Name: _____MYER RAYMOND _____ ID number: _____#18035406_____

Applied Statistics 161.111
Assignment 2

Due date: Friday 20th May 2022

Total marks: 53 **Assessment value:** 17%

Background

Lamb birth weight is the single largest influence on survival of lambs in the first few days of life. In a trial at Massey University in the late 1990s, an equal number of Coopworth pregnant ewes (female sheep) were randomly allocated to an experiment involving two shearing treatments (shorn on day 70 vs not shorn) and four feeding regimes (HH, HM, MH, MM). The first letter refers to feed during days 70-101 of pregnancy, and the second during days 102-140. The M (maintenance) feeding group were fed at a level calculated to maintain maternal conceptus-free weight, while the H (high) feeding group were fed to gain around 100g per day above this during the relevant period.

After each lamb was born, its birthweight was recorded, as well as its rank (whether it was a single or twin), and the feeding regime and shearing treatment of its mom. The resulting dataset, to be analyzed in this assignment, contains information about 204 lambs.

The data are in the Excel file *lambs.xlsx*.

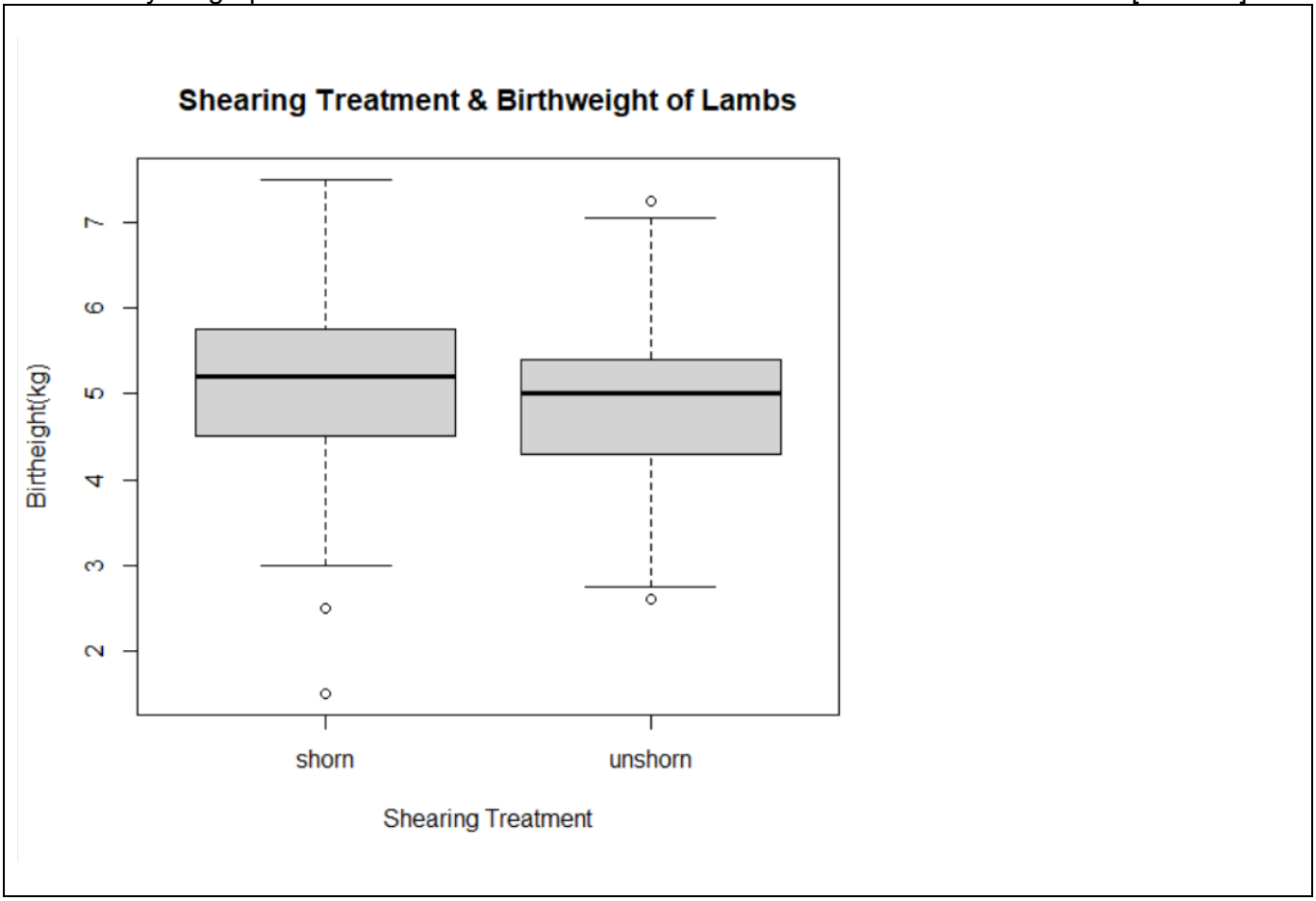
Use the data to answer the following questions in the spaces provided. You can re-size the answer spaces.

Use RStudio and incorporate the output into your answers.

Name: _____MYER RAYMOND _____ ID number: _____#18035406_____

Part A: Comparing means: Analysis of birthweight according to shearing treatment [30 marks]

A1: Use RStudio to draw side-by-side boxplots of birthweights for the lambs according to shearing treatment. Paste your graph below. [3 marks]



A2: Use RStudio to calculate numerical summaries for birthweights of lambs from shorn and unshorn ewes. Fill in the table with the values rounded appropriately. [4 marks]

Weight (kg)	Shorn	Unshorn
Number of lambs	102	102
Standard Deviation	1.0163828	0.9068687
Minimum	1.5	2.6
Lower Quartile	4.5	4.3125
Median	5.2	5
Mean	5.122353	4.919902
Upper Quartile	5.75	5.3875
Maximum	7.5	7.25

Name: _____MYER RAYMOND _____ ID number: _____#18035406_____

A3: What do the plots and numerical summaries tell you about the birthweights of Coopworth lambs from shorn and unshorn ewes? Write 1 or 2 sentences comparing each of the following: centres, spread, shape and outliers. [4 marks]

The median for shorn lambs is a weight of 5.2kg and 5 for unshorn, this is according to the numerical summary. The spread for the shorn lambs is slightly larger than that of the unshorn lambs, with an IQR of 1.25. Whereas the IQR for the unshorn lambs is 1.0745. Both boxplots appear to be normally distributed, with the unshorn having a slight negative skew. The boxplot for shorn lambs has 2 outliers both of which are less than the lower quartile. The boxplot for unshorn lambs has 2 outliers either side of the boxplot, one greater than and one less than. According to the data, we can observe that shorn sheep appear to have a higher birthweight than the unshorn sheep.

A4: One of the research goals was to investigate how the shearing treatment of pregnant ewes affects the birthweight of their lambs. Do a two-sample t-test to determine if there is any evidence that, on average, there is a difference in birthweights of lambs from shorn and unshorn Coopworth ewes.

a. Explain why a two-sample t-test is better than a t-test of differences for this experiment. [1 mark]

A two-sample t-test will be better for this experiment because the data obtained is random and because the data for birthweight/shearing treatment does not affect the other one. So, it will give us a more accurate result, by doing a two-sample t-test.

b. Step 1: Write the hypotheses. [2 marks]

The Null:
There is no difference in the true mean birthweight of shorn and unshorn lambs.
The Alternative:
There is a difference in the true mean birthweight of shorn and unshorn lambs.

c. Use RStudio to do the two-sample t-test. Paste the RStudio **code and output** below. [1 mark]

```
> t.test(lambs$birthweight~lambs$shearing_treatment, mu=0)

        Welch Two Sample t-test

data:  lambs$birthweight by lambs$shearing_treatment
t = 1.5011, df = 199.43, p-value = 0.1349
alternative hypothesis: true difference in means between group shorn and group uns
horn is not equal to 0
95 percent confidence interval:
 -0.06350855  0.46841051
sample estimates:
  mean in group shorn mean in group unshorn
             5.122353              4.919902
```

d. Step 2: State the value of the test statistic. [1 mark]

The value of the test statistic came back as 0.1349.

e. Step 3: State the statistical decision with reason. [1 mark]

The p-value was 0.1349, which is greater than 0.05. Therefore, we decide in favour of the null.

f. Step 4: Write your conclusion. [2 marks]

There is no evidence that suggest the true mean birthweight of shorn lambs is different to the birthweight of unshorn lambs.

g. Step 5: Check the conditions are met. [4 marks]

All three conditions are met.

h. Write a sentence to interpret the confidence interval. [3 marks]

We are 95% confident that shorn lambs will weigh somewhere between 0.06350855kgs less and 0.46841051kgs more than the unshorn lambs, on average.

i. Does the confidence interval confirm or add to your conclusion? Explain. [1 mark]

The confidence interval contains the value 0, this supports the conclusion that there is no difference. Therefore confirming the conclusion even more.

j. Explain how the t-test conclusion and confidence interval answer the research goal described in A4. [1 mark]

The t-test and CI help confirm the research goal by telling us that there was no difference, between the two different shearing treatments.

A5. Do your findings from the plots and summary statistics in A1, A2, and A3 contradict the results of the t-test and confidence interval above? Explain. [2 marks]

Yes, but by a very small degree. The boxplot and numerical summaries show differences, but after further investigation in the t-test and CI we can see that the data is a lot closer then originally, we presumed.

Name: _____MYER RAYMOND _____ ID number: _____#18035406_____

Part B: Analysis of rank and feed regime. [23 marks]

We are interested in whether the feeding regime affects rank (single or twin).

B1: Use RStudio to produce a **table of counts** for rank according to feeding regime. Put feeding regime as the columns and rank as the rows. [2 marks]

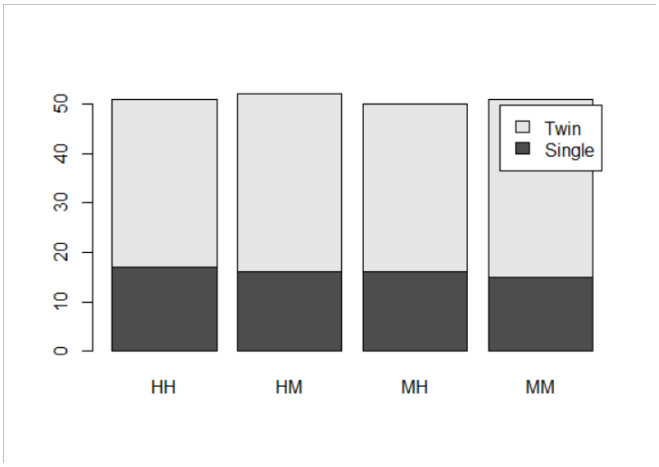
```
        HH HM MH MM
Single 17 16 16 15
Twin   34 36 34 36
```

B2: Use RStudio to produce a table of rank as a **proportion** of the lambs in each feeding regime**.** [2 marks]

```
> count.prop

             HH        HM        MH        MM
Single 0.3333333 0.3076923 0.3200000 0.2941176
Twin   0.6666667 0.6923077 0.6800000 0.7058824
```

B3: Use RStudio to produce a **stacked bar plot** of the distribution of rank as a proportion of lambs from each feeding regime. [2 marks]



B4: What do the tables and plot tell you? What does this tell you about the impact of feeding regime on rank for these lambs? [2 marks]

From the tables and plots above we can observe and conclude that there are nearly double the number of twins to single in every feeding regime. From the data, we can see that the results appear to be very close, with not much discrepancy involved.

B5. What is the (approximate) proportion of ewes giving birth to twins in each regime? [1 mark]

It is approximately 68% or 0.6825. Of ewes give birth to twins in each regime.

Name: _____MYER RAYMOND _____ ID number: _____#18035406_____

B6: Do a Chi-squared test of rank and feeding regime.

a. Step 1: Write the hypotheses. [2 marks]

The Null: Feeding regime is not related to rank.
The Alternative: Feeding regime is related to rank.

b. Use RStudio to do the Chi-Squared test. Include the RStudio **code and output** below. [1 mark]

```
> regime.table = table(lambs$feed, lambs$rank)
> chisq.test(regime.table)

        Pearson's Chi-squared test

data:  regime.table
X-squared = 0.20008, df = 3, p-value = 0.9776
```

c. Step 2: State the value of the test statistic. [1 mark]

The value of the test statistic is 0.20008

d. Step 3: State the statistical decision with reason. [1 mark]

We decide in favour of the null because the p-value of 0.9776 is greater than 0.05.

e. Step 4: Write your conclusion. [1 mark]

We have no evidence to suggest that the feeding regime is related to rank.

f. Step 5: Check the conditions are met. [4 marks]

The sample is representative.
The expected count has also been met; all values are greater than 5.
Therefore, both conditions have been met.

```
> chisq.test(regime.table)$expected

          HH       HM       MH MM
  Single 16 16.31373 15.68627 16
  Twin   35 35.68627 34.31373 35
```

g. Use RStudio to calculate the residuals. Include the RStudio **code and output** below. [1 mark]

```
> chisq.test(regime.table)$residuals

           HH        HM        MH        MM
 Single  0.25000000 -0.07767356  0.07921180 -0.25000000
 Twin   -0.16903085  0.05251691 -0.05355695  0.16903085
```

h. Do the residuals add to your conclusion? Explain. [1 mark]

They do not add to the conclusion, as seen also in the stacked bar plot. The results are far too similar, to be able to differentiate between them accurately.

i. Discuss why the conclusion of the Chi-squared test is not surprising. Your discussion should consider the data as well as the experimental design. [2 marks]

The chi-squared test is not surprising from the data, because we can observe through the experimental design that took place that the change was very low, somewhat irrelevant. The null hypothesis of; feeding regime is not related to rank stays accurate. The reason for almost double the number of twins being born compared to singles in each feeding regime has zero correlation to each other. This has been observed through the chi-squared test, where the result came back 0.20008 and a p-value of 0.9776.

+ + + + + + + +