

QUESTION#1**(i).**

```
. regress lsalary lsales lmkttval
```

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

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

The constant elasticity implies that the model is linear in elasticity.

Then, the constant elasticity model is; $\ln(\text{salary}) = \beta_0 + \beta_1 \ln(\text{sales}) + \beta_2 \ln(\text{mktval}) + u$

After regressing we get the estimated model; $\ln(\widehat{\text{Salary}}) = 4.62 + 0.16 \ln(\text{sales}) + 0.10 \ln(\text{mktval}) + u$
(N = 177, $R^2 = 0.3024$)

(ii).

```
. regress lsalary lsales lmkttval profits
```

Source	SS	df	MS			
Model	19.3509799	3	6.45032663	Number of obs =	177	
Residual	45.2952332	173	.261822157	F(3, 173) =	24.64	
Total	64.6462131	176	.367308029	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
lmkttval	.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

Adding the profit to the model (i); $\ln(\widehat{\text{Salary}}) = 4.69 + 0.16 \ln(\text{sales}) + 0.10 \ln(\text{mktval}) + 0.00 \text{profits}$
(N = 177, $R^2 = 0.2872$)

Since log cannot be negative otherwise it's undefined. Then, the profit cannot be added in logarithmic form because it contains some negative values.

Therefore, the variation in CEO salaries cannot be explained mostly by this kind of firm performance (profits).

Moreover, since R^2 is 0.2872 or approximately 0.3, implying that there are about 70% of variation in $\ln(\text{salary})$ cannot be explained. We can conclude that as we add profit to this model, profits variable has little influence on $\ln(\text{salary})$.

(iii).

. regress lsalary lsales lmktval profits ceoten

Source	SS	df	MS			
Model	20.5768102	4	5.14420254	Number of obs =	177	
Residual	44.0694029	172	.256217459	F(4, 172) =	20.08	
				Prob > F =	0.0000	
				R-squared =	0.3183	
				Adj R-squared =	0.3024	
Total	64.6462131	176	.367308029	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

Adding the variable ceoten to the model (i);

$$\ln(\widehat{Salary}) = 4.56 + 0.16\ln(\text{sales}) + 0.10\ln(\text{mktval}) + 0.00\text{profits} + 0.01\text{ceoten}$$

$$(N = 177, R^2 = 0.3183)$$

Therefore, as CEO tenure increases by one year, holding other factors constant; the estimated percentage return (or the salary) will increase by 1%.

(iv)

. pwcorr lmktval profits

	lmktval	profits
lmktval	1.0000	
profits	0.7769	1.0000

The sample Correlation coefficient between the variables $\ln(\text{mktval})$ and profits is 0.7769 or approximately 0.78. This means that $\ln(\text{mktval})$ and profits are highly correlated. Since the two variables tend to move together, then it is difficult to estimate the effect of each independent variables on $\ln(\text{salary})$. Therefore, the effect of both $\ln(\text{mktval})$ and profits are not very large.

QUESTION #2**(i). Given $\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + u_i$** **. regress lwage educ exper tenure**

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

The new model after the regression is;

$$\ln(\text{wage}) = 5.50 + 0.07\text{educ} + 0.02\text{exper} + 0.01\text{tenure} + u_i$$

Hypothesis; $H_0 : \beta_2 = \beta_3$, meaning that increase one year of workforce experience has the same impact on $\ln(\text{wage})$ as increasing one year of tenure.**(ii).** (At the 5% significance level, constructing a 95% CI.)Test the hypothesis when $\text{exper} = \text{tenure}$, we get the following information;**. test exper = tenure****(1) exper - tenure = 0**

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

As we want to estimate how $\ln(\text{wage})$ depends on both the experience and the tenure equally or not. Given that $\alpha = 0.05$, $df = 935 - 4 = 931$.From critical $t = t_{\frac{\alpha}{2}, n-k} = t_{\frac{0.05}{2}, 931} = 1.96$, and $t = 0.17$ Hence, $|t| < \text{critical } t$; $0.17 < 1.96$ Then, we do not reject the null hypothesis, meaning that one more year of experience has the same effect on $\ln(\text{wage})$ as another year of tenure, or it is statistically insignificant at 95%.By constructing a 95% CI, we can conclude that, given the confidence interval of 95%, in 95 out of 100 cases, the interval (0.0082974, 0.0184522) will contain the true β_2 .

QUESTION#3

(i). Given $\ln(\text{psoda}) = \beta_0 + \beta_1 \text{prpbck} + \beta_2 \ln(\text{income}) + \beta_3 \text{prppov} + u$
 (At 5% level of significant)

```
. regress lpsoda prpbck lnincome prppov
```

Source	SS	df	MS			
Model	.250340622	3	.083446874	Number of obs =	401	
Residual	2.62840943	397	.006620679	F(3, 397) =	12.60	
Total	2.87875005	400	.007196875	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]	
prpbck	.0728072	.0306756	2.37	0.018	.0125003	.1331141
lnincome	.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

Use

OLS to estimate the model; $\ln(\widehat{\text{psoda}}) = -1.46 + 0.07\text{prpbck} + 0.13\ln(\text{income}) + 0.38\text{prppov} + u$
 (N = 401, $R^2 = 0.0801$)

Overall significant of regression at 5% significant level ($\alpha = 0.05$)

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Using the T-test to test for β_1 ,

$$t = 2.37, \text{ and critical } t = t_{\frac{\alpha}{2}, n-k} = t_{\frac{0.05}{2}, 401-4} = t_{\frac{0.05}{2}, 397} = 1.96$$

Since, $|t| > \text{critical } t$; $2.37 > 1.96$

Hence we reject H_0 , meaning that prpbck is significant in explaining $\ln(\text{psoda})$ at the 5% level.

Using the F-test to test for β_1 ,

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\frac{0.250340622}{4-1}}{\frac{2.62840943}{401-4}} = \frac{0.250340622}{2.62840943} \times \frac{397}{3} = 12.60$$

$$\text{critical } F_{0.05, 3, 397} = 2.6$$

Since $F > \text{critical } F$; F-value is highly significant to explain $\ln(\text{psoda})$.

We reject H_0 at the 5% level

(At 1% level of significant)

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

Source	SS	df	MS			
Model	.250340622	3	.083446874	Number of obs =	401	
Residual	2.62840943	397	.006620679	F(3, 397) =	12.60	
Total	2.87875005	400	.007196875	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]	
prpbck	.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

Overall significant of regression at 1% significant level ($\alpha = 0.01$)

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Using the T-test to test for β_1 ,

$$t = 2.37, \text{ and critical } t = t_{\frac{\alpha}{2}, n-k} = t_{\frac{0.01}{2}, 401-4} = t_{\frac{0.05}{2}, 397} = 2.576$$

Since, $|t| < \text{critical } t$; $2.37 < 2.576$

Hence, we do not reject H_0 , meaning that prpbck is insignificant in explaining $\ln(\text{psoda})$ at the 1% level.

To conclude, β_1 or the variable prpbck is statistically different from zero at the 5% level, but not at the 1% level of significant.

(ii). Adding the variable $\ln(\text{hseval})$ in (i), The result of the new regression is as follows.

```
. regress lpsoda prpbck lincome prppov lhseval
```

Source	SS	df	MS			
Model	.529488085	4	.132372021	Number of obs =	401	
Residual	2.34926197	396	.00593248	F(4, 396) =	22.31	
Total	2.87875005	400	.007196875	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]	
prpbck	.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

The new model;

$$\ln(\widehat{\text{psoda}}) = -0.84 + 0.10\text{prpbck} - 0.05\ln(\text{income}) + 0.05\text{prppov} + 0.12\ln(\text{hseval}) + u$$

(N = 401, $R^2 = 0.1839$)

The coefficients can be interpreted as;

$$\beta_0 = -0.8415149$$

$$\beta_3(\text{prppov}) = 0.0521229$$

$$\beta_1(\text{prpbck}) = 0.0975502$$

$$\beta_4(\text{lhseval}) = 0.1213056$$

$$\beta_2(\text{lincome}) = -0.0529904$$

Computing the two-sided p-value for $H_0 : \beta_{\ln(\text{hseval})} = 0$,

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

$$\ln(\widehat{\text{psoda}}) = -0.84 + 0.10\text{prpbck} - 0.05\ln(\text{income}) + 0.05\text{prppov} + 0.12\ln(\text{hseval}) + u$$

p-value for β_4 is 0.00, then $P < 0.05$.

Hence, we reject H_0 , then $\ln(\text{hseval})$ is significant in explaining $\ln(\text{psoda})$ at the 5% level.

Using the T-test to test for β_4

$$t = 6.68, \text{ and critical } t = t_{\frac{\alpha}{2}, n-k} = t_{\frac{0.05}{2}, 401-5} = t_{\frac{0.05}{2}, 396} = 1.96$$

Since, $|t| > \text{critical } t$; $6.68 > 1.96$

Hence, we reject H_0 , meaning that $\ln(\text{hseval})$ is significant in explaining $\ln(\text{psoda})$ at the 5% level.

We use F-test as follows;

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

$$F = \frac{ESS_{\text{new}} - ESS_{\text{old}} / \text{number of new regressors}}{RSS_{\text{new}} / (n-k)}$$

$$F = \frac{0.529488085 - 0.250340622/1}{2.34926197/(401-5)} = \frac{0.279147463}{0.000705} = 395.953$$

$$\text{Critical } F = F_{\alpha, 1, n-k} = F_{0.05, 1, 396} = 3.84$$

Since $F > \text{critical } F$; $395.953 > 3.84$.

Hence, we reject H_0 or meaning that $\ln(\text{hseval})$ contributes to $\ln(\text{psoda})$ is statistically significant at 5% level. Also, F-value is highly significant, suggesting that the addition of $\ln(\text{hseval})$ to the model significantly increases ESS and hence the R^2 value.

(iii). From $\ln(\widehat{\text{psoda}}) = -0.84 + 0.10\text{prpbck} - 0.05\ln(\text{income}) + 0.05\text{prppov} + 0.12\ln(\text{hseval}) + u$

$$\text{p-value} = (0.004) \quad (0.001) \quad (0.159) \quad (0.699) \quad (0.00)$$

p-value for $\beta_2 \ln(\text{income})$ is 0.159, then $P > 0.05$.

Hence, we accept H_0 , then $\ln(\text{income})$ is statistically insignificant.

p-value for $\beta_3 \text{prppov}$ is 0.699, then $P > 0.05$.

Hence, we accept H_0 , then prppov is statistically insignificant.

Therefore, both variables $\ln(\text{income})$ and prppov are not individually significant. In other words, after adding the variable $\ln(\text{hseval})$ into the model, prppov and $\ln(\text{income})$ cannot explain the $\ln(\text{psoda})$ anymore.

```
. pwcorr prppov lincome
```

	prppov	lincome
prppov	1.0000	
lincome	-0.8385	1.0000

According to the table above, correlation coefficient between the variables $\ln(\text{income})$ and prppov is -0.8385. The negative correlation coefficient suggests that as $\ln(\text{mktval})$ increases, prppov tends to decrease. Since the two variables does not move together, then we can estimate the effect of each independent variables on $\ln(\text{salary})$ easier.