

Instructions: Attempt all questions. Short and specific answers are preferred. Given explanation when required, but keep it as short and simple as possible. Give only one answer to each question – if you give alternative answers, the worst answer will be graded.

Question A

The data shown in the table below come from a survey of 661 low birth weight children in Scotland from 1981 to 1988. The variables are:

- Cardiac:** mild heart problems of the mother during pregnancy,
Comps: gynaecological problems during pregnancy,
Smoking: mother smoked at least one cigarette per day during pregnancy,
BW: birth weight less than 1250g.

Cardiac		yes				no			
Comps		yes		no		yes		no	
Smoking		yes	no	yes	no	yes	no	yes	no
BW	yes	10	25	12	15	18	12	42	45
	no	7	5	22	19	10	12	202	205

- (1) (2 pts) A Poisson generalized linear model was fit to this data with the counts as the response and all four factors as predictors but with no interaction terms. This model corresponds to a particular hypothesis about the relationship between the factors. What is that hypothesis?
- (2) (2 pts) The residual deviance of the model mentioned in the previous question is 973. How many degrees of freedom does it have and should the hypothesis mentioned in the previous question be rejected? Explain.
- (3) (2 pts) The largest Pearson residual in the model described above had an absolute value of 14.6. Can we say if this is particularly large or is there no way of telling from the information given? Explain.

The half of the data corresponding to mothers with heart problems was fit using a Poisson GLM and the following sequential analysis of deviance table obtained. You may assume that the order of testing in the table makes no difference.

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi)
NULL			7	26.04	
bw	1	0.71	6	25.34	0.40
comps	1	3.86	5	21.48	0.05
smoking	1	1.47	4	20.01	0.22
bw:comps	1	13.95	3	6.06	0.00019
bw:smoking	1	4.30	2	1.76	0.04
comps:smoking	1	0.68	1	1.08	0.41

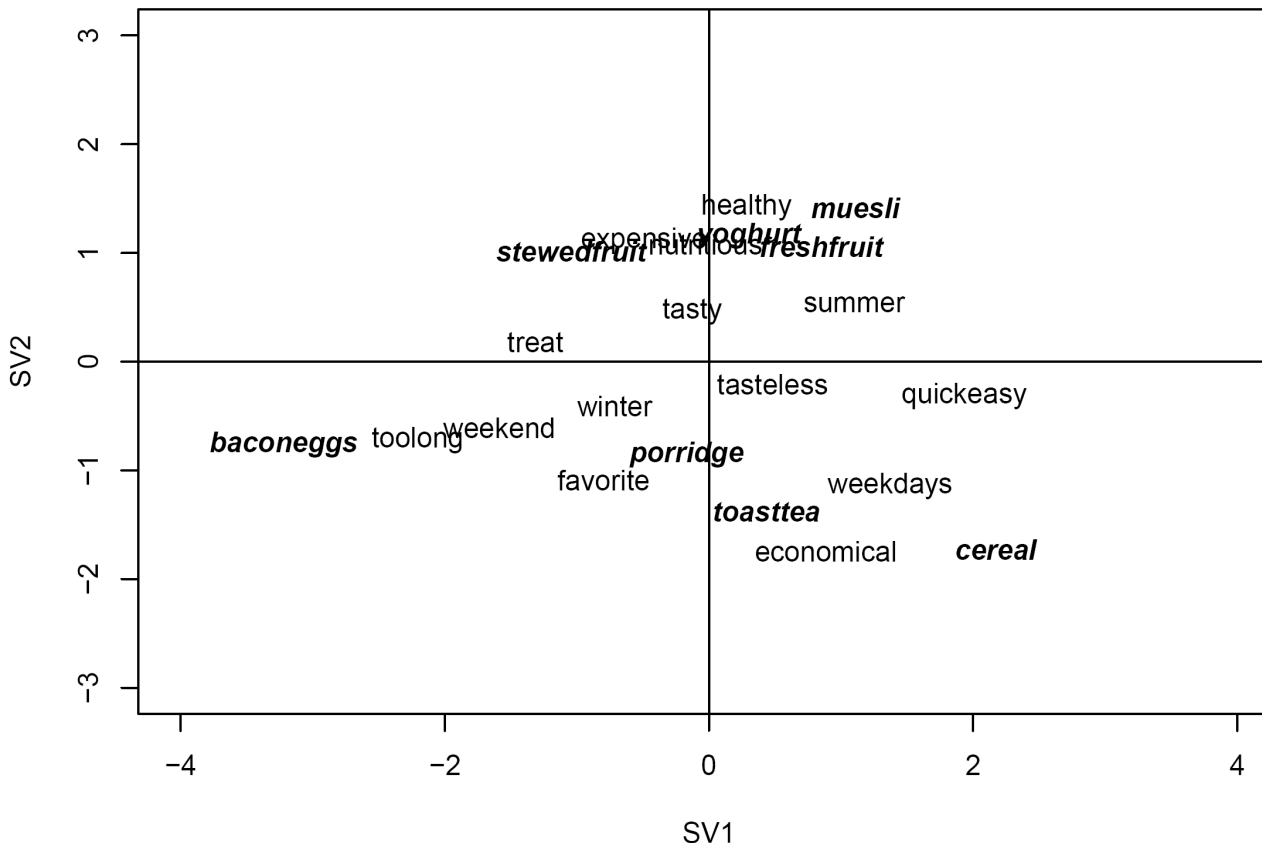
- (4) (2 pts) Describe qualitatively the nature of the relationship between the three factors.
- (5) (2 pts) In the output above, the contribution of the main effect of *bw* is not statistically significant. What does this tell us about the distribution of this variable?
- (6) (2 pts) Suppose that you wanted to assess the effect of the other three factors on whether the birth weight was low or not. Describe the model you would fit.

Question B

In a marketing study of attitudes to breakfast foods, 100 consumers were asked to associate 8 food items (including cereal, muesli, porridge, bacon & eggs, toast & tea, fresh fruit, stewed fruit, yoghurt) with 14 opinion statements (including healthy, nutritious, for summer, for winter, expensive, quick & easy, tasty, economical, for a treat, for weekdays, for weekends, tasteless, too long to prepare, family favorite). People were allowed to make more than one association. A total of 1760 associations were made. A test for independence between opinion and food was performed and rejected so that there is some association between them. A correspondence analysis was made and the result is plotted in the Figure 1.

- (7) (2 pts) Write out the model that was used to calculate the residuals for the correspondence analysis.
- (8) (2 pts) For which food are opinions most different from the rest? Explain.
- (9) (2 pts) Which food is relatively least likely to be described as “quick and easy”? Explain.
- (10) (2 pts) The opinion “tasteless” occurs close to the origin of the plot. What does this indicate about this opinion?

Figure 1



Question C

Data were collected on Broadway shows to determine what factors affect the run of the show. Summary information is shown below:

Type	Revival	Week1	run
Musical :33	No :54	Min. : 29.4	Less than 6 months:68
Musical revue: 7	Yes:41	1st Qu.: 62.4	6-12 months :11
Play :55		Median : 80.9	1-2 years :11
		Mean : 75.9	2-3 years : 1
		3rd Qu.: 91.8	More than 3 years : 4
		Max. :101.6	
		NA's : 4.0	

Week1 indicates the percentage of seats sold for the first week of performance. A proportional odds model was fit with the runs as the response and the following output obtained:

Coefficients:

	Value	Std. Error	t value
TypeMusical revue	0.340005	0.853489	0.39837
TypePlay	-1.595308	0.529629	-3.01212
RevivalYes	-0.718951	0.540672	-1.32974
Week1	0.049819	0.017694	2.81551

Intercepts:

	Value	Std. Error	t value
Less than 6 months 6-12 months	3.618	1.358	2.664
6-12 months 1-2 years	4.431	1.391	3.186
1-2 years 2-3 years	5.979	1.501	3.983
2-3 years More than 3 years	6.244	1.526	4.090

Residual Deviance: 153.42

AIC: 169.42

- (11) (2 pts) How can we interpret the values of the intercepts in a proportional odds model from the viewpoint of a latent continuous variable? Explain.
- (12) (2 pts) Only one show lasted between two and three years. Explain why it would be reasonable to combine this category with more than three years. Support your reasoning using information from the model output, rather than relying solely on the small number of observations.
- (13) (2 pts) Can musical revues be expected to perform more like musicals or plays if the purpose is to predict how long they will run? Explain.
- (14) (2 pts) Do revivals last longer? Explain.

- (15) (2 pts) This proportional odds model uses only 8 parameters. Suppose instead the response were considered nominal, and a multinomial logit model with the same predictors was fit to the data. How many parameters would the latter model require? Explain.
- (16) (2 pts) The revival of the musical “Cabaret” had 92% attendance in the first week. Based on this fitted proportional odds model, what is the predicted probability that the show lasts less than six months? (This show actually lasted more than three years).

Question D

A study was conducted to investigate the effect of rocking on the crying of new born babies with days as blocks. On each of 18 days, a baby was selected at random from the maternity ward and rocked. The other babies in the ward served as controls. At the end of the period of rocking, the number of babies crying was recorded --- the results are shown in Table 1. Conditions such as temperature varied from day to day.

Table 1

Day	No. of control babies	No. not crying	No. of experimental babies	No. not crying
1	8	3	1	1
2	6	2	1	1
3	5	1	1	1
4	6	1	1	0
5	5	4	1	1
6	9	4	1	1
7	8	5	1	1
8	8	4	1	1
9	5	3	1	1
10	9	8	1	0
11	6	5	1	1
12	9	8	1	1
13	8	5	1	1
14	5	4	1	1
15	6	4	1	1
16	8	7	1	1
17	6	4	1	0
18	8	5	1	1

- (17) (4 pts) Describe how you would model this data to answer the question of interest. Carefully specify all components in the model, including the response, distributional assumption, link function, predictors, and how the day-to-day variation is accounted for.
- (18) (2 pts) Based on your answer to (17), describe how the relevant hypothesis could be formed to test the question of interest by writing out the models corresponding to the null and alternative hypotheses. Note that the effect of the treatment factor (i.e., the effect of rocking on crying) is commonly assumed to be constant across levels of the block factor.
- (19) (2 pts) Describe in details the procedure to perform the test you presented in (18).