

**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  $\mu$ 's the same  $\mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_a$ : at least one mean is different

$F = 283.25$

P-value:  $6.5 \times 10^{-133}$

P-value says reject  $H_0$ , but this is not the appropriate test for the 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 appear to be strongly nonlinear  
 the nonlinear models do not dramatically improve  $R^2$ .

Salary may have one outlier.

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 for these variables 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
Electric	20	2787.52	139.376	9.934899
Gas	20	2738.29	136.9145	10.46487
Oil	20	2771.51	138.5755	10.59245

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	63.05802	2	31.52901	3.051961	0.055059	3.158843
Within Groups	588.8521	57	10.33074			
Total	651.9101	59				

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

$$H_0: \mu_i = \mu_j \text{ (all means are the same)}$$

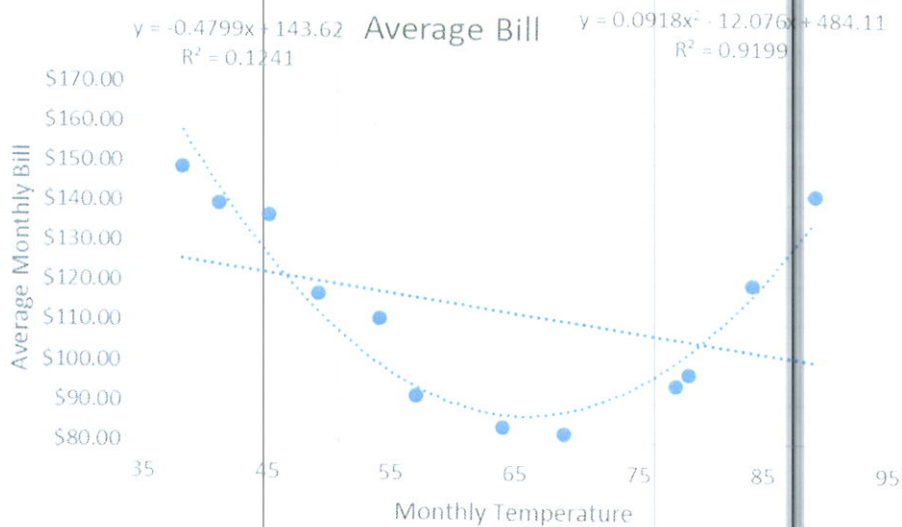
$$H_a: \mu_i \neq \mu_j \text{ for some } i \neq j \text{ (at least one mean is different)}$$

8. Using a 5% significance level, do you reject or fail to reject the null hypothesis? (4 points)

$$P\text{-value} = 0.055 > 0.05$$

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 34 degrees. (6 points)

\$ 179.65

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		ANOVA						
Multiple R	0.9157898							
R Square	0.8386710							
Adjusted R Square	0.8128584							
Standard Error	6.3022990							
Observations	30							
	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:  $H_0: R^2 = 0$  or  $\text{all } \beta_i = 0$   
 $H_a: R^2 \neq 0$  or  $\text{some } \beta_i \neq 0$

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 predicted model is 6.3 points

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

1.23  
for each increase in point in Test 4 score,  
the overall employee rating increase by 1.23 points

16. State a 90% confidence interval for the coefficient for Test #1. (6 points)

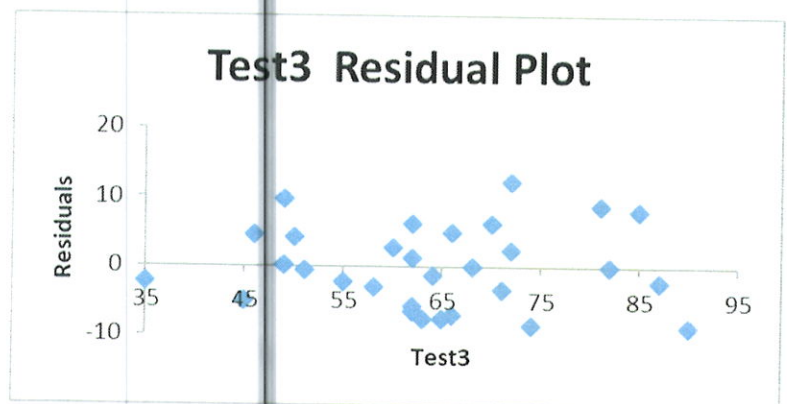
(0.41310, 0.80272)

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

no. all P-values less than  
even 1%

18. The Residual plot vs. Test #3 is shown. Does the plot appear to exhibit any problems? Why or why not? (6 points)

no. residual  
appear random w/  
equal variance



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

equal variance assumption  
linearity

20. Calculate a 90% prediction interval for the rating of a single employ with test scores  $x_1 = 71, x_2 = 65, x_3 = 98, x_4 = 79$ . (6 points)

$$\begin{aligned} Y &= -57.198 + 0.6079 * 71 + 0.48698 * 65 - 0.6186 * 98 \\ &\quad + 1.231 * 79 \\ &= 54.26 \rightarrow (43.50, 65.03) \end{aligned}$$

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

Copy the data.  
Remove the outlier from the copy  
redo the analysis.  
determine the extent of the impact of the outlier  
on the model. Use that to determine  
whether to keep the outlier 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.