Exam #2B

Summer 2019C

Part I:

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

1. To what extent do the boxplots support the equal variance assumption of ANOVA? Explain. (8 points)

Where history is known, it seems to be okay, but there is greater variance where history is N/A

2. Using the data on Salary and 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 in laymen's terms? (10 points)

H_0: all means the same

H_a: at least one mean is different

$F = 179.84$

P-value $= 3.57 \times 10^{-93} < 0.05$ reject null

all the means are not the same; history affects salary

The following questions refer to problems #2-4 from Part I:

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

$$\underset{\text{Amt. Spt.}}{Y} = -442.77 + 47.69\underset{\text{Catalogs}}{x_1} - 198.69\underset{\text{Children}}{x_2} + 0.0204\underset{\text{Salary}}{x_3}$$

4. State the R^2 value for your final model and interpret the meaning in context. (8 points)

65.8%

65.8% of variability in Amt Spent can be explained by its relationship with Catalogs, Children and Salary.

5. Interpret the meaning of the Salary slope coefficient. (6 points)

for each additional dollar of salary, customers will spend on average an additional two cents.

6. Do any of the variables from your scatterplots appear to be nonlinear? Explain. (8 points)

None are strongly nonlinear

7. Consider the residual graphs for your final equation stated in #3. Do there appear to be any problems with the model? Is the equal variance assumption satisfied? Explain. (8 points)

The equal variance assumption seems to be the most potentially problematic for all three variables

The following questions refer to problems #5 from Part I:

8. Record your χ^2 test of independence here. Clearly state the hypothesis, all key test statistics and the P-value. Interpret the results of the test in context. (10 points)

H_0: variables are independent
H_a: variables are dependent
$\chi^2 = 8.406$
P-value = 0.4938 > 0.05 fail to reject null
The variables for region and history are not dependent

The following questions refer to problems #6 from Part I:

9. State the null and alternative hypothesis for the one-sample t-test. State the test-statistic and P-value. What is the conclusion of your test? (8 points)

$H_0: \mu = 1165$
$H_a: \mu > 1165$
$T = 1.70$
P-value = 0.044 < 0.05 reject null
The amount spent is higher than in the past

The following questions refer to problems #7 from Part I:

10. Is the two-sample t-test you conducted a paired t-test or a pooled t-test? (6 points)

Pooled (independent)

11. State the null and alternative hypotheses for your two-sample t-test. What was the test-statistic and P-value? Explain the meaning of your results. (10 points)

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$T = 11.83$

P-value $= 2.478 \times 10^{-30} \ll 0.05$ reject null

There is a difference between those who do and do not own their own homes

Calculations in Excel: (1) 20 points, (2-4) 50 points, (5-7) 40 points.

Part II:

12. Suppose that the alternative hypothesis of a one-sample test of means is $H_a: \mu \neq 45$. Is the hypothesis test one-tailed or two-tailed? (6 points)

2-tailed

13. Suppose that a two-tailed test of a population proportion has a test-statistic of $z = -2.84$. Find the P-value. Use that information to determine whether the null hypothesis would be rejected at the 2% significance level. (8 points)

rejected

P-value = 0.0045

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				

14. State the null and alternative hypothesis for single-factor ANOVA using proper notation and/or terminology. (4 points)

$H_0: \mu_i = \mu_j \quad \forall\ i \neq j$

$H_a: \mu_i \neq \mu_j$ for some $i \neq j$

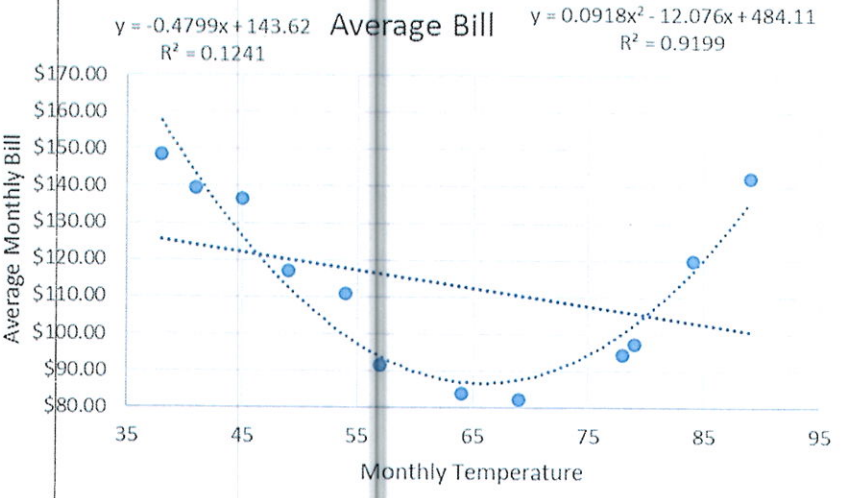
15. Using a 5% significance level, do you reject or fail to reject the null hypothesis? (4 points)

reject null

16. Interpret a Type II error in this context (of the ANOVA test above). (6 points)

we would fail to reject null and conclude all means are the same when they are infact not the same

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



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

no

18. Using the better model, predict the average monthly bill for a month with an average monthly temperature of 85 degrees. (6 points)

$120.91

19. Which variable in the scatterplot is the response variable? (4 points)

amount of bill

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

20. Interpret the coefficient reported for Test #1 in context. (4 points)

0.6079

for each one point increase in Test #1 score, the value of Y increases by 0.61 points on average

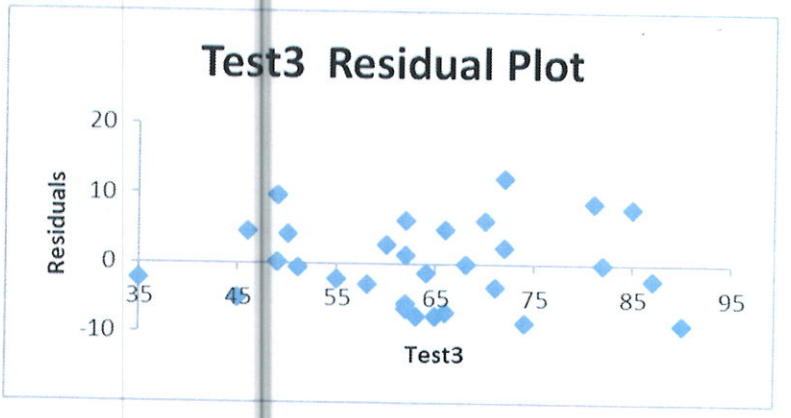
21. State a 90% confidence interval for the coefficient for Test #3. (6 points)

(-0.9223, -0.3148)

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

at 5% significance, no. all p-values less than 0.05

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



no
no patterns
random

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

linearity and equal variance

25. Predict the rating of a single employee with test scores $x_1 = 47, x_2 = 68, x_3 = 79, x_4 = 91$. (5 points)

67.74