Industrial Organization and Data Science

Instructors: Justin Rao, Affiliate Professor & Senior Researcher, Microsoft

Jacob LaRiviere, Affiliate Professor & Senior Researcher, Microsoft

Emails: justin.rao@microsoft.com

jlariv@microsoft.com

Course Assignments & Reading

Course assignments should be printed (code, output and descriptive answers) and turned in at the start of class unless otherwise noted. Feel free to work in groups but everyone is required to turn in their own work with answers written in your own words. In both calculations and complex ideas, write down each step of logic used in reaching your conclusion. Keep in mind that in most cases a good answer is one precise sentence; quality is heavily favored over quantity. This will be graded on a full credit, half credit and no credit basis. All work must be typed

Discussion questions do not need be written out ahead of time. At the beginning of each class the professors will lead a discussion around these questions. Students will be called on, potentially at random, to add their insight. This part of class will contribute heavily to your course participation grade.

Week 2, due April 13

Reading: McAfee Ch. 3

Varian's notes on price discrimination

Determinants of Store Level Price Elasticity

Optional: Ch 2-5.3 in Hermalin's lecture notes

Assignment to be turned in. Please turn in your R script and anything in bold below.

Note: you will probably not know all the relevant commands of the top of your head. Nobody does. Simply search "command in R" or "command in R examples" etc. in a search engine, and this will almost always give the answer.

1) Download the orange juice data from the course website and create an R script for this assignment.
2) Change the working directory so that R knows where to look for the data (tip: create a Econ404 folder and save datasets there). See `setwd()`. You can type `?setwd` to see the help file.
3) Read in the data, see `read.csv`. `oj` is a data frame with many variables. You can click it to explore. You can refer to any variable with `oj$var`
4) Visualizing price.
    a. Make a box plot of price.
    b. Make a box plot of log price.
    c. Make a box plot of price, but separate out each brand.
    d. Do the same for log price.

e. **What do these graphs tell you about the variation in price? Why do the log plots look different? Do you find them more/less informative.**

5) Visualizing the quantity/price relationship
    a. Plot logmove (log quantity) vs. log price. Use the `plot` command.
    b. Let's identify each brand separately on the plot
        i. First step is to assign colors `brandcol <- c("green","red","gold")` and `oj$col=brandcol[oj$brand]`. **Inspect the data frame, what did this command do?**
        ii. Use this new column to make a graph where each point is colored by the brand. **What do insights can you derive that were not apparent before?**
6) Estimating the relationship.
    a. Do a regression of log quantity on *log price*. **How well does the model fit? What is the elasticity, does it make sense?**
    b. Now add in an intercept term for each brand (add brand to the regression), **how do the results change? How should we interpret these coefficients?**
    c. Now figure out a way to allow the elasticities to differ by brand. Search "interaction terms" and "dummy variables" if you don't remember this from econometrics. Note the estimate coefficients will "offset" the base estimates. **What is the insights we get from this regression? What is the elasticity for each firm? Do the elasticities make sense?**
7) Impact of "featuring in store"
    a. Which brand is featured the most? **Make a plot to show this**.
    b. How should incorporate the feature variable into our regression? Start with an additive formulation (e.g. feature impacts sales, but not through price).
    c. Now run a model where features can impact sales and price sensitivity.
    d. Now run a model where each brand can have a different impact of being featured and a different impact on price sensitivity. **Produce a table of elasticties for each brand, one row for "featured" and one row for "not featured" (you need 6 estimates).**
8) **Overall analysis**
    a. **Based on your work, which brand has the most elastic demand, which as the least elastic?**
    b. **Do the average prices of each good match up with these insights?**
    c. **Take average prices for each brand. Use the elasticity pricing formula (you can use average values from your analysis above) to "back out" unit costs for each brand. Do the unit costs appear to be the same or different? What are your insights/reactions?**