Interpreting Indicator Variables

In this exercise we'll work with a (somewhat canonical) dataset on the prices, mileages, weights, and other characteristics of 74 automobiles. These data originally came from the April 1979 issue of Consumer Reports and from the United States Government EPA statistics on fuel consumption; they were compiled and published by Chambers et al. (1983).

To get the data, follow this link [https://github.com/nickeubank/MIDS_Data/blob/master/automobile_dataset.dta?raw=true] or go to http://www.github.com/nickeubank/MIDS_Data [http://www.github.com/nickeubank/MIDS_Data] and download the `automobile_dataset.dta` file. This is a canonical example dataset used in coding examples all over the internet, and the codebook is roughly:

```
make            Make and Model
price           Price
mpg             Mileage (mpg)
rep78           Repair Record 1978
headroom        Headroom (in.)
trunk           Trunk space (cu. ft.)
weight          Weight (lbs.)
length          Length (in.)
turn            Turn Circle (ft.)
displacement    Displacement (cu. in.)
gear_ratio      Gear Ratio
foreign         Car type
```

Indicator Variables and Omitted Variable Bias

Exercise 1

Create a new variable named `guzzler` that takes the value of 1 if the car's miles per gallon (`mpg`) is less than 18 and takes value 0 otherwise ("guzzler" is a term for a car that consumes gas very quickly, or "guzzles gas"). Regress `price` on `guzzler` and interpret the coefficients. Do gas guzzlers cost more than the other cars? How much more?

Exercise 2

Create a scatter plot of `price` against `weight` and color code your markers by the value of `guzzler` (red for `guzzler` = 1 and green for `guzzler` = 0).

Based on the graph you just created, do you think **not** controlling for `weight` might lead to omitted variable bias in the regression in Exercise 1? What is the direction of the bias?

Exercise 3

Regress `price` on `guzzler`, `weight`, `foreign`, `headroom`, and `displacement`. Interpret the coefficients. Do the regression results confirm your guess in Q3?

Exercise 4

Variable `rep78` indicates the car's repair record. The variable is poorly documented (we don't know that the value means) but take our word for it that the values from 1-5 indicate "very poor", "poor", "acceptable", "good", and "very good" record, respectively.

Create five separate indicator variables from `rep78` and regress `price` on indicators for values 2 through 5. Also control for `headroom`, `weight`, `foreign`, and `displacement`. Interpret the coefficients on the indicator for `rep78 == 3`.

(Note: You can use the `C()` method for creating indicator variables, but your

answers will only be right if the omitted category is `rep78 == 1`).

Interaction Effects

Exercise 5

You suspect that the effect of `guzzler` on `price` may be conditioned by whether or not the car is manufactured abroad. Regress `price` on `guzzler`, `foreign` **and their interaction,** controlling for `headroom`, `weight` and `displacement`. Without using mathematical language, explain to your grandma what the coefficient on the interaction term means.

Exercise 6

What is the price difference between a foreign guzzler and a foreign non-guzzler?

Exercise 7

What is the price difference between a domestic non-guzzler and a foreign non-guzzler?

Exercise 8

Regress `price` on `foreign`, `mpg` and their interaction, controlling for `headroom`, `weight` and `displacement`. Interpret the coefficients of the main independent variables. Explain in layman terms the coefficient on the interaction term.

Absolutely positively need the solutions?

Don't use this link until you've really, really spent time struggling with your code! Doing

so only results in you cheating yourself.

Link [../solutions_warning.html]