STATS 2107
Statistical Modelling and Inference II
Assignment 2

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 25th August 2017 (Week 5), 5pm.

Q1. Properties for S_p^2.

This question may be handwritten and then scanned to pdf.

Consider the independent random variables

$$Y_{ij},\ i = 1, 2;\ j = 1, 2, \ldots, n_i$$

with

$$Y_{ij} \sim N(\mu_i, \sigma^2).$$

(a) Prove that

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator for σ^2.

[3 marks]

(b) Prove that

Figure 1: Workflow for cleaning gumtree dataset.

$$\frac{(n_1+n_2-2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

[3 marks]

[Question total: 6]

Q2. Confidence intervals for σ^2

For full marks, please show all working. If you use R, please give the code

Bats are able to locate prey be emitting high-pitch sounds and listening for the echo. In a study of animal behaviour in a colony of bats, distances in cm were recorded at which the bats first detected a nearby insect. In total 65 recordings were made and the sample mean was 50.34cm and the sample standard deviation was 10.72cm.

(a) Let S^2 represent the sample variance. State the distribution of

$$\frac{(n-1)S^2}{\sigma^2}$$

[1 mark]

(b) Calculate a symmetric 95% confidence interval for σ^2.

[3 marks]

(c) Calculate a lower 95% confidence bound for σ^2.

[3 marks]

(d) Calculate an upper 95% confidence interval for σ^2.

[3 marks]

(e) The researcher would like to test

$$H_0 : \sigma^2 = 84,$$
$$H_a : \sigma^2 \neq 84.$$

Using the **appropriate** confidence interval for σ^2, would you reject or retain the null hypothesis at the 5% significance level? Justify your answer.

[3 marks]

[Question total: 13]

Q3. Cleaning the gumtree data part 2 and confidence intervals for price.

First load the cleaned dataset from Assignment 1. The workflow is given in Figure 1.

(a) For the variables age and price, check using summary statistics and histograms whether the observed values pass the stupidity test. For each variable, give your cleaning code and justification for any cleaning. For full marks, you should include both summary statistics and histograms.

Hint: dogs can live up to 25 years and can cost up to $10000

[6 marks]

(b) For each of the following variables:

- Pet offered by,
- Microchip,
- Vaccination,
- Desexing status,
- State, and
- relinquished,

write R code to calculate the 95% confidence interval for the mean price for each level of the variables. For full marks, give the R code and the output.

[12 marks]

(c) For which of the levels of the variables in part (b), do you think that the assumption of normality is unreasonable? Why?

[2 marks]

[Question total: 20]

[[Assignment total: 39]]