Statistics 201 Exam 3 – Practice Exam
Fall 2024
Confidence Intervals – Chapter 19

Disclaimer:

This practice exam is provided solely for the purpose of familiarizing you with the format and style of the Stat 201 exams. There is no explicit or implicit guarantee that the upcoming exam will ask similar questions. If you use the practice exam as your only tool to help you prepare for the upcoming exam, you most likely will not do well on the exam. You should still do the things you would have done if you did not have access to this practice exam, such as re-read the text, go over your class notes, re-work the online homework problems, and look at the list of exam topics provided and make sure that you understand all the concepts listed within it.

NOTE: These questions come from actual older exams. As such, the points on this practice exam may total more than 100 points.

HASLAM
COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

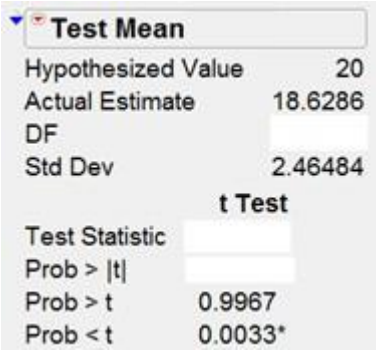
Please note: when asked to write out null and alternative hypotheses, if appropriate, be sure to use proper mathematical notation (i.e., use symbols such as $<, =, \leq, \geq, \neq, >, p, \hat{p}, \mu, \bar{x}$, etc.).

Exam Grade: _____

ON THIS EXAM, WHEN ROUNDING A FINAL NUMERICAL ANSWER , PLEASE REPORT AT LEAST 3 SIGNIFICANT DIGITS.

EXAMPLE: the value 0.000378516 should be rounded to 0.000379

1. Below is output from a hypothesis test for a single population mean based on 28 observations. The null hypothesized value, as well as the sample mean ("actual estimate") are seen below (among other numbers). Calculate the 3 missing values below. Show your work (where necessary), but fill in your final answers in the blanklines below.



Test Mean	
Hypothesized Value	20
Actual Estimate	18.6286
DF	
Std Dev	2.46484
t Test	
Test Statistic	
Prob > |t|	
Prob > t	0.9967
Prob < t	0.0033*

i) (2 points) DF : _____

ii) (2 points) Test Statistic : _____

iii) (2 points) Prob > $|tt|$: _____

2. It is generally believed that 75% of college sophomores regularly recycle. Are UT sophomores different from this widely accepted figure? A researcher at UT got the cooperation of the registrar's office and was given the contact information for a large number of randomly selected sophomores at UT (out of the approximately 5,000 UT sophomores). The researcher got responses to a short survey from 128 of these students. Based on the answers they provided, the researcher determined that 84 of the 128 students regularly recycled.

   i) (2 points) Suppose the researcher wishes to carry out a formal hypothesis test for their research question. Write out the null and alternative hypotheses they wish to test.

   ii) (6 points) Are the 3 conditions for performing a hypothesis test for a proportion met in this case? State each condition, indicate whether you think each condition is met, and explain your reasoning. Provide numerical justification whereappropriate.

   Condition 1 is: Randomization

   Is condition 1 met? (Circle one): Yes No
   Briefly explain your answer:

   Condition 2 is: 10% Condition

   Is condition 2 met? (Circle one): Yes No
   Briefly explain your answer:

   Condition 3 is:

   Is condition 3 met? (Circle one): Yes No
   Briefly explain your answer:

Question 2 (continued)

iii) (3 points) Calculate the test statistic for the hypothesis test.

iv) (2 points) Using an alpha level of 0.07, with the resulting p-value of 0.0143 what is your conclusion regarding your null hypothesis in part (i)? Circle the best answer:

Reject the null hypothesis Fail to reject the null hypothesis

v) (4 points) Write out the *full* conclusion of the hypothesis test. Be sure tostate your conclusion in the context of this problem.

vi) (2 points) What would be the next step for the research after the finding the results hypothesis test?

3. A researcher believes that British men are shorter, on average, than men from the USA. A random sample of 35 men from the USA (ages 20-50) was observed for height. The mean height was 68.1 inches and the sample standard deviation was 5.34 inches. Meanwhile, a random sample of 35 men from Britain (ages 20-50) was observed for height. The mean height was 64.6 and the sample standard deviation was 5.81 inches.

   i) (2 points) Which type of problem is this? Circle the best answer:

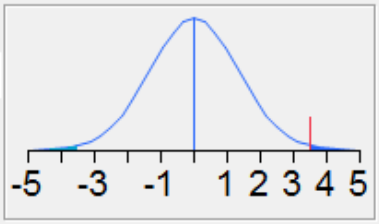
      a) One sample proportion test
      b) One sample test of a mean
      c) Two independent samples t-test
      d) Chi-Squared test of independence

   ii) (2 points) Write down the null and alternative hypotheses suggested by the first sentence of this problem (HINT: to avoid confusion later, look at the top line of the JMP output below now).

   iii) (3 points) Below is some output from JMP for a proper analysis of these data. Calculate the missing value (to the right of the word "t Ratio").

USA-Britain
Assuming unequal variances

Difference	3.49714	t Ratio	
Std Err Dif	1.33411	DF	67.51277
Upper CL Dif	6.15967	Prob > |t|	0.0108*
Lower CL Dif	0.83461	Prob > t	0.0054*
Confidence	0.95	Prob < t	0.9946



   iv) (1 point) There are 3 p-values displayed in the output. Circle the p-value that corresponds to your alternative hypothesis in part [ii] above.

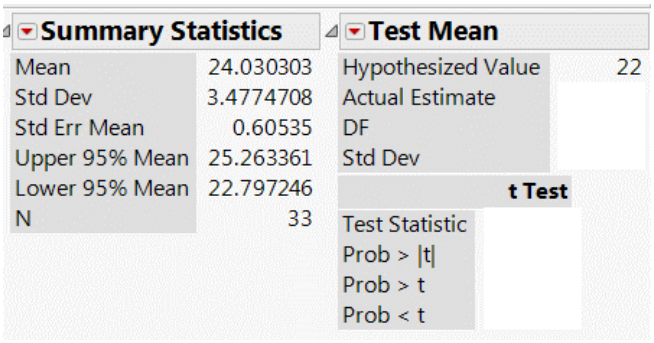
Question 3 (continued)

v) (1 point) If we use an α level of 0.01, what is your conclusion regarding your null hypothesis in part (ii)? Circle the best answer:

Reject the null hypothesis Fail to reject the null hypothesis

vi) (3 points) State the remainder of your conclusion for this hypothesis test. Be sure to state your conclusion in the context of this problem.

4. (2 points) The following is some output from JMP for a one-sample test of a population mean. The 95% confidence interval for the mean is showing, but much of the other output has been erased. The researcher wanted to do a two-sided hypothesis test, and was testing to see if the population mean is 22 (as indicated in the output).



Summary Statistics	
Mean	24.030303
Std Dev	3.4774708
Std Err Mean	0.60535
Upper 95% Mean	25.263361
Lower 95% Mean	22.797246
N	33

Test Mean	
Hypothesized Value	22
Actual Estimate	
DF	
Std Dev	

t Test
Test Statistic
Prob > |t|
Prob > t
Prob < t

Based on the output that is still showing, what is a possible value for the p-value next to "Prob > |t|"? Circle the best answer:

a) 1.0
b) 0.9999
c) 0.5015
d) 0.0501
e) 0.0051
f) The output that is still showing offers no clue as to what that p-value might be.

5. The American Banking Association is concerned that the percentage of customers who default on their auto loans is too high. They are basing this belief on a 95% confidence interval of the population proportion constructed from a random sample of loans. The confidence interval was (0.1631, 0.2329).

   i) (2 points) What is the margin of error?

   ii) (2 points) What is the sample proportion?

   iii) (4 points) Interpret this interval in the context of the problem.

   iv)(2 points) Given that the critical value $Z^*=1.96$, what is the value of the standard error used in constructing the confidence interval?

6. Suppose that the Business Analytics and Statistics Department has come up with a method to determine whether or not a student cheated on their exam. We test these simple hypotheses as follows:

H_0: The student didn't cheat
H_A: The student cheated

i) (2 points) With any test there is always the risk of making the wrong decision. Describe what a type I error is in the context of this problem.

ii) (2 points) Describe what a type II error is in the context of this problem.

iii) (2 points) If we were to give a 0 to any student who we determined to have cheated using the Business Analytics and Statistics Department's detection method, which type of error would a student typically consider more serious? Briefly explain why.

iv)(2 points) What is the symbol used to identify the probability of making a type II error? (Circle One):

α $(1-\alpha)$ β $(1-\beta)$ χ^2

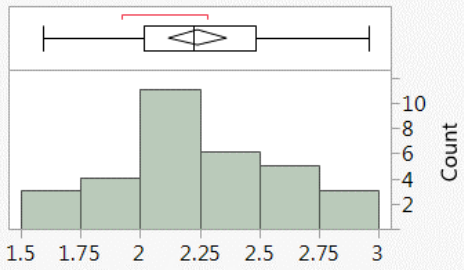
7. As patients arrive at a local hospital emergency room, as long as they are able to respond, and are not deemed to have an immediate life threatening condition, they are asked some standard questions before being asked to take a seat in the waiting room. The time it takes to complete this initial set of questions has come under scrutiny lately. The emergency room manager believes that the average amount of time it takes to get answers to these initial questions is 2 minutes. Most emergency room employees believe it takes longer than that, on average.

   A nursing school intern was asked to randomly select incoming patients over a 2 week period, and time how long it takes to get answers to these initial questions. The following are some of the results of the intern's data collection efforts.

   i) (2 points) Write out the null and alternative hypotheses suggested by the introduction to this problem.

Below is some output from the data that the intern collected:

Time in Minutes



Quantiles

100.0%	maximum	2.96
99.5%		2.96
97.5%		2.96
90.0%		2.782
75.0%	quartile	2.485
50.0%	median	2.225
25.0%	quartile	2.0125
10.0%		1.717
2.5%		1.59
0.5%		1.59
0.0%	minimum	1.59

Summary Statistics

Mean	2.2409375
Std Dev	0.3382437
Std Err Mean	0.0597936
Upper 95% Mean	2.3628873
Lower 95% Mean	2.1189877
N	32

Question 6 (continued)

ii) (3 points) Use the information in the output to calculate the value of the test statistic for this hypothesis test.

iii) (2 points) Based on your answer to part [ii] above, what would your decision be regarding the null hypothesis? Use $\alpha = 0.05$. Circle the best answer:

Reject the null hypothesis Fail to reject the null hypothesis

Briefly explain your reasoning for your selection above.
(NOTE: If you have NO ANSWER to part [ii], use (3.21) as your answer to part [ii], and use that to answer part [iii]).

iv) (3 points) State the remainder of your conclusion for this hypothesis test. Be sure to state your conclusion in the context of this problem.

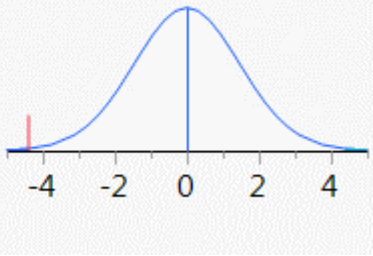
8. Below is some JMP output for a two-sample t-test. Answer the following questions based on the output provided.

t Test

Group2-Group1

Assuming unequal variances

Difference	-4.4200	t Ratio	
Std Err Dif	1.4713	DF	
Upper CL Dif	-1.4648	Prob > |t|	
Lower CL Dif	-7.3752	Prob > t	
Confidence	0.95	Prob < t	



i) (2 points) The values next to "Upper CL Dif" and "Lower CL Dif" represent the endpoints of a 95% confidence interval for the difference between the two population means. What is the "margin of error" of this confidence interval?

ii) (2 points) What was the value of tt^*_{dddd} that was used to construct this confidence interval? Please report 4 decimal places.

9. The Gallup Poll conducted a representative telephone survey during the first quarter of 2010 in the United States. Among the reported results was the following table concerning the preferred political party affiliation of respondents and their ages. We are interested in seeing if there is an association between a person's age and their political party affiliation in the United States.

Contingency Table

Age Group \ Political Party	Democrat	Independent	Rebublican	Total
18-29	131	209	51	391
30-49	140	200	79	419
50-64	141	145	102	388
65+	162	113	129	404
Total	574	667	361	1602

i) (2 points) Write the appropriate null and alternative hypothesis suggested by the statements above.

ii) (4 points) Below is some output from JMP for this analysis. Two of the values in the table have been erased (A and B). Provide the necessary arithmetic for these two missing values. Write your final answers in the blanks: A_____ B_____

Contingency Table

Age Group \ Political Party	Democrat	Independent	Rebublican	Total
18-29 Count	131	209	51	391
18-29 Expected	140.096	A	88.1092	
18-29 Cell Chi^2	0.5906	13.1143	15.6294	
30-49 Count	140	200	79	419
30-49 Expected	150.129	174.453	94.4189	
30-49 Cell Chi^2	0.6833	3.7413	2.5179	
50-64 Count	141	145	102	388
50-64 Expected	139.021	161.546	87.4332	
50-64 Cell Chi^2	0.0282	1.6946	2.4269	
65+ Count	162	113	129	404
65+ Expected	144.754	168.207	91.0387	
65+ Cell Chi^2	B	18.1195	15.8291	
Total	574	667	361	1602

Question 8 (continued)

iii) (3 points) As with any statistical technique, there are conditions that must be met for the analysis to be valid. One of the conditions of this technique (the Chi-Square test of independence) has something to do with there being enough data for the analysis to be valid. What is the name of this condition, and do you believe it is met in this case? Provide a brief explanation for your "yes" or "no" answer.

iv) (2 points) From the data given, calculate the degrees of freedom for the chi-square test of independence.

v) (2 points) Below is more output from JMP for this analysis. The Pearson ChiSquare value is 76.430. Making reference to the numbers in the Contingency Table at the bottom of the previous page, briefly explain how this number was calculated (use the space to the right of the output for your answer).

Tests

N	DF	-LogLike	RSquare (U)
1602		39.293192	0.0230

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	78.586	<.0001*
Pearson	76.430	<.0001*

vi) (1 point) Using α=0.005, what is your conclusion regarding your null hypothesis from part [i]? Circle the best answer:

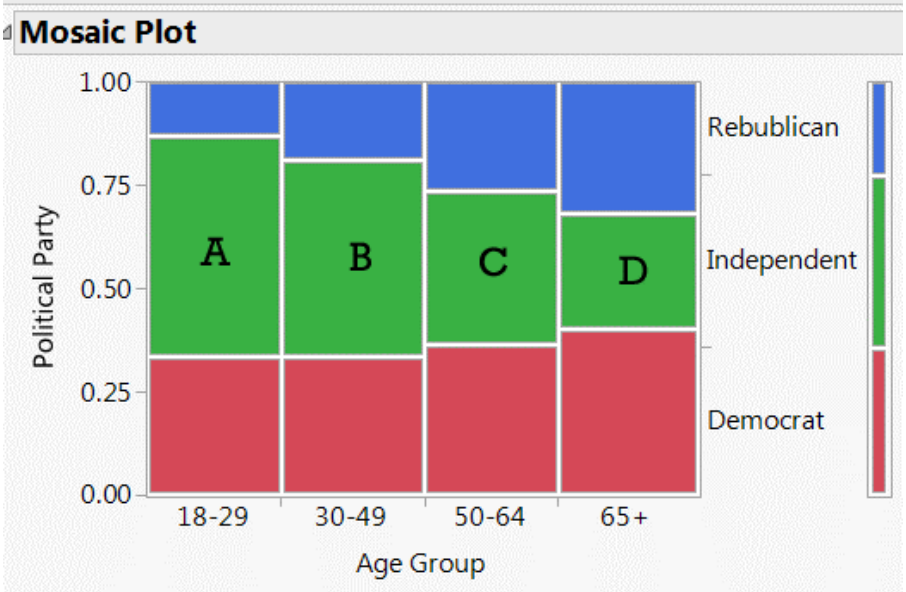
Reject the null hypothesis Fail to reject the null hypothesis

Briefly explain your reasoning for your selection above.

vii) (2 points) State the remainder of your conclusion for this hypothesis test. Be sure to state your conclusion in the context of this problem.

Question 8 (continued)

viii) (4 points) The following is the Mosaic Plot for this analysis, along with the default Contingency Table from JMP. Use the output below to help you report the percentage (to 2 decimal places) of respondents in each of the four age groups that reported they were "independents". Fill in the four blanks below with your answers. (HINT: Your answers represent the height of the bars identified as bars A, B, C and D.)



Contingency Table

Political Party

Age Group	Count / Total % / Col % / Row %	Democrat	Independent	Rebublican	Total
18-29		131	209	51	391
		8.18	13.05	3.18	24.41
		22.82	31.33	14.13	
		33.50	53.45	13.04	
30-49		140	200	79	419
		8.74	12.48	4.93	26.15
		24.39	29.99	21.88	
		33.41	47.73	18.85	
50-64		141	145	102	388
		8.80	9.05	6.37	24.22
		24.56	21.74	28.25	
		36.34	37.37	26.29	
65+		162	113	129	404
		10.11	7.05	8.05	25.22
		28.22	16.94	35.73	
		40.10	27.97	31.93	
Total		574	667	361	1602
		35.83	41.64	22.53	

18-29:_____

30-49:_____

50-64:_____

65+:_____

Question 8 (continued)

ix) (3 points) In one or two sentences, how would you describe the nature of the association that seems to exist between these two variables?

10. (3 points) Does studying really pay off? Is there a relationship between the total number of minutes students spent preparing for all 3 exams in Stat 201 and the average score they achieved on these exams? What tool or technique would you use to address this question? Circle the best answer:

    a) Box plot
    b) Hypothesis test for a population mean
    c) Decision tree
    d) Confidence interval for the difference between two population means
    e) Simple linear regression

11. (3 points) You are the regional manager of 5 large grocery stores. You let the manager of each of these 5 stores make their own decisions regarding who to hire. Since these 5 stores are within 50 miles of each other, you would expect the ethnic background of the employees at these 5 stores to be fairly similar. Is that the case? What tool or technique would you use to address this question? Circle the best answer:

    a) Histogram
    b) Chi-Square test of independence
    c) Simple linear regression
    d) Hypothesis test for a population mean
    e) Pearson's correlation coefficient (r)

True/False Questions
Circle the best answer
(2 points each)

T F "Power" represents the probability that we correctly fail to reject a true null hypothesis.

T F If, based on the same set of data, you change your level of confidence from 90% to 95%, the resulting 95% confidence interval is considered less precise than the 90% confidence interval.

T F A p-value represents the probability of seeing the results you saw, or results more unusual than that, assuming the null hypothesis is true.

T F The only way to decrease the probability of a Type II error is to increase the probability of a Type I error.

[YOU MAY REMOVE THIS SHEET]

ONE SAMPLE TESTS

Proportion

Confidence Intervals	$\hat{p} \pm z^* \times SE(\hat{p})$	$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$
Hypothesis Testing	$z_{\hat{p}} = \frac{\hat{p} - p_0}{SD(\hat{p})}$	$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}$

Mean

Confidence Intervals	$\bar{y} \pm t^*_{df} \times SE(\bar{y})$	$SE(\bar{y}) = \frac{s}{\sqrt{n}}$
Hypothesis Testing	$t_{df} = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$	$SE(\bar{y}) = \frac{s}{\sqrt{n}}$

(*df*) degrees of freedom for the t-distribution = $n - 1$

TWO SAMPLE TESTS

Difference Between Two Means

Confidence Intervals	$\bar{y}_1 - \bar{y}_2 \pm t^*_{df} \times SE(\bar{y}_1 - \bar{y}_2)$	$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$
Hypothesis Testing	$t_{df} = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE(\bar{y}_1 - \bar{y}_2)}$	$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$

(*df*) degrees of freedom for the t-distribution available from JMP output

Test of Independence

$\chi^2_{df} = Sum\ of\ \frac{(Obs-Exp)^2}{Exp}$ where $Exp = \frac{Row\ Total \times Column\ Total}{N}$ for each cell.

$Degrees\ of\ freedom = (\#Rows - 1) \times (\#Columns - 1)$