Generalized Linear Models
Models with Binary data

1. Data from Dobson (1990) show the number of bugs dead after five hours exposure to gaseous carbon disulphide at various concentrations.
A. Make a plot of death proportions versus toxic concentrations. Make a plot of empirical logits of death probabilities versus toxic concentrations. What are your conclusions?
B. Propose a transformation to toxic concentrations for better interpretation of the finally fitted model. Fit a model with the linear effect of toxic concentrations. Check its goodness-of-fit.
C. Provide interpretations for estimated model parameters.
D. Check model's goodness-of-fit using Pearson residuals.
E. Fit a new model by adding a quadratic term. Check its statistical significance. Check model's goodness-of-fit.
F. Use again Pearson residuals for visual inspection of model's goodness-of-fit.
G. Predict the death probability of a bug when is exposed to toxic concentrations 1.7709, 1.8403 and 1.8865.

Toxic concentrations	Number of bugs	Number of deaths
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

2. The following data from Dalal, Fowlkes and Hoadley (1989) show the number of O-ring failures in the 23 pre-Challenger space shuttle launches. Challenger was the shuttle that blew up on take off in 1986. Temperature is the predictor variable. The Challenger explosion occurred during takeoff at 31 degrees F. Each flight should be viewed as an independent trial. Result 1 means failure and 0 if the O-ring functioned properly.

Πτήση	Αποτέλεσμα	Θερμοκρασία
14	1	53
9	1	57
23	1	58
10	1	63
1	0	66
5	0	67
13	0	67
15	0	67
4	0	68
3	0	69
8	0	70

17	0	70
2	1	70
11	1	70
6	0	72
7	0	73
16	0	75
21	1	75
19	0	76
22	0	76
12	0	78
20	0	79
18	0	81

A. Use a suitable GLM model and link the probability of failure with temperature. Test the significance of the linear effect of temperature on the probability of failure.
B. Predict the probability of failure at the temperature that the Challenger disaster took place.
C. Can you check model's goodness-of-fit using Pearson residuals?
D. Group temperature to a new variable with two levels which express temperature values as low (temperatures <70) and high (temperatures >=70). Fit a model to test the effect of the new variable. Do your conclusions change as compared to those obtained by fitting the model in (A.)?

3. The number of deaths by leukemia or other cancers classified by radiation dose received from the Hiroshima atomic bomb. The data refer to deaths during the period 1950-1959.
A. Obtain a suitable model to describe the relationship between radiation dose and the proportion of cancer mortality rates for leukemia.
B. Examine model's goodness-of-fit.
C. Interpret obtained model parameters.

	Radiation dose (rads)					
Deaths	**0**	**1-9**	**10-49**	**50-99**	**100-199**	**200+**
Leukemia	13	5	5	3	4	18
Other cancers	378	200	151	47	31	33

4. A sample of elderly people was given a psychiatric examination to determine whether symptoms of senility were present (s=1 or s=0 if symptoms were absent). WAIS score of intelligence were also obtained (x).

x	s	x	s	x	s	x	s	x	s
9	1	7	1	7	0	17	0	13	0
13	1	5	1	16	0	14	0	13	0
6	1	14	1	9	0	19	0	9	0
8	1	13	0	9	0	9	0	15	0
10	1	16	0	11	0	11	0	10	0
4	1	10	0	13	0	14	0	11	0

14	1	12	0	15	0	10	0	12	0
8	1	11	0	13	0	16	0	4	0
11	1	14	0	10	0	10	0	14	0
7	1	15	0	11	0	16	0	20	0
9	1	18	0	6	0	14	0		

A. Make a plot of the proportion of senility symptoms for each unique value of WAIS score.
B. Fit a GLM using the logit link, considering the linear effect of WAIS score on the probability of senility symptoms.
C. Test the model's goodness-of-fit using both the scaled deviance and the Pearson's ChiSquare.
D. Use Hosmer-Lemeshow test for testing the model's goodness-of-fit. What are your conclusions?
E. Use Pearson residuals for testing goodness-of-fit.
F. Measure the model's predictive value using a ROC curve.

5. In a cutting-edge experiment medical investigators tried to associate the probability of heart failure with the average time of snoring per night. They used 2484 subjects which were classified, after 50 observation years, according to having heart failure and the average snoring hours pre night.

	Average snoring hours per night			
	0 hours	2 hours	4 hours	5 or more hours
Heart failure	24	35	21	30
No heart failure	1355	603	192	224

A. Plot empirical logits of the probability of heart failure against the average snoring hours per night.
B. Fit a logistic regression model considering the linear effect of average snoring hours. Provide model parameters interpretations.
Γ. Check model's goodness-of-fit by using the scaled deviance and the Pearson residuals.
Δ. Predict the probability of heart failure when someone has an average snoring for 1, 2 or 3 hours per night.

6. In a longitudinal study, a sample of 3182 people without cardiovascular disease were cross-classified by three factors: personality type (type A and B), cholesterol level (normal or high) and diastolic blood pressure (normal or high). Type A persons show signs of stress uneasiness and hyperactivity. Type B persons are relaxed, easygoing and normally active. We are interested into the probability of high blood pressure.

		Diastolic Blood Pressure		
Personality type	Cholesterol	Normal	High	
A	Normal	716	79	
	High	207	25	
B	Normal	819	67	
	High	186	22	

A. Fit a model which assumes the effect of personality type on the probability of high blood pressure.
B. Fit a model which assumes the effect of cholesterol on the probability of high blood pressure.
C. Fit a model which assumes the effect of personality type and cholesterol with no interaction.

D. Fit the saturated model.
E. Which model seems to best describe the data? Interpret its estimated coefficients.