Exercises: Generalized Linear Models

Exercise 1: Exam 22/02/2024

The `titanic` dataset is a collection of data about 714 passengers, and the goal is to predict the survival (`Survival`: 1 if the passenger survived, 0 if they did not) based on some personal characteristics. In particular, here we consider the ticket class (`Class`: 1 = first, 0 = second or third; dummy), the gender (`Gender`: man = 1, woman = 0; dummy), and the age (`Age`, in years). Fitting a logistic regression model in R produces the following summary:

	Estimate	Std. Error	z value	Pr(>|z|)
`(Intercept)`	1.5003	0.2462	6.09	0.0000
`Class`	2.0103	0.2479	8.11	0.0000
`Gender`	-2.5473	0.2017	??	0.0000
`Age`	-0.0299	0.0074	-4.06	??
Null deviance:	964.52			
Residual deviance:	675.14			

a) Write the corresponding theoretical model and the expression of the estimated model.

b) Write the likelihood and log-likelihood function.

c) Complete the missing values in the table. For Pr(>|z|) of `Age`, write an approximate value. What variables are statistically significant?

d) Provide an estimate of the odds for a woman aged 30 with a ticket of first class (denote this individual as “A”). How do you expect this value to change if you consider a person with the same characteristics but aged 31 (denote this individual as “B”)?

e) Provide the interpretation of the coefficient associated with the `Class` variable. Given the estimate of this coefficient, what is the effect of this covariate on the survival probability?

f) Perform a test $H_0 : \beta_{\text{class}} = 0$ vs $H_1 : \beta_{\text{class}} < 0$.

g) Define the “null deviance” and “residual deviance” in the output.

h) Perform a test about the overall significance.

Exercise 2 - Exam practice

Given a set of $n = 30$ observations, consider fitting the model $Y_i \sim \text{Bernoulli}(\pi_i)$ where $\text{logit}(\pi_i) = \beta_1 + \beta_2 x_i$, with x_i is a dummy variable taking value 1 for the first 10 observations and 0 otherwise. Fitting this model returns the following output

	Estimate	Std. Error	z value	Pr(>|z|)
`(Intercept)`	1.3863	0.5590	2.480	0.01314
`x`	-2.0794	0.7826	-2.657	0.00788

Null deviance = 47.111
Residual deviance = 39.112

Answer the following:

a) Write the likelihood, log-likelihood and score functions for (β_1, β_2). Write the fitted model.

b) Compute the estimate of the probability $\hat{\pi}$ for $x = 0$ and $x = 1$. Obtain the odds for $x = 0$ and $x = 1$ and interpret them. Give an estimate of the odds ratio and interpret it.

c) Test the hypothesis $H_0 : \beta_2 = -1$ vs $H_1 : \beta_2 < -1$.

d) What are the two quantities "Null deviance" and "Residual deviance"?

Exercise 3: Exam 03/09/2024

Consider an experiment to study the resistance to the tension of a machine component. The dataset studies how many breaks occurred during 54 replications of the experiment for two types of material (A and B) and different levels of tension (L = low; M = medium; H = high).
To study such relationship we fit a Poisson regression model. The output of the model is the following:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.6920	0.0454	81.30	0.0000
material B	-0.2060	0.0516	-3.99	0.0001
tension M	-0.3213	0.0603	-5.33	0.0000
tension H	-0.5185	0.0640	-8.11	0.0000

Null deviance: 297.37 on 53 degrees of freedom
Residual deviance: 210.39 on 50 degrees of freedom

a) Write the model formulation and assumptions.

b) Derive and explain the interpretation of the coefficient associated with the variable "material B".

c) A second model ("model B") assumes that the type of material and the level of tension do not have an impact on the number of breaks. Specify the model and perform a test to compare the model fitted in point (a) with model B.

				p			
	0.90	0.95	0.975	0.99	0.995	0.9975	0.999
z_p	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902

Table 1: Some quantiles of the Gaussian distribution: $p = \mathbb{P}(Z \leq z_p)$. Columns correspond to probabilities p.

	0.9	0.95	0.975	0.99	0.995	0.9975	0.999
$\chi^2_{1;p}$	2.7055	3.8415	5.0239	6.6349	7.8794	9.1406	10.8276
$\chi^2_{2;p}$	4.6052	5.9915	7.3778	9.2103	10.5966	11.9829	13.8155
$\chi^2_{3;p}$	6.2514	7.8147	9.3484	11.3449	12.8382	14.3203	16.2662
$\chi^2_{4;p}$	7.7794	9.4877	11.1433	13.2767	14.8603	16.4239	18.4668
$\chi^2_{5;p}$	9.2364	11.0705	12.8325	15.0863	16.7496	18.3856	20.515

Table 2: Some quantiles of the χ^2 distribution: $p = \mathbb{P}(X \leq \chi^2_{df;p})$ with $X \sim \chi^2_{df}$. Columns correspond to probabilities p. Rows correspond to different degrees of freedom df.