Tharp 620464099

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + \beta_2 \mathrm{ROE}_i + \beta_3 finance_i + \beta_4 consprod_i + \beta_5 utility_i + u_i$$

1 a)

$\log(salary_i) =$
$4.588101 + 0.2571917(\log(sales_i)) + 0.111517(ROE_i)$
$+ 0.1579564 D_{3_i} + 0.1808917 D_{4_i} - 0.2830015 D_{5_i} + u_i$

$D_{3_i} = 1$ if in financial industry; 0 for otherwise
$D_{4_i} = 1$ " consumer product " ; " — "
$D_{5_i} = 1$ " utility " ; " — "

$\beta_1 = 0.2571917$ means that if sales increases by one percent, salary is expected to increase by β_1 percent.

b)

The overall significance of this regression 99% significant from using the p-value of the F test.

Significant coefficients at 5% level: by the p-value of the t test

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_4, \hat{\beta}_5$

$d.f. = n-k = 209 - 6 = 203$

The critical value $\approx \pm 1.960$

c)

$\approx 28.3002\%$

d)

Because then we wouldn't have a baseline or when all the dummies = 0 there would be no sector for the regression. If we put all of them in the equation, stata will automatically omit one variable because it can't be estimated.

e)

There is a possibility that this could be beneficial because there could be variation in the amount of equity that companies hold in each sector, leading to different range of ROE in each sector and could possibly means that ROE affects salary of CEOs in each sector differently.

2 a)

$H_0: \hat{\beta}_1 = 0$

$H_a: \hat{\beta}_1 \neq 0$

$t_{calc} = \frac{-.5876985}{.1090181} = -5.3908$

$d.f. = n-k = 1191 - 3 = 1188$

$t_{\alpha/2}$ critical value $= \pm 1.960$

$t_{calc} >$ critical value

We can reject the null hypothesis.

95% of the time we are sure the smoking affects birth weight.

b)

$\alpha = 0.01 \quad t_{\alpha/2} = 2.576$

$$[\hat{\beta}_2 - s.e. \cdot t_{\alpha/2} < \beta_2 < \hat{\beta}_2 + s.e. \cdot t_{\alpha/2}] = 1 - \alpha = 0.99$$

$$[.0624684 - 0.0324438(2.576) < \beta_2 < 0.0624684 + 0.0324438(2.576)] = 0.99$$

$$= [-0.02110683 < \beta_2 < 0.14604363] = 0.99$$

c) $\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} = 1 - (1 - .0301)\frac{1190}{1188} = 0.0285$

d)

The overall significance of this regression 99% significant from using the p-value of the t test.

Critical value from F-test : 2.37

T-test's Critical Value : ± 1.960 $d.f. = n-k = 1191-5 = 1186$

T calc :

$\hat{\beta}_0 : \frac{118.0741}{3.500291} = 33.7327 > 1.960$; significant

$\hat{\beta}_1 : \frac{-0.5899954}{0.1106172} = 5.3291 < -1.960$; " —— "

$\hat{\beta}_2 : \frac{0.0538254}{0.0366502} = 1.4686$, $-1.960 < t_{calc} < 1.960$; not significant

$\hat{\beta}_3 : \frac{0.4436695}{0.2832896} = 1.7426$ " —— "

$\hat{\beta}_4 : \frac{-0.4379234}{0.3197377} = 1.3696$ " —— "

e)

H_0: father education & mother education has no marginal contribution to the model

H_a: otherwise

$$F_{cal} = \frac{ESS_{new} - ESS_{old} / (\#\text{ of new regressors})}{RSS_{new} / (n - K_{new})} = \frac{15827.6593 - 14536.95346 / 2}{466919.033 / (1191 - 5)}$$

$$= \frac{645.35275}{393.692271} = 1.6392314\text{5}$$

$$F_{upper\ \alpha}(2, 1186) = 3.00 > F_{cal}$$

$\alpha = 0.05$

$\therefore$ 95% of the time the education of father and mother has no marginal contribution to the model i.e. parents' education doesn't really have any impact on birth weight.

3 a, b)

$$lawge_i = \beta_1 + \beta_2 exp_i + +\beta_3 expsq_i + \beta_4 educ_i + \beta_5 age_i + \beta_6 kid6_i + \beta_7 kid18_i + u_i$$

$lwage_i$	= natural log of hourly wage of married women
exp_i	= years of experience
$expsq_i$	= years of experience squared
$educ_i$	= years of education
age_i	= age
$kid6_i$	= number of children aged 0-6 in a household
$kid18_i$	= number of children aged 6-18 in a household

Source	SS	df	MS
Model	34.8932483	6	5.815547167
Residual	188.434158	421	.446526442
Total	223.327441	427	.523015084

Number of obs	=	428
F(6 , 421)	=	13.19
Prob > F	=	0.0000
R-squared	=	0.1582
Adj R-squared	=	0.1462
Root MSE	=	.66823

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exper	.039819	.013393	2.97	0.003	.0134936	.0661444
expersq	-.0007812	.0004022	-1.94	0.053	-.0015718	9.37e-06
educ	.1078319	.0144021	7.49	0.000	.079523	.1361409
age	-.0014653	.0052925	-0.28	0.782	-.0118682	.0089377
kidslt6	-.0607106	.0887626	-0.68	0.494	-.2351836	.1137625
kidsge6	-.014591	.0278981	-0.52	0.601	-.069428	.0402459
_cons	-.4209078	.316905	-1.33	0.185	-1.043821	.2020053

c)

$$\bar{R}^2 = 1-(1-R)\frac{n-1}{n-k} = 1-(1-0{,}1582)\frac{428-1}{428-7} = 0{,}1462$$

d)

$$F_{cal} = \frac{R^2_{new} - R^2_{old} / \#\text{ of new regressors}}{1 - R^2_{new}/(n-K_{new})} = \frac{0.1582 - X / 1}{1-0.1582/(428-7)} > 3.84$$

H_0: the new regressor has no marginal contribution to the model.

new = model 3.1

old = " 3.2

$$= \frac{.1582 - X}{0.0020} > 3{,}84$$

$$.1582 - X > 0.00768$$

$$-X > 0.00768 - .1582$$

$$X < .1582 - .00768$$

$$R^2_{model\ 3.2} = X < .15052$$

e)

No, because, intuitively, higher age comes with higher wage due to more experience, knowledge etc. As I've mentioned that with higher age comes higher experience, I believe that age and experience are highly correlated which would leads to multicollinearity and become the cause of this insignificance.