This is a meant to be a very simple exposition about modeling energy usage using Stata's auto dataset. To make things work more nicely, let's pretend that this is some sort of sample of measurements, so that when we talk about average energy consumption, it will make some sense.

We'll open the **auto** dataset, and look at its structure behind the scenes. The putdocx command cannot not show Stata code or output in general. It instead is made for putting tables and other output into a report, without showing any Stata code. So, we'll have to do some work, and then report just the results.

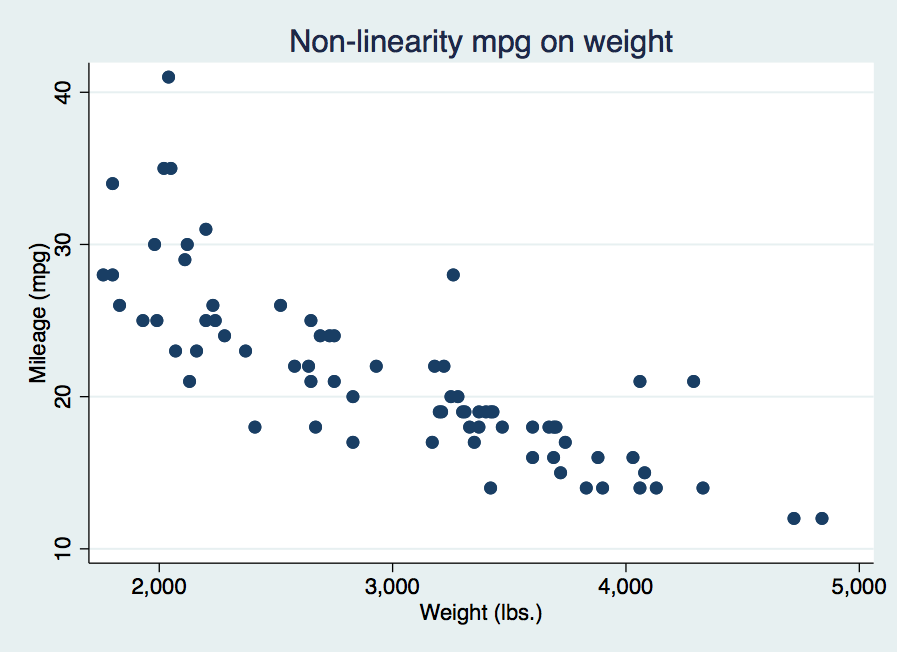
Here are some summary statisticsa about 1978 automobiles:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | N | mean | sd | min | p25 | p50 | p75 | max |
| price | 74 | 6165.256757 | 2949.495885 | 3291 | 4195 | 5006.5 | 6342 | 15906 |
| mpg | 74 | 21.2972973 | 5.78550321 | 12 | 18 | 20 | 25 | 41 |
| weight | 74 | 3019.459459 | 777.1935671 | 1760 | 2240 | 3190 | 3600 | 4840 |
| displacement | 74 | 197.2972973 | 91.83721896 | 79 | 119 | 196 | 250 | 425 |

For those unfamiliar with the system of weights and measures used in the United Stated (and Liberia), the important conversions to remember are that

1 mile = 1605 meters = 1.605 km  
1 pound = 454.6 grams = 0.4546 kg  
1 gallon = 3.7854 liters or about 15/4 liters  
1 inch = 2.54 cm  
Hence 1 cubic inch = 16.387064 cc or about 16 3/8 cc  
1 foot = 12 inches = 30.48cm  
Hence 1 cubic foot = 28316.847 cc or about 28 1/3 liters

One other oddity in the so-called traditional (or Standard or English or Imperial) system, is that energy usage is measured in miles per gallon (mpg). This is not good for analysis, because it makes for a non-linear relationship between weight and energy. This can be seen in following graph:



To make the analysis work better, we will make a variable measuring gallons use per 100 miles driven. This requires one last conversion: 1 gallon per 100 miles is about 75/32 (= 2.344) liters per 100 km.

Let's take a look at various variables by whether the cars are from the US ( *domestic*) or whether they are from outside the US ( *foreign*).

|  |  |  |  |
| --- | --- | --- | --- |
|  | mean | sd | N |
| gp100m | 5.019280218 | 1.279855839 | 74 |
| weight | 3019.459459 | 777.1935671 | 74 |
| length | 187.9324324 | 22.2663399 | 74 |
| turn | 39.64864865 | 4.399353727 | 74 |
| displacement | 197.2972973 | 91.83721896 | 74 |
| gear\_ratio | 3.014864867 | .4562870967 | 74 |

Before modelling, we should take a look to see if there could be collinearities in the predictors.

Finally, how about modelling let's first run a regression with many variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| gp100m | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
| weight | .0014428 | .000216 | 6.68 | 0.000 | .0010118 | .0018737 |
| displacement | .0012388 | .0021161 | 0.59 | 0.560 | -.0029828 | .0054603 |
| gear\_ratio | -.2037991 | .3258603 | -0.63 | 0.534 | -.8538726 | .4462744 |
| foreign | .733736 | .2301493 | 3.19 | 0.002 | .2746007 | 1.192871 |
| \_cons | .8147969 | 1.239181 | 0.66 | 0.513 | -1.657301 | 3.286895 |

We can see that, as expected, heavier cars take more energy to move, an that, as not expected, non-US cars use more gas at the same weight. It appears that we can throw out both displacement and gear\_ratio as predictors and fit a simpler model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| gp100m | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
| weight | .0016254 | .0001183 | 13.74 | 0.000 | .0013896 | .0018612 |
| foreign | .6220535 | .1997381 | 3.11 | 0.003 | .2237871 | 1.02032 |
| \_cons | -.0734839 | .4019932 | -0.18 | 0.855 | -.8750354 | .7280677 |

We can put these coefficients in a table

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | gp100m | gp100m |
| weight | 0.001\*\*\* | 0.002\*\*\* |
|  | (6.679) | (13.743) |
| displacement | 0.001 |  |
|  | (0.585) |  |
| gear\_ratio | -0.204 |  |
|  | (-0.625) |  |
| foreign | 0.734\*\*\* | 0.622\*\*\* |
|  | (3.188) | (3.114) |
| \_cons | 0.815 | -0.073 |
|  | (0.658) | (-0.183) |
| N | 74 | 74 |

From the simple model, cars from 40 years ago used 0.163 gallons per mile per extra 100 pounds, on average. Also, non-US cars use about 0.622 more gallons per mile, on average, all other things being equal.