**Name: Yuanqiu Feng**

**UmichID: yuanqiu**

What dataset are you working with: pulitzer

List 3 questions that you can ask with your dataset.

Q1: Does winning Pulitzer prizes affect newspaper readership?

Q2: Does winning more Pulitzer prizes increase readership for newspapers?

List the associated null hypothesis for each question:

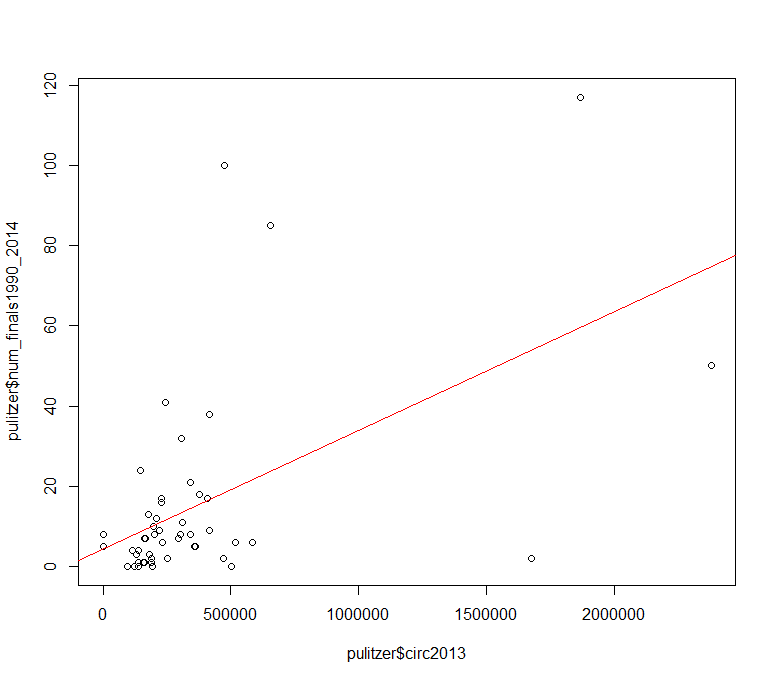
Q1: Winning Pulitzer prizes have no relationship with newspaper readership.

Q2: Winning more Pulitzer prizes does not increase readership for newspapers.

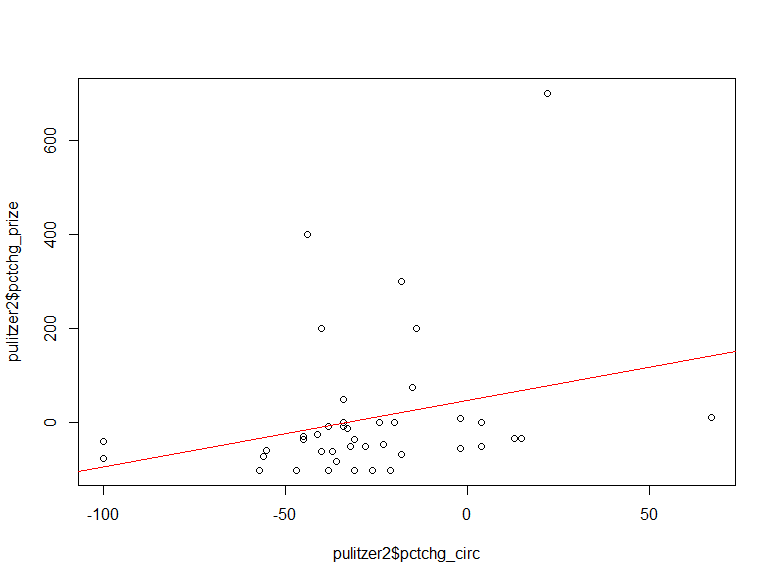
What statistical test(s) will you use to answer each of the questions:

Q1: Regression

Q2: Regression

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square). 

Q1: Total number of Pulitzer prizes won and newspaper circulation in 2013.



Q2: Percentage change in prizes won between 1990-2003 and 2004-2014 and percentage change in circulation between 2003 and 2014.

Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

Q1: Homoscedasticity, (No, BP test p-value = 0.001711)

statistical independence of errors (No, DW test p-value = 0.005545),

normality or error distribution (No, Shapiro test p-value = 9.158e-08)

Q2: Homoscedasticity, (Yes, BP test p-value = 0.4368)

statistical independence of errors (Yes, DW test p-value = 0.8222),

normality or error distribution (No, Shapiro test p-value = 0.007555)

Run the statistical test! Put your results here:

Q1: Call:

lm(formula = num\_finals1990\_2014 ~ circ2013, data = pulitzer)

Residuals:

Min 1Q Median 3Q Max

-51.903 -8.136 -5.192 2.232 81.615

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.328e+00 3.884e+00 1.114 0.271

circ2013 2.961e-05 6.850e-06 4.322 7.74e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.12 on 48 degrees of freedom

Multiple R-squared: 0.2802, Adjusted R-squared: 0.2652

F-statistic: 18.68 on 1 and 48 DF, p-value: 7.744e-05

Q2: Call:

lm(formula = pulitzer2$pctchg\_circ ~ pulitzer2$pctchg\_prize,

data = pulitzer2)

Residuals:

Min 1Q Median 3Q Max

-69.827 -11.597 -3.679 9.266 94.465

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -28.11876 4.45214 -6.316 1.89e-07 \*\*\*

pulitzer2$pctchg\_prize 0.05137 0.02938 1.748 0.0883 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.46 on 39 degrees of freedom

Multiple R-squared: 0.07268, Adjusted R-squared: 0.0489

F-statistic: 3.057 on 1 and 39 DF, p-value: 0.08828

Interpret your results!

Q1: Winning Pulitzer prizes

Q2: Winning more Pulitzer prizes did not significantly help to increase readership in newspapers.