Supplement for Lecture 10: Partitioning Variability

Load Data

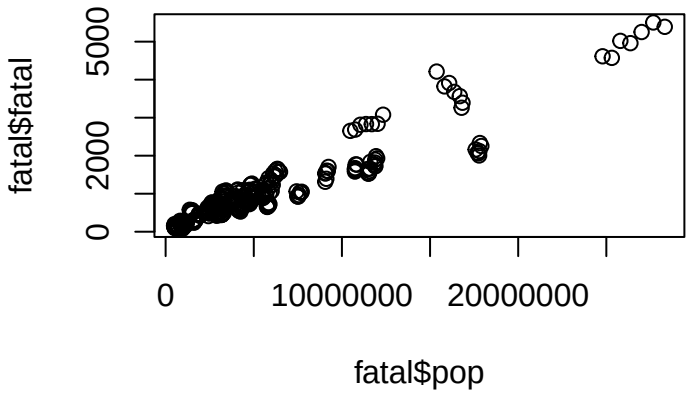
```
data("Fatalities") # Load Data

fatal = Fatalities[,c("fatal","pop","youngdrivers")]

head(fatal)
```

```
##   fatal     pop youngdrivers
## 1   839 3942002     0.211572
## 2   930 3960008     0.210768
## 3   932 3988992     0.211484
## 4   882 4021008     0.211140
## 5  1081 4049994     0.213400
## 6  1110 4082999     0.215527
```

Variables of Interest - *fatal* = Number of vehicle fatalities - *pop* = Population - *youngdrivers* = Percent of Drivers 15 - 24

Create New Variable to Adjust for Population

```
#Consider scatterplot
plot(x=fatal$pop,y=fatal$fatal)
```

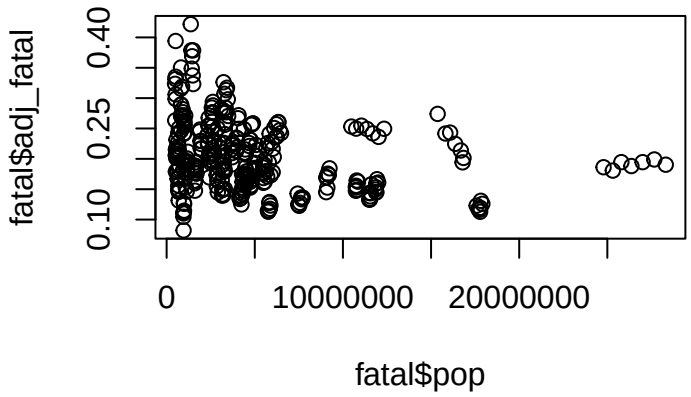



```
#Create New Variable Called adj_fatal
fatal$adj_fatal = (fatal$fatal/fatal$pop)*1000
```

```
#Remove Original Variable
fatal$fatal = NULL

#Preview Modified Dataset
head(fatal)
```

```
##       pop youngdrivers adj_fatal
## 1 3942002     0.211572  0.212836
## 2 3960008     0.210768  0.234848
## 3 3988992     0.211484  0.233643
## 4 4021008     0.211140  0.219348
## 5 4049994     0.213400  0.266914
## 6 4082999     0.215527  0.271859
```

```
#Consider new scatterplot
plot(x=fatal$pop,y=fatal$adj_fatal)
```



Output from Simple Linear Regression

```
#Model for the relationship between fatalities and proportion of young drivers.

mod = lm(adj_fatal~youngdrivers,data=fatal)
summary(mod)

#Create new variable for youngdrivers to help interpretation of slope


#Manually calculate p-value using the t-distribution
   #Find area to right and multiply by 2

#We have found significance. Hooray!!. Let's visualize the model.
```

Focus on the "t value" and "Pr(>|t|)". These are your test statistics and p-values for testing the following hypotheses:

$$H_0 : \beta_x = 0$$

$$H_a : \beta_x \neq 0$$

In class, we focused on when $x = 1$. But we could do the same test for the intercept when $x = 0$.

Confidence Interval for the Slope (and Intercept)

Interpretation of the confidence interal: I am 95 percent confident, that the (average/expected/predicted) number of vehicle fatalities (per 1000) will increase by a number between BLANK1 and BLANK2 for every 1 percent increase in the percent of young drivers.

ANOVA

```
#Manually find the p-value and check it matches
   #Want the area to the right of 19.422
```

Notice how the p-value for the F-test is identical to the p-value from the t-test. Notice how this p-value is in the output for `summary()`. Also, the last row for the *Total* is not there.

```
#Hand Calculation of SST

#Notice that this equals the sum from the ANOVA table
0.0598+1.02930
```