Name: Andrew Kinzer

UmichID: akinzer

What dataset are you working with: bad\_drivers

List 3 questions that you can ask with your dataset.

Q1: Does the number of drivers involved in fatal collisions effect the insurance premiums

Q2:Is the percentage of drivers who were speeding and involved in fatal accidents related to the number of drivers who were drunk and involved in fatal accidents?

Q3:

List the associated null hypothesis for each question:

Q1:There is no relationship between the number of drivers involved in fatal collisions and the insurance premiums

Q2: There is no relationship in the percentage of drivers involved in fatal accidetns who were speeding and who were drunk  
Q3:

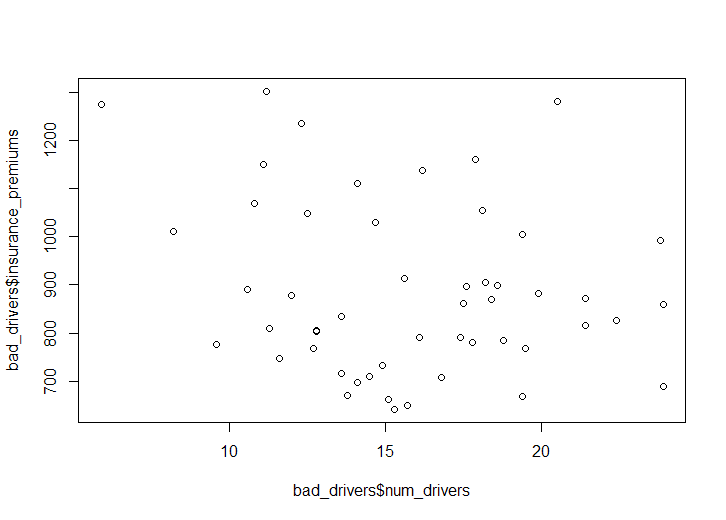
What statistical test(s) will you use to answer each of the questions:

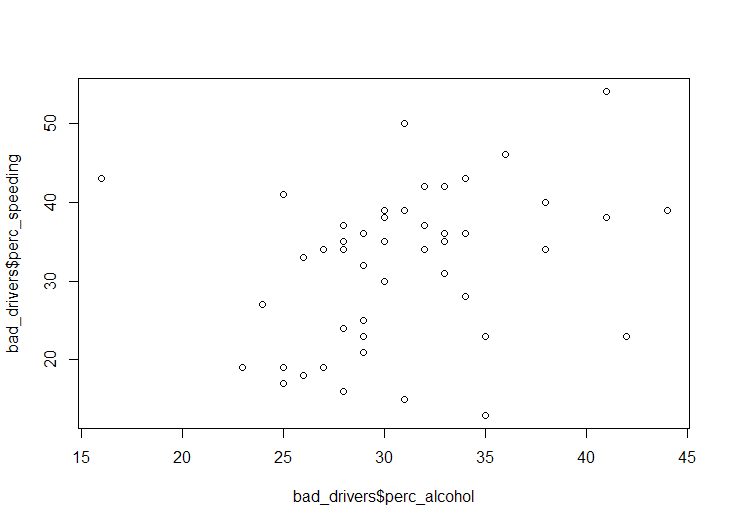
Q1: linear regression

Q2:linear regression

Q3:

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: The data does meet the assumptions required. The residuals are mostly normally distributed, and the qqplot shows the normalized residuals fall along the normal distribution line

Q2:The data does meet the assumptions required. They residuals are mostly normally distributed, and the qqplot shows the normalized residuals fall along the normal distribution line.

Q3:

Run the statistical test! Put your results here:

Q1:

Call:

lm(formula = insurance\_premiums ~ num\_drivers, data = bad\_drivers)

Residuals:

Min 1Q Median 3Q Max

-249.23 -136.43 -22.29 133.45 435.28

Coefficients:

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

(Intercept) 1023.354 98.748 10.363 6.08e-14 \*\*\*

num\_drivers -8.638 6.055 -1.427 0.16

---

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

Residual standard error: 176.5 on 49 degrees of freedom

Multiple R-squared: 0.03988, Adjusted R-squared: 0.02029

F-statistic: 2.035 on 1 and 49 DF, p-value: 0.16

Q2: Call:

lm(formula = perc\_speeding ~ perc\_alcohol, data = bad\_drivers)

Residuals:

Min 1Q Median 3Q Max

-21.043 -8.208 1.569 5.912 19.165

Coefficients:

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

(Intercept) 15.2379 7.9916 1.907 0.0624 .

perc\_alcohol 0.5373 0.2569 2.091 0.0417 \*

---

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

Residual standard error: 9.324 on 49 degrees of freedom

Multiple R-squared: 0.08194, Adjusted R-squared: 0.0632

F-statistic: 4.373 on 1 and 49 DF, p-value: 0.04172

Q3:

Interpret your results!

Q1: There is a significant negative correlation between the number of fatalities in accidents and the cost of car insurance. However, while the relationship is significant (p-value=1.08e-16) the adjusted r-square value of 0.02 indicates that only 2% of the relationship between the insurance premium and the number of drivers killed in car accidents is explained by the model.

Q2:There is a significant positive correlation between the percentage of drunk people involved in fatal accidents and the percentage of speeding people involved in fatal accidents. However, while the relationship is significant (p-value =0.042) the adjusted r-square value of 0.06 indicates that only 6% of the relationship is explained by the model.

Q3: