16th February, 2026

STK4900/9900 - Statistical Methods and Applications

Mandatory assignment 1 of 2

Submission deadline

Thursday 5th MARCH 2026, 14:30 in Canvas (canvas.uio.no).

Instructions

The assignment must be submitted as a single PDF file. If you are scanning handwritten pages, they must be clearly legible. The submission must contain your name, course and assignment number. Note that you have **one attempt** to pass the assignment. This means that there are no second attempts. All mandatory assignments in this course must be approved in the same semester, before you are allowed to take the final examination.

For the report, it is expected that you give a clear presentation with all necessary explanations. In particular, in addition to presenting all relevant numerical results and plots from an analysis, you should also discuss the conclusions you make based on them. Note also that you need to include the code used to run the analysis, either as part of the main text of the report or as an appendix (you should only include the final code, not all trial and errors). You may use any software package of your own choice, as long as you are able to answer the questions.

All aids, including collaboration, are allowed, but the submission must be written by you and reflect your understanding of the subject.

The relevant data files are available on the course web page. If you have problems accessing the data, please send an email to johanpen@math.uio.no.

Application for postponed delivery

If you need to apply for a postponement of the submission deadline due to illness or other reasons, see the part about extensions in the following page:

https://www.uio.no/english/studies/examinations/compulsory-activities/mn-math-mandatory.html

which contains general guidelines for mandatory assignments at the Department of Mathematics.

Note: To pass the assignment, the presented solutions do not have to be entirely correct, but you are expected to make a genuine effort at solving the problems (which should also be reflected by the submitted report).

Problem 1. Air Pollution

We want to study how air pollution at a measuring station at Alnabru in Oslo is related to explanatory variables such as traffic volume and meteorological conditions at the same place. Air pollution is measured by the concentration of NO2 particles. The considered data set consists of 500 observations collected by the Norwegian Public Roads Administration (Vegvesenet) and it is available at:

https://www.uio.no/studier/emner/matnat/math/STK4900/data/no2.txt

Each row (except the first) represents an observation of measurements over the following variables:

`log.no2`: logarithm of the concentration of NO2,

`log.cars`: logarithm of the number of cars per hour,

`temp`: temperature 2 meters above the ground (degrees C),

`wind.speed`: wind speed (meters/second),

`hour.of.day`: hour of the day the measurements were collected (1-24).

(a) Report the main features of variables `log.no2` and `log.cars` by numerical summaries and plots. Make a scatterplot with `log.cars` on the x-axis and `log.no2` on the y-axis. What do you see?

(b) Fit a simple linear model where `log.no2` is explained by `log.cars`. Give an interpretation of the estimated coefficients and construct a plot with the observations and the fitted line. Explain what the R^2 measure tells you.

(c) Check various residual plots to judge if the model assumptions for the model in (b) seem reasonable.

(d) Use multiple regression to study the simultaneous effect of the various predictors on `log.no2`, and find an "optimal" model for modelling `log.no2` based on the available set of measured predictors. You do not have to consider interactions, but check if you should transform

some of the variables, that is, in addition to `log.cars` which is already log-transformed.

(e) For the model you have chosen in (d), write an interpretation of the model coefficients and check if the model assumptions seem reasonable through various plots. Remember to comment all plots you include in your report.

Problem 2. Blood Pressure

We want to study how blood pressure is related to age. For this, we are going to use a data set of blood pressure measurements for 12 randomly sampled men for three different age groups:

Group 1: 30–45 years,

Group 2: 46–59 years,

Group 3: 60–75 years.

The considered data set thus consists of 36 observations and it is available at:

https://www.uio.no/studier/emner/matnat/math/STK4900/data/blood.txt

The first column `Bloodpr` represents the blood pressure measurement and the second column `age` represents the different age categories as defined above.

(a) Describe the data using boxplots and numerical summaries. From these, does it seem that blood pressure is varying across age groups?

(b) Use one-way ANOVA to answer the question in (a). Specify the hypotheses and assumptions of the considered test. Write a summary of your findings.

(c) Formulate this problem using a regression model with age group as a categorical predictor. Use treatment-contrast and the youngest group as reference. Run the analysis, interpret the results and write a conclusion. Compare with the analysis in (b).

Good luck!