BUS 310, Exam #2E, Part III, Spring 2019

Name KEY

Section

Instructions: This exam is in three parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I and you will answer questions about that work in class below; Part II is to be completed entirely in class using your computer. Part III is to be done entirely in class without your computer.

1. You may not use cell phones, and you may only access internet resources you are specifically directed to use: You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II, but not for Part III.
2. It is a violation of the honor code to communicate with other students in or out of the class during the exam, by any means. Students whose exams show evidence of coordination will be reported.
3. Show all work to support your reasoning. Primarily, this can be done in Excel. Deletion of evidence of your logical process can result in loss of credit. A significant amount of credit goes toward process, reasoning and interpretation.
4. When rounding, do not over-round. In general, do not report dollar amounts beyond the penny. Means should be rounded to one digit more than the original data; standard deviations to two digits more. Do not report fractions rounded to single digit expressions: $\frac{131}{256} \neq \frac{1}{2}$, and do not round decimals or percents to a single digit: $0.57846 \ldots \neq 60\%$ *or* 0.6. Report a minimum of two digits, up to four, unless otherwise specified in the problem.
5. If a problem asks for an explanation, state the solution clearly, then interpret or explain in addition to stating the solution, not in place of. Explanations without solutions, just as solutions without explanations, will not be awarded full credit.

Part I: At Home

This part was completed at home. You can upload the Excel file for Part I to the Part I folder in Blackboard for use during the Exam period. However, this submission will **not** be graded in this location, it must be submitted to the "**to be graded** folder" to receive credit.

Part II: In Class (with computer)

Before completing Part III, complete Part II in class. Return the paper to your instructor and put away your computer. Then pick up Part III.

Part III: In Class (without computer)

1. You may use a handheld calculator for this portion of the exam. Any calculator is fine, as long as it is not on a device that connects to the Internet. That means, you may not use the calculator on your phone or smart watch. You may also not share calculators with another student taking the exam at the same time.
2. This is Exam E.
3. Answer the questions on the paper exam. Sign the honor code statement on the next page.
4. Turn in your paper copy of the exam to your instructor. Your instructor will attach this portion of the exam to the version of Part II that you submitted previously.

Honor Code Statement:

I, ______________________________________ (print your name), agree to abide by the George Mason Honor Code and Academic Integrity Pledge: *To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University Community and with the desire for greater academic and personal achievement, I, a student member of the university community, pledge not to cheat, plagiarize, steal, or lie in matters related to academic work.* Furthermore, I have read and I agree to follow the guidelines laid out in the instructions for this exam above. I also agree not to participate in the efforts of other students to circumvent these guidelines, or to assist in their violations of the code, and will report such efforts in a timely manner.

______________________________________ ____________________

Student Signature and G# Today's Date

Part III:

1. When a hypothesis test is conducted, there are four possible combinations of outcomes: The null can be true, the null can be false, our conclusion can agree with the true state of nature, or it may not. A table of these situations is shown below. Two of these combinations are correct and two produce errors. Label all four possibilities as correct, or, if an error, which kind of error it is. (8 points)

	Nature: H_0 True	Nature: H_0 False
Conclusion: H_0 True	Correct	Type II
Conclusion: H_0 False	Type I	Correct

Below you will find calculations for a χ^2-test of salary versus gender. Use this information to answer the questions that follow.

Count of Gender	Column Labels		
Row Labels	**Female**	**Male**	**Grand Total**
High Salary	138	118	256
Huge Salary	138	75	213
Low Salary	130	187	317
Medium Salary	101	113	214
Grand Total	**507**	**493**	**1000**

Row Labels	Female	Male
High Salary	129.792	126.208
Huge Salary	107.991	105.009
Low Salary	160.719	156.281
Medium Salary	108.498	105.502

Chi-Square	30.92847335
P-Value	8.80042E-07

2. State the degrees of freedom for the test. (4 points)

4

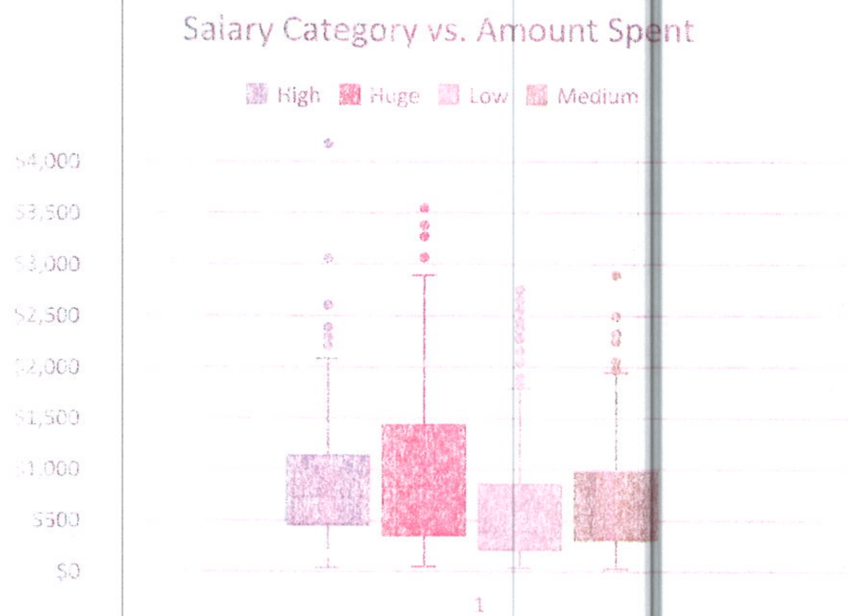
3. In the context of this problem, describe what a Type II error would mean. (6 points)

It would mean that the true state of nature is that the variables gender and salary are dependent, but we do not have sufficient evidence to prove this.

4. Explain how the value for the cell Low Salary and Female is calculated in the Expected Table. (4 points)

$$\frac{317 * 507}{1006} \quad \text{or} \quad \frac{\text{Total Low} * \text{Total Female}}{\text{Grand Total}}$$

Below you will find a boxplot and ANOVA output for a test of salary category vs. Amount Spent. Use this information to answer the questions that follow.



Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
High	256	218588	853.859375	342556.2468
Huge	213	209922	985.5492958	556203.0412
Low	317	197266	622.2902208	326807.8142
Medium	214	156552	731.5514019	310714.2016

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	18777330.51	3	6259110.171	16.63660611	1.49348E-10	2.613839375
Within Groups	374720281.9	996	376225.1826			
Total	393497612.4	999				

5. Do the graphs and table appears to support the assumption of approximately equal variances? Why or why not? (4 points)

Low, Medium & High all look about the same but huge is bigger (but less than x2)

6. Does the ANOVA test provide sufficient evidence to support the claim that salary and amount spent are related? Explain. (6 points)

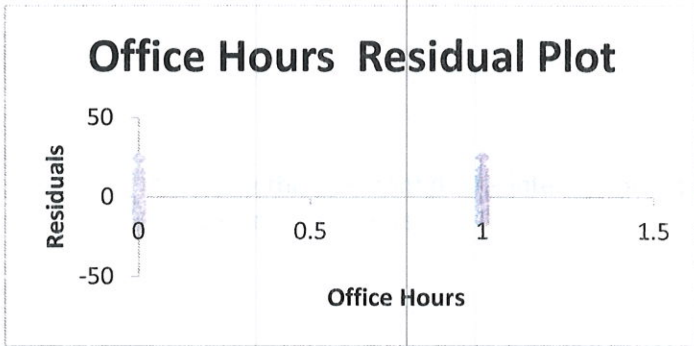
yes. The p-value is very small, so we reject the null that all means are the same so the amount sent does depend on salary category

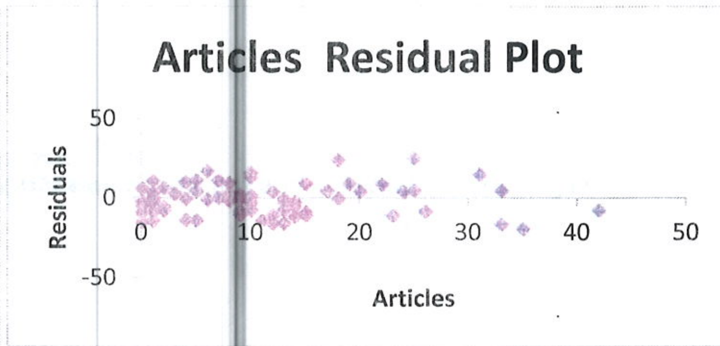
7. Based on the boxplot, which two boxplots appears to be the most different? (4 points)

Low and huge

Use the data in the residual graphs, correlation tables and regression output on the pages that follow to answer the remaining questions.

	Evaluation	*Articles*	*Office Hours*	*Salary*
Evaluation	1			
Articles	0.491503	1		
Office Hours	0.052011	-0.19992	1	
Salary	0.500421	0.898134	-0.288782371	1





8. Based on the residual plots, does the data appear to satisfy the equal variance assumption? Explain. (4 points)

yes, the spread on both graphs is pretty even

9. Based on the table of correlations, why does it make sense that the final model would include articles and evaluation as variables, but not office hours? Explain. (6 points)

the correlations for evaluations and salary are the highest compared to office hours

10. What is the final regression model and its R^2 value? (6 points)

$Y = 78.21 + 2.12X_1 - 5.16X_2$

articles   office hours

$R^2 = 0.819$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.905025681
R Square	0.819071483
Adjusted R Square	0.814045691
Standard Error	9.832671627
Observations	75

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	31513.04361	15756.52181	162.9736092	1.86317E-27
Residual	72	6961.063056	96.68143133		
Total	74	38474.10667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	78.20923201	2.297503448	34.04096393	3.99829E-46	73.62924259	82.78922143	72.13037812	84.28808591
Articles	2.119375315	0.123865903	17.11023985	4.46456E-27	1.87245315	2.366297481	1.791644478	2.447106153
Office Hours	-5.157808649	2.319402286	-2.223766304	0.029301745	-9.781452601	-0.534164696	-11.29460362	0.978986326

Standard errors: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$s_{x_1-x_2} = s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Sample sizes: $n > \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$ $n > \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ $m = n = \frac{4z_{\alpha/2}^2(\sigma_1^2+\sigma_2^2)}{w^2}$

Confidence intervals:

One sample: $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two samples (independent): $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n-1}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Test statistics:

One sample: $z\ or\ t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Two samples: dependent: $z\ or\ t = \frac{\bar{d}_0-\delta}{\frac{s_d}{\sqrt{n}}}$

Independent: $z\ or\ t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$

Degrees of freedom (two samples, unpooled) $v = \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

χ^2Tests: $\chi^2 = \sum_{all\ cells}\frac{(obs-\exp)^2}{exp}$

ANOVA: $MSE = \frac{\left(\sum_{j=1}^{J} n_j(\bar{Y}_j-\bar{\bar{Y}})^2\right)}{J-1}$ $MSS = \sum_{j=1}^{J}\frac{(n_j-1)s_j^2}{n-J}$ $F = \frac{MSE}{MSS}$