STAT 350 Post Exam 2 Material Review Problems

1. True of False Questions (explain your answer):

a) ANOVA stands for analysis of variance; therefore, it is a statistical inference procedure to test if the population variances are different.
b) One-way ANOVA can be used only when there is a single level of a factor, we use k-way ANVOA when there are k levels of a factor.
c) One-way ANOVA can be used only when there are two means to be compared.
d) In ANOVA, the null hypothesis is that the sample means are all equal.
e) In rejecting the null hypothesis in ANOVA, one can conclude that all the means are different from one another.

2. For each of the following situations, identify the response variable, the populations to be compared, and give *k*, *n* and (i) Degrees of freedom for group, for error, and for the total (ii) Null and alternative hypotheses (iii) Numerator and denominator degrees of freedom for the *F* statistic.

a) A poultry farmer is interested in reducing the cholesterol level in his marketable eggs. He wants to compare two different cholesterol-lowering drugs added to the hens' standard diet as well as an all-vegetarian diet. He assigns 25 of his hens to each of the three treatments.
b) A researcher is interested in students' opinions regarding an additional annual fee to support non-income-producing varsity sports. Students were asked to rate their acceptance of this fee on a seven-point scale. She received 94 responses, of which 31 were from students who attend varsity football or basketball games only, 18 were from students who also attend other varsity competitions, and 45 were from students who did not attend any varsity games
c) A professor wants to evaluate the effectiveness of her teaching assistants. In one class period, the 42 students were randomly divided into three equal-sized groups, and each group was taught power calculations from one of the assistants. At the beginning of the next class, each student took a quiz on power calculations, and these scores were compared.

3. Various studies have shown the benefits of massage to manage pain. In one study, 125 adults suffering from osteoarthritis of the knees were randomly assigned to one of five 8-week regimens. The primary outcome was the change in the Western Ontario and McMaster Universities Arthritis Index (WOMAC-Global). This index is used extensively to assess pain and functioning in those suffering from arthritis. Negative values indicate improvement. The following table summarizes the results of those completing the study.

Regimen	n	$\bar{x}$	s
30 min massage 1 × /wk	22	−17.4	17.9
30 min massage 2 × /wk	24	−18.4	20.7
60 min massage 1 × /wk	24	−24.0	18.4
60 min massage 2 × /wk	25	−24.0	19.8
Usual care, no massage	24	−6.3	14.6

Since you do not have the data to run, I have provided you with some of the R output below. (write down the code that was used to generate the output). Be sure to also include code and value of the critical value in the Tukey multiple comparison and any other value explicitly asked for in the questions.

```
             Df  Sum Sq   Mean Sq F value   Pr(>F)
Massage       4  5060.3   1265.09   3.728 0.00688
Residuals   114 38682.5    399.32

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pain ~ Massage)

$Massage
             diff      lwr      upr
30m1-30m2       1  -13.733   15.773
30m1-60m1     6.6   -8.173   21.373
30m1-60m2     6.6   -8.173   21.373
30m1-noc    -11.1  -25.873    3.673
30m2-60m1     5.6   -9.173   20.373
30m2-60m2     5.6   -9.173   20.373
30m2-noc    -12.1  -26.873    2.673
60m1-60m2       0  -14.773   14.733
60m1-noc    -17.7  -32.473   -2.927
60m2-noc    -17.7  -32.473   -2.927
```

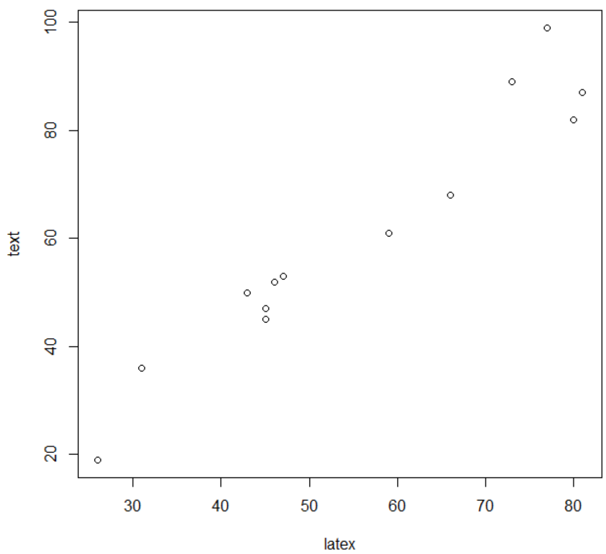
a) What proportion of adults dropped out of the study before completion?
b) Is it reasonable to use the assumption of equal standard deviations when we analyze these data? Please explain your answer.
c) Perform a hypothesis test at a 5% significance level for these data. Be sure to include the degrees of freedom(s) for the test statistic. I would suggest that you create the ANOVA table before you start. What code was used to generate the data shown above?
d) In this study, there are ten pairs of means to compare. If you do this by hand, assume that the number of observations for each case is 24. Determine the critical value for the Tukey multiple-comparisons method at a 5% significance level. Which pairs of means are found to be significantly different? You may use either the hand calculations or the data provided from the output. Please include a graphical representation of your result. Write a short summary of your analysis including what you would recommend to reduce pain. What code was used to generate the output above?

4. The editor of a statistics textbook would like to plan for the next edition. A key variable is the number of pages that will be in the final version. Text files are prepared by the authors using LaTeX, and separate files contain figures and tables. For the previous edition of the textbook, the number of pages in the LaTeX files can easily be determined, as well as the number of pages in the final version of the textbook. Here are the data:

	Chapter												
	1	2	3	4	5	6	7	8	9	10	11	12	13
LaTeX pages	77	73	59	80	45	66	81	45	47	43	31	46	26
Text pages	99	89	61	82	47	68	87	45	53	50	36	52	19

STAT 350 Post Exam 2 Material Review Problems

Scatter Plot



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-6.20176	5.71233	-1.086	0.301
latex	1.20810	0.09828	xxx	8.95e-8

Multiple R-squared: 0.9321, Adjusted R-squared: 0.926

a) What is the explanatory variable? Response variable?
b) Describe the form, direction and strength of the scatter plot above. Are there any outliers?
c) Find the equation of the least-squares regression line based on the software output.
d) Interpret the slope of the regression line.
e) What is the meaning of the y-intercept? Does this variable make sense in this situation? Please explain your response.
f) Find the predicted number of pages for the next edition if the number of LaTeX pages for a chapter is 62.
g) What proportion of the variation in Text pages is explained by LaTex pages?
h) What is the value of the correlation coefficient r?
i) Using the t test, test whether there is an association between LaTeX pages and text pages at a 10% significance level.
j) Find and interpret a 90% confidence interval for the slope β_1
k) Suppose there are 80 LaTex pages in a chapter for the new edition. Raj obtained a 95% confidence interval for the mean pages and a 95% prediction interval for the pages of the final version but did not label them. Which of the following two is the prediction interval? Please explain your answer.
   Interval A (75, 106)
   Interval B (84, 97)
k) Which interval should the editor use to plan for the next edition of the statistics textbook?
l) One new chapter has 200 LaText pages. Discuss if it is appropriate use the regression line obtained in part (c) to predict the final textbook pages.

5. Can a pretest in mathematics skills predict success in a statistics course? The 62 students in an introductory statistics class took a pretest at the beginning of the semester. The least-squares regression line for predicting the score, y, on the final exam from the pretest score, x, was y = 13.8 + 0.81x. The standard error of b_1 was 0.43.

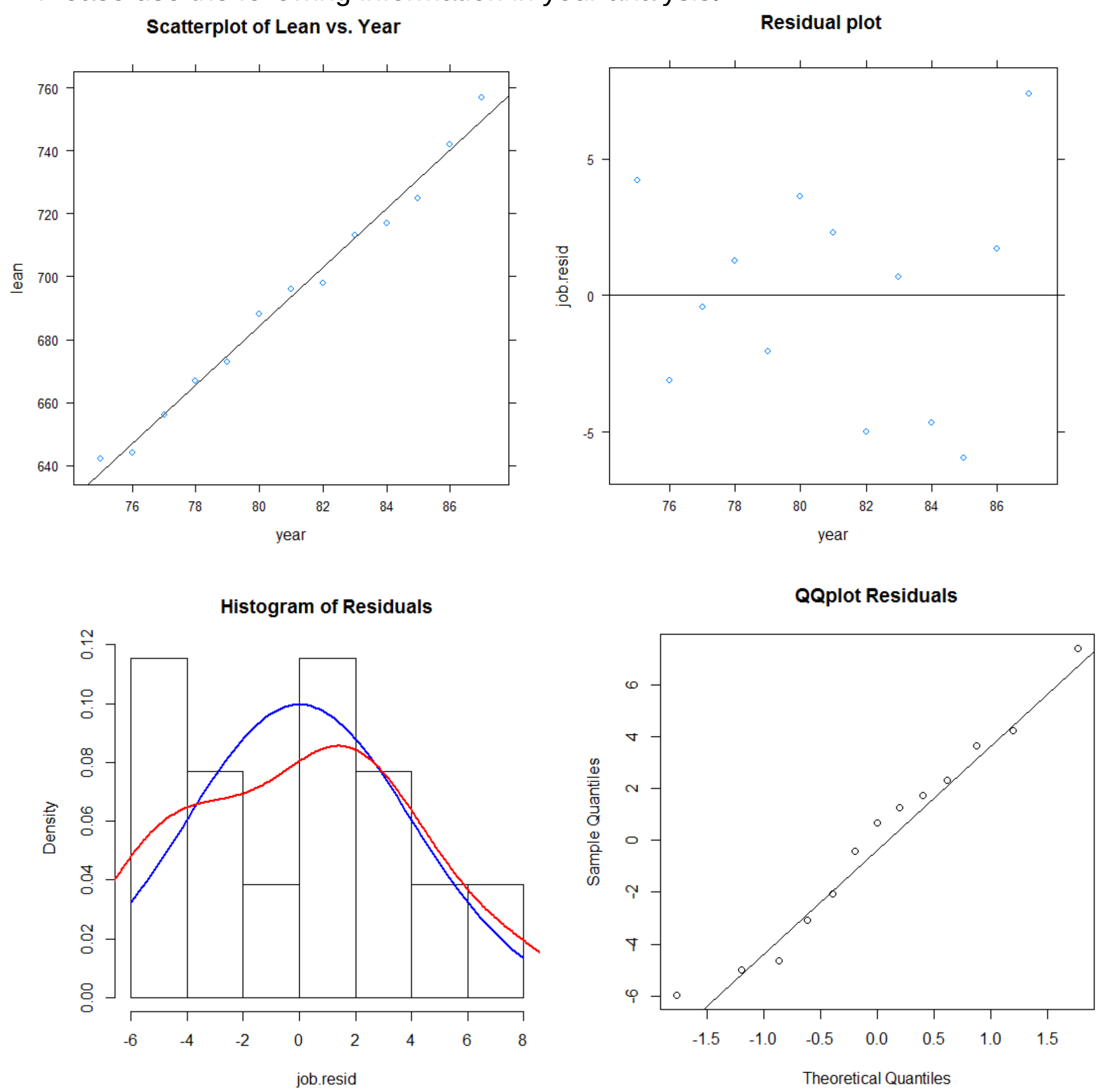
a) Perform a hypothesis test to determine if there is a linear association between the pretest score and the score on the final exam at a 5% significance level.
b) What would the decision be for the test to determine if the slope is positive at the same significance level? Explain your answer. Use the results of a) to answer this question. No hypothesis test is required.

STAT 350 Post Exam 2 Material Review Problems

6. The Leaning Tower of Pisa is an architectural wonder. Engineers concerned about the tower's stability have done extensive studies of its increasing tilt. Measurements of the lean of the tower over time provide much useful information. The following table gives measurements for the years 1975 to 1987. The variable "lean" represents the difference between where a point on the tower would be if the tower were straight and where it actually is. The data are coded as tenths of a millimeter in excess of 2.9 meters, so that the 1975 lean, which was 2.9642 meters, appears in the table as 642. Only the last two digits of the year were entered into the computer.

Year	75	76	77	78	79	80	81	82	83	84	85	86	87
Lean	642	644	656	667	673	688	696	698	713	717	725	742	757

Please use the following information in your analysis.




```
Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    xxxx        xxx       xxxx     0.0333
year           xxxx        xxx       xxxx     6.5e-12
```

$S_{XX} = 182$, $S_{YY} = 15996.77$, $S_{XY} = 1696$, $\bar{x} = 81$, $\bar{y} = 693.69$

a) Does the trend in lean over time appear to be linear? Be sure to indicate which plot(s) you are using.

b) Are the assumptions met? Please explain your answer. Be sure to indicate all plot(s) you are using for each assumption.

Please continue to do the problem no matter what you stated in parts a) and b).

c) What is the equation of the least-squares line? What is the correlation? What percent of the variation in lean is explained by this line?
d) Calculate the ANOVA table for linear regression. What percent of the variation in lean is explained by this line?
e) Using the F test, test whether there is an association between the year and the amount of Lean at a 1% significance level
f) Calculate and interpret the 99% confidence interval for the average rate of change (tenths of a millimeter per year) of the lean.
g) In 1918 the lean was 2.9071 meters. (The coded value for the year is 18, the coded value for the lean is 71.) Using the least-squares equation for the years 1975 to 1987, calculate a predicted value for the lean in 1918.
h) Do you think that the predicted value in part g) is appropriate? Please explain your answer. Use numerical and graphical summaries to support your explanation.
i) The engineers working on the Leaning Tower of Pisa were most interested in how much the tower would lean if no corrective action was taken. Use the least-squares equation to predict the tower's lean in the year 2013. (*Note:* The tower was renovated in 2001 to make sure it does not fall down.) The question implies that the engineers wanted to know what would happen if the linearity was valid in 2013.
j) Calculate and interpret the 99% confidence interval for the average lean in 2013. [In the exam, you will either be given the standard error of the mean at a point and the prediction interval or you will be given the output from both options.]
k) Calculate and interpret the 99% prediction interval for the lean in 2013. [In the exam, you will either be given the standard error of the mean at a point and the prediction interval or you will be given the output from both options.]
l) To give a margin of error for the lean in 2013, would you use a confidence interval for a mean response (j) or a prediction interval (k)? Explain your choice.