Due Date: 10:00 PM, April 3, 2023

Prof. Alexander Giessing

Last Update: March 29, 2023

Spring Quarter, 2023

Study Group: ———————————— ———————————— ————————————

Please upload your solution in a single pdf file on Canvas. Include all calculations, R-code, and figures (if applicable). All data sets are available on Canvas https://canvas.uw.edu/courses/1635461.

Question 1. A study on the optic nerve anatomy in myopia and glaucoma reported the following measurements on the area of lamina cribrosa (mm^2) from human optic nerve heads:

$$2.74 \; 2.63 \; 2.74 \; 3.85 \; 2.34 \; 2.74 \; 3.93 \; 4.21 \; 3.89$$
$$4.33 \; 3.45 \; 4.52 \; 2.43 \; 3.66 \; 2.78 \; 3.56 \; 3.03 \; 3.00$$

(a) Compute the median and IQR of the data by hand.

(b) Calculate $\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} x_i^2$.

(c) Use (b) to compute mean and standard deviation.

(d) By accident, the value 3.93 is recorded as 39.3. How much does this affect the average? How much does this affect the median?

Question 2. The data set `iris.csv` contains three plant species (setosa, virginica, versicolor) and four features measured for each sample. These quantify the morphologic variation of the iris flower in its three species, all measurements given in centimeters.

(a) Draw the histogram of the distribution of the petal length of the iris flower: use `breaks=seq(0,8,l=17)` to set the spacing of the bins.

  (i) Modify the title of the histogram and the labels of x- and y-axes to make the histogram more informative.

  (ii) What is the shape of the distribution of petal length of the iris flowers in the sample?

  (iii) What percentage of the iris flowers in the sample have a petal length of less than or equal to 2cm?

(b) Draw individual histograms of the distribution of petal length for each plant species: use `breaks= seq(0,8,l=17)` to set the spacing of the bins.

  (i) Modify the titles of the histograms and the labels of x- and y-axes to make the histograms more informative.

  (ii) What is the most striking difference between the three distributions?

  (iii) Based on the histograms, which plant species has the least variability in its petal length? Explain your reasoning. Verify your claim using the appropriate R function.

Question 3. Consider the adjusted closing prices of SP500 index, IBM stock, Apple Inc., and Johnson & Johnson from Jan. 1, 2000 to September 8, 2016.

(a) Plot the time series of the adjusted closing prices of all four assets in a single pane. Label axes and time series.

(b) Compare the distributions of the log-returns of all four assets using a side-by-side boxplot.

(c) Compute the summary statistics of the log-returns of all four assets using the function `summary()`.

(d) Compute the IQRs and the standard deviations of the log-returns of all four assets. Rank the assets according to their risk.

Question 4. Let $x_1, \ldots, x_n$ be a sample. Recall the definition of the sample standard deviation $s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$, where $\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$. Show that $s_n^2 = (n-1)^{-1} (\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2)$.

Question 5. Which of the following two statements is more probable? Explain your reasoning.

(i) A tornado devastating a city somewhere in North America next year and killing 25 people.

(ii) A tornado striking the city of LaGrange, Georgia, some time next year, causing the death of 25 people.

Question 6. We toss a fair coin three times.

(a) Write down the sample space Ω.

(b) Write down the set of outcomes corresponding to each of the following events:

   (i) A = "we throw tails exactly two times".
   (ii) B = "we throw tails at least two times".
   (iii) C = "tails did not appear *before* a head appeared".

(c) Compute the probabilities of events A, B, and C.

Question 7. Consider an urn with 3 red balls, 2 blue balls, and 7 black balls. Three balls are selected at random *with replacement*, i.e. each time a ball is selected it is returned to the urn.

(a) What is the probability of selecting one red and two black balls?

(b) What is the probability of selecting exactly two black balls?

(c) What is the probability of selecting exactly two balls of the same color?