

WEEK 7 REPORT – PROBABILITY AND DISTRIBUTIONS

Group Number: 25

Date and Time of Tutorial Session: Wednesday 4 - 5:30 pm

Group Members (First and Last name – no student numbers):

1. Katarzyna Fraser
2. Annika Tran
3. Hanna Asin
4. Ella Pustil

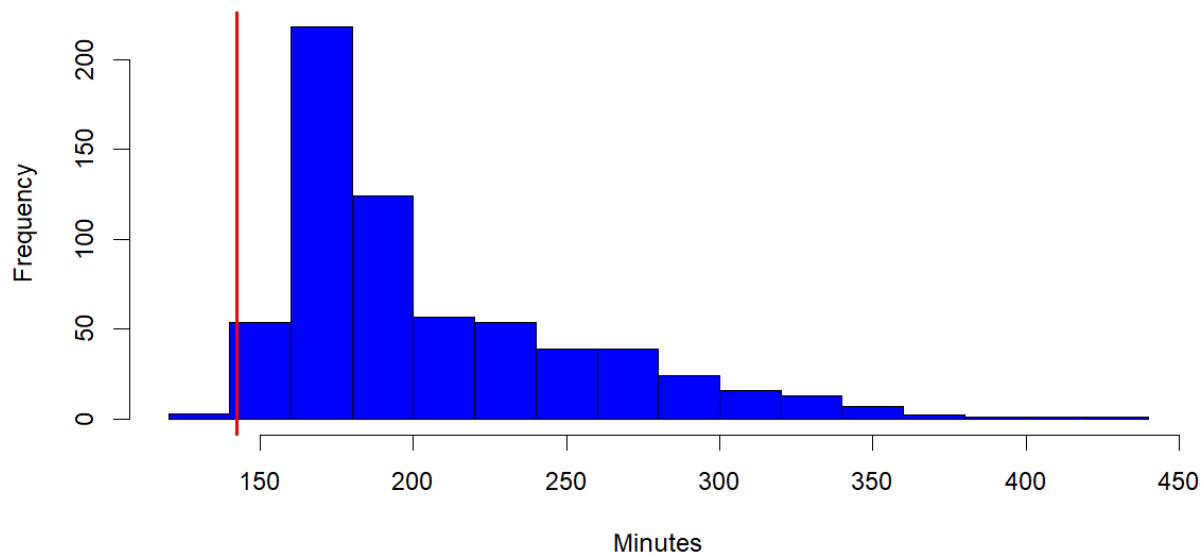
Graded out of 10 marks.

1. What are the Boston Marathon winning times for 'male' and 'female' participants registered to compete in 1968?
(1 mark)

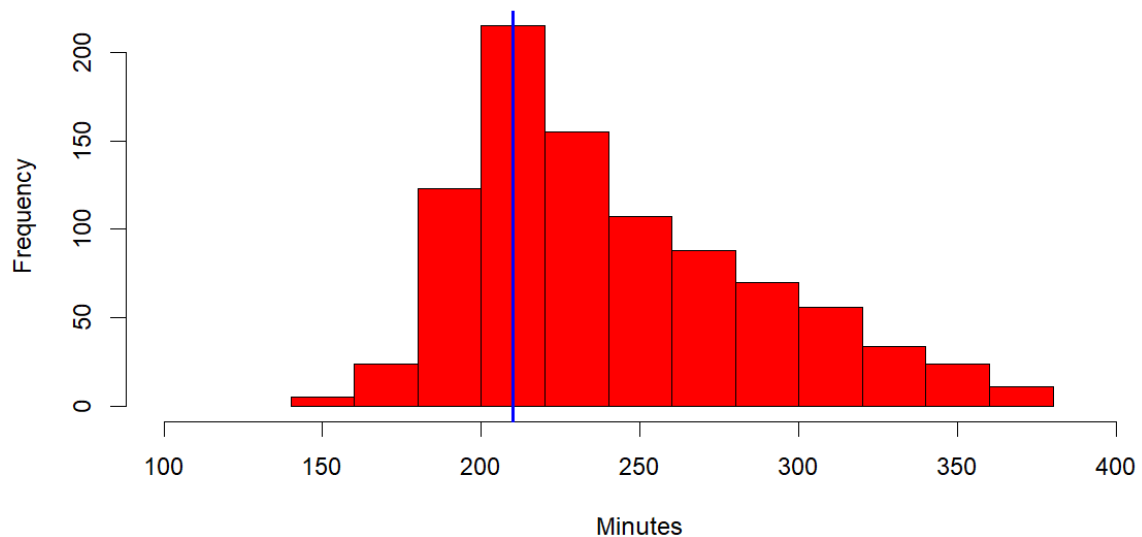
Sex	Minutes
Male	142.3
Female	210.0

2. Display your histograms for female competitors (a) and male competitors (b) in the 2019 race for under 30 years as a pair. Include a descriptive figure caption.
(2 marks)
a)

Men Under 30 Race Time in Minutes in 2019 who ran the the Boston Marathon



Women Under 30 Race Time in Minutes in 2019 who ran the the Boston Marathon



b)

Figure 1. a) Is a histogram of men under the age of 30 who ran the Boston Marathon and their race time in minutes in 2019 with a red vertical line representing the men's winning race time of 1968 and b) is a histogram woman under the age of 30 who ran the Boston Marathon and their race time in minutes in 2019 with a blue vertical line representing the women's winning race time of 1968

3. What percentage of racers in 2019 beat the winning time from 1968?
(1 mark)

Category	Percent %
Males under 30 years	0.46
Females under 30 years	28.07

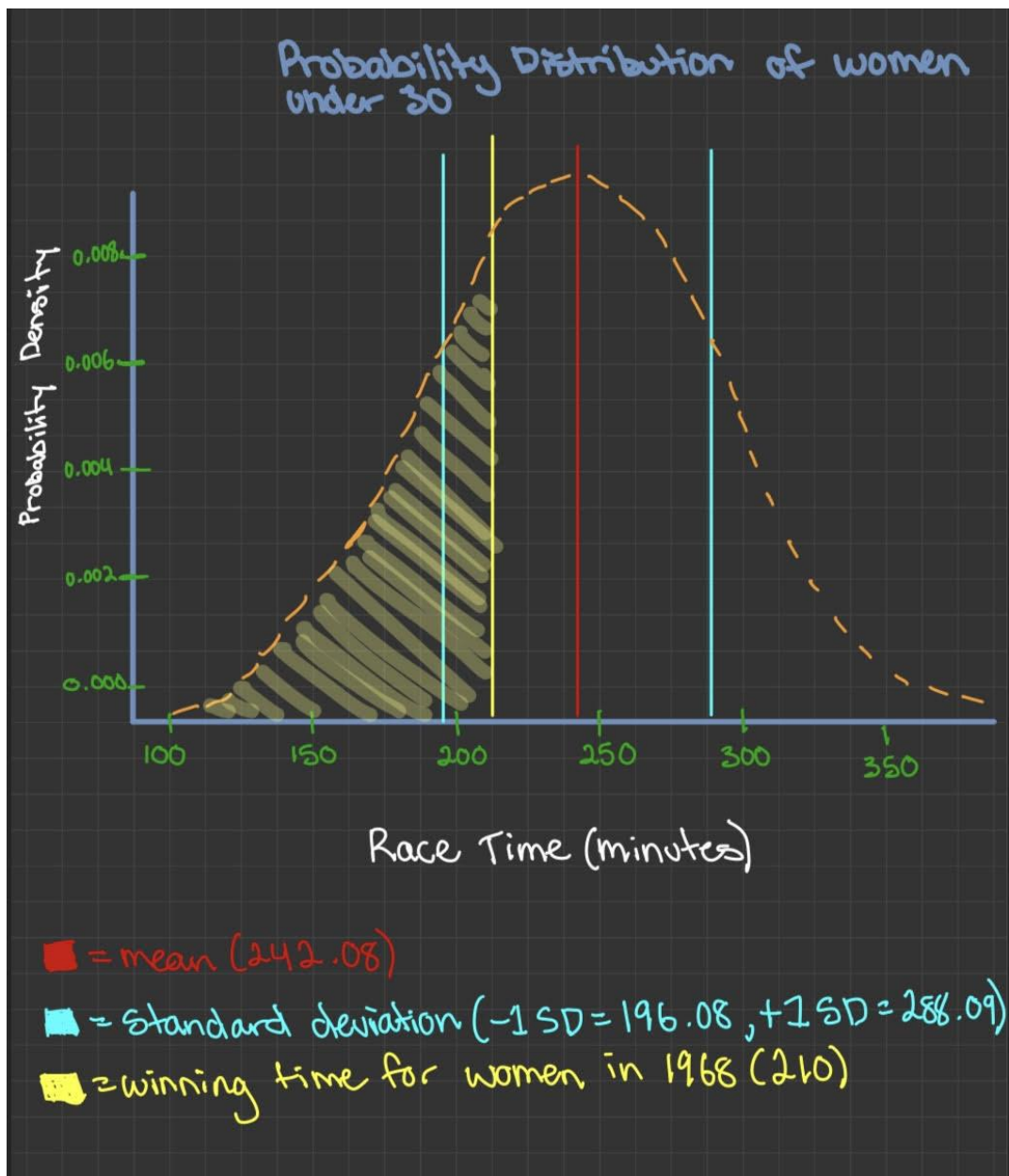
4. Calculate the mean race time and standard deviation for each category. Include these in the table below.
(1 mark)

Category	Mean	Standard deviation
Males under 30 years	206.20	49.67
Females under 30 years	242.08	46.00

5. Include a photo of your sketch of

the Normal probability distribution for females under 30 (including mean and SD's). Draw the winning time from 1968 as a line and shade in the probability of an athlete beating this time in 2019.

(2 marks)



6. a) Using the appropriate function in RStudio, calculate i) the expected probability of an athlete in 2019 beating the 1968 time if race times were normally distributed and ii) the race time (in minutes) corresponding to the fastest 10% of racers for men under 30 years and women under 30 years in 2019. Include these in the table below. (1 mark)

Category	Probability of being faster in 2019 than in 1968	Race time (minutes) of the fastest 10% of athletes in 2019
Males under 30 years	0.10	142.53

Females under 30 years	0.24	183.13
------------------------	------	--------

b) Compare your above approximation of the percentage of women under 30 being faster in 2019 than 1968 to your calculation in Question 3. Are these values similar? Explain why these two values are similar or different.
(2 marks)

The two values are different. The probability calculated of women under 30 in 2019 being faster than 1968 was predicted to be approximately 0.24% which is quite small, but the actual data shown in question 3 displays that about 28% of women under age 30 in 2019 were faster than women in 1968. The two values could be different as a bigger population of women from 1968 to 2019 started to exercise more. This could be from more gyms opening and classes marketing towards women such as Pilates.

The reason the two is different is because one is the actual data calculation with the actual data, and the other is predictive from the normal distribution.

Notes:

- Only **one** group member submits the report
- The report must be a **Word .DOC, .DOCX or .PDF** file
- Make sure everyone in the group has a copy of the report
- Double check what you have submitted!!
 - view it on OnQ to make sure everything is there and visible
- Lastly, everyone in the group needs to submit their own version of the R script.