UNIVERSITY OF OULU

YLEISEN TENTIN TENTTILOMAKE - GENERAL EXAM FORM

Opiskelija täyttää / Student fills in

Opiskelijan nimi / Student name:	**Opiskelijanumero / Student number:**

Opettaja täyttää / Lecturer fills in

Opintojakson koodi / The code of the course: 721066S	
Opintojakson (tentin) nimi / The name of the course or exam: Principles of Econometrics	
Opintopistemäärä / Credit units: 6 Mikäli kyseessä on välikoe, opintopistemääräksi täytetään 0 op. 0 ECTS Credits is used for mid-term exams.	
Tiedekunta / Faculty: Oulun yliopiston kauppakorkeakoulu / Oulu Business School	
Tentin pvm / Date of exam: 2018-12-17	**Tentin kesto tunteina / Exam in hours:** 3 h
Tentaattori(t) / Examiner(s): Sanna Huikari	**Sisäinen postiosoite / Internal address:** 6 OyKKK
Tentissä sallitut apuvälineet / The devices allowed in the exam:	
Funktiolaskin / Scientific calculator, Ohjelmoitava laskin / Programmable calculator Muu tentissä sallittu materiaali tai apuvälineet / Other material or devices, allowed in the exam: Two-sided hand-written A4-sheet, which, NB!, must be included into the answer sheet.	
Muut tenttiä koskevat ohjeet opiskelijalle (esimerkiksi kuinka moneen kysymyksen opiskelijan tulee vastata) / Other instructions for students e.g. how many questions he/she should answer: Answer all the questions. Answer either in English or in Finnish. Remember to justify your conclusion / explain your thinking. No credit for answers if you do not justify your conclusion / explain your thinking.	

The data from a random sample of high school seniors interviewed in 1980 and re-interviewed in 1986 is used in investigating the relationship between the number of completed years of education (*Ed*) for young adults and the distance from each student's high school to the nearest four-year college. (Proximity to college lowers the cost of education, so that students who live closer to a four-year college should, on average, complete more years of higher education.) Table 1 contains results from six estimated regressions. The dependent variable in regressions (1), (2), (3), (5), and (6) is the number of completed years of education (*Ed*) and in regression (4) its logarithmic transformation. All estimations were made by using the heteroskedasticity-robust standard errors. (The estimated standard errors are given in Table 1 in parentheses under the values of estimated coefficients.) The independent variables are:

Name	Description
Ed	Years of Education Completed
Female	1 = Female / 0 = Male
Black	1 = Black / 0 = Not-Black
Hispanic	1 = Hispanic / 0 = Not-Hispanic
Bytest	Base Year Composite Test Score. (These are achievement tests given to high school seniors in the sample)
Dadcoll	1 = Father is a College Graduate / 0 = Father is not a College Graduate
Momcoll	1 = Mother is a College Graduate / 0 = Mother is not a College Graduate
Incomehi	1 = Family Income > $25,000 per year / 0 = Income ≤ $25,000 per year.
Ownhome	1= Family Owns Home / 0 = Family Does not Own Home
Cue80	County Unemployment rate in 1980 (%)
Stwmfg80	State Hourly Wage in Manufacturing in 1980 ($)
Dist	Distance from 4yr College in 10's of miles
Tuition	Avg. State 4yr College Tuition in $1000's

1. (32 p.) Use the results in Table 1 to answer the following questions. Remember to show your work. No credit for answers if you do not show your work.

a) An education advocacy group argues that on average, a person's educational attainment would increase by approximately 0.15 year if distance to the nearest college is decreased by 20 miles. Is the advocacy groups' claim consistent with the estimated regression in (1)? (2 p.)

b) According to results from regression (2), what is the effect of distance to the nearest college on completed years of education? Is the effect statistically significant? What is approximately the *p*-value associated with coefficient's *t*-statistic? (3 p.)

c) Construct a 95% confidence interval for the slope coefficient of *Dist* in regression (2). (2 p.)

d) Compare the results of regressions (2) and (3). Does regression (2) seem to suffer from important omitted variable bias? (2 p.)

e) Compare the fit of the regressions (2) and (3). Which one is preferred? (2 p.)

f) Explain what is the reasoning behind including *Cue80* and *Swmfg80* eg in the regression (3). Are the signs of their estimated coefficients (+ or -) what you would have believed? Interpret the magnitudes of these coefficients. (2 p.)

g) Bob is a black male. His high school was 20 miles from the nearest college. His base-year composite test score (*Bytest*) was 58 and average state 4-year college tuition was $950. His family income in 1980 was $26,000, and his family owned a home. His mother attended college, but his father did not. The unemployment rate in his county was 7.5%, and the state average manufacturing hourly wage was $9.75. Predict Bob's years of completed schooling using the regression (3). (2 p.)

h) According to results from regression (3), how are years of education expected to change if *Dist* increases from 2 to 3 (that is 20 to 30 miles)? How are years of education expected to change if *Dist* increases from 6 to 7 (that is 60 to 70 miles)? (2 p.)

i) According to results from regression (4), how are years of education expected to change if *Dist* increases from 2 to 3 (that is 20 to 30 miles)? How are years of education expected to change if *Dist* increases from 6 to 7 (that is 60 to 70 miles)? (2 p.)

j) According to results from regression (5), how are years of education expected to change if *Dist* increases from 2 to 3 (that is 20 to 30 miles)? How are years of education expected to change if *Dist* increases from 6 to 7 (that is 60 to 70 miles)? (2 p.)

k) Is the regression (5) preferred to the regression (3)? Explain. (2 p.)

l) What does the term *DadColl* x *MomColl* in regressions (5) and (6) measure? (2 p.)

m) Mary, Jane, Alexis and Bonnie have the same values of *Dist*, *Bytest*, *Tuition*, *Female*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *Cue80* and *Stwmfg80*. Neither of Mary's parents attended college. Jane's father attended college, but her mother did not. Alexis's mother attended college, but her father did not. Both of Bonnie's parents attended college. Answer the following questions using the regression (5):

i. What does the regression predict for the difference between Jane's and Mary's years of education? (1 point)

ii. What does the regression predict for the difference between Alexis's and Mary's years of education? (1 point)

iii. What does the regression predict for the difference between Bonnie's and Mary's years of education? (1 point)

n) It has been argued that, controlling for other factors, black people and Hispanics complete more college than white people. Is this result consistent with the regression that is estimated in (3)? (2 p.)

o) Is there any evidence in regressions (1)-(6) that the effect of *Dist* on *ED* depends on the family's income? (2 p.)

2. Internal and external validity (8 p.)

a) Discuss briefly the internal validity of the regression results in Question 1. (6 p.)

b) The data set which was used in Question 1 excluded students from western states. How could you investigate the external validity of the conclusions made in Question 1? To what kind of populations and settings could inferences and conclusions made in Question 1 be possibly generalized? (2 p.)

Table 1.

Regressor	(1) *ED*	(2) *ED*	(3) *ED*	(4) *ln(ED)*	(5) *ED*	(6) *ED*
Dist	–0.073 (0.013)	–0.043 (0.012)	–0.037 (0.012)	–0.0026 (0.0009)	–0.081 (0.025)	–0.110 (0.028)
$Dist^2$	–	–	–	–	0.0047 (0.0021)	0.0065 (0.0022)
Tuition	–	–	–0.191 (0.099)	–0.014 (0.007)	–0.194 (0.099)	–0.210 (0.099)
Female	–	0.122 (0.052)	0.143 (0.050)	0.010 (0.004)	0.141 (0.050)	0.141 (0.050)
Black	–	0.278 (0.067)	0.351 (0.067)	0.026 (0.005)	0.331 (0.068)	0.333 (0.068)
Hispanic	–	0.317 (0.074)	0.362 (0.076)	0.026 (0.005)	0.329 (0.078)	0.323 (0.078)
Bytest	–	–	0.093 (0.003)	0.0067 (0.0002)	0.093 (0.003)	0.093 (0.003)
Incomehi	–	–	0.372 (0.062)	0.027 (0.004)	0.362 (0.062)	0.217 (0.090)
Ownhome	–	–	0.139 (0.065)	0.010 (0.005)	0.141 (0.065)	0.144 (0.065)
DadColl	–	–	0.571 (0.076)	0.041 (0.005)	0.654 (0.087)	0.663 (0.087)
MomColl	–	–	0.378 (0.083)	0.027 (0.006)	0.569 (0.122)	0.567 (0.122)
DadColl × *MomColl*	–	–	–	–	–0.366 (0.164)	–0.356 (0.164)
Cue80	–	–	0.029 (0.010)	0.002 (0.0007)	0.026 (0.010)	0.026 (0.010)
Stwmfg	–	–	–0.043 (0.020)	–0.003 (0.001)	–0.042 (0.020)	–0.042 (0.020)
Incomehi × *Dist*	–	–	–	–	–	0.124 (0.062)
Incomehi × $Dist^2$	–	–	–	–	–	–0.0087 (0.0062)
Intercept	13.960 (0.038)	8.465 (0.159)	8.920 (0.243)	2.266 (0.017)	9.002 (0.250)	9.042 (0.251)
F-statistics and p-values on joint hypotheses						
H_0: *Dist=0 and* $Dist^2=0$	–	–	–	–	6.00 (0.003)	8.35 (0.000)
H_0: *Incomehi* × *Dist=0* and *Incomehi* × $Dist^2=0$	–	–	–	–	–	2.34 (0.096)
SER	1.808	1.586	1.538	0.109	1.536	1.536
$\overline{R}^2$	0.007	0.235	0.281	0.283	0.283	0.283
N	3796	3796	3796	3796	3796	3796

Large-Sample Critical Values for the *t*-statistic from the Standard Normal Distibution

	Significance Level		
	10%	5%	1%
2-Sided Test (≠)			
Reject if $\|t\|$ is greater than	1.64	1.96	2.58
1-Sided Test (>)			
Reject if t is greater than	1.28	1.64	2.33
1-Sided Test (<)			
Reject if t is less than	−1.28	−1.64	−2.33

TABLE 3 Critical Values for the χ^2 Distribution

	Significance Level		
Degrees of Freedom	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48

Large-Sample Critical Values for the *F*-statistic from the $F_{m,\infty}$ Distribution

Reject if $F >$ Critical Value

	Significance Level		
Degrees of Freedom (m)	10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64