Analysis of Scientific Data – Summer Semester, 2022

Week 2 Tutorial – Solutions

Part A

This is a short quiz to check you know the important ideas about logarithms. Without using a calculator (answers should have no decimal places), simplify the following as much as possible:

1. $\log_{10} 10 = 1$

2. $\log_{10} 1000 = 3$

3. $\log_{10} 1 = 0$

4. $\log_{10} 0.01 = \log_{10} 10^{-2} = -2$

5. $\log_{10}(10^x) = x$

6. $\log_{10} 200 = \log_{10}(100 \times 2) = \log_{10} 100 + \log_{10} 2 = 2 + \log_{10} 2$

7. $\log_{10} 0$ is undefined

8. $\log_2 16 = 4$

9. $\log_2 \frac{1}{4} = -2$

10. $\log_e 2.718282 \approx 1$

Part B

A study was undertaken to evaluate the effects of an HFD on bone healing in growing female rats. Twenty-six postweaning female rats were randomly assigned to two groups (13 per group): a standard diet (SD) group and an HFD (with 60% of energy from fat) group. The rats were maintained under standard laboratory conditions at a temperature of 22 ± 2 °C and humidity of 55 ± 5 % with a 12-hour light and dark cycle and with free access to water and food. The rats received the assigned diets for 5 weeks, and in the third week they were submitted to an osteotomy procedure (i.e. bone cut) of the left tibia. At the end of the 5 weeks, the rats were euthanised and the tibiae were removed for mechanical testing. The mechanical testing involved a force begin applied to the tibia at a speed of 1.0 mm/min to generate a tension load on the osteotomy site. The maximal load (N) was measured. Lower maximal load (p = 0.040) was observed in tibiae of HFD rats compared to SD rats.

(a) What type of variable is *maximal load*? Continuous

(b) What type of variable is *diet*? Ordinal (diet ordered by fat content). Nominal would be an acceptable answer.

(c) Which variable is the variable of interest, that is the response variable? Maximal load

(d) What type of study is this? Randomised comparative experiment

(e) What would be a suitable null hypothesis associated with the p-value for maximal load? H_0: The level of dietary fat has no effect on the average bone maximal load following osteotomy in rats

(f) What are we able to conclude about the effect of diet on bone maximal load following osteotomy? There is moderate evidence against the null hypothesis (p = 0.04), suggesting that high fat diets in rats reduces the average bone maximal load.

PART C – R activity (Please note that R codes will not be provided with the solutions. It is important to attend tutorial classes to practice RStudio).

The following study comes from An Introduction to Biostatistics by Glover & Mitchell (2002).

Copper sulfate is routinely used to control algal blooms in ponds and lakes, but an ichthyologist believes that copper sulfate has an adverse effect on the gill filaments of several species of fish, reducing the number of mucus cells in these species. To test her beliefs, she recorded the number of mucus cells per square micron in the gill filaments of untreated fish and in fish exposed for 24 hours to copper sulfate at 1 mg/L. The results are given in the in the “MucusCells.csv” file in the Week 2 Tutorial Questions under Week 2 folder.

a) Read the data file into RStudio.

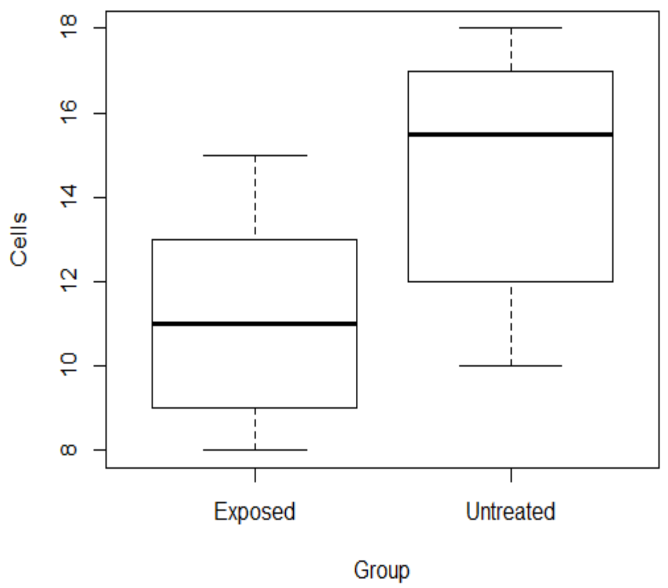
b) What is the mean number of mucus cells for all fish?
Mean = 12.75

c) What is the Interquartile Range (IQR) of mucus cells for all fish? Use five number summary in RStudio and then calculate the IQR using Q1 and Q3.

Using fivenum() in RStudio,
Q1 = 10.0; Q3 = 15.5

IQR = Q3 – Q1 = 5.5

d) Compare the distributions of mucus cells of the two groups using a side-by-side boxplot.



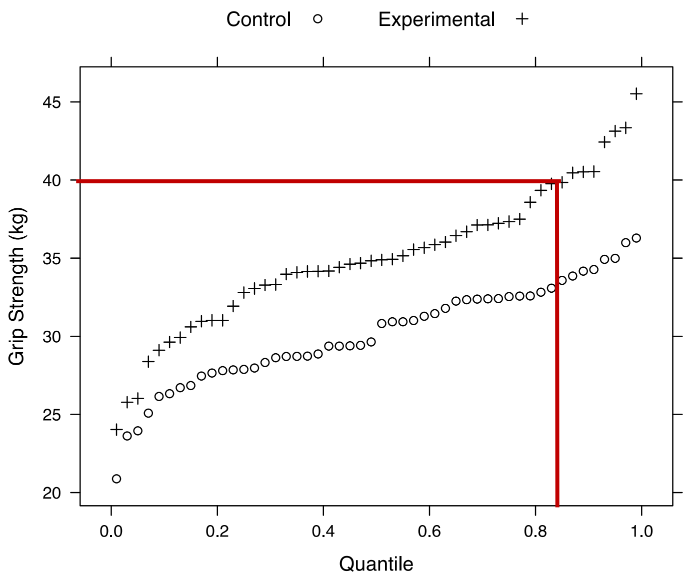
The distribution of the mucus cells in the untreated fish group is negatively skewed while the Exposed fish group is roughly symmetrical. Even though two groups show a similar variability (spread) the exposed fish group has a lower median and interquartile range (from the height of the box) than the untreated group.

e) What is the mean difference in the number of mucus cells between Untreated and Exposed groups?
Using the aggregate() in RStudio, the mean difference is 3.5

f) Which group has the higher standard deviation of mucus cells?
Using the aggregate() in RStudio, you can see that untreated group has a slightly higher standard deviation than the exposed group.

PART D

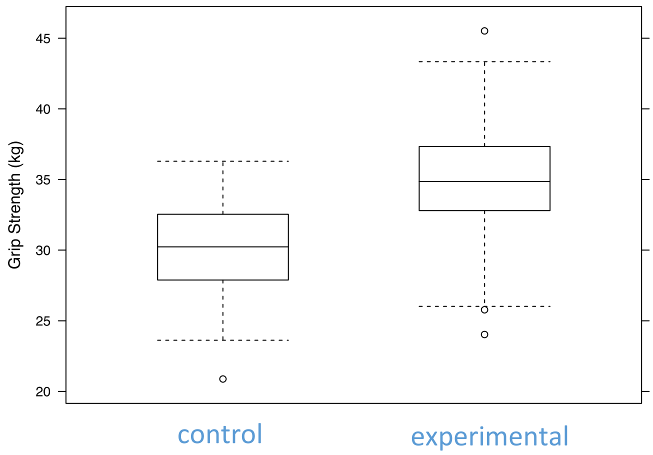
Repetitive movements and poor posture are associated with over-use of smartphones and can significantly contribute to symptoms of pain and discomfort in the upper extremities. The severity of these symptoms can be assessed through questionnaires as well as through direct measurements of grip strength (kg). A paper in 2016 presented a study on the effect of an exercise training and posturalcorrection program that aimed to help smartphone users. The researchers recruited 100 university students as potential participants in the study. The students were randomly assigned to either a control group, who received a pamphlet, or an experimental group, who received the new training program. The following quantile plot shows the measured grip strength for all subjects after the studyhad concluded:



a) Was this a blind study? Why or why not?
The study was not blind since subjects would know if they received the exercise training or not.

b) Estimate the proportion of subjects in the experimental group who had a grip strength of 40 kg or higher at the end of the study.
40 kg corresponds roughly to the 0.83 quantile (see the red line in the quantile plot). Therefore, 17% of subjects in the experimental group have a grip strength of 40kg or higher.

c) Based on the quantile plot, match the following boxplots to their corresponding group(experimental or control):



In this case, we can distinguish between the two groups based on their medians. From the quantile plot you can estimate that the medians are about 35kg (experimental) and 30kg (control).

d) Summarise the effects of the experimental treatment on the grip strength outcomes.
The experimental group has a higher median grip strength than the control group. It seems that grip strength has improved but there is also greater variability (spread) in the grip strength of the experimental group. The lowest values in the two groups are similar.

e) What issues might affect your conclusions in (d)?
It could be that the participants in the experimental group had higher initial grip strength. Randomisation would help avoid that problem, but the analysis would be improved if we knew the grip strengths of the participants at the start of the trial.

Two important points:

- Randomisation reduces the bias in the experiment due to other factors by making the groups similar in all variables, even those we did not measure.
- Our analysis can be improved if we take into account other factors (like initial strength).

f) The mean grip strength of the 50 subjects in the experimental group was 35.02 kg. One of those subjects had a grip strength of 24.03 kg. What would happen to the mean if their value was removed from the data?
It will go up. More specifically, the total grip strength of all 50 subjects was $50 \times 35.02 = 1751$ kg. If the grip strength of 24.03 is removed, then the remaining total is $1751 - 24.03 = 1726.97$ kg. The new sample mean is then $\frac{1726.97}{49} = 35.24$ kg.

g) The sample variance is the sum of the squared deviations of observations from the sample mean divided by the degrees of freedom. In later modules, we will sometimes want to obtain the sum of the squared deviations from a known sample standard deviation. For example, the standard deviation of the grip strengths of the 50 subjects in the experimental group was 4.559 kg. What was the sum of the squared deviations from the mean? What are the units of this value?

The sample standard deviation is

$$s = \sqrt{\frac{\sum(x_j - \bar{x})^2}{n-1}}$$

Rearranging this gives $\sum(x_j - \bar{x})^2 = (n-1)s^2$. For the grip strength values, this gives

$$\sum(x_j - \bar{x})^2 = (50-1)4.559^2 = 1018.44$$

The units are kg^2.

Reference

J.S Yamanaka, G.R. Yanagihara, B.L. Carlos, J. Ramos, B.B. Brancaleon, A.P. Macedo, J.P.M Issa & A.C. Shimano (2018) A high-fat diet can affect bone healing in growing rats, Journal of Bone and Mineral Metabolism, 36, 255-263. doi: 10.1007/s00774-017-0837-4