Homework 4

Statistical Inference

1- Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

   a. By resampling the population with replacement, a bootstrap distribution is created.

   b. Suppose that ANOVA at a 5% significance level rejects the null hypothesis that the means of the four groups are the same. Hence, the pairwise analysis will identify at least one pair of significantly different means.

   c. A smaller sample size is recommended when generating resampled samples.

   d. The difference between each pair of observations is the starting point for a paired analysis. Then we use these differences to make inferences.

   e. There is a reverse relation between variance within an ANOVA group and the amount of noise (variance).

   f. In a test, we randomly sample 30 items from Amazon and note the price for each. Then we visit local markets and collect the price of each of those same 30 items. This is a paired t-test.

   g. An inference is made based on the difference between two observations in a paired analysis.

   h. In order to prevent Type-II errors, you should assign a large amount of β.

2- A statistical inference class at a certain university has 500 students. The scores of 10 students were selected at random and are demonstrated in the following table.

   a. Calculate the mean and standard deviation of the sample.

   b. Calculate the margin of error.

   c. Construct a 90% confidence interval for the mean score of all the students in this class.

   d. Interpret the calculated confidence interval

76	84	69	92	58
89	73	97	85	77

3- New York is known as “the city that never sleeps”. A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below. The point estimate suggests New Yorkers sleep less than 8 hours a night on average. Is the result statistically significant?

n	$\bar{x}$	s	min	max
25	7.73	0.77	6.17	9.78

a. Write the hypotheses in symbols and words.

b. Check conditions, then calculate the test statistic, T, and the associated degrees of freedom.

c. Find and interpret the p-value in this context.

d. What is the conclusion of the hypothesis test?

e. If you were to construct a 90% confidence interval that corresponded to this hypothesis test, would you expect 8 hours to be in the interval?

4- An experiment is performed to determine whether intensive tutoring (covering a great deal of material in a fixed amount of time) is statistically different from (is more/less effective) paced tutoring (covering less material in the same amount of time). Two randomly chosen groups are tutored separately and then administered proficiency tests. Use the significance level $\alpha = 0.05$.

Group	Method	n	$\bar{x}$	s
1	Intensive	12	46.31	6.44
2	Paced	10	42.79	7.52

5- In order to conduct medical research, a medical research group is seeking participants to complete short surveys about their medical history. One survey, for example, asks about a person's family history of cancer. As part of another survey, the respondent is asked what topics were discussed during their last hospital visit. According to our data, as people sign up, they complete approximately 4 surveys on average, and the standard deviation for the number of surveys is approximately 2.2. The research group wants to try a new interface that they think will encourage new enrollees to complete more surveys, where they will randomize each enrollee to either get the new interface or the current interface. How many new enrollees do they need for each interface to detect an effect size of 0.5 surveys per enrollee if the desired power level is 80%? Assume $\alpha = 0.05$.

6- To study the effect of cigarette smoking on platelet aggregation, researchers drew blood samples from 11 individuals before and after they smoked a cigarette and measured the percentage of blood platelet aggregation. Platelets are involved in the formation of blood clots, and it is known

that smokers suffer from disorders involving blood clots more than non-smokers. Test the null hypothesis that the means before and after are the same. Use significance level $\alpha = 0.05$.

Before	After
25	27
25	29
27	37
44	56
30	46
67	82
53	57
52	61
53	80
60	59
28	43

7- The table below shows the heights of 24 men above 20 years of age from US, UK and India. We are interested in whether or not a significant difference exists between the mean heights of these three different countries. Use significance level $\alpha = 0.05$.

Country (US)	Country (UK)	Country (India)
180	185	170
183	181	183
172	180	180
178	179	175
169	164	181
179	173	183
178	180	176
180	178	167

a. Write the hypotheses for testing if the average height of three groups of men varies.
b. Calculate and conduct analysis using one-way ANOVA and complete the table.

		DF	Sum SQ	Mean SQ	F-value
Group	Class				
Error	Residuals				
	Total				

c. What is the conclusion of the test?

8- (R) The dataset "Diet" contains information about 78 people who began to follow 3 different types of diets (referred to as diets A, B, and C). This exercise aims to see which diet is most effective at losing weight.

a. Plot the three group's data using side-by-side boxplots.

b. Use ANOVA in R to determine whether there is a significant difference in the mean weight loss between groups.

c. Display and analyze the results.

d. Compare the weight loss between the two groups in R and:
   - ✓ Write the hypothesis.
   - ✓ Report the level of significance of the test and the decision about the hypothesis.
   - ✓ Estimate the size of the difference in the mean drop; use α=0.05.