Introductory Econometrics

Solutions Problem Set 3: Inference in the Linear Regression Model

Brief Solutions

The solution is aimed to help you understand where to find the numbers. There is no need for you to copy and paste the whole table when it comes to the project report. Only the result is good enough.

1

	(1)
	ltsal
educ	0.05***
	(32.88)
age	0.01*
	(2.37)
age2	-0.00
	(-0.35)
female	-0.14***
	(-17.67)
tech	-0.06***
	(-7.25)
enf18	0.01***
	(3.36)
_cons	1.06***
	(10.48)
N	4503
adj. R^2	0.258

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2 Model6: every variable has a significante effect at 5% and 10% level, except age2.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.05884	0.10107	10.48	<.0001
educ	educ	1	0.05304	0.00161	32.88	<.0001
age	agd	1	0.01196	0.00505	2.37	0.0178
age2		1	-0.00002223	0.00006311	-0.35	0.7246
female		1	-0.14342	0.00812	-17.67	<.0001
tech	tech	1	-0.06066	0.00836	-7.25	<.0001
enf18		1	0.01412	0.00421	3.36	0.0008

3 we reject H_0 at the 5% level, i.e., the model is globally significant.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	113.20063	18.86677	261.97	<.0001
Error	4496	323.79411	0.07202		
Corrected Total	4502	436.99474			

4

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	1.05884	0.10107	10.48	<.0001	0.86068	1.25699
educ	educ	1	0.05304	0.00161	32.88	<.0001	0.04988	0.05620
age	agd	1	0.01196	0.00505	2.37	0.0178	0.00207	0.02186
age2		1	-0.00002223	0.00006311	-0.35	0.7246	-0.00014595	0.00010149
female		1	-0.14342	0.00812	-17.67	<.0001	-0.15933	-0.12750
tech	tech	1	-0.06066	0.00836	-7.25	<.0001	-0.07706	-0.04426
enf18		1	0.01412	0.00421	3.36	0.0008	0.00587	0.02236

Additional: test educ=age=1

W1: from CI directly, we reject it at 5% level.

educ	educ	1	0.05304	0.00161	32.88	<.0001	0.04988	0.05620
age	agd	1	0.01196	0.00505	2.37	0.0178	0.00207	0.02186

W2: test it in SAS, we reject it at 5% level.

Test 1 Results for Dependent Variable ltsal				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	14008	194502	<.0001
Denominator	4496	0.07202		

5 We prefer M6 to M3.

Test 1 Results for Dependent Variable ltsal				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	2.30649	32.03	<.0001
Denominator	4496	0.07202		

6 Similar as in 5. Do it by yourself!

a)

For $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = F_{4,4503-7;5\%} = 2,37$
$F = 90.34 > 2,37$ (or p-value$<5\%$) then we reject H_0 at the 5% level: we prefer M_6 to M_1.

b)

For $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = F_{2,4503-5;5\%} = 3$
$F = 146.64 > 3$ (or p-value$<5\%$) then we reject H_0 at the 5% level: we prefer M_3 à M_1.

c)

For $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = F_{1,4503-4;5\%} = 3,84$
$F = 3.40 < 3,84$ (or p-value$<5\%$) then we do not reject H_0 at the 5% level: we prefer M_1 to M_2

Remark: here we can also use a Student test as there is only one restriction.

d)

For $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = F_{2,4503-3;5\%} = 3$
$F = 560.7 > 3$ (or p-value$<5\%$) then we reject H_0 at the 5% level: we prefer M_1 to M_0.

Review of concepts

Inference in Multiple Linear Regression

1 t-test: one restriction test

1) **Hypothesis Development**

$H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ (bilateral test)

H0 null hypothesis, H1 alternative hypothesis

2) Test statistic and its distribution

$$t = \frac{\widehat{\beta}_j - 0}{\widehat{\sigma}_{\beta_j}} \sim_{H_0} t_{n-(k+1)}$$

n: sample size (4503), k: number of explanatory variables (6).

Remark 1: this is because σ^2 is unkown.

Remark 2: when n→+∞, the Student distribution (t distribution) tends towards a N(0, 1) distribution (asymptotic test).

3) **Decision rule**

3.1) t-value decision rule

we reject H_0 if $|t| > t_{n-(k+1), \frac{\alpha}{2}}$; otherwise, we do not reject H_0.

- $|t|$ is computed from the estimated results. For example, $\widehat{\beta}_2 = 0.05304$ and $\widehat{\sigma_{\beta 2}} = 0.00161$, $t_{obs} = \frac{0.05304}{0.00161} = 32.88$
- $t_{n-(k+1), \frac{\alpha}{2}}$ is the critical value, which could be found in STATISTICAL TABLES. For example, when n→+∞, $t_{n-(k+1), \frac{\alpha}{2}} \simeq 1,96$ for $\alpha = 5\%$
- $|t_{obs}| > 1,96$ then we reject H_0 at the 5% level
- So the variable education has a significant effect onthe wage rate at the 5%level.

3.2) p-value decision rule

If p-value $\leq \alpha$: we reject H_0; otherwise, we do not reject H_0.

- For example, p-value=" <.0001"<0,05=5% so we reject H_0 at the 5% level. So the variable education has a significant effect onthe wage rate at the 5%level.

2 F-test: test several restrictions jointly

2.1 Whole model significance

1) **Hypothesis Development**

$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0.$

H1 :$\beta_2 = 0$ *or* $\beta_3 = 0$ *or* $\beta_4 = 0$ *or* $\beta_5 = 0$ *or* $\beta_6 = 0$ *or* $\beta_7 = 0$

In words,

H_0 : none of the explanatory variables has an effect on the wage rate (except the constant term), i.e., every coefficients is null except the intercept.

H_1 : there is at least one explanatory variable that has an effect on the wage rate, i.e., at least one of the coefficients is non-null.

2) **Test statistic and its distribution**

$$F = \frac{\sum (\widehat{y}_i - \overline{y})^2 / (k)}{\sum (y_i - \widehat{y}_i)^2 / ((n-(k+1))} = 261.97 \sim_{H_0} F_{q,n-(k+1)}$$

q =number of tested restrictions=6

n: sample size (4503), k: number of explanatory variables (6).

3) **Decision rule**

3.1) F-value decision rule

We reject H_0 at the level α if $F \geq F_{q,n-(k+1)}, \alpha$ otherwise, we do not reject H_0.

For example, in our case, F_{obs}=261.97>2.01

for $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = 2.01$ then we do reject H_0 at the 5% level, i.e., the model is globally significant.

3.2) p-value decision rule

we reject H_0 at the level α if p-value$<\alpha$ otherwise, we do not reject H_0.

For example, in our case, p-value=" <.0001"<0,05=5% so we reject H_0 at the 5% level

i.e., the model is globally significant.

2.2 Compare significance of different models

Example:

$$(M_6) \quad ltsal_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 female_i + \beta_6 tech_i + \beta_7 enf18_i + u_i,$$

where $u_i \sim \mathcal{N}(0, \sigma^2)$.

$$(M_3) \quad ltsal_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 female_i + u_i.$$

We want to test M_3 vs. M_6.

Anaylysis

Model M_3 is a nested model of model M_6, it means that M_3 is a particular case of M_6. Indeed, M_3 is equivalent to model M_6 under the constraint that $\beta_6 = \beta_7 = 0$.

1) **Hypothesis Development**

$$H_0 : \beta_6 = \beta_7 = 0 \text{ vs. } H_1 : \beta_6 \neq 0 \text{ or } \beta_7 \neq 0$$

2) **Test statistic and its distribution**

$$\frac{\left(RSS^C - RSS^{NC}\right)/q}{RSS^{NC}/(n-(k+1))} \sim_{H_0} F_{q,n-(k+1)}$$

q =number of constraints=2

n: sample size (4503), k: number of explanatory variables (6).

with RSS^C the residual sum of squares from the constrained model (here M_3)

and RSS^{NC} the residual sum of squares from the unconstrained model (here M_6)

3) **Decision rule**

3.1) F-value decision rule

We reject H_0 at the level α if $F \geq F_{q,n-(k+1)}, \alpha$ otherwise, we do not reject H_0.

For example, in our case,

For $\alpha = 5\%$, $F_{q,n-(k+1)}, \alpha = F_{2,4503-7;5\%} = 3$

$$F = \frac{(SCR^C - SCR^{NC})/q}{SCR^{NC}/(n-(k+1))} = \frac{(328.40 - 323.79)/2}{323.79/(4503-7)} = 32.03 > 3$$

Then we reject $H_0 : \beta_6 = \beta_7 = 0$ at the 5% level and we prefer M_6 to M_3.

3.2) p-value decision rule

we reject H_0 at the level α if p-value$<\alpha$ otherwise, we do not reject H_0.

For example, in our case, p-value=" <.0001"<0,05=5% so we reject H_0 at the 5% level

We prefer M6 to M3.

3 conffidence intervals

95% level confidence intervals for the β_j

$$\beta_j \in \left[\widehat{\beta}_j \pm \widehat{\sigma}_{\beta_j} * t_{n-(k+1),\frac{\alpha}{2}}\right] \text{ with } \alpha = 5\%$$

Interpretation : The 95% level confidence interval has 95% chances to contain the true value of the coefficient β_j.

Remark : $\widehat{\beta}_j$ always belong to the CI, by construction !