BUS 310, Exam #2C, Spring 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 C.
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:

1. Report on the results of your ANOVA test of the cereal box filling machines. 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. (12 points)

Ho: all means are the same $\mu_i = \mu_j$ for all $i \neq j$
Ha: at least one mean is different $\mu_i \neq \mu_j$ for one (or more) $i \neq j$

F = 326.88 P-value: $1.268 \times 10^{-37} < .05$ reject null

at least one machine fills box differently

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

No, variance 1 is 6 times larger than Machine 5's

3. Based on the results of your test, and the box plots, which filling machine(s) appears to be the most different from the others and is most in need of recalibration? (Or if no recalibration is needed.) Explain. (6 points)

Machine 4 & Machine 5 are the most different

all the machines appear to need looked at.

only Machine 1 appears to span 10 oz.

4. 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. (12 points)

Ho: The variables are independent
Ha: The variables are dependent

P-value: .30 > .05 fail to reject null

The variables month and buy category appear to be unrelated.

5. Referring back to your pivot table of the data, report the value of cell of Buy Category=Low, and Month=May, and the value of the expected count for that same cell, and explain how you calculated that value. (6 points)

Obs. Exp

45 47 $\dfrac{\text{row May total * column Low total}}{\text{Grand Total}}$

6. Consider the data on the two ads. Is the data dependent or independent? Explain. (10 points)

dependent

Since they are seen by the same person

7. Report on the results of the *t*-test. State any assumptions made about the data, and 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. (15 points)

$H_0: \mu_1 = \mu_2$ or $\delta_0 = 0$

$H_a: \mu_1 < \mu_2$ or $\delta_0 > 0$ $(\mu_2 - \mu_1 > 0)$

P-value: $4.16 \times 10^{-13} < .05$ reject null

the second ad appears to be more effective.

8. State your final regression equation and explain your reasoning as to why you chose this option. Report the R^2 value for the equation you choose. (12 points)

$y = -2570.38x_1 + 1.6965x_2$

QTY = −2570.38 (Price) + 1.6965 (Ads)

$R^2 = .9965$

P-values are < .05 for all variables

Constant eliminated due to high P-value when price included

9. What proportion of the variability in quantity sold can be explained by the variables you chose? (6 points)

99.65% (match R^2 value)

10. Use the slope for Average Price, interpret the value of the slope in the context of the problem. (8 points)

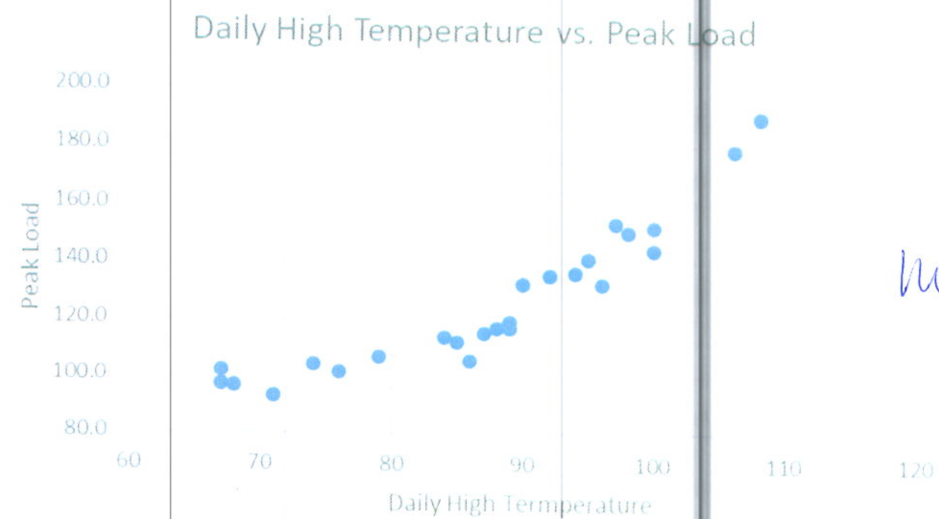
−2570.38

for each $1 price increase, the quantity sold decreases by 2570.

Calculations in Excel: (1) 25 points, (2) 25 points, (3) 20 points, (4) 40 points.

Part II:

11. Included below is a scatterplot of peak load vs. daily temperature. Based on the graph, does the data appear to be linear or nonlinear? (6 points)



nonlinear

12. Using the trendline option on the graph, compare a linear regression line (and R^2 value) with a polynomial (degree two) regression equation. Report the R^2 values and equations of both and indicate which is the better model. (9 points)

linear

$Y = 1.9765x - 47.394$

$R^2 = 0.8433$

polynomial

$Y = 0.0598x^2 - 8.295x + 385.05$

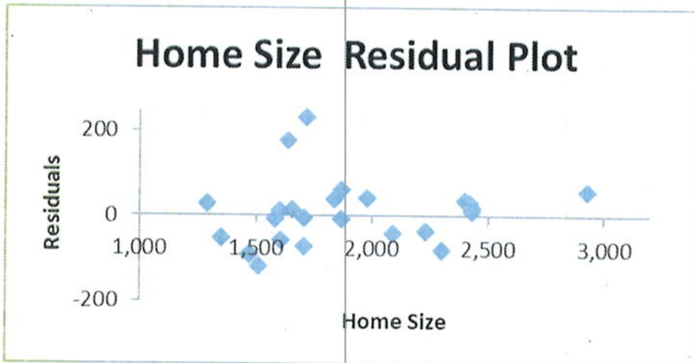
$R^2 = 0.9594$

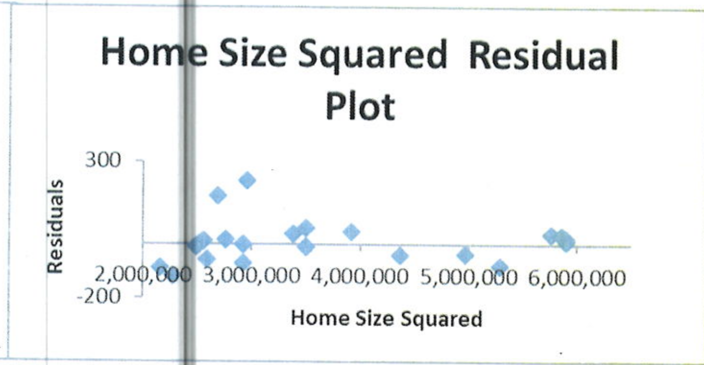
better model

13. Use the best equation to predict the peak load for an average daily temperature of 82° in your best equation. (6 points)

107.2

14. The regression output for the quadratic (polynomial degree-2 model) is shown on the next page along with the residual plots. Use this information to answer the questions that follow.





a. What information do we get from residual plots? Which regression assumptions are we testing? (6 points)

equal variance
we can also test bias, linearity/fit of model qualitatively

b. What is the standard error? Interpret the meaning of this value in context? (6 points)

91.75

st. dev. of errors (residuals) from regression equation

c. State a 95% confidence interval for the coefficient of Home Size Squared. (6 points)

(−0.000658, −0.00024)

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

$H_0: \beta_1 = 0$
$H_a: \beta_2 \neq 0$

T = 5.6955   P-value = 9.96×10^{-6} < .05 reject null

keep coeff in equation; home size does predict monthly usage

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.935647138
R Square	0.875435567
Adjusted R Square	0.864111527
Standard Error	91.75229367
Observations	25

ANOVA

	df	SS	MS	F	Significance F
Regression	2	1301627.365	650813.6827	77.30771116	1.1203E-10
Residual	22	185206.6347	8418.483394		
Total	24	1486834			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 90.0%	Upper 90.0%
Intercept	-1125.35251	407.813232	-2.759480129	0.011438868	-1971.105388	-279.5996314	-1825.626707	-425.0783128
Home Size	2.346800787	0.412042263	5.695534167	9.96046E-06	1.492277435	3.201324139	1.639264733	3.054336841
Home Size Squared	-0.000449192	0.000100737	-4.459039505	0.000196587	-0.000658108	-0.000240275	-0.000622172	-0.000276211

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) $\nu = \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

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.