GK-12 Lesson Plan

Teacher: Steven MacDonald
Period: Class: Lawrence High School Statistics Class
Date(s): 11/6/2013

SETTING THE STAGE	
Essential Question	How can outliers confuse our results?
Content Objective(s) (Student-friendly)	Students explore Anscombe's Quartet, a dataset used to illustrate the effect of large outliers on a data set, as well as the importance of graphing a dataset.
Connection to previous or future lessons	Students use R, along with the statistical concepts introduced in previous classes.
Critical Thinking Questions	Why do all four datasets have the same summary statistics?
Key Vocabulary	Correlation, mean, standard deviation, scatter plot, outlier, linear regression.
Materials Needed/Safety	Computer, R Studio
ACTIVE INSTRUCTION	
Launch (Engage)	Anscombe's quartet is a set of four datasets that all have the same mean, variance, correlation, and linear regression.
Investigation (Explore)	Students import the dataset, plot it, and then apply their programming/statistical knowledge to analyze the dataset. The students also perform a linear regression on each dataset.
TIME FOR REFLECTION	
Summarization (Explain & Extend)	Students find that all four datasets seem to have the same values, despite the wide difference in actual structure of the data. This illustrates the importance of plotting data before blindly applying statisical measures.
Homework	None

```
par(mfrow= c(2,2)) #Allows display of 4 plots at once

quartet<-read.csv("anscombes_quartet.csv", header = FALSE) #Reads the dataset into a single
#Following section divides the quartet up into individual columns. Number of the variable represents each
indivual set.
x1<-as.numeric(quartet$V1)
y1<-quartet$V2
x2<-as.numeric(quartet$V3)
y2<-quartet$V4
x3<-as.numeric(quartet$V5)
y3<-quartet$V6
x4<-as.numeric(quartet$V7)
y4<-quartet$V8

plot(x1,y1,xlim=c(4,20),ylim=c(4,13), main="1st Set")
mean(x1)
var(x1)
var(y1)
mean(y1)
cor(x1,y1)
reg1<-lm(y1~x1)
#slope-intercept formula : y = b0 + b1*x
b1<-(cor(x1,y1)*sd(y1))/sd(x1)
b0 <- mean(y1)- b1*mean(x1)
abline(a=b0, b=b1,col="Red") # a = intercept, b = slope


plot(x2,y2,xlim=c(4,20),ylim=c(4,13), main="2nd Set")
mean(x2)
var(x2)
var(y2)
mean(y2)
cor(x2,y2)
reg2<-lm(y2~x2)
abline(reg2,col="Green")

plot(x3,y3,xlim=c(4,20),ylim=c(4,13), main="3rd Set")
mean(x3)
var(x3)
var(y3)
mean(y3)
cor(x3,y3)
reg3<-lm(y3~x3)
abline(reg3,col="Blue")

plot(x4,y4,xlim=c(4,20),ylim=c(4,13), main="4th Set")
mean(x4)
```

```
var(x4)
var(y4)
mean(y4)
cor(x4,y4)
reg4<-lm(y4~x4)
abline(reg4,col="Purple")
```