Introductory Statistics

Midterm # 2

Learning Objectives covered in this midterm:

Question # 1 LOs:

- Calculate the mean and standard deviation for a sampling distribution of a sample proportion
- Check the conditions for creating a confidence interval for population proportion
- Calculate a confidence interval and explain what it means
- Find the sample size needed to have a desired margin of error for the confidence interval of a proportion

Question # 2 LOs:

- Write a null and alternative hypothesis for a hypothesis test
- Complete a one-sample z-test for proportions from hypotheses to conclusions
- Recognize Type I and Type II errors and their consequences

Question # 3 LOs:

- Calculate a confidence interval for the difference in proportions of two groups
- Make conclusions based on a confidence interval
- Complete a two-sample z-test for proportions from hypotheses to conclusions

Question # 4 LOs:

- Check the assumptions for a one-sample t confidence interval for population mean
- Calculate a confidence interval for a population mean and explain what it means
- Write the null and alternative hypotheses for tests on population means
- Complete a one-sample t-test for means from hypotheses to conclusions

Question # 5 LOs:

- Check the assumptions for a two-sample t confidence interval for population mean
- Calculate a confidence interval for the difference between two population means and explain what it means
- Complete a two-sample t-test for independent population means from hypotheses to conclusions
- Describe how a two-sample confidence interval is related to a hypothesis test for the difference between two dependent population means

Question # 6 LOs:

- Complete a two-sample t-test for dependent population means from hypotheses to conclusions

Introductory Statistics - Midterm # 2

Name: Date:

1. The Lumina Foundation and Gallup would like to investigate how important are the laws in the state where the college is located on access to reproductive health services is in your decision to stay enrolled until you have graduated. Results were based on web survey responses, conducted Oct. 26 - Nov. 17, 2022, from 6,008 U.S. adults currently enrolled in an associate degree, bachelor's degree, certificate or certification program. It is reported that 72% of currently enrolled college students report that the reproductive health laws in the state where their college is located are at least somewhat important to their decision to stay enrolled.[1]

   a. What is the point estimate for the population proportion?

   b. Calculate the mean and standard error for the sampling distribution of sample proportions in this scenario.

   c. Verify that the conditions for creating a confidence interval for population proportion are satisfied in the context of this problem.

   d. Assume the level of confidence is 95%. What is the z critical value, z*, that corresponds to the confidence level?

[1] https://news.gallup.com/poll/474365/reproductive-health-laws-factor-college-decisions.aspx

e. Calculate the margin of error for the confidence interval.

f. Using the statistical tool, find the 95% confidence interval and write the interpretation of the confidence interval in context of the problem.

g. Supposed that a news article claimed that three-quarter of enrolled college students report that the reproductive health laws in the state where their college is located are at least somewhat important to their decision to stay enrolled. Using the confidence interval, does the data support the news article claim? Explain your answer.

h. What sample sizes would be needed for a 95% confidence level and a margin of error of 1%?

i. What would happen to the 95% confidence interval if we took another sample of only 3000 U.S. adults? Explain.

2. The Lumina Foundation and Gallup would like to investigate how important are the laws in the state where the college is located on access to reproductive health services is in your decision to stay enrolled until you have graduated. Results were based on web survey responses, conducted Oct. 26 - Nov. 17, 2022, from 6,008 U.S. adults currently enrolled in an associate degree, bachelor's degree, certificate or certification program. It is reported that 72% of currently enrolled college students report that the reproductive health laws in the state where their college is located are at least somewhat important to their decision to stay enrolled.[2]

   At the 5% significance level, does the data indicate that less than three-quarter of enrolled college students report that the reproductive health laws in the state where their college is located are at least somewhat important to their decision to stay enrolled?

   a. Write the null and alternative hypotheses for this scenario.

   b. Verify that the conditions for the one-sample z-test for proportions have been met.

   c. Use the statistical tool and calculate the test statistic and P-value.

   d. Will the null hypothesis be rejected? Explain.

[2] https://news.gallup.com/poll/474365/reproductive-health-laws-factor-college-decisions.aspx

e. Write the conclusion of this hypothesis test in context of the problem.

f. In this study, did you show "statistical significance?" “practical significance?” Explain.

g. In the context of the problem, what might happen if a type I error occurs?

h. In the context of the problem, what might happen if a type II error occurs?

3. In 2020, researchers worked to develop a safe and effective vaccine against SARS-CoV-2, the coronavirus that causes COVID-19. A clinical trial was conducted with more than 30,000 adult volunteers nationwide for the Moderna COVID-19 vaccine. Participants were 18 years of age or older with no known previous SARS-CoV-2 infection. Volunteers were randomly assigned to receive either two doses of the investigational vaccine (100 micrograms each) or two shots of a saline placebo.

   The investigators recorded 196 cases of symptomatic COVID-19 among participants who received the vaccine at least 14 days after they received their second shot. Only 11 of these cases were in the group that received the vaccine, with none severe. In contrast, 185 of the cases occurred in the placebo group, 30 of which were severe.[3]

   Their results are organized in the table.

	Non-severe COVID	Severe COVID
Received the vaccine	11	0
Received the placebo	30	155

   a. Calculate the associated sample statistic for the difference in proportions of severe cases for this scenario.

   b. Using the statistical tool, find the 95% confidence interval. Identify the standard error, margin of error, confidence level, and lower and upper bounds of the confidence interval for estimating the population proportion.

[3] https://www.nih.gov/news-events/nih-research-matters/experimental-coronavirus-vaccine-highly-effective

c. Write the interpretation of the confidence interval in context of the problem.

At the 5% significance level, does the data indicate that the vaccine is effective and that it can trigger an immune response against the virus without serious side effects?

d. Write the null and alternative hypotheses for this scenario.

e. Verify that the conditions testing a two-sample z-test for difference in proportions are satisfied.

f. Use the statistical tool and calculate the test statistic and P-value.

g. Write the conclusion of this hypothesis test in context of the problem.

4. The average score on the math section of 1,737,678 test takers of the 2022 SAT is 521 out of 800.[4] For a simple random sample of 100 students, we found the same average with the standard deviation is 120.

   a. Verify the conditions for creating a confidence interval for population mean.

   b. Using the statistical tool, find the 95% confidence interval and write the interpretation of the confidence interval in context of the problem.

   c. Suppose that we want to test the hypothesis with a significance level of 0.05 that the average score on the math section of the SAT scores for a sample of 100 Lumen high school students who took the test is higher than the national average. Suppose that the average score on the math section of the SAT scores of 100 students at Lumen high school is 540 with a standard deviation of 120. What can we conclude?

[4] https://reports.collegeboard.org/media/pdf/2022-total-group-sat-suite-of-assessments-annual-report.pdf

5. The average scores on the math section of the 2022 SAT for Black students (452) are considerably lower than those of Asian students (633), Hispanic/Latino students (473), Native Hawaiian/Other Pacific Islander students (464), White students (543), American Indian/Alaska Native students (463), and Two or More Races students (543).[5]

   Let's use a two-sample t confidence interval and two-sample t-test for population means to investigate if there are significant differences on the average scores on the math section of the 2022 SAT for Black students and Hispanic/Latino students.

   Note: The number of test takers who identify as Black students on the 2020 SAT is 201,645. The number of test takers who identify as Hispanic/Latino students on the 2020 SAT is 396,422. We will use the standard deviation of the math sections of all of the 2022 SAT for both races/ethnicity, which is 120.

   a. Verify that the conditions for the population means have been met.

   b. Using the statistical tool, find the 95% confidence interval. Identify the point-estimate, standard error, margin of error, confidence level, and lower and upper bounds of the confidence interval for estimating the differences on the average scores.

   c. Write the interpretation of the confidence interval in context of the problem.

[5] https://reports.collegeboard.org/media/pdf/2022-total-group-sat-suite-of-assessments-annual-report.pdf

d. Write the null and alternative hypotheses for this scenario.

e. Use the statistical tool and calculate the test statistic and P-value.

f. Will the null hypothesis be rejected? Explain.

g. Write the conclusion of this hypothesis test in context of the problem.

h. An article has been published with the title “SAT math scores mirror and maintain racial inequity”. Based on the information provided and your analysis between two races/ethnicity, do you agree or disagree with the title of the article? Justify your answer using your statistical analyses results above.

6. Does 10K running time decrease when the runner listens to music?
   Nine runners were timed as they ran a 10K with and without listening to music. The running times in minutes are shown below.

With music	53	50	38	53	54	38	63	53	42
Without music	56	54	42	46	54	41	68	58	41

   Assume a normal distribution, what can be concluded at 0.01 level of significance?

   a. For this study, explain why t-test for the difference between two dependent population means should be used instead of the test for two independent population means.

   b. Write the null and alternative hypotheses for this scenario.

   c. Use the statistical tool and calculate the test statistic and P-value.

   d. Interpret the P-value in context of the study.

   e. Write the conclusion of this hypothesis test in context of the problem.