STATS 1000 / STATS 1004 / STATS 1504
Statistical Practice 1
Assignment 6
2020

DEADLINE:

- Wednesday 10^{th} June 2020 (Week 13) 5:00pm

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 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 word document or 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.

☐: If after the deadline, but within 24 hours, have you contacted us via the enquiry page on MyUni and then submitted your assignment online via the online submission on MyUni?

☐: 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 if the correct one, as we cannot accept other submissions after the due date.

☐: Do not write directly on the question sheet.

1. **Chi-square test in R**

   For full marks this answer must be typed in Word and plots and output included and captioned.

   Myocardial infarction is where a blockage in the blood vessels to the heart cause damage to the heart muscle. It is often referred to as a "heart attack". A 1988 study looked at the effect of aspirin to prevent death after people have suffered a myocardial infection. The dataset is on MyUni and is called `aspirin.sav`. Load the data into R, then complete the following.

   (a) Perform an Chi-square test to test for an association between treatment group and outcome. For full marks, include

      i. your R output,

      [1 mark]

      ii. the null and alternative hypotheses,

      [2 marks]

      iii. the value of the test statistic,

      [1 mark]

      iv. the degrees of freedom,

      [1 mark]

      v. the P-value, and

      [1 mark]

      vi. whether you reject or retain the null hypothesis at the 5% significance level, and why?

      [2 marks]

      [Total: 8]

2. **Linear regression in R**

   For full marks this answer must be typed in Word and plots and output included and captioned.

   One of the original uses of linear regression was to examine the relationship between the height of fathers and their sons (both in inches). The dataset `pearson.sav` is a dataset obtained by Karl Pearson[1] to look at this relation-

[1] https://en.wikipedia.org/wiki/Karl_Pearson

ship. In fact this relationship is why we call it regression[2]. Load the dataset into R and the complete the following.

(a) Produce a scatterplot of son's height (`son_height`) against father's height (`father_height`). Describe the relationship.

[3 marks]

(b) Test for a statistically significant (5% level) linear relationship between `son_height` and `father_height`. Remember to include

i. your R output,

[1 mark]

ii. the null and alternative hypotheses,

[2 marks]

iii. the observed value of test statistic,

[1 mark]

iv. the P-value, and

[1 mark]

v. your conclusion. Do you reject or retain the null hypothesis? Why? Give conclusion in context.

[3 marks]

(c) Check the assumptions of the linear regression. Remember to include captioned plots where necessary.

[8 marks]

[Total: 19]

3. **One-way ANOVA in R**

For full marks this answer must be typed in Word and plots and output included and captioned.

Many studies have suggested that there is a link between exercise and healthy bones. It is suggested that exercise stresses the bones and this causes them to get stronger.

[2] https://en.wikipedia.org/wiki/Regression_toward_the_mean

One study examined the effect of jumping on the bone density of growing rats. The rats were randomly allocated to one of three treatments: a control with no jumping, a low-jump exercise, and a high-jump exercise. After 8 weeks of 10 jumps per day, for 5 days per week, the bone density of the rats (in mg/cm^3) was measured.

In this assignment question, we will use the `density.sav` dataset to look at how to perform one-way ANOVA in R. Download this dataset from MyUni and load it into R.

(a) A boxplot of the bone density for each exercise level is given in Figure 1. Compare the distribution for each group.

[4 marks]

(b) Use a one-way ANOVA to test for a significance difference between the mean bone density for each group with the following steps:

i. Write the appropriate null and alternative hypotheses. Remember to define all parameters used.

[2 marks]

ii. Include the one-way ANOVA table in your assignment. Remember to caption it.

[1 mark]

iii. State the value of the observed test statistic.

[1 mark]

iv. What is the distribution of the test statistic if the null hypothesis is true?

[2 marks]

v. State the P-value.

[1 mark]

vi. Do you reject or retain the null hypothesis at the 5% significance level? Why?

[2 marks]

(c) Is the assumption of constant variance reasonable for this dataset? Remember to include any R output needed to support your conclusion.

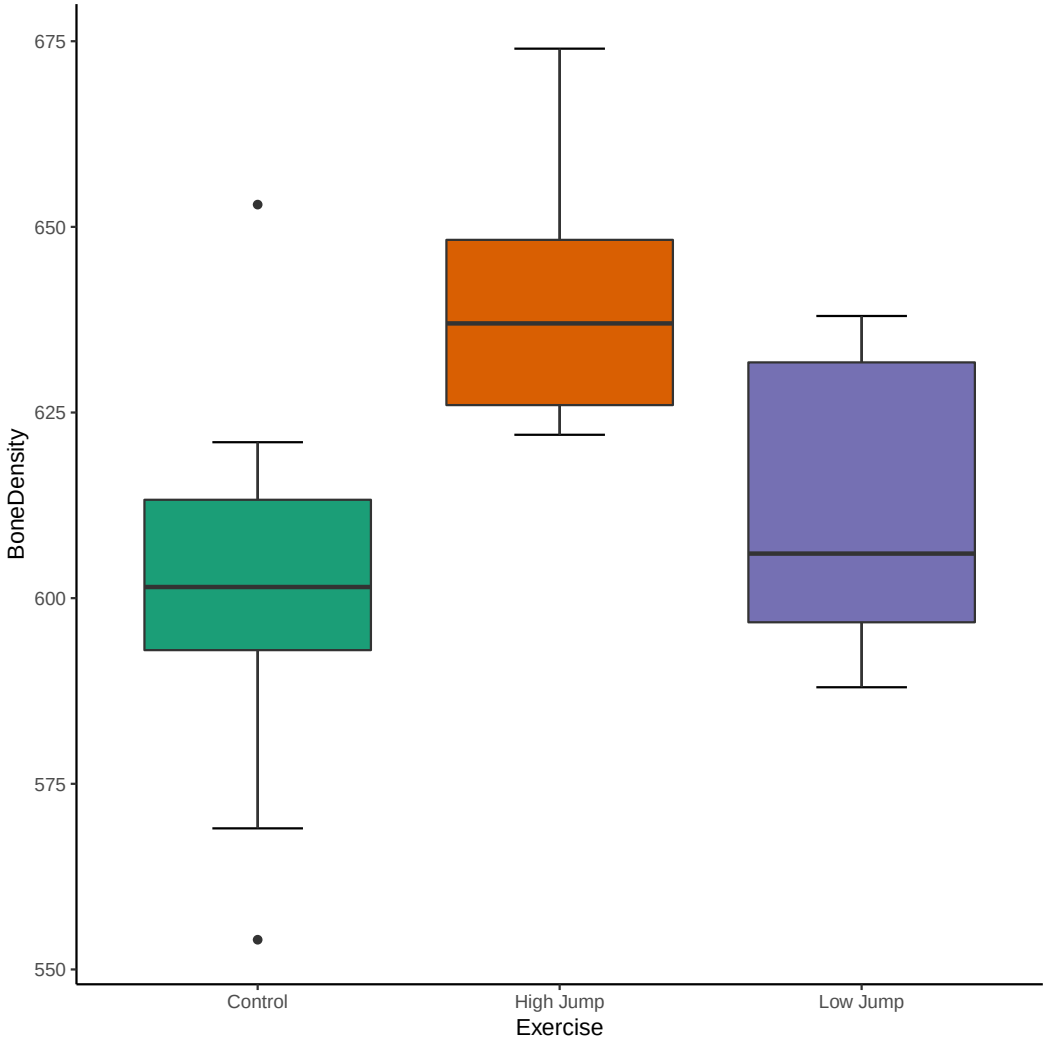

Figure 1: Boxplots of Bone Density for each Exercise for the density dataset

[3 marks]

(d) In R, produce a multiple comparisons table using a Bonferroni adjustment. Using this table, which exercises are significantly different at the 5% significance level. Remember to include and caption your table.

[3 marks]

[Total: 19]

Presentation marks

Marks for use of word and captions for all figures and all tables.

[3 marks]

[[Assignment total: 46]]