MTH6134 2023 Sample Quiz 2

(1) **Rat data**

ESSAY | marked out of 1.0 | penalty 0.10 | HTML editor

Consider the rat experiment with two delivery methods for poison, given in the data as levels of variable `w` (1:solid and 2:liquid). We use index j to refer to levels of `w`; for each level, there were six experiments performed, which are indexed by k. There is interest in studying the effect of delivery method regardless of the dose level.

To this end, fit the logistic regression model $\log\left(\frac{\pi_{jk}}{1-\pi_{jk}}\right) = \alpha_j$ to the rat data, where π_{jk} is the proportion of rats surviving the k-th experiment with the j-th delivery method. Refer to this fit as "Model 1".

(a) Write the maximum likelihood estimates $\hat{\alpha}_j$ for Model 1.

(b) Write the deviance of Model 1 and briefly explain what is its main use.

(c) Perform the usual chain of comparisons between models for these data, starting with the comparison of Model 1 against the maximal model. Depending on the result, you may want to compare a simpler model against Model 1. In each case, report the relevant deviance, the degrees of freedom, the p-value and briefly comment on the outcome of the comparison.

(d) Consider the model $Y_{jk} \sim \text{Bin}(r_{jk}, \pi_j)$. Recall maximum likelihood estimates $\hat{\pi}_j = \sum_{k=1}^{6} y_{jk} / \sum_{k=1}^{6} r_{jk}$. Compute and write these estimates for $j = 1, 2$.

(e) Link the results from the fit of Model 1 with your estimates of Question (1d). Specifically, explain and show numerically how to derive $\hat{\pi}_j$ from $\hat{\alpha}_j$.

(f) Consider a variation of Model 1 that instead of logit link, uses logarithmic link. Give the maximum likelihood estimates $\hat{\alpha}_1, \hat{\alpha}_2$ for this model. Hint: study carefully the link, you do not need to fit a second model.

Information for graders:

- (a) The estimates are $\hat{\alpha}_1 = 0.5205$ and $\hat{\alpha}_1 = 0.9808$. These are read from `m1 $coefficients` which are 0.5205, 0.4603. The estimate $\hat{\alpha}_1$ is the first entry, while $\hat{\alpha}_2$ is the addition of these two outputs.
  (b) The deviance of Model 1 is the statistic used to test the null hypothesis of data fitting Model 1 against the alternative that data fits better the maximal model. Its value 72.0885 is either computed from `deviance(m1)` or `m1$deviance`.
  (c) We start with the test H_0: data fits well Model 1 against H_1: data fits well maximal model. The deviance for this is 72.0885, with 10 degrees of freedom and p-value $1.7491009 \times 10^{-11}$. We set $\alpha = 0.05$ and as the p-value is much smaller than α, we reject H_0. The data supports better the maximal model than the two parameter Model 1.
  As we have already rejected the two parameter model, we do not compare the other hypothesis test using the null model and we stop.
  (d) The maximum likelihood estimates are $\hat{\pi}_1 = 0.6273$ and $\hat{\pi}_2 = 0.7273$.
  (e) To derive the $\hat{\pi}_j$ from $\hat{\alpha}_j$ we only need to invert the logistic link, i.e. $\hat{\pi}_j = 1/(1 + \exp(-\alpha_j))$. We have $\hat{\pi}_1 = 1/(1 + \exp(-0.5205)) = 0.62726$ and $\hat{\pi}_2 = 1/(1 + \exp(-0.9808)) = 0.72727$. Applying the logistic transformation to $\hat{\pi}_j$, we would retrieve $\hat{\alpha}_j$.
  (f) This model has link $\log(\pi_{jk}) = \alpha_j$. We use the link to have maximum likelihood estimates $\hat{\alpha}_1 = \log(0.6273) = -0.4663$ and $\hat{\alpha}_2 = \log(0.7273) = -0.3184$.

(2) **glm**

ESSAY | marked out of 1.0 | penalty 0.10 | HTML editor

You want to study whether a variable x relates to a variable y and you plan to use the function `glm()` in `R`. The data is

```
y<-c(9, 14, 8, 14, 13, 7, 9, 20)
```

and the following command will be used

```
glm(formula = y~ x, family = ???)
```

Using the text box below, write what word (or words) you could use to substitute in the place of the question marks `???` As part of your answer, very briefly explain your reasoning, using no more than two sentences. You are not expected to comment about the link but only on the family.

Information for graders:

- Options for family: poisson or binomial.
  Comment: The data are integers and although poisson seems likely, binomial cannot be ruled out. With binomial, we'd need a value for r, though.

Total of marks: 2