lsales	.2571917
roe	.0111517
finance	.1579564
consprod	.1808917
utility	-.2830015
_cons	4.588101

a. Write out the estimated regression equation for $\log(salary_i)$. Interpret the estimated coefficient associated with $\log(sales_i)$.

$\widehat{\log(\text{salary})} = b_0 + b_1 \log(\text{sale}) + b_2 \text{ROE} + b_3 \text{finance} + b_4 \text{consprod} + b_5 \text{utility}$

$= 4.588101 + 0.2571917 \log(\text{sale}) + 0.111517 \text{ROE} + 0.1579564 \text{finance} + 0.1808917 \text{consprod} - 0.2830015 \text{utility}$

Interpret the estimated coefficient associated with $\log(\text{sales}_i)$ ← ค่า b_1, b_2, b_3 ← โจทย์ข้อนี้ใช้ b_1

$\widehat{\log(\text{salary})} = 4.588101 + 0.2571917 \log(\text{sale})$

$b_1 = 0.2571917$ the elasticity of CEO saraly is 0.2571917

If the sale revenue increase for 1%, CEO salary will increase 0.257%.

b. What is the overall significance of the regression? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State **the critical value** for hypothesis testing to receive full points. สมการที่สร้างมามัน Predict Y ได้ไหม

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: At least one of $\beta_i \neq 0$; $i = 1,2,3,4,5$

P>|t|
0.000
0.010
0.077
0.034
0.005
0.000

1.) $H_0: \beta_1 = 0$ because $P>|t| = 0.000 \leq \alpha = 0.05$
$H_1: \beta_1 \neq 0$ $\therefore$ we reject $H_0: \beta_1 \neq 0$

2.) $H_0: \beta_2 = 0$ because $P>|t| = 0.010 \leq \alpha = 0.05$
$H_1: \beta_2 \neq 0$ $\therefore$ we reject $H_0: \beta_2 \neq 0$

3) $H_0: \beta_3 = 0$ because $P>|t| = 0.077 > \alpha = 0.05$
$H_1: \beta_3 \neq 0$ $\therefore$ fail to reject $H_0: \beta_3 = 0$

4) $H_0: \beta_4 = 0$ because $P>|t| = 0.034 \leq \alpha = 0.05$
$H_1: \beta_4 \neq 0$ $\therefore$ we reject $H_0: \beta_4 \neq 0$

5.) $H_0: \beta_5 = 0$ because $P>|t| = 0.005 \leq \alpha = 0.05$
$H_1: \beta_5 \neq 0$ $\therefore$ we reject $H_0: \beta_5 \neq 0$

because Prob > F = 0.0000 $\leq \alpha$ (0.05) $\therefore$ the overall is significant ← F-test : we reject H_0 : At least one of $\beta_i \neq 0$

c. Compute the approximate percentage difference in estimated salary between the utility and transportation sector, holding $sales_i$ and ROE_i fixed.

$\widehat{\log(\text{salary})} = 4.588 + 0.257 \log(\text{sale}) + 0.112 \text{ROE} + 0.158 \text{finance} + 0.181 \text{consprod} - 0.283 \text{utility}$

กำหนดให้ sale = \$ 10,000,000 (แทน 10 ลงไป)

ROE = 10%

Case the utility

$\widehat{\log(\text{salary})} = 4.588 + 0.257 \log(\text{sale}) + 0.112 \text{ROE} + 0.158 \text{finance} + 0.181 \text{consprod} - 0.283 \text{utility}$

$\widehat{\log(\text{salary})} = 4.588 + 0.257 \log 10 + 0.112(10) + 0.158(0) + 0.181(0) - 0.283(1) = 5.682\%$

Case transportation

$\widehat{\log(\text{salary})} = 4.588 + 0.257 \log 10 + 0.112(10) + 0.158(0) + 0.181(0) - 0.283(0) = 5.965\%$

Difference = 5.965% − 5.682% = 0.283%

d. Why can't we put all the sector dummies (i.e. $finance_i$, $consprod_i$, $utility_i$ and $transport_i$) in the equation? What would happen if we put all the sector dummies in the equation and use STATA run the regression anyway?

because it made colinearlity problem and STATA would show error.

e. In the above model, is there any benefit if we add interaction terms between roe and sector dummies, i.e. $\text{ROE}_i * finance_i$ and/or $\text{ROE}_i * consprod_i$ and/or $\text{ROE}_i * utility_i$?

useless because there may be a problem of multicoeffinearity between ROE and Dummmies variable

a. Based on **Model 2.1**, test whether smoking has an impact on birth weight. Show your work. (use $\alpha = 0.05$)

วิธีการ test ว่า smoking มี impact ต่อ birth weight หรือเปล่า เนื่องจาก smoking มันคือ cig ที่อยู่ติดกับ β_1 $\therefore$ เราจะ test ที่ β_1 ถ้า $\beta_1 = 0$ ไม่มี impact

$\beta_1 \neq 0$ มี impact

bwght		Coef.	Std. Err.	
cigs	b_1	-.5876985	.1090181	sb_1
faminc	b_2	.0624684	.0324438	sb_2
_cons	b_0	118.5568	1.234278	sb_0

1) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

2) test statistics

$t = \dfrac{b_1 - \beta_1}{sb_1}$ ← ดูจากตอนตั้ง H_0 กับ H_1 ได้เลย ส่วนใหญ่จะเป็น 0

$b_1 = -0.588 \qquad sb_1 = 0.109$

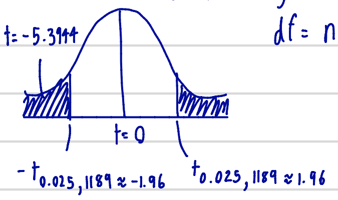
$t = \dfrac{-0.588 - 0}{0.109}$

$t = -5.3944$

3) Determine the rejection region $\alpha = 0.05$

$df = n - 2 = 1191 - 2$

$= 1189$



4) Conclusion because $t_{compute} = -5.3944$ fall into the rejection region

$\therefore$ we reject $H_0 : \beta \neq 0$ $\therefore$ smoking has an impact on birth weight.

b. Based on **Model 2.1**, construct a 99% confidence interval for β_2.

change 99% confident interval to α

$99\% = (1-\alpha)\%$ $\qquad \alpha = 1 - 0.99$

$\frac{99}{100} = 1 - \alpha$ $\qquad \alpha = 0.01$

$0.99 = 1 - \alpha$

$df = n - k - 1$

$= 1191 - 2 - 1$

$= 1188$

$b_2 \pm [t_{\frac{\alpha}{2}, n-2} \cdot sb_2]$

$0.062 \pm [t_{0.005, 1188} \cdot 0.032]$

$b_2 = 0.062$

$sb_2 = 0.032$

$n = 1191$

$-0.0204 < \beta_2 < 0.1444$

c. Would your conclusion in a) change if you use the result from **Model 2.2**? Show your work. (use $\alpha = 0.05$)

1.) $H : \beta_1 = 0$

$H : \beta_1 \neq 0$

2.) test statistics

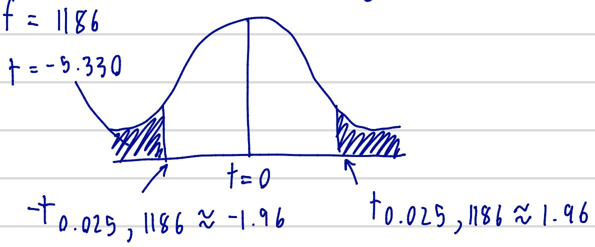
$t = \dfrac{b_1 - \beta_1}{sb_1}$

$t = \dfrac{-0.5895 - 0}{0.1106}$

$t = -5.330$

3.) determine the rejection region $\alpha = 0.05$ $k = 4$

$df = 1186$



4.) Conclusion because $t_{compute} = -5.3944$ fall into the rejection region

$\therefore$ we reject $H_0 : \beta_1 \neq 0$

$\therefore$ smoking has an impact on birth weight

d. What is the overall significance of the regression from **Model 2.2**? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State the critical value for hypothesis testing to receive full points.

1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$
H_1: At least one β_i is not equal to 0

2) test statistics because prob > F = 0.0000 $\leq \alpha = 0.05$

$\therefore$ we reject H_0: At least one β_i is not equal to 0

1.) $H: B_1 = 0$
$H: B_1 \neq 0$

2.) test statistics

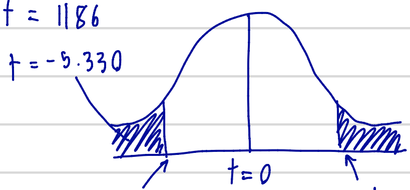
$t = \frac{b_1 - B_1}{sb_1}$

$t = \frac{-0.5895 - 0}{0.1106}$

$t = -5.330$

3.) determine the rejection region $\alpha = 0.05$ $k=4$

df = 1186



$-t_{0.025,1186} \approx -1.96$ $t_{0.025,1186} \approx 1.96$

4.) Conclusion because $t_{compute} = -5.3944$ fall into the rejection region

$\therefore$ we reject $H_0: B_1 \neq 0$

1.) $H: B_2 = 0$
$H: B_2 \neq 0$

2.) test statistics

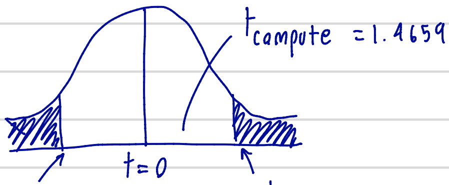
$t = \frac{b_2 - B_2}{sb_2}$

$t = \frac{0.0538 - 0}{0.0367}$

$t = 1.4659$

3.) determine the rejection region $\alpha = 0.05$ $k=4$

df = 1186



$-t_{0.025,1186} \approx -1.96$ $t_{0.025,1186} \approx 1.96$

4.) Conclusion because $t_{compute} = 1.4659$ does not fall into the rejection region

$\therefore$ fail to reject $H_0: B_2 = 0$

1.) $H: B_3 = 0$
$H: B_3 \neq 0$

2.) test statistics

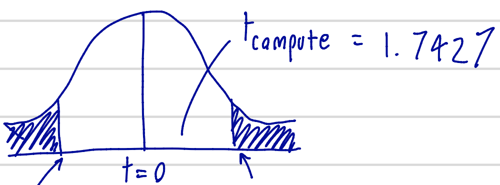
$t = \frac{b_3 - B_3}{sb_3}$

$t = \frac{0.4937 - 0}{0.2833}$

$t = 1.7427$

3.) determine the rejection region $\alpha = 0.05$ $k=4$

df = 1186



$-t_{0.025,1186} \approx -1.96$ $t_{0.025,1186} \approx 1.96$

4.) Conclusion because $t_{compute} = 1.7427$ does not fall into the rejection region

$\therefore$ fail to reject $H_0: B_3 = 0$

1.) $H: B_4 = 0$
$H: B_4 \neq 0$

2.) test statistics

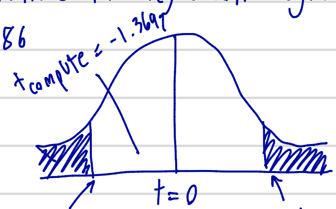
$t = \frac{b_4 - B_4}{sb_4}$

$t = \frac{-0.4379 - 0}{0.3197}$

$t = -1.3697$

3.) determine the rejection region $\alpha = 0.05$ $k=4$

df = 1186



$-t_{0.025,1186} \approx -1.96$ $t_{0.025,1186} \approx 1.96$

4.) Conclusion because $t_{compute} = -1.3697$ does not fall into the rejection region

$\therefore$ fail to reject $H_0: B_4 = 0$

a) Figure out all the degrees of freedom in this model.

b) Figure out all the sum of squares (ESS and RSS) and mean squares in this model.

ANOVA table for multiple linear regression

Source of variation	Sum of Squares	df	Mean Squares	F ratio
Regression	SSR	k	$MSR = SSR / k$	MSR / MSE
Error	SSE	$n-k-1$	$MSE = SSE / (n-k-1)$	
Total	SST	$n-1$		

Source	SS	df	MS
Model	35.3400	6	5.8900
Residual	187.9874	421	.446526442
Total	223.327441	427	6.336526

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

$MSE = \frac{SSE}{df}$ $\therefore SSE = MSE \cdot df$

$= 0.446526 \cdot 421$

$= 187.987446$

c) Figure out the adjusted R-squared ($\bar{R}^2$)

$$\text{Adjusted } R^2 = 1 - \frac{MSE}{MST}$$

$$= 1 - \frac{0.446526442}{6.336526}$$

$$= 0.9295$$

d) Given that the model above is called '**Model 3.1**', there is another competing model called '**Model 3.2**' which **an explanatory variable is excluded**, compared to '**Model 3.1**'. Though the result of estimating '**Model 3.2**' is not shown here, **what is the maximum value of R^2 from 'Model 3.2'** which will make you conclude that the excluded variable has a significant contribution in '**Model 3.1**', at the significance level of 0.05. (**Hint:** the critical value of the F-test at the significance level of 0.05 is $F_{1,421} = 3.84$)

$n = 428$ k = number of independent and dependent

m = number of eliminate

$F = \frac{[R^2_{UR} - R^2_R]/m}{[1-R^2_{UR}]/n-k}$ $3.84 = \frac{[0.1582 - R^2_R]/1}{[1-0.1582]/421}$ $0.007678 = 0.1582 - R^2_R$ $R^2_R = 0.1582 - 0.007678$ $R^2_R = 0.150522$

e) As you can see from the result, age is not significantly different from zero. In other words, age does not determine how much hourly wage would be. Does this make economic sense in your opinion? What do you think cause this insignificance?

Doesn't make economic sense because when you are a child it's mean no experience and when you are adult it's mean you have experience
$\therefore$ Problem may occur because both age and experience variable reflect the same thing.