Chapters covered: Lecture notes Chapters 9.3, 9.4, 9.5, 10.2, 10.3

Show your work to receive full credit.

Problem 1: Name the tests

For each part below answer
(i) state the name of relevant hypothesis test:

- One sample z-test for proportion p
- One sample t-test for mean μ
- Matched pairs t-test for mean of difference μ_D
- Independent two sample t-test for difference of means $\mu_1 - \mu_2$

(ii) define parameter(s) of interest and state the null and alternative hypotheses.
No need to conduct 5 step hypothesis test.

Example : To test the claim that Startbucks Coffee's daily sale in Twin Cities is higher than that in Des Moines, IA, you randomly select 30 stores from each location and calculate the sample daily sales mean for each state.
Answer : Independent two sample t-test for difference of means.
μ_1: mean daily sale of Startbucks coffee's in Twin Cities
μ_2: mean daily sale of Startbucks coffee's in Des Moines, IA
$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 > \mu_2$

a) It is known that 40% of college students in the U.S. identified Instagram as their favorite social media application. To test if this information is still accurate in 2022, you surveyed 100 college students randomly and found 32% of them identified Instagram as their favorite.

b) Forty-four sixth graders were randomly selected from a school district. Then, they were divided into 22 matched pairs, each pair having equal IQ's. One member of each pair was randomly selected to receive special training. Then all of the students were given an IQ test. The researcher wants to test if the special training improves IQ scores.

c) An experiment investigated whether cell phone use increases drivers' reaction times, using a sample of 100 college students. Students were randomly assigned to a cell phone group or to a control group, 50 to each. In a simulation of driving situations, a target flashed red or green at irregular periods. Participants pressed a brake button as soon as they detected a red light. The control group listened to the radio or audiobook while they performed the simulated driving. The cell phone group carried out a phone conversation about a political issue with someone in a separate room. The experiment measured each group's mean response time over many trials.

d) You want to know if students at the University of Minnesota sleeps more than 7 hours on average. To investigate this, you sampled a few class mates and asked how many hours they sleep on a typical week day.

e) Lucy wanted to test if the selling prices of e-book is cheaper than prices of hard cover book. She first took a random sample of 7 books. Then she looked up e-book prices and hard copy prices of those 7 books.

Problem 2: Body temperature

98.6 F is commonly referred as the average body temperature. Is this information accurate? A study was conducted in 1992 to examine this belief. The oral body temperatures of a random sample of 100 healthy adults were measured. The mean and standard deviations are $\bar{x} = 98.2F$ and $s = 0.7F$. Use this sample information to test whether the true human body temperature is equal to 98.6 or not. Use $\alpha = 0.05$.

- Assumption :
- Hypotheses :
- Test statistic: What is the distribution of test statistic? Based on the sample information provided in Problem 2, what is the value of the test statistic?
- P-value: Interpret the p-value in the context of the problem.
- Conclusion and interpretation in context.

Problem 3: Type I vs Type II error

A researcher wants to see if a new drug with potentially dangerous side effects is significantly better than the current drug. If it is found to be more effective, it will be prescribed to millions of people.

a) What does a Type I error mean in this scenario?

b) What does a type II error mean in this scenario?

c) Which type of error is worse to make in this situation? Explain.

Problem 4: Study time

A graduate teaching assistant collected data to investigate whether study time per week (average number of hour) differed between students who planned to go to graduate school and those who did not. The data were as follows:
Graduate school : 15, 7, 15, 10, 5, 5, 2, 3, 12, 16, 15, 37, 8, 14, 10, 18, 3, 25, 15, 5, 5
No graduate school : 6, 8, 15, 6, 5, 14, 10, 10, 12, 5
Use the following R command to input data.

```
grad<-c(15, 7, 15, 10, 5, 5, 2, 3, 12, 16, 15, 37, 8, 14, 10, 18, 3, 25, 15, 5, 5)
noGrad<-c(6, 8, 15, 6, 5, 14, 10, 10, 12, 5)
```

a) Use R to find sample mean and sample standard deviation for each group. Copy and paste your R output.

b) Show formula with numbers plugged in to verify that standard error for difference between the sample means is 2.16. Interpret this value.

c) What is the value of t-multiplier, $t_{\alpha/2,df=\min(n_1-1,n_2-1)}$, to construct a 99% confidence interval? Copy and paste your R command and output.

d) Show formula with numbers plugged in to verify that 99% confidence interval to estimate the difference between the two population mean is (-4.4, 9.59). Interpret.

Problem 5: Study time 2

Use the same information from Problem 4 to conduct a hypothesis test to test whether the average study hours for students who plan to go to graduate school is greater than average study hours for students who don't plan to go to graduate school. Use $\alpha = 0.1$. (Do not use `t.test()`)

- Assumption :
- Hypotheses :
- Test statistic : What is the distribution of test statistic? Based on the sample information provided in the Problem 4, what is the value of the test statistic?
- P-value
- Conclusion and interpretation in context.

Problem 6: Type 1 / Type 2 error

Refer to the conclusion from Problem 5. Based on the conclusion (reject the null or fail to reject the null), what type of error (either Type 1 or type 2) could you have made? Explain.

Problem 7: Study time 3

Use R command `t.test()` to repeat Problem 5. (Test whether students who plan to go to graduate school study more than students who don't at $\alpha = 0.1$). Copy and paste your R command and output.

Note to students : R uses "Welch's degrees of freedom" instead of minimum of $n_1 - 1$, $n_2 - 1$, hence gives a slightly different p-value from Problem 5.

Problem 8: Multiple choice

P-value of hypothesis test $H_a : \mu_1 \neq \mu_2$ is 0.03. Which of the following is always true? Explain.

(A) Using the same sample information, the 95% confidence interval for $\mu_1 - \mu_2$ will contain 0.

(B) Using the same sample information, the 99% confidence interval for $\mu_1 - \mu_2$ will contain 0.

(C) Using the same sample information, the 95% confidence interval for $\mu_1 - \mu_2$ will contain positive numbers only.