# Applied Statistics 161.111

# Assignment 2

**Due date:** Friday 15 May 2020

**Total marks: 50** **Assessment value:** 12%

## Image result for green lipped musselThe population data

## The population we are considering for this assignment are the 10,000 kuku (New Zealand green–lipped mussels) growing in a mussel farm in the Marlborough Sounds.

## Variables of interest are the length of the kuku (in millimetres), grade (small, medium or large) and sex (male or female).

## Each kuku (mussel) has a unique ID. The population consists of:

## 1948 large kuku with ID numbers from 1 to 1948.

## 4457 medium kuku with ID numbers from 1949 to 6405

## 3595 small kuku with ID numbers from 6406 to 10000.

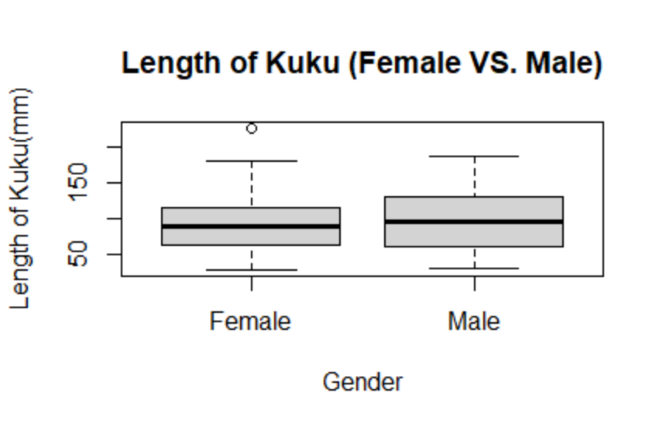
* **Use your sample data From Assignment 1.**

## If you have lost your sample data, you can get it back by clicking on the following link. <http://shiny.massey.ac.nz/dleader/DataDownload111/>

## Type your ID number into the “Student ID:” box in the Shiny app that opens.

* Use your sample data to answer the following questionsin the answer spaces provided. You can re-size the answer spaces.

**Part A: Comparing means: Analysis of kuku lengths according to sex [30 marks]**

1. Use RStudio to draw side-by-side boxplots of kuku lengths for the different sexes. [5 marks]
2. Use RStudio to calculate numerical summaries for kuku lengths for each sex. Fill in the table with the values rounded appropriately. [4 marks]

|  |  |  |
| --- | --- | --- |
| Sex | Female | Male |
| Number | 50 | 50 |
| Minimum | 28.80 | 31.6 |
| Lower Quartile | 63.95 | 62.9 |
| Median | 88.45 | 95.45 |
| Mean | 95.55 | 98.97 |
| Upper Quartile | 115.47 | 130.18 |
| Maximum | 225.10 | 187 |

1. What do the plot and numerical summaries tell you about kuku lengths of males and females in your sample? [4 marks]

|  |
| --- |
| There is a difference of 7mm in the overall median lengths between the males and females in my sample in favour of the males. However, the difference between the two means is only 3.42mm in favour of the males again. The IQR of the males is 67.28mm, while the females IQR is 51.52. The distribution is very symmetric with both boxplots, as the median lies very near the centre in both instances. There is one visible outlier in my sample, a female Kuku with a length of 225.10mm. |

1. Do a two-sample t-test to determine if there is any evidence that on average there is a difference in the length of male and female kuku.
   1. Step 1: Write the hypotheses. [2 marks]

|  |
| --- |
| The null: The true difference in means between the lengths of male and female green-lipped Kuku is equal to zero.  The Alternative: The true difference in means between the lengths of male and female green-lipped Kuku is not equal to zero |

* 1. Use RStudio to do the two-sample t-test. [1 mark]

|  |
| --- |
| data: Females$Length and Males$Length  t = -0.40836, df = 97.931, p-value = 0.6839  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  -20.06363 13.21563  sample estimates:  mean of x mean of y  95.546 98.970 |

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

|  |
| --- |
| Test statistic = -0.40836 |

1. Step 3: Make a decision based on the p-value. [1 mark]

|  |
| --- |
| The p-value in this instance is far above 0.05, thus’ I would rather trust the null hypothesis that the true difference is equal to 0, as opposed to the alternative. |

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

|  |
| --- |
| The evidence suggests that the true average lengths of Green-lipped Kuku do not differ between males and females. |

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

|  |
| --- |
| The sample was taken using a stratified random sampling method. This would mean that our sample is representative. As we have a sample group of 100, the Central Limit Theorem holds, and we can assume that the distribution of this sample is normal. |

1. Write a sentence to interpret the confidence interval. Explain how it adds to your conclusion. [4 marks]

|  |
| --- |
| I am 95% confident that the true mean difference in lengths between male and female Green-lipped Kuku is somewhere between -20.06mm and 13.22mm. This supports my conclusion, as 0mm comes between the two numbers of my confidence interval. |

1. Explain why a two-sample t-test is better than a t-test of differences here. [2 marks]

|  |
| --- |
| The samples given are completely independent and do not change gender. |

**Part B: Two Categorical Variables: Analysis of kuku** **grades according to sex. [20 marks]**

1. Use RStudio to produce a contingency table of kuku grades according to sex. [2 marks]

|  |
| --- |
|  |

1. Use RStudio to draw a side-by-side bar plot of the proportions of kuku grades according to sex. [4 marks]

|  |
| --- |
|  |

1. What do the plot and table tell you about the distribution of kuku grades according to sex for your sample. [2 marks]

|  |
| --- |
| There do appear to be slightly more large male Kuku compared to large female Kuku, and a higher amount of the medium and small female Kuku, this could suggest that there is a relationship between Kuku gender and grade, though the relationship would be very weak. |

1. Do a Chi-squared test to determine if the grade of a kuku is related to its sex.
   1. Step 1: Write the hypotheses. [2 marks]

|  |
| --- |
| The null: There is no relation between the grade of a Green-lipped Kuku and it’s gender. The alternative: There is a relation between the grade of a Green-lipped Kuku and it’s gender. |

* 1. Use RStudio to do the Chi-Squared test. [1 mark]

|  |
| --- |
| data: Kuku\_table  X-squared = 0.60702, df = 2, p-value = 0.7382 |

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

|  |
| --- |
| Test statistic = 0.60702 |

1. Step 3: Make a decision based on the p-value. [1 mark]

|  |
| --- |
| The p-value is significantly larger than 0.05, and thus’ is in favour of the null hypothesis. |

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

|  |
| --- |
| There is evidence to suggest that the grading of green-lipped Kuku is unrelated to the age of that kuku. |

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

|  |
| --- |
| As stated earlier, the Kuku were chosen with a stratified random sampling method, this would mean that the Kuku in my sample are representative of the population. Using RStudio, I also found that all the expected counts are greater than 5.   Female Male  Large 9.5 9.5  Medium 22.5 22.5  Small 18.0 18.0 |

1. Use RStudio to calculate the residuals. [1 mark]

|  |  |  |
| --- | --- | --- |
|  | Female | Male |
| Large | -0.48666 | 0.486664 |
| Medium | 0.105409 | -0.10541 |
| Small | 0.235702 | -0.235702 |

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

|  |
| --- |
| In a case where there is absolutely zero difference between what we would expect if there was no relationship between a green-lipped Kuku’s grade and age and what the actual relationship is, the residual would be zero. As we can see in the above table, the residuals instead are marginal on either side of zero. While very small differences, this would suggest that, unlike my previous conclusion, there is a marginal relationship between a green-lipped mussel’s gender and their size grade. |

+ + + + + + + +

**Appendix:**

Data in *Assignment-Data-20005898.xlsx*.