BUS 310, Final Exam A, Fall 2018

Name KEY

Section

Instructions: This exam is in two 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. 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. Be sure you are using the data file that matches the exam version you are given.

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

1. Use the work done at home to answer the Part I questions.
2. Open the file from the in-class portion of the final posted on Blackboard that corresponds to the version of the exam you have. This is Exam A.
3. Answer the questions corresponding to the data file, and any additional calculation in Excel required.
4. When you have finished answering questions on the exam, and all your answers have been recorded on the paper test for grading, upload **both** the take home Excel file **and** the in-class Excel file to the same in-class Exam folder in Blackboard for grading. Only those files submitted to the correct folder will be graded. (If in doubt, put all work in one Excel file.)
5. Turn in your paper copy of the exam to your instructor.
6. Enjoy your break!

Part I:

The following questions refer to problem #1 from Part I:

1. Write the objective function you are using to minimize production cost. State the minimum cost. (8 points)

$36{,}000x + 48{,}000y = \underline{1{,}680{,}000}$ or

$36x + 48y = \underline{1{,}680}$ in thousands

2. How many of each type of beer should be made to produce the minimum cost? (8 points)

Regular (x) = 28
Light (y) = 14

3. What is the shadow price for regular beer. Interpret the meaning of this value. (8 points)

0

this means the value of constraint is not exactly satisfied and a change in the constraint will not change result.

The following questions refer to problem #2 from Part I:

4. For your complete model, which variable had the highest P-value? State the variable name and the P-value. (8 points)

P-value = 0.95
Variable = Resident Tuition/fees

5. After eliminating all variables whose coefficients failed their t-tests, write the final regression equation you obtained, the R^2 value, and explain your reasoning for choosing it. (12 points)

$y = 20.4656x_1 + 139.696x_2 - 387.133x_3$

enroll GMAT % Int

$R^2 = 0.9923$

6. Define the term overfitting. Why is overfitting bad when developing regression models? (8 points)

trying to predict a trend in a data which is too noisy with too many variables (overly complex) to false try to fit the noise.

The predictions made are likely to be faulty

7. Did any of the surviving variables in the final model appear to be nonlinear? Why or why not? (8 points)

not especially, but there does appear to be a strong outlier (influential)

8. State a 95% confidence interval for the coefficient for Percent International in your final model. Interpret it in context. (8 points)

(−699,56, −74.70)

we are 95% confident that the true value of coefficient for % Int is between −699.56 and −74.70.

9. Interpret the meaning of the slope for Average GMAT in context. (8 points)

139.70

for each increase in average GMAT score the salary will go up by 139.70.

10. Use your equation to predict the average starting salary of a business student with an enrollment of 1234, average GMAT of 700, resident tuition 94,104, Percent International of 33, Percent Female of 32, Percent Asian of 12, Percent Minority of 13, Percent with Job Offers, 94. Construct a 95% prediction interval around that prediction. [Hint: Use you best model. If the model does not contain a particular variable, omit it as irrelevant.] (12 points)

110,266.48 midpoint

(87,559, 132,973)

11. Examine your residual graphs for your best model. Do any of the graphs indicate the variables heteroscedastic or nonlinear? Explain. (8 points)

there does appear to be an outlier, but otherwise they look good

12. Interpret the meaning of the R^2 value in the context of the problem. (8 points)

99.23% of the variability in average starting salary can be accounted for by these three variables

13. Are there any outliers in the data? Use the residuals and residual plots to determine which point is suspect. Use your standard error for the model. Find the outlier on the list of residuals produced by the regression analysis. Multiply the standard error by two. Is the absolute value of the residual larger than twice the standard error? If so, it's an outlier. If not, then it should be left in the model. Describe what you found. (15 points)

yes, associated w/ obs. #41

it is an extreme outlier

and should be pulled from model

The following questions are based on problem #3 from Part I:

14. Using data on public and private business schools, determine if the two measurements are dependent or independent. Explain your reasoning. (6 points)

they are independent
Sample sizes are not the same

15. Conduct an appropriate *t*-test to determine if private schools result in higher initial starting salaries or not. State the null and alternative hypotheses, test statistic, P-value and state the results in an English sentence understandable to a non-statistician. (12 points)

Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 > \mu_2$

T: 1.2588
P-value: 0.1063
fail to reject null

There is not sufficient evidence to think private school graduates make a higher salary than public school graduates

Calculations in Excel: (1) 30 points, (2) 50 points, (3) 25 points.

Part II:

16. A study is conducted to analyze the weekly food expenses for a family of four. One previous analysis suggests that the mean weekly food expense is \$260. Conduct a hypothesis test of means to determine if this result has changed from previous results. State the hypotheses, test statistic, P-value and conclusion. Is this sufficient evidence to think the weekly food expense is not \$260? (12 points)

Ho: $\mu = 260$
Ha: $\mu \neq 260$
T stat = -0.0514
P-value = .959 >> .05
fail to reject null

This is not good evidence to think weekly ... changed.

17. Interpret a Type I and Type II error in the context of this problem. (8 points)

Type 1: expenses have not changed but we think they did

Type 2: expenses have changed, but we do not have good evidence to prove that

18. Construct a 70% confidence interval for the mean weekly food expense. Interpret the interval in context. (8 points)

(255.10, 264.56)

we are 70% confident that the true mean weekly food expense is between 255.10 and 264.56.

19. Suppose that you wish to sample employees of a large company to determine factors that predict high inside sales commissions in order to prepare for a new training program. The company has ~~1300~~ 975 employees in this position around the world. The company wants to select 10 of them for an initial study of best practices. Eligible employees are assigned numbers from 1 to 975 based on their date of initial hire. Select a simple random sample and report the employee numbers you have selected below. (6 points)

answers will vary

808, 892, 912, 452, 591, 604,

212, 195, 195, 268, 28

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

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

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

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

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

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

Upload your completed Excel files to the Final Exam **to be graded** submission box in Blackboard, and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.