IS 609 Week 2 hw

Charley Ferrari

Friday, September 04, 2015

Page 69, Problem 12

For the scenarios presented in problem 12, identify a problem worth studying and list the variables that affect the behavior you have identified. Which variables would be neglected completely? Which might be considered as constants initially? Can you identify any submodels you would want to study in detail? Identify any data you would want collected.

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

Problem worth studying: How can we minimize the total costs of our fleet. Total costs include both maintenance costs, and the costs of replacing trucks.

Variables: Number of trucks in the fleet, Replacement cost of a truck, total amount of work required of the fleet. variables for each truck: age, number of hours of planned use, mileage variables to be neglected: Perhaps we could neglect the type of work. As long as there is a full load, what matters is how long the truck is on the road, and not where it is going.

Which might be considered as constants initially? Can you identify any submodels you would want to study in detail?

Initially, we can start with a submodel assuming that the fleet of trucks remain constant. Our model at this point is simply modeling the increasing costs of maintaining an aging fleet of vehicles, and trying to minimize that cost. The dependent variables would be the number of planned work hours for each truck. The total amount of work required of the fleet would be a constraint, and work would be allocated among trucks of varying age and mileage.

We can also introduce a longer term model allowing for the replacement of trucks. At some point, if maintenance costs are growing without bound, it will make more sense to buy a new truck with lower maintenance costs than to continue spending money on an old truck.

Data I would want collected: I would like to collect data on the maintenance spending for each truck. Using this information, I could constantly update my model, and tweak my relationship between age, mileage and expected maintenance costs.

Page 79, Problem 11

In problems 7-12, determine whether the data set supports the stated proportionality model.

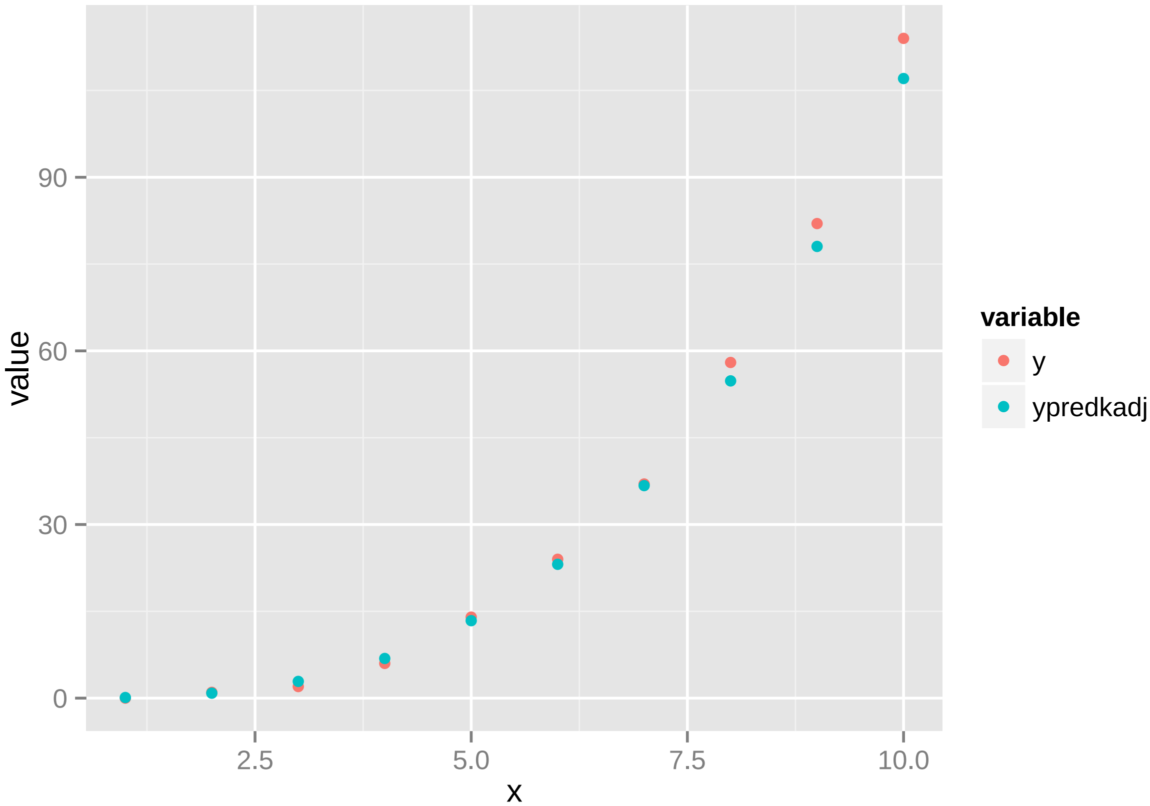
$$y \propto x^3$$

```r
q11data <- data.frame(y=c(0,1,2,6,14,24,37,58,82,114), x=c(1,2,3,4,5,6,7,8,9,10))

q11data$ypred <- q11data$x^3

q11data$k <- q11data$y/q11data$ypred

k <- mean(q11data$k)
```

```r
kadj <- mean(tail(q11data$k, -1))

q11data$ypredk <- k*(q11data$x^3)

q11data$ypredkadj <- kadj*(q11data$x^3)

q11data
```

```
##      y  x ypred          k     ypredk   ypredkadj
## 1    0  1     1 0.00000000  0.0963571   0.1070634
## 2    1  2     8 0.12500000  0.7708568   0.8565076
## 3    2  3    27 0.07407407  2.6016417   2.8907130
## 4    6  4    64 0.09375000  6.1668545   6.8520605
## 5   14  5   125 0.11200000 12.0446376  13.3829307
## 6   24  6   216 0.11111111 20.8131338  23.1257042
## 7   37  7   343 0.10787172 33.0504856  36.7227618
## 8   58  8   512 0.11328125 49.3348356  54.8164840
## 9   82  9   729 0.11248285 70.2443265  78.0492517
## 10 114 10  1000 0.11400000 96.3571009 107.0634454
```

```r
library(reshape2)

library(ggplot2)

ggplot(melt(q11data, id=c("x", "k", "ypred", "ypredk")), aes(x=x, y=value, color=variable)) +
  geom_point()
```



This data set mostly supports the stated proportionality model. I'd be a bit concerned about the 0 value for x=1, which throws off the predicted constant k. This might be a situation that is not exactly passing through the origin, but can be reasonably approximated by a function that does.

I made my graph based on an adjusted k, where I took the mean of the calculated constant disregarding the first 0 value.

Page 94, Problem 4

Lumber Cutters - 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.

Use the following data for your test:

```
q4data <- data.frame(x=c(17, 19, 20, 23, 25, 28, 32, 38, 39, 41),
                     y=c(19, 25, 32, 57, 71, 113, 123, 252, 259, 294))
```

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

a. Consider two separate assumptions, allowing each to lead to a model. Completely analyze each model.
b. Assume that all trees are right-circular cylinders, and are approximately the same height.

For this model, I'm going to assume that board feet is proportional to volume. If trees are assumed to be approximately the same height, then diameter is going to be the only variable. In this case, the area of the circle described by the diameter is what's going to control the board feet.

The volume of a right cylinder is $\pi r^2 h$, but h is constant. So:

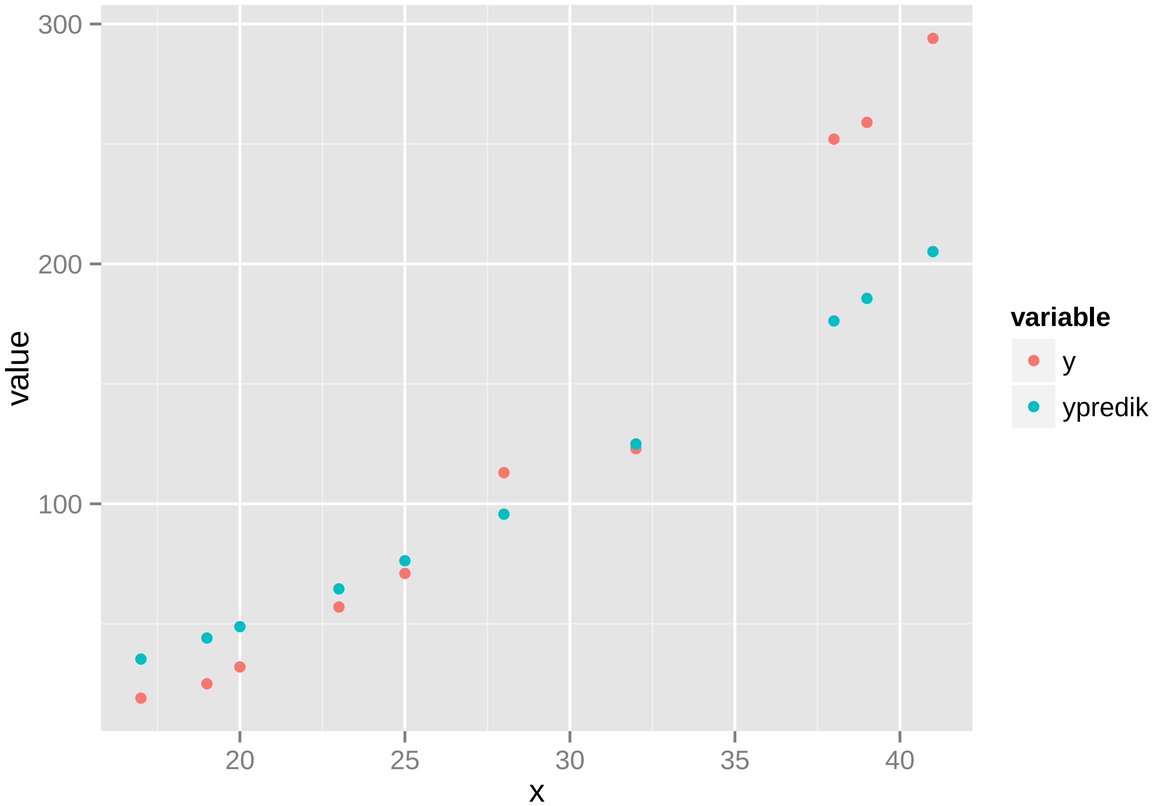
$$y \propto x^2$$

Lets test this out:

```
q4data$ypredi <- q4data$x^2

q4data$ki <- q4data$y/q4data$ypredi

ki <- mean(q4data$ki)

q4data$ypredik <- ki*(q4data$x^2)

q4data
```

```
##     x   y ypredi         ki   ypredik
## 1  17  19    289 0.06574394  35.26639
## 2  19  25    361 0.06925208  44.05248
## 3  20  32    400 0.08000000  48.81161
## 4  23  57    529 0.10775047  64.55335
## 5  25  71    625 0.11360000  76.26814
## 6  28 113    784 0.14413265  95.67075
## 7  32 123   1024 0.12011719 124.95771
## 8  38 252   1444 0.17451524 176.20990
## 9  39 259   1521 0.17028271 185.60614
## 10 41 294   1681 0.17489590 205.13078
```

```
ggplot(melt(q4data, id=c("x", "ki", "ypredi")), aes(x=x, y=value, color=variable)) + geom_point()
```



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

For this model, a new dimension is added. the volume is still $\pi r^2 h$, but h is proportional to the diameter. This new model is now:

$$y \propto x^3$$

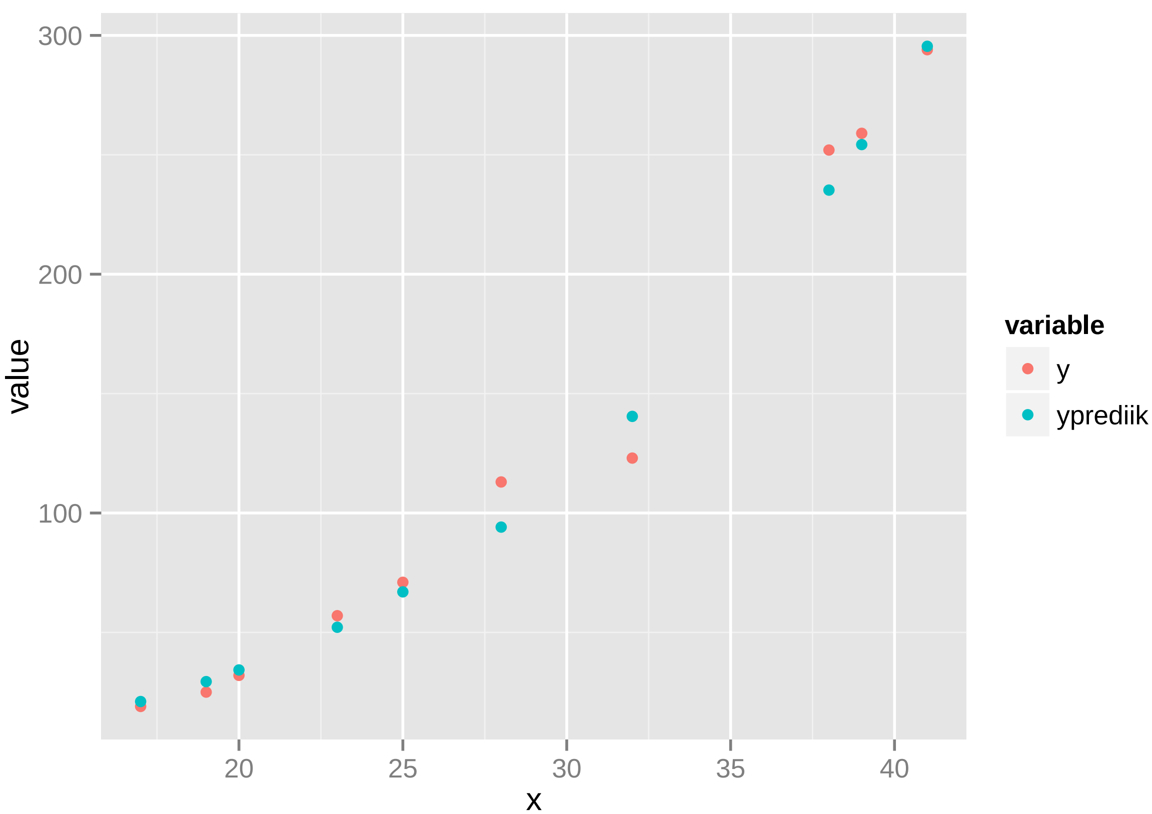
Lets test this one out:

```
q4data$ypredii <- q4data$x^3

q4data$kii <- q4data$y/q4data$ypredii

kii <- mean(q4data$kii)

q4data$ypredik <- kii*(q4data$x^3)

q4data
```

```
##     x   y ypredi         ki   ypredik ypredii         kii  ypredik
## 1  17  19    289 0.06574394  35.26639    4913 0.003867291  21.06040
## 2  19  25    361 0.06925208  44.05248    6859 0.003644846  29.40226
## 3  20  32    400 0.08000000  48.81161    8000 0.004000000  34.29334
```

```
## 4   23  57     529 0.10775047  64.55335   12167 0.004684803  52.15589
## 5   25  71     625 0.11360000  76.26814   15625 0.004544000  66.97919
## 6   28 113     784 0.14413265  95.67075   21952 0.005147595  94.10094
## 7   32 123    1024 0.12011719 124.95771   32768 0.003753662 140.46554
## 8   38 252    1444 0.17451524 176.20990   54872 0.004592506 235.21805
## 9   39 259    1521 0.17028271 185.60614   59319 0.004366223 254.28086
## 10  41 294    1681 0.17489590 205.13078   68921 0.004265754 295.44145
```

```
ggplot(melt(q4data, id=c("x", "ypredi", "ki", "ypredik", "ypredii", "kii")),
       aes(x=x, y=value, color=variable)) + geom_point()
```



b. Which model appears to be better? Why? Justify your conclusions.

Visually, the plot of the second model seems to match more closely than the plot of the first model. It would appear that it is more correct to say height is proportional to diameter, and that $y \propto x^3$ rather than $y \propto x^2$.

Intuitively, this makes sense, since taller trees would have fatter trunks.