TMA4267 - Linear statistical models

Trond Skaret Johansen

Spring semester 2024

Contents

Introduction

This is a brief summary of the course TMA4267 about linear statistical models. It includes the main content from the lecture held by ... recorded in, where some examples etc... are excluded.

The purpose of the notes is to give a good overview of the syllabus. I intend to add summaries of the lectures as I review them. I hope to include insights from projects / exercises where it is appropriate.

§

Course progress

- First reading
  - [x] Lecture 1-21
  - [ ] Lecture 22
  - [ ] Lecture 23
  - [ ] Lecture 24
  - [ ] Lecture 25
- Skikkelig TeXing
  - [ ] Lecture 1
  - [ ] Lecture 2
  - [ ] Lecture 3
  - [ ] Lecture 4
  - [ ] Lecture 5
  - [ ] Lecture 6
  - [ ] Lecture 7
  - [ ] Lecture 8
  - [ ] Lecture 9
  - [ ] Lecture 10
  - [ ] Lecture 11
  - [ ] Lecture 12
  - [ ] Lecture 13
  - [ ] Lecture 14
  - [ ] Lecture 15
  - [ ] Lecture 16
  - [ ] Lecture 17
  - [ ] Lecture 18
  - [ ] Lecture 19
  - [ ] Lecture 20
  - [ ] Lecture 21
  - [ ] Lecture 22
  - [ ] Lecture 23
  - [ ] Lecture 24
  - [ ] Lecture 25

Keywords to know

Part 1 -

dette er en test

$\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, $\widehat{\boldsymbol{\sigma}}$, $\boldsymbol{\varepsilon}$, $\widehat{\boldsymbol{\varepsilon}}$

$\widehat{\beta}$, β, σ, $\widehat{\sigma}$, ε, $\widehat{\varepsilon}$

theorem - trace formula

Theorem 1. (Trace formula)

$$\boldsymbol{\varepsilon}(\boldsymbol{Y}^T\boldsymbol{C}\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{C}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{C}\boldsymbol{\mu}$$

Proof. TODO: □

theorem - ...

Lecture 8

Theorem 2. *$\boldsymbol{Z} \sim N(0, I)$ and $\boldsymbol{R}$ symmetric and idempotent of rank r. Then*

$$\boldsymbol{Z}^T\boldsymbol{R}\boldsymbol{Z} \sim \chi^2_r.$$

Lecture 9

Assumptions

1. $\boldsymbol{X}$ is of cull column rank
2. $E\boldsymbol{\varepsilon} = \boldsymbol{0}$
3. Homostochastic: $\mathrm{Var}(\varepsilon_i) = 0 \quad \forall i$.
4. If $\boldsymbol{X}$ is random, then 2 and 3 are conditioned on $\boldsymbol{X}$.
5. Normality of errors: $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$.

... obtain least squares estimators $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2$ of $\boldsymbol{\beta}, \boldsymbol{\sigma}^2$

Residuals ...

Parameter estimation

Two approaches: LSE and MLE ...

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{k+1}}{\arg\min} \sum_{i=1}^{n} (Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2$$

... deducing that LSE and MLE give the same result ...

...

Hat matrix

ghjfiodeifgoerjfkdworw9u0gryhj

Du fulgte ikke med nei

Lecture 10

Lecture 11

Lecture 12

questions about independence. Detour into sigma algebras etc ...

Theorem 3. *Suppose X, Y are independent random variables and that f, g are two measurable functions. Then $f(X), g(Y)$ are also independent.*

ANOVA - Analysis of variance

Theorem 4. (ANOVA decomposition) *Assuming the necesarry assumptions,*

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}_{\text{SSE}}.$$

Proof. TODO: there aint space in the margin □

R2 score . . .

Lecture 13

Fictional model

"Fictional model" using x_{ij} as response for some fixed feature j.

First scalar

Lowest exponent

$$y = c_1 p^1 + c_2 p^2 + \ldots + c_n p^n + c_{n+1} x^{n+1}$$

A bunch of different scalars

... TODO:

General F-test

TODO: important to have on yellow paper

We set up a much more general problem. Let $A \in \mathbb{R}^{r \times p}$, $r < p$, $\operatorname{rank}(A) = r$, $\boldsymbol{d} \in \mathbb{R}^d$. We test the hypothesis:

$$\mathrm{H}_0 : A\boldsymbol{\beta} = \boldsymbol{d}, \qquad \mathrm{H}_1 : A\boldsymbol{\beta} \neq \boldsymbol{d}.$$

Some special cases of this general setup are.

1. $r = 1, d = 0, A = (0, \dots, 1, \dots, 0)$ with 1 at index i, gives the test
$$\mathrm{H}_0 : \beta_i = 0, \qquad \mathrm{H}_1 : \beta_i \neq 0.$$
2. $r = 1, d = 0, A = (0, \dots, 1, \dots, -1, \dots, 0)$ with 1 at index i and -1 at index j, gives the test
$$\mathrm{H}_0 : \beta_i = \beta_j, \qquad \mathrm{H}_1 : \beta_i \neq \beta_j.$$
3. $r = k, d = \mathbf{0} \in \mathbb{R}^k, A = (\mathbf{0}, \mathrm{diag}(1)) \in \mathbb{R}^{k \times p}$, gives the test
$$\mathrm{H}_0 : \beta_i = 0 \quad \forall i \in \{1, \dots, k\}, \qquad \mathrm{H}_1 : \beta_i \neq 0 \text{ for some } i \in \{1, \dots, k\}.$$

Lecture 14

Let $\mathcal{B}$ be the space of $\boldsymbol{\beta}$ satisfying H_0. The restricted problem is:
$$\widehat{\boldsymbol{\beta}}^R = \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$
Using lagrange multipliers and a bag of tricks, we obtain:
$$\widehat{\boldsymbol{\beta}}^R = \widehat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T(\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}A^T)^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{d}).$$
Denoting $\Delta = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^R$, we find:
$$\mathrm{SSE}^R = \mathrm{SSE} + \Delta^T \boldsymbol{X}^T \boldsymbol{X} \Delta$$
... IMPORTANT: the concrete expressions for the F statistic...

We claim that the under H_0, we have
$$F = \frac{\mathrm{SSE}^R - \mathrm{SSE}/r}{\mathrm{SSE}/(n-p)} \sim F_{r,n-p}.$$
Proof. what the

□

Lecture 15

... example ...

Transformations of data

Motivation: ...

box cox transformation

variance stabilising transformation

Suppose $\mu = \boldsymbol{\varepsilon}(Y_i)$ and that $\mathrm{Var}(Y_i)$ depends on μ. ...

Lecture 16

...

Lecture 17

Suppose k covariates. Then 2^k possible models from maximal:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$

to minimal:

$$Y_i = \beta_0.$$

We want to arrive at a compromise between simplisity and goodness of fit.

1. Adjusted coefficient of determination:

$$R^2_{\text{adj}} = 1 - \frac{\text{SSE}/(n-k-1)}{\text{SST}/(n-1)}$$

2.
3.
4.

example...

Multiple hypothesis testing

motivation ...

Lecture 18

. . .

FWER = probability of at least one false positive finding

... two representations

The *Bonferrony method*

The *Šidák method*

...

example 2019

. . .

example 2020

Lecture 19

Example with three groups and their means ... rewrite to regression problem ...

Analysis of varance (ANOVA)

p treatments, samples ...

Lecture 20

... cont ... + brief on two way ANOVA

Design of experiment

Lecture 21

Two level factorial design

We suppose we have k main factors $x_1, \dots, x_k$ making up a model of the form:

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_k \boldsymbol{x}_k + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

Further we suppose a feature matrix $\boldsymbol{X}$ satisfying:

1. Each column has entries ± 1.
2. The colomns are orthogonal, i.e. $\mathbf{1^T \boldsymbol{x}_i = \sum_{i=1}^n x_{ij} = 0}$ and $\boldsymbol{x}_i T \boldsymbol{x}_j = n\delta_{ij}$.

This in particular implies that we have $\boldsymbol{X}^T \boldsymbol{X} = nI_n$. Using results from regression analysis, this significantly simplifies our estimators:

TODO: expressions

Definition 1. The *main effect* of main factor j is defined as:

$$\text{effect}_j = \text{response at high level} - \text{response at low level} = 2\beta_j.$$

The estimated effect is naturally

$$\widehat{\text{effect}_j} = \text{estimated response at high level} - \text{estimated response at low level} = 2\widehat{\beta}_j.$$

To go from this to a 2^k-design, we take into account interactions of the factors modelled as products of main factors:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \beta_{1,2} x_1 x_2 + \cdots + \beta_{k-1,k} x_{k-1} k_k + \cdots + \beta_{1,2,\dots,k} x_1 \cdots x_k.$$

We extend the design matrix accordingly, and note that we still satisfy the assumptions.

TODO: example ?

Inference about effect

Need inference about σ^2... cannot use estimator from multiple linear regression since for MLR we have $\widehat{\sigma}^2 = \frac{\text{SSE}}{n-p}$ and here $n = p$. We have to resort to one of two methods.

1. neglect some effects ... then these are normally dist ... use these as estimator ...

2. Lenth's method ...

Vi tester en sitering [1].

References

[1] test. *test bok*. Ed. by Trond. UiO, 2030.

Index