**Name: Ho Hsieh**

**UmichID: 06998691**

**What dataset are you working with:** fandango and pulitzer

**List 3 questions that you can ask with your dataset.**

Q1: Is higher number of Pulitzer winners or finalists from 2004 to 2013 associate with higher increase in daily circulation of newspapers from 2004-2014?

Q2: Is the uses’ rating in fandango significantly different from the website’s (fandango’s) rating?

Q3: Is the Daily Circulation of all newspapers in 1990-2003 different from Daily Circulation in 2004-2014?

**List the associated null hypothesis for each question:**

Q1: The number of Pulitzer winners or finalists from 2004 to 2013 does not associate with the increase in daily circulation of newspapers from 2004-2014.

Q2: The users’ rating and the website’s rating in fandango is not significantly different.  
Q3: The Daily Circulation of all newspapers in 1990-2003 is not different from Daily Circulation in 2004-2014.

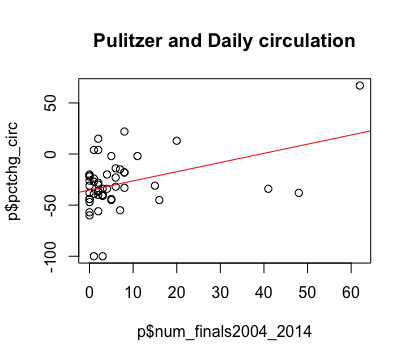
**What statistical test(s) will you use to answer each of the questions:**

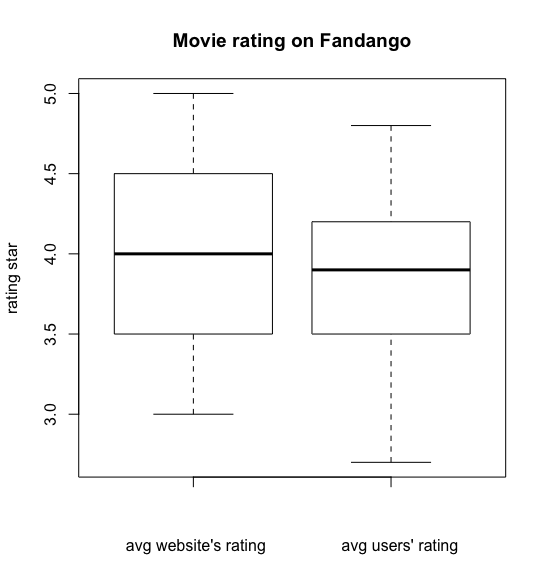
Q1: regression

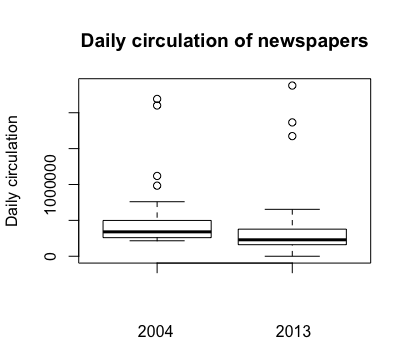
Q2: paired Welch’s t-test

Q3: Welch’s t test

**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:

Q2:

Q3:

**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:

* Linear relationship: yes

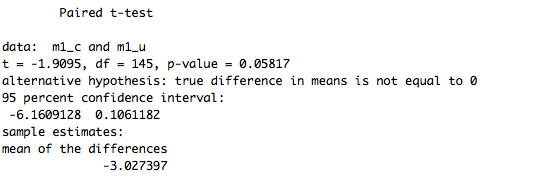
Q2:

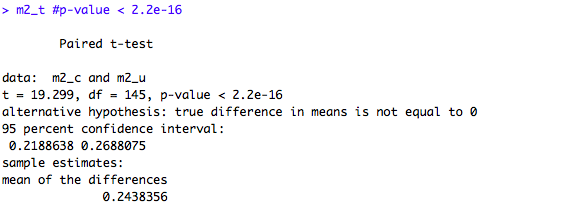
* Random sample: not known
* Independent observation: not known
* Normal distribution or sample size > 30: The data are not normally distributed, but the sample size is 146.
* Equal variance: unequal variance

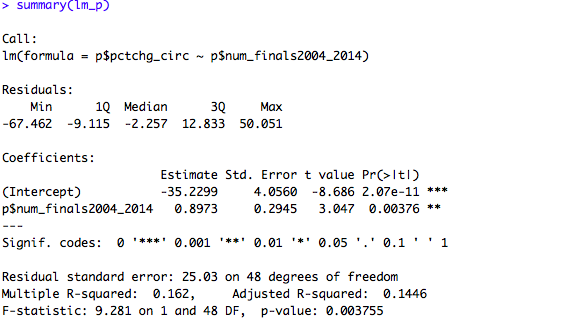
Q3:

* Random sample: not known
* Independent observation: not known
* Normal distribution or sample size > 30: The data are not normally distributed, but the sample size is 50.
* Equal variance: unequal variance

**Run the statistical test! Put your results here:**

Q1: p-value: 0.003755

Q2: p-value < 2.2e-16

Q3: p-value = 0.236

**Interpret your results!**

Q1: p-value<<0.05 which means we can reject null hypothesis and the number of Pulitzer winners or finalists from 2004 to 2013 is positively associate with the increase in daily circulation of newspapers from 2004-2014.

Q2: p-value<<0.05 which means we can reject null hypothesis and the uses’ rating in fandango is significantly different from the website’s (fandango’s) rating.

Q3: p-value>0.05 which means we cannot reject null hypothesis and the Daily Circulation of all newspapers in 1990-2003 is not different from Daily Circulation in 2004-2014.