Exam info:

The exam is based on all material that was discussed in class. So I recommend to use your class notes (the slides and the notes you took from what I wrote on the board) as the basis for your exam preparation. The exam will be a mix of practical and theoretical problems. There will be at least one question that is taken from the exercise sets. You should be able to understand and explain the R-code that was used in the class. You don't have to produce R-code yourself during the exam.

Below is a list of topics that are likely to come up at the exam. This list is not complete, but is meant to give you a better idea of what type of questions you can expect. You should be able to:

- Introduction
  - Describe the role of statistical models.
  - Describe the difference between observational and experimental studies.
  - Give an example in which association does not mean causation.
  - Describe simple ways to perform nonparametric regression, explain the idea behind the loess smoother, explain what is meant by the 'curse of dimensionality'.
  - Mention appropriate plots to examine data.
  - Mention advantages/disadvantages of the use of transformations. You should also be able to choose appropriate transformations (with a bit of trial and error).
- Multiple linear regression
  - Mention the assumptions for linear regression, including what each assumption is needed for.
  - Interpret the parameters of a regression model in a given example.
  - Explain the difference between fitting several separate simple regression models and one large multiple regression model.
  - Derive the normal equations and parameter estimates.
  - Describe the geometrical interpretation of regression.
  - Prove that (under some assumptions) the LS estimator equals the MLE.
  - Describe the distribution of the parameter estimates and test statistics (including all proofs we did related to sections 2.4 and 2.5 of the script). Also explain how the distribution of the parameter estimates depends on σ, the sample size n, the variance of independent variables, the center of the distribution of independent variables, correlation between independent variables.
  - Describe measures of model fit (SSE, $\hat{\sigma}$, R^2, adjusted R^2) and explain the idea behind these methods.
  - Perform tests (t-test and F-test), and interpret the results.
  - Construct confidence intervals, confidence bands and prediction intervals, and interpret these.

  - Use and interpret standardized regression estimates.
  - Incorporate categorical predictors in a regression model, including interpretation and inference.
  - Give the definition of interaction, and describe the difference between interaction and correlation.
  - Incorporate interaction regressors in a regression model, including interpretation and inference.
  - Explain weighted regression: why is it used, how is it used?
  - Describe, interpret and prove the two versions of the Gauss-Markov theorem.
  - Describe appropriate methods for model diagnostics. In particular, you should be able to draw and recognize Tukey-Anscombe plots for data sets that violate the assumptions of linearity or constant variance.
  - Describe outliers, leverage points, and influential data points, and methods to identify these. Describe ways to deal with such points.
  - Describe methods for model selection / variable selection.
  - Describe issues that may arise when data are aggregated, and give examples of such situations (ecological correlation).
  - Interpret all output from the R summary table of a linear model.

- Nonlinear least squares
  - Describe methods of robust regression, know advantages/disadvantages of these methods.
  - Describe and interpret logistic regression.
  - Describe and interpret Cox regression.