

Practical "Logistic regression"

P7

M. Hauptmann

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

	Influenza		
	Yes	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 1 – Suggested answer The odds for influenza among patients who received the vaccine was $(20/240)/(1-20/240)=20/220=0.09$. In gambling, this would be expressed as an odds of 1 to 11 (20 to 220), like a probability of 1 in 3 would be expressed as an odds of 1 to 2. For patients who received placebo, the odds was $80/140=0.57$, or 4 to 7. The OR for vaccine versus placebo is $20*140/(80*220)=0.16$. Therefore, patients who received the vaccine had a $(1-0.16)*100\%=84\%$ reduced odds for influenza compared with placebo, i.e., their odds was $1/0.16=6.25$ -fold lower.

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.

- 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
- 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).
- Is the mean dose to the contralateral masseter muscle significantly associated with trismus?
- 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".
- Check how correlated the two doses are. Do they provide independent information?
SPSS tip: Click Analyze – Correlate – Bivariate
- 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 and compare the results with those from logistic regression.
SPSS tip: Refer to S5 and S6.

Exercise 2 – Suggested answer

- (a) Prepare a frequency table for variable `trismus`.

```
FREQUENCIES VARIABLES=Trismus
/ORDER=ANALYSIS.
```

Post-Treatment Trismus (mouth opening < 35 mm)

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	82	83.7	83.7	83.7
1	16	16.3	16.3	100.0
Total	98	100.0	100.0	

The table indicates that 16 of 98 patients had trismus, which corresponds to an estimated probability of $16/98=0.16$, i.e., 16%.

- (b) Select dependent and independent variable and request 95% CIs and the Hosmer-Lemeshow test under "Options".

```
LOGISTIC REGRESSION VARIABLES Trismus
/METHOD=ENTER IM_mean
/PRINT=GOODFIT CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a IM_mean	.078	.024	10.158	1	.001	1.081	1.031	1.134
Constant	-4.221	.951	19.719	1	.000	.015		

a. Variable(s) entered on step 1: IM_mean.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7.327	8	.502

Contingency Table for Hosmer and Lemeshow Test

		Post-Treatment Trismus (mouth opening < 35 mm) = 0		Post-Treatment Trismus (mouth opening < 35 mm) = 1		Total
		Observed	Expected	Observed	Expected	
Step 1	1	10	9.749	0	.251	10
	2	10	9.640	0	.360	10
	3	10	9.468	0	.532	10
	4	9	9.306	1	.694	10
	5	8	8.964	2	1.036	10
	6	9	8.467	1	1.533	10
	7	9	7.974	1	2.026	10
	8	5	7.548	5	2.452	10
	9	8	6.855	2	3.145	10
	10	4	4.029	4	3.971	8

The mean dose to the ipsilateral masseter muscle is significantly positively associated with trismus. The OR is 1.08 with a 95% CI from 1.03 to 1.13, which does not include 1.0. The p-value of the estimated OR is 0.001. There is no evidence that the model does not fit the data well, based on the Hosmer-Lemeshow p-value of 0.502. The results indicate that in this population the odds of trismus increased by 8% per Gray radiation dose to the ipsilateral masseter muscle.

- (c) Fit the same model as in (b) but with the dose to the ipsilateral masseter replaced by the contralateral.

LOGISTIC REGRESSION VARIABLES Trismus

/METHOD=ENTER CM_mean

/PRINT=GOODFIT CI(95)

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a CM_mean	.076	.025	8.963	1	.003	1.079	1.027	1.134
Constant	-3.466	.732	22.420	1	.000	.031		

a. Variable(s) entered on step 1: CM_mean.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	13.803	8	.087

Contingency Table for Hosmer and Lemeshow Test

		Post-Treatment Trismus (mouth opening < 35 mm) = 0		Post-Treatment Trismus (mouth opening < 35 mm) = 1		Total
		Observed	Expected	Observed	Expected	
Step 1	1	10	9.567	0	.433	10
	2	10	9.448	0	.552	10
	3	10	9.315	0	.685	10
	4	10	9.124	0	.876	10
	5	8	8.843	2	1.157	10
	6	8	8.611	2	1.389	10
	7	5	8.199	5	1.801	10
	8	9	7.742	1	2.258	10
	9	9	7.129	1	2.871	10
	10	3	4.022	5	3.978	8

The mean dose to the contralateral masseter muscle is also significantly positively associated with trismus. The OR and 95% CI are the same as for the ipsilateral masseter. The Hosmer-Lemeshow test suggests a somewhat poorer goodness of fit but the p-value is above 0.05.

(d) Code and results for a model with both dose variables are as follows.

```
LOGISTIC REGRESSION VARIABLES Trismus
/METHOD=ENTER CM_mean IM_mean
/PRINT=CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a CM_mean	.038	.030	1.696	1	.193	1.039	.981	1.101
IM_mean	.058	.029	3.899	1	.048	1.060	1.000	1.123
Constant	-4.519	1.031	19.202	1	.000	.011		

a. Variable(s) entered on step 1: CM_mean, IM_mean.

Compared with the model with only dose to the ipsilateral masseter, adding contralateral dose slightly attenuates the effect of ipsilateral dose from 1.08 to 1.06, but it is still (borderline) significant (p=0.048). Contralateral dose does not have a significant effect in presence of ipsilateral dose (p=0.193), i.e., adding contralateral dose does not significantly improve the goodness of fit.

(e) Calculate the Pearson correlation coefficient between both dose variables.

```
CORRELATIONS
/VARIABLES=CM_mean IM_mean
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

Correlations

		Mean dose Contralat Masseter	Mean dose Ipsilateral Masseter
Mean dose Contralat Masseter	Pearson Correlation	1	.751**
	Sig. (2-tailed)		.000
	N	98	98
Mean dose Ipsilateral Masseter	Pearson Correlation	.751**	1
	Sig. (2-tailed)	.000	
	N	98	98

** . Correlation is significant at the 0.01 level (2-tailed).

The correlation is 0.75, i.e., quite high. Both variables provide more or less the same information, i.e., when together in the model only one of them retains its univariately significant effect.

(f) Add the predicted probabilities to the database and make the scatter plot

LOGISTIC REGRESSION VARIABLES Trismus

 /METHOD=ENTER IM_mean

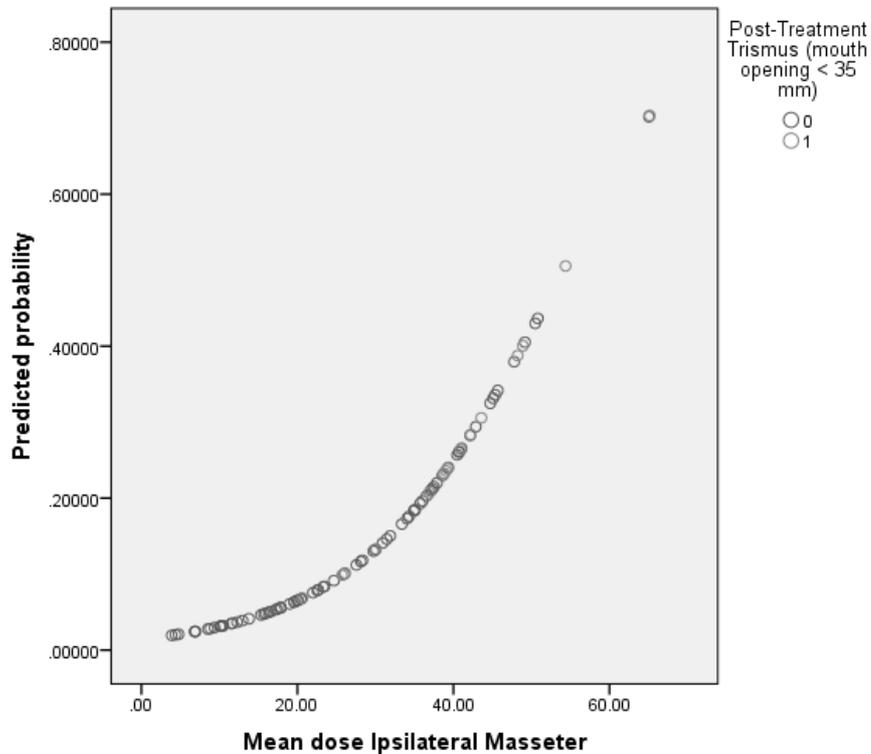
 /SAVE=PRED

 /CRITERIA=PIN(.05) POUT(.05) ITERATE(20) CUT(0.5).

GRAPH

 /SCATTERPLOT(BIVAR)=IM_mean WITH PRE_1 BY Trismus

 /MISSING=LISTWISE.



The plot shows that there are cases and non-cases with high and low predicted probabilities of trismus, although the proportion of cases appears to increase at higher dose levels.

- (g) Look at the "Variables in equation" table in the output. The predicted probability can be calculated as:

$$\text{Probability}(\text{Trismus}=1) = \frac{\exp(-4.22 + 0.078 * \text{ipsilateral dose})}{1 + \exp(-4.22 + 0.078 * \text{ipsilateral dose})}$$

For an ipsilateral dose of 75 Gray, the probability of trismus is $\frac{\exp(1.63)}{1 + \exp(1.63)} = 0.84$. This estimate has to be interpreted very carefully since a dose of 75 Gray is far beyond the range of the data from which the association is derived.

- (h) Linear regression was explained in S5 and S6. Use outcome variable `diffmouth_pct` with methods and code provided there to explore this alternative approach.