\boldsymbol{V}, \boldsymbol{U})$ of residuals versus that regressor.

3.3.2 (A2) Homoscedasticity of the errors



3.3.2 (A2) Homoscedasticity of the errors

Residual variance that depends on the response expectation

➠ scatterplot $(\widehat{\boldsymbol{Y}}, \boldsymbol{U})$ of residuals versus fitted values.

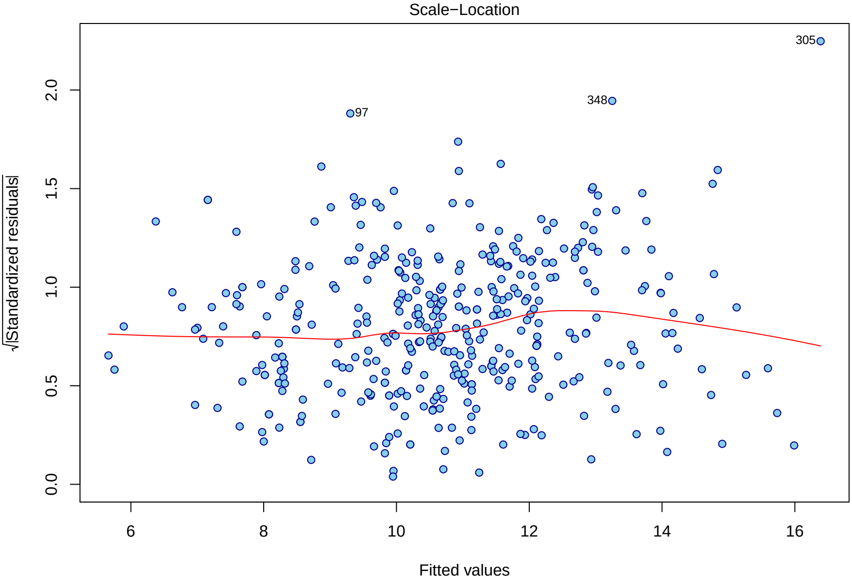
Residual variance that depends on a particular regressor $\boldsymbol{X}^j$

➠ scatterplot $(\boldsymbol{X}^j, \boldsymbol{U})$ of residuals versus that regressor.

Residual variance that depends on a regressor $\boldsymbol{V}$ not included in the model

➠ scatterplot $(\boldsymbol{V}, \boldsymbol{U})$ of residuals versus that regressor.

3.3.2 (A2) Homoscedasticity of the errors



3.3.3 (A3) Uncorrelated errors

To consider possibly correlated errors

(i) repeated observations performed on N independently behaving units/subjects;

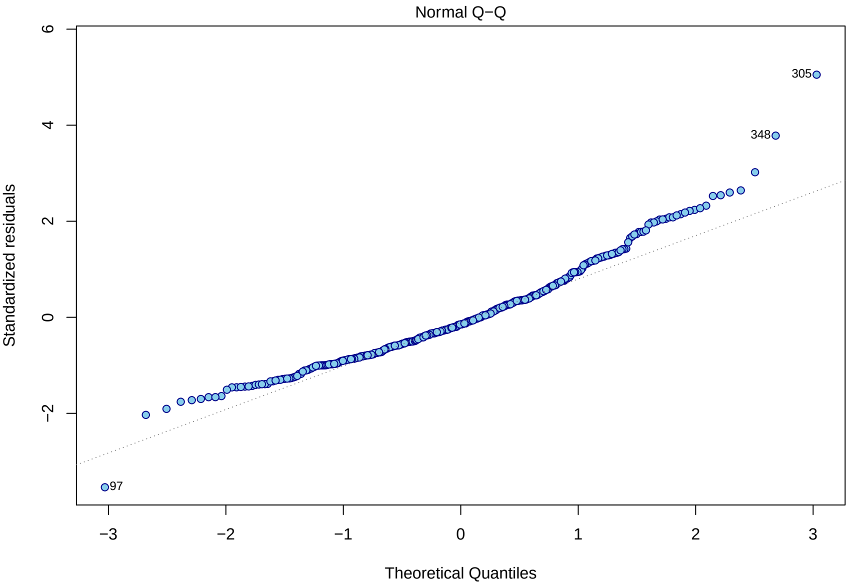
(ii) observations performed sequentially in time where the ith response value Y_i is obtained in time t_i and the observational occasions $t_1 < \cdots < t_n$ form an increasing (often equidistant) sequence.

3.3.3 (A3) Uncorrelated errors

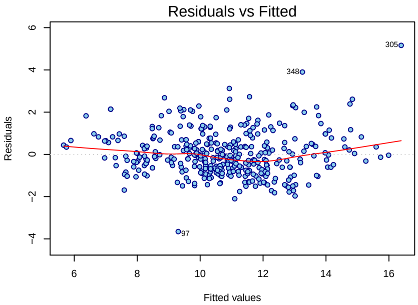
Detection of serial correlation in errors

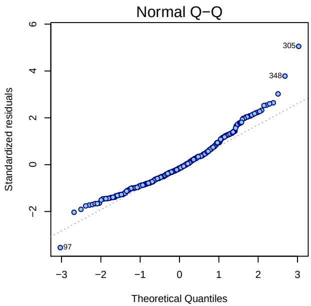
- Autocorrelation and partial autocorrelation plot based on residuals $\boldsymbol{U}$.
- Plot of delayed residuals, that is a scatterplot based on points (U_1, U_2), (U_2, U_3), ..., (U_{n-1}, U_n).

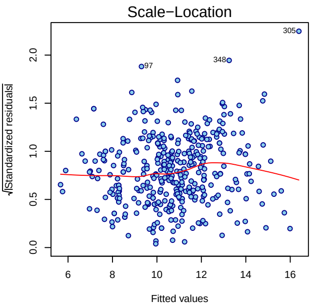
3.3.4 (A4) Normality



3.3.5 The three basic diagnostic plots





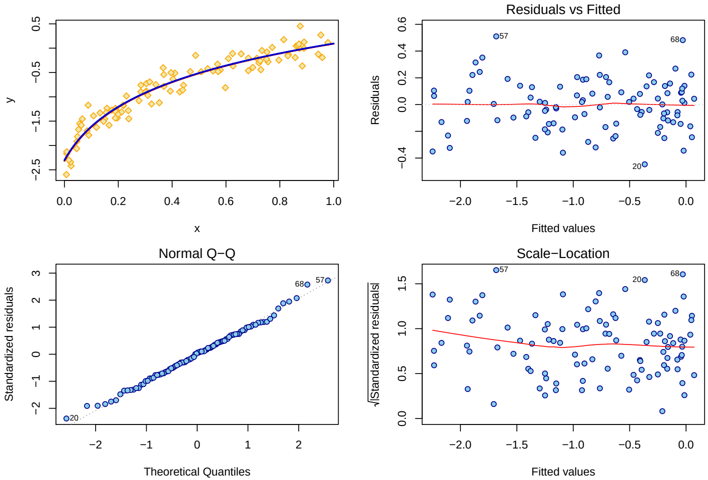


3.3.5 The three basic diagnostic plots

Correct model

True: $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, 0.2^2).$

Model: $Y = \beta_0 + \beta_1\, \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma^2).$

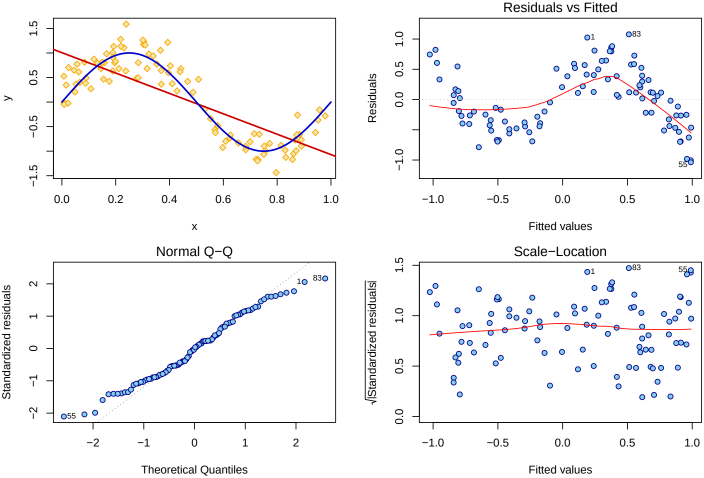


3.3.5 The three basic diagnostic plots

Incorrect regression function

True: $Y = \sin(2\pi\, x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, 0.3^2).$

Model: $Y = \beta_0 + \beta_1\, x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma^2).$

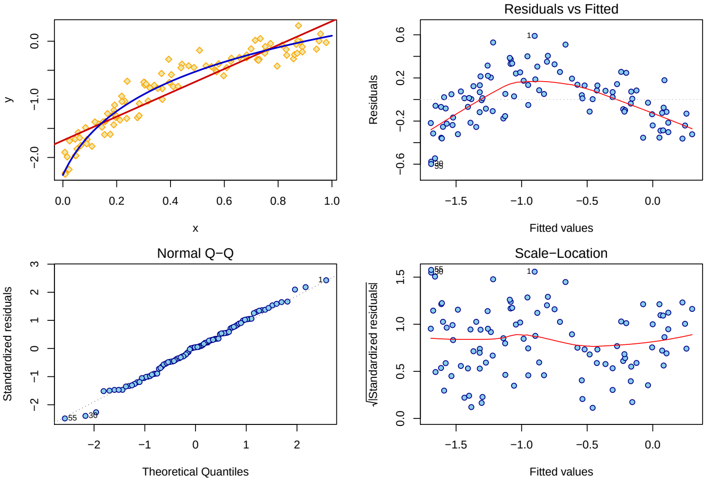


3.3.5 The three basic diagnostic plots

Incorrect regression function

True: $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, 0.2^2).$

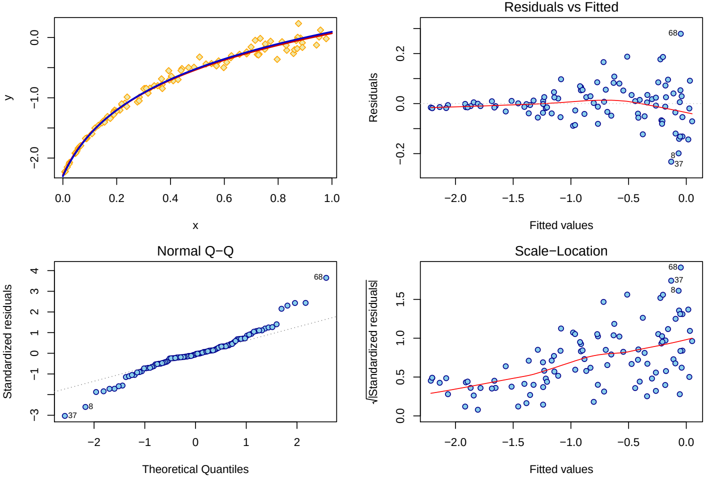
Model: $Y = \beta_0 + \beta_1\, x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma^2).$



3.3.5 The three basic diagnostic plots

Heteroscedasticity

True: $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, (0.2\,x)^2).$
Model: $Y = \beta_0 + \beta_1 \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2).$

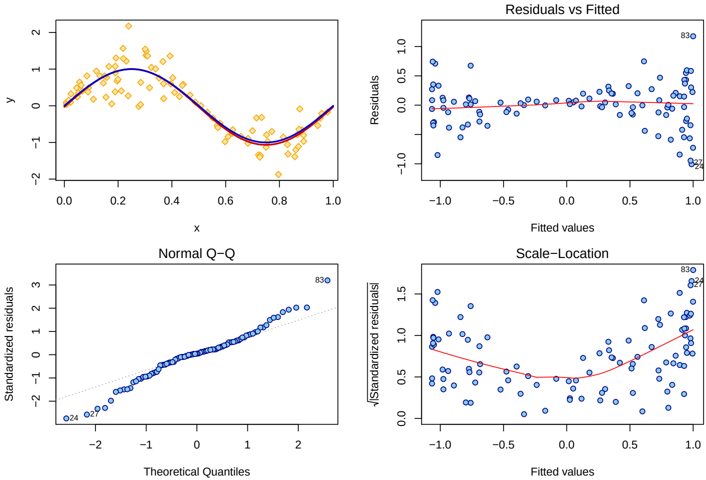


3.3.5 The three basic diagnostic plots

Heteroscedasticity

True: $Y = \sin(2\pi\, x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \{0.6\, \sin(2\pi\, x)\}^2).$

Model: $Y = \beta_0 + \beta_1\, \sin(2\pi\, x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma^2).$

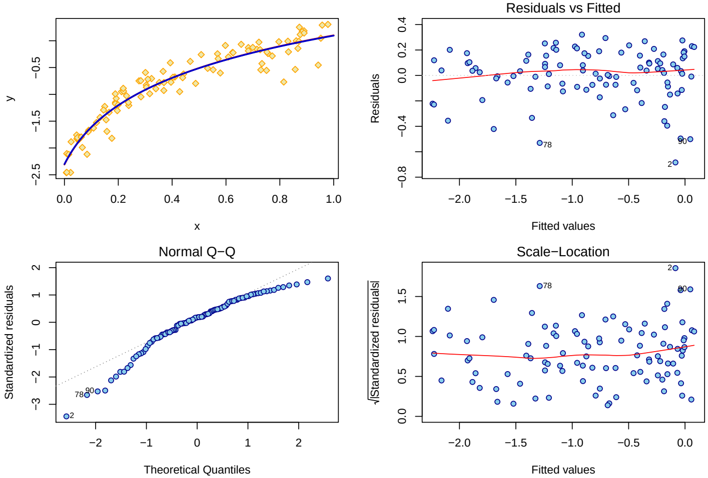


3.3.5 The three basic diagnostic plots

Nonnormal errors

True: $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim$ Gumbel.

Model: $Y = \beta_0 + \beta_1 \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$.



4

Parameterizations of Covariates

Section **4.1**

Linearization of the dependence of the response on the covariates

4.1 Linearization of the dependence

Data

$$\big(Y_i,\, \boldsymbol{Z}_i^\top\big)^\top, \qquad \boldsymbol{Z}_i = \big(Z_{i,1},\, \dots,\, Z_{i,p}\big)^\top \in \mathcal{Z} \subseteq \mathbb{R}^p,\ i = 1, \dots, n$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbb{Z} = \begin{pmatrix} \boldsymbol{Z}_1^\top \\ \vdots \\ \boldsymbol{Z}_n^\top \end{pmatrix}$$

Model

$$\mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) = \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{X}\big) = \mathbb{X}\beta,$$

$$\mathbb{X} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{t}^\top(\boldsymbol{Z}_1) \\ \vdots \\ \boldsymbol{t}^\top(\boldsymbol{Z}_n) \end{pmatrix}$$

4.1 Linearization of the dependence

Problem

Choice of $\boldsymbol{t} : \mathcal{Z} \longrightarrow \mathcal{X} \subseteq \mathbb{R}^k$,

$$\boldsymbol{t}(\boldsymbol{z}) = \big(t_0(\boldsymbol{z}), \ldots, t_{k-1}(\boldsymbol{z})\big)^\top = \big(x_0, \ldots, x_{k-1}\big)^\top = \boldsymbol{x}$$

such that

$$\begin{aligned}\mathbb{E}\big(Y \,\big|\, \boldsymbol{Z} = \boldsymbol{z}\big) &= \boldsymbol{t}^\top(\boldsymbol{z})\boldsymbol{\beta} \\ &= \beta_0\, t_0(\boldsymbol{z}) + \cdots + \beta_{k-1}\, t_{k-1}(\boldsymbol{z}) =: m(\boldsymbol{z}), \qquad \boldsymbol{z} \in \mathcal{Z}\end{aligned}$$

Section **4.2**

Parameterization of a single covariate

Definition 4.1 Parameterization of a covariate.

Let $Z_1, \ldots, Z_n$ be values of a given univariate covariate $Z \in \mathcal{Z} \subseteq \mathbb{R}$. By a *parameterization* of this covariate we mean

(i) the function $\boldsymbol{s}: \mathcal{Z} \longrightarrow \mathbb{R}^{k-1}$, $\boldsymbol{s}(z) = \bigl(s_1(z), \ldots, s_{k-1}(z)\bigr)^\top$, $z \in \mathcal{Z}$, where all $s_1, \ldots, s_{k-1}$ are non-constant functions on $\mathcal{Z}$, and

(ii) an $n \times (k-1)$ matrix $\mathbb{S}$, where

$$\mathbb{S} = \begin{pmatrix} \boldsymbol{s}^\top(Z_1) \\ \vdots \\ \boldsymbol{s}^\top(Z_n) \end{pmatrix} = \begin{pmatrix} s_1(Z_1) & \ldots & s_{k-1}(Z_1) \\ \vdots & \vdots & \vdots \\ s_1(Z_n) & \ldots & s_{k-1}(Z_n) \end{pmatrix}.$$

4.2.2 Covariate types

Numeric covariates

Covariates where a ratio of the two covariate values makes sense and a unity increase of the covariate value has an unambiguous meaning.

(i) continuous: $\mathcal{Z} \equiv$ mostly an interval in $\mathbb{R}$;

(ii) discrete: $\mathcal{Z} \equiv$ infinite countable or finite (but “large”) subset of $\mathbb{R}$.

4.2.2 Covariate types

Categorical covariates

Covariates where the ratio of the two covariate values does not necessarily make sense and a unity increase of the covariate value does not necessarily have an unambiguous meaning.

$\mathcal{Z} \equiv$ a finite (and mostly "small") set, i.e.,

$$\mathcal{Z} = \{\omega_1, \ldots, \omega_G\}.$$

$\omega_1 < \cdots < \omega_G$: somehow arbitrarily chosen labels of categories.

1. nominal: from a practical point of view, chosen values $\omega_1, \ldots, \omega_G$ are completely arbitrary.

2. ordinal: *ordering* $\omega_1 < \cdots < \omega_G$ makes sense also from a practical point of view.

Cars2004nh ($n = 425$)

```
data(Cars2004nh, package = "mffSM")
head(Cars2004nh)
```

```
                        vname type drive price.retail price.dealer   price
1           Chevrolet.Aveo.4dr    1     1        11690        10965 11327.5
2 Chevrolet.Aveo.LS.4dr.hatch    1     1        12585        11802 12193.5
3       Chevrolet.Cavalier.2dr    1     1        14610        13697 14153.5
4       Chevrolet.Cavalier.4dr    1     1        14810        13884 14347.0
5    Chevrolet.Cavalier.LS.2dr    1     1        16385        15357 15871.0
6            Dodge.Neon.SE.4dr    1     1        13670        12849 13259.5

  cons.city cons.highway consumption engine.size ncylinder horsepower
1       8.4          6.9        7.65         1.6         4        103
2       8.4          6.9        7.65         1.6         4        103
3       9.0          6.4        7.70         2.2         4        140
4       9.0          6.4        7.70         2.2         4        140
5       9.0          6.4        7.70         2.2         4        140
6       8.1          6.5        7.30         2.0         4        132

  weight      iweight  lweight wheel.base length width    ftype fdrive
1   1075 0.0009302326 6.980076        249    424   168 personal  front
2   1065 0.0009389671 6.970730        249    389   168 personal  front
3   1187 0.0008424600 7.079184        264    465   175 personal  front
4   1214 0.0008237232 7.101676        264    465   173 personal  front
5   1187 0.0008424600 7.079184        264    465   175 personal  front
6   1171 0.0008539710 7.065613        267    442   170 personal  front
```

`Cars2004nh` ($n = 425$)

```
summary(subset(Cars2004nh,
   select = c("price.retail", "price.dealer", "price", "cons.city", "cons.highway",
              "consumption", "engine.size", "horsepower", "weight",
              "wheel.base", "length", "width")))
```

```
  price.retail     price.dealer       price          cons.city
 Min.   : 10280   Min.   :  9875   Min.   : 10078   Min.   : 6.20
 1st Qu.: 20370   1st Qu.: 18973   1st Qu.: 19600   1st Qu.:11.20
 Median : 27905   Median : 25672   Median : 26656   Median :12.40
 Mean   : 32866   Mean   : 30096   Mean   : 31481   Mean   :12.36
 3rd Qu.: 39235   3rd Qu.: 35777   3rd Qu.: 37514   3rd Qu.:13.80
 Max.   :192465   Max.   :173560   Max.   :183012   Max.   :23.50
                                                    NA's   :14

  cons.highway     consumption     engine.size      horsepower
 Min.   : 5.100   Min.   : 5.65   Min.   :1.300   Min.   :100.0
 1st Qu.: 8.100   1st Qu.: 9.65   1st Qu.:2.400   1st Qu.:165.0
 Median : 9.000   Median :10.70   Median :3.000   Median :210.0
 Mean   : 9.142   Mean   :10.75   Mean   :3.208   Mean   :216.8
 3rd Qu.: 9.800   3rd Qu.:11.65   3rd Qu.:3.900   3rd Qu.:255.0
 Max.   :19.600   Max.   :21.55   Max.   :8.300   Max.   :500.0
 NA's   :14       NA's   :14

     weight        wheel.base        length          width
 Min.   : 923   Min.   :226.0   Min.   :363.0   Min.   :163.0
 1st Qu.:1412   1st Qu.:262.0   1st Qu.:450.0   1st Qu.:175.0
 Median :1577   Median :272.0   Median :472.0   Median :180.0
 Mean   :1626   Mean   :274.9   Mean   :470.6   Mean   :181.1
 3rd Qu.:1804   3rd Qu.:284.0   3rd Qu.:490.0   3rd Qu.:185.0
 Max.   :3261   Max.   :366.0   Max.   :577.0   Max.   :206.0
 NA's   :2      NA's   :2       NA's   :26      NA's   :28
```

`Cars2004nh` ($n = 425$)

```
summary(subset(Cars2004nh, select = c("type", "drive")))
```

```
      type            drive      
 Min.   :1.000   Min.   :1.000  
 1st Qu.:1.000   1st Qu.:1.000  
 Median :1.000   Median :1.000  
 Mean   :2.219   Mean   :1.692  
 3rd Qu.:3.000   3rd Qu.:2.000  
 Max.   :6.000   Max.   :3.000  
```

```
table(Cars2004nh[, "type"], useNA = "ifany")
```

```
  1   2   3   4   5   6 
242  30  60  24  49  20 
```

```
table(Cars2004nh[, "drive"], useNA = "ifany")
```

```
  1   2   3 
223 110  92 
```

Cars2004nh ($n = 425$)

```r
summary(subset(Cars2004nh, select = c("ftype", "fdrive")))
```

```
     ftype         fdrive   
 personal:242   front:223  
 wagon   : 30   rear :110  
 SUV     : 60   4x4  : 92  
 pickup  : 24              
 sport   : 49              
 minivan : 20              
```

Cars2004nh ($n = 425$)

```
summary(subset(Cars2004nh, select = "ncylinder"))
```

```
   ncylinder
 Min.   :-1.000
 1st Qu.: 4.000
 Median : 6.000
 Mean   : 5.791
 3rd Qu.: 6.000
 Max.   :12.000
```

```
table(Cars2004nh[, "ncylinder"], useNA = "ifany")
```

```
 -1   4   5   6   8  10  12
  2 134   7 190  87   2   3
```

Section **4.3**

Numeric covariate

4.3.1 Simple transformation of the covariate

Regression function

$$m(z) = \beta_0 + \beta_1\, s(z), \quad z \in \mathcal{Z},$$

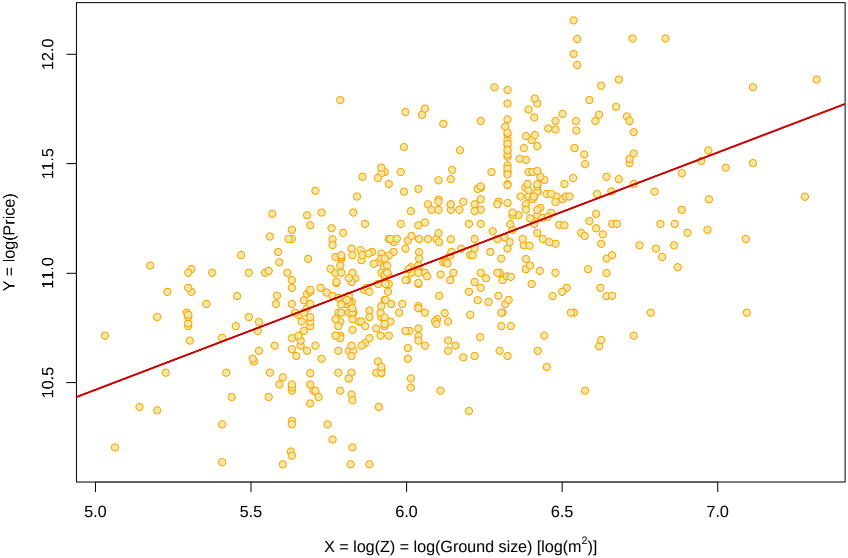
$s : \mathcal{Z} \longrightarrow \mathbb{R}$, a suitable *non-constant* function.

Reparameterizing matrix

$$\mathbb{S} = \begin{pmatrix} s(Z_1) \\ \vdots \\ s(Z_n) \end{pmatrix}.$$

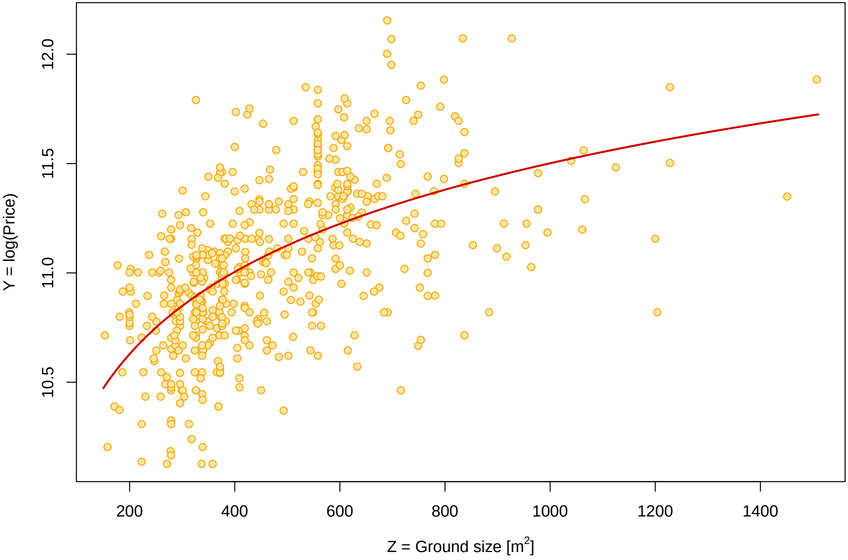
`Houses1987` ($n = 546$)

`log(price)` $\sim$ `log(ground)`, $\quad \widehat{m}(z) = 7.76 + 0.54 \log(z)$



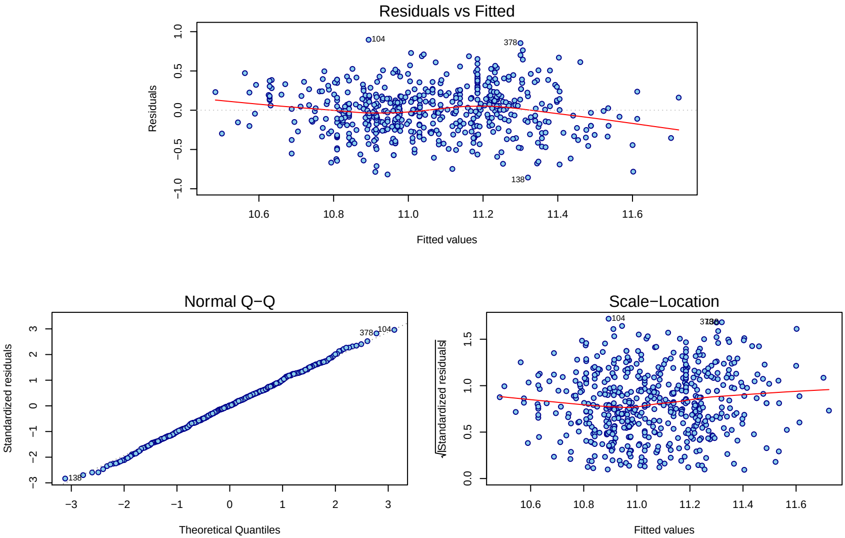
Houses1987 ($n = 546$)

log(price) $\sim$ log(ground), $\quad \widehat{m}(z) = 7.76 + 0.54\ \log(z)$



Houses1987 ($n = 546$)

log(price) ∼ log(ground), residual plots



4.3.1 Simple transformation of the covariate

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 \,+\, \beta_1 s(z), \qquad z \in \mathcal{Z}$$

Evaluation of the effect of the original covariate

$$\mathrm{H}_0: \;\; \beta_1 = 0$$

➠ t-test on regression coefficient (under normality)

Interpretation of the regression coefficients

$$\beta_1 \;=\; \mathbb{E}\big(Y \,\big|\, X = s(z)+1\big) \,-\, \mathbb{E}\big(Y \,\big|\, X = s(z)\big),$$

$$\mathbb{E}\big(Y \,\big|\, Z = z+1\big) \,-\, \mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; \beta_1\big\{s(z+1) - s(z)\big\}, \qquad z \in \mathcal{Z}$$

Houses1987 ($n = 546$)

Effect of the covariate, **interpretation** of the regression coefficients

```
summary(lm(log(price) ~ log(ground), data = Houses1987))
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-0.8571 -0.1988  0.0046  0.1929  0.8969

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.75625    0.19933   38.91   <2e-16 ***
log(ground)  0.54216    0.03265   16.61   <2e-16 ***
---

Residual standard error: 0.3033 on 544 degrees of freedom
Multiple R-squared:  0.3364,        Adjusted R-squared:  0.3351
F-statistic: 275.7 on 1 and 544 DF,  p-value: < 2.2e-16
```

4.3.2 Raw polynomials

Regression function

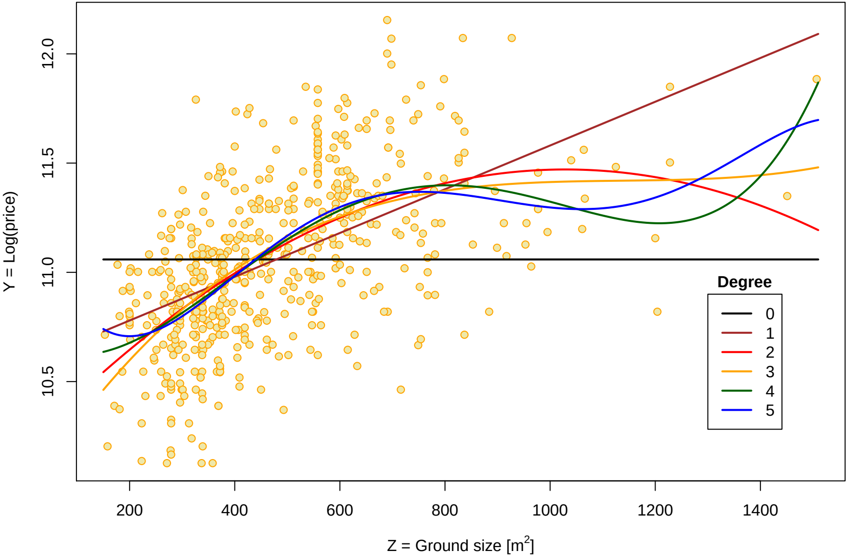
$$m(z) = \beta_0 + \beta_1 z + \cdots + \beta_{k-1} z^{k-1}, \quad z \in \mathcal{Z}.$$

Reparameterizing matrix

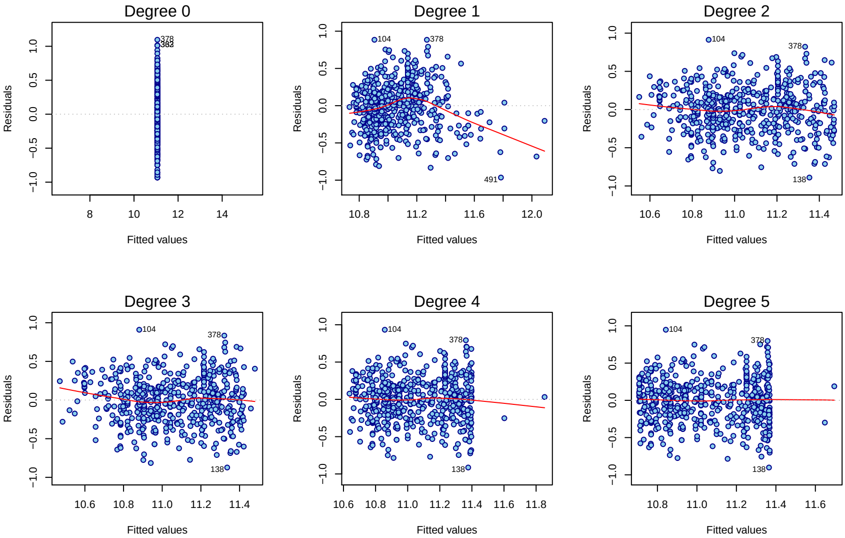
$$\mathbb{S} = \begin{pmatrix} Z_1 & \ldots & Z_1^{k-1} \\ \vdots & \vdots & \vdots \\ Z_n & \ldots & Z_n^{k-1} \end{pmatrix}.$$

Houses1987 ($n = 546$)

log(price) ∼ rawpoly(ground, d)



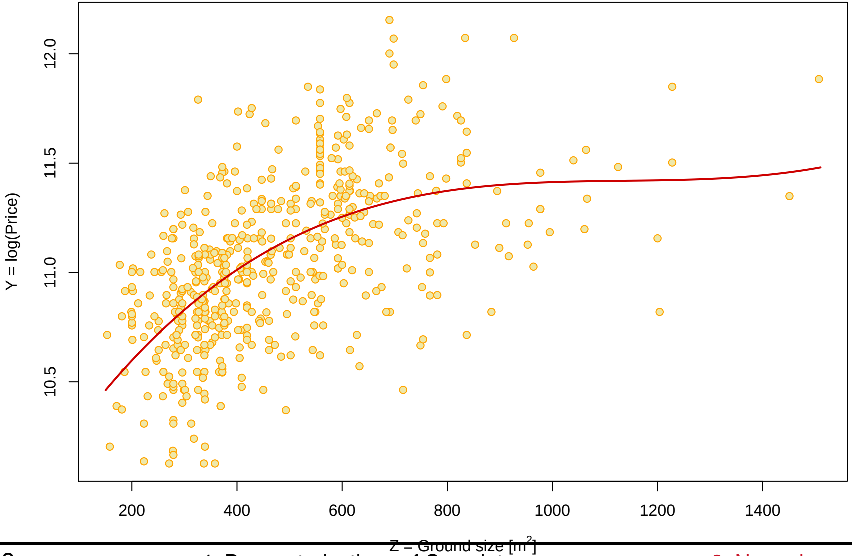
`log(price) ~ rawpoly(ground, d)`, residuals vs. fitted plots



Houses1987 ($n = 546$)

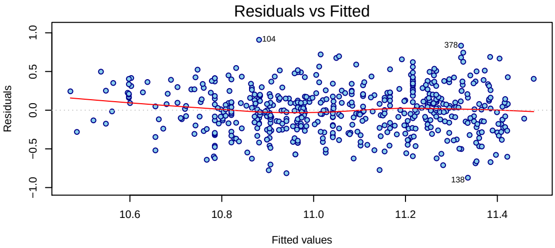
`log(price) ~ rawpoly(ground, 3),`

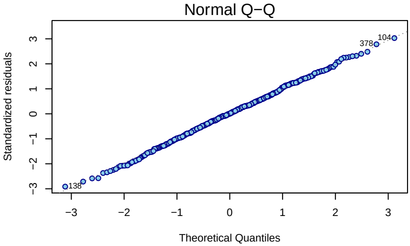
$$\widehat{m}(z) = 9.97 + 3.78 \cdot 10^{-3}\,z - 3.31 \cdot 10^{-6}\,z^2 + 9.70 \cdot 10^{-10}\,z^3$$

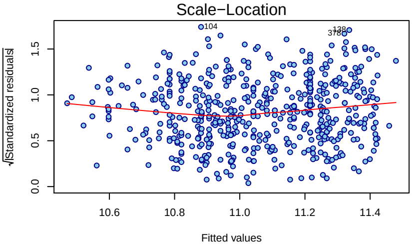


Houses1987 ($n = 546$)

log(price) ~ rawpoly(ground, 3), residual plots







4.3.2 Raw polynomials

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 + \beta_1 z + \ldots + \beta_{k-1} z^{k-1}, \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \big(\beta_1, \ldots, \beta_{k-1}\big)^\top$$

Evaluation of the effect of the original covariate

$$\mathrm{H}_0:\; \boldsymbol{\beta}^Z = \mathbf{0}_{k-1}$$

➠ Wald type test (F-test) on a subvector of regression coefficients (under normality)

≡ submodel F-test (under normality)

4.3.2 Raw polynomials

Regression function

$$\mathbb{E}\bigl(Y \,\big|\, Z = z\bigr) \;=\; m(z) \;=\; \beta_0 \;+\; \beta_1\, z \;+\; \ldots \;+\; \beta_{k-1}\, z^{k-1}, \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \bigl(\beta_1, \,\ldots,\, \beta_{k-1}\bigr)^\top$$

Interpretation of the regression coefficients

$$\mathbb{E}\bigl(Y \,\big|\, Z = z+1\bigr) \;-\; \mathbb{E}\bigl(Y \,\big|\, Z = z\bigr)$$
$$= \beta_1 \;+\; \beta_2 \left\{(z+1)^2 - z^2\right\} \;+\; \cdots \;+\; \beta_{k-1} \left\{(z+1)^{k-1} - z^{k-1}\right\},$$
$$z \in \mathcal{Z}.$$

➠ any direct reasonable interpretation?

Effect of the covariate, **interpretation** of the regression coefficients

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.87279 -0.19903  0.00212  0.19780  0.90934

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground       3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)  9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

4.3.2 Raw polynomials

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 \,+\, \beta_1\, z \,+\, \ldots \,+\, \beta_{k-1}\, z^{k-1}, \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \big(\beta_1, \,\ldots,\, \beta_{k-1}\big)^\top$$

Degree of a polynomial

Degree $d-1$ ($d < k$) is sufficient to express the regression function

$$\equiv \;\; \mathrm{H}_0 : \; \beta_d = 0 \;\;\&\;\; \ldots \;\;\&\;\; \beta_{k-1} = 0.$$

➠ Wald type test (F-test) on a subvector of regression coefficients (under normality)

$\equiv$ submodel F-test (under normality)

Degree? Cubic versus quadratic, cubic versus linear polynomial

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

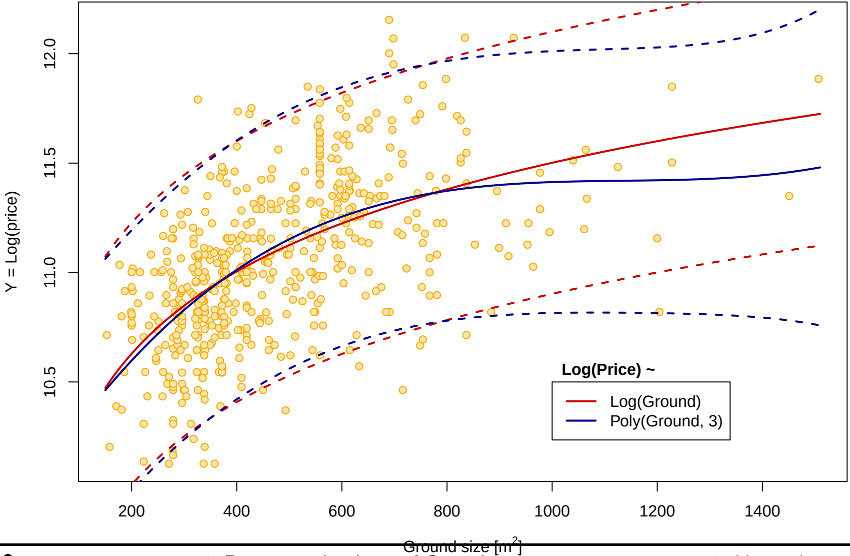
```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground       3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)  9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,       Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

```
rp3 <- lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987)
rp1 <- lm(log(price) ~ ground, data = Houses1987)
anova(rp1, rp3)
```

```
Analysis of Variance Table

Model 1: log(price) ~ ground
Model 2: log(price) ~ ground + I(ground^2) + I(ground^3)
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    544 53.186
2    542 48.968  2    4.2181 23.344 1.883e-10 ***
---
```

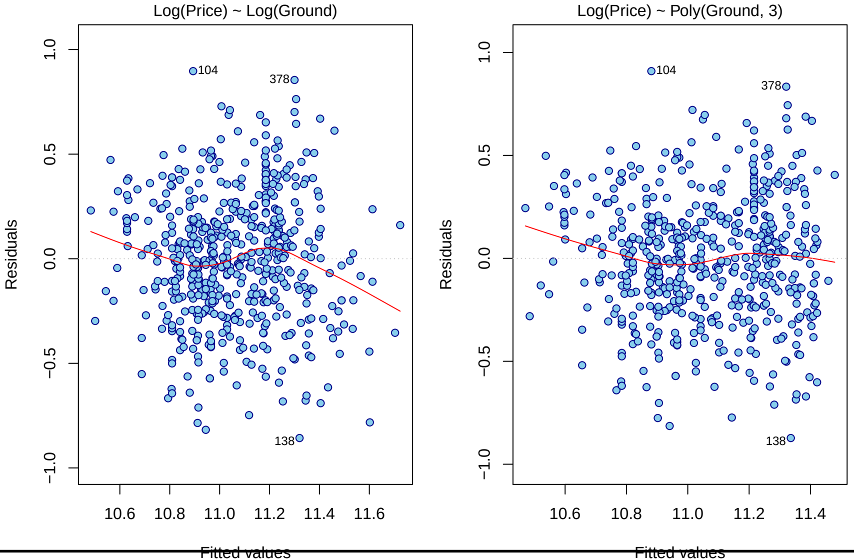
Houses1987 ($n = 546$)

log(price) $\sim$ log(ground) and log(price) $\sim$ rawpoly(ground, d),
$\widehat{m}$ with 95% prediction band



Houses1987 ($n = 546$)

`log(price)` ∼ `log(ground)` and `log(price)` ∼ `rawpoly(ground, d)`, residuals vs. fitted plots



Houses1987 ($n = 546$)

Practical importance of higher order polynomials?

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.87279 -0.19903  0.00212  0.19780  0.90934

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground       3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)  9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

4.3.3 Orthonormal polynomials

Regression function

$$m(z) = \beta_0 + \beta_1\, P^1(z) + \cdots + \beta_{k-1}\, P^{k-1}(z), \quad z \in \mathcal{Z},$$

P^j is an *orthonormal polynomial* of degree j, $j = 1, \ldots, k-1$ built above a set of the covariate datapoints $Z_1, \ldots, Z_n$.

$$P^j(z) = a_{j,0} + a_{j,1}\, z + \cdots + a_{j,j}\, z^j, \qquad j = 1, \ldots, k-1,$$

Reparameterizing matrix

$$\mathbb{S} = \Bigl(\boldsymbol{P}^1, \quad \ldots, \quad \boldsymbol{P}^{k-1}\Bigr) = \begin{pmatrix} P^1(Z_1) & \ldots & P^{k-1}(Z_1) \\ \vdots & \vdots & \vdots \\ P^1(Z_n) & \ldots & P^{k-1}(Z_n) \end{pmatrix}.$$

`log(price)` $\sim$ `orthpoly(ground, 3)`

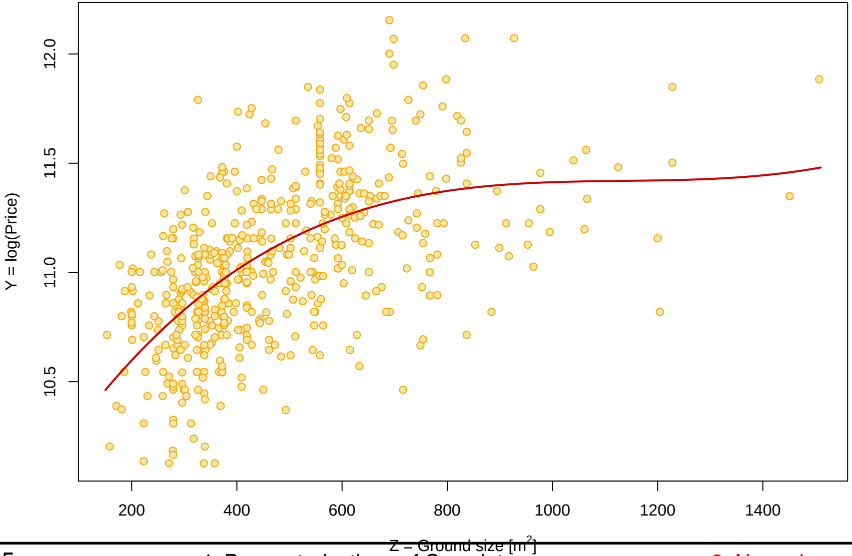
```
summary(lm(log(price) ~ poly(ground, degree = 3), data = Houses1987))
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.87279 -0.19903  0.00212  0.19780  0.90934 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               11.05896    0.01286 859.717  < 2e-16 ***
poly(ground, degree = 3)1  4.71459    0.30058  15.685  < 2e-16 ***
poly(ground, degree = 3)2 -1.96780    0.30058  -6.547 1.37e-10 ***
poly(ground, degree = 3)3  0.58811    0.30058   1.957   0.0509 .  
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471 
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

Houses1987 ($n = 546$)

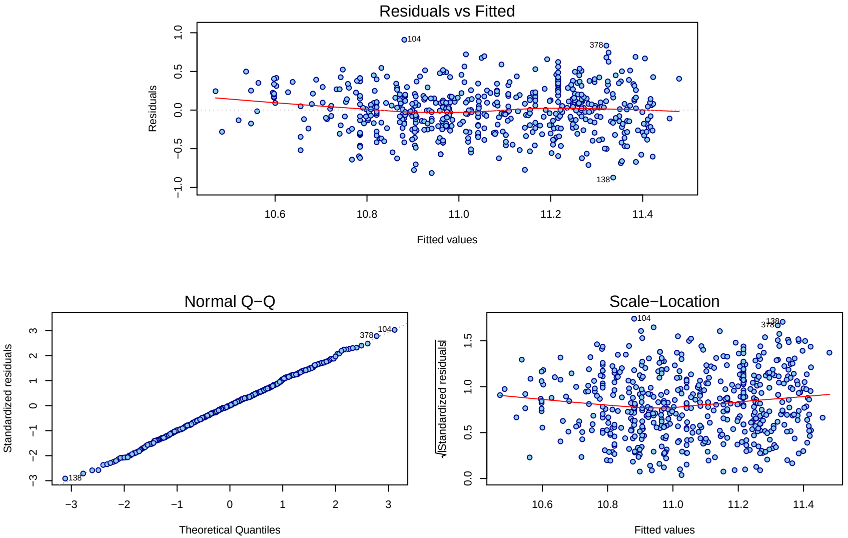
`log(price) ~ orthpoly(ground, 3),`

$$\widehat{m}(z) = 11.06 + 4.71\, P^1(z) - 1.97\, P^2(z) + 0.59\, P^3(z)$$



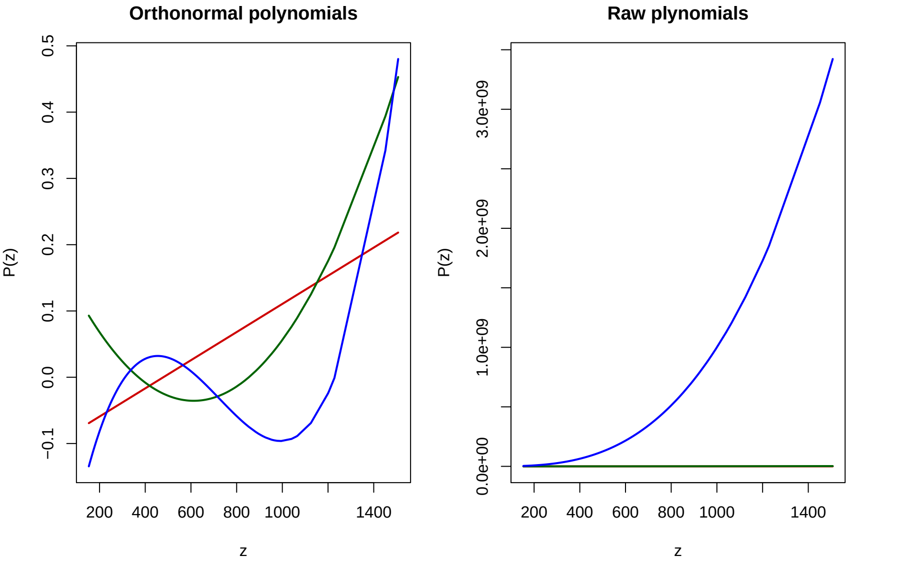
Houses1987 ($n = 546$)

log(price) ~ orthpoly(ground, 3), residual plots



Houses1987 ($n = 546$)

Basis orthonormal and raw polynomials



Advantages of orthonormal polynomials compared to raw polynomials

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground       3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)  9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(price) ~ poly(ground, degree = 3), data = Houses1987))
```

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               11.05896    0.01286 859.717  < 2e-16 ***
poly(ground, degree = 3)1  4.71459    0.30058  15.685  < 2e-16 ***
poly(ground, degree = 3)2 -1.96780    0.30058  -6.547 1.37e-10 ***
poly(ground, degree = 3)3  0.58811    0.30058   1.957   0.0509 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

4.3.3 Orthonormal polynomials

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 + \beta_1\, P^1(z) + \ldots + \beta_{k-1}\, P^{k-1}(z), \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \big(\beta_1, \ldots, \beta_{k-1}\big)^\top$$

Evaluation of the effect of the original covariate

$$\mathrm{H}_0 : \; \boldsymbol{\beta}^Z = \mathbf{0}_{k-1}$$

➠ Wald type test (F-test) on a subvector of regression coefficients (under normality)

$\equiv$ submodel F-test (under normality)

Houses1987 ($n = 546$)

Effect of the covariate (**cubic versus constant** regression function)

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground        3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)   9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(price) ~ poly(ground, degree = 3), data = Houses1987))
```

```
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               11.05896    0.01286 859.717  < 2e-16 ***
poly(ground, degree = 3)1  4.71459    0.30058  15.685  < 2e-16 ***
poly(ground, degree = 3)2 -1.96780    0.30058  -6.547 1.37e-10 ***
poly(ground, degree = 3)3  0.58811    0.30058   1.957   0.0509 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,	Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

4.3.3 Orthonormal polynomials

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 + \beta_1\, P^1(z) + \ldots + \beta_{k-1}\, P^{k-1}(z), \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \big(\beta_1, \ldots, \beta_{k-1}\big)^\top$$

Interpretation of the regression coefficients

$$\begin{aligned}\mathbb{E}\big(Y \,\big|\, Z = z+1\big) &- \mathbb{E}\big(Y \,\big|\, Z = z\big) \\ &= \beta_1 \left\{P^1(z+1) - P^1(z)\right\} + \beta_2 \left\{P^2(z+1) - P^2(z)\right\} + \cdots + \\ &\qquad \beta_{k-1} \left\{P^{k-1}(z+1) - P^{k-1}(z)\right\}, \\ &\qquad\qquad z \in \mathcal{Z}.\end{aligned}$$

➠ any direct reasonable interpretation?

4.3.3 Orthonormal polynomials

Regression function

$$\mathbb{E}\big(Y \,\big|\, Z = z\big) \;=\; m(z) \;=\; \beta_0 + \beta_1\, P^1(z) + \ldots + \beta_{k-1}\, P^{k-1}(z), \qquad z \in \mathcal{Z}$$

$$\boldsymbol{\beta}^Z \;:=\; \big(\beta_1, \ldots, \beta_{k-1}\big)^\top$$

Degree of a polynomial

Degree $d-1$ ($d<k$) is sufficient to express the regression function

$$\equiv \;\; \mathrm{H}_0: \;\beta_d = 0 \;\;\&\;\; \ldots \;\;\&\;\; \beta_{k-1} = 0.$$

➠ Wald type test (F-test) on a subvector of regression coefficients (under normality)

$\equiv$ submodel F-test (under normality)

Houses1987 ($n = 546$)

Degree? Cubic versus quadratic regression function

```
summary(lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987))
```

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.965e+00  1.371e-01  72.682  < 2e-16 ***
ground        3.784e-03  7.109e-04   5.323 1.49e-07 ***
I(ground^2) -3.306e-06  1.092e-06  -3.028  0.00258 **
I(ground^3)  9.700e-10  4.958e-10   1.957  0.05091 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,        Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

```
summary(lm(log(price) ~ poly(ground, degree = 3), data = Houses1987))
```

```
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                11.05896    0.01286 859.717  < 2e-16 ***
poly(ground, degree = 3)1  4.71459    0.30058  15.685  < 2e-16 ***
poly(ground, degree = 3)2 -1.96780    0.30058  -6.547 1.37e-10 ***
poly(ground, degree = 3)3  0.58811    0.30058   1.957   0.0509 .
---

Residual standard error: 0.3006 on 542 degrees of freedom
Multiple R-squared:  0.3507,        Adjusted R-squared:  0.3471
F-statistic: 97.57 on 3 and 542 DF,  p-value: < 2.2e-16
```

Degree? Cubic versus linear regression function

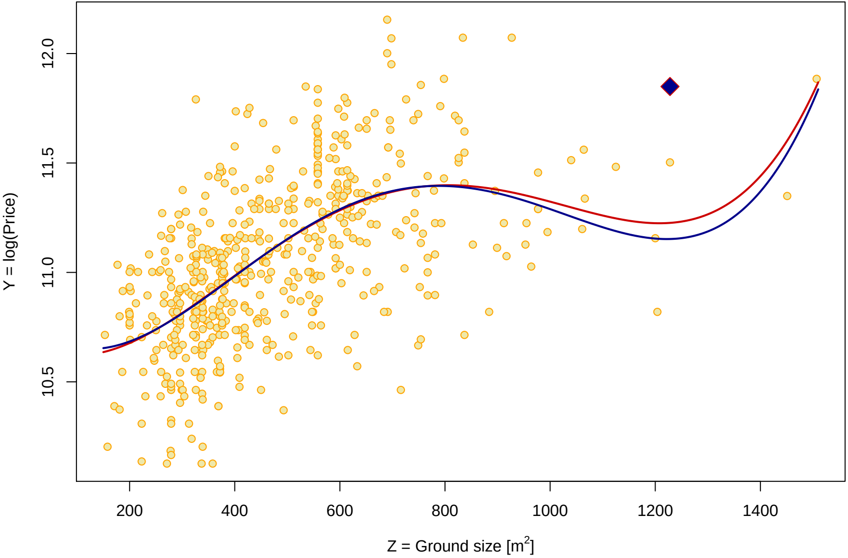
```
rp3 <- lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987)
rp1 <- lm(log(price) ~ ground, data = Houses1987)
anova(rp1, rp3)
```

```
Analysis of Variance Table

Model 1: log(price) ~ ground
Model 2: log(price) ~ ground + I(ground^2) + I(ground^3)
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    544 53.186
2    542 48.968  2    4.2181 23.344 1.883e-10 ***
---
```

```
op3 <- lm(log(price) ~ poly(ground, degree = 3), data = Houses1987)
op1 <- lm(log(price) ~ poly(ground, degree = 1), data = Houses1987)
anova(op1, op3)
```

```
Analysis of Variance Table

Model 1: log(price) ~ poly(ground, degree = 1)
Model 2: log(price) ~ poly(ground, degree = 3)
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    544 53.186
2    542 48.968  2    4.2181 23.344 1.883e-10 ***
---
```

log(price) ∼ poly(ground, 4), **global effect**



Basis splines

Definition 4.2 Basis spline with distinct knots.

Let $d \in \mathbb{N}_0$ and $\boldsymbol{\lambda} = \big(\lambda_1, \ldots, \lambda_{d+2}\big)^\top \in \mathbb{R}^{d+2}$, where $-\infty < \lambda_1 < \cdots < \lambda_{d+2} < \infty$. The *basis spline of degree d with distinct knots* $\boldsymbol{\lambda}$ is such a function $B^d(z;\, \boldsymbol{\lambda})$, $z \in \mathbb{R}$ that

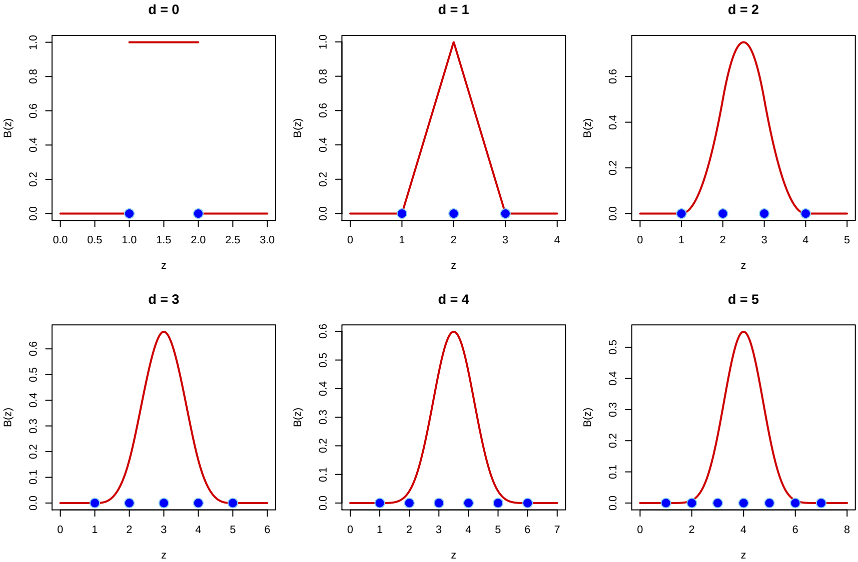
(i) $B^d(z;\, \boldsymbol{\lambda}) = 0$, for $z \le \lambda_1$ and $z \ge \lambda_{d+2}$;

(ii) On each of the intervals $(\lambda_j,\, \lambda_{j+1})$, $j = 1, \ldots, d+1$, $B^d(\cdot;\, \boldsymbol{\lambda})$ is a polynomial of degree d;

(iii) $B^d(\cdot;\, \boldsymbol{\lambda})$ has continuous derivatives up to an order $d-1$ on $\mathbb{R}$.

4.3.4 Regression splines

Some basis splines of degree $d = 0, \ldots, 5$



Definition 4.3 Basis spline with coincident left boundary knots.

Let $d \in \mathbb{N}_0$, $1 < r < d+2$ and $\boldsymbol{\lambda} = \big(\lambda_1, \ldots, \lambda_{d+2}\big)^\top \in \mathbb{R}^{d+2}$, where $-\infty < \lambda_1 = \cdots = \lambda_r < \cdots < \lambda_{d+2} < \infty$. The *basis spline of degree* d *with* r *coincident left boundary knots* $\boldsymbol{\lambda}$ is such a function $B^d(z;\, \boldsymbol{\lambda})$, $z \in \mathbb{R}$ that

(i) $B^d(z;\, \boldsymbol{\lambda}) = 0$, for $z \le \lambda_r$ and $z \ge \lambda_{d+2}$;

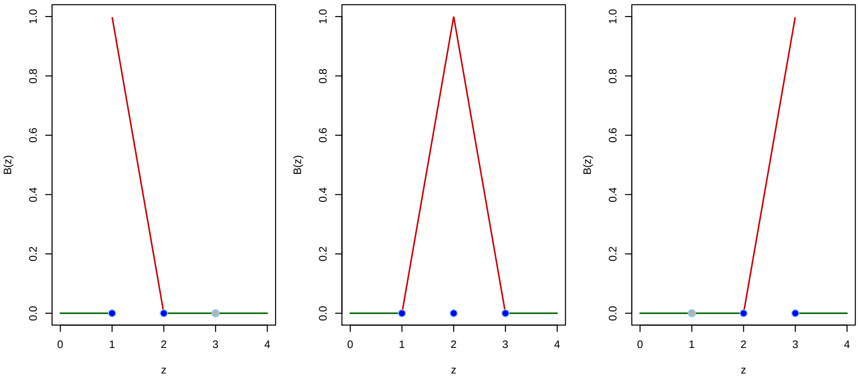
(ii) On each of the intervals $(\lambda_j,\, \lambda_{j+1})$, $j = r, \ldots, d+1$, $B^d(\cdot;\, \boldsymbol{\lambda})$ is a polynomial of degree d;

(iii) $B^d(\cdot;\, \boldsymbol{\lambda})$ has continuous derivatives up to an order $d-1$ on $(\lambda_r,\, \infty)$;

(iv) $B^d(\cdot;\, \boldsymbol{\lambda})$ has continuous derivatives up to an order $d-r$ in λ_r.

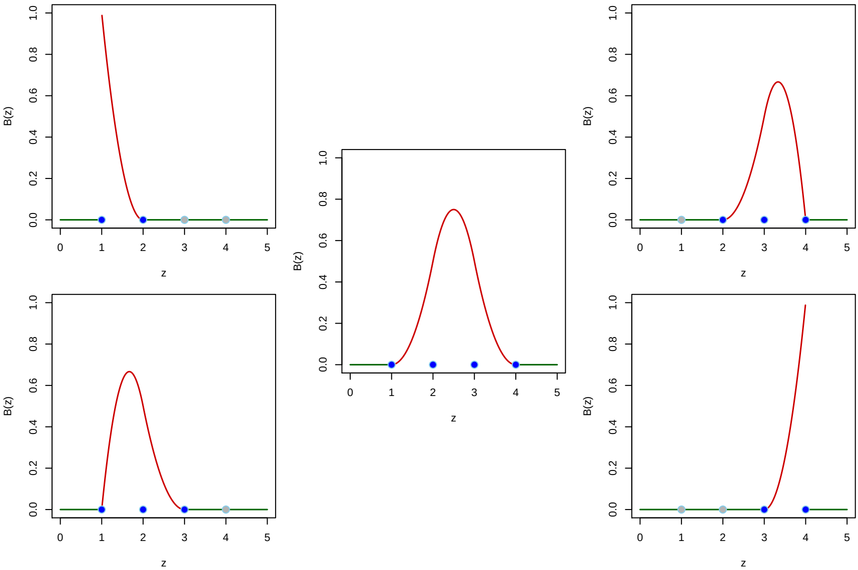
4.3.4 Regression splines

Some basis splines of degree $d = 1$ with possibly coincident boundary knots



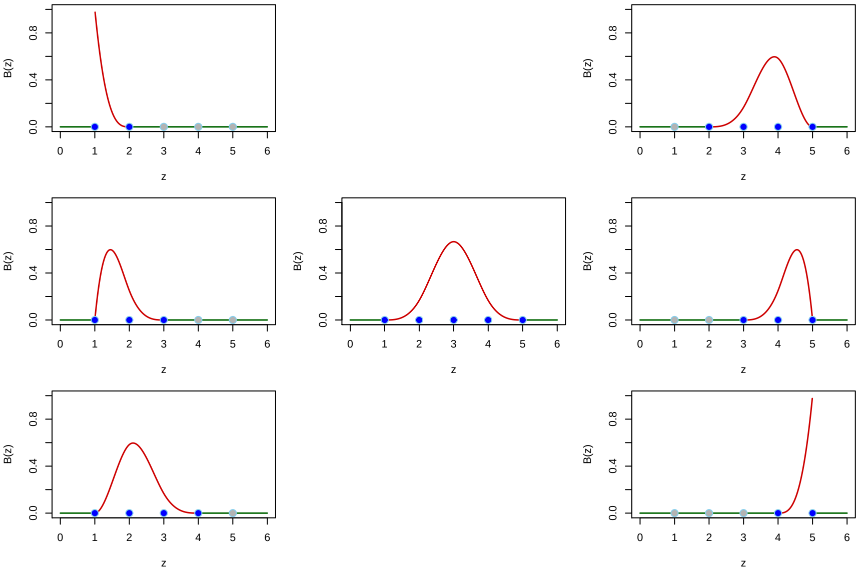
4.3.4 Regression splines

Some basis splines of degree $d = 2$ with possibly coincident boundary knots



4.3.4 Regression splines

Some basis splines of degree $d = 3$ with possibly coincident boundary knots



4.3.4 Regression splines

Basis B-splines

Previous plots showed basis B-splines.

Useful properties of a basis B-spline with knots $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d+2})^\top$:

$$B^d(z, \boldsymbol{\lambda}) > 0, \qquad \lambda_1 < z < \lambda_{d+2},$$

$$B^d(z, \boldsymbol{\lambda}) = 0, \qquad z \leq \lambda_1,\ z \geq \lambda_{d+2}.$$

Definition 4.4 Spline basis.

Let $d \in \mathbb{N}_0$, $k \geq d+1$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{k-d+1})^\top \in \mathbb{R}^{k-d+1}$, where $-\infty < \lambda_1 < \ldots < \lambda_{k-d+1} < \infty$. The *spline basis* of degree d with knots $\boldsymbol{\lambda}$ is a set of basis splines $B_1, \ldots, B_k$, where for $z \in \mathbb{R}$,

$$B_1(z) = B^d(z;\, \underbrace{\lambda_1, \ldots, \lambda_1}_{(d+1)\times}, \lambda_2),$$

$$B_2(z) = B^d(z;\, \underbrace{\lambda_1, \ldots, \lambda_1}_{d\times}, \lambda_2, \lambda_3),$$

$$\vdots$$

$$B_d(z) = B^d(z;\, \underbrace{\lambda_1, \lambda_1}_{2\times}, \lambda_2, \ldots, \lambda_{d+1}),$$

$$B_{d+1}(z) = B^d(z;\, \lambda_1, \lambda_2, \ldots, \lambda_{d+2}),$$

$$B_{d+2}(z) = B^d(z;\, \lambda_2, \ldots, \lambda_{d+3}),$$

$$\vdots$$

$$B_{k-d}(z) = B^d(z;\, \lambda_{k-2d}, \ldots, \lambda_{k-d+1}),$$

$$B_{k-d+1}(z) = B^d(z;\, \lambda_{k-2d+1}, \ldots, \underbrace{\lambda_{k-d+1}, \lambda_{k-d+1}}_{2\times}),$$

$$\vdots$$

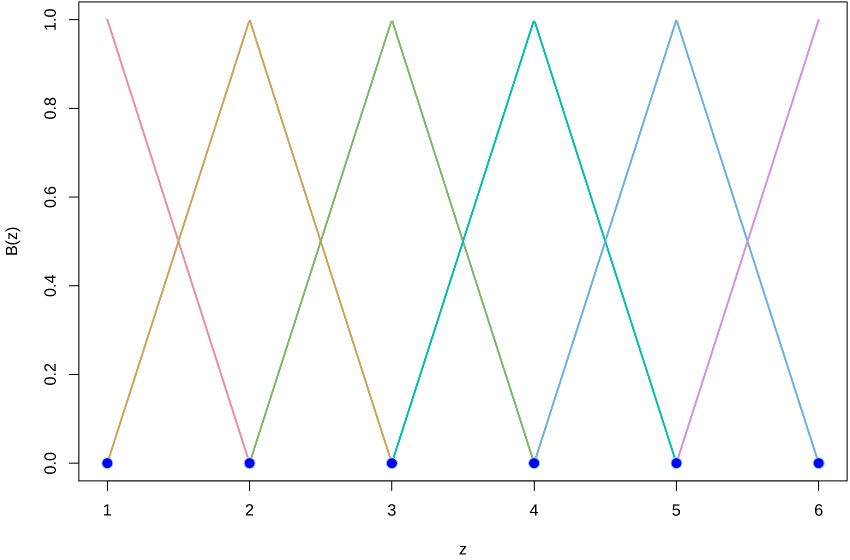
$$B_{k-1}(z) = B^d(z;\, \lambda_{k-d-1}, \lambda_{k-d} \ldots, \underbrace{\lambda_{k-d+1}, \ldots, \lambda_{k-d+1}}_{d\times}),$$

$$B_k(z) = B^d(z;\, \lambda_{k-d} \ldots, \underbrace{\lambda_{k-d+1}, \ldots, \lambda_{k-d+1}}_{(d+1)\times}).$$

$$\vdots$$

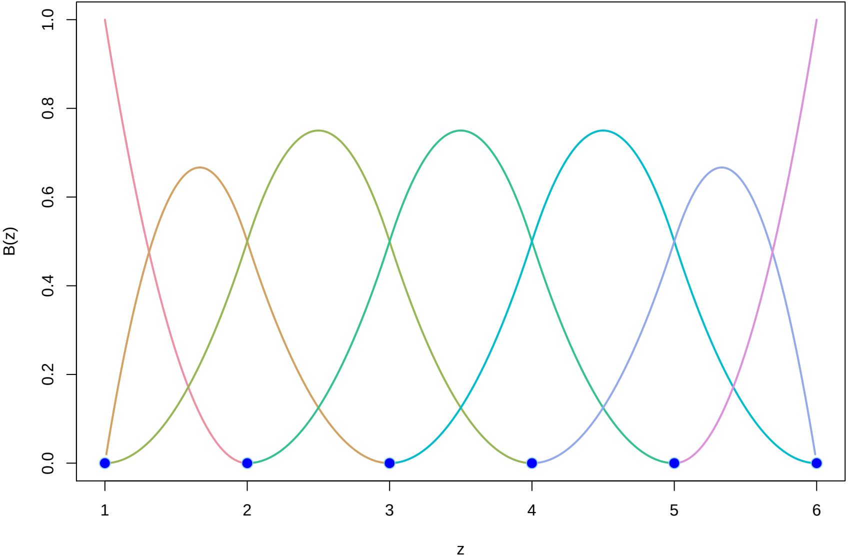
4.3.4 Regression splines

Linear B-spline basis (of degree $d = 1$)



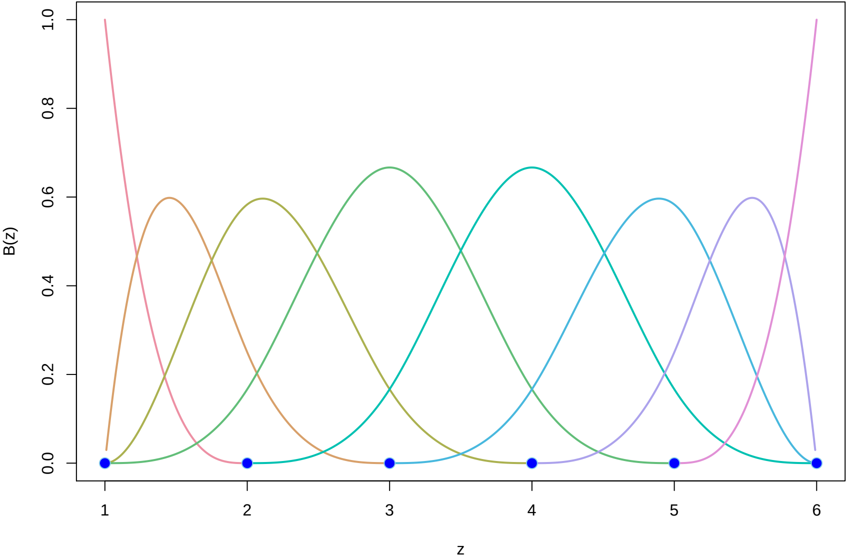
4.3.4 Regression splines

Quadratic B-spline basis (of degree $d = 2$)



4.3.4 Regression splines

Cubic B-spline basis (of degree $d = 3$)



Spline basis

Properties of the B-spline basis

(a) $$\sum_{j=1}^{k} B_j(z) = 1 \qquad \text{for all } z \in \left(\lambda_1,\, \lambda_{k-d+1}\right);$$

(b) for each $m \leq d$ there exist a set of coefficients $\gamma_1^m, \ldots, \gamma_k^m$ such that

$$\sum_{j=1}^{k} \gamma_j^m B_j(z) \text{ is on } (\lambda_1,\, \lambda_{k-d+1}) \text{ a polynomial in } z \text{ of degree } m.$$

4.3.4 Regression splines

Regression spline

Assumption:

Covariate space $\mathcal{Z} = \big(z_{min},\, z_{max}\big)$, $-\infty < z_{min} < z_{max} < \infty$.

Regression function

$$m(z) = \beta_1\, B_1(z) + \cdots + \beta_k\, B_k(z), \quad z \in \mathcal{Z},$$

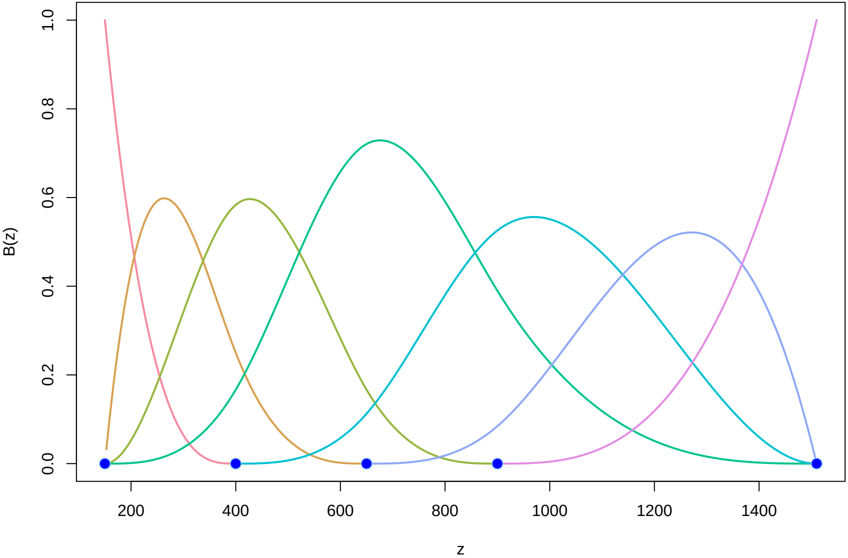
$B_1, \ldots, B_k$ is the spline basis of chosen degree $d \in \mathbb{N}_0$ composed of basis B-splines built above a set of chosen knots $\boldsymbol{\lambda} = \big(\lambda_1, \ldots, \lambda_{k-d+1}\big)^\top$, $z_{min} = \lambda_1 < \ldots < \lambda_{k-d+1} = z_{max}$.

Reparameterizing matrix

$$\mathbb{X} = \mathbb{S} = \begin{pmatrix} B_1(Z_1) & \ldots & B_k(Z_1) \\ \vdots & \vdots & \vdots \\ B_1(Z_n) & \ldots & B_k(Z_n) \end{pmatrix} =: \mathbb{B}.$$

Houses1987 ($n = 546$)

B-spline basis (cubic, $d = 3$, $\boldsymbol{\lambda} = \left(150,\ 400,\ 650,\ 900,\ 1510\right)^\top$)



Houses1987 ($n = 546$)

log(price) ∼ spline(ground, degree = 3), model matrix $\mathbb{X} = \mathbb{B}$

```
lambda.inner <- c(400, 650, 900)
lambda.bound <- c(150, 1510)
Bx <- bs(Houses1987[, "ground"],
         knots = lambda.inner, Boundary.knots = lambda.bound,
         degree = 3, intercept = TRUE)
showBx <- data.frame(ground = Houses1987[, "ground"],
                 B1 = Bx[,1], B2 = Bx[,2], B3 = Bx[,3],
                 B4 = Bx[,4], B5 = Bx[,5], B6 = Bx[,6], B7 = Bx[,7])
print(showBx)
```

```
   ground    B1    B2    B3    B4    B5 B6 B7
1     544 0.000 0.019 0.424 0.535 0.022  0  0
2     372 0.001 0.341 0.541 0.117 0.000  0  0
3     285 0.097 0.583 0.293 0.026 0.000  0  0
4     619 0.000 0.000 0.235 0.689 0.076  0  0
5     592 0.000 0.003 0.302 0.644 0.051  0  0
6     387 0.000 0.291 0.567 0.142 0.000  0  0
7     361 0.004 0.379 0.517 0.100 0.000  0  0
8     387 0.000 0.291 0.567 0.142 0.000  0  0
9     447 0.000 0.134 0.590 0.275 0.001  0  0
10    512 0.000 0.042 0.497 0.451 0.010  0  0
11    670 0.000 0.000 0.130 0.729 0.142  0  0
12    279 0.113 0.590 0.273 0.023 0.000  0  0
13    158 0.907 0.091 0.002 0.000 0.000  0  0
14    268 0.147 0.597 0.238 0.018 0.000  0  0
15    335 0.018 0.465 0.450 0.068 0.000  0  0
...
```

Houses1987 ($n = 546$)

log(price) ∼ spline(ground, degree = 3)

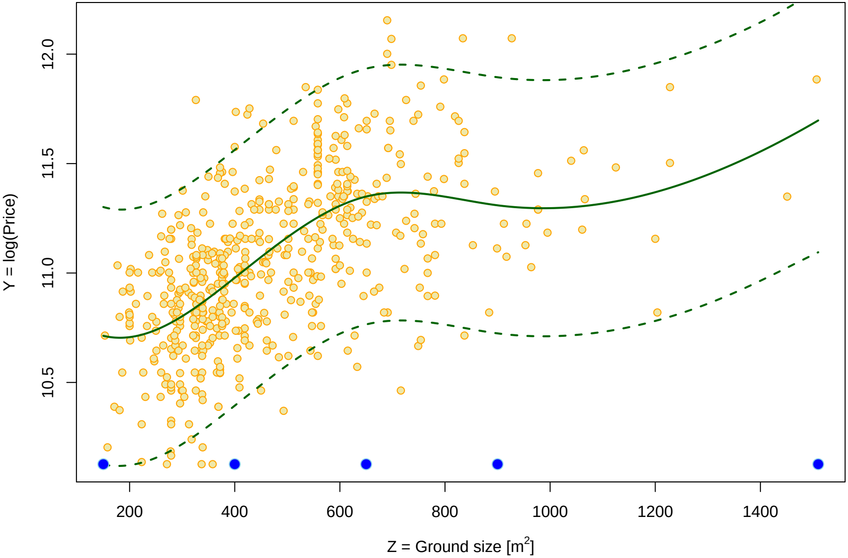
```
summary(lm(log(price) ~ Bx - 1, data = Houses1987))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.90457 -0.19497  0.00698  0.19693  0.94698

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
Bx1 10.71312    0.12078   88.70   <2e-16 ***
Bx2 10.66519    0.07956  134.06   <2e-16 ***
Bx3 10.97388    0.07464  147.03   <2e-16 ***
Bx4 11.46283    0.06699  171.11   <2e-16 ***
Bx5 11.17900    0.16773   66.65   <2e-16 ***
Bx6 11.41145    0.31448   36.29   <2e-16 ***
Bx7 11.69708    0.25076   46.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2974 on 539 degrees of freedom
Multiple R-squared:  0.9993,        Adjusted R-squared:  0.9993
F-statistic: 1.079e+05 on 7 and 539 DF,  p-value: < 2.2e-16
```

!!! R-squared’s and the F-statistic in the output do not have usual interpretation **!!!**

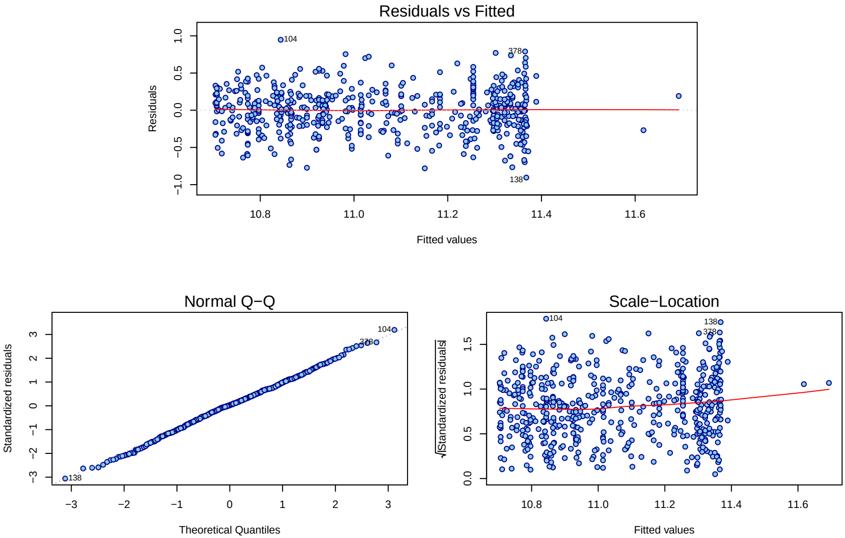
`Houses1987` ($n = 546$)

`log(price)` $\sim$ `spline(ground)`, $\widehat{m}(z) = 10.71\,B_1(z) + 10.67\,B_2(z) + 10.97\,B_3(z) + 11.46\,B_4(z) + 11.18\,B_5(z) + 11\,41\,B_6(z) + 11.70\,B_7(z)$ and the 95% prediction band



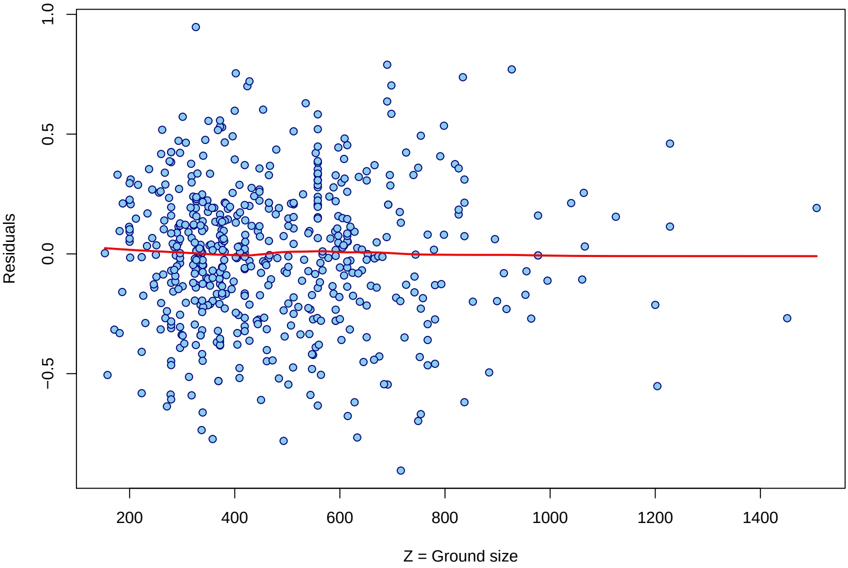
Houses1987 ($n = 546$)

log(price) ~ spline(ground), residual plots



Houses1987 ($n = 546$)

`log(price)` ∼ `spline(ground)`, residuals versus covariate plot



4.3.4 Regression splines

Regression function

$$\mathbb{E}(Y \,|\, Z = z) \;=\; m(z) \;=\; \beta_1\, B_1(z) + \ldots + \beta_k\, B_k(z), \qquad z \in \mathcal{Z}$$

Evaluation of the effect of the original covariate

Remember: $\sum_{j=1}^{k} B_j(z) = 1$ for $z \in (\lambda_1,\, \lambda_{k-d+1})$

$$\mathrm{H}_0 : \; \beta_1 = \cdots = \beta_k$$

$$\equiv \;\; \mathbb{E}(\boldsymbol{Y} \,|\, \mathbb{Z}) \in \mathcal{M}(\mathbf{1}_n) \subset \mathcal{M}(\mathbb{B})$$

➡ Submodel F-test (under normality)

Effect of the covariate

```
mB <- lm(log(price) ~ Bx - 1, data = Houses1987)
m0 <- lm(log(price) ~ 1, data = Houses1987)
anova(m0, mB)
```

```
Analysis of Variance Table

Model 1: log(price) ~ 1
Model 2: log(price) ~ Bx - 1
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    545 75.413
2    539 47.663  6     27.75 52.302 < 2.2e-16 ***
---
```

Houses1987 ($n = 546$)

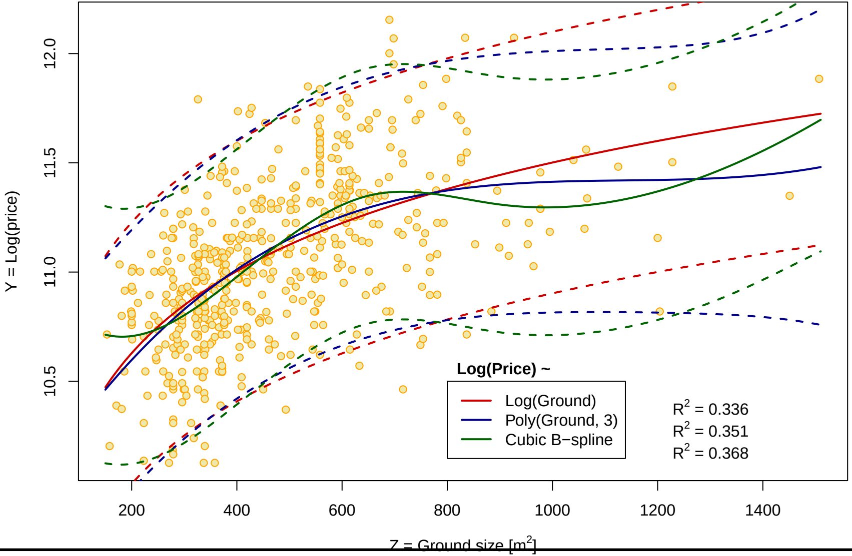
Spline better than a (global) cubic polynomial?

```
mB <- lm(log(price) ~ Bx - 1, data = Houses1987)
mpoly3 <- lm(log(price) ~ ground + I(ground^2) + I(ground^3), data = Houses1987)
anova(mpoly3, mB)
```

```
Analysis of Variance Table

Model 1: log(price) ~ ground + I(ground^2) + I(ground^3)
Model 2: log(price) ~ Bx - 1
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    542 48.968
2    539 47.663  3    1.3045 4.9174 0.002226 **
---
```

`Houses1987` ($n = 546$)

`log(price)` ~ `log(ground)`, `log(price)` ~ `poly(ground, 3)`, `log(price)` ~ `spline(ground, degree = 3)`, $\widehat{m}$ with the 95% prediction band



4.3.4 Regression splines

Regression function

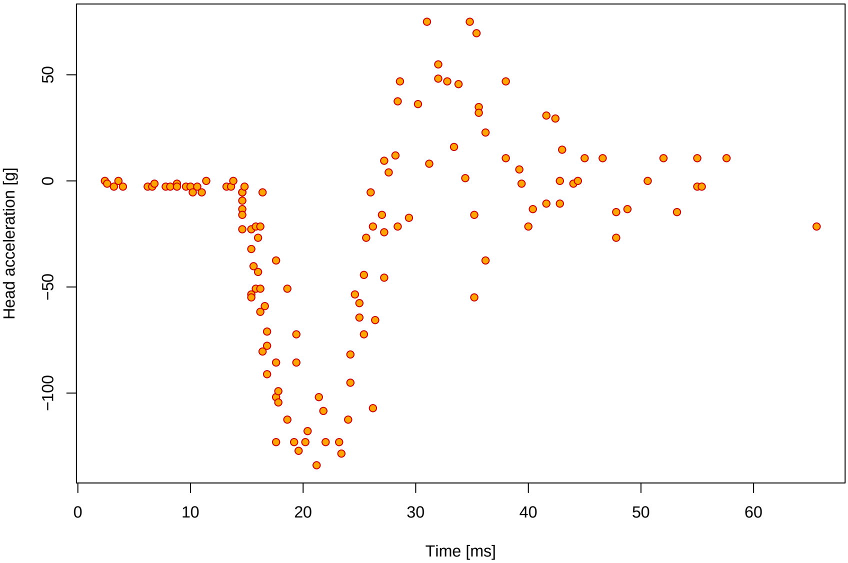
$$\mathbb{E}(Y \mid Z = z) \;=\; m(z) \;=\; \beta_1 B_1(z) + \ldots + \beta_k B_k(z), \qquad z \in \mathcal{Z}$$

Interpretation of the regression coefficients

Any direct reasonable interpretation?

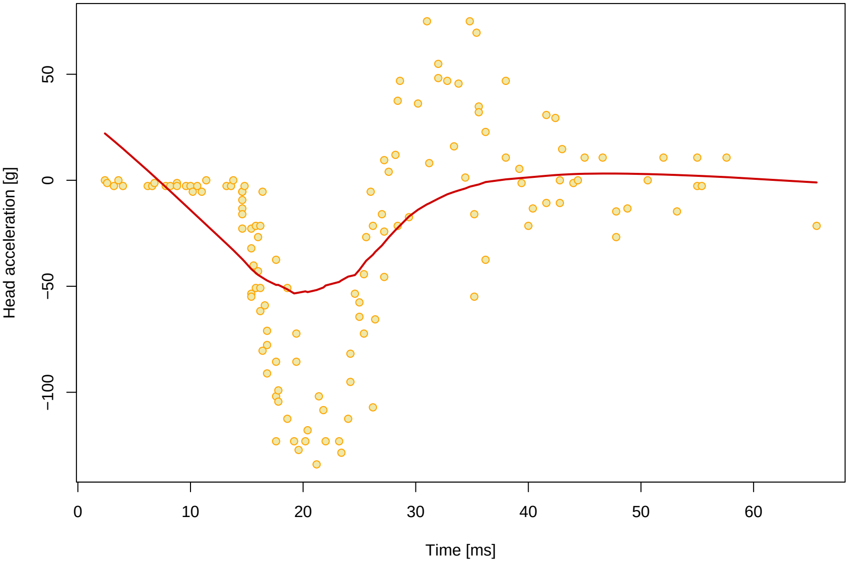
Motorcycle ($n = 133$)

haccel ∼ time



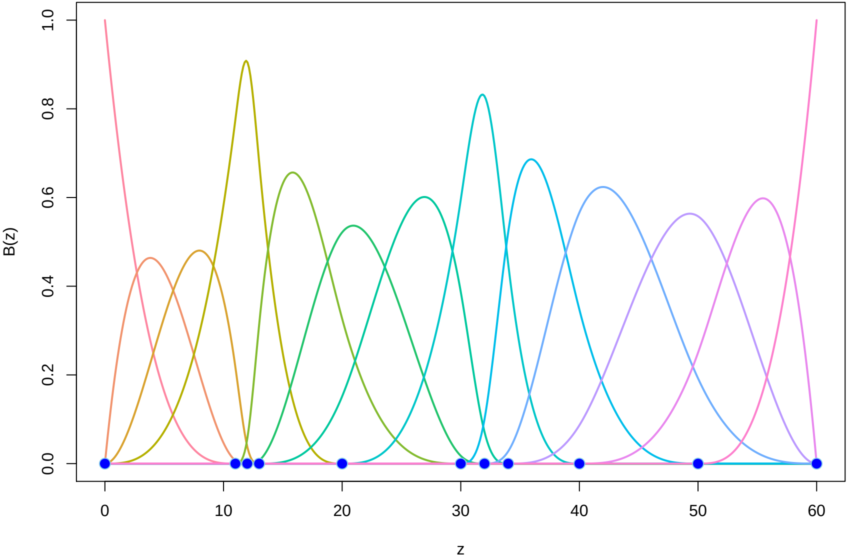
Motorcycle ($n = 133$)

haccel ~ time, scatterplot with the LOWESS smoother



Motorcycle ($n = 133$)

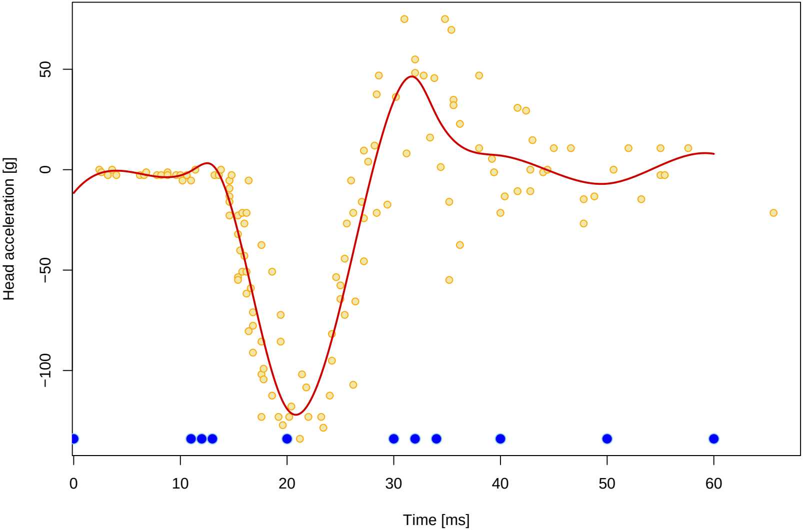
B-spline basis (cubic, $d = 3$, $\boldsymbol{\lambda} = \big(0,\ 11,\ 12,\ 13,\ 20,\ 30,\ 32,\ 34,\ 40,\ 50,\ 60\big)^\top$)



Motorcycle ($n = 133$)

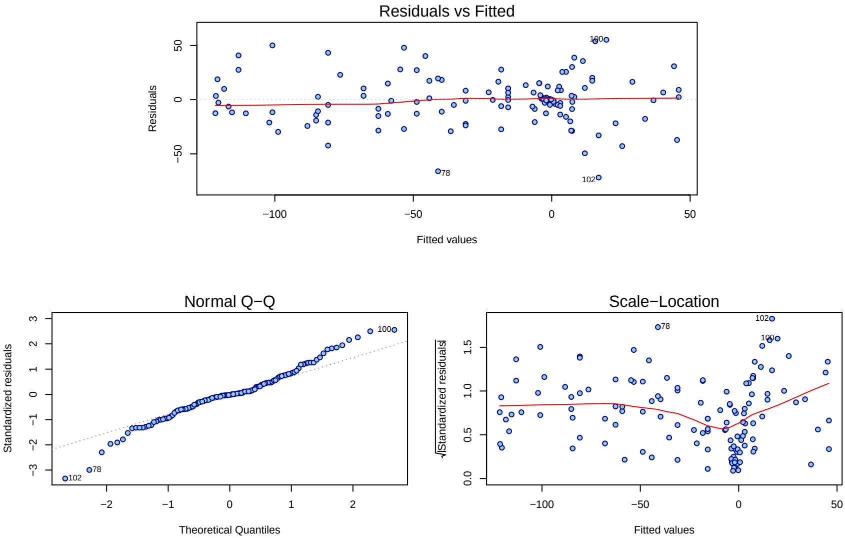
`haccel ~ spline(time),`

$$\widehat{m}(z) = -11.62\, B_1(z) + 12.45\, B_2(z) - 13.99\, B_3(z) + 2.99\, B_4(z) + 6.11\, B_5(z) - 237.28\, B_6(z) + 17.34\, B_7(z) + 53.26\, B_8(z) + 5.07\, B_9(z) + 12.72\, B_{10}(z) - 22.00\, B_{11}(z) + 11.37\, B_{12}(z) + 6.97\, B_{13}(z)$$



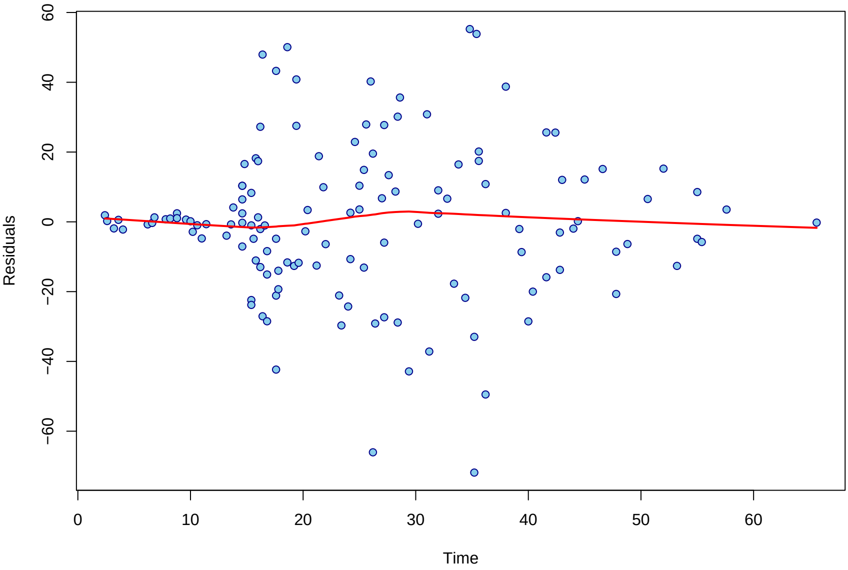
Motorcycle ($n = 133$)

haccel ∼ spline(time), residual plots



Motorcycle ($n = 133$)

haccel ~ spline(time), residuals versus covariate plot



Section **4.4**

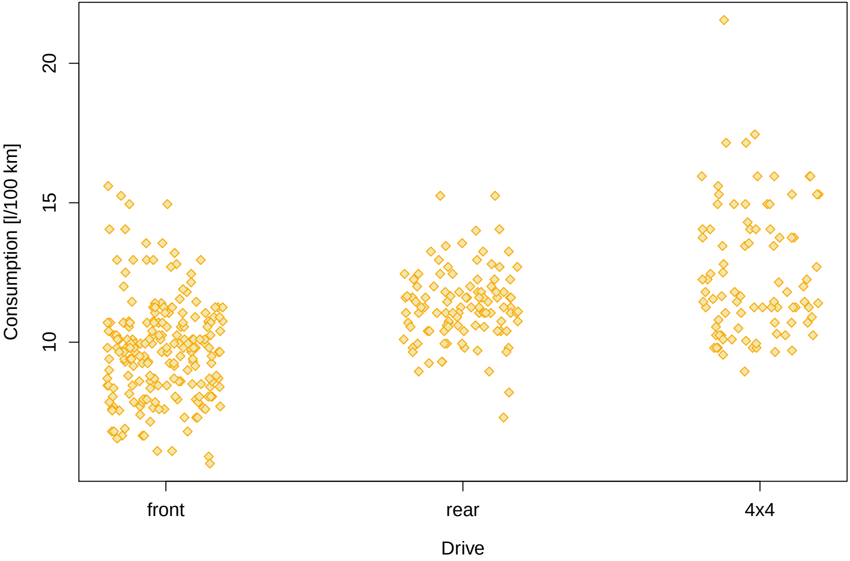
Categorical covariate

Cars2004nh (subset, $n = 409$)

`consumption ~ drive`

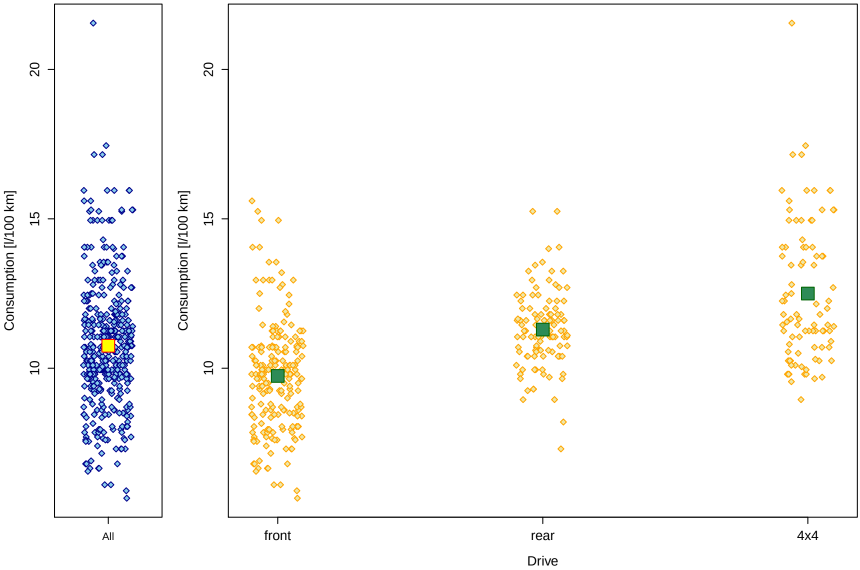
Cars2004nh (subset, $n = 409$)

consumption $\sim$ drive



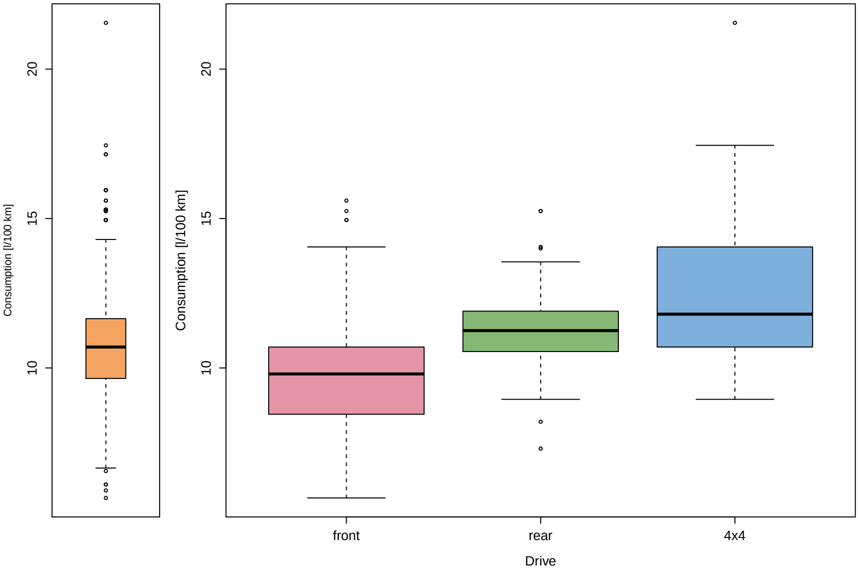
Cars2004nh (subset, $n = 409$)

consumption ∼ drive



4.4.1 Link to a *G*-sample problem

`Cars2004nh` (subset, $n = 409$)



4.4.2 Linear model parameterization of one-way classified group means

$\boldsymbol{\mu}$: the (conditional) response expectation

$$\mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{Z}\big) = \boldsymbol{\mu} := \begin{pmatrix} \mu_{1,1} \\ \vdots \\ \mu_{1,n_1} \\ -- \\ \vdots \\ -- \\ \mu_{G,1} \\ \vdots \\ \mu_{G,n_G} \end{pmatrix} = \begin{pmatrix} m_1 \\ \vdots \\ m_1 \\ -- \\ \vdots \\ -- \\ m_G \\ \vdots \\ m_G \end{pmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_1\text{-times} \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_G\text{-times} \end{matrix} = \begin{pmatrix} m_1\,\mathbf{1}_{n_1} \\ \vdots \\ m_G\,\mathbf{1}_{n_G} \end{pmatrix}.$$

4.4.3 Full-rank parameterization of one-way classified group means

$\boldsymbol{\beta} = \big(\beta_0,\, \beta_1,\, \ldots,\, \beta_{G-1}\big)^\top,\ \boldsymbol{\beta}^Z = \big(\beta_1,\, \ldots,\, \beta_{G-1}\big)^\top$

$$\begin{aligned} m_g &= \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \qquad g = 1, \ldots, G, \\ \boldsymbol{m} = \widetilde{\mathbb{X}}\boldsymbol{\beta} = \big(\boldsymbol{1}_G,\ \mathbb{C}\big)\boldsymbol{\beta} &= \beta_0\, \boldsymbol{1}_G + \mathbb{C}\boldsymbol{\beta}^Z \end{aligned}$$

$\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta}, \quad \boldsymbol{\beta} = \big(\beta_0,\, \beta_1,\, \ldots,\, \beta_{G-1}\big)^\top,\ \boldsymbol{\beta}^Z = \big(\beta_1,\, \ldots,\, \beta_{G-1}\big)^\top$

$$\mathbb{X} = \begin{pmatrix} 1 & \boldsymbol{c}_1^\top \\ \vdots & \vdots \\ 1 & \boldsymbol{c}_1^\top \\ - & - \\ \vdots & \vdots \\ - & - \\ 1 & \boldsymbol{c}_G^\top \\ \vdots & \vdots \\ 1 & \boldsymbol{c}_G^\top \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} n_1\text{-times} \\ \\ \\ \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} n_G\text{-times} \end{matrix} = \begin{pmatrix} \boldsymbol{1}_{n_1} \otimes \big(1,\ \boldsymbol{c}_1^\top\big) \\ \vdots \\ \boldsymbol{1}_{n_G} \otimes \big(1,\ \boldsymbol{c}_G^\top\big) \end{pmatrix}$$

4.4.3 Full-rank parameterization of one-way classified group means

Definition 4.5 Full-rank parameterization of a categorical covariate.

Full-rank parameterization of a categorical covariate with G levels ($G = \mathrm{card}(\mathcal{Z})$) is a choice of the $G \times (G-1)$ matrix $\mathbb{C}$ that satisfies

$$\mathrm{rank}(\mathbb{C}) = G - 1, \qquad \mathbf{1}_G \notin \mathcal{M}(\mathbb{C}).$$

4.4.3 Full-rank parameterization of one-way classified group means

$\boldsymbol{\beta} = \bigl(\beta_0,\, \beta_1,\, \ldots,\, \beta_{G-1}\bigr)^\top,\ \boldsymbol{\beta}^Z = \bigl(\beta_1,\, \ldots,\, \beta_{G-1}\bigr)^\top$

$$m_g \;=\; \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \qquad g = 1, \ldots, G,$$

$$\boldsymbol{m} = \widetilde{\mathbb{X}}\boldsymbol{\beta} = \bigl(\mathbf{1}_G,\, \mathbb{C}\bigr)\boldsymbol{\beta} \;=\; \beta_0\,\mathbf{1}_G + \mathbb{C}\boldsymbol{\beta}^Z$$

Evaluation of the effect of the categorical covariate

$\mathrm{H}_0:\ m_1 \;=\; \cdots \;=\; m_G$

$\equiv\ \mathrm{H}_0:\ \beta_1 = 0\ \ \&\ \ \cdots\ \ \&\ \ \beta_{G-1} = 0 \qquad \equiv\ \mathrm{H}_0:\ \boldsymbol{\beta}^Z = \mathbf{0}_{G-1}$

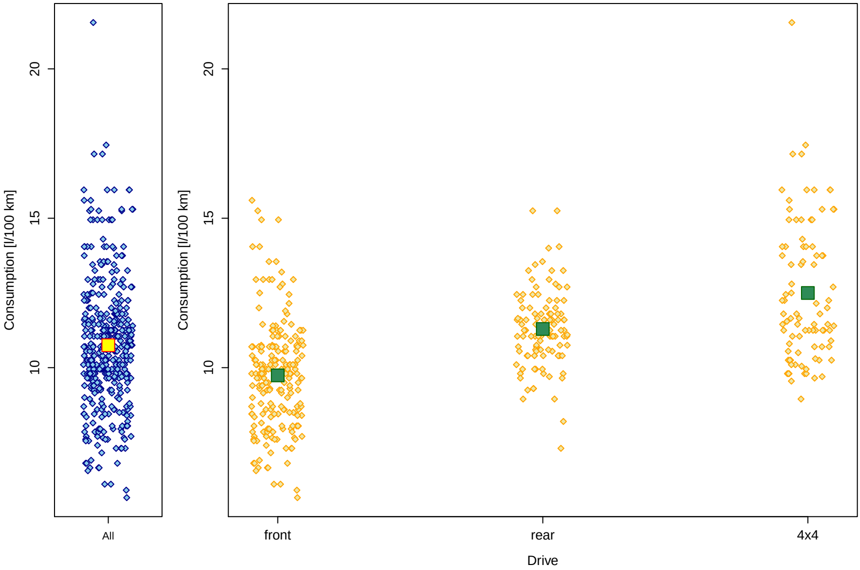
➡ Wald type test (F-test) on a subvector of regression coefficients (under normality)

$\equiv$ submodel F-test (under normality)

- $G = 2$ $\equiv$ (equal variances) two-sample t-test
- $G > 2$ $\equiv$ one-way ANOVA F-test

Cars2004nh (subset, $n = 409$, $n_{front} = 212$, $n_{rear} = 108$, $n_{4x4} = 89$)

$\overline{Y} = 10.75$, $\overline{Y}_{front} = 9.74$, $\overline{Y}_{rear} = 11.29$, $\overline{Y}_{4x4} = 12.50$



4.4.3 Full-rank parameterization of one-way classified group means

Reference group pseudocontrasts (dummy variables)

$\mathbb{C}$: `contr.treatment`

$$\mathbb{C} = \begin{pmatrix} 0 & \dots & 0 \\ 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{G-1}^\top \\ \mathbf{I}_{G-1} \end{pmatrix}$$

$\boldsymbol{m} = \beta_0 \mathbf{1}_G + \mathbb{C}\boldsymbol{\beta}^Z, \quad \boldsymbol{\beta}^Z = \big(\beta_1, \dots, \beta_{G-1}\big)^\top$

$$\begin{aligned} m_1 &= \beta_0, & \beta_0 &= m_1, \\ m_2 &= \beta_0 + \beta_1, & \beta_1 &= m_2 - m_1, \\ &\vdots & &\vdots \\ m_G &= \beta_0 + \beta_{G-1}, & \beta_{G-1} &= m_G - m_1. \end{aligned}$$

Cars2004nh (subset, $n = 409$, $n_{front} = 212$, $n_{rear} = 108$, $n_{4x4} = 89$)

$\overline{Y} = 10.75$, $\overline{Y}_{front} = 9.74$, $\overline{Y}_{rear} = 11.29$, $\overline{Y}_{4x4} = 12.50$

```
CarsNow <- subset(Cars2004nh,
       complete.cases(Cars2004nh[, c("consumption", "lweight", "engine.size")]))
mTrt <- lm(consumption ~ fdrive, data = CarsNow)
summary(mTrt)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.0913 -1.2489 -0.0440  0.9587  9.0511

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.7413     0.1247  78.149  < 2e-16 ***
fdriverear    1.5527     0.2146   7.237 2.32e-12 ***
fdrive4x4     2.7576     0.2292  12.030  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.815 on 406 degrees of freedom
Multiple R-squared:  0.2799,	Adjusted R-squared:  0.2764
F-statistic: 78.91 on 2 and 406 DF,  p-value: < 2.2e-16
```

4.4.3 Full-rank parameterization of one-way classified group means

Reference group pseudocontrasts (dummy variables)

$\mathbb{C}$: `contr.SAS`

$$\mathbb{C} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{G-1} \\ \mathbf{0}_{G-1}^{\top} \end{pmatrix}$$

$\boldsymbol{m} = \beta_0 \, \mathbf{1}_G + \mathbb{C}\boldsymbol{\beta}^Z, \quad \boldsymbol{\beta}^Z = \bigl(\beta_1, \dots, \beta_{G-1}\bigr)^{\top}$

$$\begin{aligned} m_1 &= \beta_0 + \beta_1, \\ &\vdots \\ m_{G-1} &= \beta_0 + \beta_{G-1}, \\ m_G &= \beta_0, \end{aligned} \qquad \begin{aligned} \beta_1 &= m_1 - m_G, \\ &\vdots \\ \beta_{G-1} &= m_{G-1} - m_G, \\ \beta_0 &= m_G. \end{aligned}$$

`Cars2004nh` (subset, $n = 409$, $n_{front} = 212$, $n_{rear} = 108$, $n_{4x4} = 89$)

$\overline{Y} = 10.75$, $\overline{Y}_{front} = 9.74$, $\overline{Y}_{rear} = 11.29$, $\overline{Y}_{4x4} = 12.50$

```
mSAS <- lm(consumption ~ fdrive, data = CarsNow, contrasts = list(fdrive = contr.SAS))
summary(mSAS)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.0913 -1.2489 -0.0440  0.9587  9.0511

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4989     0.1924  64.969  < 2e-16 ***
fdrive1      -2.7576     0.2292 -12.030  < 2e-16 ***
fdrive2      -1.2049     0.2598  -4.637 4.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.815 on 406 degrees of freedom
Multiple R-squared:  0.2799,       Adjusted R-squared:  0.2764
F-statistic: 78.91 on 2 and 406 DF,  p-value: < 2.2e-16
```

Sum contrasts

$\mathbb{C}$: `contr.sum`

$$\mathbb{C} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ -1 & \dots & -1 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{G-1} \\ -\mathbf{1}_{G-1}^\top \end{pmatrix}$$

$\boldsymbol{m} = \beta_0 \mathbf{1}_G + \mathbb{C}\boldsymbol{\beta}^Z, \quad \boldsymbol{\beta}^Z = \big(\beta_1, \dots, \beta_{G-1}\big)^\top, \qquad \overline{m} = \frac{1}{G}\sum_{g=1}^{G} m_g$

$$\begin{aligned} m_1 &= \beta_0 + \beta_1, \\ &\vdots \\ m_{G-1} &= \beta_0 + \beta_{G-1}, \\ m_G &= \beta_0 - \sum_{g=1}^{G-1} \beta_g, \end{aligned} \qquad \begin{aligned} \beta_0 &= \overline{m}, \\ \beta_1 &= m_1 - \overline{m}, \\ &\vdots \\ \beta_{G-1} &= m_{G-1} - \overline{m}. \end{aligned}$$

Cars2004nh (subset, $n = 409$, $n_{front} = 212$, $n_{rear} = 108$, $n_{4x4} = 89$)

$\overline{Y} = 10.75$, $\overline{Y}_{front} = 9.74$, $\overline{Y}_{rear} = 11.29$, $\overline{Y}_{4x4} = 12.50$

```r
mSum <- lm(consumption ~ fdrive, data = CarsNow, contrasts = list(fdrive = contr.sum))
summary(mSum)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.17804    0.09606 116.365   <2e-16 ***
fdrive1     -1.43677    0.12003 -11.970   <2e-16 ***
fdrive2      0.11594    0.13926   0.833    0.406

Residual standard error: 1.815 on 406 degrees of freedom
Multiple R-squared:  0.2799,        Adjusted R-squared:  0.2764
F-statistic: 78.91 on 2 and 406 DF,  p-value: < 2.2e-16
```

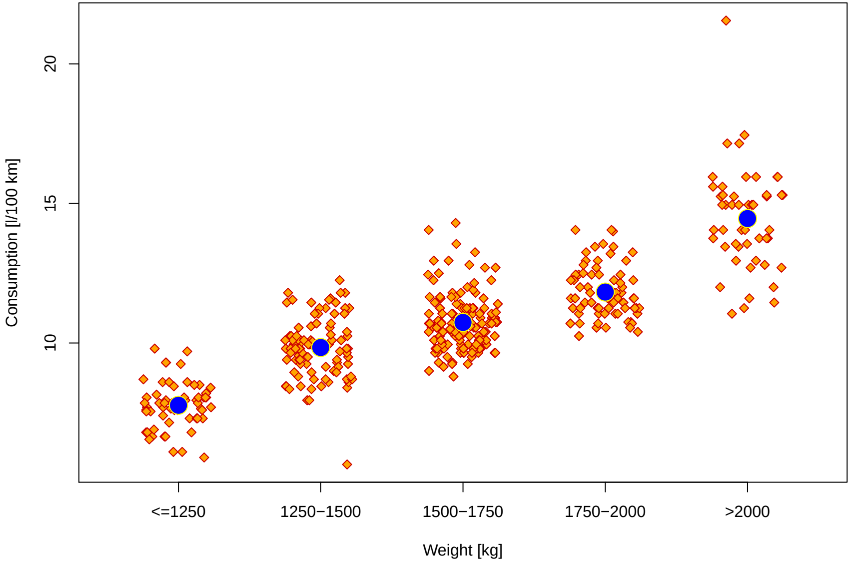
Values of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_3$

```r
alphaSum <- as.numeric(contr.sum(3) %*% coef(mSum)[-1])
names(alphaSum) <- levels(CarsNow[, "fdrive"])
print(alphaSum)
```

```
     front       rear        4x4
-1.4367702  0.1159377  1.3208326
```

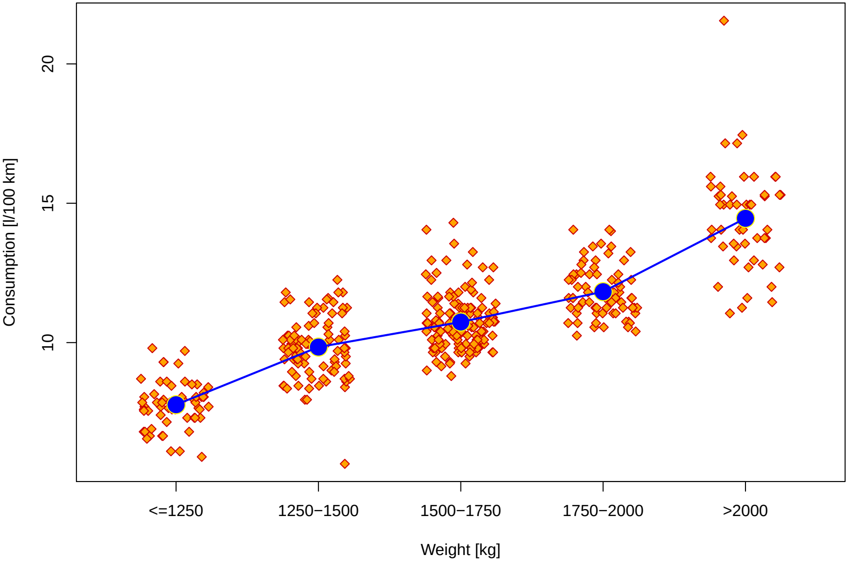
Cars2004nh (subset, $n = 409$, n's $= 57, 95, 137, 71, 49$)

consumption ∼ categorized weight



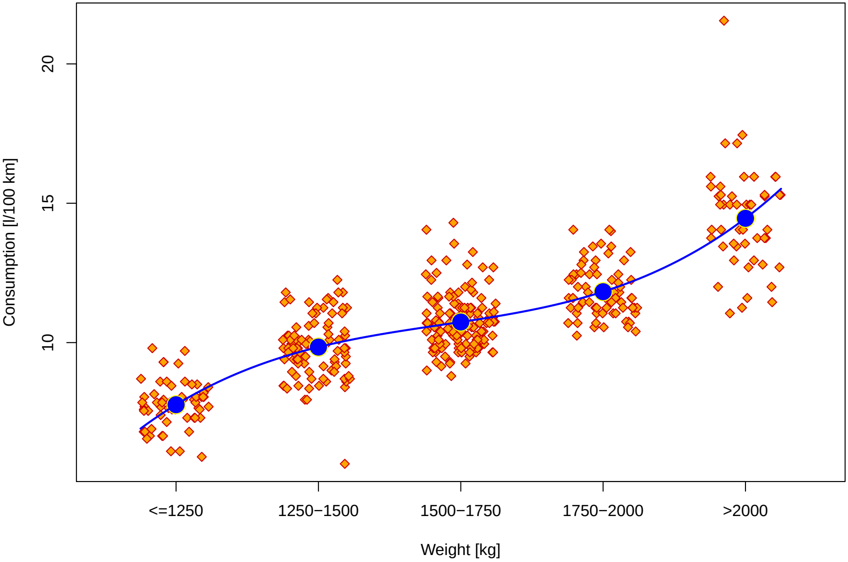
Cars2004nh (subset, $n = 409$, n's $= 57,\ 95,\ 137,\ 71,\ 49$)

$\overline{Y} = 10.75, \quad \overline{Y}_1 = 7.77, \overline{Y}_2 = 9.84, \overline{Y}_3 = 10.74, \overline{Y}_4 = 11.83, \overline{Y}_5 = 14.46$



Cars2004nh (subset, $n = 409$, n's $= 57,\ 95,\ 137,\ 71,\ 49$)

$\overline{Y} = 10.75, \quad \overline{Y}_1 = 7.77,\ \overline{Y}_2 = 9.84,\ \overline{Y}_3 = 10.74,\ \overline{Y}_4 = 11.83,\ \overline{Y}_5 = 14.46$



Orthonormal polynomial contrasts

ℂ: `contr.poly`, group means

$$
\begin{aligned}
m_1 &= m(\omega_1) = \beta_0 + \beta_1\,P^1(\omega_1) + \cdots + \beta_{G-1}\,P^{G-1}(\omega_1),\\
m_2 &= m(\omega_2) = \beta_0 + \beta_1\,P^1(\omega_2) + \cdots + \beta_{G-1}\,P^{G-1}(\omega_2),\\
&\vdots\\
m_G &= m(\omega_G) = \beta_0 + \beta_1\,P^1(\omega_G) + \cdots + \beta_{G-1}\,P^{G-1}(\omega_G),
\end{aligned}
$$

Orthonormal polynomial contrasts

$\mathbb{C}$: `contr.poly`

$$\mathbb{C} = \begin{pmatrix} P^1(\omega_1) & P^2(\omega_1) & \dots & P^{G-1}(\omega_1) \\ P^1(\omega_2) & P^2(\omega_2) & \dots & P^{G-1}(\omega_2) \\ \vdots & \vdots & \vdots & \vdots \\ P^1(\omega_G) & P^2(\omega_G) & \dots & P^{G-1}(\omega_G) \end{pmatrix},$$

- $\omega_1 < \cdots < \omega_G$:

  an *equidistant (arithmetic)* sequence of the group labels;

- $P^j(z) = a_{j,0} + a_{j,1}\,z + \cdots + a_{j,j}\,z^j, \qquad j = 1, \ldots, G-1$:

  orthonormal polynomials of degree $1, \ldots, G-1$ built above a sequence of the group labels.

Orthonormal polynomial contrasts

$\mathbb{C}$: `contr.poly`, examples

G = 2

$$\mathbb{C} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},$$

G = 3

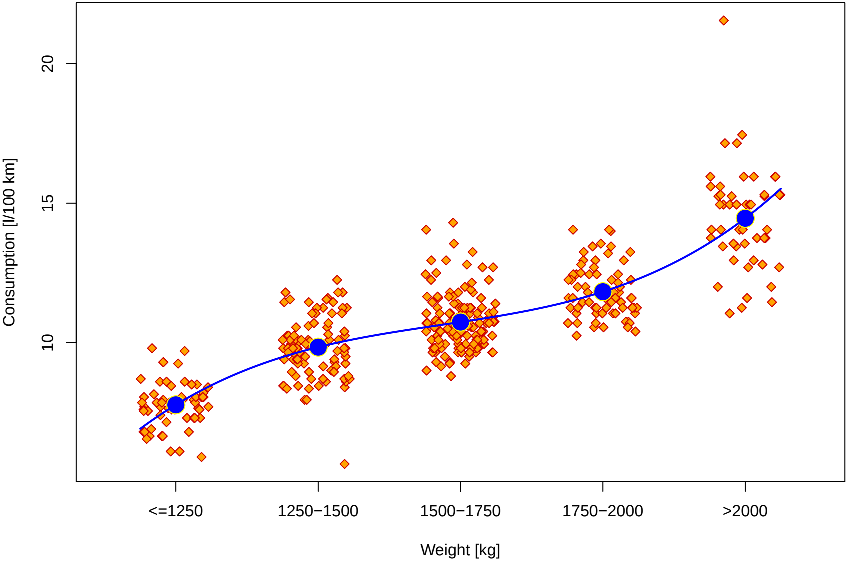
$$\mathbb{C} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix},$$

G = 4

$$\mathbb{C} = \begin{pmatrix} -\frac{3}{2\sqrt{5}} & \frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ -\frac{1}{2\sqrt{5}} & -\frac{1}{2} & \frac{3}{2\sqrt{5}} \\ \frac{1}{2\sqrt{5}} & -\frac{1}{2} & -\frac{3}{2\sqrt{5}} \\ \frac{3}{2\sqrt{5}} & \frac{1}{2} & \frac{1}{2\sqrt{5}} \end{pmatrix}.$$

Cars2004nh (subset, $n = 409$, n's $= 57,\ 95,\ 137,\ 71,\ 49$)

$\overline{Y} = 10.75, \quad \overline{Y}_1 = 7.77, \overline{Y}_2 = 9.84, \overline{Y}_3 = 10.74, \overline{Y}_4 = 11.83, \overline{Y}_5 = 14.46$



Cars2004nh (subset, $n = 409$, n's = 57, 95, 137, 71, 49)

$\overline{Y} = 10.75$, $\overline{Y}_1 = 7.77$, $\overline{Y}_2 = 9.84$, $\overline{Y}_3 = 10.74$, $\overline{Y}_4 = 11.83$, $\overline{Y}_5 = 14.46$

```
mTrt <- lm(consumption ~ fweight, data = CarsNow)
summary(mTrt)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.1900 -0.7102 -0.0400  0.6232  7.0898

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        7.7719     0.1497   51.91   <2e-16 ***
fweight1250-1500   2.0681     0.1894   10.92   <2e-16 ***
fweight1500-1750   2.9671     0.1782   16.65   <2e-16 ***
fweight1750-2000   4.0548     0.2010   20.17   <2e-16 ***
fweight>2000       6.6883     0.2202   30.37   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.13 on 404 degrees of freedom
Multiple R-squared:  0.7221,       Adjusted R-squared:  0.7193
F-statistic: 262.4 on 4 and 404 DF,  p-value: < 2.2e-16
```

```
summary(aov(consumption ~ fweight, data = CarsNow))
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
fweight       4 1341.0   335.3   262.4 <2e-16 ***
Residuals   404  516.2     1.3
```

`Cars2004nh` (subset, $n = 409$, n's $= 57, 95, 137, 71, 49$)

$\overline{Y} = 10.75$, $\overline{Y}_1 = 7.77$, $\overline{Y}_2 = 9.84$, $\overline{Y}_3 = 10.74$, $\overline{Y}_4 = 11.83$, $\overline{Y}_5 = 14.46$

```
mPoly <- lm(consumption ~ fweight, data = CarsNow,
            contrasts = list(fweight = contr.poly))
summary(mPoly)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.1900 -0.7102 -0.0400  0.6232  7.0898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.093e+01  5.975e-02 182.876  < 2e-16 ***
fweight.L    4.858e+00  1.501e-01  32.359  < 2e-16 ***
fweight.Q    3.526e-01  1.370e-01   2.574   0.0104 *
fweight.C    8.585e-01  1.320e-01   6.503 2.33e-10 ***
fweight^4   -7.193e-05  1.126e-01  -0.001   0.9995

Residual standard error: 1.13 on 404 degrees of freedom
Multiple R-squared:  0.7221,        Adjusted R-squared:  0.7193
F-statistic: 262.4 on 4 and 404 DF,  p-value: < 2.2e-16
```

```
summary(aov(consumption ~ fweight, data = CarsNow))
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
fweight       4 1341.0   335.3   262.4 <2e-16 ***
Residuals   404  516.2     1.3
```

`Cars2004nh` (subset, $n = 409$)

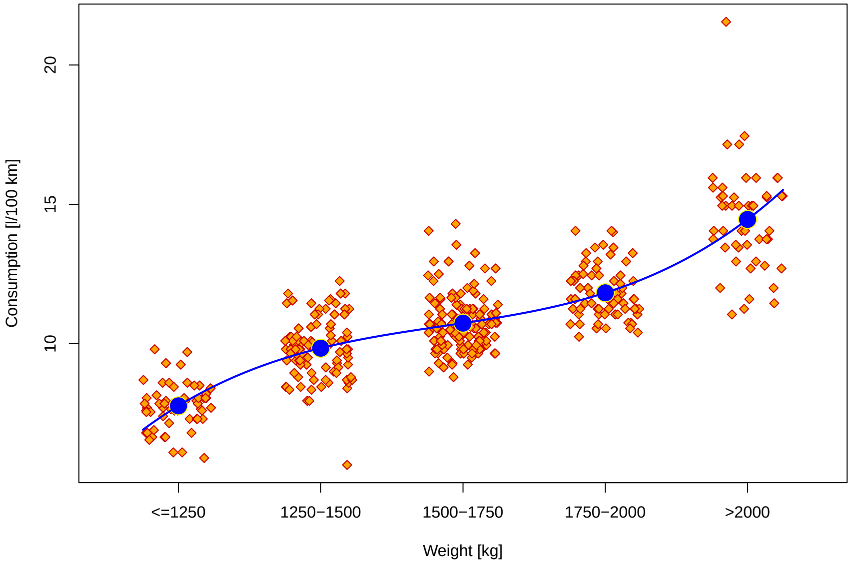
Polynomial of degree 4 based on representation of the covariate values by numbers 1, 2, 3, 4, 5, $m_g = \beta_0 + \beta_1\,g + \beta_2\,g^2 + \beta_3\,g^3 + \beta_4\,g^4$, $g = 1, \ldots, 5$

```
CarsNow <- transform(CarsNow, nweight = as.numeric(fweight))
p4 <- lm(consumption ~ nweight + I(nweight^2) + I(nweight^3) + I(nweight^4),
         data = CarsNow)
summary(p4)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.1900 -0.7102 -0.0400  0.6232  7.0898

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.177e+00  1.820e+00   1.745   0.0818 .
nweight       6.312e+00  3.274e+00   1.928   0.0546 .
I(nweight^2) -1.943e+00  1.947e+00  -0.998   0.3190
I(nweight^3)  2.265e-01  4.687e-01   0.483   0.6292
I(nweight^4) -2.507e-05  3.925e-02  -0.001   0.9995

Residual standard error: 1.13 on 404 degrees of freedom
Multiple R-squared:  0.7221,        Adjusted R-squared:  0.7193
F-statistic: 262.4 on 4 and 404 DF,  p-value: < 2.2e-16
```

Cars2004nh (subset, $n = 409$)

Is a **linear** trend adequate?



Cars2004nh (subset, $n = 409$)

Is a **linear** trend adequate?

```
p1 <- lm(consumption ~ nweight, data = CarsNow)
anova(p1, p4)
```

```
Analysis of Variance Table

Model 1: consumption ~ nweight
Model 2: consumption ~ nweight + I(nweight^2) + I(nweight^3) + I(nweight^4)
  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
1    407 577.49
2    404 516.20  3    61.291 15.99 7.667e-10 ***
```

```
anova(p1, mPoly)
```

```
Analysis of Variance Table

Model 1: consumption ~ nweight
Model 2: consumption ~ fweight
  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
1    407 577.49
2    404 516.20  3    61.291 15.99 7.667e-10 ***
```

5

Multiple Regression

Section **5.1**

Multiple covariates in a linear model

5.1.1 Additivity

Definition 5.1 Additivity of the covariate effect.

We say that a covariate Z_1 acts additively in the regression model with covariates $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top \in \mathcal{Z} \subseteq \mathbb{R}^p$ if the regression function is of the form

$$\mathbb{E}\big(Y \,\big|\, Z_1 = z_1, Z_2 = z_2, \ldots, Z_p = z_p\big) = m_1(z_1) + m_2(\boldsymbol{z}_{(-1)}),$$

where $\boldsymbol{z}_{(-1)} = (z_2, \ldots, z_p)^\top$, $m_1 : \mathbb{R} \longrightarrow \mathbb{R}$ and $m_2 : \mathbb{R}^{p-1} \longrightarrow \mathbb{R}$ are some measurable functions.

5.1.2 Interactions

Definition 5.2 Interaction terms.

Let $(Z,\, W)^\top \in \mathcal{Z} \times \mathcal{W} \subseteq \mathbb{R}^2$ be two covariates being parameterized using parameterizations $\boldsymbol{s}_Z : \mathcal{Z} \longrightarrow \mathbb{R}^{k-1}$ ($\boldsymbol{s}_Z = \big(s_Z^1,\, \dots,\, s_Z^{k-1}\big)^\top$) and $\boldsymbol{s}_W : \mathcal{W} \longrightarrow \mathbb{R}^{l-1}$ ($\boldsymbol{s}_W = \big(s_W^1,\, \dots,\, s_W^{l-1}\big)^\top$). By interaction terms based on those two parameterizations we mean elements of a vector

$$\begin{aligned}\boldsymbol{s}_{ZW}(Z,\, W) \;:=\; & \boldsymbol{s}_W^\top(W) \otimes \boldsymbol{s}_Z^\top(Z)\\ &= \big(s_Z^1(Z)\cdot s_W^1(W), \dots, s_Z^{k-1}(Z)\cdot s_W^1(W),\, \dots,\\ &\qquad s_Z^1(Z)\cdot s_W^{l-1}(W), \dots, s_Z^{k-1}(Z)\cdot s_W^{l-1}(W)\big)^\top.\end{aligned}$$

Section **5.2**

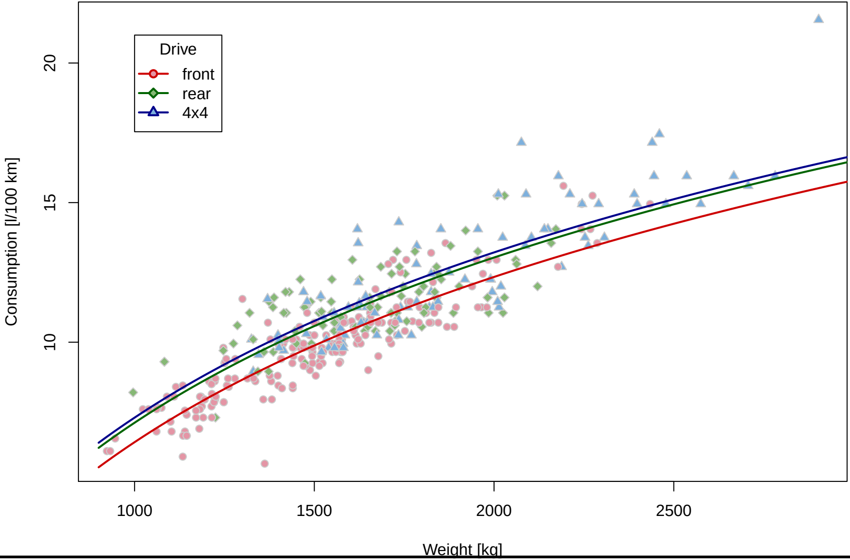
Numeric and categorical covariate

5.2.1 Additivity

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight)`,

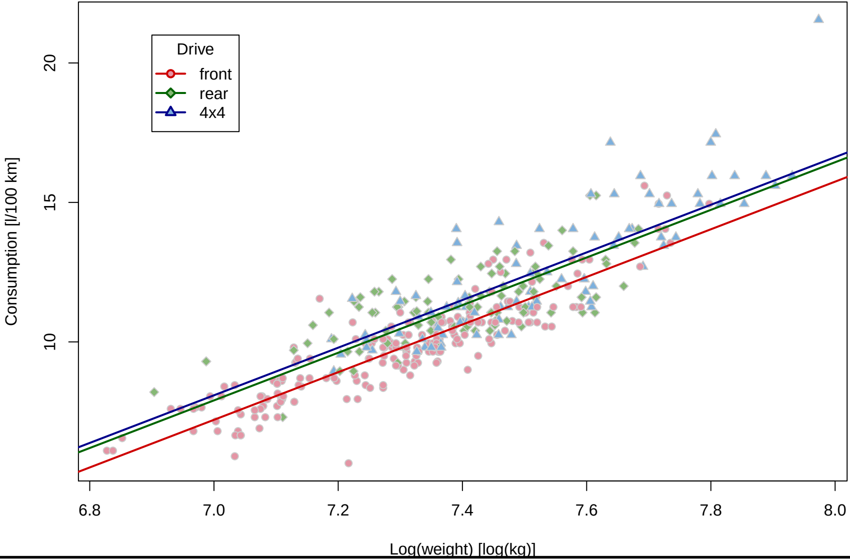
$$\widehat{m}(z, w) = -52.56 + 0.70\,\mathbb{I}[z = \text{rear}] + 0.88\mathbb{I}[z = \text{4x4}] + 8.54\,\log(w)$$



Cars2004nh (subset, $n = 409$)

`consumption ~ drive + log(weight)`,

$$\widehat{m}(z, w) = -52.56 + 0.70\,\mathbb{I}[z = \text{rear}] + 0.88\mathbb{I}[z = \text{4x4}] + 8.54\,\log(w)$$



Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight), contr.treatment param. of drive

Y: consumption *[l/100 km]*, Z: drive, W: weight *[kg]*

$$m(z, w) = \beta_0 + \beta_1^Z \, \mathbb{I}[z = \text{rear}] + \beta_2^Z \, \mathbb{I}[z = \text{4x4}] + \beta^W \log(w)$$

```
lm(consumption ~ fdrive + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.4064 -0.6649 -0.1323  0.5747  5.1533

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -52.5605     1.9627 -26.780  < 2e-16 ***
fdriverear    0.6964     0.1181   5.897 7.83e-09 ***
fdrive4x4     0.8787     0.1363   6.445 3.29e-10 ***
lweight       8.5381     0.2688  31.762  < 2e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 0.9726 on 405 degrees of freedom
Multiple R-squared:  0.7937, Adjusted R-squared:  0.7922
F-statistic: 519.5 on 3 and 405 DF,  p-value: < 2.2e-16
```

`consumption ~ drive + log(weight)`, `contr.sum` param. of `drive`

Y: consumption *[l/100 km]*, Z: drive, W: weight *[kg]*

$$m(z, w) = \beta_0 + \beta_1^Z \,\mathbb{I}[z = \text{front}] + \beta_2^Z \,\mathbb{I}[z = \text{rear}] - (\beta_1^Z + \beta_2^Z)\,\mathbb{I}[z = \text{4x4}] + \beta^W \log(w)$$

```
lm(consumption ~ fdrive + lweight, data = CarsNow,
   contrasts = list(fdrive = "contr.sum"))
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.4064 -0.6649 -0.1323  0.5747  5.1533

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -52.03547    1.99090 -26.137  < 2e-16 ***
fdrive1      -0.52504    0.07044  -7.454 5.53e-13 ***
fdrive2       0.17134    0.07465   2.295   0.0222 *
lweight       8.53810    0.26882  31.762  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9726 on 405 degrees of freedom
Multiple R-squared:  0.7937, Adjusted R-squared:  0.7922
F-statistic: 519.5 on 3 and 405 DF,  p-value: < 2.2e-16
```

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight)`, `contr.sum` param. of `drive`

Y: consumption *[l/100 km]*, Z: drive, W: weight *[kg]*

$$m(z, w) = \beta_0 + \beta_1^Z \,\mathbb{I}[z = \text{front}] + \beta_2^Z \,\mathbb{I}[z = \text{rear}] - (\beta_1^Z + \beta_2^Z)\,\mathbb{I}[z = \text{4x4}] + \beta^W \log(w)$$

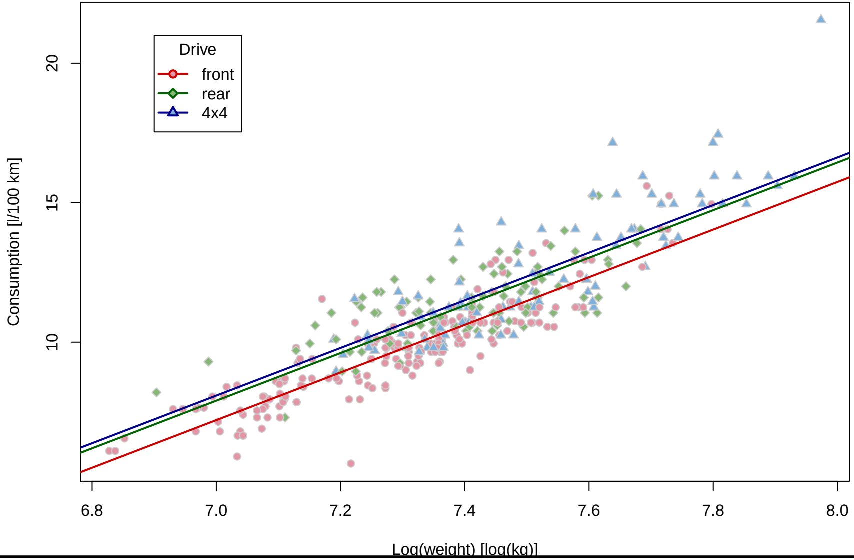
Estimates of parameters $\alpha_1^Z = \beta_1^Z$, $\alpha_2^Z = \beta_2^Z$, $\alpha_3^Z = -\beta_1^Z - \beta_2^Z$

	Estimate	Std. Error	t value	P value	Lower	Upper
front	-0.5250404	0.07043545	-7.454206	5.5325e-13	-0.66350509	-0.3865756
rear	0.1713353	0.07464863	2.295224	0.022231	0.02458813	0.3180824
4x4	0.3537051	0.08437896	4.191864	3.3999e-05	0.18782965	0.5195805

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight)`,

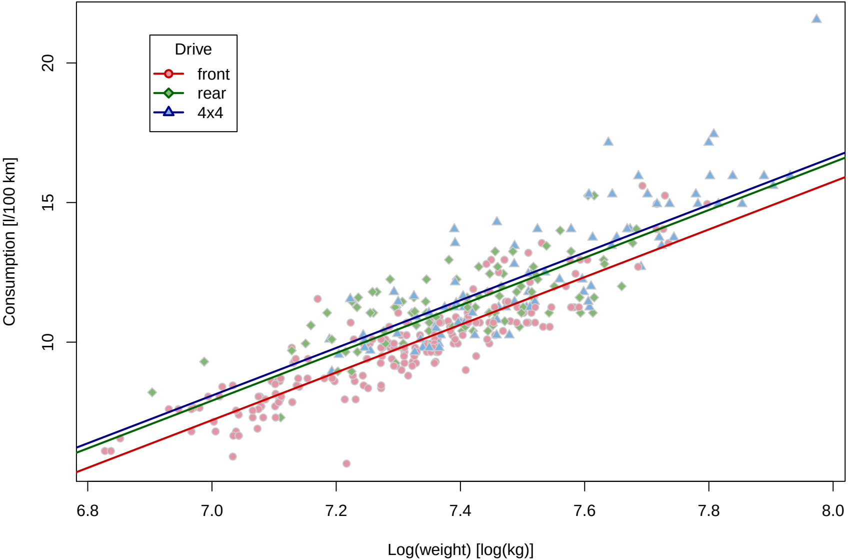
$$\widehat{m}(z, w) = -52.04 - 0.53\,\mathbb{I}[z = \text{front}] + 0.17\mathbb{I}[z = \text{rear}] + 0.35\mathbb{I}[z = \text{4x4}] + 8.54\,\log(w$$



5.2.2 Partial effects

Cars2004nh (subset, $n = 409$)

`consumption ~ drive + log(weight)`, **partial effect of log(`weight`)?**



`consumption ~ drive + log(weight)`

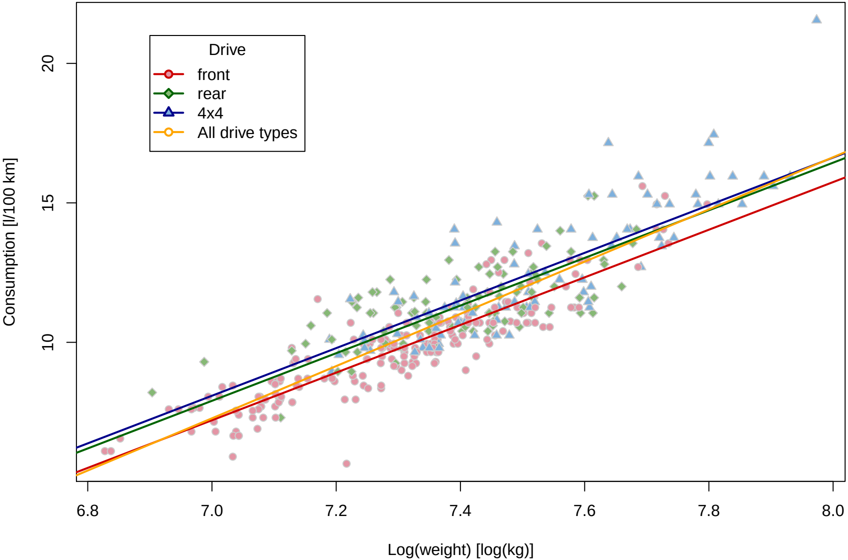
For a given `drive`, does the log(`weight`) have an effect on the mean consumption? Partial effect of log(`weight`)

```
lm(consumption ~ fdrive + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-3.4064 -0.6649 -0.1323  0.5747  5.1533 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -52.5605     1.9627 -26.780  < 2e-16 ***
fdriverear    0.6964     0.1181   5.897 7.83e-09 ***
fdrive4x4     0.8787     0.1363   6.445 3.29e-10 ***
lweight       8.5381     0.2688  31.762  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9726 on 405 degrees of freedom
Multiple R-squared:  0.7937, Adjusted R-squared:  0.7922 
F-statistic: 519.5 on 3 and 405 DF,  p-value: < 2.2e-16
```

Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight), **partial effect of drive?**



consumption ∼ drive + log(weight)

Analysis of covariance to evaluate effect of drive given log(weight)

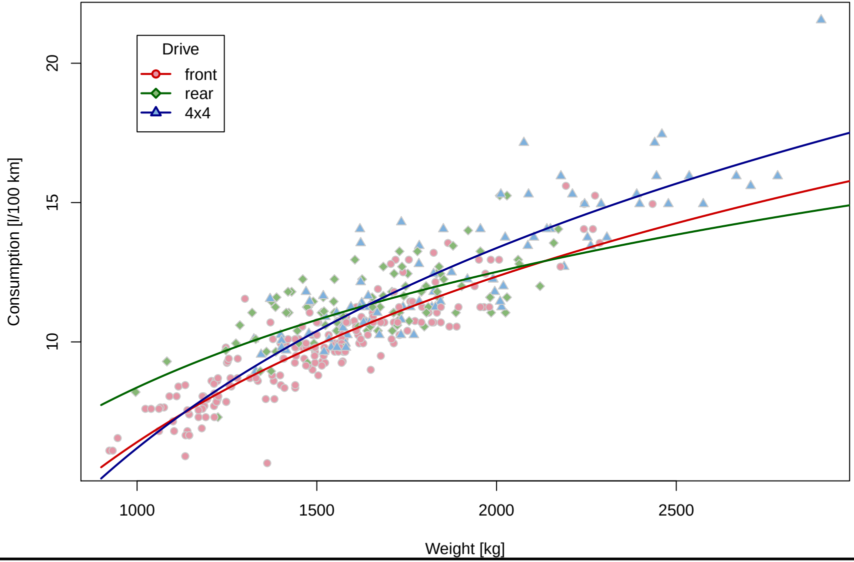
```
mAddit <- lm(consumption ~ fdrive + lweight, data = CarsNow)
mOneLine <- lm(consumption ~ lweight, data = CarsNow)
anova(mOneLine, mAddit)
```

```
Analysis of Variance Table

Model 1: consumption ~ lweight
Model 2: consumption ~ fdrive + lweight
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    407 435.68
2    405 383.10  2    52.577 27.791 4.896e-12 ***
```

Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight) + drive:log(weight),

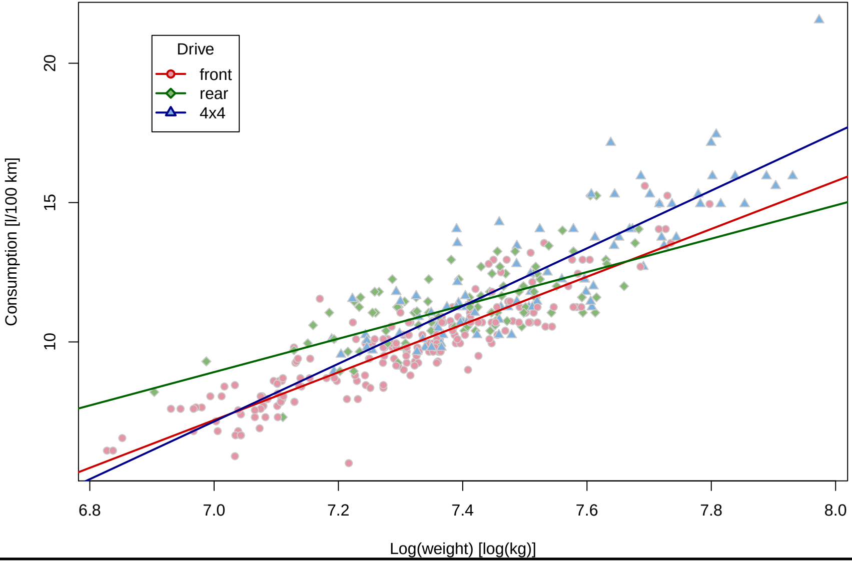
$\widehat{m}(z, w) = -52.80 + 19.84\,\mathbb{I}[z = \text{rear}] - 12.54\,\mathbb{I}[z = \text{4x4}] + 8.57\,\log(w) - 2.59\,\mathbb{I}[z = \text{rear}$



Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight) + drive:log(weight),

$\widehat{m}(z, w) = -52.80 + 19.84\,\mathbb{I}[z = \text{rear}] - 12.54\,\mathbb{I}[z = 4x4] + 8.57\,\log(w) - 2.59\,\mathbb{I}[z = \text{rear}$



consumption ~ drive + log(weight) + drive:log(weight), contr.treatment

param. of drive

Reference group pseudocontrasts for drive

$$m(z, w) = \beta_0 + \beta_1^Z\,\mathbb{I}[z = \text{rear}] + \beta_2^Z\,\mathbb{I}[z = \text{4x4}] + \beta^W\,\log(w) + \beta_1^{ZW}\,\mathbb{I}[z = \text{rear}]\,\log(w) + \beta_2^{ZW}\,\mathbb{I}[z = \text{4x4}]\,\log(w)$$

```
lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
```

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -52.8047     2.5266 -20.900  < 2e-16 ***
fdriverear          19.8445     5.1297   3.869 0.000128 ***
fdrive4x4          -12.5366     4.6506  -2.696 0.007319 **
lweight              8.5716     0.3461  24.763  < 2e-16 ***
fdriverear:lweight  -2.5890     0.6956  -3.722 0.000226 ***
fdrive4x4:lweight    1.7837     0.6240   2.858 0.004480 **
---

Residual standard error: 0.9404 on 403 degrees of freedom
Multiple R-squared:  0.8081, Adjusted R-squared:  0.8057
F-statistic: 339.4 on 5 and 403 DF,  p-value: < 2.2e-16
```

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight) + drive:log(weight)`, <u>`contr.sum`</u> param. of `drive`

Sum contrasts for `drive`

$$m(z,\, w) = \beta_0 + \beta_1^Z\, \mathbb{I}[z = \text{front}] + \beta_2^Z\, \mathbb{I}[z = \text{rear}] - (\beta_1^Z + \beta_2^Z)\, \mathbb{I}[z = \text{4x4}] + \beta^W\, \log(w)$$

$$+ \beta_1^{ZW}\, \mathbb{I}[z = \text{front}]\, \log(w) + \beta_2^{ZW}\, \mathbb{I}[z = \text{rear}]\, \log(w) - (\beta_1^{ZW} + \beta_2^{ZW})\, \mathbb{I}[z = \text{4x4}]\, \log(w)$$

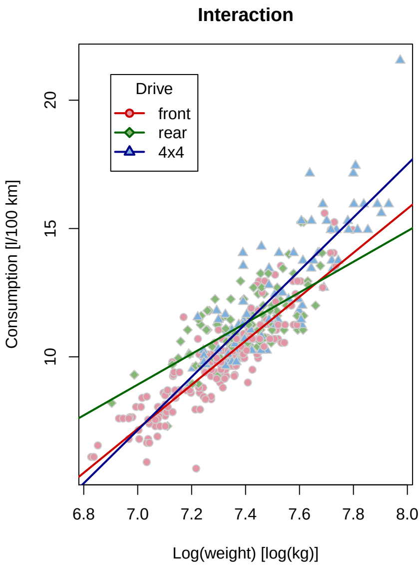
```
lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow,
   contrasts = list(fdrive = contr.sum))
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -50.3688     2.1489 -23.440  < 2e-16 ***
fdrive1          -2.4360     2.5972  -0.938    0.349
fdrive2          17.4085     3.3558   5.188 3.38e-07 ***
lweight           8.3031     0.2894  28.696  < 2e-16 ***
fdrive1:lweight   0.2684     0.3517   0.763    0.446
fdrive2:lweight  -2.3206     0.4529  -5.124 4.64e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9404 on 403 degrees of freedom
Multiple R-squared:  0.8081, Adjusted R-squared:  0.8057
F-statistic: 339.4 on 5 and 403 DF,  p-value: < 2.2e-16
```

5.2.4 Additivity or interactions?

Cars2004nh (subset, $n = 409$)

consumption ~ drive, log(weight), additivity or interactions?



consumption $\sim$ drive, log(weight), **additivity or interactions?**

Does the log(weight) have different effect on the mean consumption depending on the drive type?

```
mInter <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
mAddit <- lm(consumption ~ fdrive + lweight, data = CarsNow)
anova(mAddit, mInter)
```

```
Analysis of Variance Table

Model 1: consumption ~ fdrive + lweight
Model 2: consumption ~ fdrive + lweight + fdrive:lweight
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    405 383.1
2    403 356.4  2    26.702 15.097 4.758e-07 ***
```

Section **5.3**

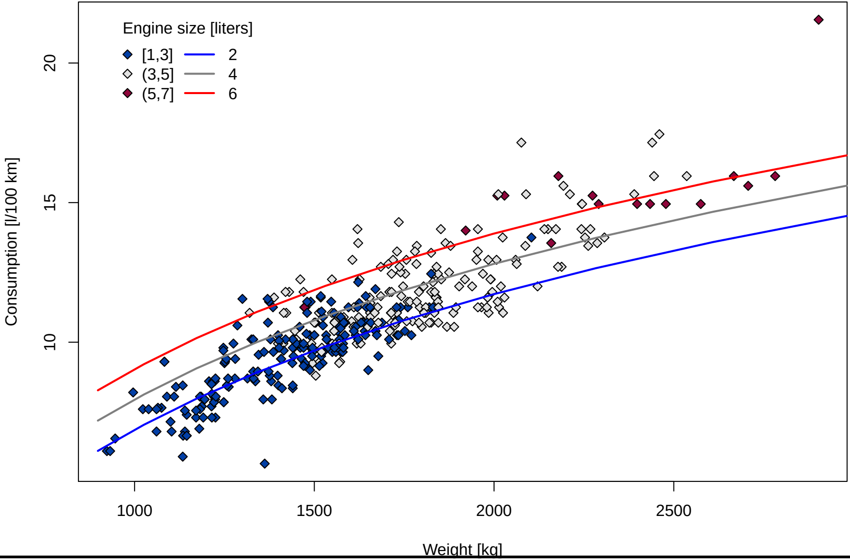
Two numeric covariates

5.3.1 Additivity

Cars2004nh (subset, $n = 409$)

`consumption ~ engine.size + log(weight)`,

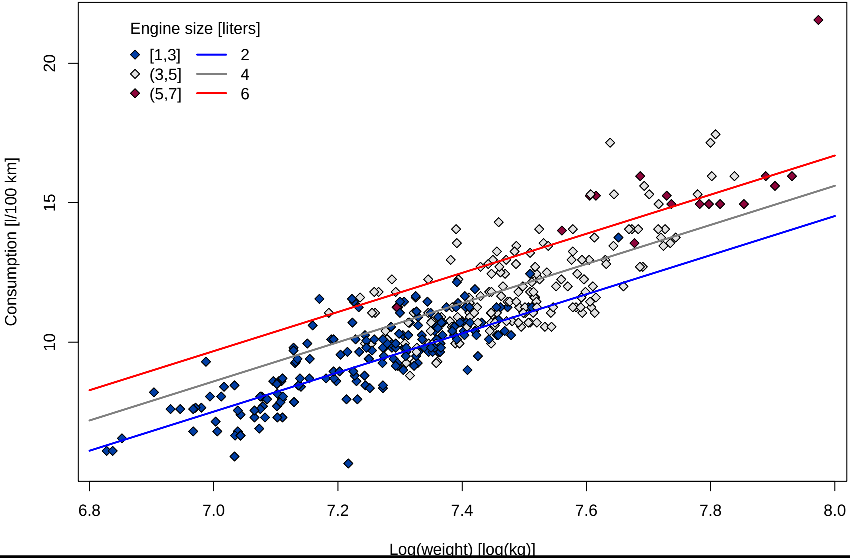
$$\widehat{m}(z, w) = -42.65 + 0.54\,z + 7.01\,\log(w)$$



Cars2004nh (subset, $n = 409$)

`consumption ~ engine.size + log(weight)`,

$$\widehat{m}(z, w) = -42.65 + 0.54\, z + 7.01\, \log(w)$$



Cars2004nh (subset, $n = 409$)

```
consumption ~ engine.size + log(weight)
```

Y: consumption *[l/100 km]*, *Z*: engine size *[l]*, *W*: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w)$$

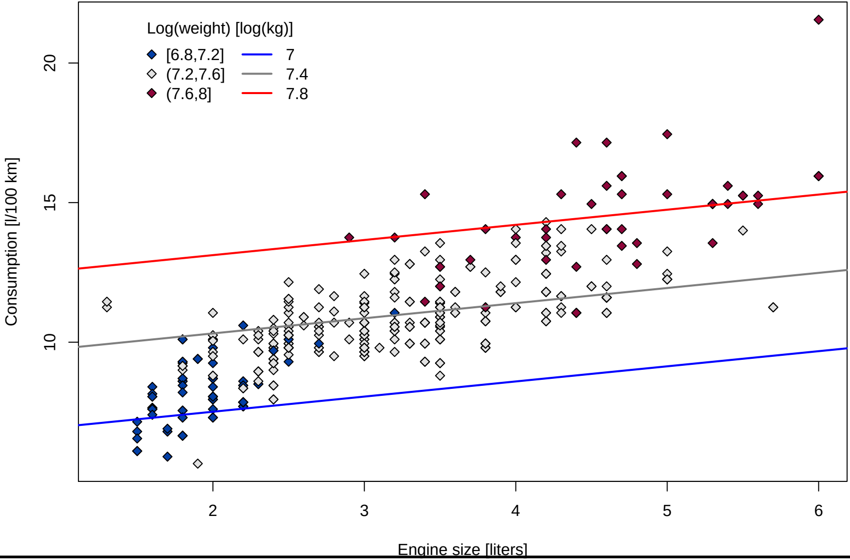
```
lm(consumption ~ engine.size + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.3243 -0.6741 -0.1286  0.5270  5.0459

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.65641    2.99243 -14.255  < 2e-16 ***
engine.size   0.54231    0.08304   6.531 1.96e-10 ***
lweight       7.01155    0.43501  16.118  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9854 on 406 degrees of freedom
Multiple R-squared:  0.7877, Adjusted R-squared:  0.7867
F-statistic: 753.3 on 2 and 406 DF,  p-value: < 2.2e-16
```

Cars2004nh (subset, $n = 409$)

consumption $\sim$ engine.size + log(weight),

$$\widehat{m}(z, w) = -42.65 + 0.54\,z + 7.01\,\log(w)$$



`Cars2004nh` (subset, $n = 409$)

`consumption` ∼ `engine.size + log(weight)`

Y: consumption *[l/100 km]*, Z: engine size *[l]*, W: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w)$$

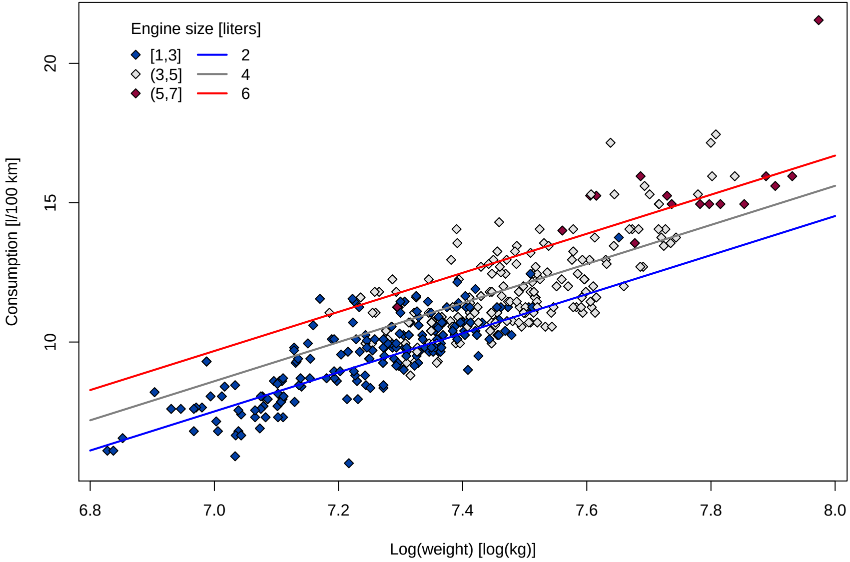
```
lm(consumption ~ engine.size + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-3.3243 -0.6741 -0.1286  0.5270  5.0459 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -42.65641    2.99243 -14.255  < 2e-16 ***
engine.size   0.54231    0.08304   6.531 1.96e-10 ***
lweight       7.01155    0.43501  16.118  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9854 on 406 degrees of freedom
Multiple R-squared:  0.7877, Adjusted R-squared:  0.7867 
F-statistic: 753.3 on 2 and 406 DF,  p-value: < 2.2e-16
```

5.3.2 Partial effects

Cars2004nh (subset, $n = 409$)

`consumption ~ engine.size + log(weight)`, **partial effect of log(`weight`)?**



`Cars2004nh` (subset, $n = 409$)

`consumption ~ engine.size + log(weight)`

Y: consumption *[l/100 km]*, *Z*: engine size *[l]*, *W*: weight *[kg]*

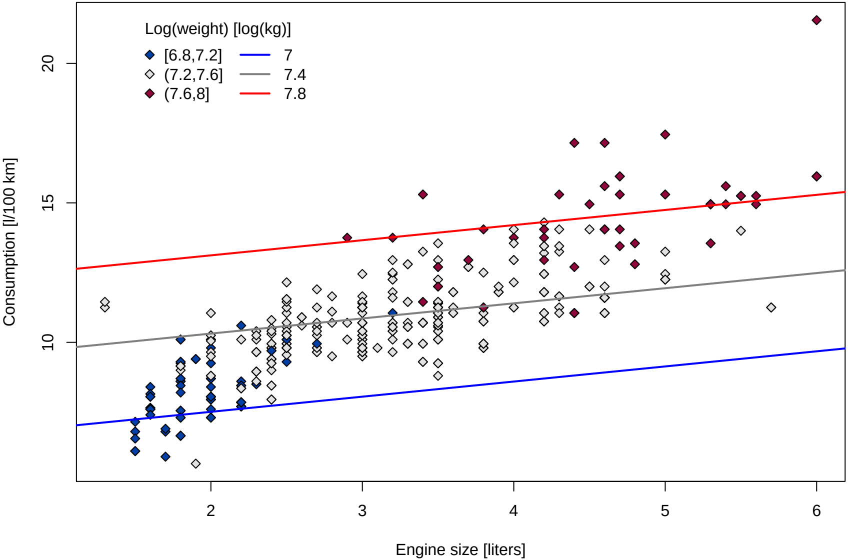
$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w)$$

```
lm(consumption ~ engine.size + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-3.3243 -0.6741 -0.1286  0.5270  5.0459 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -42.65641    2.99243 -14.255  < 2e-16 ***
engine.size   0.54231    0.08304   6.531 1.96e-10 ***
lweight       7.01155    0.43501  16.118  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9854 on 406 degrees of freedom
Multiple R-squared:  0.7877, Adjusted R-squared:  0.7867
F-statistic: 753.3 on 2 and 406 DF,  p-value: < 2.2e-16
```

Cars2004nh (subset, $n = 409$)

`consumption ~ engine.size + log(weight)`, **partial effect of engine.size?**



consumption ~ engine.size + log(weight)

Y: consumption *[l/100 km]*, Z: engine size *[l]*, W: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w)$$

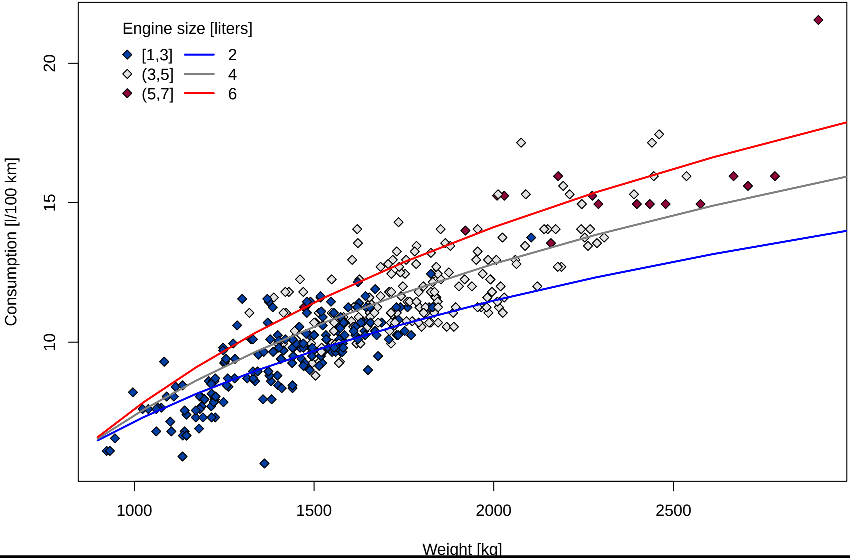
```
lm(consumption ~ engine.size + lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.3243 -0.6741 -0.1286  0.5270  5.0459

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.65641    2.99243 -14.255  < 2e-16 ***
engine.size   0.54231    0.08304   6.531 1.96e-10 ***
lweight       7.01155    0.43501  16.118  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9854 on 406 degrees of freedom
Multiple R-squared:  0.7877, Adjusted R-squared:  0.7867
F-statistic: 753.3 on 2 and 406 DF,  p-value: < 2.2e-16
```

5.3.3 Interactions

Cars2004nh (subset, $n = 409$)

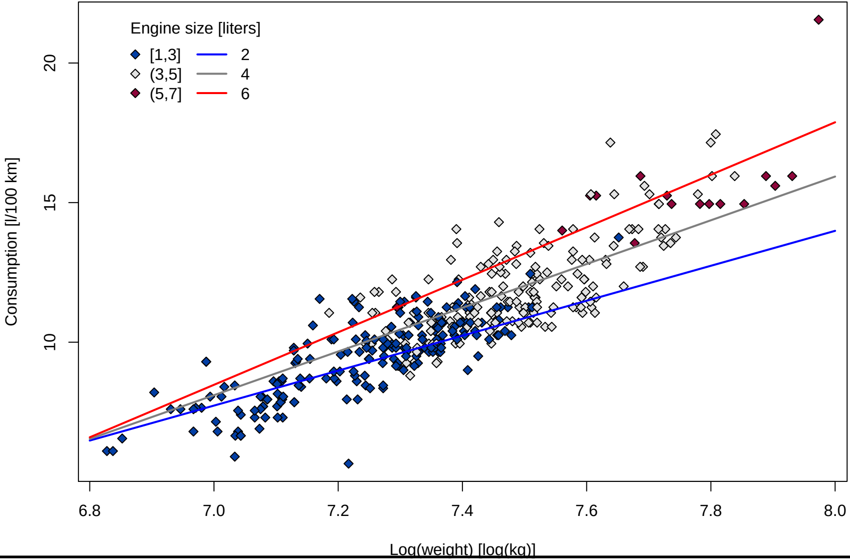
consumption ~ engine.size + log(weight) + engine.size:log(weight),
$\widehat{m}(z, w) = -25.46 - 5.32\,z + 4.69\,\log(w) + 0.79\,z\,\log(w)$



Cars2004nh (subset, $n = 409$)

consumption ~ engine.size + log(weight) + engine.size:log(weight),

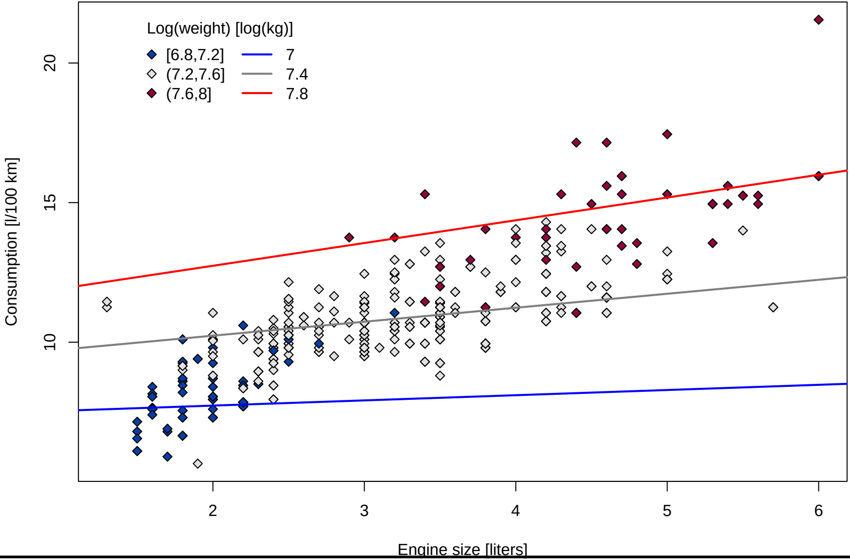
$$\widehat{m}(z, w) = -25.46 - 5.32\, z + 4.69\, \log(w) + 0.79\, z\, \log(w)$$



Cars2004nh (subset, $n = 409$)

consumption ~ engine.size + log(weight) + engine.size:log(weight),

$\widehat{m}(z, w) = -25.46 - 5.32\,z + 4.69\,\log(w) + 0.79\,z\,\log(w)$



Cars2004nh (subset, $n = 409$)

consumption ~ engine.size + log(weight) + engine.size:log(weight)

Y: consumption *[l/100 km]*, *Z*: engine size *[l]*, *W*: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w) + \beta^{ZW} z \log(w)$$

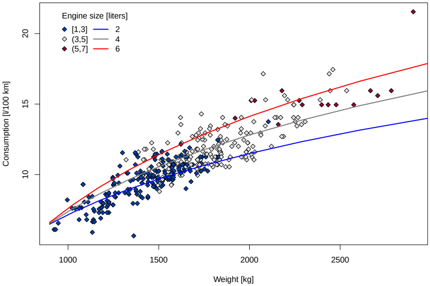
```
lm(consumption ~ engine.size + lweight + engine.size:lweight, data = CarsNow)
```

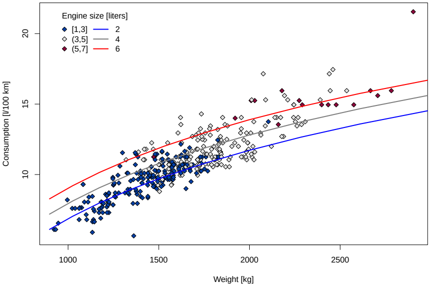
```
Residuals:
    Min      1Q  Median      3Q     Max
-3.3999 -0.6538 -0.1407  0.4779  3.9219

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -25.4574     5.1267  -4.966 1.01e-06 ***
engine.size          -5.3160     1.4338  -3.708 0.000238 ***
lweight               4.6877     0.7104   6.599 1.30e-10 ***
engine.size:lweight   0.7860     0.1921   4.092 5.15e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9669 on 405 degrees of freedom
Multiple R-squared:  0.7961,       Adjusted R-squared:  0.7946
F-statistic: 527.2 on 3 and 405 DF,  p-value: < 2.2e-16
```

5.3.4 Additivity or interactions?

Cars2004nh (subset, $n = 409$)

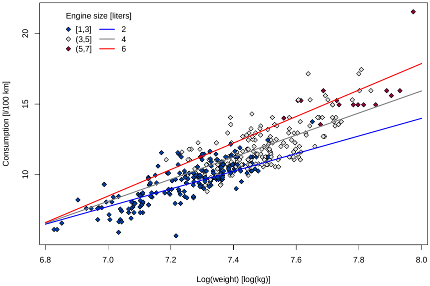
consumption $\sim$ engine.size, log(weight), **additivity or interactions?**

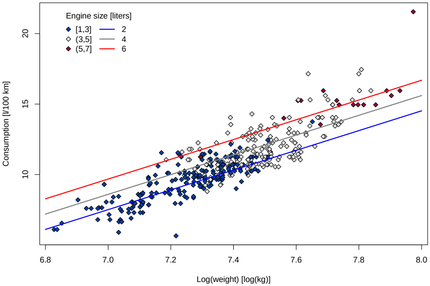




Cars2004nh (subset, $n = 409$)

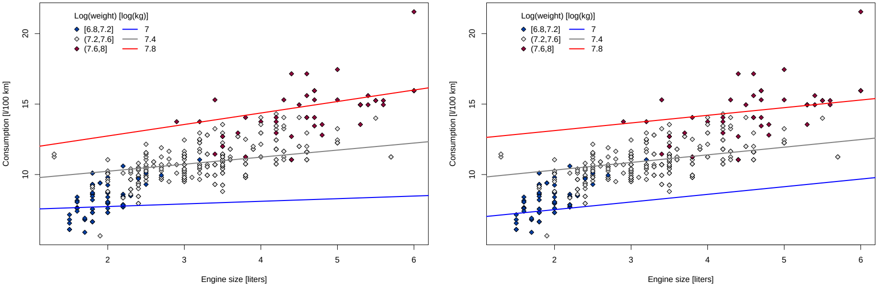
consumption ∼ engine.size, log(weight), **additivity or interactions?**





Cars2004nh (subset, $n = 409$)

consumption ∼ engine.size, log(weight), additivity or interactions?



Cars2004nh (subset, $n = 409$)

`consumption ~ engine.size + log(weight) + engine.size:log(weight)`

Y: consumption *[l/100 km]*, *Z*: engine size *[l]*, *W*: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w) + \beta^{ZW} z \log(w)$$

Does the [log]weight have different effect on the mean consumption depending on the engine size?

Does the engine size have different effect on the mean consumption depending on the [log]weight?

```
lm(consumption ~ engine.size + lweight + engine.size:lweight, data = CarsNow)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-3.3999 -0.6538 -0.1407  0.4779  3.9219 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -25.4574     5.1267  -4.966 1.01e-06 ***
engine.size          -5.3160     1.4338  -3.708 0.000238 ***
lweight               4.6877     0.7104   6.599 1.30e-10 ***
engine.size:lweight   0.7860     0.1921   4.092 5.15e-05 ***

...
```

`Cars2004nh` (subset, $n = 409$)

`consumption ~ engine.size + log(weight) + engine.size:log(weight)`

Y: consumption *[l/100 km]*, *Z*: engine size *[l]*, *W*: weight *[kg]*

$$m(z, w) = \beta_0 + \beta^Z z + \beta^W \log(w) + \beta^{ZW} z \log(w)$$

Does the [log]`weight` have different effect on the mean consumption depending on the `engine size`?

Does the `engine size` have different effect on the mean consumption depending on the [log]`weight`?

```
mAddit  <- lm(consumption ~ engine.size + lweight, data = CarsNow)
mInter  <- lm(consumption ~ engine.size*lweight, data = CarsNow)
anova(mAddit, mInter)
```

```
Analysis of Variance Table

Model 1: consumption ~ engine.size + lweight
Model 2: consumption ~ engine.size * lweight
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    406 394.26
2    405 378.60  1    15.656 16.748 5.154e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5.3.5 More complex parameterization of either covariate

Section **5.4**

Two categorical covariates

HowelsAll (subset, $n = 289$)

Covariates: gender ($G = 2$) and population ($H = 3$)

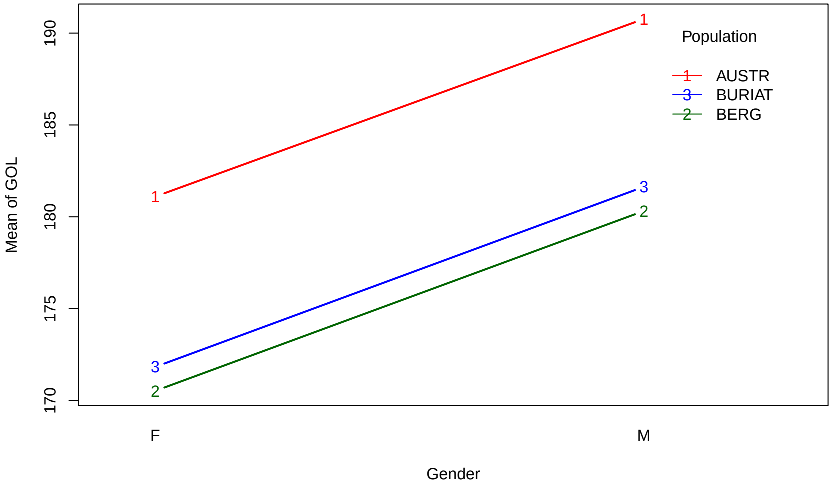
```
data(HowellsAll, package = "mffSM")
```

```
    gender popul oca gol fgender fpopul fgen.pop fpop.gen
1        1     1 123 176       M   BERG   M:BERG   BERG:M
2        1     1 115 173       M   BERG   M:BERG   BERG:M
3        1     1 117 176       M   BERG   M:BERG   BERG:M
4        1     1 113 185       M   BERG   M:BERG   BERG:M
...
57       0     1 125 171       F   BERG   F:BERG   BERG:F
58       0     1 103 178       F   BERG   F:BERG   BERG:F
59       0     1 115 165       F   BERG   F:BERG   BERG:F
60       0     1 117 169       F   BERG   F:BERG   BERG:F
...
110      1     0 109 194       M  AUSTR  M:AUSTR  AUSTR:M
112      1     0 115 188       M  AUSTR  M:AUSTR  AUSTR:M
116      1     0 115 187       M  AUSTR  M:AUSTR  AUSTR:M
117      1     0 109 196       M  AUSTR  M:AUSTR  AUSTR:M
...
192      0     0 109 186       F  AUSTR  F:AUSTR  AUSTR:F
193      0     0 115 175       F  AUSTR  F:AUSTR  AUSTR:F
194      0     0 111 185       F  AUSTR  F:AUSTR  AUSTR:F
195      0     0 113 184       F  AUSTR  F:AUSTR  AUSTR:F
...
241      1     2 118 180       M BURIAT M:BURIAT BURIAT:M
242      1     2 124 180       M BURIAT M:BURIAT BURIAT:M
243      1     2 117 183       M BURIAT M:BURIAT BURIAT:M
244      1     2 116 174       M BURIAT M:BURIAT BURIAT:M
...
295      0     2 116 175       F BURIAT F:BURIAT BURIAT:F
296      0     2 122 174       F BURIAT F:BURIAT BURIAT:F
297      0     2 113 174       F BURIAT F:BURIAT BURIAT:F
298      0     2 123 168       F BURIAT F:BURIAT BURIAT:F
...
```

5.4.1 Additivity

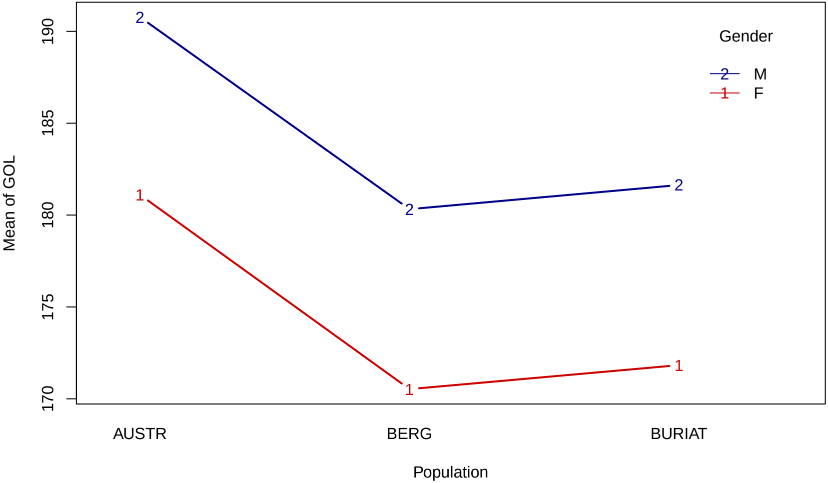
HowellsAll ($n = 289$)

gol (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$)



`HowellsAll` ($n = 289$)

`gol` (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$)



`HowelsAll` (subset, $n = 289$)

`gol ~ gender + popul`, `contr.treatment` parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z, w) = \beta_0 + \beta^Z \, \mathbb{I}[z = \text{male}] + \beta_1^W \, \mathbb{I}[w = \text{Berg}] + \beta_2^W \, \mathbb{I}[w = \text{Burjati}]$$

```
lm(gol ~ fgender + fpopul, data = HowellsAll)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-15.5400  -4.3103  -0.3103   4.4600  17.6897

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 181.0712     0.7814 231.724   <2e-16 ***
fgenderM      9.7703     0.7529  12.977   <2e-16 ***
fpopulBERG  -10.5311     0.9706 -10.850   <2e-16 ***
fpopulBURIAT  -9.2213     0.9695  -9.511   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.284 on 285 degrees of freedom
Multiple R-squared:  0.4729,	Adjusted R-squared:  0.4674
F-statistic: 85.24 on 3 and 285 DF,  p-value: < 2.2e-16
```

gol ~ gender + popul, contr.sum parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z,\, w) = \beta_0 + \beta^Z\, \mathbb{I}[z = \text{female}] - \beta^Z\, \mathbb{I}[z = \text{male}]$$
$$+ \beta_1^W\, \mathbb{I}[w = \text{Austr}] + \beta_2^W\, \mathbb{I}[w = \text{Berg}] + (-\beta_1^W - \beta_2^W)\, \mathbb{I}[w = \text{Burjati}]$$

```
options(contrasts = c("contr.sum", "contr.sum"))
lm(gol ~ fgender + fpopul, data = HowellsAll)
```

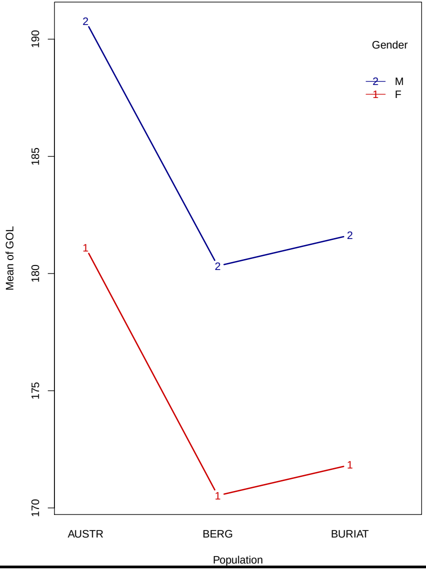
```
Residuals:
     Min       1Q   Median       3Q      Max
-15.5400  -4.3103  -0.3103   4.4600  17.6897

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 179.3722     0.3797 472.421  < 2e-16 ***
fgender1     -4.8852     0.3765 -12.977  < 2e-16 ***
fpopul1       6.5842     0.5811  11.330  < 2e-16 ***
fpopul2      -3.9470     0.5157  -7.654 3.03e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.284 on 285 degrees of freedom
Multiple R-squared:  0.4729,       Adjusted R-squared:  0.4674
F-statistic: 85.24 on 3 and 285 DF,  p-value: < 2.2e-16
```

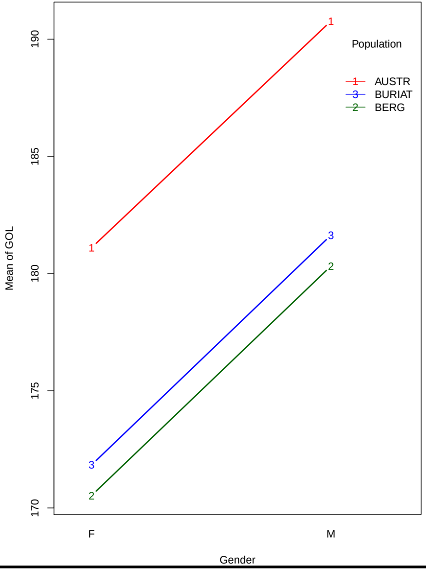
5.4.2 Partial effects

HowellsAll ($n = 289$)

gol (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$),

partial effect of gender, of population?





gol ∼ gender + popul

For a given population,
does gender **have an effect in the mean value of** gol**?**
Partial effect of gender

```
mgolAddit  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
mgolPopul  <- lm(gol ~ fpopul, data = HowellsAll)
anova(mgolPopul, mgolAddit)
```

```
Analysis of Variance Table

Model 1: gol ~ fpopul
Model 2: gol ~ fgender + fpopul
  Res.Df   RSS Df Sum of Sq     F    Pr(>F)
1    286 17904
2    285 11254  1    6649.7 168.4 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

HowelsAll (subset, $n = 289$)

gol ∼ gender + popul

For a given gender,
does population **have an effect in the mean value of** gol**?**
Partial effect of population

```
mgolAddit  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
mgolGender  <- lm(gol ~ fgender, data = HowellsAll)
anova(mgolGender, mgolAddit)
```

```
Analysis of Variance Table

Model 1: gol ~ fgender
Model 2: gol ~ fgender + fpopul
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    287 16415
2    285 11254  2    5160.7 65.345 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

HowelsAll (subset, $n = 289$)

gol $\sim$ gender + popul

F-tests of significance of both partial effects

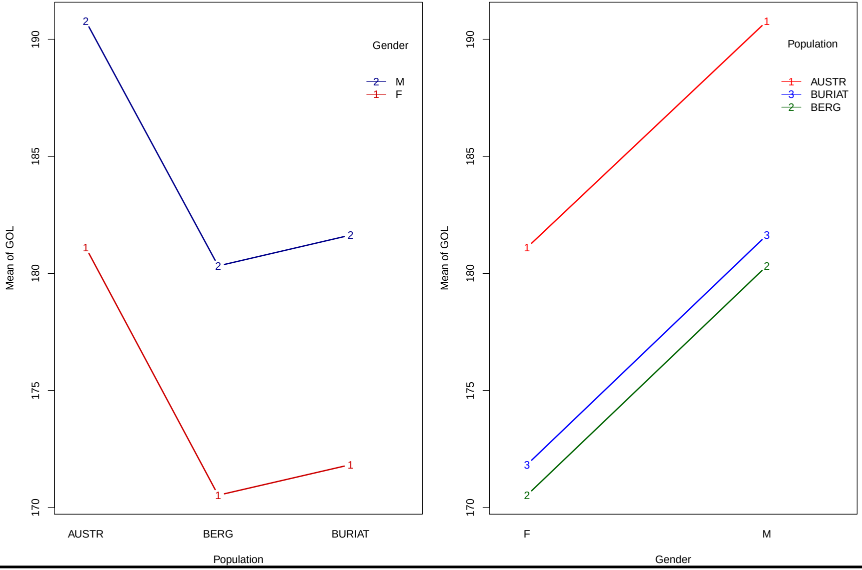
```
mgolAddit  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
drop1(mgolAddit, test = "F")
```

```
Single term deletions

Model:
gol ~ fgender + fpopul
        Df Sum of Sq   RSS    AIC F value    Pr(>F)
<none>                11254 1066.3
fgender  1    6649.7 17904 1198.5 168.396 < 2.2e-16 ***
fpopul   2    5160.7 16415 1171.4  65.345 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`HowellsAll` ($n = 289$)

`gol` (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$),

quantification of both partial effects?



`HowelsAll` (subset, $n = 289$)

`gol ~ gender + popul`, `contr.treatment` parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z, w) = \beta_0 + \beta^Z \, \mathbb{I}[z = \text{male}] + \beta_1^W \, \mathbb{I}[w = \text{Berg}] + \beta_2^W \, \mathbb{I}[w = \text{Burjati}]$$

```
lm(gol ~ fgender + fpopul, data = HowellsAll)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-15.5400  -4.3103  -0.3103   4.4600  17.6897

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 181.0712     0.7814 231.724   <2e-16 ***
fgenderM      9.7703     0.7529  12.977   <2e-16 ***
fpopulBERG  -10.5311     0.9706 -10.850   <2e-16 ***
fpopulBURIAT  -9.2213    0.9695  -9.511   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.284 on 285 degrees of freedom
Multiple R-squared:  0.4729,        Adjusted R-squared:  0.4674
F-statistic: 85.24 on 3 and 285 DF,  p-value: < 2.2e-16
```

gol ∼ gender + popul

LSE's of $\mathbb{E}\big(Y \,\big|\, Z = g_1, W = \star\big) - \mathbb{E}\big(Y \,\big|\, Z = g_2, W = \star\big)$
and $\mathbb{E}\big(Y \,\big|\, Z = \star, W = h_1\big) - \mathbb{E}\big(Y \,\big|\, Z = \star, W = h_2\big)$

```
mgolAddit  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
L  <- matrix(c(0,1,0,0, 0,0,1,0, 0,0,0,1, 0,0,-1,1), ncol = 4, byrow = TRUE)
rownames(L)  <- c("Male-Female", "Berg-Austr", "Burjati-Austr", "Burjati-Berg")
colnames(L)  <- names(coef(mgolAddit))
print(L)
```

```
              (Intercept) fgenderM fpopulBERG fpopulBURIAT
Male-Female             0        1          0            0
Berg-Austr              0        0          1            0
Burjati-Austr           0        0          0            1
Burjati-Berg            0        0         -1            1
```

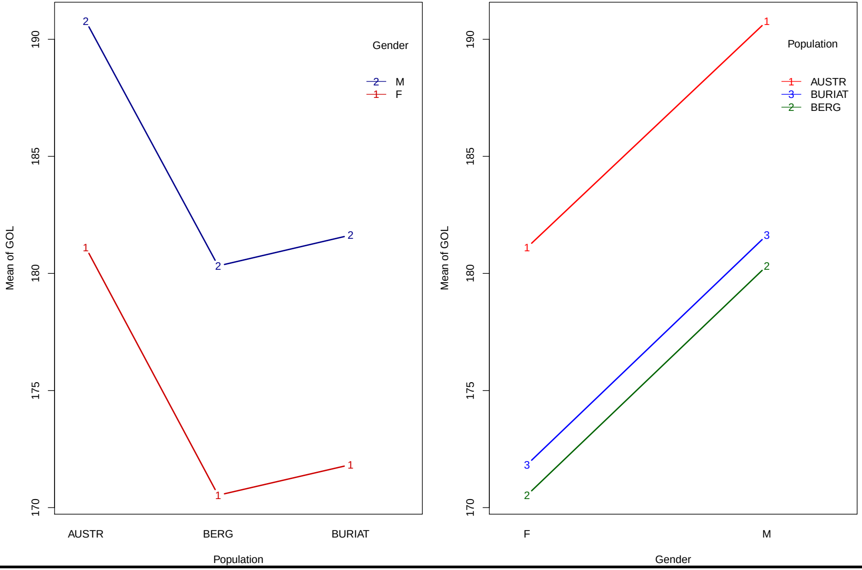
```
mffSM::LSest(mgolAddit, L = L)
```

```
                Estimate Std. Error     t value P value        Lower     Upper
Male-Female     9.770313  0.7529092  12.976750 < 2e-16   8.2883454 11.252282
Berg-Austr    -10.531148  0.9705782 -10.850385 < 2e-16 -12.4415591 -8.620737
Burjati-Austr  -9.221329  0.9695097  -9.511332 < 2e-16 -11.1296364 -7.313021
Burjati-Berg    1.309819  0.8512377   1.538723 0.12498  -0.3656911  2.985330
```

`HowellsAll` ($n = 289$)

`gol` (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$),

alternative quantification of both partial effects?



`HowelsAll` (subset, $n = 289$)

`gol ~ gender + popul`, `contr.sum` parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z, w) = \beta_0 + \beta^Z\, \mathbb{I}[z = \text{female}] - \beta^Z\, \mathbb{I}[z = \text{male}]$$
$$+ \beta_1^W\, \mathbb{I}[w = \text{Austr}] + \beta_2^W\, \mathbb{I}[w = \text{Berg}] + (-\beta_1^W - \beta_2^W)\, \mathbb{I}[w = \text{Burjati}]$$

```
options(contrasts = c("contr.sum", "contr.sum"))
lm(gol ~ fgender + fpopul, data = HowellsAll)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-15.5400  -4.3103  -0.3103   4.4600  17.6897

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 179.3722     0.3797 472.421  < 2e-16 ***
fgender1     -4.8852     0.3765 -12.977  < 2e-16 ***
fpopul1       6.5842     0.5811  11.330  < 2e-16 ***
fpopul2      -3.9470     0.5157  -7.654 3.03e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.284 on 285 degrees of freedom
Multiple R-squared:  0.4729,	Adjusted R-squared:  0.4674
F-statistic: 85.24 on 3 and 285 DF,  p-value: < 2.2e-16
```

HowelsAll (subset, $n = 289$)

gol ∼ gender + popul

LSE's of $\mathbb{E}\big(Y \,\big|\, Z = g, W = \star\big) - \frac{1}{G}\sum_{j=1}^{G} \mathbb{E}\big(Y \,\big|\, Z = j, W = \star\big)$

and $\mathbb{E}\big(Y \,\big|\, Z = \star, W = h\big) - \frac{1}{H}\sum_{j=1}^{H} \mathbb{E}\big(Y \,\big|\, Z = \star, W = j\big)$

```
options(contrasts = c("contr.sum", "contr.sum"))
mgolAdditSum  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
L  <- matrix(c(0,1,0,0, 0,-1,0,0, 0,0,1,0, 0,0,0,1, 0,0,-1,-1), ncol = 4, byrow = TRUE)
rownames(L)  <- c("Female", "Male", "Australia", "Berg", "Burjati")
colnames(L)  <- names(coef(mgolAdditSum))
print(L)
```

```
          (Intercept) fgender1 fpopul1 fpopul2
Female              0        1       0       0
Male                0       -1       0       0
Australia           0        0       1       0
Berg                0        0       0       1
Burjati             0        0      -1      -1
```

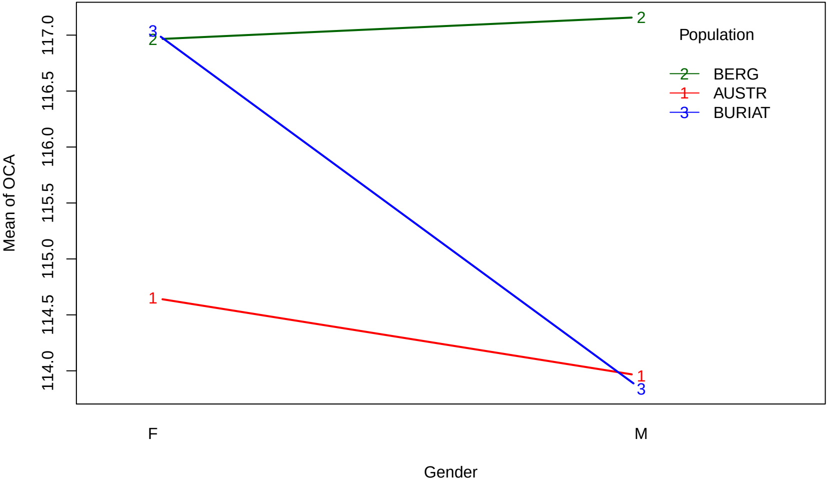
```
mffSM::LSest(mgolAdditSum, L = L)
```

```
           Estimate Std. Error    t value    P value     Lower     Upper
Female    -4.885157  0.3764546 -12.976750 < 2.22e-16 -5.626141 -4.144173
Male       4.885157  0.3764546  12.976750 < 2.22e-16  4.144173  5.626141
Australia  6.584159  0.5811231  11.330059 < 2.22e-16  5.440321  7.727997
Berg      -3.946989  0.5156772  -7.653992 3.0336e-13 -4.962008 -2.931970
Burjati   -2.637170  0.5150067  -5.120651 5.6141e-07 -3.650869 -1.623470
```

5.4.3 Interactions

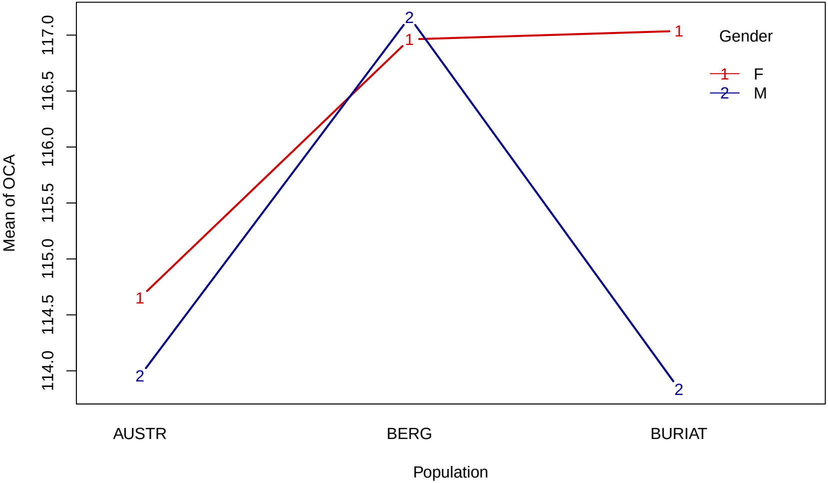
HowellsAll ($n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$)



HowellsAll ($n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$)



`HowelsAll` (subset, $n = 289$)

`oca ∼ gender + popul + gender:popul`, `contr.treatment` parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z, w) = \beta_0 + \beta^Z\,\mathbb{I}[z = \text{male}] + \beta_1^W\,\mathbb{I}[w = \text{Berg}] + \beta_2^W\,\mathbb{I}[w = \text{Burjati}]$$
$$+ \beta_1^{ZW}\mathbb{I}[z = \text{male}, w = \text{Berg}] + \beta_2^{ZW}\mathbb{I}[z = \text{male}, w = \text{Burjati}]$$

```
lm(oca ~ fgender*fpopul, data = HowellsAll)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-15.1607  -3.1607   0.0455   3.1636  13.8393 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           114.6531     0.7186 159.548   <2e-16 ***
fgenderM               -0.6985     1.2910  -0.541   0.5889    
fpopulBERG              2.3092     0.9969   2.316   0.0213 *  
fpopulBURIAT            2.3840     0.9925   2.402   0.0169 *  
fgenderM:fpopulBERG     0.8970     1.6112   0.557   0.5782    
fgenderM:fpopulBURIAT  -2.5022     1.6110  -1.553   0.1215    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.03 on 283 degrees of freedom
Multiple R-squared:  0.07842,	Adjusted R-squared:  0.06214 
F-statistic: 4.816 on 5 and 283 DF,  p-value: 0.0003046
```

HowelsAll (subset, $n = 289$)

oca ∼ gender + popul + gender:popul, contr.sum parameterisation

Z: gender (*Female, Male*), W: population (*Australia, Berg, Burjati*)

$$m(z, w) = \beta_0 + \beta^Z\,\mathbb{I}[z = \text{female}] - \beta^Z\,\mathbb{I}[z = \text{male}]$$
$$+ \beta_1^W\,\mathbb{I}[w = \text{Austr.}] + \beta_2^W\,\mathbb{I}[w = \text{Berg}] + (-\beta_1^W - \beta_2^W)\,\mathbb{I}[w = \text{Burjati}]$$
$$+ \beta_1^{ZW}\mathbb{I}[z = \text{fem.}, w = \text{Aus.}] + \beta_2^{ZW}\mathbb{I}[z = \text{fem.}, w = \text{Berg}] + (-\beta_1^{ZW} - \beta_2^{ZW})\,\mathbb{I}[z = \text{fem.}, w = \text{Bur}$$
$$- \beta_1^{ZW}\mathbb{I}[z = \text{male}, w = \text{Aus.}] - \beta_2^{ZW}\mathbb{I}[z = \text{male}, w = \text{Berg}] + (\beta_1^{ZW} + \beta_2^{ZW})\,\mathbb{I}[z = \text{male}, w = \text{Bur}$$

```
options(contrasts = c("contr.sum", "contr.sum"))
lm(oca ~ fgender + fpopul, data = HowellsAll)
```

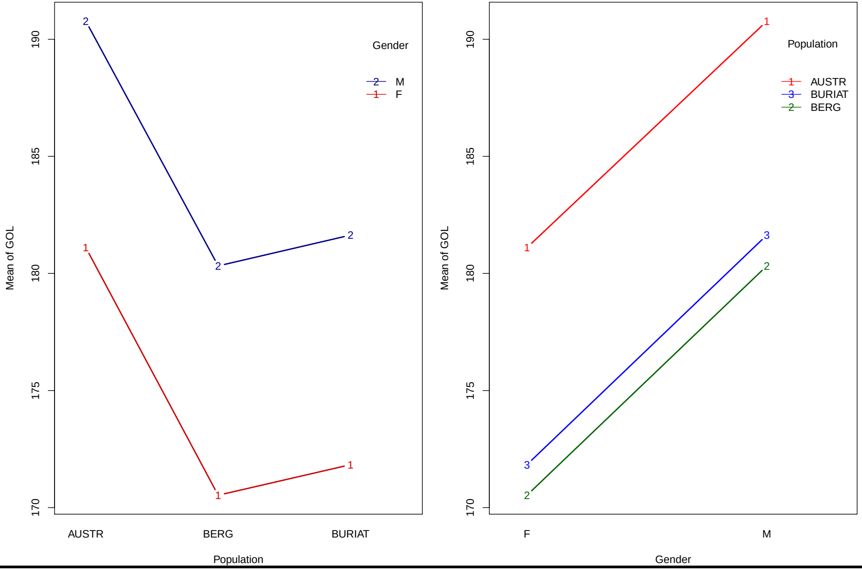
```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     115.6007     0.3129 369.455  < 2e-16 ***
fgender1          0.6168     0.3129   1.971 0.049671 *
fpopul1          -1.2969     0.4866  -2.665 0.008138 **
fpopul2           1.4608     0.4187   3.489 0.000563 ***
fgender1:fpopul1 -0.2675     0.4866  -0.550 0.582896
fgender1:fpopul2 -0.7160     0.4187  -1.710 0.088376 .
---

Residual standard error: 5.03 on 283 degrees of freedom
Multiple R-squared:  0.07842,       Adjusted R-squared:  0.06214
F-statistic: 4.816 on 5 and 283 DF,  p-value: 0.0003046
```

5.4.4 Additivity or interactions?

HowellsAll ($n = 289$)

gol (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$),

additivity or interactions?



HowelsAll (subset, $n = 289$)

gol (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$)

Do the mean gol differences between *male* and *female* depend on population?

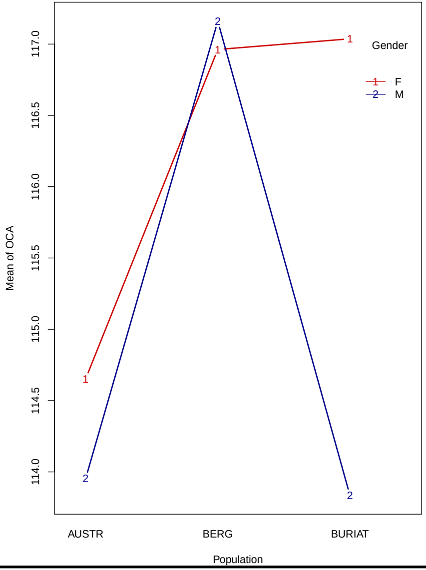
Do the mean gol differences between populations depend on gender?

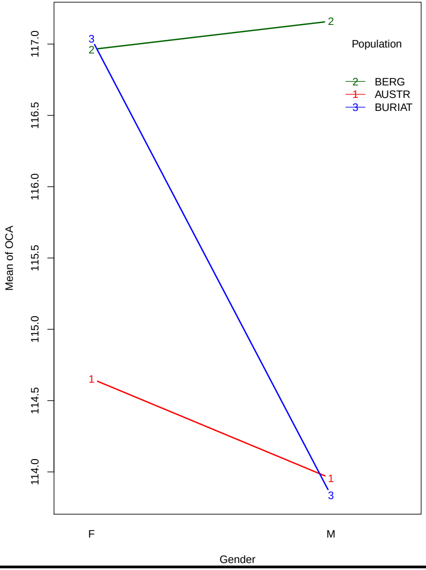
```
mgolAddit  <- lm(gol ~ fgender + fpopul, data = HowellsAll)
mgolInter  <- lm(gol ~ fgender*fpopul, data = HowellsAll)
anova(mgolAddit, mgolInter)
```

```
Analysis of Variance Table

Model 1: gol ~ fgender + fpopul
Model 2: gol ~ fgender * fpopul
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1    285 11254
2    283 11254  2   0.19404 0.0024 0.9976
```

HowellsAll ($n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$), **additivity or interactions?**





HowelsAll (subset, $n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$)

Do the mean oca differences between *male* and *female* depend on population?

Do the mean oca differences between populations depend on gender?

```
mocaAddit  <- lm(oca ~ fgender + fpopul, data = HowellsAll)
mocaInter  <- lm(oca ~ fgender*fpopul, data = HowellsAll)
anova(mocaAddit, mocaInter)
```

```
Analysis of Variance Table

Model 1: oca ~ fgender + fpopul
Model 2: oca ~ fgender * fpopul
  Res.Df  RSS Df Sum of Sq      F  Pr(>F)
1    285 7326
2    283 7161  2    165.02 3.2607 0.03981 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Section **5.5**

Multiple regression model

5.5.1 Model terms

Numeric covariate: **Simple transformation** parameterization

$\boldsymbol{s} = s : \mathcal{Z} \longrightarrow \mathbb{R}$ with

$$\mathbb{S} = \begin{pmatrix} s(Z_1) \\ \vdots \\ s(Z_n) \end{pmatrix} = (\boldsymbol{S}), \qquad \begin{aligned} \boldsymbol{X}_1 = X_1 &= s(Z_1), \\ &\vdots \\ \boldsymbol{X}_n = X_n &= s(Z_n). \end{aligned}$$

Numeric covariate: **Polynomial** parameterization

$\boldsymbol{s} = \left(s_1, \ldots, s_{k-1}\right)^\top$ such that $s_j(z) = P^j(z)$ is polynomial in z of degree j, $j = 1, \ldots, k-1$.

$$\mathbb{S} = \begin{pmatrix} P^1(Z_1) & \ldots & P^{k-1}(Z_1) \\ \vdots & \vdots & \vdots \\ P^1(Z_n) & \ldots & P^{k-1}(Z_n) \end{pmatrix} = \left(\boldsymbol{P}^1, \quad \ldots, \quad \boldsymbol{P}^{k-1}\right),$$

$$\begin{aligned} \boldsymbol{X}_1 &= \left(P^1(Z_1), \ldots, P^{k-1}(Z_1)\right)^\top, \\ &\vdots \\ \boldsymbol{X}_n &= \left(P^1(Z_n), \ldots, P^{k-1}(Z_n)\right)^\top. \end{aligned}$$

5.5.1 Model terms

Numeric covariate: **Regression spline** parameterization

$\boldsymbol{s} = \left(s_1, \ldots, s_k\right)^\top$ such that $s_j(z) = B_j(z)$, $j = 1, \ldots, k$, where $B_1, \ldots, B_k$ is the spline basis of chosen degree $d \in \mathbb{N}_0$ composed of basis B-splines built above a set of chosen knots $\boldsymbol{\lambda} = \left(\lambda_1, \ldots, \lambda_{k-d+1}\right)^\top$.

$$\mathbb{S} = \mathbb{B} = \begin{pmatrix} B_1(Z_1) & \ldots & B_k(Z_1) \\ \vdots & \vdots & \vdots \\ B_1(Z_n) & \ldots & B_k(Z_n) \end{pmatrix} = \left(\boldsymbol{B}^1, \quad \ldots, \quad \boldsymbol{B}^k\right),$$

$$\begin{aligned} \boldsymbol{X}_1 &= \left(B_1(Z_1), \ldots, B_k(Z_1)\right)^\top, \\ &\vdots \\ \boldsymbol{X}_n &= \left(B_1(Z_n), \ldots, B^k(Z_n)\right)^\top. \end{aligned}$$

5.5.1 Model terms

Categorical covariate: **(Pseudo)contrast** parameterization

- $\mathcal{Z} = \{1, \ldots, G\}$.
- $\boldsymbol{s}(z) = \boldsymbol{c}_z$, $z \in \mathcal{Z}$,
  - $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_G \in \mathbb{R}^{G-1}$

    $\equiv$ rows of a chosen (pseudo)contrast matrix $\mathbb{C}_{G \times G-1}$.

$$\mathbb{S} = \begin{pmatrix} \boldsymbol{c}_{Z_1}^\top \\ \vdots \\ \boldsymbol{c}_{Z_n}^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^1, & \ldots, & \boldsymbol{C}^{G-1} \end{pmatrix}, \qquad \begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{c}_{Z_1}, \\ &\vdots \\ \boldsymbol{X}_n &= \boldsymbol{c}_{Z_n}. \end{aligned}$$

5.5.1 Model terms

Main effect model terms

Definition 5.3 The main effect model term.

Depending on a chosen parameterization $\boldsymbol{s}: \mathcal{Z} \longrightarrow \mathbb{R}^{k^\star}$, *the main effect model term* (of order one) of a given covariate Z is defined as a transformation $\boldsymbol{t}$ with elements as follows and a matrix $\mathbb{T}$ with columns as follows.

Numeric covariate

(i) **Simple transformation** $s: \mathcal{Z} \longrightarrow \mathbb{R}$.
➡ $\boldsymbol{t} = s$ and $\mathbb{T}$ is (the only) column $\boldsymbol{S}$ of the reparameterizing matrix $\mathbb{S}$, i.e.,

$$\mathbb{T} = \mathbb{S} = \begin{pmatrix} s(Z_1) \\ \vdots \\ s(Z_n) \end{pmatrix} = \big(\boldsymbol{S}\big).$$

Definition 5.3 The main effect model term, cont'd.

(ii) **Polynomial** $\boldsymbol{s} = \left(s_1, \ldots, s_{k-1}\right)^\top$, $s_j(z) = P^j(z)$ is polynomial in z of degree j, $j = 1, \ldots, k-1$ with the reparameterizing matrix

$$\mathbb{S} = \begin{pmatrix} P^1(Z_1) & \ldots & P^{k-1}(Z_1) \\ \vdots & \vdots & \vdots \\ P^1(Z_n) & \ldots & P^{k-1}(Z_n) \end{pmatrix} = \left(\boldsymbol{P}^1, \quad \ldots, \quad \boldsymbol{P}^{k-1}\right).$$

➠ $\boldsymbol{t} = s_1 = P^1$ (linear polynomial) and $\mathbb{T}$ is the first column $\boldsymbol{P}^1$ of the reparameterizing matrix $\mathbb{S}$ that corresponds to the linear transformation of the covariate Z, i.e.,

$$\mathbb{T} = \left(\boldsymbol{P}^1\right).$$

Definition 5.3 The main effect model term, cont'd.

(iii) **Regression spline** $\boldsymbol{s} = (s_1, \ldots, s_k)^\top$, $s_j(z) = B_j(z)$, $j = 1, \ldots, k$, where $B_1, \ldots, B_k$ is the spline basis and the reparameterizing matrix is

$$\mathbb{S} = \mathbb{B} = \begin{pmatrix} B_1(Z_1) & \ldots & B_k(Z_1) \\ \vdots & \vdots & \vdots \\ B_1(Z_n) & \ldots & B_k(Z_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}^1, & \ldots, & \boldsymbol{B}^k \end{pmatrix}.$$

➠ $\boldsymbol{t} = \boldsymbol{s}$ (all basis splines) and $\mathbb{T}$ are (all) columns $\boldsymbol{B}^1, \ldots, \boldsymbol{B}^k$ of the reparameterizing matrix $\mathbb{S} = \mathbb{B}$, i.e.,

$$\mathbb{T} = \left(\boldsymbol{B}^1, \ldots, \boldsymbol{B}^k\right).$$

Definition 5.3 The main effect model term, cont'd.

Categorical covariate with $\mathcal{Z} = \{1, \dots, G\}$ parameterized by the mean of a (pseudo)contrast matrix

$$\mathbb{C} = \begin{pmatrix} \boldsymbol{c}_1^\top \\ \vdots \\ \boldsymbol{c}_G^\top \end{pmatrix},$$

i.e., $\boldsymbol{s}(z) = \boldsymbol{c}_z$, $z \in \mathcal{Z}$.

➠ $\boldsymbol{t} = \boldsymbol{s}$ (row of a chosen (pseudo)contrast matrix) and $\mathbb{T}$ are (all) columns of the corresponding reparameterizing matrix, i.e.,

$$\mathbb{T} = \mathbb{S} = \begin{pmatrix} \boldsymbol{c}_{Z_1}^\top \\ \vdots \\ \boldsymbol{c}_{Z_n}^\top \end{pmatrix} = \left(\boldsymbol{C}^1, \dots, \boldsymbol{C}^{G-1}\right).$$

Definition 5.4 The main effect model term of order j.

If a *numeric* covariate Z is parameterized using the polynomial of degree $k-1$, i.e., $\boldsymbol{s} = (s_1, \ldots, s_{k-1})^\top$, $s_j(z) = P^j(z)$, $j = 1, \ldots, k-1$, then *the main effect model term of order j*, $j = 2, \ldots, k-1$, means the element $s_j(z) = P^j(z)$ of the polynomial parameterization and a matrix $\mathbb{T}^j$ whose the only column is the jth column $\boldsymbol{P}^j$ of the reparameterizing matrix

$$\mathbb{S} = \begin{pmatrix} P^1(Z_1) & \ldots & P^{k-1}(Z_1) \\ \vdots & \vdots & \vdots \\ P^1(Z_n) & \ldots & P^{k-1}(Z_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}^1, & \ldots, & \boldsymbol{P}^{k-1} \end{pmatrix},$$

that corresponds to the polynomial of degree j, i.e.,

$$\mathbb{T}^j = (\boldsymbol{P}^j).$$

Note. The terms $\mathbb{T}, \ldots, \mathbb{T}^{j-1}$ are called as *lower order* terms included in the term $\mathbb{T}^j$.

Two-way interaction model terms

Two covariates Z and W and their main effect model terms $\boldsymbol{t}_Z$, $\mathbb{T}_Z$ and $\boldsymbol{t}_W$, $\mathbb{T}_W$.

Definition 5.5 The two-way interaction model term.

The *two-way interaction* model term means elements of a vector $\boldsymbol{t}_W \otimes \boldsymbol{t}_Z$ and a matrix $\mathbb{T}^{ZW}$, where

$$\mathbb{T}^{ZW} := \mathbb{T}_Z : \mathbb{T}_W.$$

Notes.

- The main effect model term $\mathbb{T}_Z$ and/or the main effect model term $\mathbb{T}_W$ that enters the two-way interaction may also be of a degree $j > 1$.
- Both the main effect model terms $\mathbb{T}_Z$ and $\mathbb{T}_W$ are called as *lower order* terms included in the two-way interaction term $\mathbb{T}_Z : \mathbb{T}_W$.

Higher order interaction model terms

Three covariates Z, W and V and their main effect model terms $\boldsymbol{t}_Z$, $\mathbb{T}_Z$ and $\boldsymbol{t}_W$, $\mathbb{T}_W$ and $\boldsymbol{t}_V$, $\mathbb{T}_V$.

Definition 5.6 The three-way interaction model term.

The *three-way interaction* model term means a vector $\boldsymbol{t}_V \otimes (\boldsymbol{t}_W \otimes \boldsymbol{t}_Z)$ and a matrix $\mathbb{T}^{ZWV}$, where

$$\mathbb{T}^{ZWV} := (\mathbb{T}_Z : \mathbb{T}_W) : \mathbb{T}_V.$$

Notes.

- Any of the main effect model terms $\mathbb{T}_Z$, $\mathbb{T}_W$, $\mathbb{T}_V$ that enter the three-way interaction may also be of a degree $j > 1$.
- All main effect terms $\mathbb{T}_Z$, $\mathbb{T}_W$ and $\mathbb{T}_V$ and also all two-way interaction terms $\mathbb{T}_Z : \mathbb{T}_W$, $\mathbb{T}_Z : \mathbb{T}_V$ and $\mathbb{T}_W : \mathbb{T}_V$ are called as *lower order* terms included in the three-way interaction term $\mathbb{T}^{ZWV}$.
- By induction, we could define also four-way, five-way, . . . , i.e., *higher order* interaction model terms and a notion of corresponding lower order nested terms.

5.5.2 Model formula

Symbols in a model formula

- **1**:
  intercept term in the model if this is the only term in the model (i.e., intercept only model).
- **Letter** or **abbreviation**:
  main effect of order one of a particular covariate (which is identified by the letter or abbreviation). It is assumed that chosen parameterization is either known from context or is indicated in some way (e.g., by the used abbreviation). Letters or abbreviations will also be used to indicate a response variable.
- **Power** of j, $j > 1$ (above a letter or abbreviation):
  main effect of order j of a particular covariate.
- **Colon (:)** between two or more letters or abbreviations:
  interaction term based on particular covariates.
- **Plus sign (+)**:
  a delimiter of the model terms.
- **Tilde ($\sim$)**:
  a delimiter between the response and description of the regression function.

5.5.3 Hierarchically well formulated model

Definition 5.7 Hierarchically well formulated model.

Hierarchically well formulated (HWF) model is such a model that contains an intercept term (possibly implicitely) and with each model term also all lower order terms that are nested in this term.

5.5.3 Hierarchically well formulated model

Example. Quadratic regression function

- x parameterization:

$$m_x(x) = \beta_0 + \beta_1\, x + \beta_2\, x^2$$

- Transformation $x \longrightarrow t$ ($\delta \neq 0$, $\varphi \neq 0$):

$$x = \delta\,(t - \varphi), \qquad t = \varphi + \frac{x}{\delta}$$

- t parameterization:

$$m_t(t) = \gamma_0 + \gamma_1\, t + \gamma_2\, t^2$$

$$\gamma_0 = \beta_0 - \beta_1\delta\varphi + \beta_2\delta^2\varphi^2$$
$$\gamma_1 = \beta_1\delta - 2\beta_2\delta^2\varphi$$
$$\gamma_2 = \beta_2\delta^2$$

5.5.3 Hierarchically well formulated model

Example. Quadratic regression function, no linear term

- x parameterization:

$$m_x(x) = \beta_0 + \beta_2\, x^2$$

- Transformation $x \longrightarrow t$ ($\delta \neq 0$, $\varphi \neq 0$):

$$x = \delta\,(t - \varphi), \qquad t = \varphi + \frac{x}{\delta}$$

- t parameterization:

$$m_t(t) = \gamma_0 + \gamma_1\, t + \gamma_2\, t^2$$

$$\begin{aligned} \gamma_0 &= \beta_0 + \beta_2\delta^2\varphi^2 \\ \gamma_1 &= -2\beta_2\delta^2\varphi \\ \gamma_2 &= \beta_2\delta^2 \end{aligned}$$

5.5.3 Hierarchically well formulated model

Possible reasons for not using the HWF model

- No intercept in the model
  $\equiv$ it can be assumed that the response expectation is zero if all regressors in a chosen parameterization take zero values.

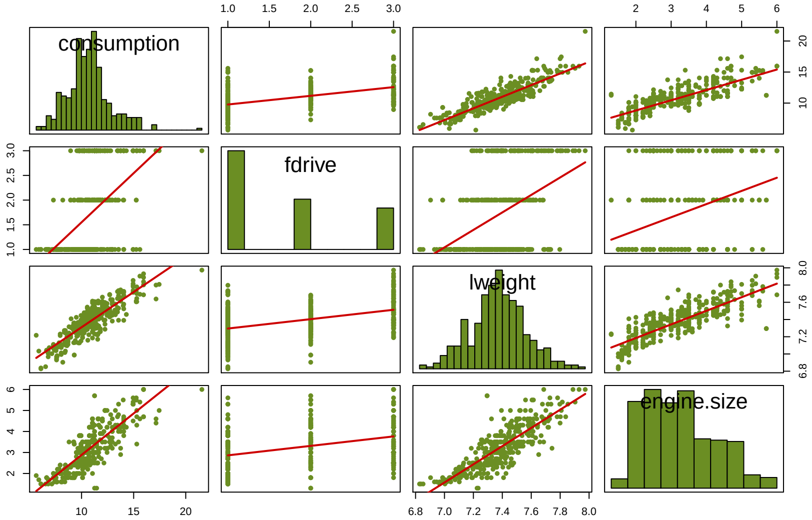
- No linear term in a model with a quadratic regression function $m(x) = \beta_0 + \beta_2 \, x^2$
  $\equiv$ it can be assumed that the regression function is a parabola with the vertex in a point $(0, \, \beta_0)$ with respect to the x parameterization.

- No main effect of one covariate in an interaction model with two numeric covariates and a regression function $m(x, \, z) = \beta_0 + \beta_1 \, z + \beta_2 \, x \, z$
  $\equiv$ it can be assumed that with $z = 0$, the response expectation does not depend on a value of x, i.e., $\mathbb{E}\big(Y \,\big|\, X = x, \, Z = 0\big) = \beta_0$ (a constant).

5.5.4 Usual strategy to specify a multiple regression model

Cars2004nh (subset, $n = 409$)

consumption ∼ drive, engine size, log(weight)



Cars2004nh (subset, $n = 409$)

consumption ~ drive + engine size + log(weight)

```
mAddit <- lm(consumption ~ fdrive + engine.size + lweight, data = CarsNow)
summary(mAddit)
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -35.84930    3.08092 -11.636  < 2e-16 ***
fdriverear    0.46260    0.11715   3.949 9.26e-05 ***
fdrive4x4     0.98198    0.13019   7.543 3.07e-13 ***
engine.size   0.56908    0.08361   6.807 3.62e-11 ***
lweight       6.03099    0.44795  13.464  < 2e-16 ***

Residual standard error: 0.9223 on 404 degrees of freedom
Multiple R-squared:  0.8149,	Adjusted R-squared:  0.8131
F-statistic: 444.8 on 4 and 404 DF,  p-value: < 2.2e-16
```

```
drop1(mAddit, test = "F")
```

```
Single term deletions

Model:
consumption ~ fdrive + engine.size + lweight
            Df Sum of Sq    RSS     AIC F value    Pr(>F)
<none>                   343.69 -61.161
fdrive       2    50.574 394.26  -9.012  29.725 9.046e-13 ***
engine.size  1    39.413 383.10 -18.758  46.330 3.625e-11 ***
lweight      1   154.205 497.89  88.436 181.267 < 2.2e-16 ***
```

Cars2004nh (subset, $n = 409$)

consumption ∼ drive + engine size + log(weight) + drive:log(weight)

```
mInter1 <- lm(consumption ~ fdrive + engine.size + lweight + fdrive:lweight, data = CarsNow)
summary(mInter1)
```

```
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -37.44459    3.22260 -11.619  < 2e-16 ***
fdriverear           22.90273    4.86163   4.711 3.40e-06 ***
fdrive4x4            -8.59853    4.42520  -1.943   0.0527 .  
engine.size           0.57588    0.08125   7.088 6.16e-12 ***
lweight               6.24702    0.46296  13.494  < 2e-16 ***
fdriverear:lweight   -3.03731    0.65971  -4.604 5.57e-06 ***
fdrive4x4:lweight     1.26748    0.59358   2.135   0.0333 *  

Residual standard error: 0.8877 on 402 degrees of freedom
Multiple R-squared:  0.8294,	Adjusted R-squared:  0.8269 
F-statistic: 325.8 on 6 and 402 DF,  p-value: < 2.2e-16
```

```
drop1(mInter1, test = "F")
```

```
Single term deletions

Model:
consumption ~ fdrive + engine.size + lweight + fdrive:lweight
               Df Sum of Sq    RSS     AIC F value    Pr(>F)    
<none>                      316.81 -90.469                      
engine.size     1    39.590 356.40 -44.308  50.236 6.159e-12 ***
fdrive:lweight  2    26.879 343.69 -61.161  17.054 7.782e-08 ***
```

Cars2004nh (subset, $n = 409$)

cons. ~ drive + eng.size + log(weight) + drive:log(wgt) + eng.size:log(wgt)

```
mInter2 <- lm(consumption ~ fdrive + engine.size + lweight + fdrive:lweight +
              engine.size:lweight, data = CarsNow)
summary(mInter2)
```

```
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -22.8398     4.9687  -4.597 5.76e-06 ***
fdriverear           27.3567     4.9219   5.558 4.98e-08 ***
fdrive4x4             4.3904     5.5249   0.795 0.427287
engine.size          -5.8845     1.6945  -3.473 0.000571 ***
lweight               4.2821     0.6873   6.230 1.18e-09 ***
fdriverear:lweight   -3.6356     0.6675  -5.446 8.98e-08 ***
fdrive4x4:lweight    -0.4836     0.7425  -0.651 0.515241
engine.size:lweight   0.8662     0.2270   3.817 0.000157 ***

Residual standard error: 0.8731 on 401 degrees of freedom
Multiple R-squared:  0.8354,       Adjusted R-squared:  0.8325
F-statistic: 290.7 on 7 and 401 DF,  p-value: < 2.2e-16
```

```
drop1(mInter2, test = "F")
```

```
consumption ~ fdrive + engine.size + lweight + fdrive:lweight +
    engine.size:lweight
                    Df Sum of Sq    RSS      AIC F value    Pr(>F)
<none>                           305.70 -103.064
fdrive:lweight       2    24.150 329.85  -75.966  15.839 2.395e-07 ***
engine.size:lweight  1    11.105 316.81  -90.469  14.567 0.0001566 ***
```

`Cars2004nh` (subset, $n = 409$)

`consumption` $\sim$ (`drive` + `engine size` + `log(weight)`)^2

```
mInter <- lm(consumption ~ (fdrive + engine.size + lweight)^2, data = CarsNow)
summary(mInter)
```

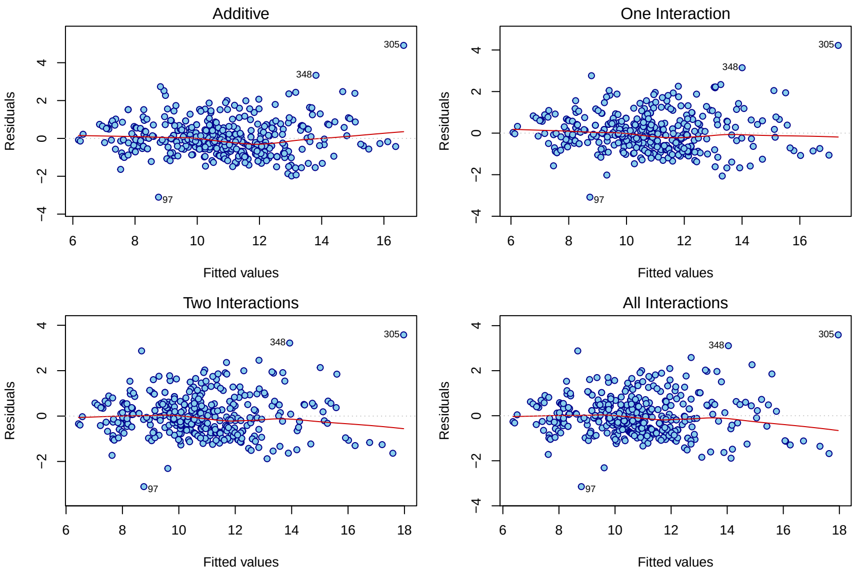
```
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -26.124609   5.776121  -4.523 8.06e-06 ***
fdriverear              26.875936   7.367167   3.648 0.000299 ***
fdrive4x4               13.308169   8.311915   1.601 0.110147
engine.size             -5.391862   1.746264  -3.088 0.002158 **
lweight                  4.757609   0.817131   5.822 1.19e-08 ***
fdriverear:engine.size   0.009665   0.182958   0.053 0.957895
fdrive4x4:engine.size    0.315489   0.216880   1.455 0.146547
fdriverear:lweight      -3.571144   1.061146  -3.365 0.000839 ***
fdrive4x4:lweight       -1.818723   1.189560  -1.529 0.127081
engine.size:lweight      0.790111   0.233312   3.386 0.000778 ***

Residual standard error: 0.8726 on 399 degrees of freedom
Multiple R-squared:  0.8364,	Adjusted R-squared:  0.8327
F-statistic: 226.7 on 9 and 399 DF,  p-value: < 2.2e-16
```

```
drop1(mInter, test = "F")
```

```
consumption ~ (fdrive + engine.size + lweight)^2
                    Df Sum of Sq    RSS      AIC F value    Pr(>F)
<none>                           303.78 -101.642
fdrive:engine.size   2    1.9215 305.70 -103.064  1.2619 0.2842440
fdrive:lweight       2    8.6863 312.46  -94.112  5.7045 0.0036085 **
engine.size:lweight  1    8.7315 312.51  -92.052 11.4684 0.0007782 ***
```

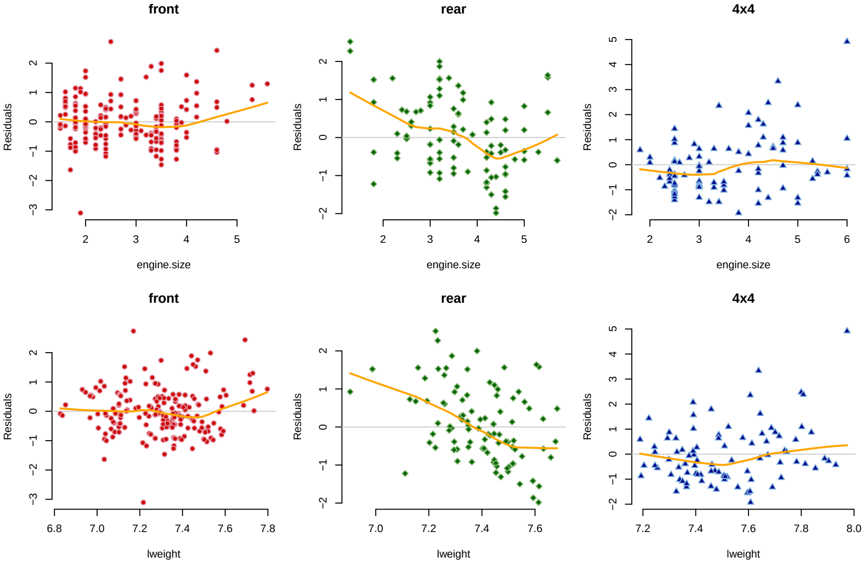
Cars2004nh (subset, $n = 409$)

consumption ∼ drive, engine size, log(weight)



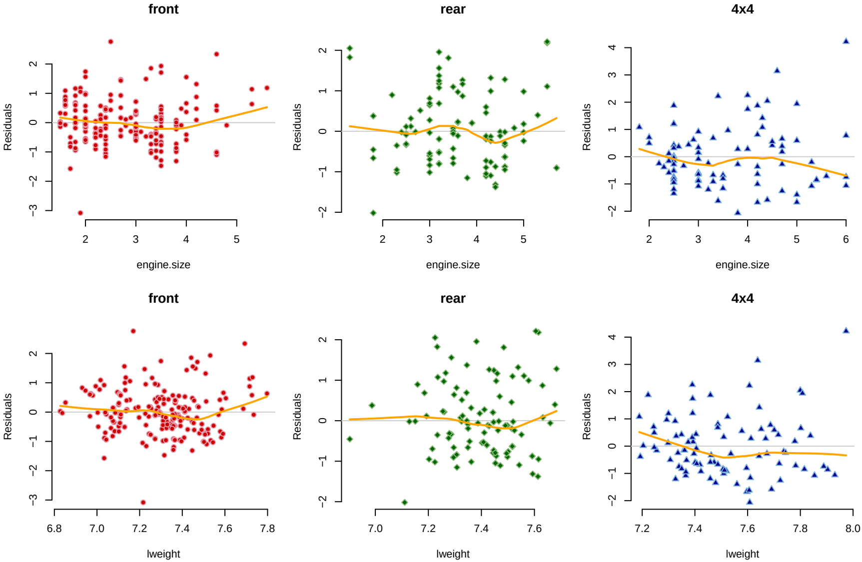
Cars2004nh (subset, $n = 409$)

consumption $\sim$ drive + engine size + log(weight)



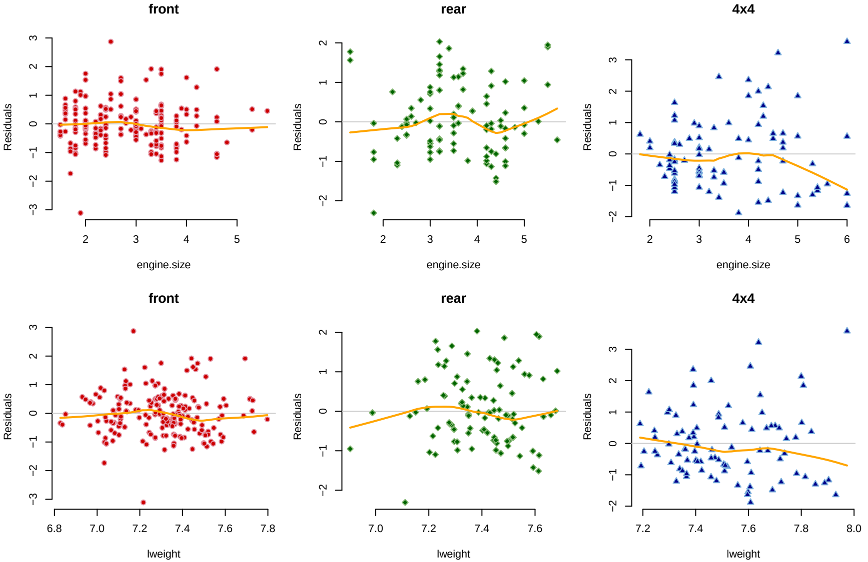
Cars2004nh (subset, $n = 409$)

consumption ~ drive + engine size + log(weight) + drive:log(weight)



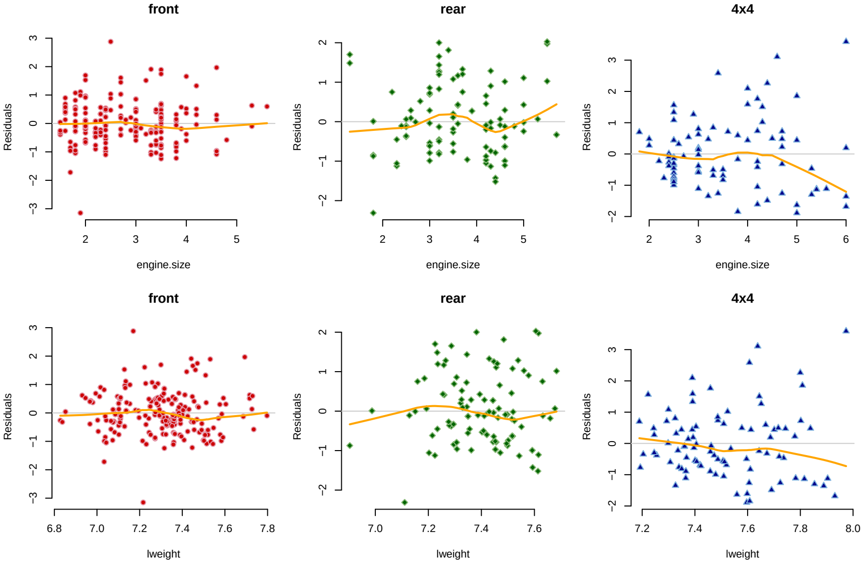
Cars2004nh (subset, $n = 409$)

cons. ∼ drive + eng.size + log(weight) + drive:log(wgt) + eng.size:log(wgt)



Cars2004nh (subset, $n = 409$)

consumption $\sim$ (drive + engine size + log(weight))^2



Cars2004nh (subset, $n = 409$)

consumption ∼ drive, engine size, log(weight)

```
anova(mAddit, mInter)
```

```
Model 1: consumption ~ fdrive + engine.size + lweight
Model 2: consumption ~ (fdrive + engine.size + lweight)^2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    404 343.69
2    399 303.78  5    39.906 10.483 1.813e-09 ***
```

```
anova(mInter1, mInter)
```

```
Model 1: consumption ~ fdrive + engine.size + lweight + fdrive:lweight
Model 2: consumption ~ (fdrive + engine.size + lweight)^2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    402 316.81
2    399 303.78  3    13.027 5.7034 0.0007864 ***
```

```
anova(mInter2, mInter)
```

```
Model 1: consumption ~ fdrive + engine.size + lweight + fdrive:lweight +
    engine.size:lweight
Model 2: consumption ~ (fdrive + engine.size + lweight)^2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    401 305.70
2    399 303.78  2    1.9215 1.2619 0.2842
```

5.5.5 ANOVA tables

`consumption ~ drive + log(weight) + drive:log(weight)`

Certain ANOVA table for the model:

$$m(z, w) = \beta_0 + \beta_1\,\mathbb{I}[z = \text{rear}] + \beta_2\,\mathbb{I}[z = \text{4x4}] + \beta_3\,\log(w)$$

$$+ \beta_4\,\mathbb{I}[z = \text{rear}]\,\log(w) + \beta_5\,\mathbb{I}[z = \text{4x4}]\,\log(w)$$

```
mInter1 <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
anova(mInter1)
```

```
Analysis of Variance Table

Response: consumption
                Df Sum Sq Mean Sq  F value    Pr(>F)
fdrive           2 519.89  259.94  293.935 < 2.2e-16 ***
lweight          1 954.26  954.26 1079.040 < 2.2e-16 ***
fdrive:lweight   2  26.70   13.35   15.097 4.758e-07 ***
Residuals      403 356.40    0.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Illustration for a model

$$\mathrm{M}_{AB}\colon \sim \mathrm{A} + \mathrm{B} + \mathrm{A}:\mathrm{B}.$$

5.5.5 ANOVA tables

Type I (sequential) ANOVA table

Order A + B + A:B

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
A	$\star$	$\mathrm{SS}(\mathrm{A} \mid 1)$	$\star$	$\star$	$\star$
B	$\star$	$\mathrm{SS}(\mathrm{A}+\mathrm{B} \mid \mathrm{A})$	$\star$	$\star$	$\star$
A:B	$\star$	$\mathrm{SS}(\mathrm{A}+\mathrm{B}+\mathrm{A}:\mathrm{B} \mid \mathrm{A}+\mathrm{B})$	$\star$	$\star$	$\star$
Residual	ν_e	SS_e	MS_e		

5.5.5 ANOVA tables

Type I (sequential) ANOVA table

Order B + A + A:B

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
B	$\star$	$\text{SS}(\text{B} \mid 1)$	$\star$	$\star$	$\star$
A	$\star$	$\text{SS}(\text{A} + \text{B} \mid \text{B})$	$\star$	$\star$	$\star$
A:B	$\star$	$\text{SS}(\text{A} + \text{B} + \text{A:B} \mid \text{A} + \text{B})$	$\star$	$\star$	$\star$
Residual	ν_e	SS_e	MS_e		

5.5.5 ANOVA tables

Type I (sequential) ANOVA table

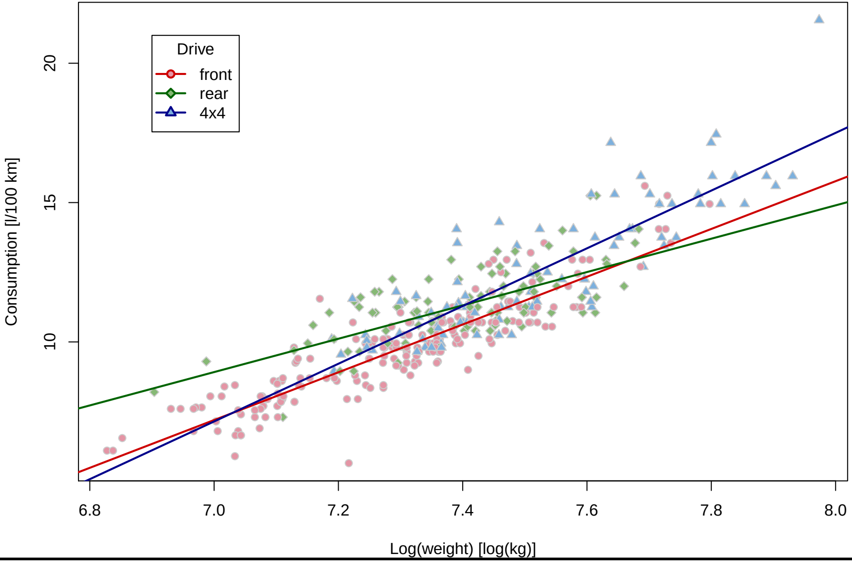
The row of the effect (term) E

- Comparison of two models $M_1 \subset M_2$
  - M_1 contains all terms included in the rows that precede the row of the term E.
  - M_2 contains the terms of model M_1 and additionally the term E.
- The sum of squares shows increase of the explained variability of the response due to the term E on top of the terms shown on the preceding rows.
- The p-value provides a significance of the influence of the term E on the response while controlling (adjusting) for all terms shown on the preceding rows.

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight) + drive:log(weight),`

$\widehat{m}(z, w) = -52.80 + 19.84\,\mathbb{I}[z = \text{rear}] - 12.54\,\mathbb{I}[z = \text{4x4}] + 8.57\,\log(w) - 2.59\,\mathbb{I}[z = \text{rea}$



`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight) + drive:log(weight)`

Reference group pseudocontrasts for `drive`

$$m(z, w) = \beta_0 + \beta_1\, \mathbb{I}[z = \text{rear}] + \beta_2\, \mathbb{I}[z = \text{4x4}] + \beta_3\, \log(w)$$
$$+ \beta_4\, \mathbb{I}[z = \text{rear}]\, \log(w) + \beta_5\, \mathbb{I}[z = \text{4x4}]\, \log(w)$$

```
mInter1 <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
anova(mInter1)
```

```
Analysis of Variance Table

Response: consumption
                Df Sum Sq Mean Sq  F value    Pr(>F)
fdrive           2 519.89  259.94  293.935 < 2.2e-16 ***
lweight          1 954.26  954.26 1079.040 < 2.2e-16 ***
fdrive:lweight   2  26.70   13.35   15.097 4.758e-07 ***
Residuals      403 356.40    0.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cars2004nh (subset, $n = 409$)

`consumption ~ log(weight) + drive + drive:log(weight)`

Reference group pseudocontrasts for drive

$$m(z, w) = \beta_0 + + \beta_1 \log(w) + \beta_2 \mathbb{I}[z = \text{rear}] + \beta_3 \mathbb{I}[z = \text{4x4}]$$

$$+ \beta_4 \mathbb{I}[z = \text{rear}] \log(w) + \beta_5 \mathbb{I}[z = \text{4x4}] \log(w)$$

```
mInter2 <- lm(consumption ~ lweight + fdrive + fdrive:lweight, data = CarsNow)
anova(mInter2)
```

```
Analysis of Variance Table

Response: consumption
                Df  Sum Sq Mean Sq  F value    Pr(>F)
lweight          1 1421.57 1421.57 1607.458 < 2.2e-16 ***
fdrive           2   52.58   26.29   29.726 9.079e-13 ***
lweight:fdrive   2   26.70   13.35   15.097 4.758e-07 ***
Residuals      403  356.40    0.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.5.5 ANOVA tables

Type II ANOVA table

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
A	$\star$	$SS(A + B \mid B)$	$\star$	$\star$	$\star$
B	$\star$	$SS(A + B \mid A)$	$\star$	$\star$	$\star$
A:B	$\star$	$SS(A + B + A{:}B \mid A + B)$	$\star$	$\star$	$\star$
Residual	ν_e	SS_e	MS_e		

Type II ANOVA table

The row of the effect (term) E

- Comparison of two models $M_1 \subset M_2$
  - M_1 is the considered (full) model without the term E and also all *higher order* terms than E that include E.
  - M_2 contains the terms of model M_1 and additionally the term E (this is the same as in type I ANOVA table).
- The sum of squares shows increase of the explained variability of the response due to the term E on top of all other terms that do not include the term E.
- The p-value provides a significance of the influence of the term E on the response while controlling (adjusting) for all other terms that do not include E.

Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight) + drive:log(weight)

Reference group pseudocontrasts for drive

$$m(z, w) = \beta_0 + \beta_1 \mathbb{I}[z = \text{rear}] + \beta_2 \mathbb{I}[z = \text{4x4}] + \beta_3 \log(w) + \beta_4 \mathbb{I}[z = \text{rear}] \log(w) + \beta_5 \mathbb{I}[z = \text{4x4}] \log(w)$$

```
mInter1 <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
car::Anova(mInter1, type = "II")
```

```
Anova Table (Type II tests)

Response: consumption
               Sum Sq  Df  F value    Pr(>F)
fdrive          52.58   2   29.726 9.079e-13 ***
lweight        954.26   1 1079.040 < 2.2e-16 ***
fdrive:lweight  26.70   2   15.097 4.758e-07 ***
Residuals      356.40 403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cars2004nh (subset, $n = 409$)

consumption ~ log(weight) + drive + drive:log(weight)

Reference group pseudocontrasts for drive

$$m(z, w) = \beta_0 + + \beta_1 \log(w) + \beta_2 \mathbb{I}[z = \text{rear}] + \beta_3 \mathbb{I}[z = \text{4x4}] + \beta_4 \mathbb{I}[z = \text{rear}] \log(w) + \beta_5 \mathbb{I}[z = \text{4x4}] \log(w)$$

```
mInter2 <- lm(consumption ~ lweight + fdrive + fdrive:lweight, data = CarsNow)
car::Anova(mInter2, type = "II")
```

```
Anova Table (Type II tests)

Response: consumption
               Sum Sq  Df  F value    Pr(>F)
lweight        954.26   1 1079.040 < 2.2e-16 ***
fdrive          52.58   2   29.726 9.079e-13 ***
fdrive:lweight  26.70   2   15.097 4.758e-07 ***
Residuals      356.40 403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.5.5 ANOVA tables

Type III ANOVA table

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
A	$\star$	$\mathrm{SS}(A + B + A{:}B \mid B + A{:}B)$	$\star$	$\star$	$\star$
B	$\star$	$\mathrm{SS}(A + B + A{:}B \mid A + A{:}B)$	$\star$	$\star$	$\star$
A:B	$\star$	$\mathrm{SS}(A + B + A{:}B \mid A + B)$	$\star$	$\star$	$\star$
Residual	ν_e	SS_e	MS_e		

5.5.5 ANOVA tables

Type III ANOVA table

The row of the effect (term) E

- Comparison of two models $M_1 \subset M_2$
  - M_1 is the considered (full) model without the term E.
  - M_2 contains the terms of model M_1 and additionally the term E (this is the same as in type I and type II ANOVA table). Due to the construction of M_1, the model M_2 is always equal to the considered (full) model.
- The submodel M_1 is not necessarily hierarchically well formulated.
- If M_1 is not HWF, interpretation of its comparison to model M_2 may depend on parameterizations of covariates included in the full model M_2. Consequently, also the interpretation of the F-test depends on the used parameterization.

`Cars2004nh` (subset, $n = 409$)

`consumption ~ drive + log(weight) + drive:log(weight)`

Reference (first) group pseudocontrasts for `drive`

$$m(z, w) = \beta_0 + \beta_1\, \mathbb{I}[z = \text{rear}] + \beta_2\, \mathbb{I}[z = \text{4x4}] + \beta_3\, \log(w) + \beta_4\, \mathbb{I}[z = \text{rear}]\, \log(w) + \beta_5\, \mathbb{I}[z = \text{4x4}]\, \log(w)$$

- β_3: slope of $\log(w)$ in group $z = \text{front}$

```
mInter <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow)
car::Anova(mInter, type = "III")
```

```
Anova Table (Type III tests)

Response: consumption
               Sum Sq  Df F value    Pr(>F)
(Intercept)    386.28   1 436.793 < 2.2e-16 ***
fdrive          26.49   2  14.979 5.310e-07 ***
lweight        542.30   1 613.216 < 2.2e-16 ***
fdrive:lweight  26.70   2  15.097 4.758e-07 ***
Residuals      356.40 403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

consumption ∼ drive + log(weight) + drive:log(weight)

Reference (last) group pseudocontrasts for drive

$$m(z, w) = \beta_0 + \beta_1 \mathbb{I}[z = \text{front}] + \beta_2 \mathbb{I}[z = \text{rear}] + \beta_3 \log(w) + \beta_4 \mathbb{I}[z = \text{front}] \log(w) + \beta_5 \mathbb{I}[z = \text{rear}] \log(w)$$

- β_3: slope of $\log(w)$ in group $z = 4\text{x}4$

```
mInterSAS <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow,
                contrasts = list(fdrive = contr.SAS))
car::Anova(mInterSAS, type = "III")
```

```
Anova Table (Type III tests)

Response: consumption
               Sum Sq  Df F value    Pr(>F)
(Intercept)    247.68   1 280.063 < 2.2e-16 ***
fdrive          26.49   2  14.979 5.310e-07 ***
lweight        351.72   1 397.714 < 2.2e-16 ***
fdrive:lweight  26.70   2  15.097 4.758e-07 ***
Residuals      356.40 403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cars2004nh (subset, $n = 409$)

consumption ~ drive + log(weight) + drive:log(weight)

Sum contrasts for drive

$$m(z, w) = \beta_0 + \beta_1\,\mathbb{I}[z = \text{front}] + \beta_2\,\mathbb{I}[z = \text{rear}] - (\beta_1 + \beta_2)\,\mathbb{I}[z = \text{4x4}] + \beta_3\,\log(w)$$
$$+ \beta_4\,\mathbb{I}[z = \text{front}]\,\log(w) + \beta_5\,\mathbb{I}[z = \text{rear}]\,\log(w) - (\beta_4 + \beta_5)\,\mathbb{I}[z = \text{4x4}]\,\log(w)$$

- β_3: mean of the slopes of $\log(w)$ in the three drive groups

```
mIntersum <- lm(consumption ~ fdrive + lweight + fdrive:lweight, data = CarsNow,
                contrasts = list(fdrive = contr.sum))
car::Anova(mIntersum, type = "III")
```

```
Anova Table (Type III tests)

Response: consumption
               Sum Sq  Df F value    Pr(>F)
(Intercept)    485.88   1 549.416 < 2.2e-16 ***
fdrive          26.49   2  14.979 5.310e-07 ***
lweight        728.22   1 823.440 < 2.2e-16 ***
fdrive:lweight  26.70   2  15.097 4.758e-07 ***
Residuals      356.40 403
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

6

Normal Linear Model

Section **6.1**

Normal linear model

6.1 Normal linear model

Definition 6.1 Normal linear model with general data.

The data $(\boldsymbol{Y}, \mathbb{X})$, satisfy a *normal linear model* if

$$\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big),$$

where $\boldsymbol{\beta} = \big(\beta_0, \ldots, \beta_{k-1}\big)^\top \in \mathbb{R}^k$ and $0 < \sigma^2 < \infty$ are unknown parameters.

6.1 Normal linear model

Lemma 6.1 Error terms in a normal linear model.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\beta,\, \sigma^2\, \mathbf{I}_n\big)$. *The error terms*

$$\boldsymbol{\varepsilon} \;=\; \boldsymbol{Y} - \mathbb{X}\beta \;=\; \big(Y_1 - \boldsymbol{X}_1^\top\beta,\, \ldots,\, Y_n - \boldsymbol{X}_n^\top\beta\big)^\top \;=\; \big(\varepsilon_1,\, \ldots,\, \varepsilon_n\big)^\top$$

then satisfy

(i) $\boldsymbol{\varepsilon} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\, \mathbf{I}_n\big)$.

(ii) $\boldsymbol{\varepsilon} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\, \mathbf{I}_n\big)$.

(iii) $\varepsilon_i \overset{i.i.d.}{\sim} \varepsilon,\; i = 1, \ldots, n,\; \varepsilon \sim \mathcal{N}\big(0,\, \sigma^2\big)$.

Section **6.2**

Properties of the least squares estimators under the normality

6.2 Properties of the LSE under the normality

Theorem 6.2 Least squares estimators under the normality.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n\big)$, $\text{rank}\big(\mathbb{X}_{n\times k}\big) = r \leq k$. *Let* $\mathbb{L}_{m\times k}$ *be a real matrix with non-zero rows* $\mathbf{l}_1^\top, \ldots, \mathbf{l}_m^\top$ *and* $\boldsymbol{\theta} := \mathbb{L}\boldsymbol{\beta} = \big(\mathbf{l}_1^\top\boldsymbol{\beta}, \ldots, \mathbf{l}_m^\top\boldsymbol{\beta}\big)^\top = \big(\theta_1, \ldots, \theta_m\big)^\top$ *be a vector of linear combinations of regression parameters.*

If additionally $r = k$, *let* $\widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}$ *be the least squares estimator of regression coefficients,* $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}} = \big(\mathbf{l}_1^\top\widehat{\boldsymbol{\beta}}, \ldots, \mathbf{l}_m^\top\widehat{\boldsymbol{\beta}}\big)^\top = \big(\widehat{\theta}_1, \ldots, \widehat{\theta}_m\big)^\top$ *and*

$$\mathbb{V} = \mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top = \big(v_{j,t}\big)_{j,t=1,\ldots,m},$$

$$\mathbb{D} = \text{diag}\left(\frac{1}{\sqrt{v_{1,1}}}, \ldots, \frac{1}{\sqrt{v_{m,m}}}\right),$$

$$T_j = \frac{\widehat{\theta}_j - \theta_j}{\sqrt{\text{MS}_e\, v_{j,j}}}, \qquad j = 1, \ldots, m,$$

$$\boldsymbol{T} = \big(T_1, \ldots, T_m\big)^\top = \frac{1}{\sqrt{\text{MS}_e}}\,\mathbb{D}\,\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big).$$

TO BE CONTINUED.

6.2 Properties of the LSE under the normality

Theorem 6.2 Least squares estimators under the normality, cont'd.

The following then holds.

(i) $\widehat{\boldsymbol{Y}} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbb{H}\big)$.

(ii) $\boldsymbol{U} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\,\mathbb{M}\big)$.

(iii) $\widehat{\boldsymbol{\theta}} \,\big|\, \mathbb{X} \sim \mathcal{N}_m\big(\boldsymbol{\theta},\, \sigma^2\,\mathbb{V}\big)$.

(iv) *Statistics $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally, given $\mathbb{X}$, independent.*

(v) *Statistics $\widehat{\boldsymbol{\theta}}$ and SS_e are conditionally, given $\mathbb{X}$, independent.*

(vi) $\dfrac{\big\|\widehat{\boldsymbol{Y}} - \mathbb{X}\boldsymbol{\beta}\big\|^2}{\sigma^2} \;\sim\; \chi^2_r$.

(vii) $\dfrac{\mathrm{SS}_e}{\sigma^2} \;\sim\; \chi^2_{n-r}$.

TO BE CONTINUED.

6.2 Properties of the LSE under the normality

Theorem 6.2 Least squares estimators under the normality, cont'd.

(viii) *For each* $j = 1, \ldots, m$, $\qquad T_j \sim \mathrm{t}_{n-r}$.

(ix) $\boldsymbol{T} \,|\, \mathbb{X} \sim \mathrm{mvt}_{m,n-r}\big(\mathbb{D}\mathbb{V}\mathbb{D}\big)$.

(x) *If additionally* $\mathrm{rank}\big(\mathbb{L}_{m\times k}\big) = m \leq r = k$ *then the matrix* $\mathbb{V}$ *is invertible and*

$$\frac{1}{m}\,\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big)^{\top}\,\big(\mathrm{MS}_e\,\mathbb{V}\big)^{-1}\,\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big) \sim \mathcal{F}_{m,\,n-r}.$$

6.2 Properties of the LSE under the normality

Consequence of Theorem 6.2: Least squares estimator of the regression coefficients in a full-rank normal linear model.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n\big)$, $\text{rank}\big(\mathbb{X}_{n\times k}\big) = k$. *Further, let*

$$\mathbb{V} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1} = \big(v_{j,t}\big)_{j,t=0,\dots,k-1},$$

$$\mathbb{D} = \text{diag}\left(\frac{1}{\sqrt{v_{0,0}}},\, \dots,\, \frac{1}{\sqrt{v_{k-1,k-1}}}\right).$$

The following then holds.

(i) $\widehat{\boldsymbol{\beta}} \,\big|\, \mathbb{X} \sim \mathcal{N}_k\big(\boldsymbol{\beta},\, \sigma^2\, \mathbb{V}\big)$.

(ii) *Statistics* $\widehat{\boldsymbol{\beta}}$ *and* SS_e *are conditionally, given* $\mathbb{X}$, *independent*.

(iii) *For each* $j = 0, \dots, k-1$, $T_j := \dfrac{\widehat{\beta}_j - \beta_j}{\sqrt{\text{MS}_e\, v_{j,j}}} \;\sim\; \text{t}_{n-k}$.

(iv) $\boldsymbol{T} := \big(T_0, \dots, T_{k-1}\big)^\top = \dfrac{1}{\sqrt{\text{MS}_e}}\, \mathbb{D}\, \big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) \;\sim\; \text{mvt}_{k,n-k}\big(\mathbb{D}\mathbb{V}\mathbb{D}\big)$, *conditionally given* $\mathbb{X}$.

(v) $\dfrac{1}{k}\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)^\top \text{MS}_e^{-1}\, \mathbb{X}^\top\mathbb{X}\, \big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) \;\sim\; \mathcal{F}_{k,\,n-k}$.

Inference on a chosen regression coefficient

$\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n\big)$, $\mathrm{rank}\big(\mathbb{X}_{n\times k}\big) = k$, $j \in \{0, \ldots, k-1\}$, $\mathbb{V} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}$

LSE of β_j: $\widehat{\beta}_j = \Big\{\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}\Big\}_j$,

Standard error: $\mathrm{S.E.}\big(\widehat{\beta}_j\big) = \sqrt{\mathrm{MS}_e\, v_{j,j}}$,

$(1-\alpha)\,100\%$ **CI**: $\big(\beta_j^L,\, \beta_j^U\big) \equiv \widehat{\beta}_j \pm \mathrm{S.E.}\big(\widehat{\beta}_j\big)\, \mathrm{t}_{n-k}\big(1-\frac{\alpha}{2}\big)$.

Test of H_0: $\beta_j = \beta_j^0$ against H_1: $\beta_j \neq \beta_j^0$ $(\beta_j^0 \in \mathbb{R})$

Test statistic:
$$T_{j,0} = \frac{\widehat{\beta}_j - \beta_j^0}{\mathrm{S.E.}\big(\widehat{\beta}_j\big)} = \frac{\widehat{\beta}_j - \beta_j^0}{\sqrt{\mathrm{MS}_e\, v_{j,j}}}.$$

Reject H_0 if
$$\big|T_{j,0}\big| \geq t_{n-k}\big(1-\tfrac{\alpha}{2}\big).$$

P-value when $T_{j,0} = t_{j,0}$:
$$p = 2\,\mathrm{CDF}_{t,\,n-k}\big(-\,|t_{j,0}|\big).$$

6.2.1 Statistical inference in a full-rank normal linear model

Simultaneous inference on a vector of regression coefficients

$\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$

LSE of β: $\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top \boldsymbol{Y}$,

$(1-\alpha)\,100\%$ **CR**:

$$\mathcal{S}(\alpha) \;=\; \Big\{\boldsymbol{\beta} \in \mathbb{R}^k :\; \left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)^\top \left(\mathsf{MS}_e^{-1}\,\mathbb{X}^\top\mathbb{X}\right)\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right) \;<\; k\,\mathcal{F}_{k,n-k}(1-\alpha)\Big\},$$

ellipsoid with
- center: $\widehat{\boldsymbol{\beta}}$,
- shape matrix: $\mathsf{MS}_e\left(\mathbb{X}^\top\mathbb{X}\right)^{-1} = \widehat{\mathsf{var}}\left(\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbb{X}\right)$,
- diameter: $\sqrt{k\,\mathcal{F}_{k,n-k}(1-\alpha)}$.

Simultaneous inference on a vector of regression coefficients

$\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big)$, $\mathsf{rank}\big(\mathbb{X}_{n\times k}\big) = k$

Test of H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}^0$ against H_1: $\boldsymbol{\beta} \neq \boldsymbol{\beta}^0$ ($\boldsymbol{\beta}^0 \in \mathbb{R}^k$)

Test statistic: $$Q_0 = \frac{1}{k}\,\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\big)^\top \mathsf{MS}_e^{-1}\,\mathbb{X}^\top\mathbb{X}\,\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\big).$$

Reject H_0 if $Q_0 \geq \mathcal{F}_{k,n-k}(1-\alpha)$.

P-value when $Q_0 = q_0$: $p = 1 - \mathsf{CDF}_{\mathcal{F},\,k,n-k}\big(q_0\big)$.

Inference on a chosen linear combination

$\mathbf{Y} \mid \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbf{I}_n)$, $\text{rank}(\mathbb{X}_{n\times k}) = r = k$, $\theta = \mathbf{l}^\top\boldsymbol{\beta}$, $\mathbf{l} \neq \mathbf{0}$

LSE of θ: $\widehat{\theta} = \mathbf{l}^\top\widehat{\boldsymbol{\beta}}$,

Standard error: $\text{S.E.}(\widehat{\theta}) = \sqrt{\text{MS}_e\, \mathbf{l}^\top \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbf{l}}$,

$(1-\alpha)\,$**100% CI:** $\big(\theta^L,\, \theta^U\big) \equiv \widehat{\theta} \pm \text{S.E.}(\widehat{\theta})\, \text{t}_{n-k}\big(1-\frac{\alpha}{2}\big)$.

Test of H_0: $\theta = \theta^0$ against H_1: $\theta \neq \theta^0$ ($\theta^0 \in \mathbb{R}$)

Test statistic: $$T_0 = \frac{\widehat{\theta} - \theta^0}{\text{S.E.}(\widehat{\theta})} = \frac{\widehat{\theta} - \theta^0}{\sqrt{\text{MS}_e\, \mathbf{l}^\top \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbf{l}}}.$$

Reject H_0 if $|T_0| \geq t_{n-k}\big(1-\frac{\alpha}{2}\big)$.

P-value when $T_0 = t_0$: $p = 2\,\text{CDF}_{t,\,n-k}\big(-\,|t_0|\big)$.

Simultaneous inference on a set of linear combinations

$\boldsymbol{Y} \mid \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$, $\text{rank}(\mathbb{X}_{n\times k}) = r = k$, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, $\text{rank}(\mathbb{L}_{m\times k}) = m \le k$

LSE of $\boldsymbol{\theta}$: $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$,

$(1-\alpha)\,100\%$ **CR:**

$$\mathcal{S}(\alpha) = \Big\{\boldsymbol{\theta} \in \mathbb{R}^m :\ \big(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\big)^\top \Big\{\text{MS}_e\,\mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1} \big(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\big) < m\,\mathcal{F}_{m,n-k}(1-\alpha)\Big\},$$

ellipsoid with
- center: $\widehat{\boldsymbol{\theta}}$,
- shape matrix: $\text{MS}_e\,\mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top = \widehat{\text{var}}\big(\widehat{\boldsymbol{\theta}} \mid \mathbb{X}\big)$,
- diameter: $\sqrt{m\,\mathcal{F}_{m,n-k}(1-\alpha)}$.

Simultaneous inference on a set of linear combinations

$\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n)$, $\text{rank}(\mathbb{X}_{n\times k}) = r = k$, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, $\text{rank}(\mathbb{L}_{m\times k}) = m \leq k$

Test of H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ against H_1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ ($\boldsymbol{\theta}^0 \in \mathbb{R}^m$)

Test statistic:
$$Q_0 = \frac{1}{m}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^\top \left\{\mathrm{MS}_e\,\mathbb{L}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{L}^\top\right\}^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right).$$

Reject H_0 if $Q_0 \geq \mathcal{F}_{m,n-k}(1-\alpha)$.

P-value when $Q_0 = q_0$: $p = 1 - \mathrm{CDF}_{\mathcal{F},\, m,n-k}\left(q_0\right)$.

Section **6.3**

Confidence interval for the model based mean, prediction interval

6.3 Confidence interval . . . , prediction interval

Theorem 6.3 Confidence interval for the model based mean, prediction interval.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$ *(full-rank model),* $\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top\boldsymbol{Y}$ *is the LSE of the regression parameters* $\boldsymbol{\beta}$*. Let* $\boldsymbol{x}_{new} \in \mathcal{X}$, $\boldsymbol{x}_{new} \neq \mathbf{0}_k$*. Let* $\varepsilon_{new} \sim \mathcal{N}(0, \sigma^2)$ *is independent of* $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}$*. Finally, let* $Y_{new} = \boldsymbol{x}_{new}^\top\boldsymbol{\beta} + \varepsilon_{new}$*. The following then holds:*

(i) *The quantity* $\widehat{\mu}_{new} := \boldsymbol{x}_{new}^\top\widehat{\boldsymbol{\beta}}$ *is the best linear unbiased estimator (BLUE) of* $\mu_{new} := \boldsymbol{x}_{new}^\top\boldsymbol{\beta}$*. The standard error of* $\widehat{\mu}_{new}$ *is*

$$\mathrm{S.E.}\left(\widehat{\mu}_{new}\right) = \sqrt{\mathrm{MS}_e\, \boldsymbol{x}_{new}^\top \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\boldsymbol{x}_{new}}$$

and the lower and the upper bound of the $(1-\alpha)\,100\%$ *confidence interval for* μ_{new} *are*

$$\left(\mu_{new}^L,\, \mu_{new}^U\right) \equiv \widehat{\mu}_{new} \pm \mathrm{S.E.}\left(\widehat{\mu}_{new}\right)\, t_{n-k}\Big(1 - \frac{\alpha}{2}\Big).$$

TO BE CONTINUED.

6.3 Confidence interval for the model based mean, prediction interval

Theorem 6.3 Confidence interval for the model based mean, prediction interval, cont'd.

(ii) *A (random) interval with the bounds*

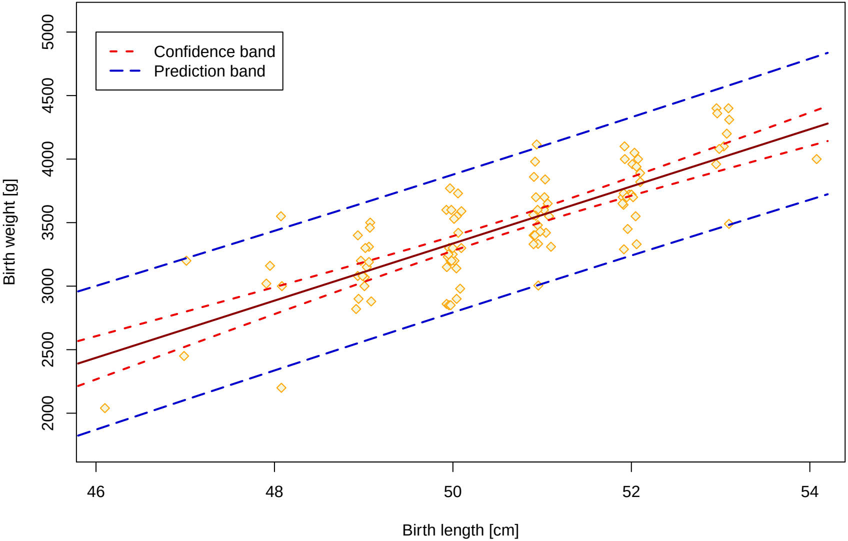
$$\left(Y_{new}^{L},\; Y_{new}^{U}\right) \equiv \widehat{\mu}_{new} \pm \text{S.E.P.}\left(\boldsymbol{x}_{new}\right) t_{n-k}\left(1 - \frac{\alpha}{2}\right),$$

where

$$\text{S.E.P.}\left(\boldsymbol{x}_{new}\right) = \sqrt{\text{MS}_e \left\{1 + \boldsymbol{x}_{new}^{\top}\left(\mathbb{X}^{\top}\mathbb{X}\right)^{-1}\boldsymbol{x}_{new}\right\}},$$

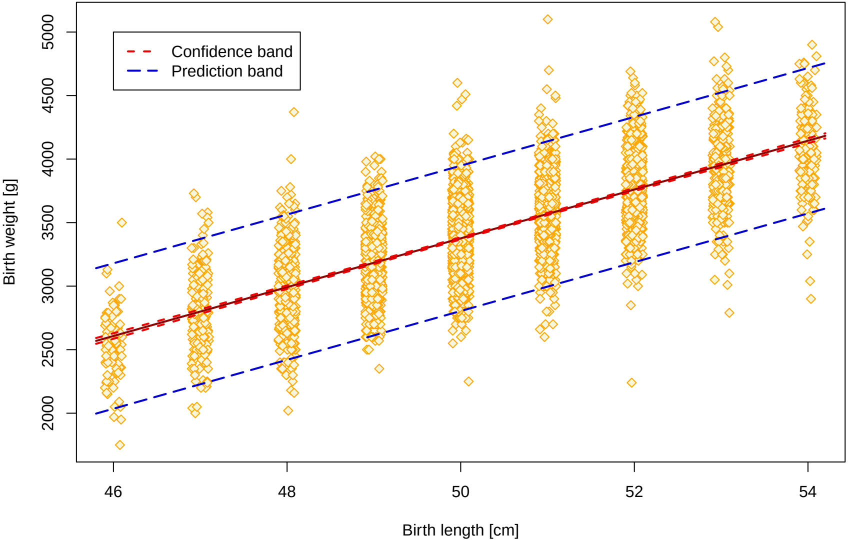
covers with the probability of $(1-\alpha)$ *the value of* Y_{new}*.*

`bweight` $\sim$ `blength`



Hosi0 ($n = 4838$)

bweight ∼ blength



Section **6.4**

Distribution of the linear hypotheses test statistics under the alternative

6.4 Distribution of the linear hypoth. test stat. under the alternative

Theorem 6.4 Distribution of the linear hypothesis test statistics under the alternative.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n),$ rank$(\mathbb{X}_{n\times k}) = k$. *Let* $\mathbf{l} \neq \mathbf{0}_k$ *and let* $\widehat{\theta} = \mathbf{l}^\top \widehat{\boldsymbol{\beta}}$ *be the LSE of the parameter* $\theta = \mathbf{l}^\top \boldsymbol{\beta}$. *Let* $\theta^0,\, \theta^1 \in \mathbb{R}$, $\theta^0 \neq \theta^1$ *and let*

$$T_0 = \frac{\widehat{\theta} - \theta^0}{\sqrt{\mathsf{MS}_e\, \mathbf{l}^\top \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbf{l}}}.$$

Then under the hypothesis $\theta = \theta^1$,

$$T_0 \,\big|\, \mathbb{X} \sim \mathrm{t}_{n-k}(\lambda), \qquad \lambda = \frac{\theta^1 - \theta^0}{\sqrt{\sigma^2\, \mathbf{l}^\top \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbf{l}}}.$$

6.4 Distribution of the linear hypoth. test stat. under the alternative

Theorem 6.5 Distribution of the linear hypotheses test statistics under the alternative.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. *Let* $\mathbb{L}_{m\times k}$ *be a real matrix with* $m \leq k$ *linearly independent rows. Let* $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ *be the LSE of the vector parameter* $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$. *Let* $\boldsymbol{\theta}^0,\, \boldsymbol{\theta}^1 \in \mathbb{R}^m$, $\boldsymbol{\theta}^0 \neq \boldsymbol{\theta}^1$ *and let*

$$Q_0 = \frac{1}{m}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\big)^\top \Big\{\text{MS}_e\, \mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\big).$$

Then under the hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}^1$,

$$Q_0 \,\big|\, \mathbb{X} \sim \mathcal{F}_{m,n-r}(\lambda), \qquad \lambda = \big(\boldsymbol{\theta}^1 - \boldsymbol{\theta}^0\big)^\top \Big\{\sigma^2\, \mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1}\big(\boldsymbol{\theta}^1 - \boldsymbol{\theta}^0\big).$$

7

Coefficient of Determination

Section **7.1**

Intercept only model

7.1 Intercept only model

Definition 7.1 Regression and total sums of squares in a linear model.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\operatorname{rank}(\mathbb{X}_{n\times k}) = r \leq k$. The following expressions define the following quantities:

(i) *Regression sum of squares* and corresponding degrees of freedom:

$$\mathrm{SS}_R = \big\| \widehat{\boldsymbol{Y}} - \overline{Y}\mathbf{1}_n \big\|^2 = \sum_{i=1}^{n} \big(\widehat{Y}_i - \overline{Y}\big)^2, \quad \nu_R = r - 1,$$

(ii) *Total sum of squares* and corresponding degrees of freedom:

$$\mathrm{SS}_T = \big\| \boldsymbol{Y} - \overline{Y}\mathbf{1}_n \big\|^2 = \sum_{i=1}^{n} \big(Y_i - \overline{Y}\big)^2, \quad \nu_T = n - 1.$$

7.1 Intercept only model

Lemma 7.1 Model with intercept only.

Let $\boldsymbol{Y} \sim \left(\mathbf{1}_n\gamma, \zeta^2\mathbf{I}_n\right)$. *Then*

(i) $\widehat{\boldsymbol{Y}} = \overline{Y}\mathbf{1}_n = \left(\overline{Y}, \ldots, \overline{Y}\right)^\top$.

(ii) $\mathrm{SS}_e = \mathrm{SS}_T$.

Section **7.2**

Models with intercept

7.2 Models with intercept

Lemma 7.2 Identity in a linear model with intercept.

Let $\boldsymbol{Y} \mid \mathbb{X} \sim (\mathbb{X}\beta,\, \sigma^2 \mathbf{I}_n)$ *where* $\mathbf{1}_n \in \mathcal{M}(\mathbb{X})$. *Then*

$$\mathbf{1}_n^\top \boldsymbol{Y} = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i = \mathbf{1}_n^\top \widehat{\boldsymbol{Y}}.$$

7.2 Models with intercept

Lemma 7.3 Breakdown of the total sum of squares in a linear model with intercept.

Let $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \left(\mathbb{X}\beta,\, \sigma^2 \mathbf{I}_n\right)$ *where* $\mathbf{1}_n \in \mathcal{M}\left(\mathbb{X}\right)$. *Then*

$$\begin{array}{ccccc} \mathsf{SS}_T & = & \mathsf{SS}_e & + & \mathsf{SS}_R \\ \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 & = & \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 & + & \sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{Y}\right)^2. \end{array}$$

Section **7.3**

Theoretical evaluation of a prediction quality of the model

7.3 Theoretical evaluation of a prediction quality of the model

Data: $\bigl(Y_i,\, \boldsymbol{X}_i^\top\bigr)^\top \overset{\text{i.i.d.}}{\sim} \bigl(Y,\, \boldsymbol{X}^\top\bigr)^\top$

Conditional response distribution

$$\mathbb{E}\bigl(Y \,\big|\, \boldsymbol{X}\bigr) = \boldsymbol{X}^\top\beta, \qquad \mathrm{var}\bigl(Y \,\big|\, \boldsymbol{X}\bigr) = \sigma^2,$$

$$\boldsymbol{Y} \,|\, \mathbb{X} \sim \bigl(\mathbb{X}\beta,\, \sigma^2 \mathbf{I}_n\bigr), \qquad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},\ \mathbb{X} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix}$$

Marginal response distribution

$$\mathbb{E}\bigl(Y\bigr) = \gamma, \qquad \mathrm{var}\bigl(Y\bigr) = \zeta^2,$$

$$\boldsymbol{Y} \sim \bigl(\mathbf{1}_n\gamma,\, \zeta^2 \mathbf{I}_n\bigr)$$

Section **7.4**

Coefficient of determination

7.4 Coefficient of determination

Unbiased estimators of the conditional and marginal resp. variances

$$\widehat{\sigma}^2 = \frac{1}{n-r}\mathrm{SS}_e = \frac{1}{n-r}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2,$$

$$\widehat{\zeta}^2 = \frac{1}{n-1}\mathrm{SS}_T = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

MLE of the conditional and marginal resp. variances under **normality**

$$\widehat{\sigma}^2_{ML} = \frac{1}{n}\mathrm{SS}_e = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2,$$

$$\widehat{\zeta}^2_{ML} = \frac{1}{n}\mathrm{SS}_T = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

7.4 Coefficient of determination

Definition 7.2 Coefficients of determination.

Consider a linear model $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\text{rank}(\mathbb{X}) = r$ where $\mathbf{1}_n \in \mathcal{M}(\mathbb{X})$.
A value

$$R^2 = 1 - \frac{\text{SS}_e}{\text{SS}_T}$$

is called the *coefficient of determination* of the linear model.
A value

$$R^2_{adj} = 1 - \frac{n-1}{n-r}\,\frac{\text{SS}_e}{\text{SS}_T}$$

is called the *adjusted coefficient of determination* of the linear model.

8

Submodels

Section **8.1**

Submodel

8.1 Submodel

Definition 8.1 Submodel.

We say that the model M_0 is the *submodel* (or the *nested model*) of the model M if

$$\mathcal{M}(\mathbb{X}^0) \subset \mathcal{M}(\mathbb{X}) \quad \text{with } r_0 < r.$$

Notation. Situation that a model M_0 is a submodel of a model M will be denoted as

$$\mathsf{M}_0 \subset \mathsf{M}.$$

8.1.1 Projection considerations

Orthonormal vector basis of $\mathbb{R}^n$

$$\begin{aligned} \mathbb{P}_{n\times n} &= (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) \\ &= (\mathbb{Q}^0, \mathbb{Q}^1, \mathbb{N}) \end{aligned}$$

$\mathbb{Q}^0_{n\times r_0}$: orthonormal vector basis of the submodel regression space

$\mathbb{Q}^1_{n\times(r-r_0)}$: orthonormal vectors such that $\mathbb{Q} := (\mathbb{Q}^0, \mathbb{Q}^1)$ is an orthonormal vector basis of the model regression space

$\mathbb{N}_{n\times(n-r)}$: orthonormal vector basis of the model residual space

$$\begin{aligned} \mathcal{M}(\mathbb{X}^0) &= \mathcal{M}(\mathbb{Q}^0) \\ \mathcal{M}(\mathbb{X}) &= \mathcal{M}(\mathbb{Q}) = \mathcal{M}\big((\mathbb{Q}^0, \mathbb{Q}^1)\big) \\ \mathcal{M}(\mathbb{X})^{\perp} &= \mathcal{M}(\mathbb{N}) \end{aligned}$$

8.1.2 Properties of submodel related quantities

Notation *(Quantities related to a submodel).*

- $\widehat{\boldsymbol{Y}}^0 = \mathbb{H}^0 \boldsymbol{Y} = \mathbb{Q}^0 {\mathbb{Q}^0}^\top \boldsymbol{Y}$:

  fitted values in the submodel (projection of $\boldsymbol{Y}$ into the submodel regression space).

- $\boldsymbol{U}^0 = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}^0 = \mathbb{M}^0 \boldsymbol{Y} = \big(\mathbb{Q}^1 {\mathbb{Q}^1}^\top + \mathbb{N}\mathbb{N}^\top\big) \boldsymbol{Y}$:

  residuals of the submodel.

- $\mathrm{SS}_e^0 = \big\|\boldsymbol{U}^0\big\|^2$:

  residual sum of squares of the submodel.

- $\nu_e^0 = n - r_0$: submodel residual degrees of freedom.

- $\mathrm{MS}_e^0 = \dfrac{\mathrm{SS}_e^0}{\nu_e^0}$: submodel residual mean square.

- $\boldsymbol{D}$: projection of the response vector $\boldsymbol{Y}$ into the space $\mathcal{M}\big(\mathbb{Q}^1\big)$

$$\boldsymbol{D} = \mathbb{Q}^1 {\mathbb{Q}^1}^\top \boldsymbol{Y} = \widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}^0 = \boldsymbol{U}^0 - \boldsymbol{U}.$$

8.1.2 Properties of submodel related quantities

Theorem 8.1 On a submodel.

Consider two linear models $\mathsf{M}\;:\;\; \boldsymbol{Y}\,|\,\mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big)$ *and* $\mathsf{M}_0\;:\;\; \boldsymbol{Y}\,|\,\mathbb{Z} \sim \big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\,\mathbf{I}_n\big)$ *such that* $\mathsf{M}_0 \subset \mathsf{M}$. *Let the submodel* M_0 *holds, i.e., let* $\mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}^0\big)$. *Then*

(i) $\widehat{\boldsymbol{Y}}^0$ *is the best linear unbiased estimator (BLUE) of a vector parameter* $\boldsymbol{\mu}^0 = \mathbb{X}^0\boldsymbol{\beta}^0 = \mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{Z}\big)$.

(ii) *The submodel residual mean square* MS_e^0 *is the unbiased estimator of the residual variance* σ^2.

(iii) *Statistics* $\widehat{\boldsymbol{Y}}^0$ *and* $\boldsymbol{U}^0$ *are conditionally, given* $\mathbb{Z}$, *uncorrelated.*

(iv) *A random vector* $\boldsymbol{D} = \widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}^0 = \boldsymbol{U}^0 - \boldsymbol{U}$ *satisfies*

$$\big\|\boldsymbol{D}\big\|^2 = \mathsf{SS}_e^0 - \mathsf{SS}_e.$$

TO BE CONTINUED.

Theorem 8.1 On a submodel, cont'd.

(v) *If additionally, a normal linear model is assumed, i.e., if* $\boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\,\mathbf{I}_n\big)$ *then the statistics* $\widehat{\boldsymbol{Y}}^0$ *and* $\boldsymbol{U}^0$ *are conditionally, given* $\mathbb{Z}$, *independent and*

$$F_0 = \frac{\dfrac{\mathrm{SS}_e^0 - \mathrm{SS}_e}{r - r_0}}{\dfrac{\mathrm{SS}_e}{n - r}} = \frac{\dfrac{\mathrm{SS}_e^0 - \mathrm{SS}_e}{\nu_e^0 - \nu_e}}{\dfrac{\mathrm{SS}_e}{\nu_e}} \sim \mathcal{F}_{r - r_0,\, n - r} = \mathcal{F}_{\nu_e^0 - \nu_e,\, \nu_e}.$$

8.1.3 Series of submodels

$$\text{Model } \mathrm{M}_0: \; \boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X}^0 \boldsymbol{\beta}^0, \, \sigma^2 \, \mathbf{I}_n\right),$$

$$\text{Model } \mathrm{M}_1: \; \boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X}^1 \boldsymbol{\beta}^1, \, \sigma^2 \, \mathbf{I}_n\right),$$

$$\text{Model } \mathrm{M}: \; \boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X} \boldsymbol{\beta}, \, \sigma^2 \, \mathbf{I}_n\right),$$

8.1.3 Series of submodels

Notation. Quantities derived while assuming a particular model

- $\widehat{\boldsymbol{Y}}^0$, $\boldsymbol{U}^0$, SS_e^0, ν_e^0, MS_e^0:
  quantities based on the (sub)model M_0: $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\mathbf{I}_n\big)$;

- $\widehat{\boldsymbol{Y}}^1$, $\boldsymbol{U}^1$, SS_e^1, ν_e^1, MS_e^1:
  quantities based on the (sub)model M_1: $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}^1\boldsymbol{\beta}^1,\, \sigma^2\mathbf{I}_n\big)$;

- $\widehat{\boldsymbol{Y}}$, $\boldsymbol{U}$, SS_e, ν_e, MS_e:
  quantities based on the model M: $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big)$.

Theorem 8.2 On submodels.

Consider three normal linear models $\mathsf{M}: \ \boldsymbol{Y}\,|\,\mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big)$, $\mathsf{M}_1: \ \boldsymbol{Y}\,|\,\mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^1\boldsymbol{\beta}^1,\, \sigma^2\,\mathbf{I}_n\big)$, $\mathsf{M}_0: \ \boldsymbol{Y}\,|\,\mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\,\mathbf{I}_n\big)$ *such that* $\mathsf{M}_0 \subset \mathsf{M}_1 \subset \mathsf{M}$. *Let the (smallest) submodel* M_0 *hold, i.e., let* $\mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}^0\big)$. *Then*

$$F_{0,1} = \frac{\dfrac{\mathsf{SS}_e^0 - \mathsf{SS}_e^1}{r_1 - r_0}}{\dfrac{\mathsf{SS}_e}{n - r}} = \frac{\dfrac{\mathsf{SS}_e^0 - \mathsf{SS}_e^1}{\nu_e^0 - \nu_e^1}}{\dfrac{\mathsf{SS}_e}{\nu_e}} \sim \mathcal{F}_{r_1 - r_0,\, n-r} = \mathcal{F}_{\nu_e^0 - \nu_e^1,\, \nu_e}.$$

Notation *(Differences when dealing with a submodel).*

M_A and M_B: two models distinguished by symbols "A" and "B" such that $\mathsf{M}_A \subset \mathsf{M}_B$.

$$\boldsymbol{D}\bigl(\mathsf{M}_B \,\big|\, \mathsf{M}_A\bigr) = \boldsymbol{D}\bigl(B \,\big|\, A\bigr) := \widehat{\boldsymbol{Y}}^B - \widehat{\boldsymbol{Y}}^A = \boldsymbol{U}^A - \boldsymbol{U}^B.$$

$$\mathsf{SS}\bigl(\mathsf{M}_B \,\big|\, \mathsf{M}_A\bigr) = \mathsf{SS}\bigl(B \,\big|\, A\bigr) := \mathsf{SS}_e^A - \mathsf{SS}_e^B.$$

8.1.4 Statistical test to compare nested models

F-test on a submodel based on Theorem 8.1

Consider two *normal* linear models:

Model M_0: $\quad \boldsymbol{Y} \,|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\,\mathbf{I}_n\big),$

Model M: $\quad \boldsymbol{Y} \,|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big),$

where $\mathrm{M}_0 \subset \mathrm{M}$, and a set of statistical hypotheses:

H_0: $\quad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}^0\big)$

H_1: $\quad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}\big) \setminus \mathcal{M}\big(\mathbb{X}^0\big),$

8.1.4 Statistical test to compare nested models

F-test on a submodel based on Theorem 8.2

Consider three *normal* linear models:

Model M_0: $\quad \boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^0\boldsymbol{\beta}^0,\, \sigma^2\,\mathbf{I}_n\big)$,

Model M_1: $\quad \boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}^1\boldsymbol{\beta}^1,\, \sigma^2\,\mathbf{I}_n\big)$,

Model M: $\quad \boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big)$,

where $\mathsf{M}_0 \subset \mathsf{M}_1 \subset \mathsf{M}$, and a set of statistical hypotheses:

H_0: $\quad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}^0\big)$

H_1: $\quad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}^1\big) \setminus \mathcal{M}\big(\mathbb{X}^0\big)$,

Section **8.2**

Omitting some regressors

8.2 Omitting some regressors

Lemma 8.3 Effect of omitting some regressors.

Consider a couple (model – submodel), where the submodel is obtained by omitting some regressors from the model. The following then holds.

(i) *If* $\mathcal{M}(\mathbb{X}^1) \perp \mathcal{M}(\mathbb{X}^0)$ *then*

$$\boldsymbol{D} = \mathbb{X}^1 \big(\mathbb{X}^{1\top}\mathbb{X}^1\big)^{-1}\mathbb{X}^{1\top}\boldsymbol{Y} =: \widehat{\boldsymbol{Y}}^1,$$

which are the fitted values from a linear model $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}^1\boldsymbol{\beta}^1,\, \sigma^2\mathbf{I}_n\big)$.

(ii) *If for given* $\mathbb{Z}$, *the conditional distribution* $\boldsymbol{Y} \,\big|\, \mathbb{Z}$ *is continuous, i.e., has a density with respect to the Lebesgue measure on* $\big(\mathbb{R}^n,\, \mathcal{B}_n\big)$ *then*

$$\boldsymbol{D} \neq \mathbf{0}_n \quad \textit{and} \;\; \mathrm{SS}_e^0 - \mathrm{SS}_e > 0 \;\; \textit{almost surely.}$$

Section **8.3**

Linear constraints

8.3 Linear constraints

Definition 8.2 Submodel given by linear constraints.

We say that the model M_0 is a *submodel given by linear constraints* $\mathbb{L}\boldsymbol{\beta} = \boldsymbol{\theta}^0$ of model M: $\boldsymbol{Y} \mid \mathbb{Z} \sim (\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$, if the response expectation $\mathbb{E}(\boldsymbol{Y} \mid \mathbb{Z})$ under the model M_0 is assumed to lie in a space $\mathcal{M}(\mathbb{X};\, \mathbb{L}\boldsymbol{\beta} = \boldsymbol{\theta}^0)$, where $\mathbb{L}_{m\times k}$ is a real matrix with m linearly independent rows, $m < k$ and $\boldsymbol{\theta}^0 \in \mathbb{R}^m$ is a given vector.

Notation. A submodel given by linear constraints will be denoted as

$$\mathsf{M}_0 : \boldsymbol{Y} \mid \mathbb{Z} \sim (\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n),\ \mathbb{L}\boldsymbol{\beta} = \boldsymbol{\theta}^0.$$

8.3 Linear constraints

Definition 8.3 Fitted values, residuals, residual sum of squares, rank of the model and residual degrees of freedom in a submodel given by linear constraints.

Let $\boldsymbol{b}^0 \in \mathbb{R}^k$ minimize $\mathsf{SS}(\beta) = \big\|\boldsymbol{Y} - \mathbb{X}\beta\big\|^2$ over $\beta \in \mathbb{R}^k$ subject to $\mathbb{L}\beta = \boldsymbol{\theta}^0$. For the submodel $\mathsf{M}_0 : \boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}\beta,\, \sigma^2\mathbf{I}_n\big),\; \mathbb{L}\beta = \boldsymbol{\theta}^0$, the following quantities are defined as follows:

Fitted values:	$\widehat{\boldsymbol{Y}}^0 := \mathbb{X}\boldsymbol{b}^0$.
Residuals:	$\boldsymbol{U}^0 := \boldsymbol{Y} - \widehat{\boldsymbol{Y}}^0$.
Residual sum of squares:	$\mathsf{SS}_e^0 := \big\|\boldsymbol{U}^0\big\|^2$.
Rank of the model:	$r_0 = k - m$.
Residual degrees of freedom:	$\nu_e^0 := n - r_0$.

8.3 Linear constraints

Theorem 8.4 On a submodel given by linear constraints.

Let $\mathsf{M}_0 \,:\, \boldsymbol{Y}\,|\,\mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big),\; \mathbb{L}\boldsymbol{\beta} = \boldsymbol{\theta}^0$ *be a submodel given by linear constraints of a model* $\mathsf{M} : \boldsymbol{Y}\,|\,\mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big)$. *Then*

(i) *There is a unique minimizer* $\boldsymbol{b}^0$ *to* $\mathsf{SS}(\boldsymbol{\beta}) = \big\|\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\big\|^2$ *subject to* $\mathbb{L}\boldsymbol{\beta} = \boldsymbol{\theta}^0$ *and is given as*

$$\boldsymbol{b}^0 = \widehat{\boldsymbol{\beta}} - \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\{\mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1}\big(\mathbb{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}^0\big),$$

where $\widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}$ *is the (unconstrained) least squares estimator of the vector* $\boldsymbol{\beta}$.

(ii) *The fitted values* $\widehat{\boldsymbol{Y}}^0$ *can be expressed as*

$$\widehat{\boldsymbol{Y}}^0 = \widehat{\boldsymbol{Y}} - \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\{\mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1}\big(\mathbb{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}^0\big).$$

(iii) *The vector* $\boldsymbol{D} = \widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}^0$ *satisfies*

$$\big\|\boldsymbol{D}\big\|^2 = \mathsf{SS}_e^0 - \mathsf{SS}_e = \big(\mathbb{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}^0\big)^\top\Big\{\mathbb{L}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top\Big\}^{-1}\big(\mathbb{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}^0\big).$$

8.3.1 F-statistic to verify a set of linear constraints

$$F_0 = \frac{\dfrac{\mathsf{SS}_e^0 - \mathsf{SS}_e}{k - r_0}}{\dfrac{\mathsf{SS}_e}{n-k}} = \frac{\dfrac{(\mathbb{L}\widehat{\boldsymbol\beta} - \boldsymbol\theta^0)^\top \left\{\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top\right\}^{-1} (\mathbb{L}\widehat{\boldsymbol\beta} - \boldsymbol\theta^0)}{m}}{\dfrac{\mathsf{SS}_e}{n-k}}$$

$$= \frac{1}{m} (\mathbb{L}\widehat{\boldsymbol\beta} - \boldsymbol\theta^0)^\top \left\{\mathsf{MS}_e\, \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top\right\}^{-1} (\mathbb{L}\widehat{\boldsymbol\beta} - \boldsymbol\theta^0)$$

$$= \frac{1}{m} (\widehat{\boldsymbol\theta} - \boldsymbol\theta^0)^\top \left\{\mathsf{MS}_e\, \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top\right\}^{-1} (\widehat{\boldsymbol\theta} - \boldsymbol\theta^0),$$

8.3.2 t-statistic to verify a linear constraint

$$F_0 = \frac{1}{m} \left(\widehat{\theta} - \theta^0\right) \left\{\mathsf{MS}_e\, \mathbf{l}^\top \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbf{l}\right\}^{-1} \left(\widehat{\theta} - \theta^0\right)$$

$$= \left(\frac{\widehat{\theta} - \theta^0}{\sqrt{\mathsf{MS}_e\, \mathbf{l}^\top \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbf{l}}}\right)^2 = T_0^2,$$

where

$$T_0 = \frac{\widehat{\theta} - \theta^0}{\sqrt{\mathsf{MS}_e\, \mathbf{l}^\top \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbf{l}}}$$

Section **8.4**

Overall F-test

8.4 Overall F-test

Lemma 8.5 Overall F-test.

Assume a normal linear model $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\beta, \sigma^2 \mathbf{I}_n)$, $\operatorname{rank}(\mathbb{X}_{n \times k}) = r > 1$ *where* $\mathbf{1}_n \in \mathcal{M}(\mathbb{X})$*. Let* R^2 *be its coefficient of determination. The submodel F-statistic to compare model* $\mathsf{M}: \boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\beta, \sigma^2 \mathbf{I}_n)$ *and the only intercept model* $\mathsf{M}_0: \boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbf{1}_n\gamma, \sigma^2 \mathbf{I}_n)$ *takes the form*

$$F_0 = \frac{R^2}{1 - R^2} \cdot \frac{n - r}{r - 1}.$$

9

Checking Model Assumptions

9 Checking Model Assumptions

Data

$$\left(Y_i,\, \boldsymbol{Z}_i^\top\right)^\top,\ \boldsymbol{Z}_i = \left(Z_{i,1},\, \dots,\, Z_{i,p}\right)^\top \in \mathcal{Z} \subseteq \mathbb{R}^p,\ i = 1, \dots, n$$

First set of regressors

$$\boldsymbol{X}_i = \boldsymbol{t}_X(\boldsymbol{Z}_i),\ i = 1, \dots, n, \qquad \text{for some transformation } \boldsymbol{t}_X : \mathbb{R}^p \longrightarrow \mathbb{R}^k$$

$$\Rightarrow \qquad \mathbb{X}_{n\times k} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix} = \left(\boldsymbol{X}^0,\, \dots,\, \boldsymbol{X}^{k-1}\right)$$

Second set of regressors

$$\boldsymbol{V}_i = \boldsymbol{t}_V(\boldsymbol{Z}_i),\ i = 1, \dots, n, \qquad \text{for some transformation } \boldsymbol{t}_V : \mathbb{R}^p \longrightarrow \mathbb{R}^l$$

$$\Rightarrow \qquad \mathbb{V}_{n\times l} = \begin{pmatrix} \boldsymbol{V}_1^\top \\ \vdots \\ \boldsymbol{V}_n^\top \end{pmatrix} = \left(\boldsymbol{V}^1,\, \dots,\, \boldsymbol{V}^l\right)$$

9 Checking Model Assumptions

Assumptions behind $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$

With $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta} = \big(Y_1 - \boldsymbol{X}_1^\top \boldsymbol{\beta}, \dots, Y_n - \boldsymbol{X}_n^\top \boldsymbol{\beta}\big)^\top = \big(\varepsilon_1, \dots, \varepsilon_n\big)^\top$,

$\boldsymbol{X}_i = t(\boldsymbol{Z}_i)$

1. Correct regression function

   $(\mathbb{E}\big(Y_i \,\big|\, \boldsymbol{Z}_i\big) = \boldsymbol{X}_i^\top \boldsymbol{\beta}$ for some $\boldsymbol{\beta}$, $\quad \mathbb{E}\big(\varepsilon_i \,\big|\, \boldsymbol{Z}_i\big) = 0)$.

2. (Conditional) homoscedasticity of errors

   $(\mathrm{var}\big(Y_i \,\big|\, \boldsymbol{Z}_i\big) = \mathrm{var}\big(\varepsilon_i \,\big|\, \boldsymbol{Z}_i\big) = \sigma^2 = \mathrm{const})$.

3. (Conditionally) uncorrelated/independent errors $\varepsilon_1, \dots, \varepsilon_n$.

4. (Conditionally) normal errors

   $(Y_i \,\big|\, \boldsymbol{Z}_i \sim \mathcal{N}, \quad \varepsilon_i \,\big|\, \boldsymbol{Z}_i \sim \mathcal{N})$.

Section **9.1**

Model with added regressors

9.1 Model with added regressors

Quantities derived under model M: $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$

$$\boldsymbol{b} = \left(\mathbb{X}^\top \mathbb{X}\right)^{-} \mathbb{X}^\top \boldsymbol{Y},$$

$$\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top \boldsymbol{Y} \qquad \text{(if } \mathbb{X} \text{ is of full-rank),}$$

$$\mathbb{H} = \mathbb{X}\left(\mathbb{X}^\top \mathbb{X}\right)^{-} \mathbb{X}^\top = \left(h_{i,t}\right)_{i,t=1,\dots,n},$$

$$\widehat{\boldsymbol{Y}} = \mathbb{H}\boldsymbol{Y} = \left(\widehat{Y}_1, \dots, \widehat{Y}_n\right)^\top,$$

$$\mathbb{M} = \mathbf{I}_n - \mathbb{H} = \left(m_{i,t}\right)_{i,t=1,\dots,n},$$

$$\boldsymbol{U} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \mathbb{M}\boldsymbol{Y} = \left(U_1, \dots, U_n\right)^\top,$$

$$\mathrm{SS}_e = \left\|\boldsymbol{U}\right\|^2.$$

9.1 Model with added regressors

Quantities derived under model M_g: $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\ \sigma^2 \mathbf{I}_n\big)$, $\mathbb{G} = \big(\mathbb{X},\ \mathbb{V}\big)$

$$\big(\boldsymbol{b}_g^\top,\ \boldsymbol{c}_g^\top\big)^\top = \big(\mathbb{G}^\top\mathbb{G}\big)^-\,\mathbb{G}^\top\boldsymbol{Y},$$

$$\big(\widehat{\boldsymbol{\beta}}_g^\top,\ \widehat{\boldsymbol{\gamma}}_g^\top\big)^\top = \big(\mathbb{G}^\top\mathbb{G}\big)^{-1}\,\mathbb{G}^\top\boldsymbol{Y} \qquad \text{(if } \mathbb{G} \text{ is of full-rank)},$$

$$\mathbb{H}_g = \mathbb{G}\big(\mathbb{G}^\top\mathbb{G}\big)^-\mathbb{G}^\top = \big(h_{g,i,t}\big)_{i,t=1,\dots,n},$$

$$\widehat{\boldsymbol{Y}}_g = \mathbb{H}_g\boldsymbol{Y} = \big(\widehat{Y}_{g,1},\dots,\widehat{Y}_{g,n}\big)^\top,$$

$$\mathbb{M}_g = \mathbf{I}_n - \mathbb{H}_g = \big(m_{g,i,t}\big)_{i,t=1,\dots,n},$$

$$\boldsymbol{U}_g = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}_g = \mathbb{M}_g\boldsymbol{Y} = \big(U_{g,1},\dots,U_{g,n}\big)^\top,$$

$$\mathrm{SS}_{g,e} = \big\|\boldsymbol{U}_g\big\|^2.$$

9.1 Model with added regressors

Lemma 9.1 Model with added regressors.

Quantities derived while assuming model $\mathsf{M}: \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$ *and quantities derived while assuming model* $\mathsf{M}_g: \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\, \sigma^2 \mathbf{I}_n\big)$ *are mutually in the following relationship.*

$$\begin{aligned} \widehat{\boldsymbol{Y}}_g &= \widehat{\boldsymbol{Y}} + \mathbb{M}\mathbb{V}\big(\mathbb{V}^\top \mathbb{M}\mathbb{V}\big)^{-}\mathbb{V}^\top \boldsymbol{U} \\ &= \mathbb{X}\boldsymbol{b}_g + \mathbb{V}\boldsymbol{c}_g, \qquad \text{for some } \boldsymbol{b}_g \in \mathbb{R}^k,\ \boldsymbol{c}_g \in \mathbb{R}^l. \end{aligned}$$

Vectors $\boldsymbol{b}_g$ *and* $\boldsymbol{c}_g$ *such that* $\widehat{\boldsymbol{Y}}_g = \mathbb{X}\boldsymbol{b}_g + \mathbb{V}\boldsymbol{c}_g$ *satisfy:*

$$\begin{aligned} \boldsymbol{c}_g &= \big(\mathbb{V}^\top \mathbb{M}\mathbb{V}\big)^{-}\mathbb{V}^\top \boldsymbol{U}, \\ \boldsymbol{b}_g &= \boldsymbol{b} - \big(\mathbb{X}^\top\mathbb{X}\big)^{-}\mathbb{X}^\top\mathbb{V}\boldsymbol{c}_g \qquad \text{for some } \boldsymbol{b} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-}\mathbb{X}^\top\boldsymbol{Y}. \end{aligned}$$

Finally

$$\mathrm{SS}_e - \mathrm{SS}_{e,g} = \big\|\mathbb{M}\mathbb{V}\boldsymbol{c}_g\big\|^2.$$

Section **9.2**

Correct regression function

9.2 Correct regression function

Assumed model

$$\mathsf{M}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\ \sigma^2 \mathbf{I}_n\big),$$

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}: \quad \mathbb{E}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \mathbf{0}_n,$$

$$\operatorname{var}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \sigma^2 \mathbf{I}_n.$$

Assumption (A1) on a correct regression function

$$\mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) \in \mathcal{M}\big(\mathbb{X}\big), \qquad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) = \mathbb{X}\boldsymbol{\beta} \quad \text{for some } \boldsymbol{\beta} \in \mathbb{R}^k,$$

$$\mathbb{E}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \mathbf{0}_n \qquad \Big(\Longrightarrow \quad \mathbb{E}\big(\boldsymbol{\varepsilon}\big) = \mathbf{0}_n\Big).$$

$$\text{(A1)} \qquad \Longrightarrow \qquad \mathbb{E}\big(\boldsymbol{U} \,\big|\, \mathbb{Z}\big) = \mathbf{0}_n$$

9.2.1 Partial residuals

Model with a removed jth regressor, $j \in \{1, \dots, k-1\}$

$$\mathbb{X}^{(-j)} = \text{matrix } \mathbb{X} \text{ without the column } \boldsymbol{X}^j,$$

$$\boldsymbol{\beta}^{(-j)} = \left(\beta_0, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_{k-1}\right)^\top,$$

$$\mathrm{M}^{(-j)}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}^{(-j)}\boldsymbol{\beta}^{(-j)}, \, \sigma^2 \mathbf{I}_n\right),$$

$$\mathbb{M}^{(-j)} := \mathbf{I}_n - \mathbb{X}^{(-j)} \left(\mathbb{X}^{(-j)\top}\mathbb{X}^{(-j)}\right)^{-1} \mathbb{X}^{(-j)\top},$$

$$\boldsymbol{U}^{(-j)} := \mathbb{M}^{(-j)} \boldsymbol{Y}.$$

Assumption: $\mathrm{rank}\left(\mathbb{X}_{n \times k}\right) = k$, $\boldsymbol{X}^0 = \mathbf{1}_n$

$\Rightarrow \quad \mathrm{rank}\left(\mathbb{X}^{(-j)}\right) = k - 1 \quad \Rightarrow$

(i) $\boldsymbol{X}^j \notin \mathcal{M}\left(\mathbb{X}^{(-j)}\right)$;

(ii) $\boldsymbol{X}^j \neq \mathbf{0}_n$;

(iii) $\boldsymbol{X}^j$ is not a multiple of a vector $\mathbf{1}_n$.

9.2.1 Partial residuals

Definition 9.1 Partial residuals.

A vector of jth partial residuals of model M is a vector

$$\boldsymbol{U}^{part,j} = \boldsymbol{U} + \widehat{\beta}_j \, \boldsymbol{X}^j = \begin{pmatrix} U_1 + \widehat{\beta}_j X_{1,j} \\ \vdots \\ U_n + \widehat{\beta}_j X_{n,j} \end{pmatrix}.$$

Note. We have

$$\begin{aligned} \boldsymbol{U}^{part,j} &= \boldsymbol{U} + \widehat{\beta}_j \boldsymbol{X}^j \\ &= \boldsymbol{Y} - \big(\mathbb{X}\widehat{\boldsymbol{\beta}} - \widehat{\beta}_j \boldsymbol{X}^j\big) \\ &= \boldsymbol{Y} - \big(\widehat{\boldsymbol{Y}} - \widehat{\beta}_j \boldsymbol{X}^j\big). \end{aligned}$$

Lemma 9.2 Property of partial residuals.

Let $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbf{I}_n\big)$, $\text{rank}\big(\mathbb{X}_{n\times k}\big) = k$, $\boldsymbol{X}^0 = \mathbf{1}_n$, $\boldsymbol{\beta} = \big(\beta_0,\, \dots,\, \beta_{k-1}\big)^\top$. *Let* $\widehat{\beta}_j$ *be the LSE of* β_j, $j \in \big\{1,\, \dots,\, k-1\big\}$. *Let us consider a linear model (regression line with covariates* $\boldsymbol{X}^j$*) with*

- *the* j*th partial residuals* $\boldsymbol{U}^{part,j}$ *as response;*
- *a matrix* $\big(\mathbf{1}_n,\, \boldsymbol{X}^j\big)$ *as the model matrix;*
- *regression coefficients* $\boldsymbol{\gamma}_j = \big(\gamma_{j,0},\, \gamma_{j,1}\big)^\top$.

The least squares estimators of parameters $\gamma_{j,0}$ *and* $\gamma_{j,1}$ *are*

$$\widehat{\gamma}_{j,0} = 0, \quad \widehat{\gamma}_{j,1} = \widehat{\beta}_j.$$

9.2.1 Partial residuals

Notation. Response, regressor and partial residuals means

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad \overline{X}^j = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}, \qquad \overline{U}^{part,j} = \frac{1}{n}\sum_{i=1}^{n} U_i^{part,j}.$$

If $\boldsymbol{X}^0 = \mathbf{1}_n$ (model with intercept), we have

$$0 = \sum_{i=1}^{n} U_i = \sum_{i=1}^{n}\Big(U_i^{part,j} - \widehat{\beta}_j X_{i,j}\Big), \qquad \frac{1}{n}\sum_{i=1}^{n} U_i^{part,j} = \widehat{\beta}_j\Big(\frac{1}{n}\sum_{i=1}^{n} X_{i,j}\Big),$$

$$\overline{U}^{part,j} = \widehat{\beta}_j\,\overline{X}^j.$$

Definition 9.2 Shifted partial residuals.

A vector of jth *response-mean* partial residuals of model M is a vector

$$\boldsymbol{U}^{part,j,Y} = \boldsymbol{U}^{part,j} + \Big(\overline{Y} - \widehat{\beta}_j\overline{X}^j\Big)\,\mathbf{1}_n.$$

A vector of jth *zero-mean* partial residuals of model M is a vector

$$\boldsymbol{U}^{part,j,0} = \boldsymbol{U}^{part,j} - \widehat{\beta}_j\overline{X}^j\,\mathbf{1}_n.$$

Interpretation of partial residuals

$\boldsymbol{U}^{part,j} \equiv$ a response vector from which we removed a possible effect of all remaining regressors

Dependence of $\boldsymbol{U}^{part,j}$ on $\boldsymbol{X}^j$ shows

- a *net* effect of the *j*th regressor on the response $\boldsymbol{Y}$;
- a *partial* effect of the *j*th regressor on the response $\boldsymbol{Y}$ which is *adjusted* for the effect of the remaining regressors.

Use of partial residuals

Diagnostic tool $\rightarrow$ on a scatterplot $\big(\boldsymbol{X}^j,\ \boldsymbol{U}^{part,j}\big)$, the points should lie along a line (Lemma 9.2)

Visualization $\rightarrow$ on a scatterplot $\big(\boldsymbol{X}^j,\ \boldsymbol{U}^{part,j}\big)$, the slope of the fitted line is equal to $\widehat{\beta}_j$ (Lemma 9.2)

Cars2004nh (subset, $n = 409$)

`consumption` $\sim$ `log(weight)` $+$ `engine size` $+$ `horsepower`

```
m <- lm(consumption ~ lweight + engine.size + horsepower, data = CarsNow)
summary(m)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.1174 -0.6923 -0.1127  0.5473  5.2275

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.353265   2.948614 -14.364  < 2e-16 ***
lweight       6.935604   0.428971  16.168  < 2e-16 ***
engine.size   0.352687   0.096730   3.646 0.000301 ***
horsepower    0.003983   0.001085   3.672 0.000273 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9706 on 405 degrees of freedom
Multiple R-squared:  0.7946,	Adjusted R-squared:  0.793
F-statistic: 522.1 on 3 and 405 DF,  p-value: < 2.2e-16
```

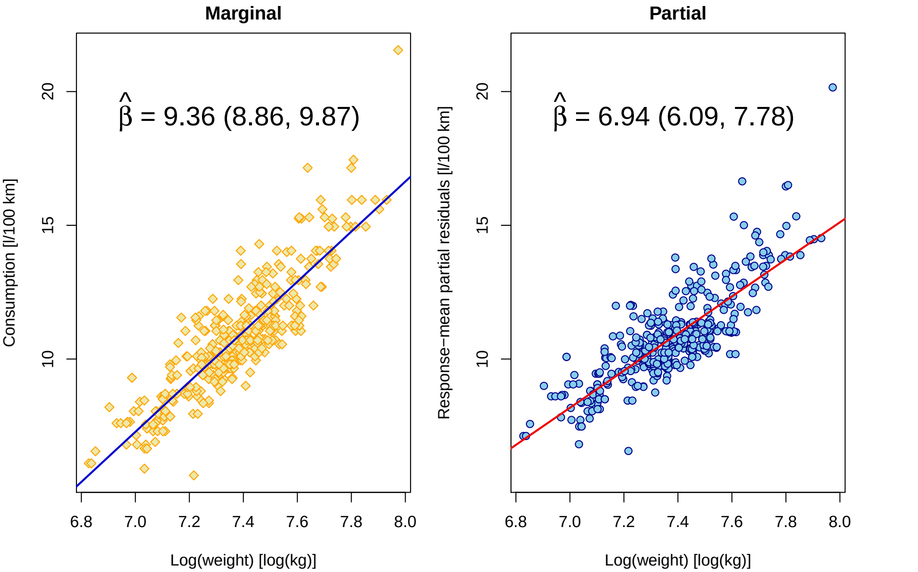
Consumption: $\overline{Y} = 10.75$, Log(weight): $\overline{X}^1 = 7.37$,

Engine size: $\overline{X}^2 = 3.18$,

Horsepower: $\overline{X}^3 = 215.8$.

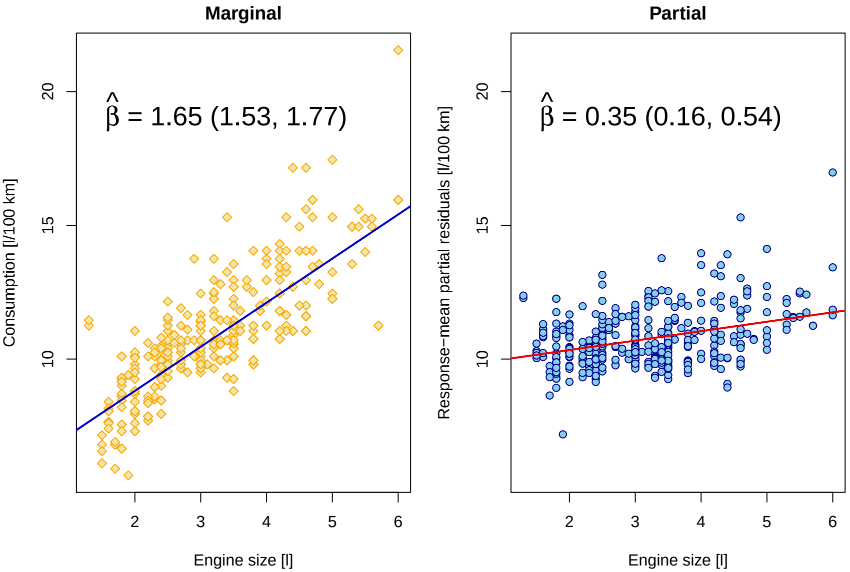
Cars2004nh (subset, $n = 409$)

consumption ~ log(weight) + engine size + horsepower



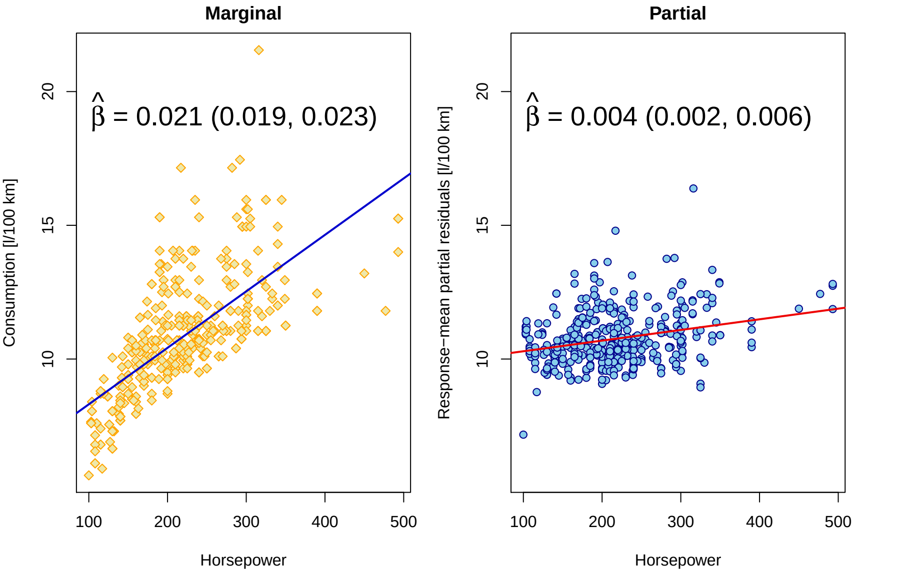
Cars2004nh (subset, $n = 409$)

consumption $\sim$ log(weight) + engine size + horsepower



Cars2004nh (subset, $n = 409$)

consumption ∼ log(weight) + engine size + horsepower



Policie ($n = 50$)

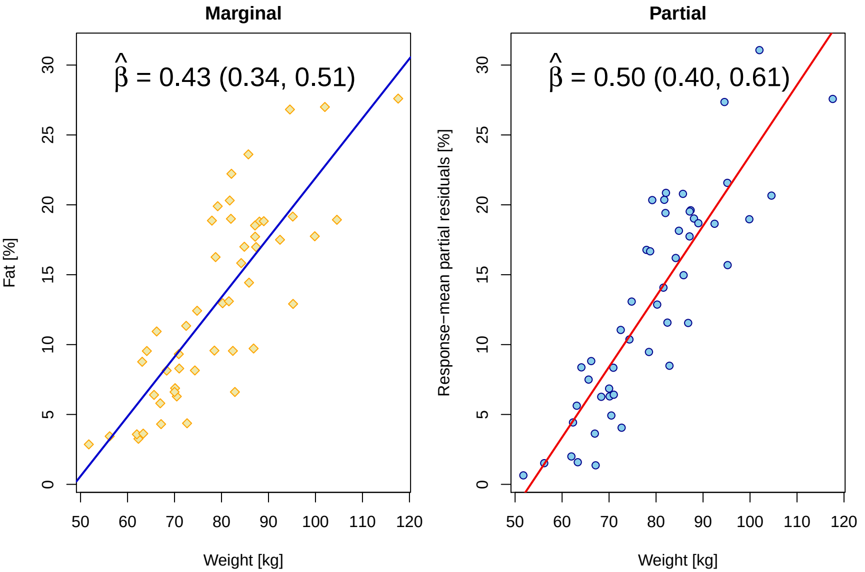
fat $\sim$ weight + height

```
summary(mHeWe <- lm(fat ~ weight + height, data = Policie))
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-6.4011 -2.9482 -0.0211  2.3072  7.2968

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.55309   15.24621   1.086   0.2831
weight       0.50418    0.05095   9.896 4.49e-13 ***
height      -0.24362    0.09728  -2.504   0.0158 *
---

Residual standard error: 3.731 on 47 degrees of freedom
Multiple R-squared:  0.714,	Adjusted R-squared:  0.7018
F-statistic: 58.66 on 2 and 47 DF,  p-value: 1.681e-13
```

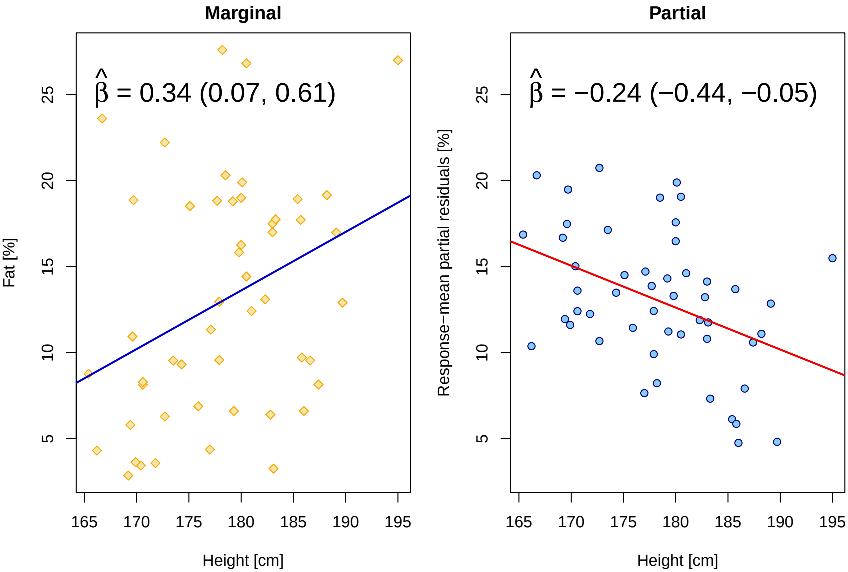
Policie ($n = 50$)

fat $\sim$ weight + height



Policie ($n = 50$)

fat ~ weight + height



9.2.2 Test for linearity of the effect

Without loss of generality:

$$\mathbb{X} = \left(\mathbf{1}_n,\ \mathbb{X}^0,\ \boldsymbol{X}^j\right).$$

9.2.2 Test for linearity of the effect

More general parameterization of the jth regressor

$\boldsymbol{X}^j \in \mathcal{M}(\mathbb{V}), \qquad \text{rank}(\mathbb{V}) \geq 2$

Submodel M: $\big(\mathbf{1}_n,\, \mathbb{X}^0,\, \boldsymbol{X}^j\big) = \mathbb{X}$;

(Larger) model M_g: $\big(\mathbf{1}_n,\, \mathbb{X}^0,\, \mathbb{V}\big)$.

Possibilities for a choice of $\mathbb{V}$:

- polynomial of degree $d \geq 2$ based on the regressor $\boldsymbol{X}^j$;
- regression spline of degree $d \geq 1$ based on the regressor $\boldsymbol{X}^j$.

consumption ~ log(weight) + engine.size + horsepower

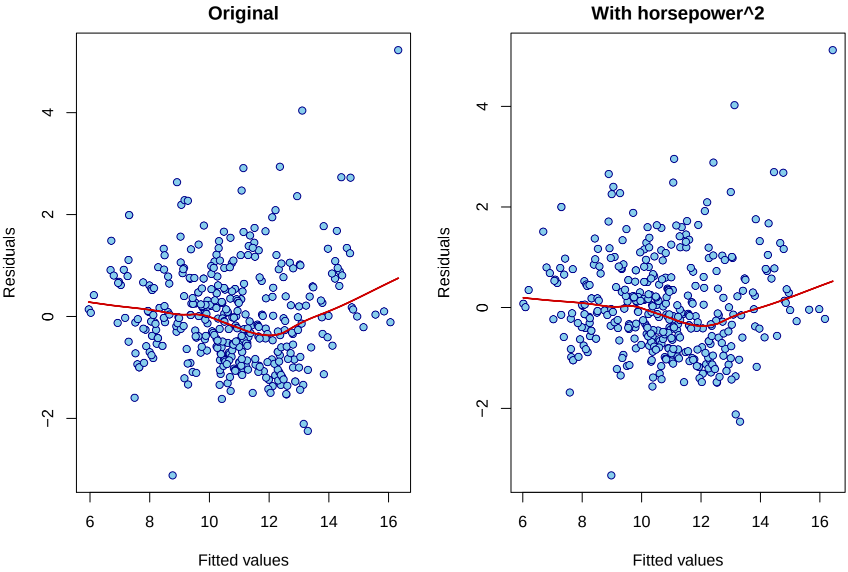
Quadratic term added for horsepower

```
mh2 <- lm(consumption ~ lweight + engine.size + horsepower + I(horsepower^2),
          data = CarsNow)
summary(mh2)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.3298 -0.6501 -0.1307  0.5178  5.1163

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -4.386e+01  3.065e+00 -14.308  < 2e-16 ***
lweight          7.249e+00  4.641e-01  15.621  < 2e-16 ***
engine.size      3.482e-01  9.652e-02   3.607 0.000348 ***
horsepower      -2.578e-03  3.914e-03  -0.659 0.510515
I(horsepower^2)  1.221e-05  7.001e-06   1.744 0.081873 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9682 on 404 degrees of freedom
Multiple R-squared:  0.7961,        Adjusted R-squared:  0.7941
F-statistic: 394.3 on 4 and 404 DF,  p-value: < 2.2e-16
```

Cars2004nh (subset, $n = 409$)

consumption $\sim$ log(weight) + engine size + horsepower

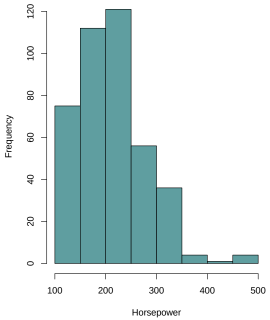


consumption ~ log(weight) + engine.size + horsepower

Cubic spline parameterization of horsepower (knots: 100, 200, 300, 500)

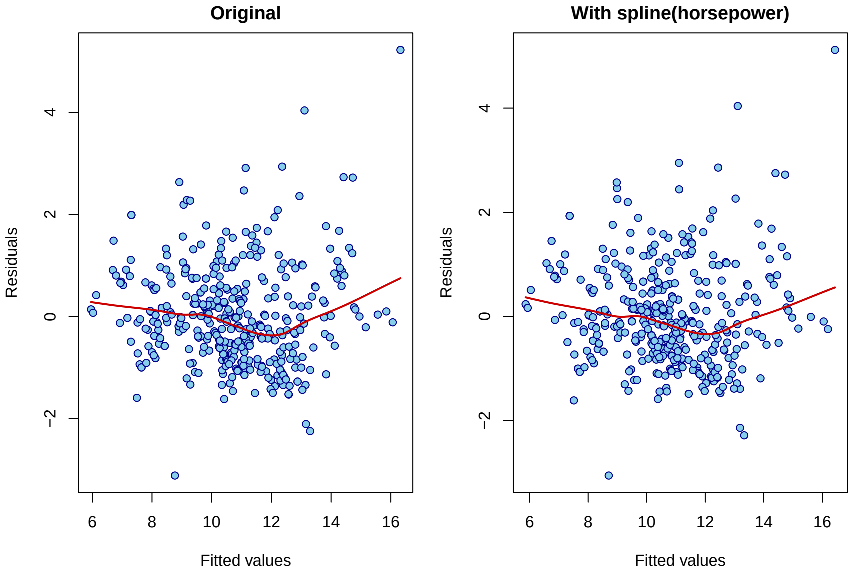
```
library("splines")
knots <- c(100, 200, 300, 500)
inner <- knots[-c(1, length(knots))]
bound <- knots[c(1, length(knots))]
hB <- bs(CarsNow[, "horsepower"], knots = inner, Boundary.knots = bound, degree = 3,
         intercept = TRUE)
mhB <- lm(consumption ~ -1 + lweight + engine.size + hB, data = CarsNow)
summary(mhB)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.0533 -0.6471 -0.1273  0.5095  5.1164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
lweight       7.19154    0.48080  14.958  < 2e-16 ***
engine.size   0.36108    0.09911   3.643 0.000304 ***
hB1         -43.88205    3.25963 -13.462  < 2e-16 ***
hB2         -43.40426    3.32369 -13.059  < 2e-16 ***
hB3         -43.58750    3.39894 -12.824  < 2e-16 ***
hB4         -43.18531    3.38594 -12.754  < 2e-16 ***
hB5         -41.93832    3.43966 -12.193  < 2e-16 ***
hB6         -41.83870    3.37295 -12.404  < 2e-16 ***
...
```



Cars2004nh (subset, $n = 409$)

consumption $\sim$ log(weight) + engine size + horsepower

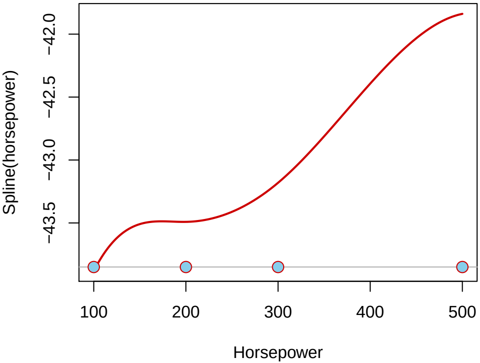


Cars2004nh (subset, $n = 409$)

consumption ∼ log(weight) + engine.size + horsepower

Cubic spline parameterization of horsepower (knots: 100, 200, 300, 500)

```
m <- lm(consumption ~ lweight +
           engine.size +
           horsepower,
           data = CarsNow)
anova(m, mhB)
```




```
Analysis of Variance Table

Model 1: consumption ~ lweight +
           engine.size + horsepower
Model 2: consumption ~ -1 + lweight +
           engine.size + hB

  Res.Df    RSS Df Sum of Sq     F Pr(>F)
1    405 381.56
2    401 377.08  4    4.4797 1.191 0.3142
```

Categorization of the jth regressor

Categorization of the jth regressor

Bounds: $x_j^{low} < \min_i X_{i,j}, \qquad \max_i X_{i,j} < x_j^{upp},$

Division: $\lambda_0 = x_j^{low} < \lambda_1 < \cdots < \lambda_{H-1} < x_j^{upp} = \lambda_H,$

Intervals and their representatives:

$$\mathcal{I}_h = \big(\lambda_{h-1},\, \lambda_h\big], \quad x_h \in \mathcal{I}_h, \quad h = 1, \ldots, H,$$

Categorized covariate: $X_i^{j,cut} = x_h \quad \equiv \quad X_i^j \in \mathcal{I}_h, \qquad h = 1, \ldots, H.$

$\mathbb{V}$ based on (pseudo)contrasts for $\boldsymbol{X}^{j,cut}$ if that is viewed as categorical

Submodel M: $\big(\mathbf{1}_n,\, \mathbb{X}^0,\, \boldsymbol{X}^{j,cut}\big);$

(Larger) model M_g: $\big(\mathbf{1}_n,\, \mathbb{X}^0,\, \mathbb{V}\big).$

`Cars2004nh` (subset, $n = 409$)

`consumption ~ log(weight) + engine.size + horsepower`

Categorized `horsepower` (100–150, 150–200, 250–300, 300–500)

```
BREAKS <- c(0, 150, 200, 250, 300, 500)
CarsNow <- transform(CarsNow, horseord = cut(horsepower, breaks = BREAKS))
levels(CarsNow[, "horseord"])[1] <- "[100, 150]"
table(CarsNow[, "horseord"])
```

```
[100, 150]  (150,200]  (200,250]  (250,300]  (300,500]
        75        112        121         56         45
```

`horsepower` categories represented by **midpoints**

```
MIDS <- c(125, 175, 225, 275, 400)
CarsNow <- transform(CarsNow, horsemid = as.numeric(horseord))
CarsNow[, "horsemid"] <- MIDS[CarsNow[, "horsemid"]]
table(CarsNow[, "horsemid"])
```

```
125 175 225 275 400
 75 112 121  56  45
```

Cars2004nh (subset, $n = 409$)

consumption ∼ log(weight) + engine.size + horsepower

Larger model (horsepower as categorical, reference group pseudocontrasts)

```
mhord <- lm(consumption ~ lweight + engine.size + horseord, data = CarsNow)
summary(mhord)
```

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -43.4282     3.1974 -13.582  < 2e-16 ***
lweight             7.1578     0.4676  15.307  < 2e-16 ***
engine.size         0.3312     0.0981   3.376 0.000806 ***
horseord(150,200]   0.3928     0.1637   2.400 0.016852 *
horseord(200,250]   0.2206     0.1832   1.204 0.229119
horseord(250,300]   0.5249     0.2338   2.245 0.025332 *
horseord(300,500]   1.0871     0.2626   4.140 4.23e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9628 on 402 degrees of freedom
Multiple R-squared:  0.7994,	Adjusted R-squared:  0.7964
F-statistic:   267 on 6 and 402 DF,  p-value: < 2.2e-16
```

```
consumption ~ log(weight) + engine.size + horsepower
```

Submodel (`horsepower` intervals represented by midpoints)

```
mhmid <- lm(consumption ~ lweight + engine.size + horsemid, data = CarsNow)
summary(mhmid)
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -43.121394   2.944142 -14.647  < 2e-16 ***
lweight        7.057884   0.427803  16.498  < 2e-16 ***
engine.size    0.338626   0.096994   3.491 0.000534 ***
horsemid       0.003519   0.009049   3.889 0.000118 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9687 on 405 degrees of freedom
Multiple R-squared:  0.7954,	Adjusted R-squared:  0.7938
F-statistic: 524.7 on 3 and 405 DF,  p-value: < 2.2e-16
```

F-test on a submodel

```
anova(mhmid, mhord)
```

```
Model 1: consumption ~ lweight + engine.size + horsemid
Model 2: consumption ~ lweight + engine.size + horseord
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    405 380.07
2    402 372.61  3    7.4566 2.6816 0.04653 *
```

Cars2004nh (subset, $n = 409$)

consumption ∼ log(weight) + engine.size + horsepower

Approximate submodel (original horsepower values)

```
m <- lm(consumption ~ lweight + engine.size + horsepower, data = CarsNow)
summary(m)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.353265   2.948614 -14.364  < 2e-16 ***
lweight       6.935604   0.428971  16.168  < 2e-16 ***
engine.size   0.352687   0.096730   3.646 0.000301 ***
horsepower    0.003983   0.001085   3.672 0.000273 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9706 on 405 degrees of freedom
Multiple R-squared:  0.7946,	Adjusted R-squared:  0.793
F-statistic: 522.1 on 3 and 405 DF,  p-value: < 2.2e-16
```

Approximate F-test on a submodel

```
anova(m, mhord)
```

```
Model 1: consumption ~ lweight + engine.size + horsepower
Model 2: consumption ~ lweight + engine.size + horseord
  Res.Df    RSS Df Sum of Sq     F  Pr(>F)
1    405 381.56
2    402 372.61  3    8.9427 3.216 0.02285 *
```

Drawback of tests for linearity of the effect

- Linearity of the effect of the jth regressor $\equiv$ null hypothesis
- Linearity of the effect can be rejected but never confirmed

Section **9.3**

Homoscedasticity

9.3 Homoscedasticity

Assumed model

$$\mathsf{M}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right),$$

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}: \quad \mathbb{E}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \mathbb{E}\big(\boldsymbol{\varepsilon}\big) = \mathbf{0}_n,$$
$$\mathsf{var}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \mathsf{var}\big(\boldsymbol{\varepsilon}\big) = \sigma^2 \mathbf{I}_n.$$

Assumption (A2) of homoscedasticity

$$\mathsf{var}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big) = \sigma^2\, \mathbf{I}_n, \qquad \mathsf{var}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{Z}\big) = \sigma^2\, \mathbf{I}_n, \qquad \left(\Longrightarrow \;\; \mathsf{var}\big(\boldsymbol{\varepsilon}\big) = \sigma^2 \mathbf{I}_n\right),$$

for some $\sigma^2 > 0$.

(A1) & (A2) $\quad \Longrightarrow \quad \mathsf{var}\big(\boldsymbol{U} \,\big|\, \mathbb{Z}\big) = \sigma^2\, \mathbb{M}, \qquad \mathbb{M} = \mathbf{I}_n - \mathbb{X}\big(\mathbb{X}^\top \mathbb{X}\big)^{-} \mathbb{X}^\top$

9.3.1 Tests of homoscedasticity

Considered hypotheses

H_0: $\operatorname{var}(\varepsilon_i \mid \boldsymbol{Z}_i) = \text{const}$,

H_1: $\operatorname{var}(\varepsilon_i \mid \boldsymbol{Z}_i) =$ certain function of some factor(s).

9.3.2 Score tests of homoscedasticity

Model under the **NULL** hypothesis

Full-rank normal linear model:

$$\mathsf{M}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big), \quad \operatorname{rank}\big(\mathbb{X}_{n\times k}\big) = k,$$

Model under the **ALTERNATIVE** hypothesis

Generalization of a general normal linear model:

$$\mathsf{M}_{hetero}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbb{W}^{-1}\big),$$

$\mathbb{W} = \operatorname{diag}(w_1, \dots, w_n), \qquad w_i^{-1} = \tau(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, \boldsymbol{Z}_i),\ i = 1, \dots, n,$

τ: a known function ($\boldsymbol{\lambda} \in \mathbb{R}^q$, $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{z} \in \mathbb{R}^p$), such that

$$\tau(\mathbf{0},\, \boldsymbol{\beta},\, \boldsymbol{z}) = 1, \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^k,\ \boldsymbol{z} \in \mathbb{R}^p.$$

9.3.2 Score tests of homoscedasticity

Breusch-Pagan test

$\boldsymbol{x} = \boldsymbol{t}_X(\boldsymbol{z}) \equiv$ regressors of model M

$$\tau(\lambda,\, \boldsymbol{\beta},\, \boldsymbol{z}) = \tau(\lambda,\, \boldsymbol{\beta},\, \boldsymbol{x}) = \exp\bigl(\lambda\, \boldsymbol{x}^\top \boldsymbol{\beta}\bigr)$$

$$\begin{aligned} \mathrm{H}_0: &\quad \lambda = 0, \\ \mathrm{H}_1: &\quad \lambda \neq 0. \end{aligned}$$

- One-sided tests with H_1: $\lambda > 0$ (or $\lambda < 0$) also possible
- Test not robust against violation of the normality assumption
- Koenker (1981): modified version of the test being robust towards non-normality

  $\Rightarrow$ (Koenker's) studentized Breusch-Pagan test

9.3.2 Score tests of homoscedasticity

Linear dependence on the regressors

$\boldsymbol{w} = \boldsymbol{t}_W(\boldsymbol{z})$: given transformation of the covariates

$$\tau(\lambda, \beta, \boldsymbol{z}) = \tau(\lambda, \boldsymbol{w}) = \exp\left(\boldsymbol{\lambda}^\top \boldsymbol{w}\right)$$

$$\begin{aligned} H_0: &\quad \boldsymbol{\lambda} = \boldsymbol{0}, \\ H_1: &\quad \boldsymbol{\lambda} \neq \boldsymbol{0}. \end{aligned}$$

Score tests of homoscedasticity in the R software

(i) `ncvTest` (abbreviation for a "non-constant variance test") from package `car`

(ii) `bptest` from package `lmtest` (allows also for the Koenker's studentized version)

9.3.3 Some other tests of homoscedasticity

Goldfeld-Quandt

G-sample tests of homoscedasticity

Applicable mainly in a context of ANOVA models.

- **Bartlett**
- **Levene**
- **Brown-Forsythe**
- **Fligner-Killeen**

Section **9.4**

Normality

9.4 Normality

Assumed model

$$\mathsf{M}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big),\ \operatorname{rank}(\mathbb{X}_{n\times k}) = r \leq k,$$

$$\Longrightarrow \quad \varepsilon_i = Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}\ \text{ satisfy } \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n.$$

Assumption (A4) of normality

$$\varepsilon_i \,\big|\, \mathbb{Z} \overset{\text{indep.}}{\sim} \mathcal{N}, \qquad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}$$

(A1) & (A2) & (A3)
& (A4)

$$\Longrightarrow \qquad \boldsymbol{U} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\, \mathbb{M}\big),$$

$$\Longrightarrow \qquad U_i^{std} \,\big|\, \mathbb{Z} \sim \quad \big(0,\, 1\big), \quad i = 1, \dots, n.$$

9.4 Normality

Reminder of notation

- Hat matrix: $\mathbb{H} = \mathbb{X}(\mathbb{X}^\top\mathbb{X})^-\mathbb{X}^\top = (h_{i,t})_{i,t=1,\ldots,n}$;
- Projection matrix into the residual space $\mathcal{M}(\mathbb{X})^\perp$: $\mathbb{M} = \mathbf{I}_n - \mathbb{H} = (m_{i,t})_{i,t=1,\ldots,n}$;
- Residuals: $\boldsymbol{U} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \mathbb{M}\boldsymbol{Y}(U_1, \ldots, U_n)^\top$;
- Residual sum of squares: $\mathrm{SS}_e = \|\boldsymbol{U}\|^2$;
- Residual mean square: $\mathrm{MS}_e = \frac{1}{n-r}\,\mathrm{SS}_e$;
- Standardized residuals: $\boldsymbol{U}^{std} = (U_1^{std}, \ldots, U_n^{std})^\top$, where

$$U_i^{std} = \frac{U_i}{\sqrt{\mathrm{MS}_e\, m_{i,i}}}, \qquad i = 1, \ldots, n \quad (\text{if } m_{i,i} > 0).$$

9.4.1 Tests of normality

Under normality of errors $\varepsilon_1, \ldots, \varepsilon_n$

$$\Longrightarrow \qquad \boldsymbol{U} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbf{0}_n,\, \sigma^2\, \mathbb{M}\big),$$

$$\Longrightarrow \qquad U_i^{std} \,\big|\, \mathbb{Z} \sim \quad (0,\, 1), \qquad i = 1, \ldots, n.$$

Approximate approaches to test

$$\mathrm{H}_0: \text{ distribution of } \varepsilon_1, \ldots, \varepsilon_n \text{ is normal.}$$

$\Rightarrow$ Apply any of classical tests of normality

(Shapiro-Wilk, Lilliefors, Anderson-Darling, . . .) to

(i) Raw residuals $U_1, \ldots, U_n$;

(ii) Standardized residuals $U_1^{std}, \ldots, U_n^{std}$.

Section **9.5**

Uncorrelated errors

9.5 Uncorrelated errors

Assumed model

$$\mathsf{M}: \quad \boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big),$$

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}: \quad \mathbb{E}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{X}\big) = \mathbb{E}\big(\boldsymbol{\varepsilon}\big) = \mathbf{0}_n,$$
$$\mathsf{var}\big(\boldsymbol{\varepsilon} \,\big|\, \mathbb{X}\big) = \mathsf{var}\big(\boldsymbol{\varepsilon}\big) = \sigma^2 \mathbf{I}_n.$$

Assumption (A3) of uncorrelated errors

$$\mathsf{cov}\big(\varepsilon_i,\, \varepsilon_l \,\big|\, \mathbb{X}\big) = 0, \;\; i \neq l \qquad \Big(\Longrightarrow \;\; \mathsf{cov}\big(\varepsilon_i,\, \varepsilon_l\big) = 0, \;\; i \neq l\Big).$$

9.5 Uncorrelated errors

Typical situations when uncorrelated errors cannot be taken for granted

(i) **Time series**: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ obtained at (equidistant) time points $t_1 < \ldots < t_n$

$\Longrightarrow$ serial dependence.

(ii) **Repeated measurements** on one subject/unit: $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^\top$, $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})^\top$, $i = 1, \ldots, n$,

i (identification of a subject) not used as a covariate.

In the following

Test for uncorrelated errors will be developed for situation when **ordering** of observations expressed by indeces $1, \ldots, n$ has a practical meaning and may induce dependence between $\varepsilon_1, \ldots, \varepsilon_n$.

9.5.1 Durbin-Watson test

Model under the **NULL** hypothesis

$$
\begin{aligned}
\mathsf{M}: \quad & Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, && i = 1, \ldots, n,\\
& \mathbb{E}\big(\varepsilon_i \,\big|\, \mathbb{X}\big) = 0, \quad \mathsf{var}\big(\varepsilon_i \,\big|\, \mathbb{X}\big) = \sigma^2, && i = 1, \ldots, n,\\
& \mathsf{cor}\big(\varepsilon_i,\, \varepsilon_l \,\big|\, \mathbb{X}\big) = 0, && i \neq l.
\end{aligned}
$$

Model under the **ALTERNATIVE** hypothesis

$$
\begin{aligned}
\mathsf{M}_{AR}: \quad & Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, && i = 1, \ldots, n,\\
& \varepsilon_1 = \eta_1, \quad \varepsilon_i = \varrho\, \varepsilon_{i-1} + \eta_i, && i = 2, \ldots, n,\\
& \mathbb{E}\big(\eta_i \,\big|\, \mathbb{X}\big) = 0, \quad \mathsf{var}\big(\eta_i \,\big|\, \mathbb{X}\big) = \sigma^2, && i = 1, \ldots, n,\\
& \mathsf{cor}\big(\eta_i,\, \eta_l \,\big|\, \mathbb{X}\big) = 0, && i \neq l,
\end{aligned}
$$

$-1 < \varrho < 1$: additional unknown parameter of the model.

9.5.1 Durbin-Watson test

Durbin-Watson test statistic

$\boldsymbol{U} = (U_1, \ldots, U_n)^\top$: residuals from model M.

$$DW = \frac{\sum_{i=2}^{n} (U_i - U_{i-1})^2}{\sum_{i=1}^{n} U_i^2}.$$

- Distribution of DW under H_0: $\varrho = 0$ depends on a model matrix $\mathbb{X}$
  ➠ not possible to derive (and tabulate) critical values in full generality.
- R function `dwtest[lmtest]`:
  p-values from approximations (Farebrother, 1980, 1984)
- R function `durbinWatsonTest[car]`:
  p-values from a general simulation based method bootstrap
- One-sided tests (with H_1: $\varrho > 0$) frequent in practice

Section **9.6**

Transformation of response

9.6 Transformation of response

Heteroscedasticity and/or non-normality for original response

➠often the following model is correct (perhaps wrong but useful):

Normal linear model for transformed response

$$\boldsymbol{Y}^\star \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\big),$$

$$\boldsymbol{Y}^\star = \big(t(Y_1), \ldots, t(Y_n)\big)^\top,$$

for suitable $t : \mathbb{R} \longrightarrow \mathbb{R}$, chosen (non-linear) transformation

WARNING, interpretation of the regression function

$$m(\boldsymbol{x}) = \mathbb{E}\big(t(Y) \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \quad \neq \quad t\Big(\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)\Big)$$

9.6.1 Prediction based on a model with transformed response

Aim: predict Y_{new}, given $\boldsymbol{X}_{new} = \boldsymbol{x}_{new}$, assume: t is strictly increasing.

1. $\widehat{Y}^{\star}_{new}$ and $\left(\widehat{Y}^{\star,L}_{new},\ \widehat{Y}^{\star,U}_{new}\right)$:

   point and interval (with a coverage of $1-\alpha$) prediction for

   $Y^{\star}_{new} = t(Y_{new})$

   based on the model $t(Y) = \boldsymbol{X}^{\top}\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \,\big|\, \boldsymbol{X} \sim \mathcal{N}(0,\, \sigma^2)$.

2. Interval

   $\left(\widehat{Y}^{L}_{new},\ \widehat{Y}^{U}_{new}\right) = \left(t^{-1}\left(\widehat{Y}^{\star,L}_{new}\right),\ t^{-1}\left(\widehat{Y}^{\star,U}_{new}\right)\right)$

   covers a value of Y_{new} with a probability of $1-\alpha$.

3. $\widehat{Y}_{new} = t^{-1}\left(\widehat{Y}^{\star}_{new}\right)$: point prediction.

9.6.2 Log-normal model

Log-normal linear model

$$\log(Y_i) = \boldsymbol{X}_i^\top \beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

$$\varepsilon_i \,\big|\, \mathbb{X} \overset{\text{indep.}}{\sim} \mathcal{N}(0, \sigma^2),$$

Multiplicative model for the original response

$$Y_i = \exp\bigl(\boldsymbol{X}_i^\top \beta\bigr)\,\eta_i, \qquad i = 1, \ldots, n,$$

$$\eta_i \,\big|\, \mathbb{X} \overset{\text{indep.}}{\sim} \mathcal{LN}(0, \sigma^2),$$

Moments of the log-normal distribution

$$M := \mathbb{E}(\eta_i) = \mathbb{E}(\eta_i \,|\, \mathbb{X}) = \exp\Bigl(\frac{\sigma^2}{2}\Bigr) \qquad > 1 \quad (\text{with } \sigma^2 > 0),$$

$$V := \operatorname{var}(\eta_i) = \operatorname{var}(\eta_i \,|\, \mathbb{X}) = \bigl\{\exp(\sigma^2) - 1\bigr\}\exp(\sigma^2).$$

9.6.2 Log-normal model

Conditional expectation and variance of the response (given $\boldsymbol{X} = \boldsymbol{x}$, $\boldsymbol{x} \in \mathcal{X}$)

$$\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) = M \exp\big(\boldsymbol{x}^\top \boldsymbol{\beta}\big),$$

$$\mathrm{var}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) = V \exp\big(2\,\boldsymbol{x}^\top \boldsymbol{\beta}\big) = V \cdot \left(\frac{\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}{M}\right)^2.$$

Features of the log-normal model

1. Response (conditional) distribution is skewed (log-normal).
2. Response (conditional) variance increases with the expectation.

9.6.2 Log-normal model

Interpretation of regression coefficients

$$\boldsymbol{x} = \big(x_0,\,\dots,\,x_j\,\dots,\,x_{k-1}\big)^\top \in \mathcal{X},$$

$$\boldsymbol{x}^{j(+1)} := \big(x_0,\,\dots,\,x_j+1\,\dots,\,x_{k-1}\big)^\top \in \mathcal{X},$$

$$\boldsymbol{\beta} = \big(\beta_0,\,\dots,\,\beta_{k-1}\big)^\top.$$

Ratio of the two expectations

$$\frac{\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}^{j(+1)}\big)}{\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)} = \frac{M \exp\big(\boldsymbol{x}^{j(+1)}{}^\top \boldsymbol{\beta}\big)}{M \exp\big(\boldsymbol{x}^\top \boldsymbol{\beta}\big)} = \exp(\beta_j).$$

9.6.2 Log-normal model

Interpretation of regression coefficients

Example. Log-normal model used with one-way classification

$$\mathbb{E}\bigl(\log(Y)\,\big|\,Z=g\bigr) = \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z,\ g = 1,\ \ldots,\ G$$

$\boldsymbol{c}_1^\top,\ \ldots,\ \boldsymbol{c}_G^\top$: rows of the (pseudo)contrast matrix

Ratio of the two group means

$$\frac{\mathbb{E}\bigl(Y\,\big|\,Z=g\bigr)}{\mathbb{E}\bigl(Y\,\big|\,Z=h\bigr)} = \frac{M\,\exp\bigl(\beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z\bigr)}{M\,\exp\bigl(\beta_0 + \boldsymbol{c}_h^\top \boldsymbol{\beta}^Z\bigr)} = \exp\Bigl\{\bigl(\boldsymbol{c}_g^\top - \boldsymbol{c}_h^\top\bigr)\boldsymbol{\beta}^Z\Bigr\}$$

$$= \exp\Bigl\{\mathbb{E}\bigl(\log(Y)\,\big|\,Z=g\bigr) - \mathbb{E}\bigl(\log(Y)\,\big|\,Z=h\bigr)\Bigr\}, \qquad g \neq h$$

10

Consequences of a Problematic Regression Space

10 Consequences of a Problematic Regression Space

Data

$\big(Y_i,\,\boldsymbol{Z}_i^\top\big)^\top$, $\boldsymbol{Z}_i = \big(Z_{i,1},\,\dots,\,Z_{i,p}\big)^\top \in \mathcal{Z} \subseteq \mathbb{R}^p$, $i = 1,\dots,n$

Response vector and the model matrix

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbb{X}_{n\times k} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix} = \big(\mathbf{1}_n,\,\boldsymbol{X}^1,\,\dots,\,\boldsymbol{X}^{k-1}\big),$$

$$\boldsymbol{X}_i = \boldsymbol{t}_X(\boldsymbol{Z}_i), \qquad i = 1,\,\dots,\,n$$

Full-rank linear model with intercept assumed

$$\boldsymbol{Y}\,\big|\,\mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\,\sigma^2\mathbf{I}_n\big), \qquad \text{rank}\big(\mathbb{X}\big) = k < n,$$

$\equiv$ Model matrix $\mathbb{X}$ sufficient to write $\mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{Z}\big) = \mathbb{E}\big(\boldsymbol{Y}\,\big|\,\mathbb{X}\big) = \mathbb{X}\boldsymbol{\beta}$

for some $\boldsymbol{\beta} = \big(\beta_0,\,\dots,\,\beta_{k-1}\big)^\top \in \mathbb{R}^k$

Section **10.1**

Multicollinearity

SVD of the model matrix $\mathbb{X}$

$$\mathbb{X} = \mathbb{U}\,\mathbb{D}\,\mathbb{V}^\top = \sum_{j=0}^{k-1} d_j\,\boldsymbol{u}_j\,\boldsymbol{v}_j^\top, \qquad \mathbb{D} = \mathrm{diag}(d_0,\,\ldots,\,d_{k-1})$$

- $\mathbb{U}_{n\times k} = \bigl(\boldsymbol{u}_0,\,\ldots,\,\boldsymbol{u}_{k-1}\bigr)$:
  the first k orthonormal eigenvectors of the $n \times n$ matrix $\mathbb{X}\mathbb{X}^\top$.
- $\mathbb{V}_{k\times k} = \bigl(\boldsymbol{v}_0,\,\ldots,\,\boldsymbol{v}_{k-1}\bigr)$:
  (all) orthonormal eigenvectors of the $k \times k$ (invertible) matrix $\mathbb{X}^\top\mathbb{X}$.
- $d_j = \sqrt{\lambda_j}, \quad j = 0, \ldots, k-1$, where $\lambda_0 \geq \cdots \geq \lambda_{k-1} > 0$ are
  - the first k eigenvalues of the matrix $\mathbb{X}\mathbb{X}^\top$;
  - (all) eigenvalues of the matrix $\mathbb{X}^\top\mathbb{X}$, i.e.,

$$\mathbb{X}^\top\mathbb{X} = \sum_{j=0}^{k-1} \lambda_j\,\boldsymbol{v}_j\,\boldsymbol{v}_j^\top = \mathbb{V}\,\boldsymbol{\Lambda}\,\mathbb{V}^\top, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_0,\,\ldots,\,\lambda_{k-1})$$

$$= \sum_{j=0}^{k-1} d_j^2\,\boldsymbol{v}_j\,\boldsymbol{v}_j^\top = \mathbb{V}\,\mathbb{D}^2\,\mathbb{V}^\top.$$

10.1.2 Multicollinearity and its impact on precision of the LSE

LSE in a linear model $\boldsymbol{Y} \,|\, \mathbb{Z} \sim (\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$, rank$(\mathbb{X}_{n\times k}) = k$

(i) $\widehat{\boldsymbol{Y}} = \big(\widehat{Y}_1, \ldots, \widehat{Y}_n\big)^\top = \mathbb{H}\boldsymbol{Y} \qquad \big(\mathbb{H} = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\big)$:

BLUE of $\quad \boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} = \mathbb{E}\big(\boldsymbol{Y} \,|\, \mathbb{Z}\big) \qquad$ with $\quad \mathrm{var}\big(\widehat{\boldsymbol{Y}} \,|\, \mathbb{Z}\big) = \sigma^2\,\mathbb{H}$;

(ii) $\widehat{\boldsymbol{\beta}} = \big(\widehat{\beta}_0, \ldots, \widehat{\beta}_n\big)^\top = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y}$:

BLUE of $\quad \boldsymbol{\beta} \qquad$ with $\quad \mathrm{var}\big(\widehat{\boldsymbol{\beta}} \,|\, \mathbb{Z}\big) = \sigma^2\,\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}$.

Multicollinearity

- No impact on precision of LSE of $\boldsymbol{\mu} = \mathbb{E}\big(\boldsymbol{Y} \,|\, \mathbb{Z}\big)$
- Possibly serious inflation of the standard errors of LSE of $\boldsymbol{\beta}$

Lemma 10.1 Bias in estimation of the squared norms.

Let $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\text{rank}(\mathbb{X}_{n\times k}) = k$. *The following then holds.*

$$\mathbb{E}\Big(\big\|\widehat{\boldsymbol{Y}}\big\|^2 - \big\|\mathbb{X}\boldsymbol{\beta}\big\|^2 \,\Big|\, \mathbb{Z}\Big) = \sigma^2\, k,$$

$$\mathbb{E}\Big(\big\|\widehat{\boldsymbol{\beta}}\big\|^2 - \big\|\boldsymbol{\beta}\big\|^2 \,\Big|\, \mathbb{Z}\Big) = \sigma^2\, \text{tr}\Big\{\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\Big\}.$$

Notation, linear model $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\ \sigma^2 \mathbf{I}_n\big)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$,
$\mathbb{X} = \big(\mathbf{1}_n,\ \boldsymbol{X}^1,\ \dots,\ \boldsymbol{X}^{k-1}\big)$, $\boldsymbol{X}^j = \big(X_{1,j},\ \dots,\ X_{n,j}\big)^\top$, $j = 1, \dots, k-1$

Response sample mean: $\overline{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i$;

Square root of the total sum of squares:

$$T_Y = \sqrt{\sum_{i=1}^{n}\big(Y_i - \overline{Y}\big)^2} = \big\|\boldsymbol{Y} - \overline{Y}\mathbf{1}_n\big\|;$$

Fitted values: $\widehat{\boldsymbol{Y}} = \big(\widehat{Y}_1,\ \dots,\ \widehat{Y}_n\big)^\top$;

Coefficient of determination:

$$R^2 = 1 - \frac{\big\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\big\|^2}{\big\|\boldsymbol{Y} - \overline{Y}\mathbf{1}_n\big\|^2} = 1 - \frac{\big\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\big\|^2}{T_Y^2}.$$

Residual mean square: $\mathrm{MS}_e = \dfrac{1}{n-k}\,\big\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\big\|^2.$

10.1.3 Variance inflation factor and tolerance

For $j = 1, \ldots, k-1$: Notation, linear model M_j, where response $= \boldsymbol{X}^j$, model matrix $= \mathbb{X}^{(-j)} = \left(\mathbf{1}_n,\, \boldsymbol{X}^1,\, \ldots,\, \boldsymbol{X}^{j-1},\, \boldsymbol{X}^{j+1},\, \ldots,\, \boldsymbol{X}^{k-1}\right)$

Column sample mean: $$\overline{X}^j = \frac{1}{n}\sum_{i=1}^{n} X_{i,j};$$

Square root of the total sum of squares from model M_j:

$$T_j = \sqrt{\sum_{i=1}^{n}\left(X_{i,j} - \overline{X}^j\right)^2} = \left\|\boldsymbol{X}^j - \overline{X}^j\mathbf{1}_n\right\|;$$

Fitted values from model M_j: $\widehat{\boldsymbol{X}}^j = \left(\widehat{X}_{1,j},\, \ldots,\, \widehat{X}_{n,j}\right)^\top;$

Coefficient of determination from model M_j:

$$R_j^2 = 1 - \frac{\left\|\boldsymbol{X}^j - \widehat{\boldsymbol{X}}^j\right\|^2}{\left\|\boldsymbol{X}^j - \overline{X}^j\mathbf{1}_n\right\|^2} = 1 - \frac{\left\|\boldsymbol{X}^j - \widehat{\boldsymbol{X}}^j\right\|^2}{T_j^2}.$$

10.1.3 Variance inflation factor and tolerance

If data $\big(Y_i,\, X_{i,1},\, \ldots,\, X_{i,k-1}\big)^\top \overset{\text{i.i.d.}}{\sim} \big(Y,\, X_1,\, \ldots,\, X_{k-1}\big)^\top$:

- R^2: a squared value of a sample coefficient of multiple correlation between Y and $\boldsymbol{X} := \big(X_1,\, \ldots,\, X_{k-1}\big)^\top$.
- R_j^2 $(j = 1,\, \ldots,\, k-1)$:
  a squared value of a sample coefficient of multiple correlation between X_j and $\boldsymbol{X}_{(-j)} := \big(X_1,\, \ldots,\, X_{j-1},\, X_{j+1},\, \ldots,\, X_{k-1}\big)^\top$.

R_j^2 close to 1

- $\boldsymbol{X}^j$ is close to being a linear combination of columns of $\mathbb{X}^{(-j)}$ (remaining columns of the model matrix)

  ⇒ $\boldsymbol{X}^j$ is collinear with the remaining columns of the model matrix

$R_j^2 = 0$

- $\boldsymbol{X}^j$ is orthogonal to all remaining non-intercept regressors
- the jth regressor represented by the random variable X_j is multiply uncorrelated with the remaining regressors represented by the random vector $\boldsymbol{X}_{(-j)}$.

Theorem 10.2 Estimated variances of the LSE of the regression coefficients.

For a given dataset for which a linear model $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$, $\mathrm{rank}(\mathbb{X}_{n \times k}) = k$, $\boldsymbol{X} = \left(\mathbf{1}_n, \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{k-1}\right)$ *is applied, diagonal elements of the matrix* $\widehat{\mathrm{var}}\left(\widehat{\boldsymbol{\beta}} \,|\, \mathbb{Z}\right) = \mathrm{MS}_e \left(\mathbb{X}^\top \mathbb{X}\right)^{-1}$, *can also be calculated, for* $j = 1, \ldots, k-1$, *as*

$$\widehat{\mathrm{var}}\left(\widehat{\beta}_j \,|\, \mathbb{Z}\right) = \left(\frac{T_Y}{T_j}\right)^2 \cdot \frac{1 - R^2}{n - k} \cdot \frac{1}{1 - R_j^2}.$$

Definition 10.1 Variance inflation factor and tolerance.

For given $j = 1, \ldots, k-1$, the *variance inflation factor* and the tolerance of the jth regressor of the linear model $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$, $\mathrm{rank}(\mathbb{X}_{n\times k}) = k$ are values VIF_j and Toler_j, respectively, defined as

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}, \qquad \mathrm{Toler}_j = 1 - R_j^2 = \frac{1}{\mathrm{VIF}_j}.$$

10.1.3 Variance inflation factor and tolerance

Interpretation and use of VIF

$(1-\alpha)\,100\%$ confidence interval for β_j, $j = 1, \ldots, k-1$ (under normality)

$$\widehat{\beta}_j \;\pm\; \mathrm{t}_{n-k}\Bigl(1-\frac{\alpha}{2}\Bigr)\,\sqrt{\widehat{\mathrm{var}}\bigl(\widehat{\beta}_j\,\big|\,\mathbb{Z}\bigr)},$$

$$\widehat{\beta}_j \;\pm\; \mathrm{t}_{n-k}\Bigl(1-\frac{\alpha}{2}\Bigr)\,\frac{T_Y}{T_j}\,\sqrt{\frac{1-R^2}{n-k}}\,\sqrt{\mathrm{VIF}_j}.$$

Variance inflation factor

$$\mathrm{VIF}_j = \left(\frac{\mathrm{Vol}_j}{\mathrm{Vol}_{0,j}}\right)^2,$$

$\mathrm{Vol}_j =$ length (volume) of the confidence interval for β_j;

$\mathrm{Vol}_{0,j} =$ length (volume) of the confidence interval for β_j if it was $R_j^2 = 0$.

10.1.4 Basic treatment of multicollinearity

Especially if interest in inference on β's (evaluation of the covariate effects):

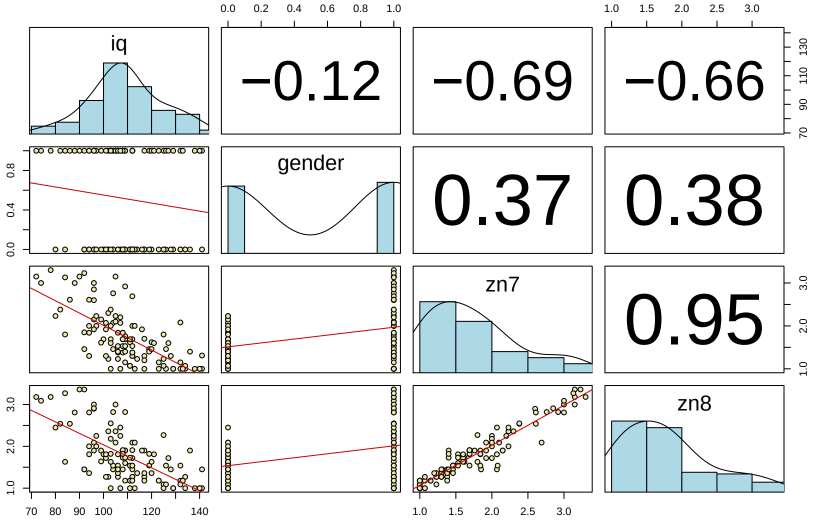
- Do not include mutually highly correlated regressors in one model.
  - At first step, basic decision based on sample correlation coefficients.
  - In some (especially econometric) literature, rules of thumb are applied like *"Regressors with a correlation (in absolute value) higher than 0.80 should not be included together in one model."*
  - Such rules should never be applied in an automatic manner (why just 0.80 and not 0.79, . . . ?)
- Deep analyzis of mutual relationships among regressors must precede any regression modelling!

10.1.4 Basic treatment of multicollinearity

Especially if interest in inference on β's (evaluation of the covariate effects):

- Decisions of which regressors are collinear and should be removed can also be based on (generalized) variance inflation factors and possibly values of standardized regression coefficients (see Proof of Theorem 10.2) that are comparable among regressors (higher value of $\beta_j^\star$ means higher practical importance of a particular regressor).
- Regularization methods (Ridge regression, LASSO, ..., *not covered by this course*).

iq ~ gender + zn7 + zn8



IQ ($n = 111$)

iq ∼ gender + zn7 + zn8

```
summary(m1 <- lm(iq ~ gender + zn7 + zn8, data = IQ))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-22.1677  -7.5243  -0.4338   7.1780  26.4095

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  138.222      3.119  44.314  < 2e-16 ***
gender         4.563      2.221   2.055  0.04232 *
zn7          -16.767      5.536  -3.029  0.00308 **
zn8           -1.149      5.557  -0.207  0.83658
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.81 on 107 degrees of freedom
Multiple R-squared:  0.4943,	Adjusted R-squared:  0.4801
F-statistic: 34.87 on 3 and 107 DF,  p-value: 8.472e-16
```

```
library("car")
vif(m1)
```

```
  gender      zn7      zn8
 1.16923 11.26866 11.40240
```

IQ ($n = 111$)

iq ∼ gender + zn7

```r
(sm27 <- summary(m27 <- lm(iq ~ gender + zn7, data = IQ)))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-21.9606  -7.4290  -0.1927   7.0047  26.5244

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  138.093      3.043  45.376   <2e-16 ***
gender         4.513      2.198   2.054   0.0424 *
zn7          -17.852      1.765 -10.116   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.77 on 108 degrees of freedom
Multiple R-squared:  0.4941,	Adjusted R-squared:  0.4848
F-statistic: 52.74 on 2 and 108 DF,  p-value: < 2.2e-16
```

```r
vif(m27)
```

```
 gender     zn7
1.15531 1.15531
```

IQ ($n = 111$)

$\texttt{iq} \sim \texttt{gender} + \texttt{zn8}$

```
(sm28 <- summary(m28 <- lm(iq ~ gender + zn8, data = IQ)))
```

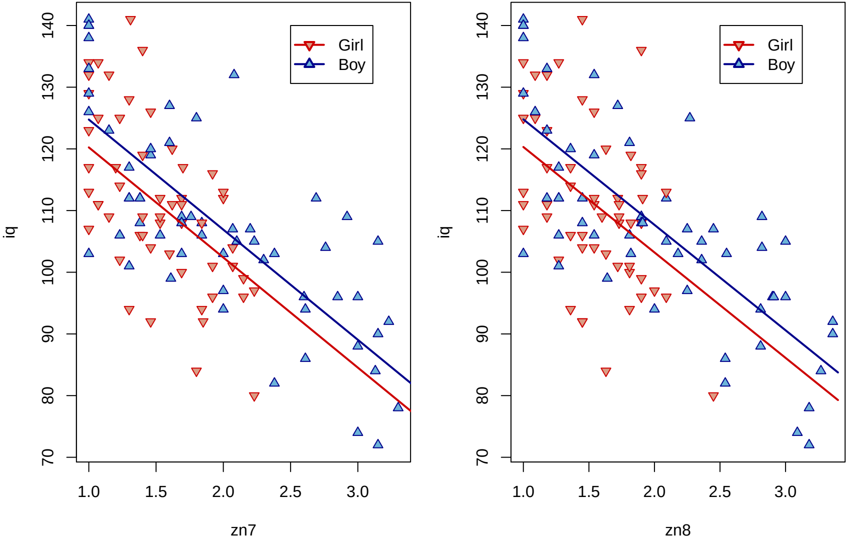
```
Residuals:
     Min       1Q   Median       3Q      Max
-25.5378  -7.9585  -0.0763   7.1273  31.0778

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  137.402      3.223  42.634  < 2e-16 ***
gender         4.474      2.303   1.943   0.0547 .
zn8          -17.095      1.846  -9.263 2.21e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.22 on 108 degrees of freedom
Multiple R-squared:  0.451,	Adjusted R-squared:  0.4408
F-statistic: 44.36 on 2 and 108 DF,  p-value: 8.673e-15
```

```
vif(m28)
```

```
  gender      zn8
1.169022 1.169022
```

IQ ($n = 111$)

iq ~ gender + znX



Section **10.2**

Misspecified regression space

10.2.1 Omitted and irrelevant regressors

Data $\left(Y_i, \boldsymbol{Z}_i^\top\right)^\top, \; i = 1, \dots, n$

$\Rightarrow$ Two sets of regressors:

$$\boldsymbol{X}_i = \boldsymbol{t}_X(\boldsymbol{Z}_i) \qquad \longrightarrow \mathbb{X}_{n\times k} = \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix} = \left(\boldsymbol{X}^0, \dots, \boldsymbol{X}^{k-1}\right)$$

$$\boldsymbol{V}_i = \boldsymbol{t}_V(\boldsymbol{Z}_i) \qquad \longrightarrow \mathbb{V}_{n\times l} = \begin{pmatrix} \boldsymbol{V}_1^\top \\ \vdots \\ \boldsymbol{V}_n^\top \end{pmatrix} = \left(\boldsymbol{V}^1, \dots, \boldsymbol{V}^l\right)$$

Assumptions

$$\mathrm{rank}\left(\mathbb{X}_{n\times k}\right) = k, \quad \mathrm{rank}\left(\mathbb{V}_{n\times l}\right) = l,$$

$$\text{for} \quad \mathbb{G}_{n\times(k+l)} := \left(\mathbb{X}, \mathbb{V}\right), \quad \mathrm{rank}\left(\mathbb{G}\right) = k + l < n.$$

10.2.1 Omitted and irrelevant regressors

Omitted important regressors

- M_{XV} is correct (with $\gamma \neq \mathbf{0}_l$) but inference based on M_X.
  - β estimated using M_X;
  - σ^2 estimated using M_X;
  - prediction based on fitted M_X.

Irrelevant regressors included in a model

- M_X is correct but inference based on M_{XV}.
  - β estimated using M_{XV};
  - σ^2 estimated using M_{XV};
  - prediction based on fitted M_{XV}.

10.2.1 Omitted and irrelevant regressors

Quantities derived under model M_X: $\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right)$

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_X &= \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top\boldsymbol{Y} = \left(\widehat{\beta}_{X,0},\, \dots,\, \widehat{\beta}_{X,k-1}\right)^\top,\\
\mathbb{H}_X &= \mathbb{X}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top,\\
\mathbb{M}_X &= \mathbf{I}_n - \mathbb{H}_X,\\
\widehat{\boldsymbol{Y}}_X &= \mathbb{H}_X\boldsymbol{Y} = \mathbb{X}\widehat{\boldsymbol{\beta}}_X = \left(\widehat{Y}_{X,1}, \dots, \widehat{Y}_{X,n}\right)^\top,\\
\boldsymbol{U}_X &= \boldsymbol{Y} - \widehat{\boldsymbol{Y}}_X = \mathbb{M}_X\boldsymbol{Y} = \left(U_{X,1}, \dots, U_{X,n}\right)^\top,\\
\mathsf{SS}_{e,X} &= \left\|\boldsymbol{U}_X\right\|^2,\\
\mathsf{MS}_{e,X} &= \frac{\mathsf{SS}_{e,X}}{n-k}.
\end{aligned}
$$

10.2.1 Omitted and irrelevant regressors

Quantities derived under model M_{XV}: $\boldsymbol{Y} \,|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\, \sigma^2\mathbf{I}_n\big)$, $\mathbb{G} = \big(\mathbb{X},\, \mathbb{V}\big)$

$$\big(\widehat{\boldsymbol{\beta}}_{XV}^\top,\, \widehat{\boldsymbol{\gamma}}_{XV}^\top\big)^\top = \big(\mathbb{G}^\top\mathbb{G}\big)^{-1}\mathbb{G}^\top\boldsymbol{Y},$$

$$\widehat{\boldsymbol{\beta}}_{XV} = \big(\widehat{\beta}_{XV,0},\, \dots,\, \widehat{\beta}_{XV,k-1}\big)^\top, \qquad \widehat{\boldsymbol{\gamma}}_{XV} = \big(\widehat{\gamma}_{XV,1},\, \dots,\, \widehat{\gamma}_{XV,l}\big)^\top,$$

$$\mathbb{H}_{XV} = \mathbb{G}\big(\mathbb{G}^\top\mathbb{G}\big)^{-1}\mathbb{G}^\top,$$

$$\mathbb{M}_{XV} = \mathbf{I}_n - \mathbb{H}_{XV},$$

$$\widehat{\boldsymbol{Y}}_{XV} = \mathbb{H}_{XV}\boldsymbol{Y} = \mathbb{X}\widehat{\boldsymbol{\beta}}_{XV} + \mathbb{V}\widehat{\boldsymbol{\gamma}}_{XV} = \big(\widehat{Y}_{XV,1},\dots,\widehat{Y}_{XV,n}\big)^\top,$$

$$\boldsymbol{U}_{XV} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}_{XV} = \mathbb{M}_{XV}\boldsymbol{Y} = \big(U_{XV,1},\dots,U_{XV,n}\big)^\top,$$

$$\mathsf{SS}_{e,XV} = \big\|\boldsymbol{U}_{XV}\big\|^2,$$

$$\mathsf{MS}_{e,XV} = \frac{\mathsf{SS}_{e,XV}}{n-k-l}.$$

Consequence of Lemma 9.1: Relationship between the quantities derived while assuming the two models.

Quantities derived while assuming models M_X *and* M_{XV} *are mutually in the following relationships:*

$$\begin{aligned}\widehat{\boldsymbol{Y}}_{XV} - \widehat{\boldsymbol{Y}}_X &= \mathbb{M}_X \mathbb{V}\big(\mathbb{V}^\top \mathbb{M}_X \mathbb{V}\big)^{-1} \mathbb{V}^\top \boldsymbol{U}_X, \\ &= \mathbb{X}\big(\widehat{\boldsymbol{\beta}}_{XV} - \widehat{\boldsymbol{\beta}}_X\big) + \mathbb{V}\widehat{\boldsymbol{\gamma}}_{XV}, \\ \widehat{\boldsymbol{\gamma}}_{XV} &= \big(\mathbb{V}^\top \mathbb{M}_X \mathbb{V}\big)^{-1} \mathbb{V}^\top \boldsymbol{U}_X, \\ \widehat{\boldsymbol{\beta}}_{XV} - \widehat{\boldsymbol{\beta}}_X &= -\big(\mathbb{X}^\top \mathbb{X}\big)^{-1} \mathbb{X}^\top \mathbb{V}\widehat{\boldsymbol{\gamma}}_{XV}, \\ \mathsf{SS}_{e,X} - \mathsf{SS}_{e,XV} &= \big\|\mathbb{M}_X \mathbb{V}\widehat{\boldsymbol{\gamma}}_{XV}\big\|^2,\end{aligned}$$

$$\mathbb{H}_{XV} = \mathbb{H}_X + \mathbb{M}_X \mathbb{V}\left(\mathbb{V}^\top \mathbb{M}_X \mathbb{V}\right)^{-1} \mathbb{V}^\top \mathbb{M}_X.$$

Lemma 10.3 Variance of the LSE in the two models.

Irrespective of whether M_X *or* M_{XV} *holds, the covariance matrices of the fitted values and the LSE of the regression coefficients satisfy the following:*

$$\operatorname{var}\big(\widehat{\boldsymbol{Y}}_{XV} \,\big|\, \mathbb{Z}\big) - \operatorname{var}\big(\widehat{\boldsymbol{Y}}_{X} \,\big|\, \mathbb{Z}\big) \geq 0,$$

$$\operatorname{var}\big(\widehat{\boldsymbol{\beta}}_{XV} \,\big|\, \mathbb{Z}\big) - \operatorname{var}\big(\widehat{\boldsymbol{\beta}}_{X} \,\big|\, \mathbb{Z}\big) \geq 0.$$

10.2.2 Prediction quality of the fitted model

Data and Model

Data: $\big(Y_i,\, \boldsymbol{Z}_i^\top\big)^\top$, $\boldsymbol{Z}_i = \big(Z_{i,1}, \ldots, Z_{i,p}\big)^\top \in \mathcal{Z} \subseteq \mathbb{R}^p$, $i = 1, \ldots, n$

$\equiv$ random sample from a distribution of $\big(Y,\, \boldsymbol{Z}^\top\big)^\top$,

$\boldsymbol{Z} = \big(Z_1, \ldots, Z_p\big)^\top$.

Model: $\mathbb{E}\big(Y \,\big|\, \boldsymbol{Z}\big) = m(\boldsymbol{Z})$, $\quad \mathrm{var}\big(Y \,\big|\, \boldsymbol{Z}\big) = \sigma^2$,

Unknowns: parameters in m, $\sigma^2 > 0$.

Replicated response

- $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$: values of $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ in data.
- New data: $\left(Y_{n+i}, \boldsymbol{Z}_{n+i}^\top\right)^\top \overset{\text{i.i.d.}}{\sim} \left(Y, \boldsymbol{Z}\right)^\top$, $i = 1, \ldots, n$,

  independent of (old data) $\left(Y_i, \boldsymbol{Z}_i^\top\right)^\top$, $i = 1, \ldots, n$

  with the response vector $\boldsymbol{Y} = \left(Y_1, \ldots, Y_n\right)^\top$.
- **AIM**: Predict Y_{n+i} given $\boldsymbol{Z}_{n+i} = \boldsymbol{z}_i$, $i = 1, \ldots, n$

$$\boldsymbol{Y}_{new} = \left(Y_{n+1}, \ldots, Y_{n+n}\right)^\top \equiv \text{replicated response vector}$$

if Y_{n+i} generated by the conditional distribution $Y \,\big|\, \boldsymbol{Z} = \boldsymbol{z}_i$.

Prediction of replicated response

$\widehat{\boldsymbol{Y}}_{new} := \left(\widehat{Y}_{n+1}, \ldots, \widehat{Y}_{n+n}\right)^\top$: prediction of $\boldsymbol{Y}_{new}$ based on the assumed model fitted using the original data $\boldsymbol{Y}$ with $\boldsymbol{Z}_1 = \boldsymbol{z}_1, \ldots, \boldsymbol{Z}_n = \boldsymbol{z}_n$

➠ $\widehat{\boldsymbol{Y}}_{new}$ is some statistic of $\boldsymbol{Y}$ (and $\mathbb{Z}$).

Evaluation of quality of prediction by MSEP, differences as compared to Sec. 7.3

- Value of a random vector rather than of a random variable predicted now

  ➠ $\text{MSEP} = \sum \text{MSEP}_i$
- Interest in knowing on how the prediction performs if new data contain the same covariate values as the old data

  ➠ all statements will be calculated conditionally given $\mathbb{Z}$
- Sample variability induced by estimation of parameters will be taken into account

Prediction of replicated response

Definition 10.2 Quantification of a prediction quality of the fitted regression model.

Prediction quality of the fitted regression model will be evaluated by the *mean squared error of prediction (MSEP)* defined as

$$\mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new}\big) = \sum_{i=1}^{n} \mathbb{E}\Big\{\big(\widehat{Y}_{n+i} - Y_{n+i}\big)^2 \,\Big|\, \mathbb{Z}\Big\},$$

where the expectation is with respect to the $(n+n)$-dimensional conditional distribution of the vector $\big(\boldsymbol{Y}^\top, \boldsymbol{Y}_{new}^\top\big)^\top$ given

$$\mathbb{Z} = \begin{pmatrix} \boldsymbol{Z}_1^\top \\ \vdots \\ \boldsymbol{Z}_n^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{Z}_{n+1}^\top \\ \vdots \\ \boldsymbol{Z}_{n+n}^\top \end{pmatrix}.$$

TO BE CONTINUED.

Prediction of replicated response

Definition 10.2 Quantification of a prediction quality of the fitted regression model, cont'd.

Additionally, we define the *averaged mean squared error of prediction (AMSEP)* as

$$\mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new}\big) = \frac{1}{n}\,\mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new}\big).$$

10.2.2 Prediction quality of the fitted model

Prediction of replicated response in a linear model

Linear model

$$\begin{aligned}
\boldsymbol{\mu} &= \big(\mu_1,\,\dots,\,\mu_n\big)^\top \\
&= \mathbb{E}\big(\boldsymbol{Y}\,\big|\,\boldsymbol{Z}_1 = \boldsymbol{z}_1,\,\dots,\,\boldsymbol{Z}_n = \boldsymbol{z}_n\big) \\
&= \mathbb{E}\big(\boldsymbol{Y}_{new}\,\big|\,\boldsymbol{Z}_{n+1} = \boldsymbol{z}_1,\,\dots,\,\boldsymbol{Z}_{n+n} = \boldsymbol{z}_n\big)
\end{aligned}$$

is

$$\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} = \big(\boldsymbol{x}_1^\top\boldsymbol{\beta},\,\dots,\,\boldsymbol{x}_n^\top\boldsymbol{\beta}\big)^\top, \quad \mathbb{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{t}_X^\top(\boldsymbol{z}_1) \\ \vdots \\ \boldsymbol{t}_X^\top(\boldsymbol{z}_n) \end{pmatrix}$$

10.2.2 Prediction quality of the fitted model

Prediction of replicated response in a linear model

Best linear unbiased prediction

Just another variant of Gauss-Markov theorem:

$\text{MSEP}\big(\widehat{\boldsymbol{Y}}_{new}\big)$ is subject to

(i) linearity ($\widehat{\boldsymbol{Y}}_{new} = \boldsymbol{a} + \mathbb{A}\boldsymbol{Y}$ for some $\boldsymbol{a}$ and $\mathbb{A}$);

(ii) unbiasedness ($\mathbb{E}\big(\widehat{\boldsymbol{Y}}_{new} \,\big|\, \mathbb{Z}\big) = \boldsymbol{\mu}$)

minimized for

$$\widehat{\boldsymbol{Y}}_{new} = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-}\mathbb{X}^\top\boldsymbol{Y} = \widehat{\boldsymbol{Y}} =: \widehat{\boldsymbol{\mu}}$$

➠ best linear unbiased prediction (BLUP) of $\boldsymbol{Y}_{new}$

Lemma 10.4 Mean squared error of the BLUP in a linear model.

In a linear model, the mean squared error of the best linear unbiased prediction can be expressed as

$$\mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new}\big) \;=\; n\,\sigma^2 + \sum_{i=1}^{n} \mathrm{MSE}\big(\widehat{Y}_i\big),$$

where

$$\mathrm{MSE}\big(\widehat{Y}_i\big) = \mathbb{E}\Big\{\big(\widehat{Y}_i - \mu_i\big)^2 \,\Big|\, \mathbb{Z}\Big\}, \qquad i = 1, \dots, n,$$

is the mean squared error of $\widehat{Y}_i$ if this is viewed as estimator of μ_i, $i = 1, \dots, n$.

Correct model

$$\mathsf{M}_{XV}: \quad \boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\ \sigma^2 \mathbf{I}_n\right), \qquad \text{with } \boldsymbol{\gamma} \neq \mathbf{0}_l$$

Properties of LSE derived under the correct model $\mathsf{M}_{XV}: \ \boldsymbol{Y} \,|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\ \sigma^2 \mathbf{I}_n\right)$

$$\begin{aligned}
\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{XV} \,\middle|\, \mathbb{Z}\right) &= \boldsymbol{\beta}, \\
\mathbb{E}\left(\widehat{\boldsymbol{Y}}_{XV} \,\middle|\, \mathbb{Z}\right) &= \mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma} =: \boldsymbol{\mu}, \\
\sum_{i=1}^{n} \mathrm{MSE}\left(\widehat{Y}_{XV,i}\right) &= \sum_{i=1}^{n} \mathrm{var}\left(\widehat{Y}_{XV,i} \,\middle|\, \mathbb{Z}\right) = \mathrm{tr}\left(\mathrm{var}\left(\widehat{\boldsymbol{Y}}_{XV} \,\middle|\, \mathbb{Z}\right)\right) = \mathrm{tr}\left(\sigma^2\, \mathbb{H}_{XV}\right) \\
&= \sigma^2\,(k + l), \\
\mathbb{E}\left(\mathrm{MS}_{e,XV} \,\middle|\, \mathbb{Z}\right) &= \sigma^2.
\end{aligned}$$

Lemma 10.5 Properties of the LSE in a model with omitted regressors.

Let $\mathsf{M}_{XV}: \; \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\, \sigma^2 \mathbf{I}_n\big)$ *hold, i.e.,* $\boldsymbol{\mu} := \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z}\big)$ *satisfies*

$$\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma}$$

for some $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{\gamma} \in \mathbb{R}^l$.

Then the least squares estimators derived while assuming model $\mathsf{M}_X: \; \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big)$ *attain the following properties:*

$$\begin{aligned}
\mathbb{E}\big(\widehat{\boldsymbol{\beta}}_X \,\big|\, \mathbb{Z}\big) &= \boldsymbol{\beta} + \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\mathbb{V}\boldsymbol{\gamma}, \\
\mathbb{E}\big(\widehat{\boldsymbol{Y}}_X \,\big|\, \mathbb{Z}\big) &= \boldsymbol{\mu} - \mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}, \\
\sum_{i=1}^{n} \mathrm{MSE}\big(\widehat{Y}_{X,i}\big) &= k\,\sigma^2 + \big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2, \\
\mathbb{E}\big(\mathrm{MS}_{e,X} \,\big|\, \mathbb{Z}\big) &= \sigma^2 + \frac{\big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2}{n-k}.
\end{aligned}$$

10.2.3 Omitted regressors

Least squares estimators

Omitted regressors

$$\widehat{\boldsymbol{\beta}}_X = \widehat{\boldsymbol{\beta}}_{XV} + \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{V} \widehat{\boldsymbol{\gamma}}_{XV}$$

$$\text{bias}\big(\widehat{\boldsymbol{\beta}}_X\big) \;=\; \mathbb{E}\big(\widehat{\boldsymbol{\beta}}_X - \boldsymbol{\beta} \,\big|\, \mathbb{Z}\big) = \;\left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{V} \boldsymbol{\gamma}$$

(i) $\mathbb{X}^\top \mathbb{V} = \mathbf{0}_{k \times l}$

- $\widehat{\boldsymbol{\beta}}_X = \widehat{\boldsymbol{\beta}}_{XV}$;
- $\text{bias}\big(\widehat{\boldsymbol{\beta}}_X\big) = \mathbf{0}_k$.

(ii) $\mathbb{X}^\top \mathbb{V} \neq \mathbf{0}_{k \times l}$

- $\widehat{\boldsymbol{\beta}}_X$ is a *biased* estimator of $\boldsymbol{\beta}$.

Omitted regressors

Compare $\widehat{\boldsymbol{Y}}_{new,X} = \widehat{\boldsymbol{Y}}_X$ and $\widehat{\boldsymbol{Y}}_{new,XV} = \widehat{\boldsymbol{Y}}_{XV}$

$$\begin{aligned}\mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new,XV}\big) &= n\,\sigma^2 + k\,\sigma^2 + l\,\sigma^2,\\ \mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new,X}\big) &= n\,\sigma^2 + k\,\sigma^2 + \big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2.\end{aligned}$$

$$\begin{aligned}\mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,XV}\big) &= \sigma^2 + \frac{k}{n}\sigma^2 + \frac{l}{n}\sigma^2,\\ \mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,X}\big) &= \sigma^2 + \frac{k}{n}\sigma^2 + \frac{1}{n}\big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2.\end{aligned}$$

- The term $\big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2$ might be huge compared to $l\,\sigma^2$.
- $\frac{l}{n}\,\sigma^2 \to 0$ with $n \to \infty$ (while increasing the number of predictions).
- $\frac{1}{n}\big\|\mathbb{M}_X\mathbb{V}\boldsymbol{\gamma}\big\|^2$ does not necessarily tend to zero with $n \to \infty$.

10.2.3 Omitted regressors

Estimator of the residual variance

Omitted regressors

$$\text{bias}\left(\text{MS}_{e,X}\right) = = \mathbb{E}\left(\text{MS}_{e,X} - \sigma^2 \,\middle|\, \mathbb{Z}\right) = \frac{\left\|\mathbb{M}_X \mathbb{V}\gamma\right\|^2}{n-k}$$

10.2.4 Irrelevant regressors

Correct model

$$\begin{aligned} & \mathsf{M}_X: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\right) \\ \equiv \quad & \mathsf{M}_{XV}: \quad \boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \left(\mathbb{X}\boldsymbol{\beta} + \mathbb{V}\boldsymbol{\gamma},\, \sigma^2 \mathbf{I}_n\right), \qquad \text{with } \boldsymbol{\gamma} = \mathbf{0}_l \end{aligned}$$

Properties of LSE derived under the two models

$$
\begin{aligned}
\mathbb{E}\big(\widehat{\boldsymbol{\beta}}_X \,\big|\, \mathbb{Z}\big) = \mathbb{E}\big(\widehat{\boldsymbol{\beta}}_{XV} \,\big|\, \mathbb{Z}\big) &= \boldsymbol{\beta}, \\
\mathbb{E}\big(\widehat{\boldsymbol{Y}}_X \,\big|\, \mathbb{Z}\big) = \mathbb{E}\big(\widehat{\boldsymbol{Y}}_{XV} \,\big|\, \mathbb{Z}\big) &= \mathbb{X}\boldsymbol{\beta} =: \boldsymbol{\mu}, \\
\sum_{i=1}^{n} \text{MSE}\big(\widehat{Y}_{X,i}\big) &= \sum_{i=1}^{n} \text{var}\big(\widehat{Y}_{X,i} \,\big|\, \mathbb{Z}\big) = \text{tr}\Big(\text{var}\big(\widehat{\boldsymbol{Y}}_X \,\big|\, \mathbb{Z}\big)\Big) \\
&= \text{tr}\big(\sigma^2\, \mathbb{H}_X\big) = \sigma^2\, k, \\
\sum_{i=1}^{n} \text{MSE}\big(\widehat{Y}_{XV,i}\big) &= \sum_{i=1}^{n} \text{var}\big(\widehat{Y}_{XV,i} \,\big|\, \mathbb{Z}\big) = \text{tr}\Big(\text{var}\big(\widehat{\boldsymbol{Y}}_{XV} \,\big|\, \mathbb{Z}\big)\Big) \\
&= \text{tr}\big(\sigma^2\, \mathbb{H}_{XV}\big) = \sigma^2\,(k + l), \\
\mathbb{E}\big(\text{MS}_{e,X} \,\big|\, \mathbb{Z}\big) = \mathbb{E}\big(\text{MS}_{e,XV} \,\big|\, \mathbb{Z}\big) &= \sigma^2.
\end{aligned}
$$

Least squares estimators

Irrelevant regressors

$$\mathsf{MSE}\big(\widehat{\boldsymbol{\beta}}_{XV}\big) - \mathsf{MSE}\big(\widehat{\boldsymbol{\beta}}_{X}\big)$$

$$= \mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{XV} - \boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{XV} - \boldsymbol{\beta}\big)^\top \,\Big|\, \mathbb{Z}\Big\} - \mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{X} - \boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{X} - \boldsymbol{\beta}\big)^\top \,\Big|\, \mathbb{Z}\Big\}$$

$$= \mathsf{var}\big(\widehat{\boldsymbol{\beta}}_{XV} \,\big|\, \mathbb{Z}\big) - \mathsf{var}\big(\widehat{\boldsymbol{\beta}}_{X} \,\big|\, \mathbb{Z}\big)$$

$$= \sigma^2\Big[\Big\{\mathbb{X}^\top\mathbb{X} - \mathbb{X}^\top\mathbb{V}(\mathbb{V}^\top\mathbb{V})^{-1}\mathbb{V}^\top\mathbb{X}\Big\}^{-1} - (\mathbb{X}^\top\mathbb{X})^{-1}\Big] \geq 0$$

(i) $\mathbb{X}^\top\mathbb{V} = \mathbf{0}_{k\times l}$

- $\widehat{\boldsymbol{\beta}}_X = \widehat{\boldsymbol{\beta}}_{XV}$ and $\mathsf{var}\big(\widehat{\boldsymbol{\beta}}_X \,\big|\, \mathbb{Z}\big) = \mathsf{var}\big(\widehat{\boldsymbol{\beta}}_{XV} \,\big|\, \mathbb{Z}\big)$

⇒ irrelevant regressors do not influence quality of the LSE of $\boldsymbol{\beta}$

(ii) $\mathbb{X}^\top\mathbb{V} \neq \mathbf{0}_{k\times l}$

- $\widehat{\boldsymbol{\beta}}_{XV}$ is worse $\widehat{\boldsymbol{\beta}}_X$ in terms of its variability
- difference in quality might be huge (multicollinearity. . .)

Irrelevant regressors

Compare $\widehat{\boldsymbol{Y}}_{new,X} = \widehat{\boldsymbol{Y}}_X$ and $\widehat{\boldsymbol{Y}}_{new,XV} = \widehat{\boldsymbol{Y}}_{XV}$

$$\begin{aligned}\mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new,XV}\big) &= n\,\sigma^2 + (k+l)\,\sigma^2,\\ \mathrm{MSEP}\big(\widehat{\boldsymbol{Y}}_{new,X}\big) &= n\,\sigma^2 + k\,\sigma^2.\end{aligned}$$

$$\begin{aligned}\mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,XV}\big) &= \sigma^2 + \frac{k+l}{n}\,\sigma^2,\\ \mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,X}\big) &= \sigma^2 + \frac{k}{n}\,\sigma^2.\end{aligned}$$

- Both $\mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,XV}\big) \to \sigma^2$ and $\mathrm{AMSEP}\big(\widehat{\boldsymbol{Y}}_{new,X}\big) \to \sigma^2$ as $n \to \infty$
- Use of M_{XV} (which for finite n provides worse prediction than M_X) eliminates problem of omitted important covariates that leads to biased predictions with possibly even worse MSEP and AMSEP than that of model M_{XV}

10.2.5 Summary

Interest in estimation of the regression coefficients and inference on them

- Omitting important regressors which are (multiply) correlated with regressors of main interest
  ➠ bias in estimation of β.

- Inclusion of irrelevant regressors which are (multiply) correlated with regressors of main interest
  ➠ possible multicollinearity and inflation of standard errors of $\widehat{\beta}$.

- Regressors which are (multiply) uncorrelated with regressors of main interest influence neither bias nor variability of $\widehat{\beta}$ irrespective of whether they are omitted or irrelevantly included.

10.2.5 Summary

Interest in prediction

- Omitting important regressors
  - ➠ biased prediction
  - ➠ the AMSEP not tending to the optimal value of σ^2 with $n \to \infty$
- Including irrelevant regressors
  - ➠ the AMSEP tending to the optimal value of σ^2 with $n \to \infty$
  - ➠ negligible difference of a quality of prediction compared to a model with irrelevant regressors omitted from the model

11

Unusual Observations

11 Unusual Observations

$$\mathsf{M}: \ \boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n\big), \quad \mathrm{rank}(\mathbb{X}_{n\times k}) = k,$$
$$t \in \{1, \dots, n\}$$

Standard notation

- $\widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top \boldsymbol{Y} = \big(\widehat{\beta}_0, \, \dots, \, \widehat{\beta}_{k-1}\big)^\top$: LSE of the vector $\boldsymbol{\beta}$;
- $\mathbb{H} = \mathbb{X}\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top = \big(h_{i,t}\big)_{i,t=1,\dots,n}$: the hat matrix;
- $\mathbb{M} = \mathbf{I}_n - \mathbb{H} = \big(m_{i,t}\big)_{i,t=1,\dots,n}$: the residual projection matrix;
- $\widehat{\boldsymbol{Y}} = \mathbb{H}\boldsymbol{Y} = \mathbb{X}\widehat{\boldsymbol{\beta}} = \big(\widehat{Y}_1, \, \dots, \, \widehat{Y}_n\big)^\top$: the vector of fitted values;
- $\boldsymbol{U} = \mathbb{M}\boldsymbol{Y} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \big(U_1, \, \dots, \, U_n\big)^\top$: the residuals;
- $\mathrm{SS}_e = \big\|\boldsymbol{U}\big\|^2$: the residual sum of squares;
- $\mathrm{MS}_e = \frac{1}{n-k}\,\mathrm{SS}_e$ is the residual mean square;
- $\boldsymbol{U}^{std} = \big(U_1^{std}, \, \dots, \, U_n^{std}\big)^\top$: vector of standardized residuals,

$$U_i^{std} = \frac{U_i}{\sqrt{\mathrm{MS}_e\, m_{i,i}}}, \ i = 1, \dots, n.$$

Section **11.1**

Leave-one-out and outlier model

11.1 Leave-one-out and outlier model

Definition 11.1 Leave-one-out model.

The tth leave-one-out model is a linear model

$$\mathsf{M}_{(-t)}: \quad \boldsymbol{Y}_{(-t)} \,\big|\, \mathbb{X}_{(-t)} \sim \big(\mathbb{X}_{(-t)}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_{n-1}\big).$$

Definition 11.2 Outlier model.

The tth outlier model is a linear model

$$\mathsf{M}_t^{out}: \quad \boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta} + \boldsymbol{j}_t \gamma_t^{out},\, \sigma^2 \mathbf{I}_n\big).$$

11.1 Leave-one-out and outlier model

Lemma 11.1 Three equivalent statements.

While assuming $\operatorname{rank}(\mathbb{X}_{n\times k}) = k$*, the following three statements are equivalent:*

(i) $\operatorname{rank}(\mathbb{X}) = \operatorname{rank}\left(\mathbb{X}_{(-t)}\right) = k$, *i.e.,* $\boldsymbol{x}_t \in \mathcal{M}\left(\mathbb{X}_{(-t)}^\top\right)$;

(ii) $m_{t,t} > 0$;

(iii) $\operatorname{rank}\left(\mathbb{X}, \boldsymbol{j}_t\right) = k + 1$.

11.1 Leave-one-out and outlier model

Quantities related to $\mathsf{M}_{(-t)}$: $\quad \widehat{\boldsymbol{\beta}}_{(-t)}, \widehat{\boldsymbol{Y}}_{(-t)}, \mathsf{SS}_{e,(-t)}, \mathsf{MS}_{e,(-t)}, \ldots$.

Quantities related to M_t^{out}: $\quad \widehat{\boldsymbol{\beta}}_t^{out}, \widehat{\boldsymbol{Y}}_t^{out}, \mathsf{SS}_{e,t}^{out}, \mathsf{MS}_{e,t}^{out}, \ldots$.

Solutions to normal equations in model M_t^{out} (the LSE of $\big((\boldsymbol{\beta}_t^{out})^\top, \gamma_t^{out}\big)^\top$):

$$\big((\widehat{\boldsymbol{\beta}}_t^{out})^\top, \widehat{\gamma}_t^{out}\big)^\top.$$

11.1 Leave-one-out and outlier model

Lemma 11.2 Equivalence of the outlier model and the leave-one-out model.

1. *The residual sums of squares in models* $\mathsf{M}_{(-t)}$ *and* M_t^{out} *are the same, i.e.,*

$$\mathsf{SS}_{e,(-t)} = \mathsf{SS}_{e,t}^{out}.$$

2. *Vector* $\widehat{\boldsymbol{\beta}}_{(-t)}$ *solves the normal equations of model* $\mathsf{M}_{(-t)}$ *if and only if a vector* $\left((\widehat{\boldsymbol{\beta}}_t^{out})^\top, \widehat{\gamma}_t^{out}\right)^\top$ *solves the normal equations of model* M_t^{out}*, where*

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_t^{out} &= \widehat{\boldsymbol{\beta}}_{(-t)}, \\ \widehat{\gamma}_t^{out} &= Y_t - \boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}}_{(-t)}. \end{aligned}$$

11.1 Leave-one-out and outlier model

Notation: Leave-one-out least squares estimators of the response expectations

If $m_{t,t} > 0$ for all $t = 1, \ldots, n$:

$$\widehat{Y}_{[t]} := \boldsymbol{x}_t^\top \widehat{\beta}_{(-t)}, \quad t = 1, \ldots, n,$$

$$\widehat{\boldsymbol{Y}}_{[\bullet]} := \left(\widehat{Y}_{[1]}, \ldots, \widehat{Y}_{[n]}\right)^\top.$$

11.1 Leave-one-out and outlier model

Calculation of quantities of the outlier and the leave-one-out models

Application of Lemma 9.1

If $m_{t,t} > 0$

$$
\begin{aligned}
\widehat{\gamma}_t^{out} &= \frac{U_t}{m_{t,t}}, \\
\widehat{\boldsymbol{\beta}}_t^{out} &= \widehat{\boldsymbol{\beta}} - \frac{U_t}{m_{t,t}} \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \boldsymbol{x}_t, \\
\widehat{\boldsymbol{Y}}_t^{out} &= \widehat{\boldsymbol{Y}} + \frac{U_t}{m_{t,t}} \boldsymbol{m}_t, \\
\mathrm{SS}_e - \mathrm{SS}_{e,t}^{out} &= \frac{U_t^2}{m_{t,t}} = \mathrm{MS}_e \left(U_t^{std}\right)^2,
\end{aligned}
$$

$\boldsymbol{m}_t$: the tth column (and row as well) of the residual project. matrix $\mathbb{M}$.

Lemma 11.3 Quantities of the outlier and leave-one-out model expressed using quantities of the original model.

Suppose that for given $t \in \{1, \ldots, n\}$, $m_{t,t} > 0$. *The following quantities of the outlier model* M_t^{out} *and the leave-one-out model* $\mathsf{M}_{(-t)}$ *are expressable using the quantities of the original model* M *as follows.*

$$\begin{aligned}
\widehat{\gamma}_t^{out} &= Y_t - \boldsymbol{x}_t^\top \widehat{\beta}_{(-t)} = Y_t - \widehat{Y}_{[t]} = \frac{U_t}{m_{t,t}}, \\
\widehat{\boldsymbol{\beta}}_{(-t)} = \widehat{\boldsymbol{\beta}}_t^{out} &= \widehat{\boldsymbol{\beta}} - \frac{U_t}{m_{t,t}} \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \boldsymbol{x}_t, \\
\mathrm{SS}_{e,(-t)} = \mathrm{SS}_{e,t}^{out} &= \mathrm{SS}_e - \frac{U_t^2}{m_{t,t}} = \mathrm{SS}_e - \mathrm{MS}_e \left(U_t^{std}\right)^2, \\
\frac{\mathrm{MS}_{e,(-t)}}{\mathrm{MS}_e} = \frac{\mathrm{MS}_{e,t}^{out}}{\mathrm{MS}_e} &= \frac{n - k - \left(U_t^{std}\right)^2}{n - k - 1}.
\end{aligned}$$

11.1 Leave-one-out and outlier model

Definition 11.3 Deleted residual.

If $m_{t,t} > 0$, then the quantity

$$\widehat{\gamma}_t^{out} = Y_t - \widehat{Y}_{[t]} = \frac{U_t}{m_{t,t}}$$

is called the tth *deleted residual* of the model M.

Section **11.2**

Outliers

11.2 Outliers

$m_{t,t} > 0$

T-statistic to test $\mathrm{H}_0: \ \gamma_t^{out} = 0$ in the tth outlier model M_t^{out} (if normality assumed):

$$T_t = \frac{\widehat{\gamma}_t^{out}}{\sqrt{\widehat{\mathrm{var}}\left(\widehat{\gamma}_t^{out}\right)}} = \text{some calculation} = \frac{Y_t - \widehat{Y}_{[t]}}{\sqrt{\mathrm{MS}_{e,(-t)}}} \sqrt{m_{t,t}}$$

$$= \text{some calculation} = \frac{U_t}{\sqrt{\mathrm{MS}_{e,(-t)}\, m_{t,t}}}.$$

Under $\mathrm{H}_0: \ \gamma_t^{out} = 0$

$$T_t \sim \mathrm{t}_{n-k-1}.$$

11.2 Outliers

Definition 11.4 Studentized residual.

If $m_{t,t} > 0$, then the quantity

$$T_t = \frac{Y_t - \widehat{Y}_{[t]}}{\sqrt{\mathrm{MS}_{e,(-t)}}}\,\sqrt{m_{t,t}} = \frac{U_t}{\sqrt{\mathrm{MS}_{e,(-t)}\, m_{t,t}}}$$

is called the tth *studentized residual* of the model M.

Expression of the studentized residual using the standardized residual

Use of identity $\frac{\mathrm{MS}_{e,(-t)}}{\mathrm{MS}_e} = \frac{n-k-(U_t^{std})^2}{n-k-1}$:

$$T_t = \sqrt{\frac{n-k-1}{n-k-\left(U_t^{std}\right)^2}}\;U_t^{std}.$$

11.2 Outliers

Lemma 11.4 On studentized residuals.

Let $\boldsymbol{Y} \mid \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, *where* $\text{rank}(\mathbb{X}_{n \times k}) = k < n$. *Let further* $n > k + 1$. *Let for given* $t \in \{1, \ldots, n\}$ $m_{t,t} > 0$. *Then*

1. *The* t*th studentized residual* T_t *follows the Student t-distribution with* $n - k - 1$ *degrees of freedom.*
2. *If additionally* $n > k + 2$ *then* $\mathbb{E}(T_t) = 0$.
3. *If additionally* $n > k + 3$ *then* $\text{var}(T_t) = \dfrac{n - k - 1}{n - k - 3}$.

11.2 Outliers

Test for outliers

M_t^{out}: $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\beta + \boldsymbol{j}_t\gamma_t^{out},\, \sigma^2\mathbf{I}_n\big)$

H_0: $\gamma_t^{out} = 0$,

H_1: $\gamma_t^{out} \neq 0$

M: $\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\beta,\, \sigma^2\mathbf{I}_n\big)$

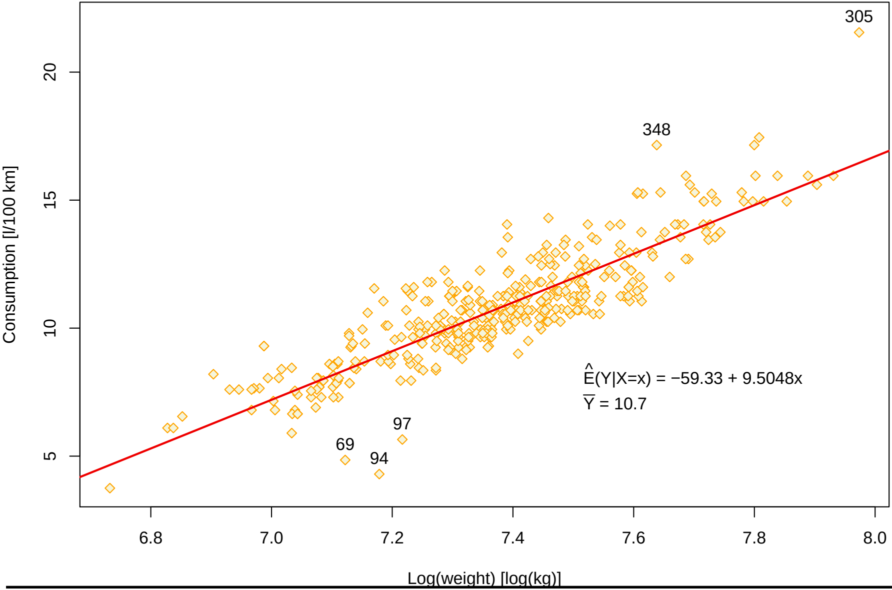
H_0: tth observations is not outlier of model M,

H_1: tth observations is outlier of model M,

- Under H_0: $T_t \sim \mathrm{t}_{n-k-1}$.
- Multiple testing problem!

Cars2004 (subset, $n = 412$), consumption ~ log(weight)

Observations with five highest absolute values of studentized residuals



Standardized, studentized and deleted residuals

Standardized residuals $U_1^{std}, \ldots, U_n^{std}$

```
m1 <- lm(consumption ~ lweight, data = CarsUsed)
rstandard(m1)
```

```
          1           2            3            4            5            6
0.600003668  0.683558025 -0.237013632 -0.437157041 -0.237013632 -0.491068598  ...
```

Studentized residuals $T_1, \ldots, T_n$

```
rstudent(m1)
```

```
          1           2            3            4            5            6
0.599534780  0.683113271 -0.236740634 -0.436725391 -0.236740634 -0.490613671  ...
```

Deleted residuals $\widehat{\gamma}_1^{out}, \ldots, \widehat{\gamma}_n^{out}$

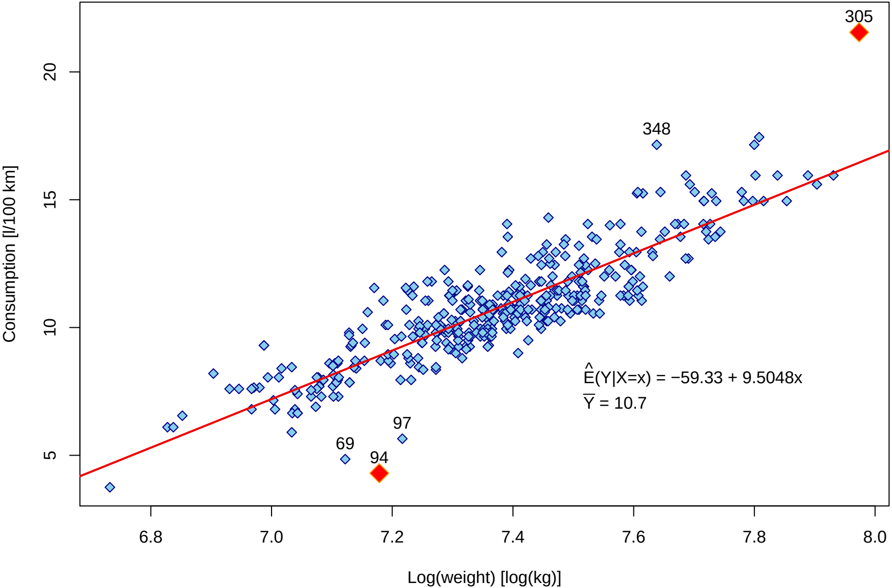
```
residuals(m1) / (1 - hatvalues(m1))
```

```
          1           2            3            4            5            6
0.646454917  0.736641641 -0.254845546 -0.469869858 -0.254845546 -0.528142442  ...
```

Observations with five highest absolute values of studentized residuals

```
                                 vname fhybrid consumption  lweight weight
305 Hummer.H2                               No       21.55 7.973500   2903
94  Toyota.Prius.4dr.(gas/electric)        Yes        4.30 7.178545   1311
348 Land.Rover.Discovery.SE                 No       17.15 7.638198   2076
97  Volkswagen.Jetta.GLS.TDI.4dr            No        5.65 7.216709   1362
69  Honda.Civic.Hybrid                     Yes        4.85 7.122060   1239
    .4dr.manual.(gas/electric)

                                 vname     gamma        Tt    PvalUnadj  PvalBonf
305 Hummer.H2                            5.223712   4.953073    0.000001  0.000441
94  Toyota.Prius.4dr.(gas/electric)     -4.618542  -4.396641    0.000014  0.005782
348 Land.Rover.Discovery.SE              3.910233   3.693509    0.000251  0.103499
97  Volkswagen.Jetta.GLS.TDI.4dr        -3.623890  -3.420244    0.000689  0.283692
69  Honda.Civic.Hybrid                  -3.531883  -3.327145    0.000957  0.394186
    .4dr.manual.(gas/electric)
```

Identified outliers



To know about outliers

- Two or more outliers next to each other can hide each other.
- A notion of outlier is always relative to considered model (also in other areas of statistics). Observation which is outlier with respect to one model is not necessarily an outlier with respect to some other model.
- Especially in large datasets, few outliers are not a problem provided they are not at the same time also influential for statistical inference.
- In our context (of a normal linear model), presence of outliers may indicate that the error distribution is some distribution with heavier tails than the normal distribution.
- Outlier can also suggest that a particular observation is a data-error.

Treatment of outliers

NEVER, NEVER, NEVER exclude "outliers" from the analysis in an automatic manner.

If some observation is indicated to be an outlier, it should always be explored:

- Is it a data-error? If yes, try to correct it, if this is impossible, no problem (under certain assumptions) to exclude it from the data.
- Is the assumed model correct and it is possible to find a physical/practical explanation for occurrence of such unusual observation?
- If an explanation is found, are we interested in capturing such artefacts by our model or not?
- Do the outlier(s) show a serious deviation from the model that cannot be ignored (for the purposes of a particular modelling)?
- ⋮

Treatment of outliers

Often, identification of outliers with respect to some model is of primary interest:

- Example: model for amount of credit card transactions over a certain period of time depending on some factors (age, gender, income, . . .).

  Model found to be correct for a “standard” population (of clients).

  Outlier with respect to such model $\equiv$ potentially a fraudulent use of the credit card.

If the closer analysis of “outliers” suggest that the assumed model is not satisfactory capturing the reality we want to capture (it is not useful), some other model (maybe not linear, maybe not normal) must be looked for.

Section **11.3**

Leverage points

11.3 Leverage points

Terminology Leverage

A diagonal element $h_{t,t}$ ($t = 1, \ldots, n$) of the hat matrix $\mathbb{H}$ is called the *leverage* of the tth observation.

11.3 Leverage points

Interpretation of the leverage

Model with intercept and the column means

$$\mathbb{X} = \left(\mathbf{1}_n,\, \boldsymbol{x}^1,\, \dots,\, \boldsymbol{x}^{k-1}\right) = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k-1} \end{pmatrix}$$

$$\overline{x}^1 = \frac{1}{n}\sum_{i=1}^{n} x_{i,1}, \quad \dots, \quad \overline{x}^{k-1} = \frac{1}{n}\sum_{i=1}^{n} x_{i,k-1}$$

Non-intercept columns centered

$$\widetilde{\mathbb{X}} = \left(\boldsymbol{x}^1 - \overline{x}^1\mathbf{1}_n, \quad \dots, \quad \boldsymbol{x}^{k-1} - \overline{x}^{k-1}\mathbf{1}_n\right) = \begin{pmatrix} x_{1,1} - \overline{x}^1 & \dots & x_{1,k-1} - \overline{x}^{k-1} \\ \vdots & \vdots & \vdots \\ x_{n,1} - \overline{x}^1 & \dots & x_{n,k-1} - \overline{x}^{k-1} \end{pmatrix},$$

$$\mathcal{M}\left(\mathbb{X}\right) = \mathcal{M}\left(\mathbf{1}_n,\, \widetilde{\mathbb{X}}\right), \qquad \mathbf{1}_n^\top \widetilde{\mathbb{X}} = \mathbf{0}_{k-1}^\top.$$

11.3 Leverage points

Interpretation of the leverage

The hat matrix

$$\mathbb{H} = \left(\mathbf{1}_n, \widetilde{\mathbb{X}}\right) \left\{ \left(\mathbf{1}_n, \widetilde{\mathbb{X}}\right)^\top \left(\mathbf{1}_n, \widetilde{\mathbb{X}}\right) \right\}^{-1} \left(\mathbf{1}_n, \widetilde{\mathbb{X}}\right)^\top$$

$$= \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top + \widetilde{\mathbb{X}} \left(\widetilde{\mathbb{X}}^\top \widetilde{\mathbb{X}}\right)^{-1} \widetilde{\mathbb{X}}^\top$$

The leverage

$$h_{t,t} = \frac{1}{n} +$$

$$\left(x_{t,1} - \overline{x}^1, \ \ldots, \ x_{t,k-1} - \overline{x}^{k-1}\right) \left(\widetilde{\mathbb{X}}^\top \widetilde{\mathbb{X}}\right)^{-1} \left(x_{t,1} - \overline{x}^1, \ \ldots, \ x_{t,k-1} - \overline{x}^{k-1}\right)^\top$$

11.3 Leverage points

High value of a leverage

$\mathbb{Q}$: $n \times k$ matrix with the orthonormal basis of the regression space $\mathcal{M}(\mathbb{X})$

$$\sum_{i=1}^{n} h_{i,i} = \mathrm{tr}(\mathbb{H}) = \mathrm{tr}\big(\mathbb{Q}\mathbb{Q}^\top\big) = \mathrm{tr}\big(\mathbb{Q}^\top\mathbb{Q}\big) = \mathrm{tr}(\mathbf{I}_k) = k.$$

Mean value of the leverage

$$\overline{h} = \frac{1}{n}\sum_{i=1}^{n} h_{i,i} = \frac{k}{n}.$$

R function `influence.measures` rule-of-thumb

tth observation is a leverage point if

$$h_{t,t} > \frac{3\,k}{n}.$$

11.3 Leverage points

Influence of leverage points

$$\mathrm{var}\big(U_t \,\big|\, \mathbb{X}\big) = \mathrm{var}\big(Y_t - \widehat{Y}_t \,\big|\, \mathbb{X}\big) = \sigma^2\, m_{t,t} = \sigma^2\,(1 - h_{t,t}), \qquad t = 1, \ldots, n.$$

- High leverage $\quad\Longrightarrow\quad$ low $\mathrm{var}\big(U_t \,\big|\, \mathbb{X}\big) = \mathrm{var}\big(Y_t - \widehat{Y}_t \,\big|\, \mathbb{X}\big)$

➡ the tth fitted value is forced to be close to the observed response value.

Leverages and influence measures

Leverages $h_{1,1}, \ldots, h_{n,n}$

```
m1 <- lm(consumption ~ lweight, data = CarsUsed)
hatvalues(m1)
```

```
          1           2           3           4           5           6
0.011453373 0.011892770 0.007436292 0.006688146 0.007436292 0.007916965   ...
```

Influence measures

```
influence.measures(m1)
```

```
Influence measures of
         lm(formula = consumption ~ lweight, data = CarsUsed) :

       dfb.1_  dfb.lwgh     dffit cov.r   cook.d     hat inf
1    5.81e-02 -5.73e-02  0.064533 1.015 2.09e-03 0.01145   *
2    6.78e-02 -6.69e-02  0.074943 1.015 2.81e-03 0.01189   *
3   -1.71e-02  1.68e-02 -0.020491 1.012 2.10e-04 0.00744
4   -2.92e-02  2.86e-02 -0.035836 1.011 6.43e-04 0.00669
5   -1.71e-02  1.68e-02 -0.020491 1.012 2.10e-04 0.00744
6   -3.71e-02  3.65e-02 -0.043827 1.012 9.62e-04 0.00792
7   -4.59e-02  4.50e-02 -0.055070 1.010 1.52e-03 0.00732
8    7.70e-03 -7.56e-03  0.009196 1.012 4.24e-05 0.00749
9   -2.15e-02  2.11e-02 -0.025596 1.012 3.28e-04 0.00758
...
```

Potentially influential observations

```
summary(influence.measures(m1))
```

```
Potentially influential observations of
         lm(formula = consumption ~ lweight, data = CarsUsed) :

    dfb.1_ dfb.lwgh dffit   cov.r   cook.d hat
1     0.06  -0.06     0.06    1.01_*  0.00   0.01
2     0.07  -0.07     0.07    1.01_*  0.00   0.01
17    0.07  -0.07     0.07    1.01_*  0.00   0.01
39   -0.01   0.01    -0.01    1.02_*  0.00   0.01
47    0.07  -0.07     0.07    1.02_*  0.00   0.02_*
48    0.09  -0.09     0.10    1.02_*  0.00   0.02_*
49    0.06  -0.06     0.06    1.02_*  0.00   0.02_*
69   -0.21   0.20    -0.26_*  0.96_*  0.03   0.01
70   -0.14   0.14    -0.14    1.03_*  0.01   0.03_*
94   -0.21   0.20    -0.30_*  0.92_*  0.04   0.00
97   -0.13   0.13    -0.21_*  0.95_*  0.02   0.00
204  -0.05   0.06     0.14    0.98_*  0.01   0.00
270   0.20  -0.20     0.22_*  0.99    0.02   0.01
271   0.20  -0.20     0.22_*  0.99    0.02   0.01
278   0.05  -0.04     0.12    0.98_*  0.01   0.00
294   0.21  -0.21     0.23_*  1.00    0.03   0.02_*
295  -0.02   0.02     0.02    1.02_*  0.00   0.01
301   0.00   0.00    -0.01    1.02_*  0.00   0.01
302   0.00   0.00     0.00    1.01_*  0.00   0.01
...
```

Leverage points

$\frac{3\,k}{n} = 0.0146$

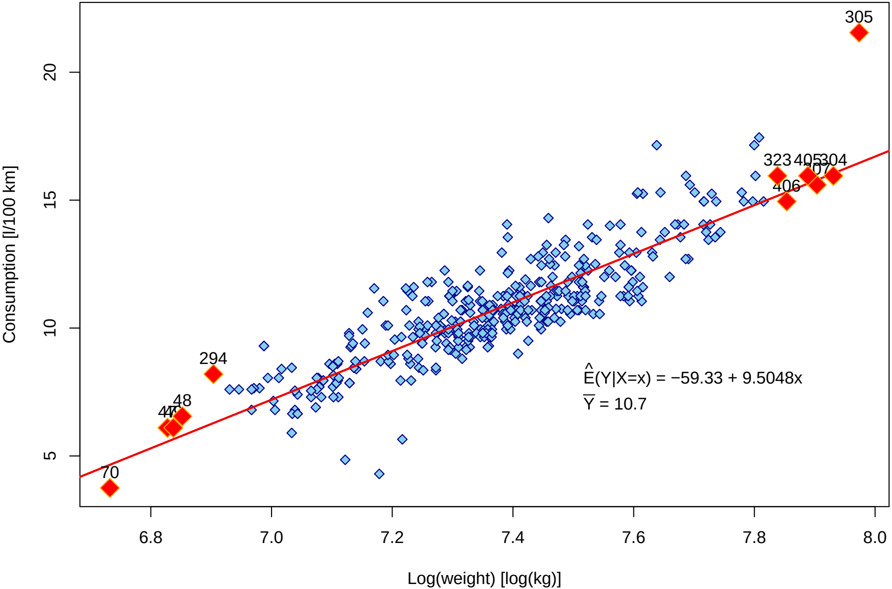
```
sum(hatvalues(m1) > 3 * k / n)
```

```
[1] 11
```

```
                              vname consumption weight  lweight          h
47           Toyota.Echo.2dr.manual        6.10    923 6.827629 0.01992471
48             Toyota.Echo.2dr.auto        6.55    946 6.852243 0.01836889
49                 Toyota.Echo.4dr        6.10    932 6.837333 0.01930270
70   Honda.Insight.2dr.(gas/electric)      3.75    839 6.732211 0.02664081
294 Toyota.MR2.Spyder.convertible.2dr      8.20    996 6.903747 0.01534760
304           GMC.Yukon.XL.2500.SLT       15.95   2782 7.930925 0.02132481
305                        Hummer.H2      21.55   2903 7.973500 0.02429502
307         Lincoln.Navigator.Luxury      15.60   2707 7.903596 0.01953240
323                     Lexus.LX.470      15.95   2536 7.838343 0.01561382
405           Cadillac.Escalade.EXT       15.95   2667 7.888710 0.01859360
406         Chevrolet.Avalanche.1500      14.95   2575 7.853605 0.01648470
```

Cars2004 (subset, $n = 412$), consumption ~ log(weight)

Leverage points



Section **11.4**

Influential diagnostics

11.4 Influential diagnostics

- Both outliers and leverage points not necessarily a problem
- Problem if any of observations have “too high” influence on quantities of primary interest
- **Influential diagnostics** $\equiv$ quantification of how the LSE related quantities change if calculated using a dataset without a particular observation (leave-one-out diagnostics)

11.4 Influential diagnostics

Full model

$$\mathsf{M}:\ \boldsymbol{Y} \,\big|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\ \sigma^2 \mathbf{I}_n\big), \qquad \mathrm{rank}\big(\mathbb{X}_{n\times k}\big) = k$$

Leave-one-out model ($t = 1, \ldots, n$)

$$\mathsf{M}_{(-t)}:\ \boldsymbol{Y}_{(-t)} \,\big|\, \mathbb{X}_{(-t)} \sim \big(\mathbb{X}_{(-t)}\boldsymbol{\beta},\ \sigma^2 \mathbf{I}_{n-1}\big).$$

Assumption (for given t): $m_{t,t} > 0$

$$\Longrightarrow \mathrm{rank}\big(\mathbb{X}_{(-t)}\big) = \mathrm{rank}\big(\mathbb{X}\big) = k.$$

Influence measures

```
m1 <- lm(consumption ~ lweight, data = CarsUsed)
influence.measures(m1)
```

```
Influence measures of
         lm(formula = consumption ~ lweight, data = CarsUsed) :

       dfb.1_  dfb.lwgh     dffit cov.r   cook.d     hat inf
1    5.81e-02 -5.73e-02  0.064533 1.015 2.09e-03 0.01145   *
2    6.78e-02 -6.69e-02  0.074943 1.015 2.81e-03 0.01189   *
3   -1.71e-02  1.68e-02 -0.020491 1.012 2.10e-04 0.00744
4   -2.92e-02  2.86e-02 -0.035836 1.011 6.43e-04 0.00669
5   -1.71e-02  1.68e-02 -0.020491 1.012 2.10e-04 0.00744
...
```

```
summary(influence.measures(m1))
```

```
Potentially influential observations of
         lm(formula = consumption ~ lweight, data = CarsUsed) :

    dfb.1_ dfb.lwgh dffit   cov.r   cook.d hat
1    0.06  -0.06     0.06    1.01_*  0.00   0.01
2    0.07  -0.07     0.07    1.01_*  0.00   0.01
17   0.07  -0.07     0.07    1.01_*  0.00   0.01
39  -0.01   0.01    -0.01    1.02_*  0.00   0.01
47   0.07  -0.07     0.07    1.02_*  0.00   0.02_*
48   0.09  -0.09     0.10    1.02_*  0.00   0.02_*
...
```

11.4.1 DFBETAS

LSE's of $\boldsymbol{\beta}$ (rank$(\mathbb{X}) = $ rank$(\mathbb{X}_{(-t)}) = k$) in M and $\mathsf{M}_{(-t)}$

$$\mathsf{M}: \quad \widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_0, \ldots, \widehat{\beta}_{k-1}\right)^\top = \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top \boldsymbol{Y},$$

$$\mathsf{M}_{(-t)}: \quad \widehat{\boldsymbol{\beta}}_{(-t)} = \left(\widehat{\beta}_{(-t),0}, \ldots, \widehat{\beta}_{(-t),k-1}\right)^\top = \left(\mathbb{X}_{(-t)}{}^\top\mathbb{X}_{(-t)}\right)^{-1}\mathbb{X}_{(-t)}{}^\top \boldsymbol{Y}_{(-t)}.$$

Influence of the tth observation on the LSE of $\boldsymbol{\beta}$ (Lemma 11.3)

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(-t)} = \frac{U_t}{m_{t,t}}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\boldsymbol{x}_t$$

DFBETAS ($t = 1, \dots, n$, $j = 0, \dots, k-1$)

$$\mathrm{DFBETAS}_{t,j} := \frac{\widehat{\beta}_j - \widehat{\beta}_{(-t),j}}{\sqrt{\mathrm{MS}_{e,(-t)}\, v_{j,j}}} = \frac{U_t}{m_{t,t}\sqrt{\mathrm{MS}_{e,(-t)}\, v_{j,j}}}\, \boldsymbol{v}_t^\top \boldsymbol{x}_t$$

$$(\mathbb{X}^\top \mathbb{X})^{-1} = \begin{pmatrix} \boldsymbol{v}_0^\top \\ \vdots \\ \boldsymbol{v}_{k-1}^\top \end{pmatrix} = \begin{pmatrix} v_{0,0} & \cdots & v_{0,k-1} \\ \vdots & \vdots & \vdots \\ v_{k-1,0} & \cdots & v_{k-1,k-1} \end{pmatrix}$$

R function `influence.measures` rule-of-thumb

tth observation is influential with respect to the LSE of the jth regression coefficient if

$$\left|\mathrm{DFBETAS}_{t,j}\right| > 1.$$

DFBETAS

```
dfbetas(m1)
```

```
    (Intercept)      lweight
1    0.058079251 -0.057288572
2    0.067760218 -0.066859700
3   -0.017131716  0.016817978
4   -0.029182966  0.028603518
5   -0.017131716  0.016817978
6   -0.037145548  0.036495821
7   -0.045873896  0.045023905
8    0.007702297 -0.007562061
9   -0.021494294  0.021106330
10   0.009424138 -0.009254036
...
```

Maximal absolute values of DFBETAS for each regressor

```
apply(abs(dfbetas(m1)), 2, max)
```

```
(Intercept)     lweight
  0.7344821   0.7415123
```

11.4.2 DFFITS

LSE's of $\mu_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} = \mathbb{E}\big(Y_t \,\big|\, \boldsymbol{X}_t = \boldsymbol{x}_t\big)$ in M and $\mathsf{M}_{(-t)}$

$$\mathsf{M}: \quad \widehat{Y}_t = \boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top \mathbb{X}\big)^{-1} \mathbb{X}^\top \boldsymbol{Y},$$

$$\mathsf{M}_{(-t)}: \quad \widehat{Y}_{[t]} = \boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}}_{(-t)}, \qquad \widehat{\boldsymbol{\beta}}_{(-t)} = \big(\mathbb{X}_{(-t)}{}^\top \mathbb{X}_{(-t)}\big)^{-1} \mathbb{X}_{(-t)}{}^\top \boldsymbol{Y}_{(-t)}.$$

Expression of $\widehat{\boldsymbol{\beta}}_{(-t)}$ from Lemma 11.3

$$\widehat{Y}_{[t]} = \boldsymbol{x}_t^\top \left\{ \widehat{\boldsymbol{\beta}} - \frac{U_t}{m_{t,t}} \big(\mathbb{X}^\top \mathbb{X}\big)^{-1} \boldsymbol{x}_t \right\} = \widehat{Y}_t - \frac{U_t}{m_{t,t}} \boldsymbol{x}_t^\top \big(\mathbb{X}^\top \mathbb{X}\big)^{-1} \boldsymbol{x}_t$$

$$= \widehat{Y}_t - U_t \frac{h_{t,t}}{m_{t,t}}$$

Influence of the tth observation on the LSE of μ_t

$$\widehat{Y}_t - \widehat{Y}_{[t]} = U_t \frac{h_{t,t}}{m_{t,t}}$$

11.4.2 DFFITS

DFFITS ($t = 1, \ldots, n$)

$$\text{DFFITS}_t \;:=\; \frac{\widehat{Y}_t \;-\; \widehat{Y}_{[t]}}{\sqrt{\text{MS}_{e,(-t)}\, h_{t,t}}}$$

$$= \frac{h_{t,t}}{m_{t,t}}\, \frac{U_t}{\sqrt{\text{MS}_{e,(-t)}\, h_{t,t}}} \;=\; \sqrt{\frac{h_{t,t}}{m_{t,t}}}\, \frac{U_t}{\sqrt{\text{MS}_{e,(-t)}\, m_{t,t}}} \;=\; \sqrt{\frac{h_{t,t}}{m_{t,t}}}\; T_t$$

R function `influence.measures` rule-of-thumb

tth observation excessively influences the LSE of its expectation if

$$\left|\text{DFFITS}_t\right| > 3\sqrt{\frac{k}{n-k}}.$$

DFFITS

```
dffits(m1)
```

```
           1            2             3             4             5
 0.0645330957  0.0749431929 -0.0204914092 -0.0358359160 -0.0204914092  ...
```

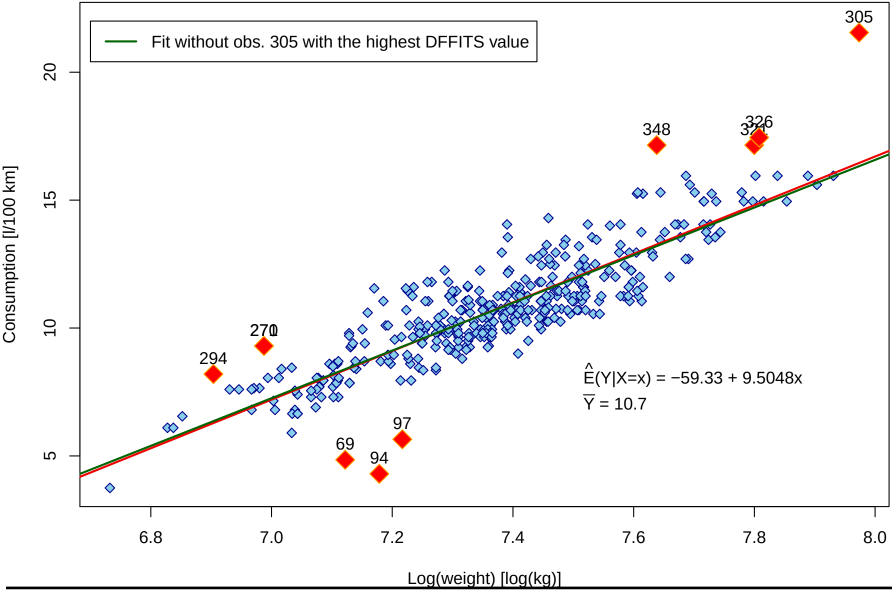
$$3\sqrt{\frac{k}{n-k}} = 0.2095$$

```
sum(abs(dffits(m1)) > 3 * sqrt(k / (n-k)))
```

```
[1] 10
```

```
                                  vname consumption weight  lweight     dffits
69               Honda.Civic.Hybrid.4dr        4.85   1239 7.122060 -0.2598440
                  manual.(gas/electric)
94       Toyota.Prius.4dr.(gas/electric)        4.30   1311 7.178545 -0.2984834
97          Volkswagen.Jetta.GLS.TDI.4dr        5.65   1362 7.216709 -0.2114462
270      Mazda.MX-5.Miata.convertible.2dr        9.30   1083 6.987490  0.2216790
271   Mazda.MX-5.Miata.LS.convertible.2dr        9.30   1083 6.987490  0.2216790
294     Toyota.MR2.Spyder.convertible.2dr        8.20    996 6.903747  0.2254823
305                             Hummer.H2       21.55   2903 7.973500  0.7815812
321            Land.Rover.Range.Rover.HSE       17.15   2440 7.799753  0.2597672
326                     Mercedes-Benz.G500       17.45   2460 7.807917  0.2892681
348               Land.Rover.Discovery.SE       17.15   2076 7.638198  0.3049335
```

Large DFFITS values



11.4.3 Cook distance

LSE's of $\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} = \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{X}\big)$ in M and $\mathsf{M}_{(-t)}$

$$\mathsf{M}: \quad \widehat{\boldsymbol{Y}} = \mathbb{X}\widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} = \big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{X}^\top\boldsymbol{Y},$$

$$\mathsf{M}_{(-t)}: \quad \widehat{Y}_{(-t\bullet)} = \mathbb{X}\widehat{\boldsymbol{\beta}}_{(-t)}, \qquad \widehat{\boldsymbol{\beta}}_{(-t)} = \big(\mathbb{X}_{(-t)}{}^\top\mathbb{X}_{(-t)}\big)^{-}\mathbb{X}_{(-t)}{}^\top\boldsymbol{Y}_{(-t)}.$$

Remind difference

$$\widehat{\boldsymbol{Y}}_{(-t\bullet)} = \mathbb{X}\widehat{\boldsymbol{\beta}}_{(-t)} = \begin{pmatrix} \boldsymbol{x}_1^\top\widehat{\boldsymbol{\beta}}_{(-t)} \\ \vdots \\ \boldsymbol{x}_n^\top\widehat{\boldsymbol{\beta}}_{(-t)} \end{pmatrix}, \qquad \widehat{\boldsymbol{Y}}_{[\bullet]} = \begin{pmatrix} \widehat{Y}_{[1]} \\ \vdots \\ \widehat{Y}_{[n]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^\top\widehat{\boldsymbol{\beta}}_{(-1)} \\ \vdots \\ \boldsymbol{x}_n^\top\widehat{\boldsymbol{\beta}}_{(-n)} \end{pmatrix},$$

$\widehat{\boldsymbol{Y}}_{(-t)} = \mathbb{X}_{(-t)}\widehat{\boldsymbol{\beta}}_{(-t)}$ is a subvector of length $n-1$ of a vector $\widehat{\boldsymbol{Y}}_{(-t\bullet)}$ of length n.

11.4.3 Cook distance

Influence of the tth observation on the LSE of μ

$$\left\|\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(-t\bullet)}\right\|^2 = \text{some calculations} = \frac{h_{t,t}}{m_{t,t}^2}\,U_t^2.$$

Cook distance ($t = 1, \ldots, n$)

$$D_t := \frac{1}{k\,\mathrm{MS}_e}\left\|\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(-t\bullet)}\right\|^2$$

$$= \frac{1}{k}\,\frac{h_{t,t}}{m_{t,t}}\,\frac{U_t^2}{\mathrm{MS}_e\,m_{t,t}} = \frac{1}{k}\,\frac{h_{t,t}}{m_{t,t}}\,\left(U_t^{std}\right)^2$$

- $0 < h_{t,t} = 1 - m_{t,t} < 1$,
  $h_{t,t}/m_{t,t}$ increases with $h_{t,t}$ and is high for leverage points.
- $\left(U_t^{std}\right)^2$ is high for outliers.

11.4.3 Cook distance

Remember

$$\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top \boldsymbol{Y},$$

$$\widehat{\boldsymbol{\beta}}_{(-t)} = \left(\mathbb{X}_{(-t)}{}^\top \mathbb{X}_{(-t)}\right)^{-1} \mathbb{X}_{(-t)}{}^\top \boldsymbol{Y}_{(-t)}.$$

Cook distance expressed differently ($t = 1, \ldots, n$)

$$D_t = \text{directly from definition} = \frac{\left(\widehat{\boldsymbol{\beta}}_{(-t)} - \widehat{\boldsymbol{\beta}}\right)^\top \mathbb{X}^\top \mathbb{X}\left(\widehat{\boldsymbol{\beta}}_{(-t)} - \widehat{\boldsymbol{\beta}}\right)}{k\,\mathsf{MS}_e}.$$

$1 - \alpha$ confidence region for $\boldsymbol{\beta}$ derived from model M while assuming normality

$$\mathcal{C}(\alpha) = \left\{\boldsymbol{\beta} : \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^\top \mathbb{X}^\top \mathbb{X}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) < k\,\mathsf{MS}_e\,\mathcal{F}_{k,n-k}(1-\alpha)\right\}.$$

11.4.3 Cook distance

Link between the Cook distance and the confidence region for β derived from model M

$$\widehat{\beta}_{(-t)} \in \mathcal{C}(\alpha) \qquad \text{if and only if} \qquad D_t < \mathcal{F}_{k,n-k}(1-\alpha).$$

R function `influence.measures` rule-of-thumb

tth observation excessively influences the LSE of the full response expectation μ if

$$D_t \; > \; \mathcal{F}_{k,n-k}(0.50).$$

Cook distance

```
cooks.distance(m1)
```

```
           1            2            3            4            5
0.0020855185 0.0028118990 0.0002104334 0.0006433764 0.0002104334   ...
```

$$\mathcal{F}_{k,n-k}(0.50) = 0.6943$$

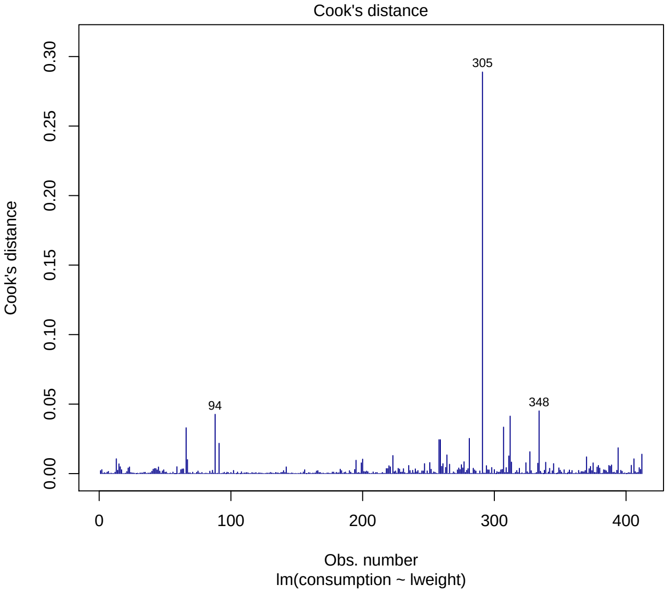
Maximal Cook distance

```
max(cooks.distance(m1))
```

```
[1] 0.288855
```

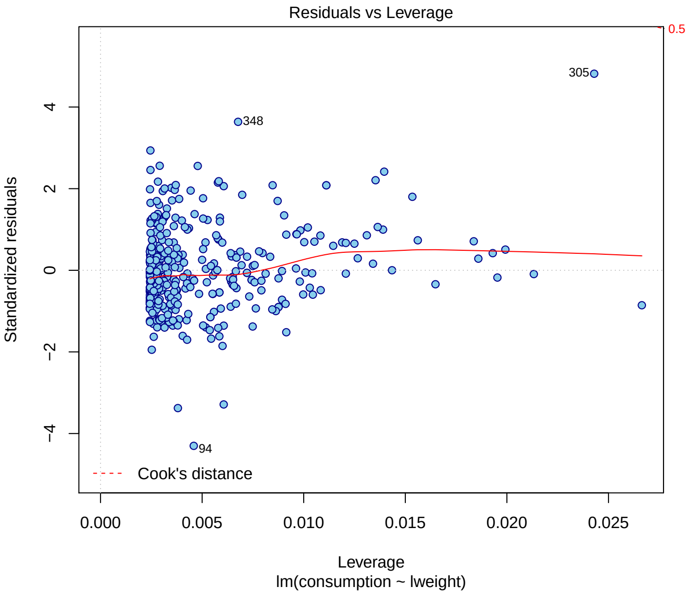
Cars2004 (subset, $n = 412$), `consumption ~ log(weight)`

diagnostic plot (`plot(m1, which = 4)`)



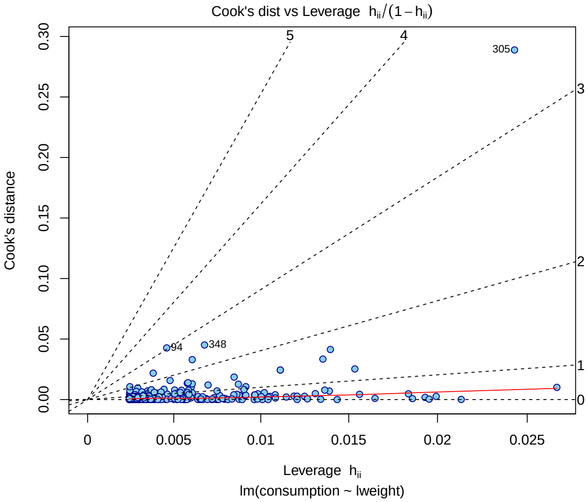
Cars2004 (subset, $n = 412$), `consumption ~ log(weight)`

R diagnostic plot (`plot(m1, which = 5)`)



Cars2004 (subset, $n = 412$), consumption $\sim$ log(weight)

diagnostic plot (plot(m1, which = 6))



The x-axis shows values of $h_{i,i}/(1 - h_{i,i})$ and not $h_{i,i}$.
Contours are related to the values of $U_t^{std}/\sqrt{k}$.

11.4.4 COVRATIO

LSE's of β $(\mathrm{rank}(\mathbb{X}) = \mathrm{rank}(\mathbb{X}_{(-t)}) = k)$ in M and $\mathrm{M}_{(-t)}$

$$\begin{aligned}
\mathrm{M}: \qquad \widehat{\boldsymbol{\beta}} &= \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top \boldsymbol{Y}, \\
\mathrm{M}_{(-t)}: \qquad \widehat{\boldsymbol{\beta}}_{(-t)} &= \left(\mathbb{X}_{(-t)}{}^\top\mathbb{X}_{(-t)}\right)^{-1}\mathbb{X}_{(-t)}{}^\top \boldsymbol{Y}_{(-t)}.
\end{aligned}$$

Estimated covariance matrices of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{(-t)}$

$$\begin{aligned}
\widehat{\mathrm{var}}\left(\widehat{\boldsymbol{\beta}} \,\middle|\, \mathbb{X}\right) &= \mathrm{MS}_e \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}, \\
\widehat{\mathrm{var}}\left(\widehat{\boldsymbol{\beta}}_{(-t)} \,\middle|\, \mathbb{X}\right) &= \mathrm{MS}_{e,(-t)} \left(\mathbb{X}_{(-t)}^\top\mathbb{X}_{(-t)}\right)^{-1}.
\end{aligned}$$

11.4.4 COVRATIO

Influence of the tth observation ($t = 1, \ldots, n$) on the precision of the LSE of the vector of regression coefficients

$$\text{COVRATIO}_t = \frac{\det\Bigl\{\widehat{\text{var}}\bigl(\widehat{\beta}_{(-t)} \,\big|\, \mathbb{X}\bigr)\Bigr\}}{\det\Bigl\{\widehat{\text{var}}\bigl(\widehat{\beta} \,\big|\, \mathbb{X}\bigr)\Bigr\}}$$

$$= \text{some calculations} = \frac{1}{m_{t,t}}\left\{\frac{n-k-\bigl(U_t^{std}\bigr)^2}{n-k-1}\right\}^k.$$

R function `influence.measures` rule-of-thumb

tth observation excessively influences the precision of the LSE of the regression coefficients if

$$\bigl|1 - \text{COVRATIO}_t\bigr| > 3\,\frac{k}{n-k}.$$

COVRATIO

COVRATIO

```
covratio(m1)
```

```
       1        2        3        4        5        6
1.014754 1.014674 1.012147 1.010719 1.012147 1.011724  ...
```

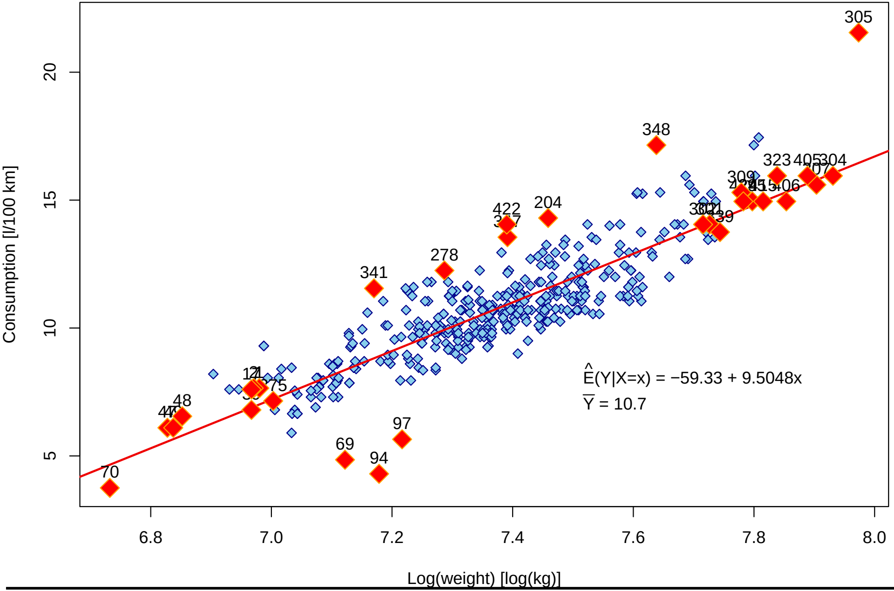
$3\,\frac{k}{n-k} = 0.0146$

```
sum(abs(1 - covratio(m1)) > 3 * (k / (n-k)))
```

```
[1] 31
```

```
                               vname consumption weight  lweight  covratio
1                 Chevrolet.Aveo.4dr        7.65   1075 6.980076 1.0147544
2        Chevrolet.Aveo.LS.4dr.hatch        7.65   1065 6.970730 1.0146741
17       Hyundai.Accent.GT.2dr.hatch        7.60   1061 6.966967 1.0149481
39                Scion.xA.4dr.hatch        6.80   1061 6.966967 1.0171433
47             Toyota.Echo.2dr.manual       6.10    923 6.827629 1.0240384
48               Toyota.Echo.2dr.auto       6.55    946 6.852243 1.0211810
49                    Toyota.Echo.4dr       6.10    932 6.837333 1.0237925
69                Honda.Civic.Hybrid        4.85   1239 7.122060 0.9584411
         .4dr.manual.(gas/electric)
70 Honda.Insight.2dr.(gas/electric)       3.75    839 6.732211 1.0287100
...
305                         Hummer.H2      21.55   2903 7.973500 0.9166531
...
```

COVRATIO value far from 1



11.4.5 Final remarks

- All presented influence measures should be used sensibly.
- Depending on what is the purpose of the modelling, different types of influence are differently harmful.
- There is certainly no need to panic if some observations are marked as “influential”!

12

Model Building

13

Analysis of Variance

Section **13.1**

One-way classification

13.1 One-way classification

One-way classified group means

$$m(g) = \mathbb{E}\big(Y \,\big|\, Z = g\big) =: m_g, \qquad g = 1, \ldots, G$$

Data sorted according to the value of Z

$$\begin{aligned}
Z_1 &= \cdots = Z_{n_1} &= 1,\\
Z_{n_1+1} &= \cdots = Z_{n_1+n_2} &= 2,\\
&\vdots\\
Z_{n_1+\cdots+n_{G-1}+1} &= \cdots = Z_n &= G.
\end{aligned}$$

Double subscript

$$\begin{aligned}
Z = 1: \quad \boldsymbol{Y}_1 &= \big(Y_{1,1}, \ldots, Y_{1,n_1}\big)^\top &&= \big(Y_1, \ldots, Y_{n_1}\big)^\top,\\
\vdots \qquad\quad &\vdots && \vdots\\
Z = G: \quad \boldsymbol{Y}_G &= \big(Y_{G,1}, \ldots, Y_{G,n_G}\big)^\top &&= \big(Y_{n_1+\cdots+n_{G-1}+1}, \ldots, Y_n\big)^\top.
\end{aligned}$$

13.1 One-way classification

Linear model

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_G \end{pmatrix}, \qquad \mathbb{E}(\boldsymbol{Y} \,|\, \mathbb{Z}) = \begin{pmatrix} m_1 \, \mathbf{1}_{n_1} \\ \vdots \\ m_G \, \mathbf{1}_{n_G} \end{pmatrix} =: \boldsymbol{\mu}, \quad \mathrm{var}(\boldsymbol{Y} \,|\, \mathbb{Z}) = \sigma^2 \, \mathbf{I}_n$$

Differences between the group means

$$\theta_{g,h} := m_g - m_h, \qquad g, h = 1, \ldots, G, \; g \neq h,$$

Principal null hypothesis to be tested

$$\mathrm{H}_0: \; m_1 = \cdots = m_G,$$

$$\mathrm{H}_0: \; \theta_{g,h} = 0, \quad g, h = 1, \ldots, G, \; g \neq h.$$

Factor effects

Definition 13.1 Factor effects in a one-way classification.

By *factor effects* in case of a one-way classification we understand the quantities $\eta_1, \ldots, \eta_G$ defined as

$$\eta_g = m_g - \overline{m}, \qquad g = 1, \ldots, G,$$

where $\overline{m} = \dfrac{1}{G}\sum_{h=1}^{G} m_h$ is the mean of the group means.

Principal null hypothesis to be tested

$$\mathrm{H}_0: \; m_1 = \cdots = m_G,$$

$$\mathrm{H}_0: \eta_g = 0, \quad g = 1, \ldots, G,$$

Regression space

$$\left\{ \begin{pmatrix} m_1 \, \mathbf{1}_{n_1} \\ \vdots \\ m_G \, \mathbf{1}_{n_G} \end{pmatrix} : \; m_1, \ldots, m_G \in \mathbb{R} \right\} \subseteq \mathbb{R}^n.$$

13.1.2 One-way ANOVA model

Full-rank parameterization

$$m_g \;=\; \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \qquad g = 1, \ldots, G$$

with $k = G$, $\boldsymbol{\beta} = \bigl(\beta_0, \underbrace{\boldsymbol{\beta}^Z}_{(\beta_1, \ldots, \beta_{G-1})^\top}\bigr)^\top$,

where $\mathbb{C} = \begin{pmatrix} \boldsymbol{c}_1^\top \\ \vdots \\ \boldsymbol{c}_G^\top \end{pmatrix}$ is a chosen $G \times (G-1)$ (pseudo)contrast matrix.

13.1.3 Least squares estimation

Lemma 13.1 Least squares estimation in one-way ANOVA linear model.

The fitted values and the LSE of the group means in a one-way ANOVA linear model are equal to the group sample means:

$$\widehat{m}_g = \widehat{Y}_{g,j} = \frac{1}{n_g}\sum_{l=1}^{n_g} Y_{g,l} =: \overline{Y}_{g\bullet}, \qquad g = 1, \ldots, G,\ j = 1, \ldots, n_g.$$

That is,

$$\widehat{\boldsymbol{m}} := \begin{pmatrix} \widehat{m}_1 \\ \vdots \\ \widehat{m}_G \end{pmatrix} = \begin{pmatrix} \overline{Y}_{1\bullet} \\ \vdots \\ \overline{Y}_{G\bullet} \end{pmatrix}, \qquad \widehat{\boldsymbol{Y}} = \begin{pmatrix} \overline{Y}_{1\bullet}\mathbf{1}_{n_1} \\ \vdots \\ \overline{Y}_{G\bullet}\mathbf{1}_{n_G} \end{pmatrix}.$$

If additionally normality is assumed, i.e., $\boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$, *where* $\boldsymbol{\mu} = \left(m_1 \mathbf{1}_{n_1}^\top, \ldots, m_G \mathbf{1}_{n_G}^\top\right)^\top$, *then* $\widehat{\boldsymbol{m}} \mid \mathbb{Z} \sim \mathcal{N}_G(\boldsymbol{m}, \sigma^2\, \mathbb{V})$, *where*

$$\mathbb{V} = \begin{pmatrix} \frac{1}{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{n_G} \end{pmatrix}.$$

13.1.3 Least squares estimation

LSE of regression coefficients and their linear combinations

Full-rank parameterization $m_g = \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z$, $\boldsymbol{\beta}^Z = \left(\beta_1, \ldots, \beta_{G-1}\right)^\top$

$$\boldsymbol{m} = \beta_0 \mathbf{1}_G + \mathbb{C}\boldsymbol{\beta}^Z$$

LSE of the differences between the group means

$$\widehat{\theta}_{g,h} = \overline{Y}_{g\bullet} - \overline{Y}_{h\bullet}, \qquad g,\, h = 1, \ldots, G$$

LSE of the factor effects

$$\widehat{\eta}_g = \overline{Y}_{g\bullet} - \frac{1}{G}\sum_{h=1}^{G} \overline{Y}_{h\bullet}, \qquad g = 1, \ldots, G$$

13.1.4 Within and between groups sums of squares...

Sums of squares

Overall sample mean

$$\overline{Y} = \frac{1}{n}\sum_{g=1}^{G}\sum_{j=1}^{n_g} Y_{g,j} = \frac{1}{n}\sum_{g=1}^{G} n_g \overline{Y}_{g\bullet}.$$

Within groups sum of squares (= residual sum of squares)

$$\mathrm{SS}_e = \left\| \boldsymbol{Y} - \widehat{\boldsymbol{Y}} \right\|^2 = \sum_{g=1}^{G}\sum_{j=1}^{n_g}\left(Y_{g,j} - \widehat{Y}_{g,j}\right)^2 = \sum_{g=1}^{G}\sum_{j=1}^{n_g}\left(Y_{g,j} - \overline{Y}_{g\bullet}\right)^2,$$

$$\nu_e = n - G,$$

Between groups sum of squares (= regression sum of squares)

$$\mathrm{SS}_R = \left\| \widehat{\boldsymbol{Y}} - \overline{Y}\mathbf{1}_n \right\|^2 = \sum_{g=1}^{G}\sum_{j=1}^{n_g}\left(\widehat{Y}_{g,j} - \overline{Y}\right)^2 = \sum_{g=1}^{G} n_g\left(\overline{Y}_{g\bullet} - \overline{Y}\right)^2,$$

$$\nu_R = G - 1.$$

One-way ANOVA F-test

Submodel $\boldsymbol{Y} \mid \mathbb{Z} \sim \mathcal{N}_n\bigl(\mathbf{1}_n\beta_0,\, \sigma^2\mathbf{I}_n\bigr) \quad \equiv \quad m_1 = \cdots = m_G$

$$\begin{aligned} \mathrm{SS}_e^0 &= \ldots \\ F &= \ldots \end{aligned}$$

One-way ANOVA table

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
Factor	$G-1$	SS_R	MS_R	F	p
Residual	$n-G$	SS_e	MS_e		

Section **13.2**

Two-way classification

13.2 Two-way classification

Two-way classified group means

$$m(g,\, h) = \mathbb{E}\big(Y \,\big|\, Z = g,\, W = h\big) =: m_{g,h},$$

$$g = 1, \ldots, G,\; h = 1, \ldots, H$$

Sample sizes

$$n = \sum_{g=1}^{G} \sum_{h=1}^{H} n_{g,h}$$

Assumption:

$n_{g,h} > 0$ (almost surely) for all $g = 1, \ldots, G$, $h = 1, \ldots, H$

13.2 Two-way classification

Covariate matrix and overall response vector

$$
\begin{pmatrix} Z_1 & W_1 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ Z_n & W_n \end{pmatrix}
=
\left(\begin{array}{cc}
Z_{1,1,1} & W_{1,1,1} \\
\vdots & \vdots \\
Z_{1,1,n_{1,1}} & W_{1,1,n_{1,1}} \\
\hdashline
\vdots & \vdots \\
\hdashline
Z_{G,1,1} & W_{G,1,1} \\
\vdots & \vdots \\
Z_{G,1,n_{G,1}} & W_{G,1,n_{G,1}} \\
\hdashline
\vdots & \vdots \\
\vdots & \vdots \\
\hdashline
Z_{1,H,1} & W_{1,H,1} \\
\vdots & \vdots \\
Z_{1,H,n_{1,H}} & W_{1,H,n_{1,H}} \\
\hdashline
\vdots & \vdots
\end{array}\right)
=
\left(\begin{array}{cc}
1 & 1 \\
\vdots & \vdots \\
1 & 1 \\
\hdashline
\vdots & \vdots \\
\hdashline
G & 1 \\
\vdots & \vdots \\
G & 1 \\
\hdashline
\vdots & \vdots \\
\vdots & \vdots \\
\hdashline
1 & H \\
\vdots & \vdots \\
1 & H \\
\hdashline
\vdots & \vdots
\end{array}\right),
\quad
\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ Y_n \end{pmatrix}
=
\left(\begin{array}{c}
Y_{1,1,1} \\
\vdots \\
Y_{1,1,n_{1,1}} \\
\hdashline
\vdots \\
\hdashline
Y_{G,1,1} \\
\vdots \\
Y_{G,1,n_{G,1}} \\
\hdashline
\vdots \\
\vdots \\
\hdashline
Y_{1,H,1} \\
\vdots \\
Y_{1,H,n_{1,H}} \\
\hdashline
\vdots
\end{array}\right)
$$

13.2 Two-way classification

Response random variables with $(Z, W)^\top = (g, h)^\top$

$$\mathbf{Y}_{g,h} = \left(Y_{g,h,1}, \ldots, Y_{g,h,n_{g,h}}\right)^\top$$

Overall response vector

$$\mathbf{Y} = \left(\mathbf{Y}_{1,1}^\top, \ldots, \mathbf{Y}_{G,1}^\top, \ \ldots, \ \mathbf{Y}_{1,H}^\top, \ldots, \mathbf{Y}_{G,H}^\top\right)^\top$$

Vector of two-way classified group means

$$\mathbf{m} = \left(m_{1,1}, \ldots, m_{G,1}, \ldots \ldots, m_{1,H}, \ldots, m_{G,H}\right)^\top$$

13.2 Two-way classification

Sample sizes by values of Z and W

$$n_{g\bullet} = \sum_{h=1}^{H} n_{g,h}, \quad g = 1, \ldots, G, \qquad n_{\bullet h} = \sum_{g=1}^{G} n_{g,h}, \quad h = 1, \ldots, H$$

Means of the group means

$$\overline{m} := \frac{1}{G \cdot H} \sum_{g=1}^{G} \sum_{h=1}^{H} m_{g,h},$$

$$\overline{m}_{g\bullet} := \frac{1}{H} \sum_{h=1}^{H} m_{g,h}, \qquad g = 1, \ldots, G,$$

$$\overline{m}_{\bullet h} := \frac{1}{G} \sum_{g=1}^{G} m_{g,h}, \qquad h = 1, \ldots, H$$

13.2 Two-way classification

Response variables

Z	W: 1	$\dots$	H
1	$\boldsymbol{Y}_{1,1} = \left(Y_{1,1,1}, \dots, Y_{1,1,n_{1,1}}\right)^\top$	$\vdots$	$\boldsymbol{Y}_{1,H} = \left(Y_{1,H,1}, \dots, Y_{1,H,n_{1,H}}\right)^\top$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
G	$\boldsymbol{Y}_{G,1} = \left(Y_{G,1,1}, \dots, Y_{G,1,n_{G,1}}\right)^\top$	$\vdots$	$\boldsymbol{Y}_{G,H} = \left(Y_{G,H,1}, \dots, Y_{G,H,n_{G,H}}\right)^\top$

Group means

Z	W: 1	$\dots$	H	$\bullet$
1	$m_{1,1}$	$\vdots$	$m_{1,H}$	$\overline{m}_{1\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
G	$m_{G,1}$	$\vdots$	$m_{G,H}$	$\overline{m}_{G\bullet}$
$\bullet$	$\overline{m}_{\bullet 1}$	$\dots$	$\overline{m}_{\bullet H}$	$\overline{m}$

Sample sizes

Z	W: 1	$\dots$	H	$\bullet$
1	$n_{1,1}$	$\vdots$	$n_{1,H}$	$n_{1\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
G	$n_{G,1}$	$\vdots$	$n_{G,H}$	$n_{G\bullet}$
$\bullet$	$n_{\bullet 1}$	$\dots$	$n_{\bullet H}$	n

Linear model

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_{1,1} \\ \vdots \\ \boldsymbol{Y}_{G,1} \\ \vdots \\ \boldsymbol{Y}_{1,H} \\ \vdots \\ \boldsymbol{Y}_{G,H} \end{pmatrix}, \qquad \mathbb{E}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z},\, \mathbb{W}\big) = \begin{pmatrix} m_{1,1}\,\mathbf{1}_{n_{1,1}} \\ \vdots \\ m_{G,1}\,\mathbf{1}_{n_{G,1}} \\ \vdots \\ m_{1,H}\,\mathbf{1}_{n_{1,H}} \\ \vdots \\ m_{G,H}\,\mathbf{1}_{n_{G,H}} \end{pmatrix} =: \boldsymbol{\mu}, \quad \mathsf{var}\big(\boldsymbol{Y} \,\big|\, \mathbb{Z},\, \mathbb{W}\big) = \sigma^2\,\mathbf{I}_n$$

13.2.1 Parameters of interest

The mean of the group means

$$\overline{m} = \frac{1}{G \cdot H} \sum_{g=1}^{G} \sum_{h=1}^{H} m_{g,h}$$

- **Designed experiment**: $\overline{m} =$ the mean outcome if the experiment is performed with all combinations of the input factors Z and W, each combination equally replicated
- $Y =$ industrial production: $\overline{m} =$ the mean production as if all combinations of inputs are equally often used in the production process

13.2.1 Parameters of interest

The means of the means by the first or the second factor

$\overline{m}_{1\bullet}, \ldots, \overline{m}_{G\bullet},$ and $\overline{m}_{\bullet 1}, \ldots, \overline{m}_{\bullet H}$

- **Designed experiment**: $\overline{m}_{g\bullet}$ = the mean outcome if we fix the factor Z on its level g and perform the experiment while setting the factor W to all possible levels (each equally replicated)
- Y = industrial production: $\overline{m}_{g\bullet}$ = the mean production as if the Z input is set to g but all possible values of the second input W are equally often used in the production process

13.2.1 Parameters of interest

Differences between the means of the means by the first or the second factor

$$\theta_{g_1,g_2\bullet} := \overline{m}_{g_1\bullet} - \overline{m}_{g_2\bullet}, \qquad g_1,\, g_2 = 1, \dots, G,\; g_1 \neq g_2,$$

$$\theta_{\bullet h_1,h_2} := \overline{m}_{\bullet h_1} - \overline{m}_{\bullet h_2}, \qquad h_1,\, h_2 = 1, \dots, H,\; h_1 \neq h_2$$

- **Designed experiment**: $\theta_{g_1,g_2\bullet}$ $(g_1 \neq g_2) =$ the mean difference between the outcome values if we fix the factor Z to its levels g_1 and g_2, repectively and perform the experiment while setting the factor W to all possible levels (each equally replicated)
- $Y =$ industrial production: $\theta_{g_1,g_2\bullet}$ $(g_1 \neq g_2) =$ difference between the mean productions with Z set to g_1 and g_2, respectively while using all possible values of the second input W equally often in the production process

13.2.1 Parameters of interest

Definition 13.2 Factor main effects in two-way classification.

Consider a two-way classification based on factors Z and W. By *main effects* of the factor Z, we understand quantities $\eta_1^Z, \ldots, \eta_G^Z$ defined as

$$\eta_g^Z := \overline{m}_{g\bullet} - \overline{m}, \qquad g = 1, \ldots, G.$$

By *main effects* of the factor W, we understand quantities $\eta_1^W, \ldots, \eta_H^W$ defined as

$$\eta_h^W := \overline{m}_{\bullet h} - \overline{m}, \qquad h = 1, \ldots, H.$$

Interaction model

Interaction model M_{ZW}: $\sim \mathsf{Z} + \mathsf{W} + \mathsf{Z}:\mathsf{W}$

$$
\begin{aligned}
m_{g,h} &= \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z + \boldsymbol{d}_h^\top \boldsymbol{\beta}^W + (\boldsymbol{d}_h^\top \otimes \boldsymbol{c}_g^\top)\boldsymbol{\beta}^{ZW},\\
&= \alpha_0 + \alpha_g^Z + \alpha_h^W + \alpha_{g,h}^{ZW},\\
&\qquad g = 1, \ldots, H,\; h = 1, \ldots, H.
\end{aligned}
$$

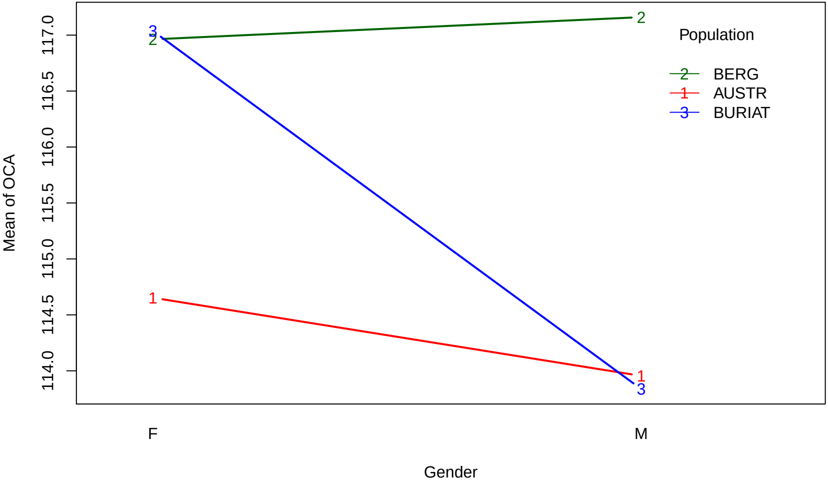
$$\text{Rank} = G \cdot H \qquad \text{if } n_{g,h} > 0 \text{ for all } (g,\ h).$$

Regression coefficients

$$
\begin{aligned}
&\beta_0, \quad \boldsymbol{\beta}^Z = \left(\beta_1^Z, \ldots, \beta_{G-1}^Z\right)^\top, \quad \boldsymbol{\beta}^W = \left(\beta_1^W, \ldots, \beta_{H-1}^W\right)^\top,\\
&\boldsymbol{\beta}^{ZW} = \left(\beta_{1,1}^{ZW}, \ldots, \beta_{G-1,1}^{ZW}, \;\ldots,\; \beta_{1,H-1}^{ZW}, \ldots, \beta_{G-1,H-1}^{ZW}\right)^\top\\
&\alpha_0 = \beta_0,\\
&\alpha_g^Z = \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \qquad g = 1, \ldots, G,\\
&\alpha_h^W = \boldsymbol{d}_h^\top \boldsymbol{\beta}^W, \qquad h = 1, \ldots, H,\\
&\alpha_{g,h}^{ZW} = (\boldsymbol{d}_h^\top \otimes \boldsymbol{c}_g^\top)\boldsymbol{\beta}^{ZW}, \qquad g = 1, \ldots, G,\; h = 1, \ldots, H.
\end{aligned}
$$

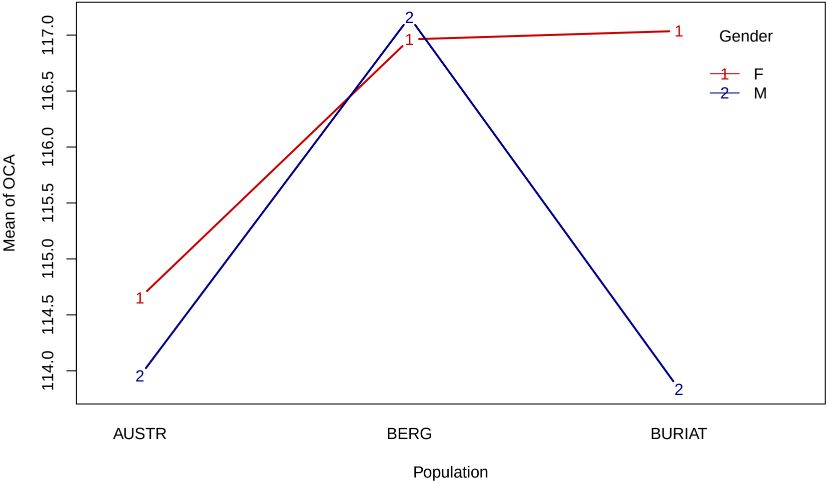
Howells ($n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$)



Howells ($n = 289$)

oca (occipital angle) $\sim$ gender ($G = 2$) and population ($H = 3$)



Additive model

Additive model M_{Z+W}: $\sim \mathrm{Z} + \mathrm{W}$

$$m_{g,h} = \alpha_0 + \alpha_g^Z + \alpha_h^W,$$

$$= \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z + \boldsymbol{d}_h^\top \boldsymbol{\beta}^W, \qquad g = 1, \dots, H,\ h = 1, \dots, H$$

Rank $= G + H - 1$ if $n_{g\bullet} > 0$ for all g and $n_{\bullet h} > 0$ for all h.

Additive model implies

- For each $g_1 \neq g_2$, $m_{g_1,h} - m_{g_2,h}$ does not depend on h,

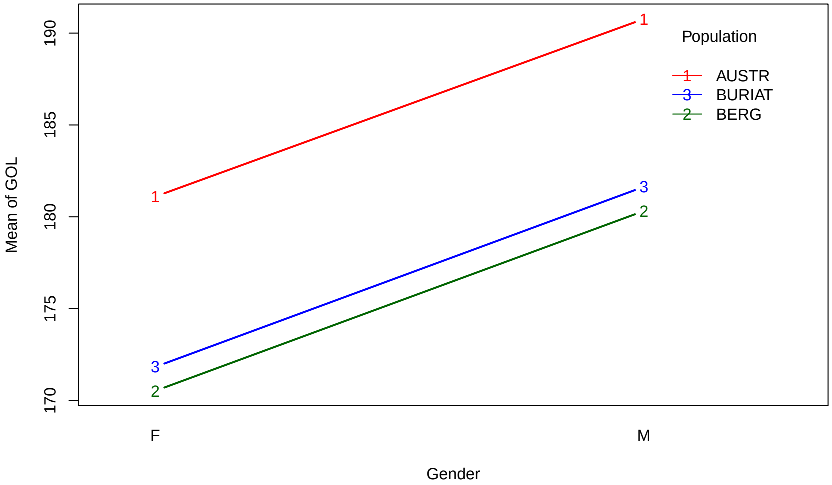
$$m_{g_1,h} - m_{g_2,h} = \overline{m}_{g_1\bullet} - \overline{m}_{g_2\bullet} = \eta_{g_1}^Z - \eta_{g_2}^Z = \theta_{g_1,g_2\bullet} = \alpha_{g_1}^Z - \alpha_{g_2}^Z$$
$$= \left(\boldsymbol{c}_{g_1} - \boldsymbol{c}_{g_2}\right)^\top \boldsymbol{\beta}^Z$$

- For each $h_1 \neq h_2$, $m_{g,h_1} - m_{g,h_2}$ does not depend on g,

$$m_{g,h_1} - m_{g,h_2} = \overline{m}_{\bullet h_1} - \overline{m}_{\bullet h_2} = \eta_{h_1}^W - \eta_{h_2}^W = \theta_{\bullet h_1,h_2} = \alpha_{h_1}^W - \alpha_{h_2}^W$$
$$= \left(\boldsymbol{d}_{h_1} - \boldsymbol{d}_{h_2}\right)^\top \boldsymbol{\beta}^W$$

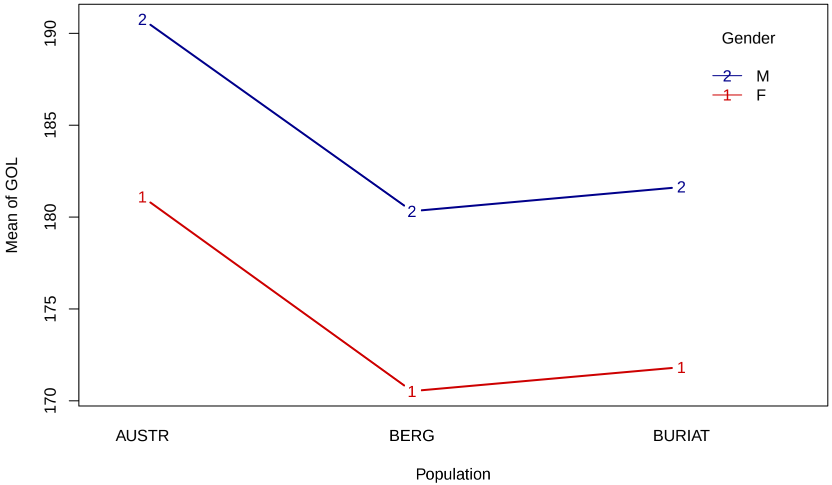
`Howells` ($n = 289$)

`gol` (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$)



Howells ($n = 289$)

gol (glabell-occipital length) $\sim$ gender ($G = 2$) and population ($H = 3$)



Model of effect of Z only

Model of effect of Z only M_Z: $\sim$ Z

$$m_{g,h} = \alpha_0 + \alpha_g^Z,$$

$$= \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \qquad g = 1, \dots, H,\ h = 1, \dots, H$$

$$\text{Rank} = G \qquad \text{if } n_{g\bullet} > 0 \text{ for all } g.$$

Model of effect of Z only implies

- For each $g = 1, \dots, G$ $\quad m_{g,1} = \cdots = m_{g,H} = \overline{m}_{g\bullet}$
- $\overline{m}_{\bullet 1} = \cdots = \overline{m}_{\bullet H}$

Model of effect of W only

Model of effect of W only $\mathrm{M}_W: \sim \mathrm{W}$

$$m_{g,h} = \alpha_0 + \alpha_h^W,$$

$$= \beta_0 + \boldsymbol{d}_h^\top \boldsymbol{\beta}^W, \qquad g = 1, \ldots, H, \; h = 1, \ldots, H$$

$$\text{Rank} = H \qquad \text{if } n_{\bullet h} > 0 \text{ for all } h.$$

Model of effect of W only implies

- For each $h = 1, \ldots, H$ $\quad m_{1,h} = \cdots = m_{G,h} = \overline{m}_{\bullet h}$
- $\overline{m}_{1\bullet} = \cdots = \overline{m}_{G\bullet}$

13.2.2 Two-way ANOVA models

Intercept only model

Intercept only model M_0 : ~ 1

$$
\begin{aligned}
m_{g,h} &= \alpha_0, \\
&= \beta_0, \qquad g = 1, \ldots, H,\ h = 1, \ldots, H
\end{aligned}
$$

$$\text{Rank} = 1 \qquad \text{if } n > 0.$$

Summary

Two-way ANOVA models

Model	Rank	Requirement for Rank
$\mathrm{M}_{ZW}: \sim \mathrm{Z}+\mathrm{W}+\mathrm{Z}:\mathrm{W}$	$G \cdot H$	$n_{g,h} > 0$ for all $g = 1, \ldots, G$, $h = 1, \ldots,$
$\mathrm{M}_{Z+W}: \sim \mathrm{Z}+\mathrm{W}$	$G+H-1$	$n_{g\bullet} > 0$ for all $g = 1, \ldots, G$, $n_{\bullet h} > 0$ for all $h = 1, \ldots, H$
$\mathrm{M}_{Z}: \sim \mathrm{Z}$	G	$n_{g\bullet} > 0$ for all $g = 1, \ldots, G$
$\mathrm{M}_{W}: \sim \mathrm{W}$	H	$n_{\bullet h} > 0$ for all $h = 1, \ldots, H$
$\mathrm{M}_{0}: \sim 1$	1	$n > 0$

13.2.3 Least squares estimation

Notation: Sample means in two-way classification

$$\overline{Y}_{g,h\bullet} := \frac{1}{n_{g,h}} \sum_{j=1}^{n_{g,h}} Y_{g,h,j}, \qquad g = 1, \ldots, G,\ h = 1, \ldots, H,$$

$$\overline{Y}_{g\bullet} := \frac{1}{n_{g\bullet}} \sum_{h=1}^{H} \sum_{j=1}^{n_{g,h}} Y_{g,h,j} = \frac{1}{n_{g\bullet}} \sum_{h=1}^{H} n_{g,h} \overline{Y}_{g,h\bullet}, \qquad g = 1, \ldots, G,$$

$$\overline{Y}_{\bullet h} := \frac{1}{n_{\bullet h}} \sum_{g=1}^{G} \sum_{j=1}^{n_{g,h}} Y_{g,h,j} = \frac{1}{n_{\bullet h}} \sum_{g=1}^{G} n_{g,h} \overline{Y}_{g,h\bullet}, \qquad h = 1, \ldots, H,$$

$$\overline{Y} := \frac{1}{n} \sum_{g=1}^{G} \sum_{h=1}^{H} \sum_{j=1}^{n_{g,h}} Y_{g,h,j} = \frac{1}{n} \sum_{g=1}^{G} n_{g\bullet} \overline{Y}_{g\bullet} = \frac{1}{n} \sum_{h=1}^{H} n_{\bullet h} \overline{Y}_{\bullet h}.$$

13.2.3 Least squares estimation

Lemma 13.2 Least squares estimation in two-way ANOVA linear models.

The fitted values and the LSE of the group means in two-way ANOVA linear models are given as follows (always for $g = 1, \ldots, G$, $h = 1, \ldots, H$, $j = 1, \ldots, n_{g,h}$).

(i) ***Interaction model*** $\mathsf{M}_{ZW}: \ \sim \mathsf{Z} + \mathsf{W} + \mathsf{Z}:\mathsf{W}$

$$\widehat{m}_{g,h} = \widehat{Y}_{g,h,j} = \overline{Y}_{g,h\bullet}.$$

(ii) ***Additive model*** $\mathsf{M}_{Z+W}: \ \sim \mathsf{Z} + \mathsf{W}$

$$\widehat{m}_{g,h} = \widehat{Y}_{g,h,j} = \overline{Y}_{g\bullet} + \overline{Y}_{\bullet h} - \overline{Y},$$

but only in case of balanced data ($n_{g,h} = J$ for all $g = 1, \ldots, G$, $h = 1, \ldots, H$).

TO BE CONTINUED.

Lemma 13.2 Least squares estimation in two-way ANOVA linear models, cont'd.

(iii) ***Model of effect of Z only*** $\mathsf{M}_Z: \;\sim \mathsf{Z}$

$$\widehat{m}_{g,h} = \widehat{Y}_{g,h,j} = \overline{Y}_{g\bullet}.$$

(iv) ***Model of effect of W only*** $\mathsf{M}_W: \;\sim \mathsf{W}$

$$\widehat{m}_{g,h} = \widehat{Y}_{g,h,j} = \overline{Y}_{\bullet h}.$$

(v) ***Intercept only model*** $\mathsf{M}_0: \;\sim 1$

$$\widehat{m}_{g,h} = \widehat{Y}_{g,h,j} = \overline{Y}.$$

Consequence of Lemma 13.2: LSE of the means of the means in the interaction and the additive model with balanced data.

With balanced data ($n_{g,h} = J$ for all $g = 1, \ldots, G$, $h = 1, \ldots, H$), the LSE of the means of the means by the first factor (parameters $\overline{m}_{1\bullet}, \ldots, \overline{m}_{G\bullet}$) or by the second factor (parameters $\overline{m}_{\bullet 1}, \ldots, \overline{m}_{\bullet H}$) satisfy in both the interaction and the additive two-way ANOVA linear models the following:

$$\widehat{\overline{m}}_{g\bullet} = \overline{Y}_{g\bullet}, \qquad g = 1, \ldots, G,$$
$$\widehat{\overline{m}}_{\bullet h} = \overline{Y}_{\bullet h}, \qquad h = 1, \ldots, H.$$

If additionally normality is assumed then $\widehat{\overline{\boldsymbol{m}}}^Z := \big(\widehat{\overline{m}}_{1\bullet}, \ldots, \widehat{\overline{m}}_{G\bullet}\big)^\top$ and $\widehat{\overline{\boldsymbol{m}}}^W := \big(\widehat{\overline{m}}_{\bullet 1}, \ldots, \widehat{\overline{m}}_{\bullet H}\big)^\top$ satisfy

$$\widehat{\overline{\boldsymbol{m}}}^Z \,|\, \mathbb{Z}, \mathbb{W} \sim \mathcal{N}_G\big(\overline{\boldsymbol{m}}^Z, \sigma^2\,\mathbb{V}^Z\big), \qquad \widehat{\overline{\boldsymbol{m}}}^W \,|\, \mathbb{Z}, \mathbb{W} \sim \mathcal{N}_H\big(\overline{\boldsymbol{m}}^W, \sigma^2\,\mathbb{V}^W\big),$$

Consequence of Lemma 13.2: LSE of the means of the means in the interaction and the additive model with balanced data, cont'd.

where

$$\overline{\boldsymbol{m}}^Z = \begin{pmatrix} \overline{m}_{1\bullet} \\ \vdots \\ \overline{m}_{G\bullet} \end{pmatrix}, \quad \mathbb{V}^Z = \begin{pmatrix} \frac{1}{JH} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{JH} \end{pmatrix},$$

$$\overline{\boldsymbol{m}}^W = \begin{pmatrix} \overline{m}_{\bullet 1} \\ \vdots \\ \overline{m}_{\bullet H} \end{pmatrix}, \quad \mathbb{V}^W = \begin{pmatrix} \frac{1}{JG} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{JG} \end{pmatrix}.$$

13.2.4 Sums of squares and ANOVA tables with balanced data

Sums of squares

With **balanced** data

$$\mathrm{SS}(\mathsf{Z}+\mathsf{W}+\mathsf{Z}:\mathsf{W}\,|\,\mathsf{Z}+\mathsf{W}) = \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{g,h\bullet} - \overline{Y}_{g\bullet} - \overline{Y}_{\bullet h} + \overline{Y}\right)^2,$$

$$\begin{aligned}
\mathrm{SS}(\mathsf{Z}+\mathsf{W}\,|\,\mathsf{W}) &= \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{g\bullet} + \overline{Y}_{\bullet h} - \overline{Y} - \overline{Y}_{\bullet h}\right)^2 = \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{g\bullet} - \overline{Y}\right)^2,\\
\mathrm{SS}(\mathsf{Z}+\mathsf{W}\,|\,\mathsf{Z}) &= \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{g\bullet} + \overline{Y}_{\bullet h} - \overline{Y} - \overline{Y}_{g\bullet}\right)^2 = \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{\bullet h} - \overline{Y}\right)^2,\\
\mathrm{SS}(\mathsf{Z}\,|\,1) &= \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{g\bullet} - \overline{Y}\right)^2,\\
\mathrm{SS}(\mathsf{W}\,|\,1) &= \sum_{g=1}^{G}\sum_{h=1}^{H} J\left(\overline{Y}_{\bullet h} - \overline{Y}\right)^2
\end{aligned}$$

13.2.4 Sums of squares and ANOVA tables with balanced data

Sums of squares

Notation: Sums of squares in two-way classification

For balanced data

$$SS_Z := \sum_{g=1}^{G}\sum_{h=1}^{H} J\big(\overline{Y}_{g\bullet} - \overline{Y}\big)^2,$$

$$SS_W := \sum_{g=1}^{G}\sum_{h=1}^{H} J\big(\overline{Y}_{\bullet h} - \overline{Y}\big)^2,$$

$$SS_{ZW} := \sum_{g=1}^{G}\sum_{h=1}^{H} J\big(\overline{Y}_{g,h\bullet} - \overline{Y}_{g\bullet} - \overline{Y}_{\bullet h} + \overline{Y}\big)^2,$$

$$SS_T := \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{j=1}^{J} \big(Y_{g,h,j} - \overline{Y}\big)^2,$$

$$SS_e^{ZW} := \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{j=1}^{J} \big(Y_{g,h,j} - \overline{Y}_{g,h\bullet}\big)^2.$$

13.2.4 Sums of squares and ANOVA tables with balanced data

Lemma 13.3 Breakdown of the total sum of squares in a balanced two-way classification.

In case of a balanced two-way classification, the following identity holds

$$\mathrm{SS}_T = \mathrm{SS}_Z + \mathrm{SS}_W + \mathrm{SS}_{ZW} + \mathrm{SS}_e^{ZW}.$$

ANOVA tables

Type I as well as type II ANOVA table for two-way classification with **balanced** data

Effect (Term)	Degrees of freedom	Effect sum of squares	Effect mean square	F-stat.	P-value
Z	$G - 1$	SS_Z	$\star$	$\star$	$\star$
W	$H - 1$	SS_W	$\star$	$\star$	$\star$
Z : W	$GH - G - H + 1$	SS_{ZW}	$\star$	$\star$	$\star$
Residual	$n - GH$	SS_e^{ZW}	$\star$		

Section **13.3**

Higher-way classification

14

Simultaneous Inference in a Linear Model

Section **14.1**

Basic simultaneous inference

14.1 Basic simultaneous inference

Matrix $\mathbb{L}_{m\times k}$: $m \leq k$;

its rows – vectors $\mathbf{l}_1, \ldots, \mathbf{l}_m \in \mathbb{R}^k$ linear. independent

Confidence region for $\boldsymbol{\theta}$ with a coverage of $1-\alpha$, $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}} =$ LSE of $\boldsymbol{\theta}$

$$\left\{\boldsymbol{\theta} \in \mathbb{R}^m : \; \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)^\top \left\{\mathsf{MS}_e\, \mathbb{L}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{L}^\top\right\}^{-1} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \; < \; m\,\mathcal{F}_{m,n-k}(1-\alpha)\right\}$$

Test of H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}^0$

$$Q_0 = \frac{1}{m}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^\top \left\{\mathsf{MS}_e\, \mathbb{L}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{L}^\top\right\}^{-1} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)$$

$$C(\alpha) = \left[\mathcal{F}_{m,n-k}(1-\alpha),\; \infty\right)$$

P-value if $Q_0 = q_0$: $p = 1 - \mathrm{CDF}_{\mathcal{F},\,m,\,n-k}(q_0)$

Section **14.2**

Multiple comparison procedures

14.2.1 Multiple testing

Definition 14.1 Multiple testing problem, elementary null hypotheses, global null hypothesis.

A testing problem with the null hypothesis

$$\mathsf{H}_0: \quad \theta_1 = \theta_1^0 \quad \& \quad \dots \quad \& \quad \theta_m = \theta_m^0,$$

is called the *multiple testing problem* with the m *elementary hypotheses*

$$\mathsf{H}_1: \quad \theta_1 = \theta_1^0, \quad \dots, \quad \mathsf{H}_m: \quad \theta_m = \theta_m^0.$$

Hypothesis H_0 is called in this context also as a *global null hypothesis*.

Notation

$\mathsf{H}_0 \equiv \mathsf{H}_1 \; \& \; \dots \; \& \; \mathsf{H}_m$ or $\mathsf{H}_0 \equiv \mathsf{H}_1, \dots, \mathsf{H}_m$ or $\mathsf{H}_0 = \bigcap_{j=1}^{m} \mathsf{H}_j$

Example. Multiple testing problem for one-way classified group means

One-way classified group means,

$\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$, $\boldsymbol{\beta} = \left(\beta_0,\, \boldsymbol{\beta}^{Z\top}\right)^\top$

- One categorical covariate $Z \in \mathcal{Z} = \{1, \ldots, G\}$.
- $\mathbb{X} \equiv n \times G$ model matrix derived from a (pseudo)contrast parameterization $\mathbb{C}$ ($G \times (G-1)$ matrix) of Z.
- $m_g := \mathbb{E}\big(Y \,|\, Z = g\big) = \beta_0 + \boldsymbol{c}_g^\top \boldsymbol{\beta}^Z, \quad g = 1, \ldots, G.$
- $\mathrm{H}_0: m_1 = \cdots = m_G$

  $\mathrm{H}_{1,2}: m_1 - m_2 = 0, \quad \ldots, \quad \mathrm{H}_{G-1,G}: m_{G-1} - m_G = 0$

  $\mathrm{H}_{1,2}: \theta_{1,2} = 0, \quad \ldots, \quad \mathrm{H}_{G-1,G}: \theta_{G-1,G} = 0$

$$\theta_{g,h} = m_g - m_h = \left(\boldsymbol{c}_g - \boldsymbol{c}_h\right)^\top \boldsymbol{\beta}^Z, \qquad g = 1, \ldots, G-1,\; h = g+1, \ldots, G$$

14.2.2 Simultaneous confidence intervals

Suppose that a distribution of the random vector $\boldsymbol{D}$ depends on a (vector) parameter $\boldsymbol{\theta} = \big(\theta_1, \ldots, \theta_m\big)^\top \in \Theta_1 \times \cdots \times \Theta_m = \Theta \subseteq \mathbb{R}^m$.

Definition 14.2 Simultaneous confidence intervals.

(Random) intervals $\big(\theta_j^L, \theta_j^U\big)$, $j = 1, \ldots, m$, where $\theta_j^L = \theta_j^L(\boldsymbol{D})$ and $\theta_j^U = \theta_j^U(\boldsymbol{D})$, $j = 1, \ldots, m$, are called *simultaneous confidence intervals* for parameter $\boldsymbol{\theta}$ with a coverage of $1 - \alpha$ if for any $\boldsymbol{\theta}^0 = \big(\theta_1^0, \ldots, \theta_m^0\big)^\top \in \Theta$,

$$\mathsf{P}\Big(\big(\theta_1^L, \theta_1^U\big) \times \cdots \times \big(\theta_m^L, \theta_m^U\big) \ni \boldsymbol{\theta}^0;\ \boldsymbol{\theta} = \boldsymbol{\theta}^0\Big) \geq 1 - \alpha.$$

14.2.2 Simultaneous confidence intervals

Example. Bonferroni simultaneous confidence intervals

- For each $j = 1, \ldots, m$, $\quad \big(\theta_j^L,\, \theta_j^U\big)$:
  a classical confidence interval for θ_j with a coverage of $1 - \frac{\alpha}{m}$

$$\forall\, j = 1, \ldots, m,\ \forall\, \theta_j^0 \in \Theta_j : \quad \mathsf{P}\Big(\big(\theta_j^L,\, \theta_j^U\big) \ni \theta_j^0;\, \theta_j = \theta_j^0\Big) \ \geq\ 1 - \frac{\alpha}{m}.$$

- $\forall\, j = 1, \ldots, m,\ \forall\, \theta_j^0 \in \Theta_j : \quad \mathsf{P}\Big(\big(\theta_j^L,\, \theta_j^U\big) \not\ni \theta_j^0;\, \theta_j = \theta_j^0\Big) \ \leq\ \dfrac{\alpha}{m}$
- For any $\boldsymbol{\theta}^0 \in \Theta$

$$\mathsf{P}\Big(\exists\, j = 1, \ldots, m : \quad \big(\theta_j^L,\, \theta_j^U\big) \not\ni \theta_j^0;\, \boldsymbol{\theta} = \boldsymbol{\theta}^0\Big)$$

$$\leq \sum_{j=1}^{m} \mathsf{P}\Big(\big(\theta_j^L,\, \theta_j^U\big) \not\ni \theta_j^0;\, \boldsymbol{\theta} = \boldsymbol{\theta}^0\Big) \leq \sum_{j=1}^{m} \frac{\alpha}{m} \ =\ \alpha.$$

14.2.3 Multiple comparison procedure, P-values adjusted for multiple comparison

Let for each $0 < \alpha < 1$ a procedure be given to construct the simultaneous confidence intervals $\bigl(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\bigr)$, $j = 1, \ldots, m$, for parameter $\boldsymbol{\theta}$ with a coverage of $1 - \alpha$. Let for each $j = 1, \ldots, m$, the procedure creates intervals satisfying a monotonicity condition

$$1 - \alpha_1 < 1 - \alpha_2 \qquad \Longrightarrow \qquad \bigl(\theta_j^L(\alpha_1),\, \theta_j^U(\alpha_1)\bigr) \subseteq \bigl(\theta_j^L(\alpha_2),\, \theta_j^U(\alpha_2)\bigr).$$

Definition 14.3 Multiple comparison procedure.

Multiple comparison procedure (MCP) for a multiple testing problem with the elementary null hypotheses $\mathsf{H}_j\colon\ \theta_j = \theta_j^0$, $j = 1, \ldots, m$, based on given procedure for construction of simultaneous confidence intervals for parameter $\boldsymbol{\theta}$ is the testing procedure that for given $0 < \alpha < 1$

(i) rejects the global null hypothesis $\mathsf{H}_0\colon\ \boldsymbol{\theta} = \boldsymbol{\theta}^0$ if and only if

$$\bigl(\theta_1^L(\alpha),\, \theta_1^U(\alpha)\bigr) \times \cdots \times \bigl(\theta_m^L(\alpha),\, \theta_m^U(\alpha)\bigr) \not\ni \boldsymbol{\theta}^0;$$

(ii) for $j = 1, \ldots, m$, rejects the jth elementary hypothesis $\mathsf{H}_j\ :\ \theta_j = \theta_j^0$ if and only if

$$\bigl(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\bigr) \not\ni \theta_j^0.$$

14.2.3 Multiple comparison procedure, P-values adjusted for multiple comparison

Definition 14.4 P-values adjusted for multiple comparison.

P-values adjusted for multiple comparison for a multiple testing problem with the elementary null hypotheses $\mathsf{H}_j:\ \theta_j = \theta_j^0,\ j = 1,\ldots,m$, based on given procedure for construction of simultaneous confidence intervals for parameter $\boldsymbol{\theta}$ are values $p_1^{adj},\ldots,p_m^{adj}$ defined as

$$p_j^{adj} = \inf\Big\{\alpha:\ \big(\theta_j^L(\alpha),\,\theta_j^U(\alpha)\big) \not\ni \theta_j^0\Big\},\qquad j = 1,\ldots,m.$$

For given α, $0 < \alpha < 1$

- MCP rejects $\mathsf{H}_j:\ \theta_j = \theta_j^0$ ($j = 1,\ldots,m$) if and only if $p_j^{adj} \le \alpha$.
- MCP rejects $\mathsf{H}_0:\ \boldsymbol{\theta} = \boldsymbol{\theta}^0$

$\equiv$ at least one elementary hypothesis rejected

$$\equiv \min\big\{p_1^{adj},\ldots,p_m^{adj}\big\} \le \alpha$$

$\Longrightarrow$ P-value of the test of H_0: $p^{adj} := \min\big\{p_1^{adj},\ldots,p_m^{adj}\big\}$

14.2.3 Multiple comparison procedure, P-values adjusted for multiple comparison

Example. Bonferroni MCP, Bonferroni adjusted P-values

For given α, $0 < \alpha < 1$

- For each $j = 1, \ldots, m$, $\quad \big(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\big)$:
  a classical confidence interval for θ_j with a coverage of $1 - \frac{\alpha}{m}$

$$\forall\, j = 1, \ldots, m,\ \forall\, \theta_j^0 \in \Theta_j: \quad \mathsf{P}\Big(\big(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\big) \ni \theta_j^0;\, \theta_j = \theta_j^0\Big) \ \geq\ 1 - \frac{\alpha}{m}.$$

  $\equiv$ Bonferroni simultaneous confidence intervals for $\boldsymbol{\theta}$

  with a coverage of $1 - \alpha$

- For $j = 1, \ldots, m$, p_j^{uni}: a P-value related to the (single) test of the (jth elementary) hypothesis $\mathrm{H}_j: \ \theta_j = \theta_j^0$ being dual to the confidence interval $\big(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\big)$

$$p_j^{uni} = \inf\left\{\frac{\alpha}{m}:\ \big(\theta_j^L(\alpha),\, \theta_j^U(\alpha)\big) \not\ni \theta_j^0\right\}.$$

14.2.4 Bonferroni simultaneous inference in a normal linear model

$\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n), \qquad \mathrm{rank}(\mathbb{X}_{n\times k}) = k < n$

Linear comb. of regr. param.:	$\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta} = \left(\mathbf{l}_1^\top\boldsymbol{\beta},\, \ldots,\, \mathbf{l}_m^\top\boldsymbol{\beta}\right)^\top = \left(\theta_1,\, \ldots,\, \theta_m\right)^\top$
LSE:	$\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}} = \left(\mathbf{l}_1^\top\widehat{\boldsymbol{\beta}},\, \ldots,\, \mathbf{l}_m^\top\widehat{\boldsymbol{\beta}}\right)^\top = \left(\widehat{\theta}_1,\, \ldots,\, \widehat{\theta}_m\right)^\top$
Residual mean square:	MS_e

Bonferroni simultaneous confidence intervals (coverage $1-\alpha$)

$$\theta_j^L(\alpha) = \mathbf{l}_j^\top\widehat{\boldsymbol{\beta}} - \sqrt{\mathsf{MS}_e\, \mathbf{l}_j^\top \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbf{l}_j}\;\; \mathrm{t}_{n-k}\!\left(1 - \frac{\alpha}{2\,m}\right),$$

$$\theta_j^U(\alpha) = \mathbf{l}_j^\top\widehat{\boldsymbol{\beta}} + \sqrt{\mathsf{MS}_e\, \mathbf{l}_j^\top \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbf{l}_j}\;\; \mathrm{t}_{n-k}\!\left(1 - \frac{\alpha}{2\,m}\right), \qquad j = 1, \ldots, m.$$

Bonferroni adjusted P-values, $\mathrm{H}_j:\ \theta_j = \theta_j^0,\ j = 1, \ldots, m$

$$p_j^B = \min\left\{2\,m\,\mathrm{CDF}_{t,\,n-k}\big(-\,|t_{j,0}|\big),\ 1\right\}, \qquad j = 1, \ldots, m,$$

$$t_{j,0} = \frac{\mathbf{l}_j^\top\widehat{\boldsymbol{\beta}} - \theta_j^0}{\sqrt{\mathsf{MS}_e\, \mathbf{l}_j^\top \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbf{l}_j}}$$

Section **14.3**

Tukey's T-procedure

Lemma 14.1 Studentized range.

Let $T_1, \ldots, T_m$ *be a random sample from* $\mathcal{N}(\mu, \sigma^2)$, $\sigma^2 > 0$. *Let*

$$R = \max_{j=1,\ldots,m} T_j - \min_{j=1,\ldots,m} T_j$$

be the range of the sample. Let S^2 *be the estimator of* σ^2 *such that* S^2 *and* $\boldsymbol{T} = (T_1, \ldots, T_m)^\top$ *are independent and*

$$\frac{\nu\, S^2}{\sigma^2} \sim \chi^2_\nu \quad \textit{for some} \quad \nu > 0.$$

Let

$$Q = \frac{R}{S}.$$

The distribution of the random variable Q *then depends on neither* μ, *nor* σ.

14.3.1 Tukey's pairwise comparisons theorem

Definition 14.5 Studentized range.

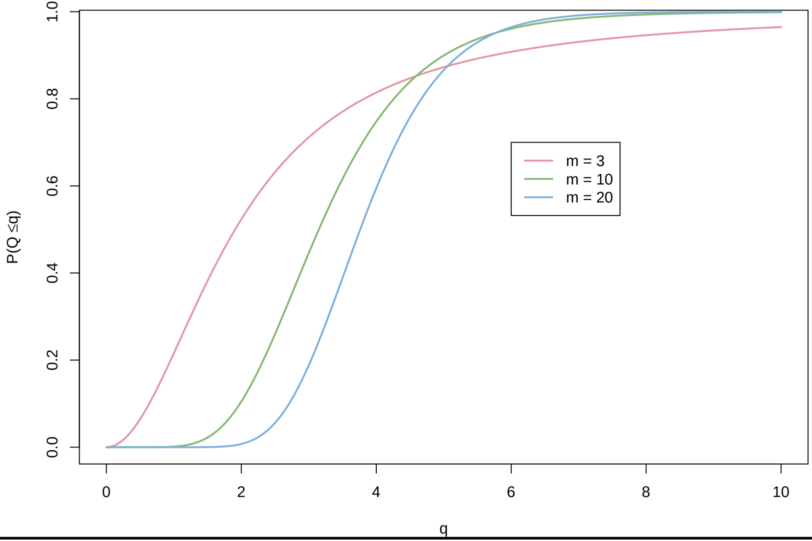
The random variable $Q = \dfrac{R}{S}$ from Lemma 14.1 will be called *studentized range* of a sample of size m with ν degrees of freedom and its distribution will be denoted as $\mathsf{q}_{m,\nu}$.

Notation.

- For $0 < p < 1$, the $p\,100\%$ quantile of the random variable Q with distribution $\mathsf{q}_{m,\nu}$ will be denoted as $\mathsf{q}_{m,\nu}(p)$.
- The distribution function of the random variable Q with distribution $\mathsf{q}_{m,\nu}$ will be denoted $\mathrm{CDF}_{\mathsf{q},m,\nu}(\cdot)$.

Studentized range: distribution functions

For $m = 3, 10, 20$ and $\nu = m - 1$, R: `ptukey(q, m, nu)`



Studentized range: selected quantiles

For $m = 3, 10, 20$ and $\nu = m - 1$, R: `qtukey(p, m, nu)`

```
p <- c(0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975)
quants <- data.frame(p = p,
                     q3  = round(qtukey(p, 3, 2), 4),
                     q10 = round(qtukey(p, 10, 9), 4),
                     q20 = round(qtukey(p, 20, 19), 4))
colnames(quants) <- c("p", paste("m = ", c(3, 10, 20), sep = ""))
print(quants)
```

```
      p    m = 3   m = 10   m = 20
1 0.025   0.3050   1.5291   2.2698
2 0.050   0.4370   1.7270   2.4650
3 0.100   0.6351   1.9800   2.7087
4 0.250   1.1007   2.4726   3.1664
5 0.500   1.9082   3.1494   3.7626
6 0.750   3.3080   4.0107   4.4724
7 0.900   5.7326   5.0067   5.2315
8 0.950   8.3308   5.7384   5.7518
9 0.975  11.9365   6.4790   6.2498
```

Theorem 14.2 Tukey's pairwise comparisons theorem, balanced version.

Let $T_1, \dots, T_m$ *be independent random variables and let* $T_j \sim \mathcal{N}(\mu_j,\ v^2\sigma^2)$, $j = 1, \dots, m$, *where* $v^2 > 0$ *is a known constant. Let* S^2 *be the estimator of* σ^2 *such that* S^2 *and* $\boldsymbol{T} = \big(T_1, \dots, T_m\big)^\top$ *are independent and*

$$\frac{\nu\, S^2}{\sigma^2} \sim \chi^2_\nu \quad \text{for some} \quad \nu > 0.$$

Then

$$\mathrm{P}\Big(\text{for all } j \neq l\text{:} \quad \big|T_j - T_l - (\mu_j - \mu_l)\big| < \mathrm{q}_{m,\nu}(1-\alpha)\sqrt{v^2\ S^2}\Big) = 1 - \alpha.$$

14.3.1 Tukey's pairwise comparisons theorem

Theorem 14.3 Tukey's pairwise comparisons theorem, general version.

Let $T_1, \dots, T_m$ *be independent random variables and let* $T_j \sim \mathcal{N}(\mu_j,\ v_j^2\sigma^2)$, $j = 1, \dots, m$, *where* $v_j^2 > 0$, $j = 1, \dots, m$ *are known constants. Let* S^2 *be the estimator of* σ^2 *such that* S^2 *and* $\boldsymbol{T} = (T_1, \dots, T_m)^\top$ *are independent and*

$$\frac{\nu\, S^2}{\sigma^2} \sim \chi^2_\nu \quad \textit{for some} \quad \nu > 0.$$

Then

$$\mathsf{P}\left(\textit{for all } j \neq l \quad \big|T_j - T_l - (\mu_j - \mu_l)\big| < \mathrm{q}_{m,\nu}(1-\alpha)\sqrt{\frac{v_j^2 + v_l^2}{2}\, S^2}\right) \geq 1 - \alpha.$$

Proof. See Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *The Annals of Statistics*, **12**(1), 61–75. ❑

14.3.2 Tukey's honest significance differences (HSD)

Assumptions

$\boldsymbol{T} = \big(T_1, \ldots, T_m\big)^\top \sim \mathcal{N}_m(\boldsymbol{\mu}, \sigma^2\,\mathbb{V})$

- $\boldsymbol{\mu} = \big(\mu_1, \ldots, \mu_m\big)^\top \in \mathbb{R}^m$, $\sigma^2 > 0$: unknown parameters;
- $\mathbb{V} = \mathrm{diag}\big(v_1^2, \ldots, v_m^2\big)$: known diagonal matrix.

S^2: estimator of σ^2,

- S^2 and $\boldsymbol{T}$ independent;
- $\nu\, S^2/\sigma^2 \sim \chi^2_\nu$ for some $\nu > 0$.

Multiple comparison problem

$$\theta_{j,l} = \mu_j - \mu_l, \qquad j = 1, \ldots, m-1,\ l = j+1, \ldots, m,$$
$$\boldsymbol{\theta} = \big(\theta_{1,2}, \theta_{1,3}, \ldots, \theta_{m-1,m}\big)^\top$$

$m^\star = \binom{m}{2}$ elementary hypotheses

$$\mathrm{H}_{j,l}:\quad \theta_{j,l} = \theta^0_{j,l}, \qquad j = 1, \ldots, m-1,\ l = j+1, \ldots, m,$$

for some $\boldsymbol{\theta}^0 = \big(\theta^0_{1,2}, \theta^0_{1,3}, \ldots, \theta^0_{m-1,m}\big)^\top \in \mathbb{R}^{m^\star}$.

14.3.2 Tukey's honest significance differences (HSD)

Theorem 14.4 Tukey's honest significance differences.

Random intervals given by

$$\theta^{TL}_{j,l}(\alpha) = T_j - T_l - \mathrm{q}_{m,\nu}(1-\alpha)\sqrt{\frac{v_j^2+v_l^2}{2}\,S^2},$$

$$\theta^{TU}_{j,l}(\alpha) = T_j - T_l + \mathrm{q}_{m,\nu}(1-\alpha)\sqrt{\frac{v_j^2+v_l^2}{2}\,S^2}, \qquad j<l.$$

are simultaneous confidence intervals for parameters $\theta_{j,l} = \mu_j - \mu_l$, $j = 1,\dots,m-1$, $l = j+1,\dots,m$ *with a coverage of* $1-\alpha$.

In the balanced case of $v_1^2 = \cdots = v_m^2$, *the coverage is exactly equal to* $1-\alpha$, *i.e., for any* $\boldsymbol{\theta}^0 \in \mathbb{R}^{m^\star}$

$$\mathrm{P}\Big(\text{for all } j \neq l \quad \big(\theta^{TL}_{j,l}(\alpha),\ \theta^{TU}_{j,l}(\alpha)\big) \ni \theta^0_{j,l};\ \boldsymbol{\theta} = \boldsymbol{\theta}^0\Big) = 1-\alpha.$$

Related P-values for a multiple testing problem with elementary hypotheses $H_{j,l}:\ \theta_{j,l} = \theta^0_{j,l}$, $\theta^0_{j,l} \in \mathbb{R}$, $j<l$, *adjusted for multiple comparison are given by*

$$p^T_{j,l} = 1 - \mathrm{CDF}_{\mathrm{q},m,\nu}\Big(\big|t^0_{j,l}\big|\Big), \qquad j<l,$$

where $t^0_{j,l}$ *is a value of* $T_{j,l}(\theta^0_{j,l}) = \dfrac{T_j - T_l - \theta^0_{j,l}}{\sqrt{\frac{v_j^2+v_l^2}{2}\,S^2}}$ *attained with given data.*

14.3.3 Tukey's HSD in a linear model

$\boldsymbol{Y} \,\big|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbf{I}_n)$, rank$(\mathbb{X}_{n\times k}) = k < n$

- $\mathbb{L}_{m\times k}$: a matrix with non-zero rows $\mathbf{l}_1^\top, \ldots, \mathbf{l}_m^\top$,

$$\boldsymbol{\eta} := \mathbb{L}\boldsymbol{\beta} = \big(\mathbf{l}_1^\top\boldsymbol{\beta}, \ldots, \mathbf{l}_m^\top\boldsymbol{\beta}\big)^\top = \big(\eta_1, \ldots, \eta_m\big)^\top.$$

- $\mathbb{L}$ such that $\mathbb{V} := \mathbb{L}\,\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\mathbb{L}^\top = \big(v_{j,l}\big)_{j,l=1,\ldots,m}$
  is **diagonal** with $v_j^2 := v_{j,j}$, $j = 1, \ldots, m$.

Properties of LSE (conditionally given $\mathbb{X}$)

$$\boldsymbol{T} = \widehat{\boldsymbol{\eta}} := \big(\mathbf{l}_1^\top\widehat{\boldsymbol{\beta}}, \ldots, \mathbf{l}_m^\top\widehat{\boldsymbol{\beta}}\big)^\top = \mathbb{L}\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_m\big(\boldsymbol{\eta},\, \sigma^2\mathbb{V}\big),$$

$$\frac{(n-k)\mathrm{MS}_e}{\sigma^2} \sim \chi^2_{n-k},$$

$\widehat{\boldsymbol{\eta}}$ and MS_e independent.

14.3.3 Tukey's HSD in a linear model

One-way classification

$\boldsymbol{Y} = \big(Y_{1,1}, \ldots, Y_{G,n_G}\big)^\top$, $n = \sum_{g=1}^{G} n_g$

$$Y_{g,j} \sim \mathcal{N}(m_g,\ \sigma^2),$$

$$Y_{g,j} \text{ independent for } g = 1, \ldots, G,\ j = 1, \ldots, n_g,$$

LSE of group means and their properties (with random covariates conditionally given the covariate values)

$$\boldsymbol{T} := \begin{pmatrix} \overline{Y}_1 \\ \vdots \\ \overline{Y}_G \end{pmatrix} \sim \mathcal{N}_G\left(\begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix},\quad \sigma^2 \begin{pmatrix} \frac{1}{n_1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{1}{n_G} \end{pmatrix}\right).$$

$$\frac{\nu_e\,\mathrm{MS}_e}{\sigma^2} \sim \chi^2_{\nu_e} \quad \text{with } \nu_e = n - G, \qquad \mathrm{MS}_e \text{ and } \boldsymbol{T} \text{ independent}$$

14.3.3 Tukey's HSD in a linear model

Two-way classification, **BALANCED** data

$\boldsymbol{Y} = \big(Y_{1,1,1}, \ldots, Y_{G,H,n_{G,H}}\big)^\top$, $n_{g,h} = J$ for all g, h, $n = GHJ$

$$Y_{g,h,j} \sim \mathcal{N}(m_{g,h},\ \sigma^2),$$

$$Y_{g,h,j} \text{ independent for } g = 1,\ldots,G,\ h = 1,\ldots,H,\ j = 1,\ldots,J,$$

LSE of the means of the group means and their properties (with random covariates conditionally)

Both interaction and additive model:

$$\boldsymbol{T} := \begin{pmatrix} \overline{Y}_{1\bullet} \\ \vdots \\ \overline{Y}_{G\bullet} \end{pmatrix} \sim \mathcal{N}_G\left(\begin{pmatrix} \overline{m}_{1\bullet} \\ \vdots \\ \overline{m}_{G\bullet} \end{pmatrix},\ \sigma^2 \begin{pmatrix} \frac{1}{JH} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{JH} \end{pmatrix}\right),$$

$$\frac{\nu_e^\star\,\mathrm{MS}_e^\star}{\sigma^2} \sim \chi^2_{\nu_e^\star}, \qquad \mathrm{MS}_e^\star \text{ and } \boldsymbol{T} \text{ independent}$$

Section **14.4**

Hothorn-Bretz-Westfall procedure

14.4.1 Max-abs-t distribution

Definition 14.6 Max-abs-t-distribution.

Let $\boldsymbol{T} = \big(T_1, \ldots, T_m\big)^\top \sim \mathrm{mvt}_{m,\nu}(\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive *semidefinite* matrix. The distribution of a random variable

$$H = \max_{j=1,\ldots,m} |T_j|$$

will be called the max-abs-t-distribution of dimension m with ν degrees of freedom and a scale matrix $\boldsymbol{\Sigma}$ and will be denoted as $\mathrm{h}_{m,\nu}(\boldsymbol{\Sigma})$.

Notation.

- For $0 < p < 1$, the $p\,100\%$ quantile of the distribution $\mathrm{h}_{m,\nu}(\boldsymbol{\Sigma})$ will be denoted as $\mathrm{h}_{m,\nu}(p;\,\boldsymbol{\Sigma})$. That is, $\mathrm{h}_{m,\nu}(p;\,\boldsymbol{\Sigma})$ is the number satisfying

$$\mathrm{P}\Big(\max_{j=1,\ldots,m} |T_j| \le \mathrm{h}_{m,\nu}(p;\,\boldsymbol{\Sigma})\Big) = p.$$

- The distribution function of the random variable with distribution $\mathrm{h}_{m,\nu}(\boldsymbol{\Sigma})$ will be denoted $\mathrm{CDF}_{\mathrm{h},m,\nu}(\cdot;\,\boldsymbol{\Sigma})$.

14.4.2 General multiple comparison procedure for a linear model

$\boldsymbol{Y} \mid \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, rank$(\mathbb{X}_{n\times k}) = k < n$

- $\mathbb{L}_{m\times k}$: a matrix with non-zero rows $\mathbf{l}_1^\top, \ldots, \mathbf{l}_m^\top$,

$$\boldsymbol{\theta} := \mathbb{L}\boldsymbol{\beta} = \left(\mathbf{l}_1^\top\boldsymbol{\beta}, \ldots, \mathbf{l}_m^\top\boldsymbol{\beta}\right)^\top = \left(\theta_1, \ldots, \theta_m\right)^\top.$$

- We allow for: $m > k$;
  linearly dependent rows in $\mathbb{L}$;
  matrix $\mathbb{V} := \mathbb{L}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{L}^\top$ neither diagonal nor invertible.

(Standard) notation

- $\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{X}^\top\boldsymbol{Y}$
- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}} = \left(\mathbf{l}_1^\top\widehat{\boldsymbol{\beta}}, \ldots, \mathbf{l}_m^\top\widehat{\boldsymbol{\beta}}\right)^\top = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_m\right)^\top$: LSE of $\boldsymbol{\theta}$
- $\mathbb{V} = \mathbb{L}\left(\mathbb{X}^\top\mathbb{X}\right)^{-1}\mathbb{L}^\top = \left(v_{j,l}\right)_{j,l=1,\ldots,m}$
- $\mathbb{D} = \text{diag}\left(\dfrac{1}{\sqrt{v_{1,1}}}, \ldots, \dfrac{1}{\sqrt{v_{m,m}}}\right)$
- MS_e: the residual mean square of the model with $\nu_e = n - k$ degrees of freedom

14.4.2 General MCP for a linear model

Properties of LSE

For $j = 1, \ldots, m$ (both conditionally given $\mathbb{X}$ and unconditionally as well):

$$Z_j := \frac{\widehat{\theta}_j - \theta_j}{\sqrt{\sigma^2\, v_{j,j}}} \;\sim\; \mathcal{N}(0,\,1), \qquad\qquad T_j := \frac{\widehat{\theta}_j - \theta_j}{\sqrt{\mathrm{MS}_e\, v_{j,j}}} \;\sim\; \mathrm{t}_{n-k}.$$

Conditionally given $\mathbb{X}$:

$$\boldsymbol{Z} = \bigl(Z_1,\,\ldots,\,Z_m\bigr)^\top = \frac{1}{\sqrt{\sigma^2}}\,\mathbb{D}\,\bigl(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\bigr) \;\sim\; \mathcal{N}_m\bigl(\mathbf{0}_m,\,\mathbb{DVD}\bigr),$$

$$\boldsymbol{T} = \bigl(T_1,\,\ldots,\,T_m\bigr)^\top = \frac{1}{\sqrt{\mathrm{MS}_e}}\,\mathbb{D}\,\bigl(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\bigr) \;\sim\; \mathrm{mvt}_{m,\,n-k}\bigl(\mathbb{DVD}\bigr).$$

Theorem 14.5 Hothorn-Bretz-Westfall MCP for linear hypotheses in a normal linear model.

Random intervals given by

$$\theta_j^{HL}(\alpha) = \widehat{\theta}_j - \mathrm{h}_{m,\,n-k}(1-\alpha;\, \mathbb{DVD})\,\sqrt{\mathsf{MS}_e\, v_{j,j}},$$
$$\theta_j^{HU}(\alpha) = \widehat{\theta}_j + \mathrm{h}_{m,\,n-k}(1-\alpha;\, \mathbb{DVD})\,\sqrt{\mathsf{MS}_e\, v_{j,j}}, \qquad j = 1,\,\dots,\,m.$$

are simultaneous confidence intervals for parameters $\theta_j = \mathbf{l}_j^\top \boldsymbol{\beta}$, $j = 1,\dots,m$, *with an exact coverage of* $1-\alpha$, *i.e., for any* $\boldsymbol{\theta}^0 = \left(\theta_1^0,\dots,\theta_m^0\right)^\top \in \mathbb{R}^m$

$$\mathsf{P}\Big(\text{for all } j = 1,\dots,m \quad \big(\theta_j^{HL}(\alpha),\ \theta_j^{HU}(\alpha)\big) \ni \theta_j^0;\ \boldsymbol{\theta} = \boldsymbol{\theta}^0\Big) = 1-\alpha.$$

Related P-values for a multiple testing problem with elementary hypotheses $H_j:\ \theta_j = \theta_j^0$, $\theta_j^0 \in \mathbb{R}$, $j = 1,\dots,m$, *adjusted for multiple comparison are given by*

$$p_j^H = 1 - \mathrm{CDF}_{\mathrm{h},m,n-k}\Big(\big|t_j^0\big|;\, \mathbb{DVD}\Big), \qquad j = 1,\dots,m,$$

where t_j^0 *is a value of* $T_j(\theta_j^0) = \frac{\widehat{\theta}_j - \theta_j^0}{\sqrt{\mathsf{MS}_e\, v_{j,j}}}$ *attained with given data.*

Section **14.5**

Confidence band for the regression function

14.5 Confidence band for the regression function

$\left(Y_i, \boldsymbol{Z}_i^\top\right)^\top \overset{\text{i.i.d.}}{\sim} \left(Y, \boldsymbol{Z}^\top\right)^\top, \; i = 1, \ldots, n$

Model matrix $\mathbb{X}$ based on a known transformation $\boldsymbol{t} : \mathbb{R}^p \longrightarrow \mathbb{R}^k$ of the covariates $\mathbb{Z}$.

$$\boldsymbol{Y} \,\big|\, \mathbb{Z} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\big), \qquad \text{rank}\big(\mathbb{X}_{n\times k}\big) = k,$$

$$Y_i \,\big|\, \boldsymbol{Z}_i \sim \mathcal{N}\big(\boldsymbol{X}_i^\top \boldsymbol{\beta}, \sigma^2\big), \qquad \boldsymbol{X}_i = \boldsymbol{t}(\boldsymbol{Z}_i), \; i = 1, \ldots, n,$$

$$\varepsilon_i = Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta} \overset{\text{i.i.d.}}{\sim} \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Regression function

$$\mathbb{E}\big(Y \,\big|\, \boldsymbol{X} = \boldsymbol{t}(\boldsymbol{z})\big) = \mathbb{E}\big(Y \,\big|\, \boldsymbol{Z} = \boldsymbol{z}\big) = m(\boldsymbol{z}) = \boldsymbol{t}^\top(\boldsymbol{z})\boldsymbol{\beta}, \qquad \boldsymbol{z} \in \mathbb{R}^p$$

Confidence interval for the model based mean

For any $\boldsymbol{z} \in \mathbb{R}^p$, any $\boldsymbol{\beta}^0 \in \mathbb{R}^k$, $\sigma_0^2 > 0$,

$$\mathsf{P}\left(\boldsymbol{t}^\top(\boldsymbol{z})\widehat{\boldsymbol{\beta}} \pm \mathrm{t}_{n-k}\left(1 - \frac{\alpha}{2}\right)\sqrt{\mathrm{MS}_e\, \boldsymbol{t}^\top(\boldsymbol{z})\big(\mathbb{X}^\top\mathbb{X}\big)^{-1}\boldsymbol{t}(\boldsymbol{z})} \ni \boldsymbol{t}^\top(\boldsymbol{z})\boldsymbol{\beta}^0; \; \boldsymbol{\beta} = \boldsymbol{\beta}^0, \sigma^2 = \sigma_0^2\right) = 1 - \alpha.$$

14.5 Confidence band for the regression function

Theorem 14.6 Confidence band for the regression function.

Let $\bigl(Y_i,\,\boldsymbol{Z}_i^\top\bigr)^\top$, $i = 1,\dots,n$, *be i.i.d. random vectors such that* $\boldsymbol{Y}\,\big|\,\mathbb{Z} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\,\sigma^2\mathbf{I}_n)$, *where* $\mathbb{X}$ *is the* $n \times k$ *model matrix based on a known transformation* $\boldsymbol{t} : \mathbb{R}^p \longrightarrow \mathbb{R}^k$ *of the covariates* $\boldsymbol{Z}_1,\,\dots,\,\boldsymbol{Z}_n$. *Let* $\operatorname{rank}\bigl(\mathbb{X}_{n\times k}\bigr) = k$. *Finally, let for all* $\boldsymbol{z} \in \mathbb{R}^p$ $\boldsymbol{t}(\boldsymbol{z}) \neq \mathbf{0}_k$. *Then for any* $\boldsymbol{\beta}^0 \in \mathbb{R}^k$, $\sigma_0^2 > 0$,

$$\mathsf{P}\Bigl(\textit{for all } \boldsymbol{z} \in \mathbb{R}^p \quad \boldsymbol{t}^\top(\boldsymbol{z})\widehat{\boldsymbol{\beta}} \pm \sqrt{k\,\mathcal{F}_{k,\,n-k}(1-\alpha)\,\mathsf{MS}_e\,\boldsymbol{t}^\top(\boldsymbol{z})\bigl(\mathbb{X}^\top\mathbb{X}\bigr)^{-1}\boldsymbol{t}(\boldsymbol{z})} \ni \boldsymbol{t}^\top(\boldsymbol{z})\boldsymbol{\beta}^0; \quad \boldsymbol{\beta} = \boldsymbol{\beta}^0,\,\sigma^2 = \sigma_0^2\Bigr) = 1-\alpha.$$

14.5 Confidence band for the regression function

Half width of the confidence band

Band **FOR** the regression function (overall coverage)

$$\sqrt{k\,\mathcal{F}_{k,n-k}(1-\alpha)\;\;\mathsf{MS}_e\;\boldsymbol{t}^\top(\boldsymbol{z})(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{t}(\boldsymbol{z})}.$$

Band **AROUND** the regression function (pointwise coverage)

$$\mathrm{t}_{n-k}\Big(1-\frac{\alpha}{2}\Big)\;\;\sqrt{\mathsf{MS}_e\;\boldsymbol{t}^\top(\boldsymbol{z})(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{t}(\boldsymbol{z})}$$

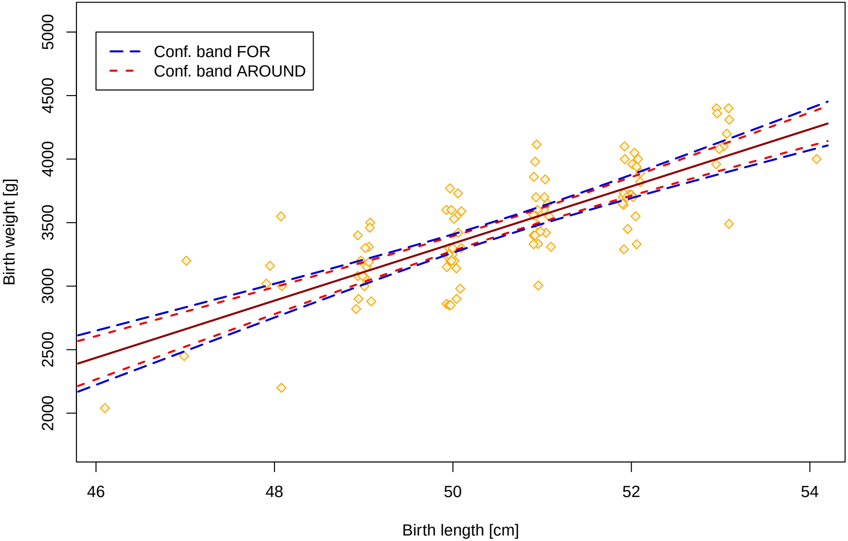
$$=\;\sqrt{\mathcal{F}_{1,n-k}(1-\alpha)\;\;\mathsf{MS}_e\;\boldsymbol{t}^\top(\boldsymbol{z})(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{t}(\boldsymbol{z})},$$

For $k \geq 2$, and any $\nu > 0$,

$$k\,\mathcal{F}_{k,\nu}(1-\alpha) > \mathcal{F}_{1,\nu}(1-\alpha)$$

Kojeni ($n = 99$)

bweight $\sim$ blength



15

General Linear Model

15 General Linear Model

Definition 15.1 General linear model.

The data $(\boldsymbol{Y}, \mathbb{X})$ satisfy a *general linear model* if

$$\mathbb{E}(\boldsymbol{Y} \,|\, \mathbb{X}) = \mathbb{X}\boldsymbol{\beta}, \qquad \mathsf{var}(\boldsymbol{Y} \,|\, \mathbb{X}) = \sigma^2\, \mathbb{W}^{-1},$$

where $\boldsymbol{\beta} \in \mathbb{R}^k$ and $0 < \sigma^2 < \infty$ are unknown parameters and $\mathbb{W}$ is a *known* positive definite matrix.

Notation: $\boldsymbol{Y} \,|\, \mathbb{X} \sim \big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\mathbb{W}^{-1}\big)$.

15 General Linear Model

Example: Regression based on sample means

Data (we would like to have):

$$\left(\widetilde{Y}_{1,1},\ldots,\widetilde{Y}_{1,w_1},\, \boldsymbol{X}_1\right),$$
$$\ldots,$$
$$\left(\widetilde{Y}_{n,1},\ldots,\widetilde{Y}_{n,w_n},\, \boldsymbol{X}_n\right)$$

Observable data:

$$Y_1 = \frac{1}{w_1}\sum_{j=1}^{w_1} \widetilde{Y}_{1,j}, \quad \ldots, \quad Y_n = \frac{1}{w_n}\sum_{j=1}^{w_n} \widetilde{Y}_{n,j}$$

and the related covariates/regressors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$

15 General Linear Model

Theorem 15.1 Generalized least squares.

Assume a general linear model $\boldsymbol{Y} \,|\, \mathbb{X} \sim (\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbb{W}^{-1})$, *where* $\mathrm{rank}(\mathbb{X}_{n\times k}) = k < n$. *The following then holds:*

(i) *A vector*

$$\widehat{\boldsymbol{Y}}_G := \mathbb{X}\big(\mathbb{X}^\top \mathbb{W}\mathbb{X}\big)^{-1}\mathbb{X}^\top \mathbb{W}\,\boldsymbol{Y}$$

is the best linear unbiased estimator (BLUE) of a vector parameter $\boldsymbol{\mu} := \mathbb{E}\big(\boldsymbol{Y} \,|\, \mathbb{X}\big) = \mathbb{X}\boldsymbol{\beta}$, *and*

$$\mathrm{var}\big(\widehat{\boldsymbol{Y}}_G \,|\, \mathbb{X}\big) = \sigma^2\, \mathbb{X}\big(\mathbb{X}^\top \mathbb{W}\mathbb{X}\big)^{-1}\mathbb{X}^\top.$$

If further $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbb{W}^{-1})$ *then*

$$\widehat{\boldsymbol{Y}}_G \,|\, \mathbb{X} \sim \mathcal{N}_n\Big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2\, \mathbb{X}\big(\mathbb{X}^\top \mathbb{W}\mathbb{X}\big)^{-1}\mathbb{X}^\top\Big).$$

TO BE CONTINUED.

Theorem 15.1 Generalized least squares, cont'd.

(ii) *Let* $\mathbf{l} \in \mathbb{R}^k$, $\mathbf{l} \neq \mathbf{0}_k$ *and let*

$$\widehat{\boldsymbol{\beta}}_G := (\mathbb{X}^\top \mathbb{W} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{W} \boldsymbol{Y}.$$

Then $\widehat{\theta}_G = \mathbf{l}^\top \widehat{\boldsymbol{\beta}}_G$ *is the best linear unbiased estimator (BLUE) of* θ *with*

$$\mathrm{var}(\widehat{\theta}_G \,|\, \mathbb{X}) = \sigma^2 \mathbf{l}^\top (\mathbb{X}^\top \mathbb{W} \mathbb{X})^{-1} \mathbf{l}.$$

If further $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n(\mathbb{X}\boldsymbol{\beta}, \, \sigma^2 \mathbb{W}^{-1})$ *then*

$$\widehat{\theta}_G \,|\, \mathbb{X} \sim \mathcal{N}\big(\theta, \, \sigma^2 \mathbf{l}^\top (\mathbb{X}^\top \mathbb{W} \mathbb{X})^{-1} \mathbf{l}\big).$$

TO BE CONTINUED.

Theorem 15.1 Generalized least squares, cont'd.

(iii) *The vector*

$$\widehat{\boldsymbol{\beta}}_G := \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{W} \boldsymbol{Y}$$

is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ *with*

$$\mathrm{var}\big(\widehat{\boldsymbol{\beta}}_G \,\big|\, \mathbb{X}\big) = \sigma^2 \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1}.$$

If additionally $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbb{W}^{-1}\big)$ *then*

$$\widehat{\boldsymbol{\beta}}_G \,\big|\, \mathbb{X} \sim \mathcal{N}_k\Big(\boldsymbol{\beta},\, \sigma^2 \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1}\Big).$$

TO BE CONTINUED.

15 General Linear Model

Theorem 15.1 Generalized least squares, cont'd.

(iv) *The statistic*

$$\mathrm{MS}_{e,G} := \frac{\mathrm{SS}_{e,G}}{n-k},$$

where

$$\mathrm{SS}_{e,G} := \left\| \mathbb{W}^{\frac{1}{2}} \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G \right) \right\|^2 = \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G \right)^\top \mathbb{W} \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G \right),$$

is the unbiased estimator of the residual variance σ^2.
If additionally $\boldsymbol{Y} \,|\, \mathbb{X} \sim \mathcal{N}_n\big(\mathbb{X}\boldsymbol{\beta},\, \sigma^2 \mathbb{W}^{-1}\big)$ *then*

$$\frac{\mathrm{SS}_{e,G}}{\sigma^2} \sim \chi^2_{n-k},$$

and the statistics $\mathrm{SS}_{e,G}$ *and* $\widehat{\boldsymbol{Y}}_G$ *are conditionally, given* $\mathbb{X}$, *independent.*

15 General Linear Model

Terminology.

- $\widehat{\boldsymbol{Y}}_G = \mathbb{X}\left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{W} \boldsymbol{Y}$:
  the vector of the generalized fitted values.

- $\mathrm{SS}_{e,G} = \left\| \mathbb{W}^{\frac{1}{2}} \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G\right) \right\|^2 = \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G\right)^\top \mathbb{W} \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_G\right)$:
  the generalized residual sum of squares.

- $\mathrm{MS}_{e,G} = \dfrac{\mathrm{SS}_{e,G}}{n-k}$:
  the generalized mean square.

- The statistic $\widehat{\boldsymbol{\beta}}_G = \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{W} \boldsymbol{Y}$ in a full-rank general linear model:
  the generalized least squares (GLS) estimator of the regression coefficients.

Data `Kojeni` and `wKojeni`

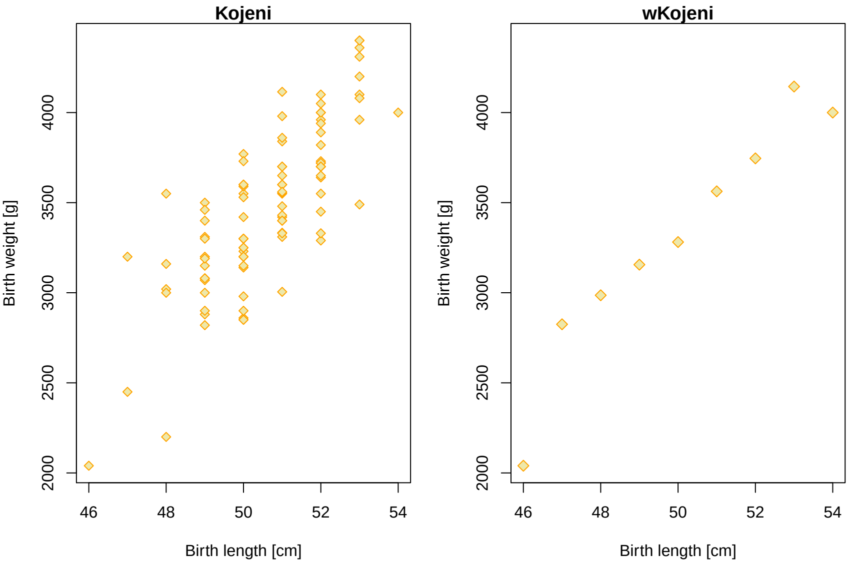
Kojeni

- Data on $n = 99$ newborn children.
- Y: birth weight (`bweight`).
- X: birth length (`blength`)
  - Only (nine) discrete values 46, 47, ..., 54 [cm] appear in data due to rounding.

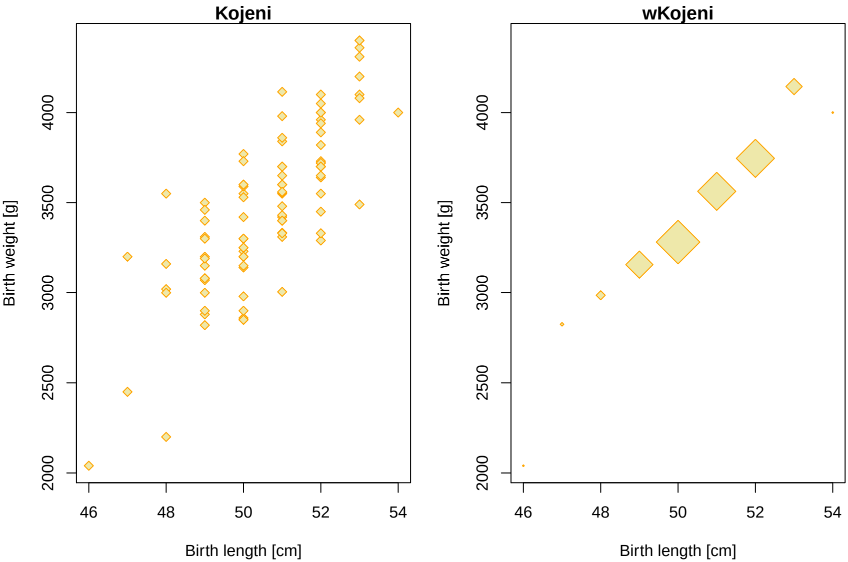
wKojeni

- $n = 9$.
- Y: average birth weight of all children from data `Kojeni` with the same birth length.

Data `Kojeni` and `wKojeni`



Data Kojeni and wKojeni



Data Kojeni

bweight ∼ blength

Ordinary least squares using complete data Kojeni

```
m1 <- lm(bweight ~ blength, data = Kojeni)
summary(m1)
confint(m1)
```

```
### summary(m1):
Call:
lm(formula = bweight ~ blength, data = Kojeni)

Residuals:
    Min      1Q  Median      3Q     Max
-685.93 -152.83  -30.76  196.83  664.07

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7905.80     895.45  -8.829 4.52e-14 ***
blength       224.83      17.69  12.709  < 2e-16 ***
 ---

Residual standard error: 271.7 on 97 degrees of freedom
Multiple R-squared:  0.6248,       Adjusted R-squared:  0.6209
F-statistic: 161.5 on 1 and 97 DF,  p-value: < 2.2e-16

### confint(m1):
                 2.5 %     97.5 %                        2.5 %     97.5 %
(Intercept) -9683.0226 -6128.5847          blength  189.7184   259.9372
```

Data wKojeni

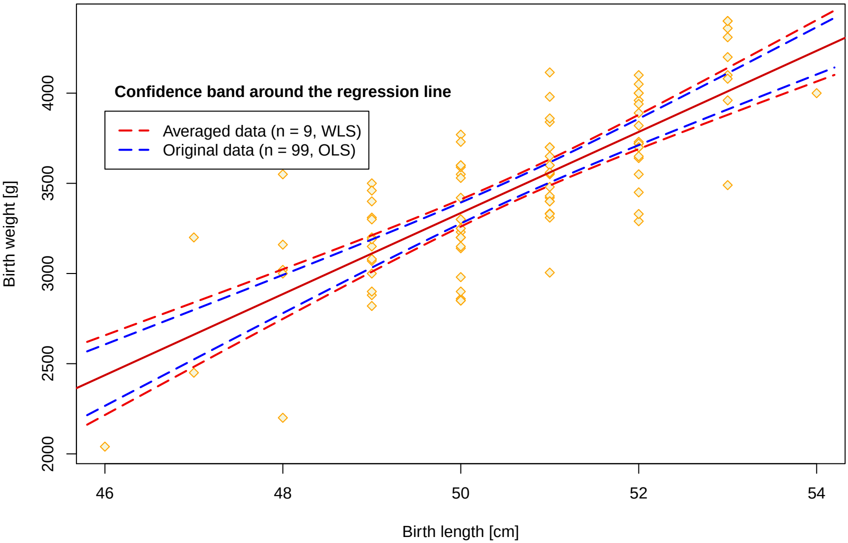
bweight ∼ blength

Weighted least squares using averaged data wKojeni

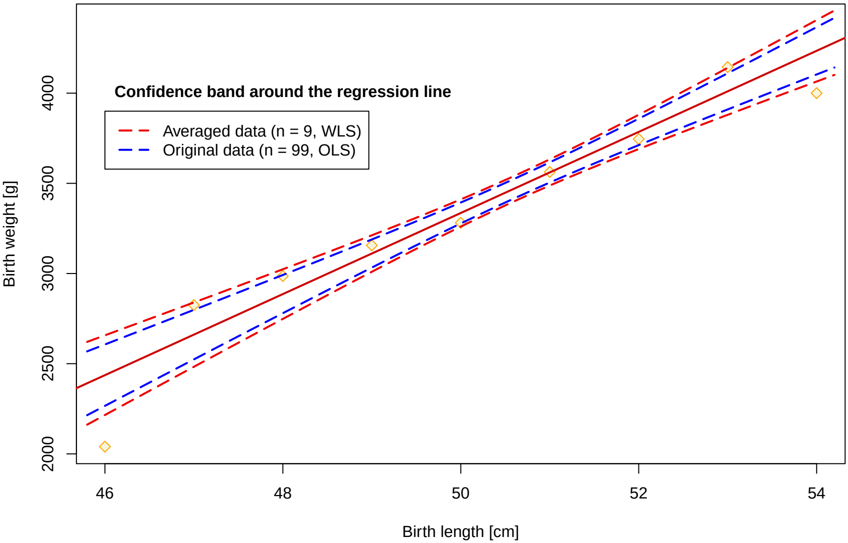
```
wm1 <- lm(bweight ~ blength, weights = w, data = wKojeni)
summary(wm1)
confint(wm1)
```

```
### summary(wm1):
Call:
lm(formula = bweight ~ blength, data = wKojeni, weights = w)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-396.28 -234.90   10.75  223.76  403.12

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7905.80     975.42  -8.105 8.39e-05 ***
blength       224.83      19.27  11.667 7.68e-06 ***
---

Residual standard error: 295.9 on 7 degrees of freedom
Multiple R-squared:  0.9511,        Adjusted R-squared:  0.9441
F-statistic: 136.1 on 1 and 7 DF,  p-value: 7.676e-06

### confint(wm1):
                  2.5 %     97.5 %                          2.5 %     97.5 %
(Intercept) -10212.3079 -5599.2995          blength   179.2623   270.3934
```

Data `Kojeni` and `wKojeni`



Data `Kojeni` and `wKojeni`



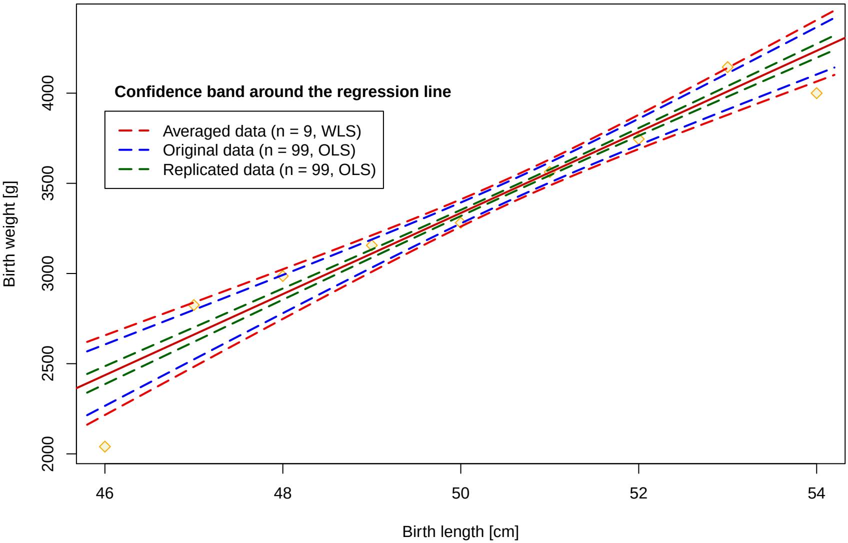
bweight ∼ blength

Ordinary least squares for data replicated from wKojeni

```
replKojeni <- data.frame(bweight = rep(wKojeni[, "bweight"], wKojeni[, "w"]),
                         blength = rep(wKojeni[, "blength"], wKojeni[, "w"]))
m1repl <- lm(bweight ~ blength, data = replKojeni)
summary(m1repl)
confint(m1repl)
```

```
### summary(m1repl):
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7905.804    262.033  -30.17   <2e-16 ***
blength       224.828      5.177   43.43   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 79.5 on 97 degrees of freedom
Multiple R-squared:  0.9511,	Adjusted R-squared:  0.9506
F-statistic:  1886 on 1 and 97 DF,  p-value: < 2.2e-16

### confint(m1repl):
                 2.5 %     97.5 %                  2.5 %     97.5 %
(Intercept)  -8425.8658 -7385.7416     blength   214.5539   235.1018
```

Data Kojeni and wKojeni



16

Asymptotic Properties of the LSE and Sandwich Estimator

Section **16.1**

Assumptions and setup

16.1 Assumptions and setup

Assumption (A0)

(i) Let $(Y_1, \boldsymbol{X}_1^\top)^\top, (Y_2, \boldsymbol{X}_2^\top)^\top, \ldots$ be a *sequence* of $(1+k)$-dimensional *independent and identically distributed (i.i.d.)* random vectors being distributed as a generic random vector $(Y, \boldsymbol{X}^\top)^\top$,
$(\boldsymbol{X} = (X_0, X_1, \ldots, X_{k-1})^\top$,
$\boldsymbol{X}_i = (X_{i,0}, X_{i,1}, \ldots, X_{i,k-1})^\top, i = 1, 2, \ldots)$;

(ii) Let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{k-1})^\top$ be an unknown k-dimensional real parameter;

(iii) Let $\mathbb{E}(Y \mid \boldsymbol{X}) = \boldsymbol{X}^\top \boldsymbol{\beta}$.

Notation: error terms

We denote
$$\varepsilon = Y - \boldsymbol{X}^\top \boldsymbol{\beta},$$
$$\varepsilon_i = Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}, \quad i = 1, 2, \ldots.$$

16.1 Assumptions and setup

Assumption (A1)

Let the covariate random vector $\boldsymbol{X} = \left(X_0, \ldots, X_{k-1}\right)^\top$ satisfy

(i) $\mathbb{E}\left|X_j X_l\right| < \infty, \quad j, l = 0, \ldots, k-1;$

(ii) $\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) = \mathbb{W}$, where $\mathbb{W}$ is a positive definite matrix.

Notation: covariates second and first mixed moments

Let $\mathbb{W} = \left(w_{j,l}\right)_{j,l=0,\ldots,k-1}$. We have,

$$w_j^2 := w_{j,j} = \mathbb{E}\left(X_j^2\right), \qquad j = 0, \ldots, k-1,$$

$$w_{j,l} = \mathbb{E}\left(X_j X_l\right), \qquad j \neq l.$$

Let

$$\mathbb{V} := \mathbb{W}^{-1} = \left(v_{j,l}\right)_{j,l=0,\ldots,k-1}.$$

16.1 Assumptions and setup

Notation: Data of size n

For $n \geq 1$:

$$\boldsymbol{Y}_n := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbb{X}_n := \begin{pmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{pmatrix}, \qquad \mathbb{W}_n := \mathbb{X}_n^\top \mathbb{X}_n = \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^\top,$$

$$\mathbb{V}_n := \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1} \text{ (if it exists).}$$

Lemma 16.1 Consistent estimator of the second and first mixed moments of the covariates.

Let assumpions (A0) and (A1) hold. Then

$$\frac{1}{n}\mathbb{W}_n \xrightarrow{a.s.} \mathbb{W} \qquad \text{as } n \to \infty,$$

$$n\mathbb{V}_n \xrightarrow{a.s.} \mathbb{V} \qquad \text{as } n \to \infty.$$

16.1 Assumptions and setup

Assumption (A2 homoscedastic)

Let the conditional variance of the response satisfy

$$\sigma^2(\boldsymbol{X}) := \operatorname{var}(Y \mid \boldsymbol{X}) = \sigma^2,$$

where $\infty > \sigma^2 > 0$ is an unknown parameter.

Assumption (A2 heteroscedastic)

Let $\sigma^2(\boldsymbol{X}) := \operatorname{var}(Y \mid \boldsymbol{X})$ satisfy $\mathbb{E}\{\sigma^2(\boldsymbol{X})\} < \infty$ and also for each $j, l = 0, \ldots, k-1$,

$$\mathbb{E}\{\sigma^2(\boldsymbol{X}) X_j X_l\} < \infty.$$

Notation

$$\mathbb{W}^\star := \mathbb{E}\{\sigma^2(\boldsymbol{X}) \boldsymbol{X}\boldsymbol{X}^\top\}$$

Section **16.2**

Consistency of LSE

16.2 Consistency of LSE

Will be shown

(i) Strong consistency of $\widehat{\boldsymbol{\beta}}_n$, $\widehat{\boldsymbol{\theta}}_n$, $\widehat{\boldsymbol{\xi}}_n$ (LSE's regression coefficients or their linear combinations).

- No need of normality;
- No need of homoscedasticity.

(ii) Strong consistency of $\mathsf{MS}_{e,n}$ (unbiased estinator of the residual variance).

- No need of normality.

16.2 Consistency of LSE

Theorem 16.2 Strong consistency of LSE.

Let assumptions (A0), (A1) and (A2 heteroscedastic) hold.

Then

$$\widehat{\beta}_n \xrightarrow{a.s.} \beta \qquad \text{as } n \to \infty,$$

$$\mathbf{l}^\top \widehat{\beta}_n = \widehat{\theta}_n \xrightarrow{a.s.} \theta = \mathbf{l}^\top \beta \qquad \text{as } n \to \infty,$$

$$\mathbb{L}\widehat{\beta}_n = \widehat{\xi}_n \xrightarrow{a.s.} \xi = \mathbb{L}\beta \qquad \text{as } n \to \infty.$$

16.2 Consistency of LSE

Theorem 16.3 Strong consistency of the mean squared error.

Let assumptions (A0), (A1), (A2 homoscedastic) hold.

Then

$$\mathrm{MS}_{e,n} \xrightarrow{a.s.} \sigma^2 \quad \text{as } n \to \infty.$$

Section **16.3**

Asymptotic normality of LSE under homoscedasticity

16.3 Asymptotic normality of LSE under homoscedasticity

Reminder

$$\mathbb{V} = \left\{ \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}^\top \right) \right\}^{-1}$$

Theorem 16.4 Asymptotic normality of LSE in homoscedastic case.

Let assumptions (A0), (A1), (A2 homoscedastic) hold. Further, let $\mathbb{E}\left|\varepsilon^2 X_j X_l\right| < \infty$ *for each* $j,\, l = 0, \ldots, k-1$.

Then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_k\left(\mathbf{0}_k,\ \sigma^2\,\mathbb{V}\right) \qquad as\ n \to \infty,$$

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}_1\left(0,\ \sigma^2\,\mathbf{l}^\top \mathbb{V}\,\mathbf{l}\right) \qquad as\ n \to \infty,$$

$$\sqrt{n}\left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_m\left(\mathbf{0}_m,\ \sigma^2\,\mathbb{L}\,\mathbb{V}\,\mathbb{L}^\top\right) \qquad as\ n \to \infty.$$

16.3.1 Asymptotic validity of the classical inference under homoscedasticity but non-normality

For $n \geq n_0 > k$ ($\mathbb{L}$ is a matrix with m rows and k columns)

$$T_n := \frac{\widehat{\theta}_n - \theta}{\sqrt{\mathsf{MS}_{e,n}\,\mathbf{l}^\top \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}\mathbf{l}}},$$

$$Q_n := \frac{1}{m}\,\frac{\left(\widehat{\xi}_n - \xi\right)^\top \left\{\mathbb{L}\left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}\mathbb{L}^\top\right\}^{-1}\left(\widehat{\xi}_n - \xi\right)}{\mathsf{MS}_{e,n}}.$$

Consequence of Theorem 16.4: Asymptotic distribution of t- and F-statistics.

Under assumptions of Theorem 16.4:

$$T_n \xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\,1) \qquad \text{as } n \to \infty,$$

$$m\,Q_n \xrightarrow{\mathcal{D}} \chi^2_m \qquad \text{as } n \to \infty.$$

16.3.1 Asymptotic validity of the classical inference under homoscedasticity but non-normality

Confidence interval for θ based on the $\mathcal{N}(0,\,1)$ distribution

$$\mathcal{I}_n^{\mathcal{N}} := \Bigl(\widehat{\theta}_n \,-\, u(1-\alpha/2)\,\sqrt{\mathsf{MS}_{e,n}\,\mathbf{l}^\top\bigl(\mathbb{X}_n^\top\mathbb{X}_n\bigr)^{-1}\mathbf{l}},\quad \widehat{\theta}_n \,+\, u(1-\alpha/2)\,\sqrt{\mathsf{MS}_{e,n}\,\mathbf{l}^\top\bigl(\mathbb{X}_n^\top\mathbb{X}_n\bigr)^{-1}\mathbf{l}}\Bigr)$$

Confidence interval for θ based on the t_{n-k} distribution

$$\mathcal{I}_n^{\mathrm{t}} := \Bigl(\widehat{\theta}_n \,-\, \mathrm{t}_{n-k}(1-\alpha/2)\,\sqrt{\mathsf{MS}_{e,n}\,\mathbf{l}^\top\bigl(\mathbb{X}_n^\top\mathbb{X}_n\bigr)^{-1}\mathbf{l}},\quad \widehat{\theta}_n \,+\, \mathrm{t}_{n-k}(1-\alpha/2)\,\sqrt{\mathsf{MS}_{e,n}\,\mathbf{l}^\top\bigl(\mathbb{X}_n^\top\mathbb{X}_n\bigr)^{-1}\mathbf{l}}\Bigr)$$

16.3.1 Asymptotic validity of the classical inference under homoscedasticity but non-normality

Asymptotic coverage (for any $\theta^0 \in \mathbb{R}$)

$$\mathsf{P}\big(\mathcal{I}_n^{\mathcal{N}} \ni \theta^0;\ \theta = \theta^0\big) \longrightarrow 1 - \alpha \qquad \text{as } n \to \infty,$$

$$\mathsf{P}\big(\mathcal{I}_n^{\mathrm{t}} \ni \theta^0;\ \theta = \theta^0\big) \longrightarrow 1 - \alpha \qquad \text{as } n \to \infty.$$

16.3.1 Asymptotic validity of the classical inference under homoscedasticity but non-normality

Confidence ellipsoid for ξ based on the χ^2_m distribution

$$\mathcal{K}_n^{\chi} := \Big\{ \boldsymbol{\xi} \in \mathbb{R}^m : \quad \big(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\big)^\top \Big\{ \mathsf{MS}_{e,n} \mathbb{L} \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1} \mathbb{L}^\top \Big\}^{-1} \big(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\big) < \chi^2_m(1-\alpha) \Big\}$$

Confidence ellipsoid for ξ based on the $\mathcal{F}_{m,n-k}$ distribution

$$\mathcal{K}_n^{\mathcal{F}} := \Big\{ \boldsymbol{\xi} \in \mathbb{R}^m : \quad \big(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\big)^\top \Big\{ \mathsf{MS}_{e,n} \mathbb{L} \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1} \mathbb{L}^\top \Big\}^{-1} \big(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\big) < m\,\mathcal{F}_{m,n-k}(1-\alpha) \Big\}$$

16.3.1 Asymptotic validity of the classical inference under homoscedasticity but non-normality

Asymptotic coverage (for any $\xi^0 \in \mathbb{R}^m$)

$$\mathsf{P}\big(\mathcal{K}_n^{\chi} \ni \xi^0;\ \xi = \xi^0\big) \longrightarrow 1 - \alpha \qquad \text{as } n \to \infty,$$

$$\mathsf{P}\big(\mathcal{K}_n^{\mathcal{F}} \ni \xi^0;\ \xi = \xi^0\big) \longrightarrow 1 - \alpha \qquad \text{as } n \to \infty.$$

Section **16.4**

Asymptotic normality of LSE under heteroscedasticity

16.4 Asymptotic normality of LSE under heteroscedasticity

Reminder

$$\mathbb{V} = \left\{\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^\top)\right\}^{-1}, \qquad \mathbb{W}^\star = \mathbb{E}\{\sigma^2(\boldsymbol{X})\,\boldsymbol{X}\boldsymbol{X}^\top\}.$$

Theorem 16.5 Asymptotic normality of LSE in heteroscedastic case.

Let assumptions (A0), (A1), (A2 heteroscedastic) hold. Further, let $\mathbb{E}\left|\varepsilon^2 X_j X_l\right| < \infty$ for each $j,\ l = 0, \ldots, k-1$.

Then

$$\begin{aligned}
\sqrt{n}\big(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\big) &\xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}_k,\ \mathbb{V}\mathbb{W}^\star\mathbb{V}) && \text{as } n \to \infty,\\
\sqrt{n}\big(\widehat{\theta}_n - \theta\big) &\xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\ \mathbf{l}^\top\mathbb{V}\mathbb{W}^\star\mathbb{V}\,\mathbf{l}) && \text{as } n \to \infty,\\
\sqrt{n}\big(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}\big) &\xrightarrow{\mathcal{D}} \mathcal{N}_m(\mathbf{0}_m,\ \mathbb{L}\,\mathbb{V}\mathbb{W}^\star\mathbb{V}\,\mathbb{L}^\top) && \text{as } n \to \infty.
\end{aligned}$$

16.4 Asymptotic normality of LSE under heteroscedasticity

Residuals and related quantities based on a model for data of size n

M_n: $\boldsymbol{Y}_n \,\big|\, \mathbb{X}_n \;\sim\; \left(\mathbb{X}_n\boldsymbol{\beta},\, \sigma^2\,\mathbf{I}_n\right)$

- Hat matrix: $\mathbb{H}_n = \mathbb{X}_n\left(\mathbb{X}_n^\top\mathbb{X}_n\right)^{-1}\mathbb{X}_n^\top$;
- Residual projection matrix: $\mathbb{M}_n = \mathbf{I}_n - \mathbb{H}_n$;
- Diagonal elements of matrix $\mathbb{H}_n$: $h_{n,1},\,\ldots,\,h_{n,n}$;
- Diagonal elements of matrix $\mathbb{M}_n$: $m_{n,1} = 1 - h_{n,1},\,\ldots,\,m_{n,n} = 1 - h_{n,n}$;
- Residuals: $\boldsymbol{U}_n = \mathbb{M}_n\boldsymbol{Y}_n = \left(U_{n,1},\,\ldots,\,U_{n,n}\right)^\top$.

16.4 Asymptotic normality of LSE under heteroscedasticity

Reminder

- $\mathbb{V}_n = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top\right)^{-1} = \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}$.
- Under assumptions (A0) and (A1): $n\mathbb{V}_n \xrightarrow{\text{a.s.}} \mathbb{V}$ as $n \to \infty$.

Theorem 16.6 Sandwich estimator of the covariance matrix.

Let assumptions (A0), (A1), (A2 heteroscedastic) hold. Let additionally, for each $s, t, j, l = 0, \dots, k-1$

$$\mathbb{E}\left|\varepsilon^2 X_j X_l\right| < \infty, \qquad \mathbb{E}\left|\varepsilon X_s X_j X_l\right| < \infty, \qquad \mathbb{E}\left|X_s X_t X_j X_l\right| < \infty.$$

Then

$$n\mathbb{V}_n \mathbb{W}_n^\star \mathbb{V}_n \xrightarrow{a.s.} \mathbb{V}\,\mathbb{W}^\star\,\mathbb{V} \qquad \text{as } n \to \infty,$$

where for $n = 1, 2, \dots,$

$$\mathbb{W}_n^\star = \sum_{i=1}^{n} U_{n,i}^2 \boldsymbol{X}_i \boldsymbol{X}_i^\top = \mathbb{X}_n^\top \boldsymbol{\Omega}_n \mathbb{X}_n,$$

$$\boldsymbol{\Omega}_n = \text{diag}\left(\omega_{n,1}, \dots, \omega_{n,n}\right), \qquad \omega_{n,i} = U_{n,i}^2, \quad i = 1, \dots, n.$$

16.4 Asymptotic normality of LSE under heteroscedasticity

Heteroscedasticity consistent (sandwich) estimator of the covariance matrix

$$\mathbb{V}_n\,\mathbb{W}_n^{\star}\,\mathbb{V}_n = \underbrace{\left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}\mathbb{X}_n^\top}_{\text{bread}}\ \underbrace{\boldsymbol{\Omega}_n}_{\text{meat}}\ \underbrace{\mathbb{X}_n\left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}}_{\text{bread}}$$

Alternative sorts of meat for sandwich

- $\nu_1, \nu_2, \ldots$: real sequence such that $\frac{\nu_n}{n} \to 1$ as $n \to \infty$.
- $\boldsymbol{\delta}_n = \left(\delta_{n,1}, \ldots, \delta_{n,n}\right)^\top$, $n = 1, 2, \ldots$: suitable sequence of real numbers.

$$\boldsymbol{\Omega}_n^{HC} := \mathrm{diag}\left(\omega_{n,1}, \ldots, \omega_{n,n}\right),$$

$$\omega_{n,i} = \frac{n}{\nu_n}\,\frac{U_{n,i}^2}{m_{n,i}^{\delta_{n,i}}}, \qquad\qquad i = 1, \ldots, n.$$

ν_n: degrees of freedom of the sandwich.

16.4 Asymptotic normality of LSE under heteroscedasticity

Alternative sorts of meat for sandwich

HC0: $\omega_{n,i} = U_{n,i}^2$ White (1980),

HC1: $\omega_{n,i} = \dfrac{n}{n-k}\, U_{n,i}^2$ MacKinnon and White (1985),

HC2: $\omega_{n,i} = \dfrac{U_{n,i}^2}{m_{n,i}}$ MacKinnon and White (1985),

HC3: $\omega_{n,i} = \dfrac{U_{n,i}^2}{m_{n,i}^2}$ MacKinnon and White (1985),

HC4: $\omega_{n,i} = \dfrac{U_{n,i}^2}{m_{n,i}^{\delta_{n,i}}}$ Cribari-Neto(2004),

$$\delta_{n,i} = \min\left\{4,\ \frac{h_{n,i}}{\overline{h}_n}\right\}.$$

16.4.1 Heteroscedasticity consistent asymptotic inference

For $n \geq n_0 > k$ ($\mathbb{L}$ is a matrix with m rows and k columns)

$$\mathbb{V}_n^{HC} := \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1} \mathbb{X}_n^\top \, \boldsymbol{\Omega}_n^{HC} \, \mathbb{X}_n \left(\mathbb{X}_n^\top \mathbb{X}_n\right)^{-1}.$$

$\boldsymbol{\Omega}_n^{HC}$: sequence of the meat matrices that lead to the heteroscedasticity consistent estimator of the covariance matrix of the LSE $\widehat{\boldsymbol{\beta}}_n$.

$$T_n^{HC} := \frac{\widehat{\theta}_n - \theta}{\sqrt{\mathbf{l}^\top \mathbb{V}_n^{HC} \mathbf{l}}},$$

$$Q_n^{HC} := \frac{1}{m} \left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}\right)^\top \left(\mathbb{L}\mathbb{V}_n^{HC}\mathbb{L}^\top\right)^{-1} \left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}\right).$$

16.4.1 Heteroscedasticity consistent asymptotic inference

Consequence of Theorems 16.5 and 16.6: Heteroscedasticity consistent asymptotic inference.

Under assumptions of Theorem 16.5 and 16.6:

$$
\begin{aligned}
T_n^{HC} &\xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\, 1) \qquad \text{as } n \to \infty, \\
m\, Q_n^{HC} &\xrightarrow{\mathcal{D}} \chi^2_m \qquad \text{as } n \to \infty.
\end{aligned}
$$

16.4.1 Heteroscedasticity consistent asymptotic inference

Confidence interval for θ based on the $\mathcal{N}(0,\,1)$ distribution

$$\mathcal{I}_n^{\mathcal{N}} := \Bigl(\widehat{\theta}_n \;-\; u(1-\alpha/2)\,\sqrt{\mathbf{l}^\top\,\mathbb{V}_n^{HC}\,\mathbf{l}},\quad \widehat{\theta}_n \;+\; u(1-\alpha/2)\,\sqrt{\mathbf{l}^\top\,\mathbb{V}_n^{HC}\,\mathbf{l}}\Bigr)$$

Confidence interval for θ based on the t_{n-k} distribution

$$\mathcal{I}_n^{\mathrm{t}} := \Bigl(\widehat{\theta}_n \;-\; \mathrm{t}_{n-k}(1-\alpha/2)\,\sqrt{\mathbf{l}^\top\,\mathbb{V}_n^{HC}\,\mathbf{l}},\quad \widehat{\theta}_n \;+\; \mathrm{t}_{n-k}(1-\alpha/2)\,\sqrt{\mathbf{l}^\top\,\mathbb{V}_n^{HC}\,\mathbf{l}}\Bigr)$$

Asymptotic coverage (for any $\theta^0 \in \mathbb{R}$)

$$\mathsf{P}\bigl(\mathcal{I}_n^{\mathcal{N}} \ni \theta^0;\; \theta = \theta^0\bigr) \longrightarrow 1-\alpha \qquad \text{as } n \to \infty,$$

$$\mathsf{P}\bigl(\mathcal{I}_n^{\mathrm{t}} \ni \theta^0;\; \theta = \theta^0\bigr) \longrightarrow 1-\alpha \qquad \text{as } n \to \infty.$$

16.4.1 Heteroscedasticity consistent asymptotic inference

Confidence ellipsoid for ξ based on the χ^2_m distribution

$$\mathcal{K}_n^{\chi} := \left\{ \boldsymbol{\xi} \in \mathbb{R}^m : \ \left(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\right)^{\top} \left(\mathbb{L}\,\mathbb{V}_n^{HC}\mathbb{L}^{\top}\right)^{-1} \left(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\right) \ < \ \chi^2_m(1-\alpha) \right\}$$

Confidence ellipsoid for ξ based on the $\mathcal{F}_{m,n-k}$ distribution

$$\mathcal{K}_n^{\mathcal{F}} := \left\{ \boldsymbol{\xi} \in \mathbb{R}^m : \ \left(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\right)^{\top} \left(\mathbb{L}\,\mathbb{V}_n^{HC}\mathbb{L}^{\top}\right)^{-1} \left(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\right) \ < \ m\,\mathcal{F}_{m,n-k}(1-\alpha) \right\}$$

Asymptotic coverage (for any $\boldsymbol{\xi}^0 \in \mathbb{R}^m$)

$$\mathsf{P}\left(\mathcal{K}_n^{\chi} \ni \boldsymbol{\xi}^0;\ \boldsymbol{\xi} = \boldsymbol{\xi}^0\right) \longrightarrow 1-\alpha \qquad \text{as } n \to \infty,$$

$$\mathsf{P}\left(\mathcal{K}_n^{\mathcal{F}} \ni \boldsymbol{\xi}^0;\ \boldsymbol{\xi} = \boldsymbol{\xi}^0\right) \longrightarrow 1-\alpha \qquad \text{as } n \to \infty.$$