BUS 310, Exam #2B, 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 B.
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 problems refer to problem #1 of Part I:

1. Report on the results of your ANOVA test of the types of Experience Training and their influence on sales. State your null and alternative hypotheses, your test-statistic and P-value, and the conclusion of your test. Give a sentence to explain the meaning of the test in context understandable by a lay person. (10 points)

Ho: all means are the same
Ha: at least one mean is different
F-stat = 21.168 P-value: $4.559 \times 10^{-10} < .05$
reject null; different training methods do impact sales

2. Examine your boxplots for Experience Training. Is the equal variance assumption approximately satisfied? Why or why not? (6 points)

no, the equal variance assumptions do not appear to be met.

Experience #2 is much higher than others compared to range of data.

3. Describe the results of your two-sample t-test. State the null and alternative hypotheses and interpret the result in the context of the problem. Compare the results to the comparative boxplots. Do they appear to agree? (8 points)

Ho: $\mu_1 = \mu_2$ (means are equal)
Ha: $\mu_1 \neq \mu_2$ (means are not equal)
T-stat: 2.83 P-value: 0.00588 < .05 reject null
presentation type does appear to influence results.
agrees w/ plot

4. Based on the results of your tests, and the box plots, which training method or presentation type appears to have the biggest effect on sales? Explain. How would a sales manager incorporate this information into their management strategy? (6 points)

it appears that presentation style one and experience training 4 have the largest means and so are most effective.

The follow questions refer to problems #2 from Part I:

5. Report on the findings of your χ^2-test of independence. State the null and alternative hypotheses, your test statistic and P-value, and the conclusion. Give a sentence that summarizes the meaning of the test that a lay person can understand. (10 points)

Ho: CardType and Region are independent
Ha: CardType ≠ Region are dependent
P-value: 0.965 >> .05 fail to reject null
Cardtype and region are independent (not related)

6. Referring back to your pivot table of the data, report the value of cell of Card Type=ElecMart, and Region=NorthEast, and the value of the expected count for that same cell, and explain how you calculated that value. (6 points)

27 observed
27.025 expected
⇓
Column total * row total / grand total = 115*94/400

The following questions refer to problem #3 from Part I:

7. Is the heating bill dependent or independent of home type? Explain. [Hint: this question is not about a specific hypothesis test. This question is about your knowledge of the data and the real world.] (5 points)

independent
Since the question does not say
they are matched or paired

8. Report on the results of the t-test. State the type of test conducted, the null and alternative hypotheses, the test-statistic and P-value, and the conclusion of the test. Summarize the results in a single sentence that can explain the results in context to a lay person unfamiliar with statistics. (10 points)

t-test, 2 sample, unpaired, pooled

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

T-test: -.073097

P-value: 0.4677 > 7.05

fail to reject null

home type does not affect cost of electric bill

The following questions refer to problem #4 from Part I:

9. State your best-fit final regression equation and explain your reasoning as to why you chose this option. Report the R^2 value for the equation you choose. (10 points)

$y = 0.08168x$ x is Miles Driven $R^2 = .99407$

P-value for constant and Age of Truck were over 0.05

10. What proportion of the variability in quantity sold can be explained by the variables you chose? This question should be consistent with your previous results. (6 points)

99.4%

11. Using your final model reported above, interpret the value of the slope for Miles Driven in the context of the problem. (8 points)

for each additional mile driven approximately $0.08 is added to maintenance cost

12. Using your final model reported above, predict the maintenance cost for a truck that drives 7,000 miles and is 9 years old, if the trend continues. (6 points)

$571.78

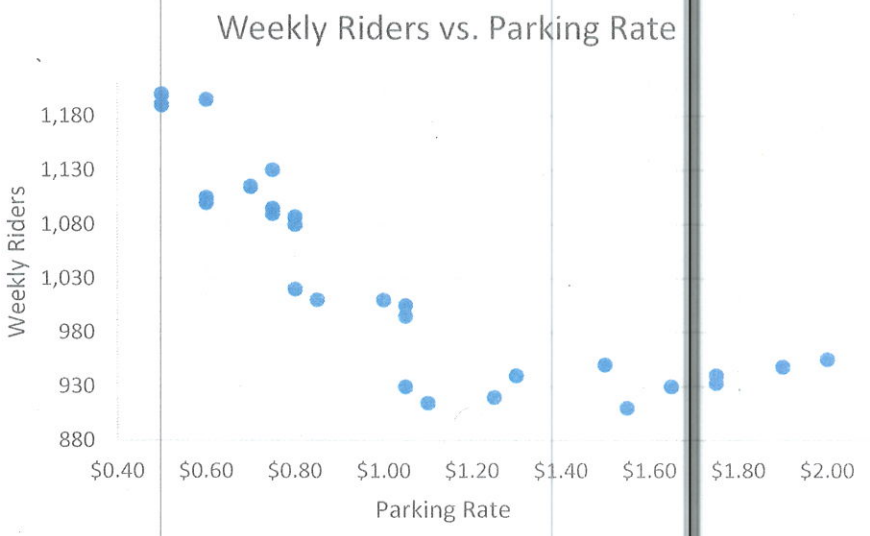
13. Create a scatterplot of Age of Truck vs. Miles Driven. Together with your other scatterplots, provide at least two possible problems with the data. Are independent variables actually dependent? Are the graphs nonlinear? Are there any outliers? (8 points)

residuals appear to have a slight curve although nonlinear graphs do not provide much improvement for R^2 value based on linear model of trendline
2 possible mild outliers in final model

Calculations in Excel: (1) 30 points, (2) 25 points, (3) 25 points, (4) 30 points.

Part II:

14. Included below is a scatterplot of weekly riders vs. daily parking rate at the park-n-ride. Based on the graph, does the data appear to be linear or nonlinear? (6 points)



nonlinear

15. Does the general trend of the graph appear to be positive (increasing) or negative (decreasing)? (5 points)

decreasing

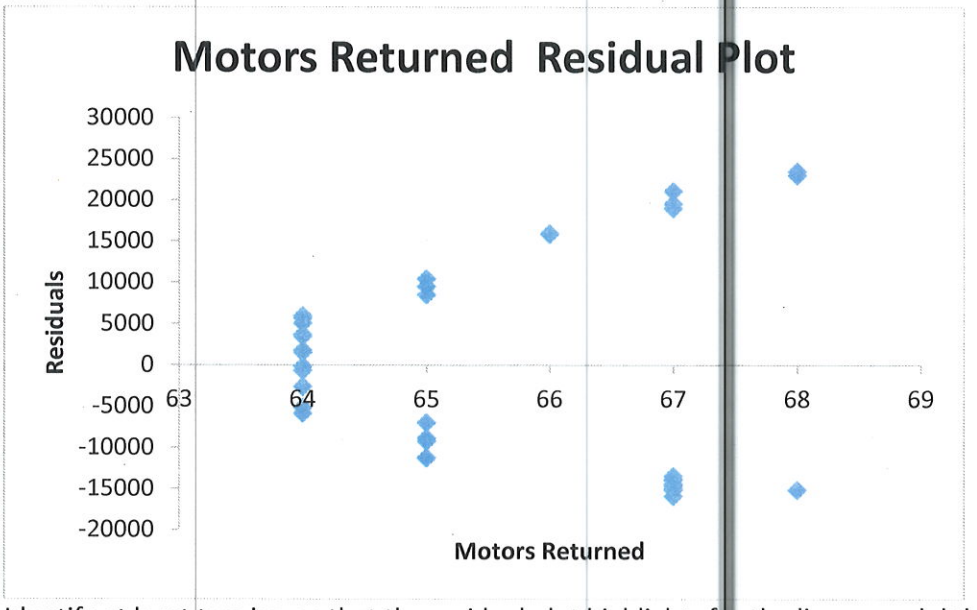
16. The R^2 value for the linear trendline is 0.6801. What is the value of the correlation? (4 points)

-0.82468

17. The linear trendline for this graph is $y = -170.83x + 1208.7$. Use the equation to predict the number of riders, if the trend continues, if parking rates rise to $8.00. Does this value make sense? Why or why not? (6 points)

# of rides is predicted to be -328 or -329.
This does not make sense since we can't have negative rides

18. The following graph displays the residual plot for the relationship between inspection costs and motors returned.



Identify at least two issues that the residual plot highlights for the linear model that produced these residuals. Explain what features of the graph are related to each issue. (8 points)

nonlinear (pattern)
heteroscedastic
possible outliers

19. The regression output for the multiple regression model predicting ridership from four variables is shown on the page following this problem. Use this information to answer the questions that follow.
    a. What is the standard error? Interpret the meaning of this value in context? (6 points)

21.04 ~ average distance of observed value from prediction w/ model

b. State a 99% confidence interval for the coefficient of Price Per Ride. (6 points)

$(-300.41, -26.36)$

c. Conduct a hypothesis test on the coefficient of Population in the equation. State the hypothesis, test statistic and P-value, and interpret the results in the context of the problem. (8 points)

$H_0: \beta_{population} = 0$

$H_a: \beta_{population} \neq 0$

coefficient for population is non-zero - keep in eq.

T-test: 58.0779 P-value: $1.868 \times 10^{-26} << .05$

d. Interpret the meaning of the slope coefficient for Parking Rate in the context of the problem. (6 points)

192.84

For each $1 increase in parking rate, # of rides increases by about 193

e. Write the equation for the multiple regression model. (6 points)

$y = -163.39x_1 + 0.687x_2 + 0.0448x_3 + 192.84x_4$

x_1 ↓ price per ride; x_2 ↓ population; x_3 ↓ income; x_4 ↓ parking rate

f. Use the model above to predict the number of weekly riders for $0.25 price per ride, 1800 population, $5,500 income and $1.25 parking. (8 points)

1,190.58

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9998222
R Square	0.9996445
Adjusted R Square	0.9561198
Standard Error	21.042083
Observations	27

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	28633278	7158320	16167.1564	8.4419E-38
Residual	23	10183.69	442.7692		
Total	27	28643462			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Price per Ride	-163.386	48.81101	-3.34732	0.00279269	-264.3593027	-62.41276	-300.4149	-26.3571336
Population	0.6871328	0.011831	58.07791	1.8681E-26	0.662658021	0.7116075	0.653919	0.720346994
Income	-0.044813	0.008022	-5.58622	1.103E-05	-0.061408014	-0.028218	-0.067334	-0.0222924
Parking Rate	192.84261	35.06938	5.498888	1.365E-05	120.2960656	265.38916	94.39108	291.2941499

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 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.