Homework 2

Statistical Inference, Fall 2021

1- Sixty students have lined up to enter the final exam of the Graph Theory course. Each of these 60 people has a card corresponding to a class chair. Unfortunately, the first person loses her card and then randomly chooses a chair and sits on it. Each of the subsequent people will sit on his/her seat if it is empty. Otherwise, they will select a vacant seat at random and sit on it. What is the probability that the last person to enter the class will sit in his/her chair?

2- Gloria has become the project manager of a software company recently. She has decided to give each software developer a day off whenever it is the birthday of at least one of them. Except for these days, the developers work a 365-day year. Gloria wants to increase the productivity of her company. How many developers should she hire to reach this goal?

3- Benedict is a pizza salesperson. He sells on average 20 pizzas on a round of his route.

   a. Explain why the distribution of the number of pizzas follows Poisson.

   b. What is the chance that he sells an even number of pizzas?

   c. **(R)** Do the calculations of part b in the form of an R script.

4- There are N students currently enrolled in the Statistical Inference course. All students will vote on the Elearn website and they will either vote for date A or date B as potential exam dates. Ali has unofficially asked the students and realized that *pN* of the students support date A, and *(1-p)N* of the students will vote for date B, where N is known and $p \in (0, 1)$. Unfortunately, because the Elearn website has some technical problems (as always), each student will randomly and independently be kicked out of the website with a probability of 0.5, and therefore, will not vote. Let X_A be the number of people who have successfully voted for date A and let X_b be the number of people who have successfully voted for date B.

   a. What are the **mean** and **standard deviation** of random variables X_A and X_B in terms of p and N?

   b. Show that when N is large enough, the fraction of students who voted for date A should be very close to p.

   Now suppose that we have fixed the website and want to repeat the vote. This time, some students are "busy" and will vote with probability $\frac{1}{4}$ and will not participate with probability $\frac{3}{4}$. Others are "free" and will vote with probability $\frac{3}{4}$ and will not participate with probability $\frac{1}{4}$. Suppose that a fraction q_A of the supporters of date A are busy and $1 - q_A$ are free, and a fraction q_B of the supporters of date B are busy, and $1 - q_B$ are free.

   c. show that:

   $$E[X_A] = \frac{1}{4} q_A \, p \, N + \frac{3}{4} (1 - q_A) \, p \, N$$

   $$E[X_B] = \frac{1}{4} q_B \, (1 - p) \, N + \frac{3}{4} (1 - q_B) \, (1 - p) \, N$$

   d. Does the statement in part b hold here as well?

Homework 2

Statistical Inference, Fall 2021

5- Answer the following questions.

a. What is the difference between the "Independence" and "Disjointness" of random variables?
b. Is it possible for two disjoint variables to be independent as well? If yes, explain when it happens.
c. There are ***n*** children in a particular family. Find the ***n*** for which the events A and B are independent:

A: "The family has children of both sexes."

B: "There is at most one boy."

6- A family has two children, each of the children is either a girl or boy. Find the probability of both children being boys in the following scenarios:

a. We asked their mother, "Do you have at least one son?" and she answered, "Yes!".
b. We asked their mother, "Do you have at least one son named 'Ali' ?" and she answered "Yes!". (Assume that if the family has a son, they name him "Ali" with probability α).
c. Are the answers for the two scenarios different? Explain why.

7- Below is the structure of a 16-Team Tennis Tournament (Single Elimination) where "Federer" and "Nadal" both will participate.

a. What is the chance that Federer and Nadal will meet in a match during the tournament?

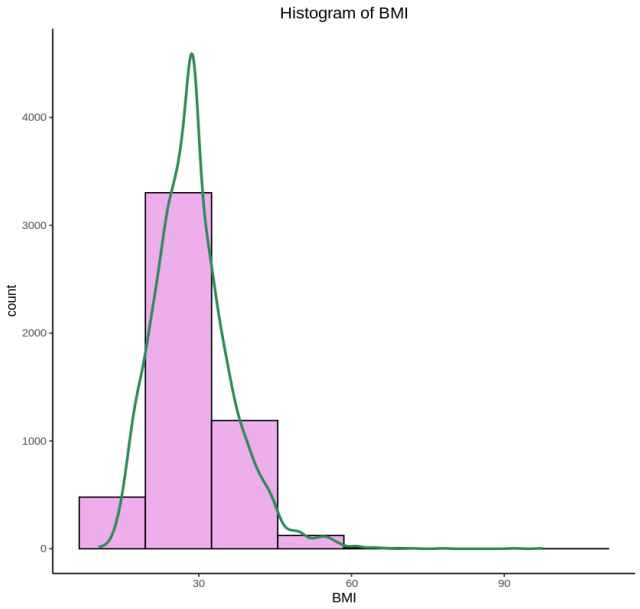
b. Solve the problem for a tournament with 2^n participants instead of 16 teams.

Homework 2

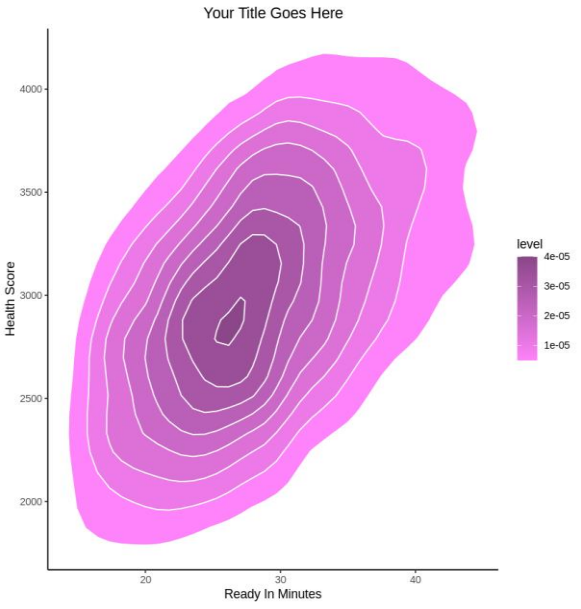
Statistical Inference, Fall 2021

8- **(R, ggplot2)** In this question, you are going to use the "Foods" dataset. This dataset consists of some details about 1,700 foods. ***Note that you must use the ggplot2 library to draw the diagrams.***

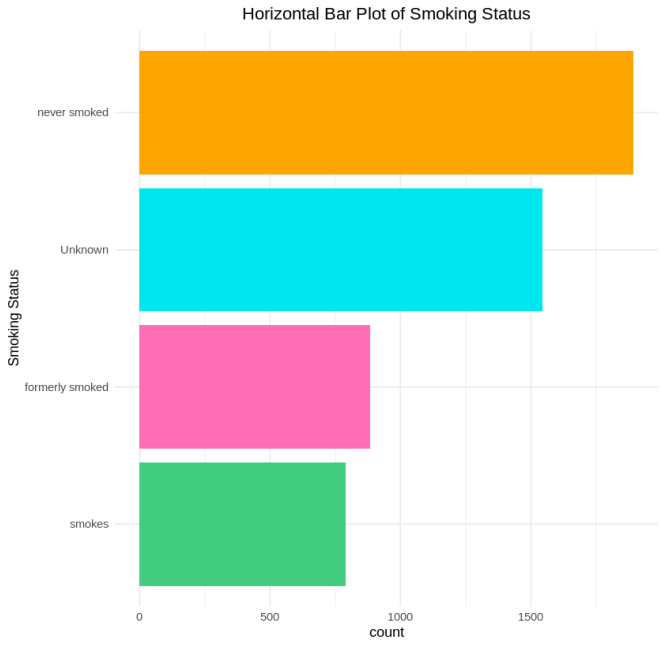
a. Plot the histogram for "pricePerServing" with an appropriate bin size, then overlay that with the density curve. Your diagram must be similar to the following figure.



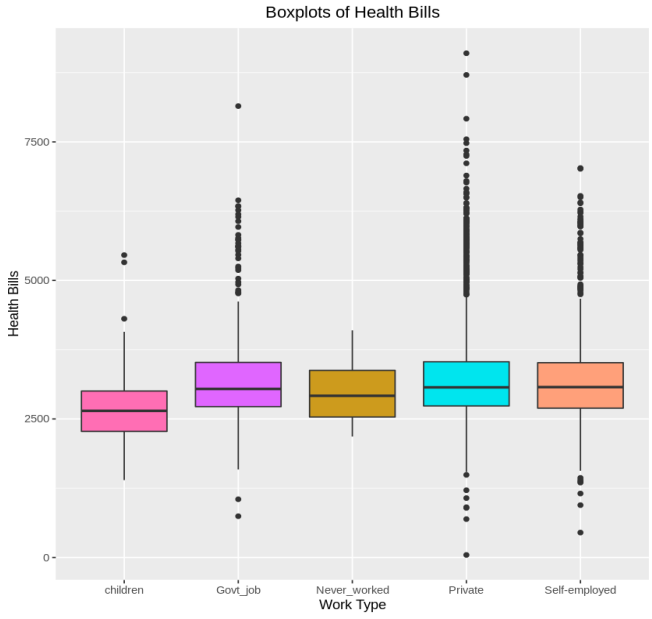
b. Draw the 2D density plot of "healthScore" and "readyInMinutes". Your output must be similar to the following image.



c. Sort the categories in “dishType” by their frequencies, then draw a horizontal barplot to show the result. Your output must be similar to the following image.



d. Draw the separate boxplots of the “healthScore” variable for each “dishType". Your diagram must look like the following image.



e. Draw the mosaic plot of “veryHealthy” and “dairyFree”. Your output must be similar to the following image. Please pay attention to all of the details you can see in this figure.