

(Practice) Regression Problem 1:

Consider the following regression output:

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	4.6755	0.3404	13.735
x	-0.4099	0.3958	-1.036

on 11 degrees of freedom

- Using the t-table provided. Test the null hypothesis that the slope is zero using the significance level $\alpha = 0.01$. Explain your steps and include a conclusion.

- Using the t-table provided. Find the 95% confidence interval of the slope.

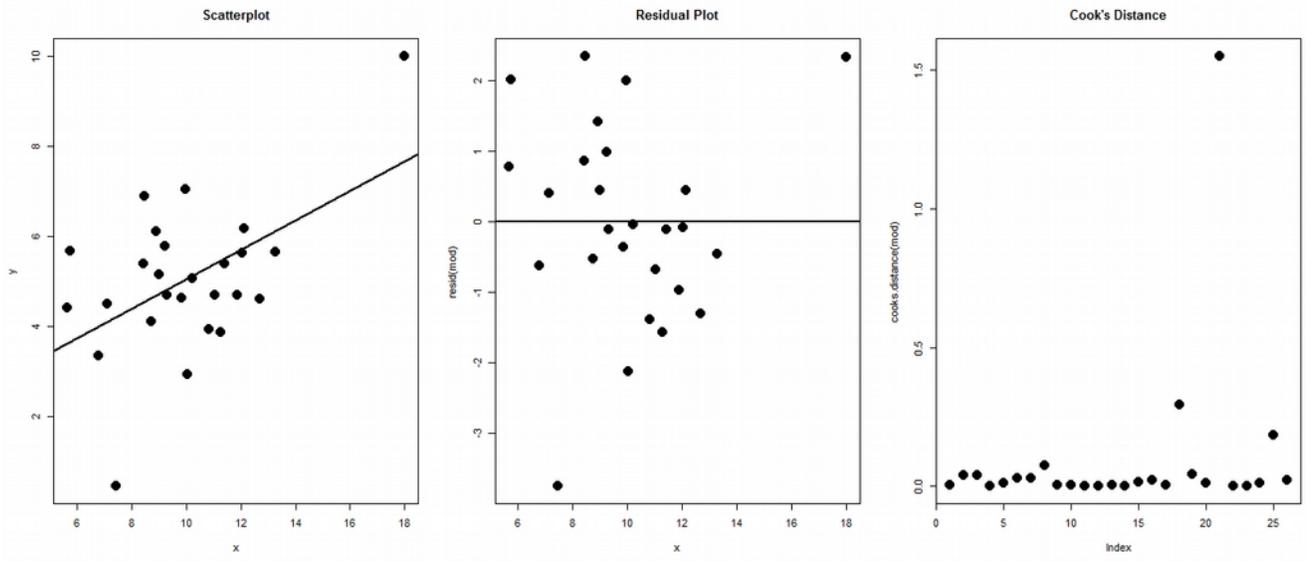
The Pearson correlation coefficient between these two variables is -0.298, how much of the variance in y is explained by x?

(Practice) Diagnostics:

Here is a scatterplot, residual plot, and Cook's distance plot from a regression.

a) **(3 pts)** Identify a potential problem, and explain how you found it.

b) **(2 pts)** Suggest what could be done to better model the relationship between X and Y?



(Practice) Regression:

Consider the following regression output:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11.4580	1.2753	8.984	1.16e-13	***
x	2.6957	0.1776	15.180	< 2e-16	***

Write the regression equation.

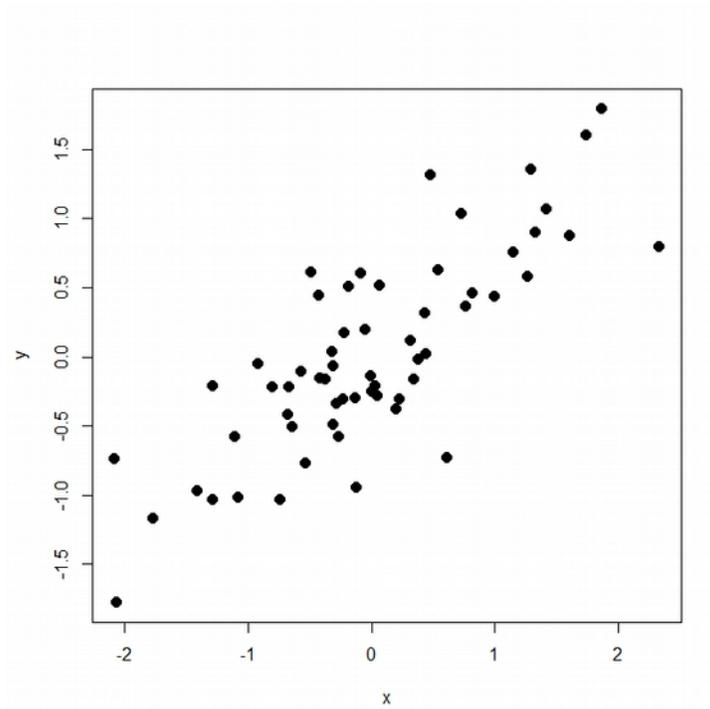
A new observation with a value $x = 4.0$, using this regression equation, what is the best estimate for the prediction of y for this observation?

$y = 44$ is outside the prediction interval for this y value, which is (2.51 , 41.96).

Is it outside the confidence interval for the mean of Y values, inside the interval, or is it impossible to tell? Explain.

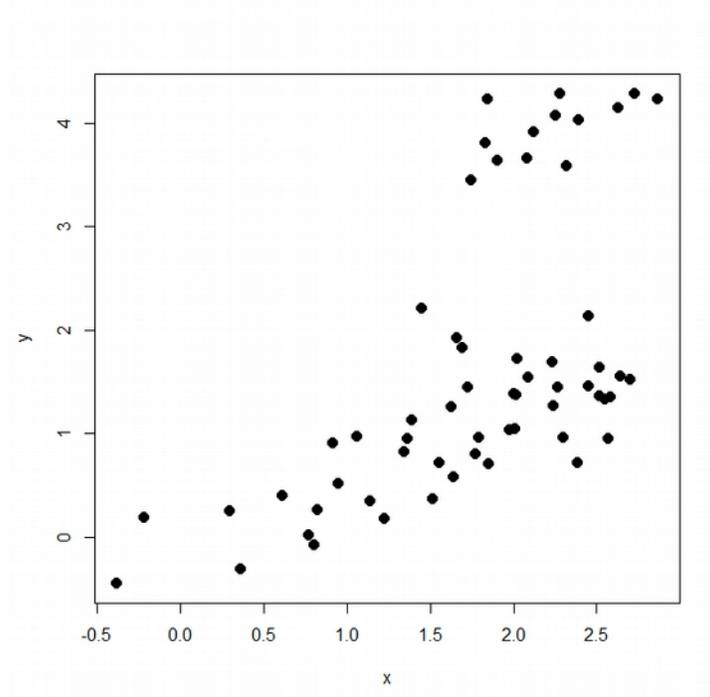
(Practice) Correlation (Without any problems, about +0.7):

- Estimate the Pearson correlation between these variables. Anywhere within 0.2 of the real answer is acceptable.



(Practice) Correlation (Curved, about +0.4 pearson, 0.7 spearman Practice):

- The Pearson correlation coefficient between these two variables is 0.592, would the Spearman correlation be stronger, weaker, or about the same? Justify.



(Practice) One-Way ANOVA:

- From this ANOVA, is there sufficient evidence, at the $\alpha = 0.05$ significance level, to suggest that the five means are different? Why or why not?

Response: y

```
          Df Sum Sq Mean Sq F value Pr(>F)
group    4    6175  1543.81   2.0916 0.0943 .
Residuals 40595    738.09
```

Group	N	Mean	SD
Blue	4	9.67	13.19
Green	11	41.96	20.89
Yellow	9	27.68	12.98
Orange	31	25.08	33.15
Red	5	4.76	18.55

From the information provided, how many degrees of freedom remain for residuals?

Are there any features of this data that may interfere with the anova results and your conclusion?

A T-test comparing the means of the Green and Red groups shows strong evidence of a difference. How could this happen with the p-value from the above ANOVA, which does not show any differences?

Welch Two Sample t-test

```
data: y[which(group == "green")] and y[group == "red"]
t = 3.5703, df = 8.7714, p-value = 0.006276
```

Simplified T-Table

(1 tail)	0.025	0.005
(2 tail)	0.05	0.01
df		
1	12.7065	63.6551
2	4.3026	9.9247
3	3.1824	5.8408
4	2.7764	4.6041
5	2.5706	4.0322
6	2.4469	3.7074
7	2.3646	3.4995
8	2.306	3.3554
9	2.2621	3.2498
10	2.2282	3.1693
11	2.201	3.0158
12	2.1788	3.0545
13	2.1604	3.0123
14	2.1448	2.9768
15	2.1314	2.9467
16	2.1199	2.9208
17	2.1098	2.8983
18	2.1009	2.8784
19	2.093	2.8609
20	2.086	2.854