# Data 609 Week 2

*Jason Joseph*

*September 6, 2016*

## PG 68 # 12

Which variables are the most important?

A company with a fleet of trucks faces increasing maintenance costs as the age and mileage of the trucks increase.

1. Distance covered by each truck.
2. The type of terrain each truck traverse.
3. The age of each truck. Also if a truck should be replaced at a certain age or mileage.

## PG 79 # 11

Determine whether the data set supports the stated proportionality model

```
y <-c(0,1,2,6,14,24,37,58,82,114)
x <- c(1,2,3,4,5,6,7,8,9,10)

data <- data.frame(y,x)
data$x_cube <- x^3

ggplot(data,aes(x=y,y=x_cube)) + geom_point(shape=1)+ geom_smooth(method =lm,se=FALSE)+xlab("Y values")
  ylab(expression(X^3)) +
  ggtitle(expression(paste("Plot of y vs.",x^3,sep="")))
```


### Estimating k

Estimating the slope by choosing two points (24,216) and (58,512). Therefore,

$$k = \frac{512 - 216}{58 - 24} \approx 8.70$$

The data set supports the proportionality model since the plot is approximately linear, with the line approximately passing through the origin.

### PG 94 #4

Lumber cutters wish to use readily available measurements to estimate the number of board feet of lumber in a tree. Assume they measure the diameter of the tree in inches at waist height. Develop a model that predicts board feet as a function of diameter in inches.

The variable x is the diameter of a ponderosa pine in inches, and y is the number of board feet divided by 10.

a) Assume that all trees are right-circular cylinders and are approximately the same height.

The volume of a right-circular cylinder is given as

$$V = \pi r^2 h$$

Since all of the trees are of the same height we can disregard h. Our model becomes

$$V = k\pi r^2 = kr^2$$

. Where

$$r = 0.5x$$

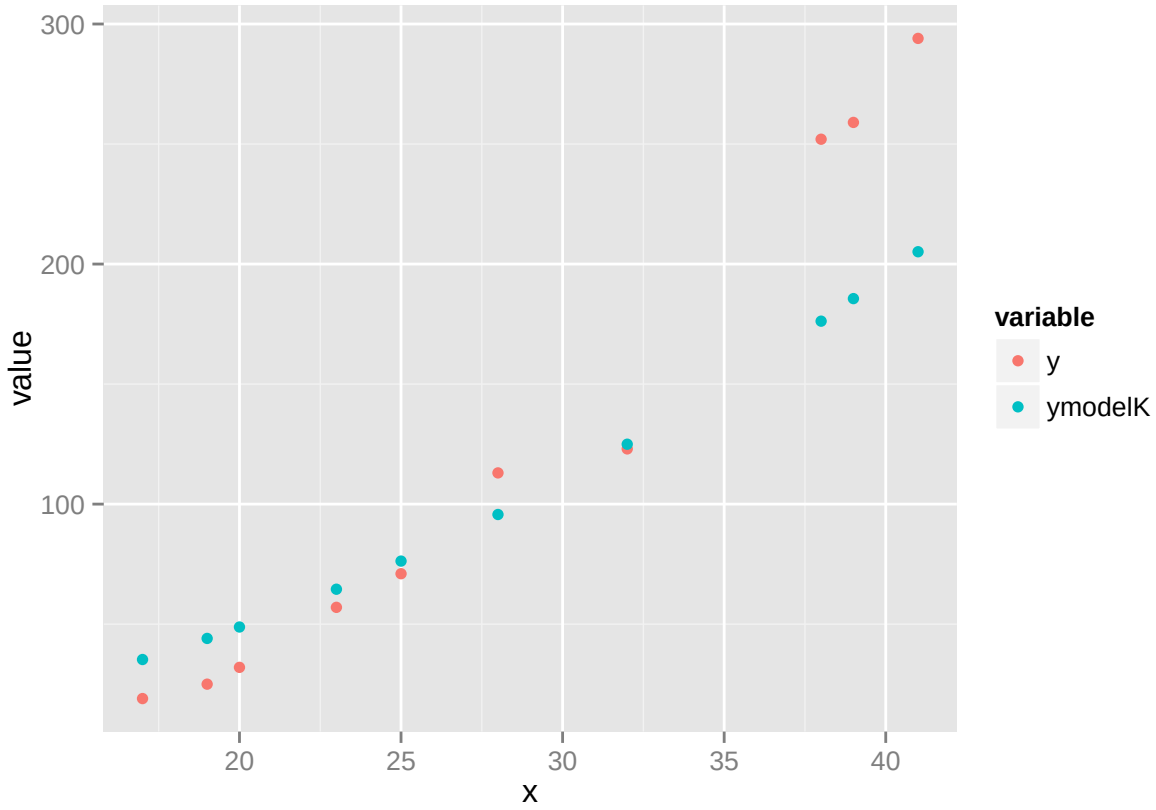
Therefore:

$$y \approx (0.5x)^2$$

```
# diameter
x <- c(17,19,20,23,25,28,32,38,39,41)

# number of board feet
y <- c(19,25,32,57,71,113,123,252,259,294)

data <- data.frame(x,y)
data$ymodel <- (0.5*data$x)^2

#calculating K
k <- mean(data$y/data$ymodel)
data$ymodelK <- k *(data$ymodel)
data
```

```
##     x   y ymodel   ymodelK
## 1  17  19  72.25  35.26639
## 2  19  25  90.25  44.05248
## 3  20  32 100.00  48.81161
## 4  23  57 132.25  64.55335
## 5  25  71 156.25  76.26814
## 6  28 113 196.00  95.67075
## 7  32 123 256.00 124.95771
## 8  38 252 361.00 176.20990
## 9  39 259 380.25 185.60614
## 10 41 294 420.25 205.13078
```

```
#plotting graph
data.long <- melt(data, id=c("x","ymodel"))
ggplot(data.long,aes(x=x, y=value, color=variable))+geom_point()
```




b) Assume that all trees are right-circular cylinders and that the height of the tree is proportional to the diameter.

As a result

$$h \approx k_2 x$$

Threfore

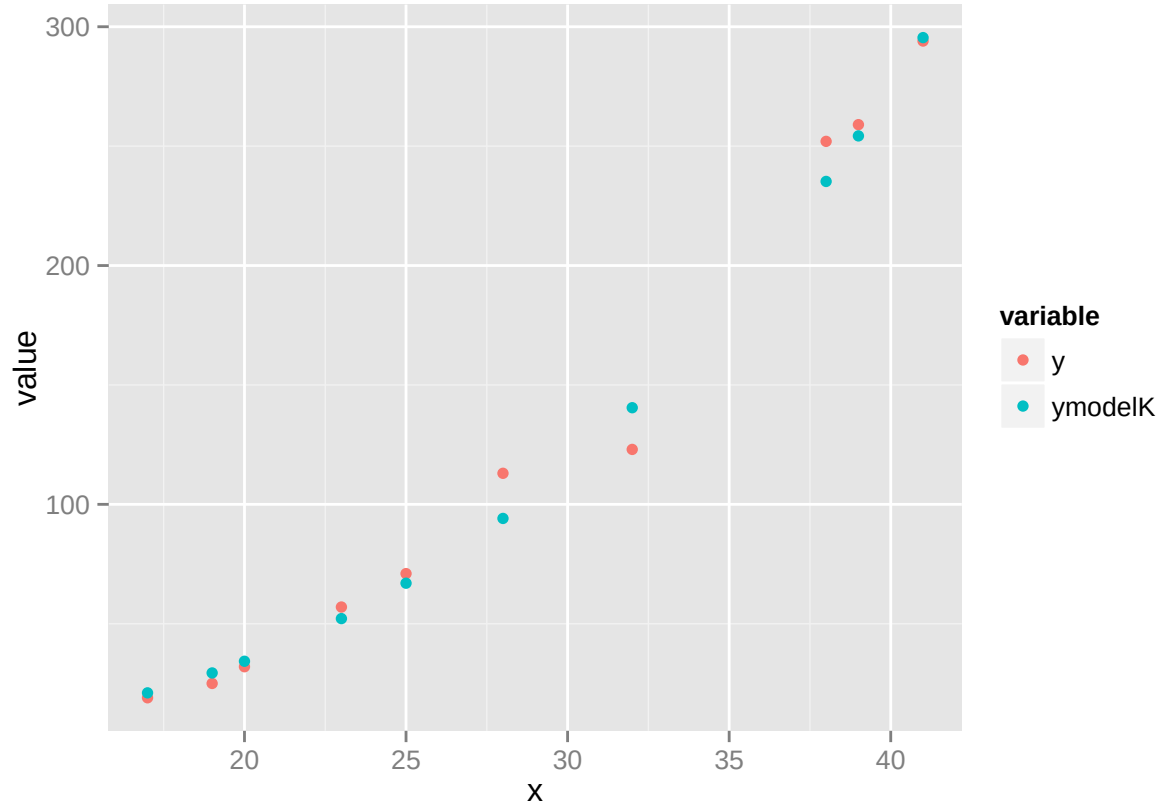
$$y \approx (0.5x)^3$$

```
# diameter
x <- c(17,19,20,23,25,28,32,38,39,41)

# number of board feet
y <- c(19,25,32,57,71,113,123,252,259,294)

data <- data.frame(x,y)
data$ymodel <- (0.5*data$x)^3

#calculating K
k <- mean(data$y/data$ymodel)
data$ymodelK <- k *(data$ymodel)
data
```

```
##     x   y   ymodel   ymodelK
## 1  17  19  614.125  21.06040
```

```
## 2  19  25  857.375  29.40226
## 3  20  32 1000.000  34.29334
## 4  23  57 1520.875  52.15589
## 5  25  71 1953.125  66.97919
## 6  28 113 2744.000  94.10094
## 7  32 123 4096.000 140.46554
## 8  38 252 6859.000 235.21805
## 9  39 259 7414.875 254.28086
## 10 41 294 8615.125 295.44145
```

```
#plotting graph
data.long <- melt(data, id=c("x","ymodel"))
ggplot(data.long,aes(x=x, y=value, color=variable))+geom_point()
```




c) The second model appears to be a better fit visually. This may be the case since it is highly unlikely that all the trees were of the same height , which was the assumption of the first model. The second model assumption that the height of the tree is proportional its trunk produced a better model.

## PG 99 #3

Discuss several factors that were completely ignored in our analysis of the gasoline mileage problem.

The analysis of the gasoline mileage problem ignored the weight of the vehicle, the number of persons in the vehicle could change during the trip. The vehicle could have been towing a load which would play an important factor in the amount of energy needed to propel the vehicle. The displacement of the vehicle's

engine could be a contributing factor since vehicles with larger engine displacement may utilize more fuel. The year the vehicle was manufactured could have been considered since newer cars are considered more efficient.

## PG 104 #2