

Fuentes

Name\_\_\_\_\_

**ST-314 Homework 5**  
**Due Friday, May June 6th at 5:00 PM**

*Please download this document. Then, complete your solutions and upload as PDF or Word Document.  
No other formats will be accepted.*

*Typing or entering answers by hand is accepted as long as solutions are presented neatly and the document is uploaded as PDF. Give the solutions in the space provided.*

**Instructions**

- The homework is due on Friday, June 6th at 5:00 PM. The homework **MUST** be uploaded to Blackboard as PDF or Word document. No other formats will be accepted.
- The homework is worth 25 points.
- Late homeworks will not be accepted under ANY circumstances.
- You will have 3 attempts to successfully upload your homework. Only your last successful attempt will be graded.
- You may work in groups of 2-3 people, but must submit individual solutions.
- You must provide complete answers in order to receive full credit.
- Please, use the space assigned to provide solutions to the problems.
- Failing to follow any of these instructions may result in a deduction of points from your total score.

**Scanners**

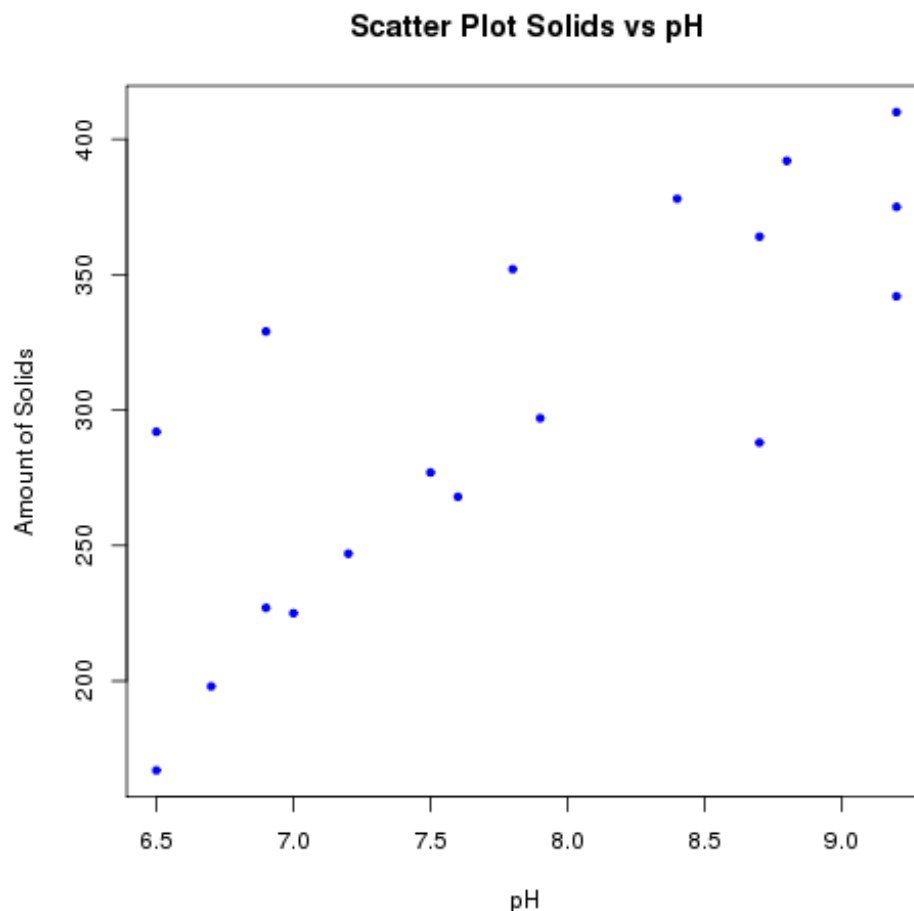
If you need to scan your homework, eScanners are located in the 2nd floor Copy Center, 1st, 3rd and 5th floors. These scanners allow you to scan documents (color, gray, b/w) in searchable PDF or quick PDF and send them to an email address or store on a flash drive. It is recommended that you scan no more than 17 pages at a time to prevent the eScanners from freezing. Your mailbox will also need to have sufficient storage space for your documents or they will be lost when sent. There is no charge for the service at this time.

**Problem 1. (25 Points)**

In this problem you will fit a simple linear regression model. Use the R-code provided with the homework to answer the following questions:

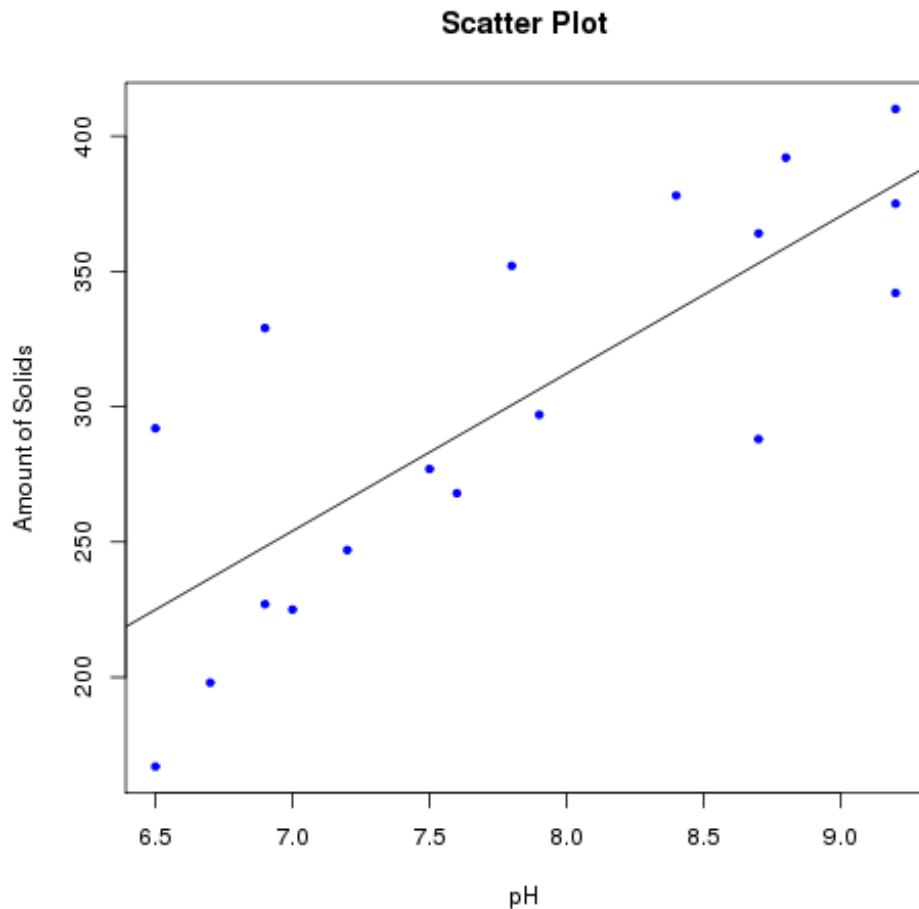
- Consider the data in the file “polymers.csv” and construct a scatter-plot of “solids” vs “pH”. Discuss whether the data follows a linear trend and the direction of the trend.
- Compute the correlation coefficient between these variables. Does the value of the correlation coefficient agree with what you observe in the scatter-plot?
- Fit a simple linear regression model using “solids” as the response variable and “pH” as the explanatory variable. Obtain the values of the estimated coefficients. Based on the R-output, do you think the value of the slope is statistically significant? Explain.
- Obtain a 95% confidence interval for the value of the slope.
- Obtain the ANOVA table for the simple linear regression model fitted in part (c).
- Obtain the value of the estimated regression coefficients by hand and check the coincide with the R-output.

a) The data shows a strong positive linear trend. However, there are several outliers that don't seem to fit the general trend of the data.



b)  $r=0.802115$  This value for the correlation coefficient agrees with the observed trend in the data. 'r' lies close to 1 which indicates a strong positive linear association.

c) By calculating the coefficient of determination ( $r^2$ ) as .6434, this points to the value of the slope being statistically significant. This means that 64.34% of the variation in the response variable can be accounted for by the linear regression of the independent variable.



d) 35.22525 81.13802

e)

Source	df	Sum of Squares	Mean Squared
pH	1	54856	54856
residuals	16	30405	1900
Total	17		

f) given:  $b_1 = r(S_y/S_x)$ ,  $r=0.802115$   $S_x(\text{from } R)=0.976$   $S_y(\text{From } R)=70.8189$

Fuentes  
 $b_1 = 58.202$

Name \_\_\_\_\_ Jared Wasinger \_\_\_\_\_

given:  $b_0 = \bar{Y} - b_1(\bar{X})$ ,  $\bar{Y}(\text{from } R) = 301.556$   $b_1 = 58.202$   $\bar{X}(\text{from } R) = 7.18667$

$b_0 = -144.446$

This values are consistent with values given by R. However,  $b_0$  is slightly off from the R value.