QUANTITATIVE METHODS 830:200:H2

SUMMER 2020

Final
8/12/2020

Name

_____

Show your work – all questions are one point except question 34 (2 points)

Rutgers Recreation states that the general student participation in intramural sports activities was 42%. If the student participation increases in the freshman class next academic year, Rutgers Recreation wants to enhance the gym facilities. During freshman orientation, a pollster arranged for a random sample of students to be surveyed. The pollster reported that 202 of 454 students in the random sample were thinking about joining an intramural team. Is this strong evidence that the participation rate may *increase*?

1.Assuming that the conditions and assumptions are met, what's the null hypothesis in words?
A. Students' participation to intramural activities will increase.
B. Students' participation to intramural activities will decrease
C. Students' participation to intramural activities will be the same as before
D. Students' participation to intramural activities will be zero

2.What's the alternative hypothesis in words?
A. Students' participation to intramural activities will increase
B. Students' participation to intramural activities will decrease
C. Students' participation to intramural activities will be the same as before
D. Students' participation to intramural activities will be zero

3.What are the null and the alternative hypothesis in mathematical terms?
A. $H_0 : p \neq 0.42$ - $H_A : p > 0.42$
B. $H_0 : p > 0.42$ - $H_A : p < 0.42$
C. $H_0 : p = 0.42$ - $H_A : p \neq 0.42$
D. $H_0 : p = 0.42$ - $H_A : p > 0.42$

4.Which kind of test do I need to carry out?
A. a one sample t-test
B. a two samples t-test
C. a two proportions z-test
D. a one proportion z-test

5.What's the value of z?
A. 1.53 B. 1.09 C. 0.58 D. 0.445

6.Is the test going to be one-tailed upper tail, one-tailed lower tail or two-tailed?
A. one-tailed upper tail
B. one-tailed lower tail
C. two-tailed
D. impossible to determine

7. You know that the *P*-value indicates a probability. In this case, how do you express the *P*-value in mathematical terms? (X stays for the value of z found in question #5 – substitute it in the formula and draw a picture of the distribution)
A. *P*-value = $P(z = X)$
B. *P*-value = $P(z < X)$
C. *P*-value = $P(z \neq X)$
D. *P*-value = $P(-X < z < X)$

8.How much is the *P*-value?
A. 0.138 B. 0.043 C. 0.978 D.0.002

9.What are your conclusions?
A. There is enough evidence to suggest that the students' participation to intramural activities will decrease.
B. There is not enough evidence to suggest that the students' participation to intramural activities will stay the same.
C. There is not enough evidence to suggest that the students' participation to intramural activities will increase.
D. There is enough evidence to suggest that the students' participation to intramural activities will increase.

10.What would be an example of Type I error in this case?
A. Saying the proportion will stay the same.
B. Saying the proportion will decrease.
C. Saying the proportion will increase.
D. Saying the proportion will be indeterminate.

Peloton offers online spin classes. These classes are always accompanied by songs, which have different features, such as song type, length and beats per minute. A Peloton instructor wondered if the songs in the high intensity classes have a number of beats per minute that is higher that the beats per minute of the songs for low intensity workouts. A random sample of songs from each group was selected; the beats per minute are listed in the chart at the right. Do these samples indicate that the number of beats per minute in the songs accompanying high intensity workouts is higher that those of the songs used in low impact classes?

Let: mean of low impact (LI) classes' song = μ_1,
mean of high intensity (HI) classes' songs = μ_2,
and assume the assumptions and conditions are all met.

Beats per Minute			
LI Songs		HI Songs	
121	119	122	120
122	121	121	118
117	122	121	121
120	119	122	123
120	119	121	118
121	118	119	120
118	120	120	124
120	123	119	
117	118		

11.What are my null and alternative hypotheses?

A. $H_0 : \mu_2 - \mu_1 = 0$ $H_A : \mu_2 - \mu_1 > 0$

B. $H_0 : \mu_2 - \mu_1 > 0$ $H_A : \mu_2 - \mu_1 < 0$

C. $H_0 : \mu_2 - \mu_1 = 0$ $H_A : \mu_2 - \mu_1 \neq 0$

D. $H_0 : \mu_2 - \mu_1 > 0$ $H_A : \mu_2 - \mu_1 = 0$

12.Which kind of test will answer the question?

A. A two samples t-test
B. A one sample t-test
C. A paired t-test
D. A two proportions z-test

Here are the summary statistics for the two samples:
$n_1 = 18$
$n_2 = 15$
$\bar{y}_1 = 119.7$ (mean beats per minute of LI songs)
$\bar{y}_2 = 120.6$ (mean beats per minute of HI songs)
$s_1 = 1.74$
$s_2 = 1.72$
$df = 30$ (from computer)

13. What's the observed difference? (use $\bar{y}_2 - \bar{y}_1$)
A. 0.023 B. 0.9 C. 0.18 D. 0.09

14. What's the standard error of the sampling distribution of the differences between the means?
A. 0.236 B. 0.468 C. 0.375 D. 0.605

15. What's the value of *t*? (use $\bar{y}_2 - \bar{y}_1$)
A. 3.374 B. 0.247 C. 0.05 D. 1.449

16. From a computer I obtain the degrees of freedom for the value of *t* for this study: 30
What's the meaning of the *P*-value?
A. The probability that a value as extreme or more extreme than the *t* value I found belongs to a *Student's t model distribution* with 30 degrees of freedom.
B. The probability that a value as extreme or more extreme than the *t* value I found does not belong to a *Student's t model distribution* with 30 degrees of freedom.
C. The probability that a value as extreme or less extreme than the *t* value I found belongs to a *Student's t model distribution* with 30 degrees of freedom.
D. The probability that a value as extreme or less extreme than the *t* value I found does not belong to a *Student's t model distribution* with 30 degrees of freedom.

17. You know that the *P*-value indicates a probability. In this case, how do you express the *P*-value in mathematical terms? (X stays for the value of *t* found in question #15 – substitute it in the formula and draw a picture of the distribution)
A. *P*-value = $P(t < X)$
B. *P*-value = $P(t > X)$
C. *P*-value = $P(t = X)$
D. *P*-value = $P(t \neq X)$

18.From the t-table, what's the approximate value of the *P*-value?
A. *P*-value > 0.10
B. 0.05 < *P*-value < 0.10
C. *P*-value < 0.05
D. 0.01 < *P*-value < 0.05

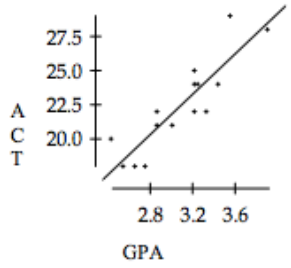
19.What are your conclusions?
A. LI's beats per minute are significantly lower
B. LI's beats per minute are significantly higher
C. There is little evidence of a difference in rates
D. There is a strong evidence of a difference in rates

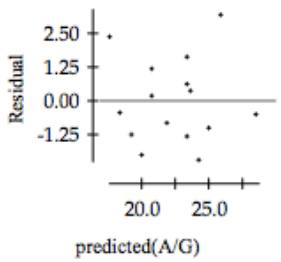
20. I want to create a 90% confidence interval. What's the *critical value t** from the t-table?
A. 1.96 B. 2 C. 1.697 D. 1.310

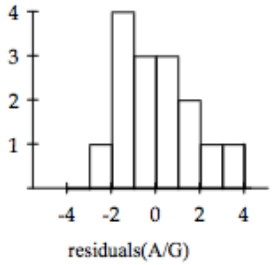
21.What's the interval?
A. (-0.13, 1.93)
B. (-0.43, 1.56)
C. (0.53, 2.94)
D. (2.13, 3.93)

A high school counselor was interested in finding out how well student grade point averages (GPA) predict ACT scores. A sample of the senior class data was reviewed to obtain GPA and ACT scores. The data are shown in the table to the right.

GPA	ACT
3.25	24
2.87	21
2.66	18
3.33	22
2.87	22
3.21	22
2.76	18
3.91	28
3.55	29
2.55	18
2.44	20
3.22	24
3.01	21
3.44	24
3.22	25







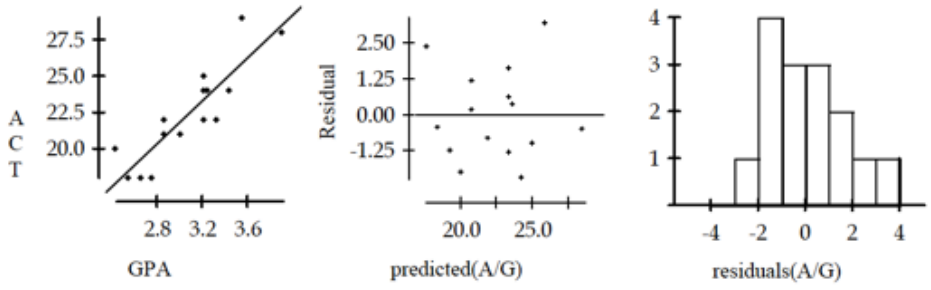
Dependent variable is: **ACT**
No Selector
R squared = 78.1% R squared (adjusted) = 76.4%
s = 1.630 with 15 - 2 = 13 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	123.041	1	123.041	46.3
Residual	34.5589	13	2.65838	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-0.427035	3.382	-0.126	0.9014
GPA	7.39697	1.087	6.80	≤ 0.0001

A Rutgers admission dean was interested in understanding whether there is an association between high school GPA and ACT scores. A sample of data from the New Brunswick High School senior class is presented in the table below.

GPA	ACT
3.25	24
2.87	21
2.66	18
3.33	22
2.87	22
3.21	22
2.76	18
3.91	28
3.55	29
2.55	18
2.44	20
3.22	24
3.01	21
3.44	24
3.22	25



Dependent variable is: **ACT**
No Selector
R squared = 78.1% R squared (adjusted) = 76.4%
s = 1.630 with 15 - 2 = 13 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	123.041	1	123.041	46.3
Residual	34.5589	13	2.65838	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-0.427035	3.382	-0.126	0.9014
GPA	7.39697	1.087	6.80	≤ 0.0001

Is there *evidence of an association* between GPA and ACT score? Test an appropriate hypothesis and state your conclusion in the proper context.

22. Assuming that the assumptions and conditions are all met, what's the null hypothesis?
A. There is some association between GPA and ACT scores
B. There is a negative association between GPA and ACT scores
C. There is a positive association between GPA and ACT scores
D. There is no association between GPA and ACT scores

23.What's the alternative hypothesis?
A. There is some association between GPA and ACT scores
B. There is a negative association between GPA and ACT scores
C. There is a positive association between GPA and ACT scores
D. There is no association between GPA and ACT scores

24.What's the regression equation?
A. *GPA* = -0.427 + 7.397(*ACT score*)
B. *ACT score* = -0.427 + 7.397(*GPA*)
C. *ACT score* = + 7.397 -0.427 (*GPA*)
D. *GPA* = + 7.397 -0.427 (*ACT score*)

25.What's the correlation coefficient?
A. 0.884 B. 0.756 C. 0.235 D. 0.967

26.What's the *P*-value corresponding to the null hypothesis?
A. 0.9014 B. 0.05 C. > 0.0004 D. < 0.0001

27.What are your conclusions?
A. There is strong evidence of an association between GPA and ACT scores
B. There is no evidence of an association between GPA and ACT score.
C. There is weak evidence of an association between GPA and ACT score.
D. There is absolute evidence of an association between GPA and ACT score.

28.How many degrees of freedom do I have?
A. 13 B. 15 C. 2 D. 17

29.What's a 95% confidence interval for the slope? [Hint: you need to get the critical value t^* from the t-table, and the standard error SE from the regression output]
A. (6.05, 10.34)
B. (3.67, 6.54)
C. (2.48, 7.43)
D. (5.05, 9.74)

30.What's the meaning of the confidence interval?
A. We are 95% confident that the ACT score decreases an average of between 5.05 and 9.74 points for every additional GPA point.
B. We are 95% confident that the ACT score increases an average of between 5.05 and 9.74 points for every additional GPA point.
C. We are 95% confident that the GPA score increases an average of between 5.05 and 9.74 points for every additional ACT point.
D. We are 90% confident that the ACT score decreases an average of between 5.05 and 9.74 points for every additional GPA point.

Medical researchers enlisted 90 subjects for an experiment comparing treatments for skin rashes. The participants were divided into three groups at random and given different types of topical creams to use over a period of three months. Unknown to them, one group received a placebo treatment, another group a cream containing the coconut oil, and the last group a cream containing hydrocortisone. After six months, physicians (who did not know which treatment each person had received) evaluated the subjects to see if their eczema had returned.
Is the rate of recurrence the same for all the groups?
The following table shows the results of the study:

	TREATMENT			
DIAGNOSIS	Placebo	Coconut oil	Hydrocortisone	Total
Eczema returned	24	22	14	60
No sign of eczema	6	8	16	30
Total	30	30	30	90

31.What's the null hypothesis?
A. The rate of recurrence is smaller for one of the three treatments
B. The rate of recurrence is greater for one of the three treatments
C. The rate of recurrence is the same for all three treatments
D. The rate of recurrence differs for some treatments

32.What's the alternative hypothesis?
A. The rate of recurrence is smaller for one of the three treatments
B. The rate of recurrence is greater for one of the three treatments
C. The rate of recurrence is the same for all three treatments
D. The rate of recurrence differs for some treatments

33.Which kind of test do you need to perform?
A. A chi-square goodness-of-fit test
B. A chi-square test of independence
C. A chi-square test of homogeneity
D. Impossible to determine at this stage

34.What's the value of the chi-square (χ^2) statistics? 2 POINTS
[Hint: first find the expected values]
A. 3.2 B. 1.8 C. 8.4 D. 9.5

35. You know that the *P*-value indicates a probability. In this case, how do you express the *P*-value in mathematical terms? (X is the value of the χ^2 statistics we found in the previous question – substitute it and draw a picture of the distribution)
A. *P*-value = $P(\chi^2 < X)$
B. *P*-value = $P(\chi^2 = X)$
C. *P*-value = $P(\chi^2 > X)$
D. *P*-value = $P(\chi^2 \neq X)$

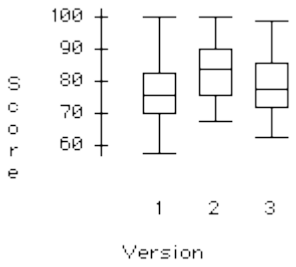
36. How many degrees of freedom do we have?
A. 3 B. 6 C. 2 D. 5

37.What's the approximate value of the *P*-value from the chi-square table ? (look at the degrees of freedom on the column, see where your statistics fall, and then read the corresponding value at the top)
A. $0.05 < P\text{-value} < 0.10$
B. $0.025 < P\text{-value} < 0.05$
C. $0.005 < P\text{-value} < 0.01$
D. $0.01 < P\text{-value} < 0.025$

38.What are your conclusions?
A. There is strong evidence that the tested treatments are all equally effective in preventing the recurrence of skin rashes.
B. There is strong evidence that the tested treatments are not all equally effective in preventing the recurrence of skin rashes.
C. There is weak evidence that the tested treatments are all equally effective in preventing the recurrence of skin rashes.
D. There is absolute evidence that the tested treatments are not all equally effective in preventing the recurrence of skin rashes.

To discourage students from collaborating during an exam, a Rutgers instructor makes three different versions of a test, printing 35 copies of each version, to have enough copies for the 105 students in her class. The 105 exams are randomly scrambled, and each student receives one copy. After the exam, the instructor is worried that one version might have been easier than the others. He uses a one-way ANOVA to test whether the average score was different for the three versions. See below the ANOVA table and a boxplot of the results. Let μ_1, μ_2, μ_3 be the mean scores of all the possible students on tests 1, 2, 3.

Source	df	Sums of Squares	Mean Square	F-ratio	P-value
Version	2	771.943	385.971	4.4317	0.0143
Error	102	8883.49	87.093		
Total	104	9655.43			



39. What hypotheses are tested by this ANOVA?

A. H_0: $\mu_1 < \mu_2 < \mu_3$ H_A: at least one version has a different mean score than the others
B. H_0: $\mu_1 = \mu_2 = \mu_3$ H_A: all three versions have a distinct mean score
C. H_0: $\mu_1 < \mu_2 < \mu_3$ H_A: all three versions have a distinct mean score
D. H_0: $\mu_1 = \mu_2 = \mu_3$ H_A: at least one version has a different mean score than the others

40. How was the *F*-ratio found?

A. MS_T / 87.093
B. MS_E / 87.093
C. MS_E / MS_T
D. 87.093 / MS_E

41. What are your conclusions?

A. At least one version of the exam has a different mean score than the others
B. I fail to reject the null hypothesis
C. The *P*-value is higher than the typical significance level of 5%
D. The three versions have all distinct mean scores

---

42. The mean square error, MSE…
A. is always a good estimate of σ^2
B. is a good estimate of σ^2 only if the samples have the same mean
C. is a good estimate of σ^2 only if the samples have different means
D. is never a good estimate

43. In ANOVA, an interaction between two factors means that…
A. both factors are related to the response variable.
B. the variances of the two factors are unequal.
C. the effect of one factor depends on the level of the other factor.
D. none of the above are correct.

44. If I increase the sample size used in my study:
A. The standard deviation will decrease
B. The sample statistics will get closer to the population parameter
C. The probability of Type I and Type II error will decrease
D. All of the above