

MATH 523 - ASSIGNMENT 2

To be handed in not later than 5pm, Friday 4th March 2011.

Please hand in during lectures, to Burnside 1225, or to the Mathematics Office Burnside 1005

The following data reflect the blood test results for 471 male volunteers who were tested for antibodies to the bacterium *Helicobacter pylori*. The men were classified into seven age groups.

		Age Group							Total
		< 30	30-34	35-39	40-44	45-49	50-54	55-65	
Test positive	y	22	26	14	30	32	23	29	176
Test negative	z	52	55	59	53	31	28	17	295
Total	m	74	81	73	83	63	51	46	471

- (a) The first model that could be fitted to these data assumes a *Poisson* model for the counts in each of the fourteen cells of the table.

Using code similar to that provided in lectures, fit Poisson GLM(s) with log link to the fourteen data investigating whether there is a significant difference between the rates of occurrences of positive tests across the different age groups. Report your results.

In this case, there are two factors, age (taking seven levels) and test (taking two levels)

8 MARKS

- (b) A second model that could be fitted assumes that the data in columns in the table can be modelled using a Binomial model; that is, in column j , the number that test positive, Y_j , are binomially distributed, $Y_j \sim \text{Binomial}(m_j, \pi_j)$, for $j = 1, \dots, 7$.

Using code similar to that provided in lectures, fit a Binomial GLM with logistic link to the seven data in columns of the table, investigating whether there is a significant difference between the rates of occurrences of positive tests across the different age groups. Report your results.

Recall that to fit a binomial model, R syntax for fitting the factor predictor age is

```
age<-gl(7,1)
glm(cbind(y,z)~age,family=binomial)
```

6 MARKS

- (c) In terms of understanding the differences in proportions of positive tests across age groups, which of these two models gives more insight? In particular, consider the estimated differences in proportions across age groups derived from models from (a) and (b). Is it possible to obtain **identical** estimates of the magnitude of the age effect? Justify your answers.

4 MARKS

- (d) Identify one further model or test that might be of interest to researchers in this analysis, giving full justification for the model or test you have selected.

2 MARKS