

Math 247: Final Exam
(100 pts)

Name: _____

1. (25 pts) **Vitamin C** A study (double-blind) was done investigating the therapeutic value of vitamin C (ascorbic acid) for treating common colds. The study (done in 1971 by Linus Pauling) was conducted during a 2-week period on a sample of 279 school children in a skiing camp in the Swiss Alps. The participants were split into two groups (assume random assignment). In one group the kids took 1 gram of vitamin C per day; in the other group, the kids took a placebo. At the end of two weeks the researchers assessed who had gotten a cold and who hadn't.

A. Is the sample (using the kids at the ski-camp) considered a simple random sample from the population of all school children? Yes No

	Cold	No Cold
Placebo	31	109
Vitamin C	17	122

The name for this type of sample is

“SAMPLE OF _____”

B. Would it be correct to apply the findings of this study to all school age children? Why or why not?

C. Is this study a Controlled, Randomized Experiment or an Observational Study? How can you tell?

D. Which of the following hypothesis tests could be used to see whether there was a significant difference in the proportion of kids who got colds, based on whether or not they took Vitamin C? (Circle the answer)

z-Test for Two Proportions

One Sample t-Test

Two Sample t-Test

Paired Difference Test

E. If Dr. Pauling wanted to test whether proportion of kids who got colds (in that order) in the Vitamin C group was significantly less than in the Placebo group what would the hypotheses be?

Write the hypothesis using words and symbols

- F. List the conditions you must check to use the test, then explain how they are or are not satisfied or assumptions you will have to make.

- G. Fill in the Compute step result from StatCrunch:

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value

- H. What is the conclusion of the hypothesis test? Include whether you can conclude a cause-and-effect relationship in your answer.

- I. Fill in the 95% confidence interval from StatCrunch:

95% confidence interval results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	L. Limit	U. Limit

- J. Interpret the confidence interval in the context of the problem. Be very specific about what it tells you about kids, colds, and placebo vs. Vitamin C.

2. (25 pts) **Quality Control.** A package of Diamond roasted almonds is supposed to contain 8 ounces. The Quality Control engineer, Janine, wants to make sure the machines are filling the packages correctly, i.e., she wants to test whether the mean weight is different from 8 ounces. She sets the significance level to .05, then she draws a random sample of 25 packages and finds that the sample has a mean of 7.8 ounces with a standard deviation of 0.7 ounces.

A. What hypothesis test should she use? (Circle your answer)

z-Test for Two Proportions

One Sample t-Test

Two Sample t-Test

Paired Difference Test

B. What are the hypotheses? Write them using math symbols then describe them, in words.

C. Is this a one-tailed or two-tailed test? One-tailed Two-tailed

D. Why is the two-tailed test considered more strict (harder to get a significant result)? (Circle your answer)

- (a) Because the P-value for the two-tailed test is half the P-value for the one-tailed test.
- (b) Because the P-value for the two-tailed test is two times the P-value for the one-tailed test.
- (c) Because the P-value doesn't matter for a two-tailed test.
- (d) Because two-tails are twice as much fun!

E. List the conditions she must check to use the test, then explain how they are or are not satisfied or assumptions she will have to make.

F. Fill in the Compute step result from StatCrunch:

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value

G. Why is the t-statistic you found in the previous answer a negative value?

- (a) Because she found a significant result.
- (b) Because she did not find a significant result.
- (c) Because the sample mean is less than the null.
- (d) Because she made a mistake.

H. Interpret the results from the hypothesis test.

I. Fill in the confidence interval calculation from StatCrunch:

95% confidence interval results:

Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit

J. Interpret the confidence interval in the context of the problem.

K. Does the confidence interval support the null or the alternative hypothesis? Explain how you can tell.

3. (25 pts) **Marketing.** Big Foods Grocery has two grocery stores located in Johnston City. One store is located on First Street and the other on Main Street and each is run by a different manager. Each manager claims that her store's layout maximizes the amounts customers will purchase on impulse. Both managers surveyed a sample of 20 of their customers and asked them how much more they spent than they planned to (so impulse buys).

The average impulse spending at First Street was \$15.92, with a standard deviation of \$4.15. The average impulse spending at Main Street was \$24.75 with a standard deviation of \$6.43.

A. What is the research question for this situation?

What is the independent variable? _____

What is the dependent variable? _____

Is this an observational study or controlled experiment? Observational Experiment

Is this design balanced or unbalanced? Balanced Unbalanced

B. Which of the hypothesis tests would you use to determine whether the average amount of impulse spending at First Street was significantly different from the average amount of spending at Main Street?

z-Test for Two Proportions

One Sample t-Test

Two Sample t-Test

Paired Difference Test

C. What are the hypotheses? Write them using math symbols then describe them, in words.

D. List the conditions you must check to use the test, then explain how they are or are not satisfied or assumptions you will have to make.

E. Fill in the Compute step result from StatCrunch:

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value

F. Interpret the results from the hypothesis test.

G. Fill in the confidence interval calculation from StatCrunch:

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit

H. Interpret the confidence interval in the context of the problem.

I. Does the confidence interval support the null or the alternative hypothesis? Explain how you can tell.

J. If there was a significant difference between impulse spending, on average, between the two stores, could you conclude that the store layout CAUSED the extra impulse buying at one store? Briefly explain.

K. Describe one possible confounder for this study.

4. (25 pts) **Drug Testing.** A study is designed to evaluate a new drug that may lower cholesterol. Fifty patients agree to participate in the study and each is asked to take the new drug for 6 weeks. Before starting the treatment, each patient's total cholesterol level is measured (BASELINE data). After taking the drug for 6 weeks, each patient's total cholesterol level is measured again (6 WEEKS data).

A. What is the independent variable? _____

What is the dependent variable? _____

B. What hypothesis test would you use for this study?

z-Test for Two Proportions

One Sample t-Test

Two Sample t-Test

Paired Difference Test

C. If you wanted to test whether the subjects' cholesterol levels, on average, were lower after 6 weeks on the drug, what should the hypotheses be? Write them in math symbols and in words.

- D. List the conditions you must check to use the test, then explain how they are or are not satisfied or assumptions you will have to make.

- E. The confidence interval is for the difference between Baseline and 6 Weeks is given below:

99% confidence interval results:

Difference	Mean	Std. Err.	DF	L. Limit	U. Limit
Baseline - 6 Weeks	16.933333	3.657	49	6.05	27.82

Interpret this confidence interval in the context of the problem. Be very specific about mean cholesterol level at Baseline vs 6 weeks later.

- F. Based on the confidence interval above, what would your conclusion be? (Circle the answer.)
- (a) There is no significant difference between the mean Baseline and After 6 Weeks cholesterol levels.
 - (b) There was a significant increase in cholesterol levels after 6 weeks on the drug.
 - (c) There was a significant decrease in cholesterol levels after 6 weeks on the drug.
 - (d) No conclusion can be made from this information.

- G. What is a major flaw in the design of this study? Circle all that apply.

- (a) There is no placebo group for comparison
- (b) The sample size is considered small.
- (c) There are potential confounders, like diet and exercise, that could have impacted the cholesterol levels.
- (d) None of the above.

- H. Based on the confidence interval and the design of the study, we can conclude which of the following.

- (a) The drug causes a reduction in cholesterol levels, on average.
- (b) The drug does not cause a reduction in cholesterol levels, on average.
- (c) There was an association between the drug and lower cholesterol but we can't say the drug caused it.
- (d) No conclusion can be made.

Extra Credit: (25 points)

Sound and Learning. Maddie is doing a study where she predicts that students will learn most effectively with a constant background sound, as opposed to an unpredictable sound or no sound at all.

She randomly selects 30 students for the study then randomly assigns them into three groups of 10.

All students study a passage of text for 30 minutes. Those in group 1 study with background sound at a constant volume in the background. Those in group 2 study with noise that changes volume periodically. Those in group 3 study with no sound at all. After studying, all students take a 10 point multiple choice test over the material. Their scores follow:

<u>group</u>	<u>test scores</u>									
1) constant sound	7	4	6	8	6	6	2	9	8	10
2) random sound	5	5	3	4	4	7	2	2	5	4
3) no sound	2	4	7	1	2	1	5	5	6	5

Here are the Summary Statistics for this data:

Column	n	Mean	Std. Dev.	Std. Error
Constant	10	6.6	2.3664319	0.74833148
Random	10	4.1	1.5238839	0.48189441
No Sound	10	3.8	2.1499354	0.67986927

A. What is the research question for this situation?

B. What is the independent variable? _____

What is the dependent variable? _____

Is this an observational study or controlled experiment? _____

C. Based on the sample means, it appears that which way of studying might be the best? _____

Since this is just sample data, we can't say these results would apply to everyone because the differences we see

between the groups may be due to just _____.

D. Which of the hypothesis tests would you use for this situation? _____

What are the hypotheses? Write them in symbols, then state what they mean in the context of the problem.

E. List the conditions she must check to use the test, then explain how they are or are not satisfied or assumptions she will have to make.

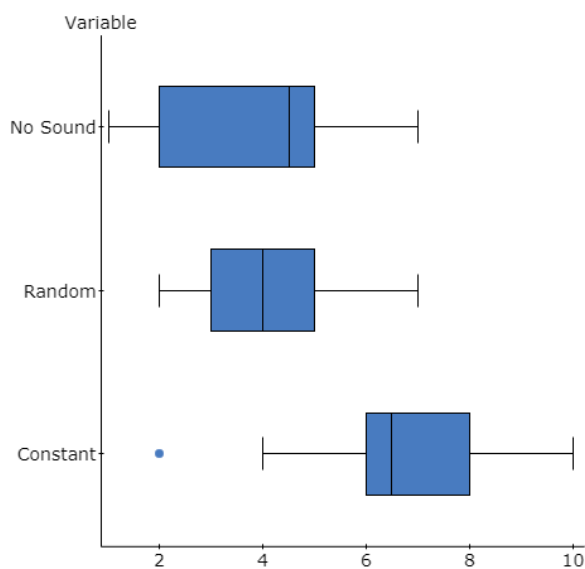
F. To check the condition of about population variances, Maddie examined the standard deviations (SD's) of the three groups (given at beginning of problem).

Which conclusion should she make? (Circle the answer.)

- (a) The SD's are unequal so the population variances must also be unequal.
- (b) The SD's are equal so the population variances must also be equal.
- (c) One of the SD's is more than two times another, so she shouldn't assume equal population variances.
- (d) None of the SD's is more than two times another, so she can assume equal population variances.
- (e) She can't tell anything about the population from the sample data.

G. To check the condition of normality, Maddie graphed boxplots of the sample data.

Here is the graph:



What can Maddie tell from the boxplots? (Circle the answer.)

- (a) The boxplots show the populations are perfectly symmetric, with no skewing, so she can proceed.
- (b) The boxplots show the sample data is somewhat skewed, with an outlier, but she will proceed with some reservations.
- (c) The boxplots show the populations are highly skewed, so she shouldn't proceed.
- (d) The boxplots show the sample data is perfectly symmetric, with no skewing or outliers, so she can proceed.
- (e) The boxplots can't tell her anything about the population distributions.

H. Use the given F-Table to answer the following questions

Source	DF	SS	MS	F-Stat	P-value
Columns	2	47.266667	23.633333	5.6519043	0.0089
Error	27	112.9	4.1814815		
Total	29	160.16667			

- (a) What number tells you the mean variance BETWEEN the groups? _____
- (b) What number tells you the mean variance WITHIN the groups? _____
- (c) How is the F-Stat found, using the numbers in the table? (Show the calculation and verify it matches the F-Stat in the table.)

Calculation of F-stat:

I. In general (NOT about this specific study) if the F-stat in an ANOVA test is quite small, what can you tell? (Circle the best answer.)

- (a) The signal (variance between groups) is weak relative to the noise (variance within the groups)
- (b) The P-value will be large.
- (c) The test won't show a significant difference between the groups.
- (d) All of the above.
- (e) None of the above.

J. Determine the conclusion of the hypothesis test for the Sound Study. The F-table is recopied here for convenience.

Source	DF	SS	MS	F-Stat	P-value
Columns	2	47.266667	23.633333	5.6519043	0.0089
Error	27	112.9	4.1814815		
Total	29	160.16667			

Conclusion (interpret results of hypothesis test):

Can Maddie conclude that sound during studying CAUSES a difference in how well students will do on a test, on average? Briefly explain.