HW EE325

1.) This file CEOSAL2.DTA contains data on 177 chief executive officers and can be used to examine the effects of firm performance on CEO salary

(i)
```
. regress lsalary lsales lmktval

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(  2,   174) =   37.13
       Model |  19.3365617     2  9.66828083           Prob > F      =  0.0000
    Residual |  45.3096514   174  .260400295           R-squared     =  0.2991
-------------+------------------------------           Adj R-squared =  0.2911
       Total |  64.6462131   176  .367308029           Root MSE      =  .51029

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lsales |   .1621283   .0396703     4.09   0.000     .0838315    .2404252
     lmktval |    .106708    .050124     2.13   0.035     .0077787    .2056372
       _cons |   4.620917   .2544083    18.16   0.000     4.118794    5.123041
------------------------------------------------------------------------------
```

Ans : The model of the constant elasticity: $\log(salary) = \beta 0 + \beta 1\log(sales) + \beta 2\log(mktval) + u$

The estimated model would be: $\log(salary) = 4.6209 + 0.1621\log(sales) + 0.1067\log(mktval)$

(ii) . regress lsalary lsales lmktval profits

Source	SS	df	MS
Model	19.3509799	3	6.45032663
Residual	45.2952332	173	.261822157
Total	64.6462131	176	.367308029

Number of obs =	177
F(3, 173) =	24.64
Prob > F =	0.0000
R-squared =	0.2993
Adj R-squared =	0.2872
Root MSE =	.51169

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lsales	.1613683	.0399101	4.04	0.000	.0825949	.2401416
lmktval	.0975286	.0636886	1.53	0.128	-.0281782	.2232354
profits	.0000357	.000152	0.23	0.815	-.0002643	.0003356
_cons	4.686924	.3797294	12.34	0.000	3.937425	5.436423

<u>Ans</u> (i) Profits cannot be included in log form because the company can make loss and thus the profit will be negative. But the number after log cannot be negative.

(ii) No, because these firm performance variables can explain the variation in CEO salaries only 29.93%. There is still left 70.07% that cannot be explained.

(iii) . regress lsalary lsales lmktval profits ceoten

Source	SS	df	MS
Model	20.5768102	4	5.14420254
Residual	44.0694029	172	.256217459
Total	64.6462131	176	.367308029

Number of obs =	177
F(4, 172) =	20.08
Prob > F =	0.0000
R-squared =	0.3183
Adj R-squared =	0.3024
Root MSE =	.50618

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lsales	.1622339	.0394826	4.11	0.000	.0843012	.2401667
lmktval	.1017598	.063033	1.61	0.108	-.022658	.2261775
profits	.0000291	.0001504	0.19	0.847	-.0002677	.0003258
ceoten	.0116847	.005342	2.19	0.030	.0011403	.022229
_cons	4.55778	.3802548	11.99	0.000	3.807213	5.308347

<u>Ans</u> The estimated model is:

log(salary) = 4.56 + 0.16log(sales) + 0.10log(mktval) + 0.00prof its + 0.01ceoten

This means that if CEO tenure increases by 1 year, salary would increase by 1%.

(iv)

```
. correlate profits lmktval
(obs=177)

             |  profits  lmktval
-------------+------------------
     profits |   1.0000
     lmktval |   0.7769   1.0000
```

Ans (i) Yes, these variables are highly correlated because the correlation coefficient between lmktval and profits is 0.7769.

(ii) For the OLS estimators, since the log(mktval) and profits are highly correlated meaning that if log(mktval) changes, profits will also change. Moreover you cannot hold any of these factors constant by given a change in one variable. Thus you cannot observe the pure effect of each variable, log(mktval) and profits, on CEO salaries

2.) Use the data in WAGE2.DTA for this exercise

```
      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  3,   931) =   56.97
       Model |  25.6953242     3  8.56510806           Prob > F      =  0.0000
    Residual |  139.960959   931  .150334005           R-squared     =  0.1551
-------------+------------------------------           Adj R-squared =  0.1524
       Total |  165.656283   934  .177362188           Root MSE      =  .38773

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0748638   .0065124    11.50   0.000      .062083    .0876446
       exper |   .0153285   .0033696     4.55   0.000     .0087156    .0219413
      tenure |   .0133748   .0025872     5.17   0.000     .0082974    .0184522
       _cons |   5.496696   .1105282    49.73   0.000     5.279782    5.713609
------------------------------------------------------------------------------
```

(i) Ans $H_o: \beta_2 = \beta_3$

$H_a: \beta_2 \neq \beta_3$

(ii)

```
. test exper = tenure

 ( 1)  exper - tenure = 0

       F(  1,   931) =    0.17
            Prob > F =    0.6805
```

Ans At the 5% significance level, we accept H_o: $\beta_2 = \beta_3$ because it does not fall into the rejection region as the significance level of the test is 68.05%.

3.) Use the data in DISCRIM.DTA to answer this question.

(i)

```
. regress lpsoda prpblck lincome prppov
```

Source	SS	df	MS
Model	.250340622	3	.083446874
Residual	2.62840943	397	.006620679
Total	2.87875005	400	.007196875

Number of obs =	401
F(3, 397) =	12.60
Prob > F =	0.0000
R-squared =	0.0870
Adj R-squared =	0.0801
Root MSE =	.08137

lpsoda	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
prpblck	.0728072	.0306756	2.37	0.018	.0125003	.1331141
lincome	.1369553	.0267554	5.12	0.000	.0843552	.1895553
prppov	.38036	.1327903	2.86	0.004	.1192999	.6414201
_cons	-1.463333	.2937111	-4.98	0.000	-2.040756	-.8859092

Ans The usual form is :

log (psoda) = -1.463333 + 0.728072prpblck + 0.1369553lincome + 0.38036prppov

So given 0.05 level of significant, t value of the prpblck is 2.37, it falls into the rejection region. This is because the p-value of prpblck is equal to 0.018 which is less than the rejection area of 0.05. Therefore we reject H_o at 0.05 level of significant which means that the estimated β_1 is statistically different from zero.

With 0.01 level of significant

```
. regress lpsoda prpblck lincome prppov, level (99)
```

Source	SS	df	MS
Model	.250340622	3	.083446874
Residual	2.62840943	397	.006620679
Total	2.87875005	400	.007196875

Number of obs	=	401
F(3, 397)	=	12.60
Prob > F	=	0.0000
R-squared	=	0.0870
Adj R-squared	=	0.0801
Root MSE	=	.08137

lpsoda	Coef.	Std. Err.	t	P>|t|	[99% Conf.	Interval]
prpblck	.0728072	.0306756	2.37	0.018	-.0065895	.152204
lincome	.1369553	.0267554	5.12	0.000	.067705	.2062056
prppov	.38036	.1327903	2.86	0.004	.0366628	.7240571
_cons	-1.463333	.2937111	-4.98	0.000	-2.223536	-.7031291

It can be concluded that given 0.01 level of significant, t value of the prpblck is 2.37, it does not fall into the rejection region. As the p-value of prpblck is equal to 0.018 which is greater than the rejection area of 0.01, thus we accept H_o at 0.01 level of significant which means that the estimated β_1 is equal to zero.

(ii)

```
. regress lpsoda prpblck lincome  prppov lhseval
```

Source	SS	df	MS
Model	.529488085	4	.132372021
Residual	2.34926197	396	.00593248
Total	2.87875005	400	.007196875

Number of obs	=	401
F(4, 396)	=	22.31
Prob > F	=	0.0000
R-squared	=	0.1839
Adj R-squared	=	0.1757
Root MSE	=	.07702

lpsoda	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
prpblck	.0975502	.0292607	3.33	0.001	.0400244	.155076
lincome	-.0529904	.0375261	-1.41	0.159	-.1267657	.0207848
prppov	.0521229	.1344992	0.39	0.699	-.2122989	.3165447
lhseval	.1213056	.0176841	6.86	0.000	.0865392	.1560721
_cons	-.8415149	.2924318	-2.88	0.004	-1.416428	-.2666019

Ans The coefficient of lhseval is 0.1213056. This means that if hseval increase by 1%, holding other factors constant, psoda will increase by 0.1213056%.

H_o: $\beta_4 = 0$

H_a: $\beta_4 \neq 0$

At 0.05 significance level, the t value is 6.86, which falls into the rejection region because the p-value is 0.000 less than 0.05. So we reject H_o at 0.05 level of significance.

(iii) $H_o: \beta_1 = \beta_2 = 0$

$H_a: \beta_1 \neq \beta_2 \neq 0$

```
. test (lincome=0)(prppov=0)

 ( 1)  lincome = 0
 ( 2)  prppov = 0

       F(  2,   396) =    3.52
            Prob > F =    0.0304
```

Given 0.05 level of significance, we can conclude that the two variables are jointly significant because the p-value is approximately equal to 0.0304 which is less than 0.05, thus it falls into the rejection region. Therefore we reject H_o which means that $\beta_1 \neq \beta_2 \neq 0$, or lincome and prppov are jointly significant.

(iv) The second regression which added log(hseval) in to the model is more reliable because the adjusted R-squared increases when adding this variable . The original adjusted R-squared before we added log(hseval) is equal to 0.0801, which means that only 8.01% of variation in log(psoda) can be explained by prpblck, log(income) and prppov. Later on the adjusted R-squared is improved to 0.1757 when we include the log(hseval) which means that 17.57% of the variation in log(psoda) is explained by prpblck, log(income), prppov and log(hseval).