1/29/14 Class Notes

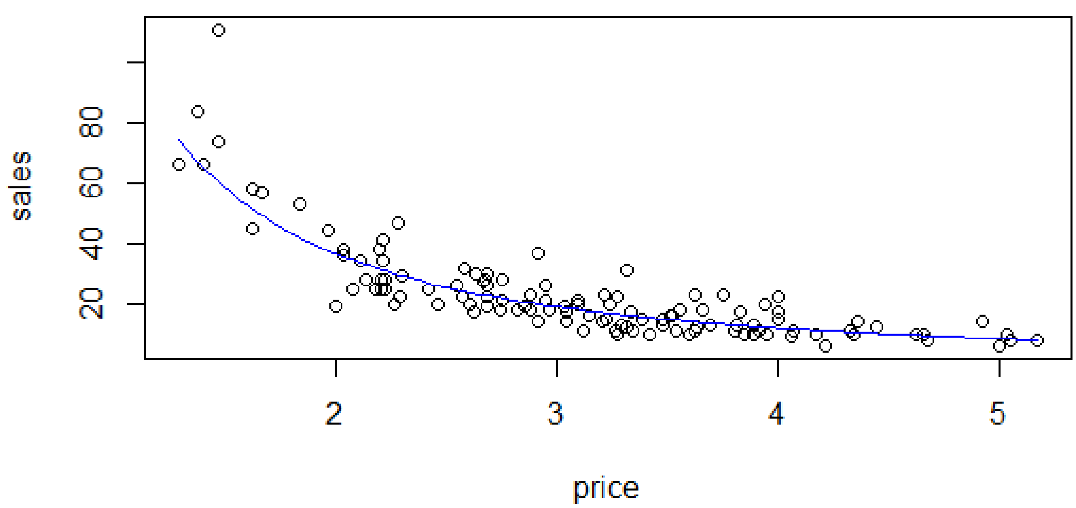
In this lecture, we focused on prediction uncertainty associated with a linear model. Data and R scripts used were milk prices and pickup truck sales on Craigslist. We worked individually in class, and then reviewed the examples during the lecture.

Optimize Profit

a. EXAMPLE: (milk.csv)
   i. How much should we charge for gallons of milk?
   ii. Start from the end; we want to maximize profit
   Let
      Y = profit
      X = price per unit → choice variable
      C = cost per unit ($1.50) → known
      F(x) = units sold → unknown: it is a function of x

   Profit = revenue – cost
   Y = f (x) * (price/unit – cost/unit)
   Y = f(x)*(x-c)
   Profit = x * f(x) – c * f(x)
   iii. Quantity sold VS. Price charged per unit
      1. Data suggests that as price increases, sales decrease
      2. DEMAND CURVE

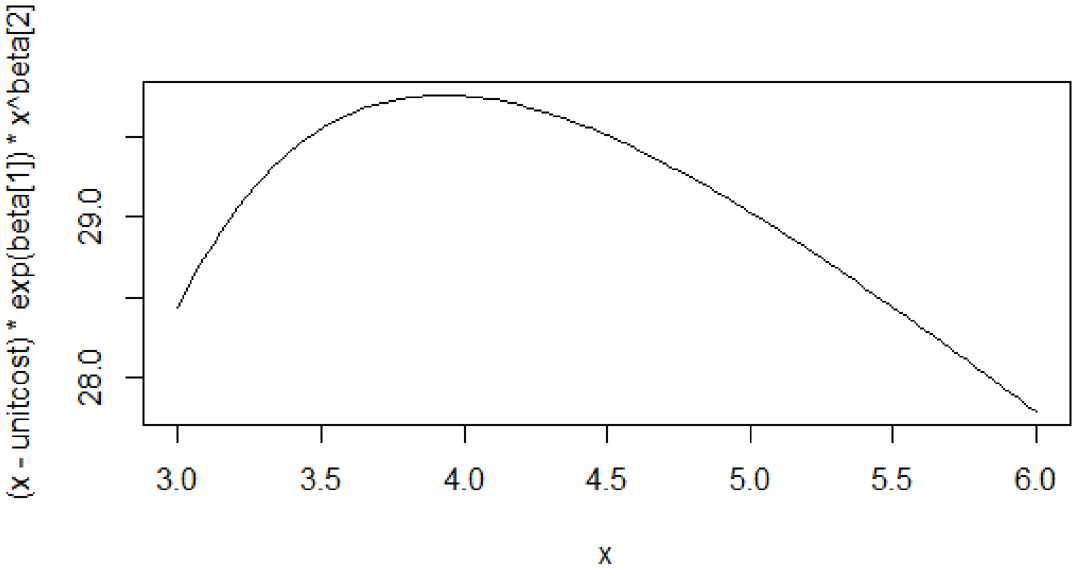
Demand Curve of Milk Prices



b. Optimization strategies

   i. Take derivative and set to 0

   ii. Plot and Point Method → Plot points and take maximum



Reducing Uncertainty

a. EXAMPLE: 4th Story (pickup.csv, pickup.r)

   i. Remember that,

$$\underbrace{Y = Bo + B1(Xi)}_{\textit{Systematic Model Fitted}} \quad + \quad \underbrace{Ei}_{\textit{Residual}}$$

   ii. Point Estimate = y* = Bo + Bx* + E*

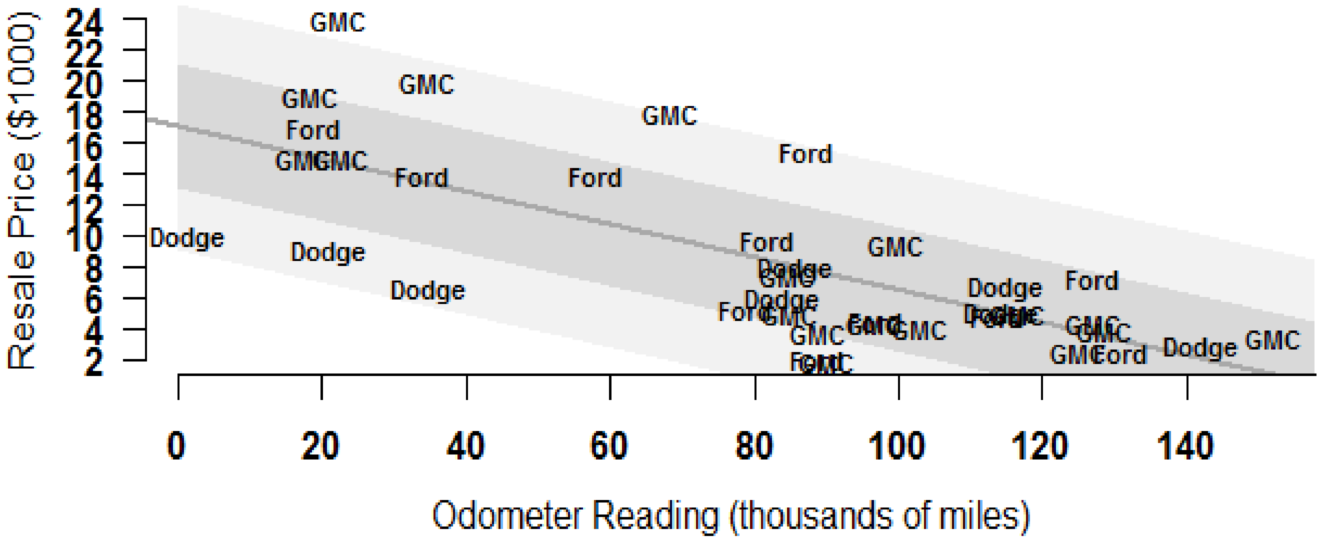
      a. Plug in 60,000 for X.

      b. The purpose of doing this is to take the past variability as measure for likely size of e*. You can use the standard deviation of the residuals.

   iii. Interval Estimate = $y^* = \hat{\beta}_0 + \hat{\beta}_1 x^* \pm S_e$

      a. We widen the interval in order to reduce the likelihood of error

      b. There are no predictions without error bars



How many data points fell outside the given window?
This gives us a measure of empirical coverage.

R^2 (Coefficient of Determination):

a. A measure of the information content on a predictor
b. How to construct:
   i. Se = Standard deviation of residuals
   ii. Sy = std dev of original y variable
   iii. Ratio of these quantities (Se/Sy) tells about predictor:
      The higher that ratio, the lower the information
      The lower the ration, the higher the information
      1 – (Se/Sy)^2 (puts on a more intuitive scale)
c. Relates to Pythagoreans Theorem:
   Some squares in statistics sum up
   Var(resid) + var(fitted Values) = var(data)

Will not dwell on this in class but spend some time understanding this in the notes.

Extra remarks:

a. Pdata() returns fraction of data that falls below given values
b. Construct intervals on transformed scales and undo predictions at the end
c. USING R AS A GRAPHING CALCULATOR
   a. curve((x-unitcost)*exp(beta[1])*x^beta[2], from=1, to=10)
   b. “from = 1, to = 10” set the boundaries of the equation