Homework 5

Statistical Inference, Spring 2021

1- We are interested in estimating the proportion of graduates at a mid-sized university who found a job within one year of completing their undergraduate degree. Suppose we conduct a survey and find out that 348 of the 400 randomly sampled graduates found jobs. The graduating class under consideration included over 4500 students.

   a. Describe the population parameter of interest. What is the value of the point estimate of this parameter?

   b. Check if the conditions for constructing a confidence interval based on these data are met.

   c. Calculate a 95% confidence interval for the proportion of graduates who found a job within one year of completing their undergraduate degree at this university, and interpret it in the context of the data.

   d. What does "95% confidence" mean?

   e. Now calculate a 99% confidence interval for the same parameter and interpret it in the context of the data.

   f. Compare the widths of the 95% and 99% confidence intervals. Which one is wider? Explain.

2- Greece has faced a severe economic crisis since the end of 2009. A Gallup poll surveyed 1,000 randomly sampled Greeks in 2011 and found that 25% of them said they would rate their lives poorly enough to be considered "suffering"

   a. Describe the population parameter of interest. What is the value of the point estimate of this parameter?

   b. Check if the conditions required for constructing a confidence interval based on these data are met.

   c. Construct a 95% confidence interval for the proportion of Greeks who are "suffering".

   d. Without doing any calculations, describe what would happen to the confidence interval if we decided to use a higher confidence level.

   e. Without doing any calculations, describe what would happen to the confidence interval if we used a larger sample.

3- A survey on 1,509 high school seniors who took the SAT and who completed an optional web survey shows that 55% of high school seniors are fairly certain that they will participate in a study abroad program in college.

   a. Is this sample a representative sample from the population of all high school seniors in the US? Explain.
   b. Let's suppose the conditions for inference are met. Even if your answer to part (a) indicated that this approach would not be reliable, this analysis might still be interesting to carry out.

Construct a 90% confidence interval for the proportion of high school seniors (of those who took the SAT) who are fairly certain they will participate in a study abroad program in college, and interpret this interval in context.

c. Based on this interval, would it be appropriate to claim that the majority of high school seniors are fairly certain that they will participate in a study abroad program in college?

4- An experiment conducted by the "Myth Busters", a science entertainment TV program on the Discovery Channel, tested if a person can be subconsciously influenced into yawning if another person near them yawns. 50 people were randomly assigned to two groups: 34 to a group where a person near them yawned (treatment) and 16 to a group where there wasn't a person yawning near them (control). The following table shows the results of this experiment.

	Group		
	Treatment	Control	Total
Yawn	10	4	14
Not Yawn	24	12	36
Total	34	16	50

A simulation was conducted to understand the distribution of the test statistic under the assumption of independence: having someone yawn near another person has no influence on if the other person will yawn. In order to conduct the simulation, a researcher wrote yawn on 14 index cards and not yawn on 36 index cards to indicate whether or not a person yawned. Then he shuffled the cards and dealt them into two groups of size 34 and 16 for treatment and control, respectively. He counted how many participants in each simulated group yawned in an apparent response to a nearby yawning person, and calculated the difference between the simulated proportions of yawning as $\hat{P}_{trtmt;sim} - \hat{P}_{ctrl;sim}$.

a. What are the hypotheses?
b. Calculate the observed difference between the yawning rates under the two scenarios.
c. Estimate the p-value and determine the conclusion of the hypothesis test.

5- Aria noted that in the previous year, 50 percent of workers at the milk factory where she works had 0-15 minute wait times, 30 percent had 16-30 minute wait times, 10 percent had 31-45 minute wait times, and 10% percent had wait times that were 46 or more minutes long. He wondered if the worker's wait times this year still followed this distribution, so he took a random sample of 150 workers and recorded the lengths of their wait times. Here are his results:

Wait time (minutes)	0-15	16-30	31-45	46+
workers	75	55	15	5

He wants to use these results to carry out a chi-squares test to determine if the distribution of wait times this year differs from last year.

a. What are the hypotheses?
b. What are the values of the test statistic and P-value for Aria's test?
c. **(R)** Calculate P-value and compare it with part b.

6- A student wanted to see if different breeds of cats preferred toys of certain colors. The student presented each cat with 3 toys, identical except that each was a different color, and recorded which toy the cat played with the most in a set time period. Here are the outcomes and partial results of a chi-square test.

Chi-square test: Breed vs. toy color

	Yellow	Purple	Black	Total
Manx	16	30	14	60
Thai	27	23	10	60
Toyger	29	16	15	60
Total	72	69	39	180

They want to use these results to carry out a $\chi 2$ test for homogeneity. Assume that all conditions for inference were met.

a. What is the expected count for the row corresponding to Toyger?
b. Write the hypotheses for the test you identified.
c. What are the values of the test statistic and P-value for this test?

7- **(R)** The experimenters investigated whether people can recognize the smell of their coffee. 24 practitioners of People who drink coffee were kept in an isolated room. The coffees used by these practitioners are served in the room. The practitioner should be able to identify her/his coffee. There were a total of 300 trials, of which the practitioner correctly recognize his/her smell of coffee 125 times.

a. What is the null hypothesis, i.e., how often would we expect the participants to be correct by chance (in raw number and percentage)?
b. Using a chi-square test, what do you conclude about whether practitioners recognize the smell of their coffee?

8- **(R)** Consider the dataset named "survey" in the "MASS" library. The Smoke column records the students smoking habit, while the "Exer" column records their exercise level. Test the hypothesis whether the students smoking habit is independent of their exercise level at .05 significance level.

Note: You are not allowed to use "inference.R" package in this homework.