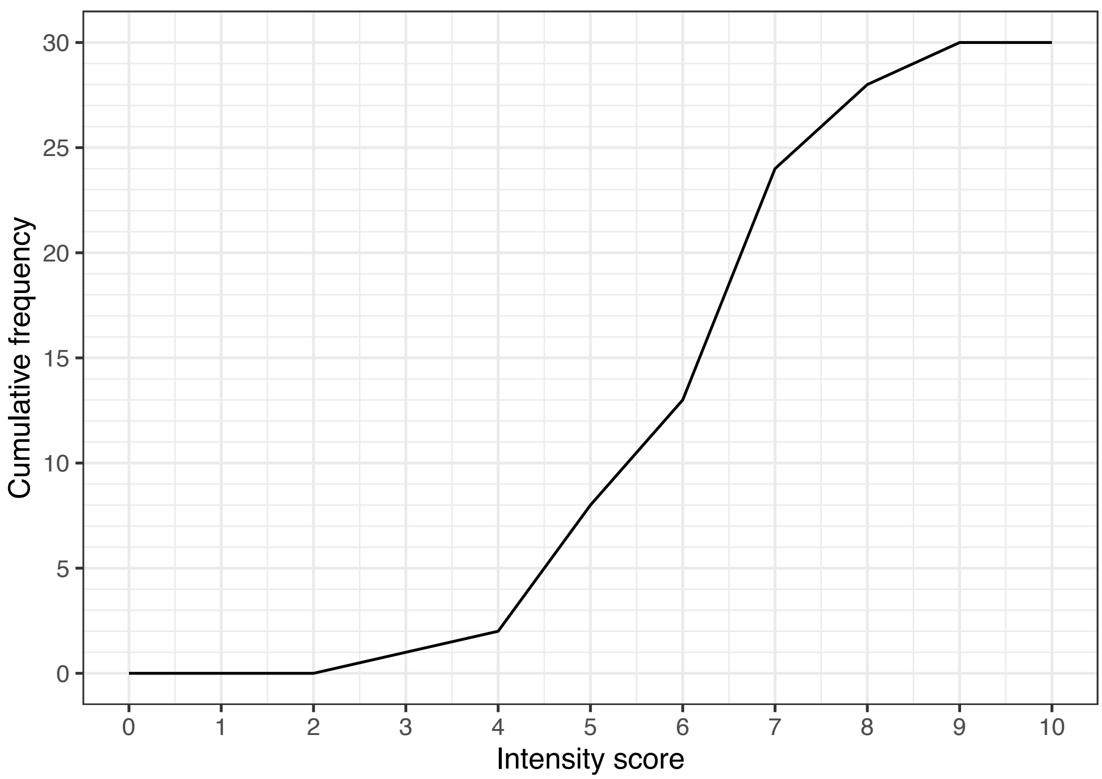
SM-1402 Basic Statistics Assignment 3 – Semester 2 2019/20

This assignment counts for 30% of your summative assessment. Attempt all questions and submit your solutions online via Canvas by 12:00 pm on **Tuesday, 21 April 2020**. The total marks attainable is 50. Late submissions will be penalised.

1. Naja runs a spin cycle class at the gym. After the class, she asks the participants to rate the intensity of the class on a scale of 0 to 10, with 0 being the lowest intensity possible (class is not challenging at all), and 10 being the highest intensity (class is extremely challenging). Ratings are given in whole numbers only. She collects the data and draws a cumulative frequency polygon, as shown below.



   (a) (1 mark) How many participants in Naja's class were there altogether?

   (b) (3 marks) What was the minimum, maximum, and the range of intensity scores, according to the responses from the class?

   (c) (4 marks) Calculate the mean and median intensity score.

   (d) (5 marks) Calculate the standard deviation, the quartiles and the interquartile range of the data.

2. (12 marks) Suppose that a the normal rate of infection for a certain disease in cattle is 25%. To test a new serum which may prevent infection, three experiments are carried out. The test for infection is not always valid for some particular cattle, so the experimental results are "inconclusive" to some degree–we cannot always tell whether a cow is infected or not. The results of the three experiments are

(a) 10 animals are injected; all 10 remain free from infection.

(b) 17 animals are injected; more than 15 remain free from infection and there are two doubtful cases.

(c) 23 animals are infected; more than 20 remain free from infection and there are three doubtful cases.

Which experiment provides the strongest evidence in favour of the serum? Explain your answer.

Hint: In each case, calculate the probabilities that out of n cows, more than x remain free from infection under the assumption that the serum has no effect at all, i.e. cows get infected at the normal rate of infection.

3. Let X be a continuous, uniform random variable that lies in the interval $[0, 1]$. We write $X \sim \text{Unif}(0, 1)$, with $f(x) = 1$ for $0 \leq x \leq 1$.

   (a) (2 marks) Determine the cdf $F(x) = \text{P}(X \leq x)$.

   (b) (6 marks) Using the cdf or otherwise, calculate the following probabilities

      i. $\text{P}(X > 1)$

      ii. $\text{P}(X > 0.2)$

      iii. $\text{P}(X \geq 0.2)$

      iv. $\text{P}(X^2 > 0.04)$

   (c) (4 marks) Obtain the mean and variance of X.

4. (a) (5 marks) Let $X \sim \text{N}(\mu_x, \sigma_x^2)$ and $Y \sim \text{N}(\mu_y, \sigma_y^2)$. Given that X and Y are independent of each other, what is

      i. $\text{E}(X - Y)$?

      ii. $\text{Var}(X - Y)$?

      iii. The distribution of the random variable $Z = X - Y$?

   (b) (4 marks) A company manufactures rods whose diameters are normally distributed with a mean of 5 mm and a standard deviation of 0.05 mm. It also drills holes to receive the rods and the diameters of these holes are normally distributed with a mean of 5.2 mm and a standard deviation of 0.07 mm. The rods are allocated to the holes at random. What proportion of rods will fit into the holes?

   (c) (4 marks) Suppose that $X = \log(Y)$ and that $X \sim \text{N}(0.5, 0.16)$. What is the probability of $\text{P}(Y > 1)$?