Name: Yilun Zhao

UmichID: 53756840 (zyilun)

What dataset are you working with: airline\_safety

List 3 questions that you can ask with your dataset.

Q1: Is total number of fatal accidents from 1985-1999 equal to total number of fatal accidents from 2000-2014

Q2: Is available seat kilometers flown every week related to total number of incidents from 2000 -2014.

Q3: Is the number of incidents from 1985-1999 equal to the number of incidents from 2000-2014

List the associated null hypothesis for each question:

Q1: total number of fatal accidents from 1985-1999 equal to total number of fatal accidents from 2000-2014

Q2: There is no relationship between available seat kilometers flown every week and total number of incidents from 2000 -2014.  
Q3: The number of incidents from 1985-1999 is equal to the number of incidents from 2000-2014

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

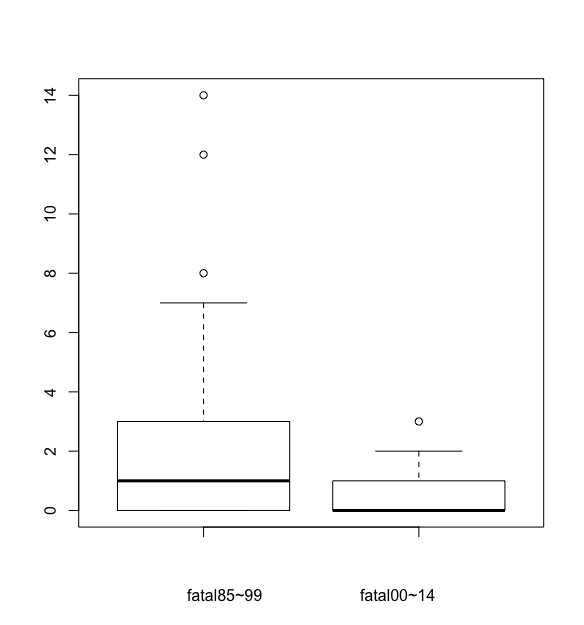
Q1: t-test

Q2: linear regression

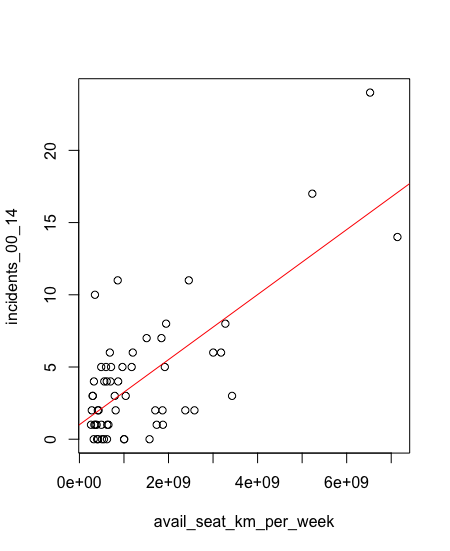
Q3: t-test

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).

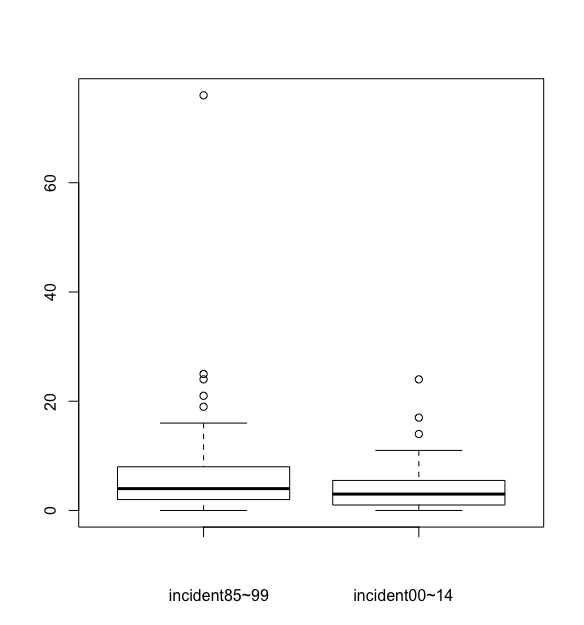
Q1:



Q2:



Q3:



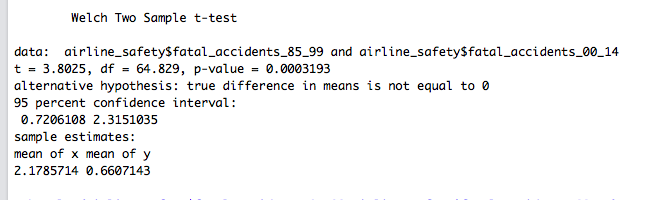
Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

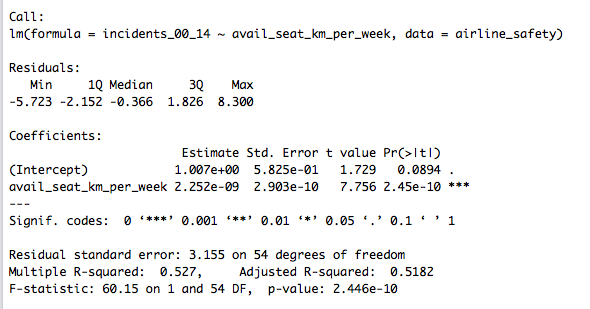
Q1: random sampled, independent observation, sample size is larger than 30 or this is a normal distribution, equal sample size, unequal variance. Meet the assumption.

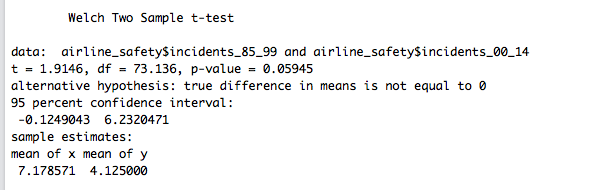
Q2: meet linear regression requests

Q3: random sampled, independent observation, sample size is larger than 30 or this is a normal distribution, equal sample size, unequal variance. Meet the assumption.

Run the statistical test! Put your results here:

Q1: 

Q2: 

Q3: 

Interpret your results!

Q1: p-value is lower than 0.05, so the number of fatal accidents from 1985 to 1999 is significantly different from the number of fatal accidents from 2000 to 2014.

Q2: The p-value is lower than 0.05, so the incidents number from 2000 to 2014 is positively linear related to available seat kilometers flown every week.

Q3: p-value is higher than 0.05, so the number of incidents from 1985 to 1999 is not significantly different from the number of incidents from 2000 to 2014.