

Tentin päivämäärä / Date of exam: 7.11.2013	
Opintojakson koodi, nimi ja tentin numero / The code and the name of the course and number of the exam: 721060A Ekonometrian perusteet / Introduction to Econometrics (1)	
Tentaattori(t)/ Examiner(s): Sanna Huikari	
Sallitut apuvälineet / The devices allowed in the exam: <input checked="" type="checkbox"/> Laskin (ei graafinen/ohjelmoitava)/Calculator (not graphic, programmable) <input type="checkbox"/> Sanakirja/Dictionary <input type="checkbox"/> Muu materiaali, tarkennettu alla/Other material, specified below	
Tenttiin vastaaminen / Please answer the questions <input checked="" type="checkbox"/> suomeksi/ in Finnish	<input checked="" type="checkbox"/> englanniksi/ in English NB: Only if your first language is not Finnish (e.g. Exchange students)
Kysymyspaperi on palautettava / Paper with exam questions must be returned: <input type="checkbox"/> Kyllä/Yes <input checked="" type="checkbox"/> Ei/No	

1. Vastaa tai selitä lyhyesti (6 p.)

- Ekonometria
- Mitä hyötyä suurten lukujen laista ja keskeisestä raja-arvolauseesta on ekonometriassa?
- Dummy-muuttuja
- Mihin instrumenttimuuttuja-menetelmää käytetään? Mitkä kaksi ehtoa instrumenttimuuttujan on täytettävä, että se on validi?

2. Kolikkoa heitetään kaksi kertaa. Olkoon Y kruunujen lukumäärä. (3 p.)

- Piirrä Y :n tiheys(todennäköisyys)funktio.
- Laske Y :n odotusarvo ja varianssi.

3. Vastaa seuraaviin kysymyksiin. (6 p.)

- Sinua pyydetään tutkimaan työntekijöiden kouluttamisen (mitattuna koulutukseen käytettyinä tunteina työntekijää kohden viikossa) vaikutusta työntekijöiden tuottavuuteen (tuotos työntekijää kohden tunnissa) eräissä tuotantolaitoksessa. Kuvaile:
 - Millainen olisi ideaali satunnaistettu koejärjestely, jolla voisi mainittua vaikutusta mitata?
 - Millaisella poikkileikkaushavaintoaineistolla voisit tutkia mainittua vaikutusta?
- Hahmottele regressiofunktion kuvaaja, joka on kasvava (positiivinen kulmakerroin), ja joka on jyrkempi pienemmille X :n arvoille ja loivempi suuremmille X :n arvoille. Millaisella epälineaarisella regressiolla voisit tällaista muotoa mallintaa? Anna esimerkki taloudellisesta relaatiosta, jolla mahdollisesti voisi olla tällainen muoto.

4. (9 p.) Vuonna 1980 suoritettiin kyselytutkimus 3796 lukiolaiselle (*am. high school seniors*), jotka on valittu tutkimukseen satunnaisotannalla. Samalle joukolle on suoritettu kyselytutkimus uudestaan vuonna 1986.

Käyttämällä saatua aineistoa on haluttu tutkia sitä, kuinka opiskelijan lukion ja lähimmän 4-vuotisen collegen (korkeakoulu, jossa voi suorittaa alemman tutkinnon) välinen etäisyys vaikuttaa nuoren aikuisen kouluttautumiseen. (Collegen läheisyys alentaa kouluttautumisen kustannuksia, joten opiskelijoiden, jotka asuvat lähempänä 4-vuotista collegea pitäisi keskimäärin suorittaa enemmän ylemmän asteen koulutusta.)

Taulukko 1 sivulla 4 sisältää viiden estimoidun regression tulokset. Selitettävä muuttuja on regressioissa (1), (3), (4) ja (5) kouluttautumiseen käytetyt vuodet (*ED*) ja regressiossa (2) sama muuttuja logaritmoituna. Kaikki estimoinnit on suoritettu käyttäen heteroskedastisuus-robusteja keskivirheitä. Selittävien muuttujien kuvaukset:

Nimi	Kuvaus
ED	Koulutusvuodet
Female	1 = Nainen / 0 = Mies
Black	1 = Musta / 0 = Ei-Musta
Hispanic	1 = Latino / 0 = Ei-Latino
Bytest	Lukiolaisille suoritetun tasokokeen pistemäärä
DadColl	1 = Isä on valmistunut collegesta / 0 = Isä ei ole valmistunut collegesta
MomColl	1 = Äiti on valmistunut collegesta / 0 = Äiti ei ole valmistunut collegesta
Incomehi	1 = Perheen tulot > \$25,000 vuodessa / 0 = Perheen tulot ≤ \$25,000 vuodessa
Ownhome	1 = Perheellä omistusasunto / 0 = Perheellä ei omistusasuntoa
Urban	1 = Koulu sijaitsee kaupunkialueella / 0 = Koulu ei sijaitse kaupunkialueella
Cue80	Piirikunnan työttömyysaste vuonna 1980
Stwmfg80	Osavaltion keskimääräinen teollisuudessa työskentelevien tuntipalkka vuonna 1980
Dist	Etäisyys 4-vuotiseen collegen (kymmenissä mailleissa)
Tuition	4-vuotisen collegen keskimäär. lukuvuosimaksu osavaltiossa (tuhansissa dollareissa)

- a) Regression (1) tulosten mukaan, kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kahdesta kolmeen (eli etäisyys muuttuu 20 mailista 30 mailiin)? Kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kuudesta seitsemään (eli etäisyys muuttuu 60 mailista 70 mailiin)?
- b) Regression (2) tulosten mukaan, kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kahdesta kolmeen (eli etäisyys muuttuu 20 mailista 30 mailiin)? Kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kuudesta seitsemään (eli etäisyys muuttuu 60 mailista 70 mailiin)?
- c) Regression (3) tulosten mukaan, kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kahdesta kolmeen (eli etäisyys muuttuu 20 mailista 30 mailiin)? Kuinka koulutukseen käytettyjen vuosien voidaan keskimäärin odottaa muuttuvan, jos *Dist*-muuttujan arvo muuttuu kuudesta seitsemään (eli etäisyys muuttuu 60 mailista 70 mailiin)?
- d) Onko regressiomalli (1) mielestäsi parempi kuin regressiomalli (3)? Perustele.

- e) Mitä vuorovaikutustermi *DadColl* x *MomColl* regressioissa (4) ja (5) mittaa? Mikä ajatus on ollut sen taustalla, että termi on haluttu lisätä malliin?
- f) Muuttujat *Dist*, *Bytest*, *Tuition*, *Female*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *Cue80* ja *Stwmfg80* saavat samat arvot Marylle, Janelle, Alexisille ja Bonnielle. Kumpikaan Maryn vanhemmista ei ole suorittanut collegea. Janen isä on suorittanut collegen, mutta äiti ei. Alexisin äiti on suorittanut collegen, mutta isä ei. Bonnien molemmat vanhemmat ovat suorittaneet collegen. Käytä regression (4) tuloksia vastataksesi seuraaviin kysymyksiin:
- Mikä on regressiotulosten antama ennuste Janen ja Maryn koulutusvuosien erolle?
 - Mikä on regressiotulosten antama ennuste Alexisin ja Maryn koulutusvuosien erolle?
 - Mikä on regressiotulosten antama ennuste Bonnien ja Maryn koulutusvuosien erolle?
- g) Onko regressiotuloksista löydettävissä tukea sille ajatukselle, että se, kuinka etäisyys (*Dist*) vaikuttaa koulutusvuosiin (*ED*) riippuisi perheen tuloista?
- h) Esitä lyhyt yhteenveto etäisyyden (*Dist*) vaikutuksesta koulutusvuosiin (*ED*) perustuen estimointituloksiin.

5. Sisäinen ja ulkoinen validiteetti (6 p.)

- Arvioi niiden mallien sisäistä validiteettia, joita käytit vastatessasi kysymykseen 4. h).
- Tehtävän 4 havaintoaineisto ei sisällä havaintoja Yhdysvaltojen läntisistä osavaltioista. Kuinka tutkisit kohdassa 4 h.) tekemiesi päätelmien ulkoista validiteettia, jos sinulla olisi vastaavanlainen havaintoaineisto käytettävissä myös läntisistä osavaltioista?

Taulukko 1. / Table 1.

<i>Selittäjä / Regressor</i>	(1) <i>ED</i>	(2) <i>ln(ED)</i>	(3) <i>ED</i>	(4) <i>ED</i>	(5) <i>ED</i>
<i>Dist</i>	-0.037** (0.012)	-0.0026** (0.0009)	-0.081** (0.025)	-0.081** (0.025)	-0.110** (0.028)
<i>Dist</i> ²	-	-	0.0046* (0.0021)	0.0047* (0.0021)	0.0065* (0.0022)
<i>Tuition</i>	-0.191 (0.099)	-0.014* (0.007)	-0.193* (0.099)	-0.194* (0.099)	-0.210* (0.099)
<i>Female</i>	0.143** (0.050)	0.010** (0.004)	0.143** (0.050)	0.141** (0.050)	0.141** (0.050)
<i>Black</i>	0.351** (0.067)	0.026** (0.005)	0.334** (0.068)	0.331** (0.068)	0.333** (0.068)
<i>Hispanic</i>	0.362** (0.076)	0.026** (0.005)	0.333** (0.078)	0.329** (0.078)	0.323** (0.078)
<i>Bytest</i>	0.093** (0.003)	0.0067** (0.0002)	0.093** (0.003)	0.093** (0.003)	0.093** (0.003)
<i>Incomehi</i>	0.372** (0.062)	0.027** (0.004)	0.369** (0.062)	0.362** (0.062)	0.217* (0.090)
<i>Ownhome</i>	0.139* (0.065)	0.010* (0.005)	0.143* (0.065)	0.141* (0.065)	0.144* (0.065)
<i>DadColl</i>	0.571** (0.076)	0.041** (0.005)	0.561** (0.077)	0.654** (0.087)	0.663** (0.087)
<i>MomColl</i>	0.378** (0.083)	0.027** (0.006)	0.378** (0.083)	0.569** (0.122)	0.567** (0.122)
<i>DadColl × MomColl</i>	-	-	-	-0.366* (0.164)	-0.356* (0.164)
<i>Cue80</i>	0.029** (0.010)	0.002** (0.0007)	0.026** (0.010)	0.026** (0.010)	0.026** (0.010)
<i>Stwmfg</i>	-0.043* (0.020)	-0.003* (0.001)	-0.043* (0.020)	-0.042* (0.020)	-0.042* (0.020)
<i>Incomehi × Dist</i>	-	-	-	-	0.124* (0.062)
<i>Incomehi × Dist</i> ²	-	-	-	-	-0.0087 (0.0062)
<i>Vakio / Intercept</i>	8.920** (0.243)	2.266** (0.017)	9.012** (0.250)	9.002** (0.250)	9.042** (0.251)

F-testisuure ja yhteishypoteesien p-arvot /

F-statistic and p-values on joint hypotheses

(a) <i>Dist ja/and Dist</i> ²	-	-	6.08 (0.002)	6.00 (0.003)	8.35 (0.000)
(b) Vuorovaikutustermit / Interaction terms <i>Incomehi × Dist ja/and Incomehi × Dist</i> ²	-	-	-	-	2.34 (0.096)
<i>SER</i>	1.538	0.109	1.537	1.536	1.536
\bar{R}^2	0.281	0.283	0.282	0.283	0.283
<i>N</i>	3796	3796	3796	3796	3796

Tilastollisesti merkitsevä * = 5 %:n ja ** = 1 %:n merkitsevyytasolla.

Significant at the * = 5% and ** = 1% significance level.

Questions in English:

1. Answer or shortly define the meaning. (6 p.)

- a) Econometrics
- b) What role does the law of large numbers and the central limit theorem play in econometrics?
- c) Dummy variable
- d) What is the use of instrumental variables regression? What are the two conditions that must hold for an instrument to be valid?

2. Let Y denote the number of "heads" that occur when two coins are tossed. (3 p.)

- a) Derive the probability distribution of Y .
- b) Derive the mean and variance of Y .

3. Answer the following questions. (6 p.)

- a) You are asked to study the causal effect of hours spent on employee training (measured in hours per worker per week) in a manufacturing plant on the productivity of its workers (output per worker per hour). Describe:
 - i) An ideal randomized controlled experiment to measure this causal effect.
 - ii) An observational cross-sectional data set with which you could study this effect.
- b) Sketch a regression function that is increasing (has a positive slope) and is steeper for small values of X but less steep for large values of X . Explain how you would specify a nonlinear regression to model this shape. Can you think of an economic relationship with a shape like this?

4. (9 p.) The data from a random sample of high school seniors interviewed in 1980 and re-interviewed in 1986 was 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 in page 4 contains results from five estimated regressions. The dependent variable in regressions (1), (3), (4) and (5) is the number of completed years of education (ED) and in regression (2) its logarithmic transformation. All estimations were made by using the heteroskedasticity-robust standard errors. 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
Urban	1 = School in Urban Area / 0 = School not in Urban Area
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

- a) According to results from regression (1), 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)?
- b) According to results from regression (2), 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)?
- c) 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)?
- d) Do you prefer the regression (3) to the regression (1)? Explain.
- e) What does the interaction term $DadColl \times MomColl$ in regressions (4) and (5) measure? What is the intuition behind adding that variable in regression?
- f) 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 (4):
- What does the regression predict for the difference between Jane's and Mary's years of education?
 - What does the regression predict for the difference between Alexis's and Mary's years of education?
 - What does the regression predict for the difference between Bonnie's and Mary's years of education?
- g) Is there any evidence that the effect of Dist on ED depends on the family's income?
- h) Summarize the effect of Dist on years of education based on estimated regressions.

5. Internal and external validity (6 p.)

- a) Discuss the internal validity of the regressions that you used to answer Question 4. h).
- b) The data set which was used in Question 4 excluded students from western states. How would you investigate the (geographic) external validity of the conclusions that you reached in Question 4. h), if the data from western states was available?