Practical "Logistic regression"
P7

M. Hauptmann

Exercise 1 Consider an influenza vaccination trial with the following data.

	Influenza Yes	Influenza No	
Vaccine	20	220	240
Placebo	80	140	220

Use a simple calculator and determine the odds for influenza in each group (vaccine and placebo) and the odds ratio (OR) without confidence interval. How do you interpret the OR?

Exercise 2 Dataset `trismus.sav` includes head and neck cancer patients treated with chemoradiation. These patients are at risk for developing trismus (reduced mouth opening <35 mm). Trismus is a side-effect of treatment that impairs eating, speech and oral hygiene. Mouth opening is controlled by the bilateral masseter muscle (left and right). For each patient, information is available on the mean radiotherapy dose to the ipsilateral muscle (side of the tumor, i.e., the muscle receiving the highest dose) and to the contralateral muscle (i.e., the muscle receiving the lowest dose). The objective of this exercise is to evaluate whether the dose delivered to the masseter muscle is predictive for developing trismus.

(a) Construct a frequency table of the variable `trismus` (1=trismus present). What is the probability of developing trismus for a patient based on this table?
SPSS tip: Click Analyze – Descriptive Statistics – Frequencies

(b) Is the mean dose to the ipsilateral masseter muscle significantly associated with trismus? What is the OR? Does the 95% confidence interval (CI) around the OR include 1? Test the goodness of fit.
SPSS tip: Click Analyze – Regression – Binary logistic. Select dependent and independent variable and request 95% CIs and the Hosmer-Lemeshow test under "Options". In the output, you can find EXP(B), the estimate of the OR, in the table "Variables in the equation" (at the bottom of the output).

(c) Is the mean dose to the contralateral masseter muscle significantly associated with trismus?

(d) Does a model with mean dose to the ipsilateral masseter significantly improve when mean dose to the contralateral masseter is added?
SPSS tip: Select both dose variables and make sure the "Method:" is set to "Enter".

(e) Check how correlated the two doses are. Do they provide independent information?
SPSS tip: Click Analyze – Correlate – Bivariate

(f) Fit again a model with dose to the ipsilateral masseter as covariate, but this time save the predicted probabilities to the database. Plot the dose to the ipsilateral masseter against the predicted probabilities of trismus. Do only the trismus cases have higher estimated probabilities of trismus?
SPSS tip: Define the model as before, click on "Save" and select "Probabilities" under "Predicted Values". Predicted probabilities will be added to your database (last column). Make a scatter plot with ipsilateral dose on the horizontal axis and the predicted probabilities on the vertical axis (Click Graphs - Legacy Dialogs – Scatter/Dot – Simple Scatter and select variable `trismus` at "Set markers by").

(g) Construct a formula for the probability of trismus from the estimated regression coefficients. Calculate the probability of trismus for a patient who received a mean dose of 75 Gy to the ipsilateral trismus. Do you find this a reliable estimate? Compare with the plot you made and explain your conclusion.

(h) Explore the data using linear regression by modelling the relative reduction in mouth opening as an outcome (variable `diffmouth_pct`) and compare the results with those from logistic regression.
SPSS tip: Refer to S5 and S6.