RA Task

Andreas Kraft

University of Chicago Booth School of Business

This exercise is meant to assess your programming, statistics, and economic reasoning abilities. The task should take you about 2-4 hours, but there is no time limit.

You should submit a PDF with your answers to the questions, including any figures and tables. You should also submit a separate PDF file with your code. Your code should be well-commented, so that it is easy for someone else to understand the operations that you perform. You should code in Python or R; please explicitly state which language at the top of your code.

1 Data and Goals

There are three CSV files containing raw data on vehicle purchases in Texas, obtained from the Texas Department of Vehicles. Some variables in the data have been changed or removed to preserve anonymity, but the vehicle ID variable is a unique identifier for each vehicle. Importantly, the dataset has also been reduced to only include observations that had sales prices between $10,500 and $13,500. The three files are:

1. purchases.csv, which includes a unique vehicle ID, the acquisition cost the car dealership paid to acquire the vehicle, and the acquisition date.
2. sale.csv, which includes a unique vehicle ID, the sales price the car dealership received when selling the vehicle, and the date of the sale.
3. vehicle_information.csv, which includes the unique vehicle ID and an ID for each make, an ID for each model, and the model year.

Goal: Consumers often exhibit left-digit bias and treat numbers right below a round number significantly differently than those right above the threshold. You are interested in seeing how this type of consumer inattention potentially affects firm outcomes in the car market.

2 Summarizing Data and Preliminary analysis

1. We cannot directly observe if dealers make any investments in refurbishing vehicles. However, we observe when the car dealership acquired the vehicle and the subsequent selling date. Generate an additional variable that measures the time (in days) between the acquisition and sales date. Drop all observations not sold within half a year of the initial acquisition date.

2. Create a clearly labeled table of summary statistics that describe the data. These should include the mean, standard deviation, min, and max of the sales price, the acquisition cost, the model year, and the odometer reading.

3. Create a table that lists the five most sold make ids and the five most sold model IDs.

4. Generate a histogram of the sales price of vehicles. Be sure to pick a reasonable bandwidth to create an informative figure.

5. Briefly explain what the distribution looks like.

6. There is a uniform tax rate on vehicle transactions in Texas added to the sales price. Assuming all vehicles in the dataset are priced at or above the *Standard Presumptive Values* and sold without a trade-in, generate a new variable that denotes the price of each vehicle, including the tax.

7. Generate a histogram of the sales price (including tax).

3 Firm Profits

1. To obtain a measure of operating profit for each vehicle, generate a variable that measures the difference between the sales price and the acquisition cost. Present a density plot that presents the operating profit for the vehicles in the sample. Be sure to include a note if you exclude outliers to make the figure more informative.

2. To understand the relationship between the acquisition cost and sales price, it is helpful to create a visual representation. To do so, assign each vehicle to a "bucket," where each bucket has a range of $50. The first bucket contains vehicles between $10,500 and $10,549. The next bucket contains vehicles between $10,550 and $10,599, and so on. In total, you will have 60 such buckets. Calculate the total number of transactions, the mean acquisition price, and the mean operating profit for each bucket.

3. Generate a plot that shows the mean acquisition price for each bucket and a separate plot that shows the number of transactions for each bucket (Hint: The plot should contain 60 dots and have a structure like the attached example).

4. Using a similar approach, plot the average operating profit for each bucket. The y-axis should now measure the operating profit.

5. Perhaps selling vehicles right below a round threshold takes longer. Create a plot similar to the one in 2.3 to look at the average time between acquisition and sale for each "bucket" of vehicles.

6. Run a regression of the operating profit on the time-to-sale variable you have created in 1.1. Create a clearly-labeled table with the regression output and briefly interpret the coefficients.