MATH255 - Autumn 2023 Computer Lab - Week 10

Note. Question 1 is your Lab Preparation exercise this week. Your answers must be submitted as a pdf document before the start of your lab.

Key Results from Lectures

- Probability of union. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Probability of intersection. $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
- Probability of intersection (independent). $P(A \cap B) = P(A)P(B)$
- Conditional probability. $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Bayes' Rule. $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}$
- Mean. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Variance. $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Correlation coefficient. $r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$
- Best-fit line. $y = mx + b$, $m = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, b = \bar{y} - m\bar{x}$

1. Below are three incomplete two-way tables describing Events A, B and C in the same sample space.

	B	B^c	
A	0.1	0.3	
A^c	0.4		

	C	C^c	
B	0.15		
B^c			0.5
		0.4	

	C	C^c	
A			
A^c		0.15	0.6
	0.6		

   (a) Fill in the blank cells in all the tables with the proper values.

   (b) With the additional information $P(A \cap B \cap C) = 0.05$, use (a) to draw a Venn diagram in R with the three events labelled ('category' command) and all regions filled by using three colours of your choice ('fill' command).

2. (a) You wake up one day with red spots on your face. You call a doctor, who tells you that 90% of people who have smallpox have your symptoms and the chance of a healthy person developing your symptoms is 8.1%. Since smallpox is fatal, you are naturally terrified. However, after you think for a while, you realise that the information the doctor gave you isn't really what you want to know. You do some research online and learn that the prevalence of smallpox in the general population is one in a thousand. What is it that you want to know, and what is the answer?

   (b) Now suppose you are the doctor, confronted with a patient with red spots. The symptoms are consistent with chickenpox, but also with smallpox. You know that 80% of people with chickenpox have these symptoms. Given that the prevalence of chickenpox in the general population is one in ten, how likely is it that your patient has chickenpox?

3. The `iris` data is pre-loaded into RStudio.

   (a) Using the first four columns only (sepal length/width and petal length/width), create a scatterplot matrix with the `pairs` command.

   (b) Choose one of your graphs that appears to show a linear correlation. Make a scatterplot of just that one and insert the best-fit line (recall last week's commands for the stackloss data). Are there any outliers or influential points?

   (c) Calculate the correlation coefficient for your selected data. Is your assumption of linear correlation correct?

   (d) What is the equation of your least-squares line? Hint: give the `lm(...)` portion of your `abline` command a variable name and type `summary(your variable name)`. There is a lot of information there, but the Estimate column gives you the intercept and slope of your line.

   (e) Choose a value of the independent variable of your line that is outside the range (to the right, but still a reasonable flower petal/sepal size) of your plot and predict what the other variable value would be.