BUS 310, Exam #2A, Fall 2017 Name KEY

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:

1. Do the boxplots support the equal variance assumption of ANOVA? (8 points)

no, the boxplots do not support the equal variance assumption

2. Using the data on Amount Spent vs. History, conduct a one-way ANOVA test. Record null and alternative hypotheses, the F-statistic, and the P-value below. What do you conclude from this test (taking your answer above into account)? (15 points)

H_0: all μ the same $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_a: at least one mean different
$F = 283.25$
P-value: 6.5×10^{-133}

P-value says reject H_0
However, this is not the appropriate test for data.

3. Write the equation for your final multiple regression model here. (8 points)

$$Y = -442.77 + 47.69X_1 - 198.69X_2 + .0204X_3$$

4. Interpret the meaning of each slope coefficient. (16 points)

as the # of Catalogs increases (by 1), the amount of Sales goes up by $47.69.

as the # of Children increases (by 1), the amount of Sales goes down by $198.69.

as the Salary goes up (by $1), the amount of Sales goes up by $0.02

5. Do any of the variables from your scatterplots appear to be nonlinear? Are there any outliers? Explain. (16 points)

None of the models appears strongly nonlinear. The nonlinear models do not dramatically improve R^2.

Salary may have one possible outliers

6. Consider the residual graph for your final equation in #3. Do there appear to be any problems with the model? Explain. (8 points)

Because the data is discrete, many of the residual graphs are unusual. There may be a problem of unequal variance of all the variables, but it is not dramatic.

Calculations in Excel: (1) 25 points, (2) 50 points, (3) 20 points, (4) 20 points.

Part II:

Use the ANOVA table below to answer the questions that follow.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
National	10	98	9.8	4.177778
Competitor 1	10	113	11.3	4.011111
Competitor 2	10	126	12.6	4.044444

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	39.26667	2	19.63333	4.814714	0.016282	3.354131
Within Groups	110.1	27	4.077778			
Total	149.3667	29				

7. State the null and alternative hypothesis for single-factor ANOVA using proper notation. (4 points)

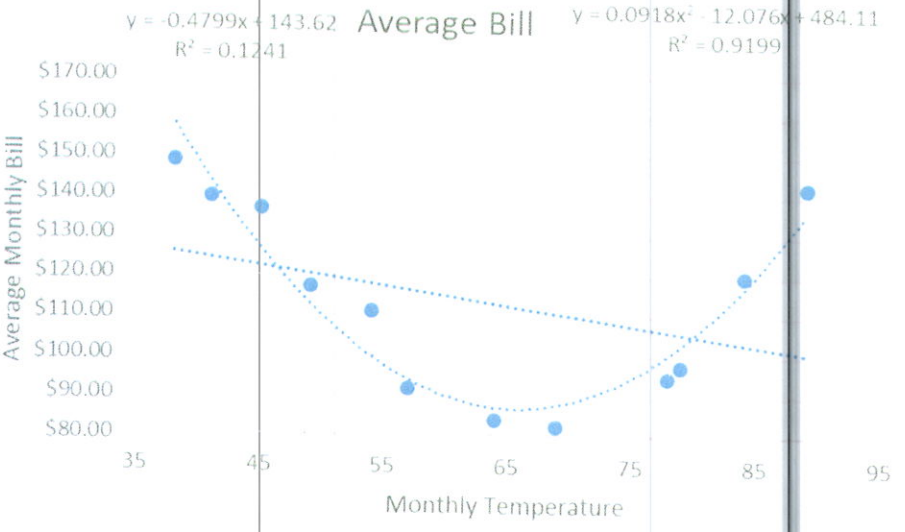
$H_0: \mu_i = \mu_j$ for all i, j (all means the same)

$H_a: \mu_i \neq \mu_j$ for some $i \neq j$ (at least one mean is different)

8. Using a 1% significance level, do you reject or fail to reject the null hypothesis? (4 points)

P-value = 0.016 > .01 fail to reject H_0

Use the scatterplot below of temperature and average electric bill to answer the questions that follow.



9. Based on the scatterplot, is the linear model an appropriate model for the data? (4 points)

NO

10. Using the R^2 value for the better model, interpret this value? (8 points)

$R^2 = .9199$

roughly 92% of the variability in average monthly energy bill can be explained by the change in average monthly temperature.

11. Using the better model, predict the average monthly bill for a month with an average monthly temperature of 94 degrees. (6 points)

$160.11

12. Which variable in the scatterplot is the explanatory variable? (4 points)

average monthly temperature

Use the multiple regression output and the residual plots to answer the following questions. The data predicts an overall employee rating based on the results of four tests.

Regression Statistics	
Multiple R	0.9157898
R Square	0.8386710
Adjusted R Square	0.8128584
Standard Error	6.3022990
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	5161.9923	1290.498	32.49072	1.43375E-09
Residual	25	992.97432	39.71897		
Total	29	6154.9666			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	-57.198345	10.438851	-5.479	1.08E-05	-78.6975	-35.6991	-75.0293	-39.3673
Test1	0.6079173	0.1140467	5.330	1.59E-05	0.37303	0.84280	0.41310	0.80272
Test2	0.4869782	0.1466522	3.3206	0.00276	0.18494	0.78901	0.23647	0.73748
Test3	-0.6185604	0.1778282	-3.478	0.001864	-0.98480	-0.25231	-0.92231	-0.31480
Test4	1.2308779	0.1956889	6.2899	1.4E-06	0.82784	1.63390	0.89661	1.56514

13. State the null and alternative hypotheses for a multiple regression analysis. (6 points)

model as a whole: $H_0: R^2 = 0$ $H_a: R^2 \neq 0$ or all $\beta_i = 0$ some $\beta_i \neq 0$

coefficients $H_0: \beta_i = 0$ $H_a: \beta_i \neq 0$

14. Interpret the standard error of the test. (4 points)

SE = 6.302

the average distance of an observation from the result predicted by the model is 6.3 points

15. Interpret the coefficient reported for Test #2 in context. (4 points)

0.4869782

for each increase in Test #2 score, the overall employee rating is expected to go up by .49 points.

16. State a 90% confidence interval for the coefficient for Test #4. (6 points)

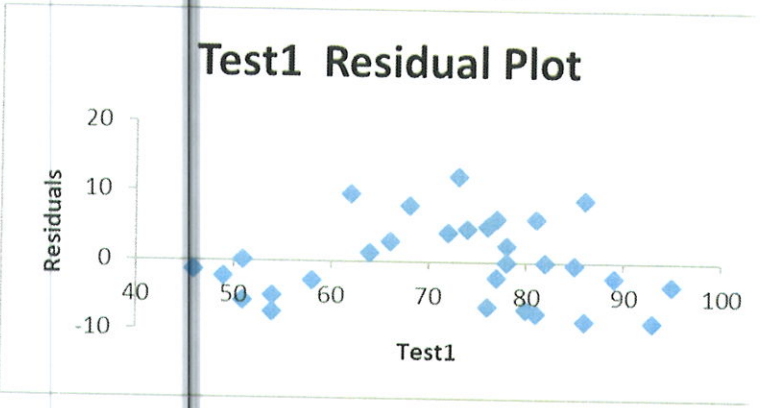
(0.89661, 1.56514)

17. Can any coefficients be eliminated from the model? Why or why not? (6 points)

no. all P-values are less than even 1%

18. The Residual plot vs. Test #1 is shown. Does the plot appear to exhibit any problems? Why or why not? (6 points)

no. all are random w/ equal variance - no patterns



19. What assumption of regression models is being tested in the residual plot? (6 points)

equal variance
linearity

20. Calculate a 95% prediction interval for the rating of a single employ with test scores $x_1 = 75, x_2 = 56, x_3 = 88, x_4 = 74$. (6 points)

$Y = -57.198 + 0.6079 * 75 + 0.48698 * 56 +$
$-0.6186 * 88 + 1.2318 * 74$

$= 52.34 \rightarrow (39.36, 65.32)$

21. Explain the procedures for dealing with an outlier. (6 points)

Copy the data
remove the outliers from the copy.
redo the analysis.
determine the extent of the impact of the outliers on the model. use that to determine whether to keep the outliers out or keep it in.

Upload your completed Excel files to the Exam #2 submission box in Blackboard, and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.