Questions for the individual oral examination, Spring 2024
MASM22/FMSN30/FMSN40: Linear and Logistic Regression (with Data Gathering)

A. Compulsory questions for all grades

In order to Pass you must answer these questions.

Models and assumptions

- State the model, including the assumptions, for multiple *Linear* and for multiple *Logistic* regression and interpret the parameters.

  Includes stating the assumptions for the *x*-variables, the distribution of Y, how the linear predictor $\mathbf{x}\beta$ enters the distribution and how the random variability is assumed to behave.

Parameter estimation

- Describe the principle behind the estimation methods used for estimating the β-parameters in *Linear* and in *Logistic* regression, respectively.

- Write down the relevant Residual sum of squares / Likelihood function and solve as far as the normal equations. State the expressions for the resulting β-estimates. Estimate σ in *Linear* regression.

B. Additional questions for grades VG, 4, 5

In order to get a higher grade than G/3 you must first answer the compulsory questions above, and then a random sample of the questions below.

Properties of parameter estimates

1. Derive the distribution of the β-estimates in *Linear* regression. State the (asymptotic) distribution of the β-estimates in *Logistic* regression.
   Also define and describe the use of t-test and Wald test and construct the corresponding confidence interval for β_j.

The linear predictor

2. Use the distribution of the β-estimates to derive the (asymptotic) distribution of the linear predictor $\mathbf{x}_0\hat{\beta}$ and construct a confidence interval for $\mathbf{x}_0\beta$.
   Linear: Construct prediction intervals for new observations and describe the conceptual difference between a confidence interval and a prediction interval.
   Logistic: Construct confidence intervals for probabilities.

Sums of Squares and log-likelihood

3. *Linear:* Perform a variance decomposition of the sum of squares and state the idea, definition and use of global and partial F-tests.
   Logistic: State the idea, definition and use

of likelihood ratio tests, including the partial likelihood based confidence interval for β_j.

Residuals and Influence

4. Define the hat matrix and the leverage of individual observations and describe their interpretation.
   Also define Cook's distance and describe its use.
5. *Linear:* Derive the distribution of the residual vector, $\mathbf{e}$, and define and motivate standardized och studentized residuals and describe their use.
   Logistic: Define and motivate Pearson and deviance residuals and describe their use.

Model selection and Goodness of fit

6. State the idea, definition and use of
   - VIF,
   - R^2, R^2_{adj}, R^2_{McFadden}, $R^2_{\text{McFadden, adj}}$,
   - AIC and BIC.

   Logistic: Define sensitivity and specificity and describe their use.
   Also, describe the idea behing the ROC-curve and AUC and their use.