TMA4267 Linear statistical models

18. march 2025

Thea Bjørnland

TMA4267 Linear statistical models

About the course

Examination arrangement

Examination arrangement: School exam

Grade: Letter grades

Evaluation	Weighting	Duration	Examination aids
School exam	100/100	4 hours	C

Course content

~~Random vectors. Multivariate normal distribution. Multiple linear regression. Analysis of variance. Multiple hypothesis testing~~. Design of experiments.

Design of Experiments

by
John Tyssedal, NTNU

Experimentation is an old discipline, but modern design of experiments theory dates back to the pioneering work of Ronald Aylmer Fisher (1890-1962) at the Rothamsted Experimental Station, where he became a statistician in 1919. Rothamsted Experimental Station was an agricultural research institute. Fisher soon experienced the problems by trying to analyse haphazardly collected data and realized the advantage it would be to collect these in a planned and controlled manner. In 1935 he published his famous book *Design of Experiments*. Surprisingly fast his ideas found its way into industry, but there they seemed to have an obstacle for success. Agricultural experiments tend to be large in scale, having several variables with many levels for each variable that needed to be replicated. And they may take a long time to complete. Experiments in industry can be expensive, thereby cost considerations need to be taken into account. On the other side, in contrast to agriculture where one sows in the spring and harvests in the autumn, experiments in industry often give immediate response and new experiments can be planned and performed the next week. In the late 1940s George Box (1919-2013) discovered that sequential experimentation, where in each step smaller experiments with few levels for each factors were performed, much faster could bring a production process closer to optimal operational conditions. The analysis of such designs also relies more heavily on regression modelling. So despite the evolution we will start with introducing the ideas for industrial experimentation first and then move on to more classic Design of Experiments theory.

Two-level factorial designs

In the regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ the design matrix $\mathbf{X}$ has a decisive impact on how easy it is to find a good model. Especially we have seen (chapter 12.7) that if the columns in the design matrix, $\mathbf{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_k$, are orthogonal, the vector of the estimators for the coefficients is given by:

First: why not observational studies?

https://scientific-publishing.webshop.elsevier.com/research-process/observational-study-design-and-types/

https://en.wikipedia.org/wiki/Observational_study

First: why not observational studies?

Example 1:
In Norway, the rate of occurrence of newborn babies with severe conditions is greater in large hospitals than in small.
Are smaller hospitals safer?

Example 2:
Children who live in homes with many books do better in school.
Should all homes have more books?

Design of experiments

- We (the experimenters) decide the values of the independent variables (exposures, covariates)
- We randomize the experiment to avoid confounding or omitted variable bias
- We take informed decisions so that we are aware of potential biases
- In TMA4267: **2-level factorial designs**

  Using: Linear regression, design matrix with orthogonal columns, ANOVA with effect coding, estimation and hypothesis testing

Plan

- Today (18.3):
  Basic introduction to 2-level factorial experiments
  After todays lecture you should start planning your own experiment

- Thursday 20.3:
  2-level factorial experiments with interactions (plots and visualisation)

- Tuesday 25.3:
  More theory and examples

- Thursday 27.3, Tuesday 1.4, Thursday 3.4:
  Some lecturing on DOE, but also in-class help with your projects
  (come prepared to ask questions)

- You can also use exercise classes for getting help with your project

- Deadline: April 4th

TMA4267 Spring 2025 Assignment 3: Design of experiments

About

The topic is design of experiments (DOE). The purpose is to provide insight and training in planning, performing and analysing a statistical experiment, as well as to report the results.

Requirements of the hand-in

- The project can be done alone or in groups of two or three.
- The report should be no longer than 8 pages.
- The R code you use (including your collected observations) should be added to the report (but not presented as part of the report)

In short: 5 main steps of the assignment

1. Use DOE to plan an experiment with at least 3 factors each at 2 levels. You need to do a minimum of 16 single experiments (2^3 in duplicate, 2^4 or fractions of 2^5 or higher). The response needs to be continuous so that the experiment can be analysed with linear regression.
2. Plan how the collection of data can be done, focus on genuine run replicates and randomization. Discuss if you need blocking.
3. Collect your data.
4. Analyse the data.
5. Report on findings.