**Analytics Primer**

**Quiz 3**

***Honor Pledge:***

*I certify that I have not received or given unauthorized aid in taking this exam.*

*Signed:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Printed Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**There are 15 questions. The first 10 are worth 2 points each. The last 6 are worth different point values. There will be some True/False, Multiple Choice, and Short Answer questions.**

Use the following scatter plot to answer questions 1 and 2.



1. From the scatter plot we say that the two variables have approximately the following relationship:

a. Positive Nonlinear

b. Positive Linear

c. Negative Linear

d. No Relationship

2. What would be the ***best*** estimate of the correlation coefficient between the two variables in the scatter plot?

a. -0.89.

b. -1.00.

c. +0.64.

d. 0.

3. In your own words, describe what correlation is. Do not give an equation, but explain what it means if we say two variables are correlated.

4. List two assumptions for the *simple* linear regression model.

1.

2.

5. In simple linear regression, the sign of the slope coefficient must be the same as the sign of the correlation coefficient between the independent and dependent variables.

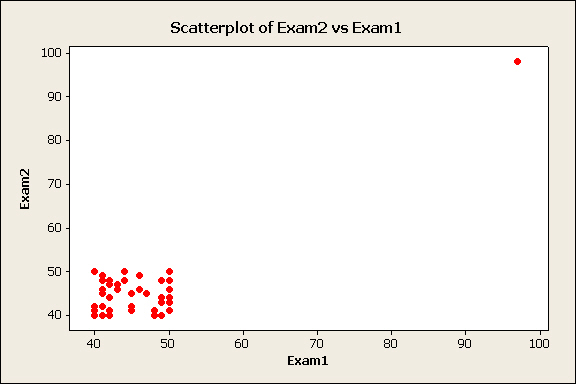
a. True.

b. False.

A teacher of a statistics course recorded the scores of 50 students on the two exams

she gave this semester. She plotted the following scatter plot. Use this to answer

question 6.



**A**

6. There is an outlier in this data (marked with the letter “A”). If we removed this outlier the following would happen to the correlation coefficient:

a. Get closer to zero.

b. Get closer to one.

c. Stay the same.

d. Need more information.

7. A marketing researcher develops two models to try and explain the usage of products at a convenience store. His first model has an R2 = 0.722 and adjusted R2 = 0.703. His second model is the same as the first model with one additional variable. His second model has an R2 = 0.727 and adjusted R2 = 0.691. What can we conclude?

a. Something is calculated incorrectly because the adjusted R2 is always higher than the R2.

b. The second model is probably better because it has a higher R2 value.

c. The additional variable in the second model probably hinders the overall model more than helps it.

d. You cannot compare adjusted R2 values between models.

8. Why do we calculate adjusted R2 values in multiple regression instead of just R2 values?

a. Mathematically R2 values always increase for extra variables added to a model.

b. R2 values only describe a simple linear regression, while adjusted R2 values describe a multiple linear regression.

c. The adjusted R2 only increases if the addition of another variable outweighs the loss of degrees of freedom.

d. Both a and c, but not b.

A data analyst is trying to interpret the following results of a simple linear regression with an R2 = 0.84. Use these results to answer question 9.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Std. Err.** | **DF** | **T-Stat** | **P-Value** |
| Intercept | 22.57 | 3.53 | 48 | 6.394 | <0.0001 |
| Slope | -9.82 | 1.32 | 48 | -7.439 | <0.0001 |

9. What is the value of the correlation coefficient r?

1. 0.917.
2. 0.84.
3. -0.917.
4. Not enough information.

10. If I were to refer to a regression line as the least squares regression line, what is meant by the term “least squares” here?

An analyst in charge of mutual funds for a major financial firm is trying to use 2009

profits (in millions of dollar) from different corporations to help explain the prices of

bonds (in dollars) issued by those same corporations in 2009. The analyst calculated

the following simple regression model using a sample of 76 corporations. Use this information to answer questions 11 and 12.

11. (10 points) Conduct a hypothesis test to see if the slope of the true regression line equals zero with a significance level of 0.05. Do not forget to include your statement of hypothesis, test statistic, p-value, decision rule, and conclusion.

12. (5 points) Calculate a 90% confidence interval for the slope of the regression line in the above problem.

A data analyst for a major car manufacturer is trying to develop a pricing model for the company’s cars. The analyst came up with the following model from a sample of 54 cars. Use this information to answer questions 13 – 15.

13. (5 points) Interpret the coefficient for in terms of the problem.

14. (5 points) Calculate the R2 value and the adjusted R2 value.

15. (5 points) Calculate the F-statistic for overall model significance in this model. Do not compute the entire hypothesis test, just the F-statistic for the hypothesis test of overall model significance.