Homework 1

Data set used: - cars89.dta

1. Report sample mean and sample std.dev of variables – Price, weight, horsepower, seating and engine displacement

```
. sum price weight horsepower seating displacement

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        107    16800.64    9527.994       5666      57183
      weight |        107    2828.093    552.4292       1713       4209
  horsepower |        107    124.6729    40.45094         53        245
     seating |        107    5.037383     1.11529          2          8
displacement |        107    159.4393    66.13126         61        350
```

2. Hypothesis test, Population mean price = 19000

   Null Hypothesis: H0: Bj =19000

   Alternative Hypothesis: H1: Bj <> 19000

   Calculate t-stat

```
. sum price

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        107    16800.64    9527.994       5666      57183

. di (r(mean)-19000)/(r(sd)/sqrt(r(N)))
-2.3877433
```

   Calculate p-Value

```
. di 2*ttail(106,2.3877)
.0187245
```

   As central limit theorem(CLT) implies, sample is greater than 30 which is a normal distribution, so considering the critical value 1.65(90%), 1.96(95%) and 2.58(99%).

   Reject in 90% and 95% as population mean price is 19000 and Fail to reject in 99%.

Extra verification,

```
. ttest price==19000

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   price |     107    16800.64    921.1059    9527.994    14974.45    18626.82
------------------------------------------------------------------------------
    mean = mean(price)                                            t =  -2.3877
Ho: mean = 19000                                 degrees of freedom =      106

    Ha: mean < 19000           Ha: mean != 19000             Ha: mean > 19000
 Pr(T < t) = 0.0094       Pr(|T| > |t|) = 0.0187          Pr(T > t) = 0.9906
```

3. Generated variable “Imported” with dummy values 1 for imported and 0 for local manufactured, (Based on information from Google)

```
gen imported=1 if make=="Audi"
replace imported=0 if make=="Acura"
replace imported=0 if make=="BMW"
replace imported=0 if make=="Buick"
replace imported=0 if make=="Volkswagen"
replace imported=0 if make=="Toyota"
replace imported=0 if make=="Subaru"
replace imported=0 if make=="Nissan"
replace imported=0 if make=="Mercedes"
replace imported=0 if make=="Lincoln"
replace imported=0 if make=="Hyundai"
replace imported=0 if make=="Honda"
replace imported=0 if make=="Ford"
replace imported=0 if make=="Dodge"
replace imported=0 if make=="Chevrolet"
replace imported=0 if make=="Cadillac"
replace imported=1 if make=="Volvo"
replace imported=1 if make=="Sterling"
replace imported=1 if make=="Saab"
replace imported=1 if make=="Pontiac"
replace imported=1 if make=="Peugot"
replace imported=1 if make=="Oldsmobile"
replace imported=1 if make=="Mitsubishi"
replace imported=1 if make=="Merkur"
replace imported=1 if make=="Mercury"
replace imported=1 if make=="Mazda"
replace imported=1 if make=="Isuzu"
replace imported=1 if make=="Geo"
replace imported=1 if make=="Eagle"
replace imported=1 if make=="Daihatsu"
replace imported=1 if make=="Chrysler"
```

Estimate treatment effect is the difference of average outcome from treated and average outcome from non-treated group,

```
. ****Calculate ATE ****
. sum price if imported==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |         36    15247.17    5491.535       6397      28030

. sum price if imported==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |         71    17588.31    10978.25       5666      57183

. di 15247.17-17588.31
-2341.14
```

4. Using the means difference to determine the statistical significance of average treatment effect of a car being imported,

Null Hypothesis,

H0 : Bj = 0

Alternative Hypothesis,

H1: Bj <> 0

Using the summary statistic from question-3, calculate t-stat for average treatment effect,

```
. *** Calculate t-stat from summary statistic obtained in Q.3 ****
. di ((15247.17-17588.31)-0)/sqrt((5491.535)^2/36+(10978.25)^2/71)
-1.470356

. *** t-stat = -1.47 < 1.65,1.96 or 2.58- Fail to reject in 90%, 95% or 99%

. ttest price, by(imported) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      71    17588.31    1302.879    10978.25     14989.8    20186.82
       1 |      36    15247.17    915.2558    5491.535     13389.1    17105.23
---------+--------------------------------------------------------------------
combined |     107    16800.64    921.1059    9527.994    14974.45    18626.82
------------------------------------------------------------------------------
```

```
     diff |             2341.143     1592.227               -815.9499    5498.236
------------------------------------------------------------------------------
     diff = mean(0) - mean(1)                                        t =   1.4704
Ho: diff = 0                        Satterthwaite's degrees of freedom =  104.996
```

As central limit theorem(CLT) implies, sample is greater than 30 which is a normal distribution, so considering the critical value 1.65(90%), 1.96(95%) and 2.58(99%).

Fail to reject in 90%, 95% or 99%.

5. Estimate the equation,

$$price = \beta_0 + \beta_1 weight + \beta_2 horsepower + \varepsilon$$

```
. reg price weight horsepower

      Source |       SS           df       MS      Number of obs   =       107
-------------+----------------------------------   F(2, 104)       =     68.93
       Model |  5.4850e+09         2  2.7425e+09   Prob > F        =    0.0000
    Residual |  4.1380e+09       104  39788428.5   R-squared       =    0.5700
-------------+----------------------------------   Adj R-squared   =    0.5617
       Total |  9.6230e+09       106  90782666.9   Root MSE        =    6307.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   5.767519   1.648051     3.50   0.001     2.499372    9.035667
  horsepower |   111.4845   22.50706     4.95   0.000     66.85213    156.1168
       _cons |  -13409.54   3258.986    -4.11   0.000    -19872.23   -6946.849
------------------------------------------------------------------------------
```

6. Test hypothesis increase of one horsepower correspondence with a price increase of $70

*** Not sure about this, if I got the hypothesis rt ***

Null Hypothesis,

H0: Bj = 70

Alternative Hypothesis,

H1: Bj <> 70

```
. *** Calculate t-stat ***
. di (111.4845-70)/22.50706
1.8431772
```

Reject in 90%, Fail to reject in 95% and 99%

```
. *** Calculate p-value ***
. di 2*ttail(107,1.8431772)
.06807101

. *** p-value = .0681 < 0.10, but > 0.05 and 0.01,Reject in 90%, but fail to
. *** reject in 95% and 99%
```

7. Estimate the equation,

$$price = \beta_0 + \beta_1 weight + \beta_2 horsepower + \beta_3 displacement + \beta_4 seating + \beta_5 imported + \varepsilon$$

```
. reg price weight horsepower displacement seating imported
```

Source	SS	df	MS
Model	6.5286e+09	5	1.3057e+09
Residual	3.0944e+09	101	30637243.9
Total	9.6230e+09	106	90782666.9

Number of obs	=	107
F(5, 101)	=	42.62
Prob > F	=	0.0000
R-squared	=	0.6784
Adj R-squared	=	0.6625
Root MSE	=	5535.1

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
weight	14.75433	2.115436	6.97	0.000	10.55788	18.95079
horsepower	94.77561	26.29188	3.60	0.000	42.61959	146.9316
displacement	-66.10356	16.17708	-4.09	0.000	-98.19454	-34.01259
seating	-2456.172	627.9554	-3.91	0.000	-3701.867	-1210.478
imported	-880.0065	1170.044	-0.75	0.454	-3201.058	1441.045
_cons	-13533.69	3678.084	-3.68	0.000	-20830.02	-6237.358

All the variables are statistically significant ($p<0.05$), For every lb. of weight the price goes up by \$14.75, For every increment of horsepower the price increase by \$94.76, For engine displacement and every seating the price gets decreased as in coefficient, and since imported is a dummy variable, non-imported is baseline-constant and the difference of _cons and imported coefficient is the value impact if the car is imported.

8. Predict the price,

   Car model -Mercedes Benz, considered to be local manufactured.

   Engine displacement – 180.8 cubic inches

   Seating- 5

   Horsepower – 178

Weight – 3131 lb

Using the estimate coefficient from Q.7,

```
. di -13533.69+(3131*14.75433)+(178*94.77561)+(180.8*(-66.10356))+(5*(-2456.172
> ))+(0*(-880.0065))
```

=$25,299.792

9. Assumption #3: - Error term has mean zero and no correlation between the explanatory variable, Above assumption need to be true to use the model from Q.8 to make an active prediction.

   Yes, above is a reasonable assumption for active prediciton because there is some omitted variable that impact the price and are highly correlated with explanatory variable weights, horsepower, etc. that is in error terms, Example of omitted variables – “mpgcity” City mileage per gallon.