STATS 2107
Statistical Modelling and Inference II
Assignment 5

Jono Tuke

Semester 2 2017

CHECKLIST

- ☐: Have you shown all of your working, including probability notation where necessary?
- ☐: Have you given all numbers to 3 decimal places?
- ☐: Have you included all R output and plots to support your answers where necessary?
- ☐: Have you included all of your R code?
- ☐: Have you made sure that all plots and tables each have a caption?
- ☐: If before the deadline, have you submitted your assignment via the online submission on MyUni?
- ☐: Is your submission a single pdf file - correctly orientated, easy to read? If not, penalties apply.
- ☐: Penalties for more than one document - 10% of final mark for each extra document. Note that you may resubmit and your final version is marked, but the final document should be a single file.
- ☐: Penalties for late submission - within 24 hours 40% of final mark. After 24 hours, assignment is not marked and you get zero.
- ☐: Assignments emailed instead of submitted by the online submission on MyUni will not be marked and will recieve zero.
- ☐: Have you checked that the assignment submitted is the correct one, as we cannot accept other submissions after the due date?

Due date: Friday 20th October 2017 (Week 11), 5pm.

Q1 Exponential distribution

Suppose $y_1,\ y_2,\ \ldots, y_n$ are independent exponential observations with probability density function

$$f(y;\theta) = \frac{1}{\theta}e^{-y/\theta} \quad \text{for } y \geq 0, \quad \theta > 0.$$

(a) Write down the log-likelihood $\ell(\theta;\boldsymbol{y})$.

[2 marks]

(b) Hence, find the maximum likelihood estimate $\hat{\theta}$, and

[2 marks]

(c) the Fisher information I_θ.

[3 marks]

[Question total: 7]

Q2 Binomial distribution

Consider a single binomial observation x from $\text{Bin}(n, \theta)$ where the number of trials is n and the probability of success is p.

(a) Give the log-likelihood $\ell(\theta; x)$.

[1 mark]

(b) Find the Score function, the Fisher information, and the MLE, $\hat{\theta}$.

[6 marks]

(c) Find expressions for the log-likelihood ratio test statistic, G^2, and the score test statistic, U, for testing the null hypothesis $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$.

[5 marks]

(d) State the asymptotic distributions of G^2 and U, respectively, under H_0.

[2 marks]

[Question total: 14]

Q3 Poisson distribution

Suppose $y_1, y_2, \ldots, y_n$ are independent $Po(\lambda)$ observations.

Hint: You may use the log-likelihood, score function, and Fisher information for λ, and the MLE $\hat{\lambda}$, from the lecture notes to answer the following questions.

(a) State an approximate $100(1-\alpha)\%$ confidence interval for λ.

[2 marks]

(b) Let $\phi = \log \lambda$.

(i) Write down the log-likelihood $\ell_\phi(\phi; \mathbf{y})$.

[2 marks]

(ii) Hence find the maximum likelihood estimate $\hat{\phi}$ and the Fisher information $i_n(\phi)$.

[2 marks]

(iii) In lectures, we showed that the score test statistic to test $H_0 : \lambda = \lambda_0$ is

$$U = \frac{\bar{Y} - \lambda_0}{\sqrt{\lambda_0/n}}.$$

Let $\phi_0 = \log \lambda_0$. Using your answer to (ii), show that U', the score test statistic for testing $H_0 : \phi = \phi_0$, is the same as U.

[3 marks]

[Question total: 9]

Q4 Linear modelling question

This question must be typed up for full marks. Please include your code and output to show your working.

In this question, you are going to examine the relationship between price and the predictors:

- age and
- state.

To do this complete the following:

a. Load the gumtree data in `2.gumtree.rds`
b. We are only going to consider dogs that have been sold as opposed to given away. We will consider all dogs with a price greater than $10 to be sold and the rest to be given away. Filter your dataset to include only that dogs whose price is greater than $10.

[1 mark]

c. Fit a model with age as a predictor and price as the response variable (Model 1). Are the assumptions of the linear model reasonable (include the appropriate plots to justify your conclusions)?

[5 marks]

d. Fit a model with age as a predictor and log(price) as the response variable (Model 2). Are the assumptions of the linear model reasonable (include the appropriate plots to justify your conclusions)?

[5 marks]

e. Using Model 2, calculate a 95% prediction interval for a dog with an age of 1 year.

[2 marks]

f. Fit a model with state as a predictor and log(price) as the response variable (Model 3). Are the assumptions of the linear model reasonable (include the appropriate plots to justify your conclusions)?

   Leave the observations with state NA in the dataset for the model.

[5 marks]

g. In Model 3, which state is used as the reference level?

[2 marks]

h. In Model 3, which state is significantly different to the reference level at the 5% significance level?

[1 mark]

i. Using Model 3, predict the price of a dog from South Australia. Also give the 95% prediction interval for the price of a dog from South Australia.

[2 marks]

[Question total: 23]

[[Assignment total: 53]]