

# Econ 3385: Measuring Market Power

## Problem Set 1

For this problem set, you may use any programming language or statistical software you like. I recommend R, Stata, Matlab, or Python. Please print out the code you used to generate your results alongside a writeup of your answers.

The file `whales_demand.csv` contains data on the prices and quantities of whale oil sold in the United States every year from 1804 to 1914. Whale oil was used for lighting, and was a primary fuel for much of the 19th century until the development of the crude oil and natural gas industry. The idea of this problem set is to get you familiar with manipulating and analyzing patterns in standard data used for estimating demand and supply.

The variables in the dataset are as follows.

- `year`
- `price_sperm`: The price of a gallon of sperm oil (that is, oil from sperm whales).
- `price_oil`: The price of a gallon of oil not from sperm whales.
- `price_sperm_real`: Sperm oil price adjusted for inflation (in 1914 dollars).
- `price_oil_real`: Non-sperm oil price adjusted for inflation (in 1914 dollars).
- `sperm`: Quantity of sperm oil sold (in 100s of barrels; 1 barrel = 42 gallons).
- `oil`: Quantity of non-sperm oil sold (in 100s of barrels; 1 barrel = 42 gallons).
- `shipwrecks`: the probability of a whale ship being lost at sea in the last 10 years.

## Questions

1. Create a table that summarizes the distribution of the real prices and quantities for sperm and oil. (e.g. mean, median, standard deviation etc.)
  - What do these statistics suggest about demand and supply for these two products?
2. Generate a scatter plot of `oil` against `price_oil_real`, and a scatter plot of `sperm` against `price_sperm_real`.
  - Do you observe any correlations? Is this surprising or not?
3. Generate a scatter plot of `price_oil_real` against `price_sperm_real`.
  - Do you observe any correlation? What might this suggest?
4. Now let's run some regressions. Consider the following equations:

$$Q_y^{oil} = \alpha_1 + \beta_1 P_y^{oil} + \epsilon_y$$
$$Q_y^{sperm} = \alpha_2 + \beta_2 P_y^{sperm} + \epsilon_y$$

- Let's interpret these as demand equations. What sign should we expect the parameters  $\beta_1$  and  $\beta_2$  to have? Why?
  - Now estimate these equations by running two regressions: (1) `oil` on `price_oil_real` and (2) `sperm` on `price_sperm_real` (HINT: use `lm()` in R or `regress` in Stata).
  - Report the estimated parameters and standard errors in a table.
  - Do the estimates make sense? Why or why not?
5. Why might the variable `shipwrecks` be useful for estimating demand?