

QUESTIONS FOR ORAL EXAMINATION
MASM22/FMSN30(N40): LINEAR AND LOGISTIC REGRESSION (WITH
DATA GATHERING), 2019

A. Model and assumptions (compulsory)

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 $X\beta$ enters the distribution and how the random variability is assumed to behave.

B. Parameter estimation (compulsory)

Describe the principle behind the estimation method used for estimating the β -parameters in Linear and in Logistic regression.

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

C. Properties of the parameter estimates (random selection)

(a) State the distribution of the β -estimates.

(b) + (a) *Linear*: Derive the distribution of the β -estimates.

(c) + (a) Define and describe the use of the t-test/Wald test.

(d) + (a) Construct a confidence interval for β .

D. The linear predictor (random selection)

(a) State the (asymptotic) distribution of the linear predictor $X\hat{\beta}$.

(b) + (a) Derive the (asymptotic) distribution of the linear predictor $X\hat{\beta}$.

(c) + (a) Construct confidence intervals for the linear predictor $X\beta$.

(d) + (a) *Linear*: Construct prediction intervals for new observations and describe the conceptual difference between a confidence interval and a prediction interval.

(e) + (a) *Logistic*: Construct confidence intervals for probabilities.

E. Sums of Squares/Deviance (random selection)

(a) ANOVA decomposition of the sum of squares.

(b) Define the Deviance.

(c) + (a) *Linear*: The idea, definition and use of global and partial F-tests.

(d) + (b) *Logistic*: The idea, definition and use of likelihood ratio tests.

(e) + (a) + (b) The idea, definition and use of R^2 , R^2_{adj} / $R^2_{\text{Cox-Snell}}$, $R^2_{\text{Nagelkerke}}$.

(f) + (a) + (b) The idea, definition and use of AIC/BIC.

F. Residuals (random selection)

(a) Define the hat matrix and the leverage of individual observations.

(b) + (a) Define and motivate the different types of residuals (ordinary, standardized, studentized / Pearson, deviance) and describe their uses.

(c) + (a) *Linear*: Derive the distribution of the (ordinary) residuals.

G. Influence (random selection)

(a) Define Cook's distance and describe its use.

(b) Define df_{β} and describe their use.