Pset2_2025_V3_Blank

February 19, 2025

1 Spring 2025 ENVECON/IAS 118 - Introductory Applied Econometrics Problem Set 2

1.1 Due on Gradescope, Midnight February 23

2 Submission Instructions

Go to the file dropdown menu and select the "Save and export notebook as" dropdown menu. In this menu make sure to select "PDF", "Webpdf" or "PDF via Chrome" (if that option appears instead).

The figures in the problem statement may not show up in the pdf you generate. Once you have downloaded this pdf, make sure it shows all your answers and upload it to Gradescope: https://www.gradescope.com/courses/927499

2.1 Question 1: What factors affect the demand for fish in the New York fresh fish market?

This question is an adapted case from a paper published in the Journal of Economic Perspectives (JEP, 2006) titled ''Markets: The Fulton Fish Market,'' by Kathryn Graddy. Full citation: Journal of Economic Perspectives, vol. 20, no. 2, Spring 2006 (pp. 207–220).

Graddy uses data price and quantity data to estimate fish demand in the New York fish market. (But you do not need to read the paper to complete the problem set):

In this problem set, we study the demand for fish for different days of the week in the New York Fish Market using a sample of 111 daily observations.

2.2 Data Description

The data for this exercise come from daily fish quantity sold and price data. The data also has a column identifying which day of the week the observation pertains to, that is, if it is for a Monday, Tuesday, …, or Friday. It also identifies weather variables for each daily observation.

The variables included in the `fishData.dta` that are required for this question are:

- `date` a numeric variable identifying a date
- `day1` : dummy equal to 1 if it is a Monday, 0 otherwise
- `day2` : dummy equal to 1 if it is a Tuesday, 0 otherwise

- `day3` : dummy equal to 1 if it is a Wednesday, 0 otherwise

- `day4` : dummy equal to 1 if it is a Thursday, 0 otherwise

- `day5` : dummy equal to 1 if it is a Friday, 0 otherwise

- `p`: price of fish in dollars per pound

- `q`: quantity sold in pounds

- `rainy`: dummy equal to 1 if it is a rainy day, 0 otherwise

- `cold`: dummy equal to 1 if it is a cold day, 0 otherwise

- `stormy`: dummy equal to 1 if it is stormy at see for fishermen that day, 0 otherwise

- `windspeed`: wind speed in knots

2.3 Question 1 : Descriptive Statistics

Load the dataset `fishdata.dta`. Notice that this is a `.dta` file so you will need to use the `haven` package. Use the `head()` function to have a look at the dataset.

```
[ ]: library(haven)
# Read in data
mydata <- read_dta("fishData.dta")
head(mydata)
```

2.3.1 a) Create log of price and log of q Variables

Please create the log of q and the log of p and add them to the dataframe.

```
[ ]: # Type code here
```

2.3.2 b) Descriptive statistics

i) Weather characteristic How many days are in your data? How many days are rainy days, and how many are not rainy days?

Hint: Use the tidyverse (dplyr) method (after loading the appropriate package in the preamble): filter() trims dataframe to observations with only rainy days and those with only non-rainy days

```
[ ]: # Type code here
```

Type answer here

ii) What is the mean price on rainy days? Hint: check for the `mean()` syntax in this website: https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/mean .

```
[ ]: # Type code here
```

Type answer here

iii) Measures of dispersion Compute the standard deviation of the q and p variables.

You can use canned functions for the standard deviations.

```
[ ]: # Type code here
```

Type answer here

2.3.3 b) Histogram for q

i) Plot a histogram (Hint: use the `hist()` command) of q, with 10 bins.

```
[ ]: # Type code here
```

Type answer here

ii) In the same graph, plot the histogram for q in red for rainy days and in blue for non rainy days What do you conclude when comparing the histogram for rainy and non rainy days

```
[ ]: # Type code here
```

Type answer here

2.3.4 c) Comparisons of q for rainy and non-rainy days

i) Means of quantity Calculate the mean of q between rainy- and non-rainy days and their standard errors. Compare the two means, do they seem substantially different?

Hint: Use the tidyverse (dplyr) method (after loading the appropriate package in the preamble): filter() trims dataframe to observations with only rainy days and those with only non-rainy days

```
[ ]: # Type code here
```

Type answer here

ii) Test Statistics Create a test statistic for the difference of the mean of q between rainy- and non-rainy days. Use a two-tail test. Is the difference statistically significant at the 0.90 confidence level?

```
[ ]: # Type code here
```

Type answer here

2.4 Question 2: Effect of price on quantity

In this part of the exercise, we will estimate the effect of price on quantity sold each day. Consider the two following models:

Model (1): $q = \beta_0 + \beta_1 \, p + u$

Model (2): $q = \beta_0 + \beta_1 \, p + \beta_2 \, rainy + u$

3

2.4.1 a) Estimation

Estimate equations (1) and (2) with `lm()`.

```
[ ]: # Type code here
```

2.4.2 b) Interpretation

Interpret each of the estimated parameters associated with the covariates (other than the constant) of Model (2) - as sign and size (for now, no need to talk about significance)

As an illustration, the estimated beta_0 is the predicted quantity sold when price is zero and it is not raining, holding all else constant.

Type answer here

2.4.3 c) Omitted Variable Bias

How did your estimate of $\hat{\beta}_1$ change between equation (1) and equation (2)? Without performing any calculations, what information does this give you about the correlation between rainy days and price in the sample? (Explain your reasoning in no more than 4 sentences.)

Type answer here

2.4.4 d) Prediction

i) Predict the expected quantity sold, if it is a rainy day and price is equal to 2 using your estimates from Model (2).

```
[ ]: # Type code here
```

Type answer here

ii) Predict the expected quantity sold, if it is not a rainy day and price is equal to 2 using your estimates from Model (2).

```
[ ]: # Type code here
```

Type answer here

2.5 Question 3: Hypothesis testing

2.5.1 (a) Specify and estimate a model that allows you to test the following two hypotheses:

1) The windspeed does not affect the quantity of fish sold holding price, rain, and cold days constant.

2) A rainy day has the same effect on quantity sold than a cold day holding price and windspeed constant.

```
[ ]: # Type code here
```

2.5.2 (b) Hypothesis 1

Given the estimated Model, what can you conclude about your first hypothesis? Use the five steps of hypothesis testing.

```
[ ]: # Type code here
```

Type answer here

2.5.3 (c) Omitting Windspeed

If you omit windspeed from the above model, what happens to the estimate of price? Why is that? explain with the omitted variable formula argument

```
[ ]: # Type code here
```

Type answer here

2.6 Question 3

The data in the sample are for weekdays only, from Monday to Friday. What would happen if you ran the following model?

$q = \beta_0 + \beta_1 \, p + \beta_2 \, rainy + day1 \, \beta_1 + day2 \, \beta_2 + day3 \, \beta_3 + day4 \, \beta_4 + day5 \, \beta_5 + u$

Please explain why that is.

```
[ ]: # Type your code here if you wish to estimate to illustrate
     # but you do not need to estimate the model to answer this question
     # A simple answer below suffices
```

Type answer here

3 THE END