

Institutt for samfunnsøkonomi

Eksamensoppgave i SØK2900 – Empirisk metode Empirical methods in Economics

Faglig kontakt under eksamen: Lars-Erik Borge

Tlf.: 73 59 19 41

Eksamensdato: 5. juni 2013

Eksamenstid: 4 timer

Sensurdato: 26. juni 2013

Tillatte hjelpemidler: Flg formelsamling: Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske. Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin.
Enkel kalkulator Citizen SR-270x, HP 30S eller SR-270X College

Annen informasjon: Eksamensoppgaven består av 5 oppgaver med delspørsmål som alle skal besvares.

Målform/språk: Bokmål, nynorsk og engelsk

Antall sider: 10 (inkl. forside)

Antall sider vedlegg: 3

Oppgave 1 (25 %)

En fotballklubb eksperimenterer med ulike billettpriser for å undersøke sammenhengen mellom billettpris og antall tilskuere. Observasjonene for de 10 hjemmekampene i sesongen 2012 var som i tabellen under.

Billettpris (kroner), X_i	Antall tilskuere, Y_i
150	12400
200	11100
100	13200
175	12000
125	14300
150	11900
175	11000
250	10000
225	10500
175	11500

Basert på disse observasjonene kan vi beregne følgende:

$$\sum X_i = 1725, \sum Y_i = 117900, \sum (X_i - \bar{X})^2 = 18062.5, \sum (Y_i - \bar{Y})^2 = 14769000, \sum (X_i - \bar{X})(Y_i - \bar{Y}) = -465250$$

- Bruk informasjonen over til å beregne gjennomsnitt, median, varians og standardavvik for billettpris og antall tilskuere.
- Beregn kovarians og korrelasjon mellom billettpris og antall tilskuere. Hva gir kovarians og korrelasjon uttrykk for?
- Økonomisjefen i fotballklubben benytter dataene over til å estimere regresjonsmodellen $Y = a + bX + \varepsilon$, der a og b er parametere og ε er et stokastisk restledd. Estimer a og b ved bruk av minste kvadraters metode og finn modellens forklaringskraft.
- Økonomisjefen ønsker å undersøke påstanden om at økt billettpris reduserer antall tilskuere. Formuler og gjennomfør en test av denne påstanden. Presiser forutsetningene du baserer testen på og bruk et signifikansnivå på 5 prosent.

Oppgave 2 (25 %)

I januar 2012 var gjennomsnittlig boligformue for alle husholdninger i Trondheim 2,5 millioner kroner. For å undersøke om boligformuen hadde økt ble det i januar 2013 gjennomført en utvalgsundersøkelse blant 300 husholdninger.

- Beskriv hvordan du vil estimere populasjonsgjennomsnittet basert på informasjon om boligformue fra utvalgsundersøkelsen. Hvilke egenskaper har den foreslåtte estimatoren?

- b) Du får vite at gjennomsnittlig boligformue i utvalget var 2,75 millioner kroner og at standardavviket var 1,25 millioner kroner. Gir utvalgsundersøkelsen grunnlag for å si at gjennomsnittlig boligformue har økt fra januar 2012 til januar 2103? Bruk et signifikansnivå på 1 prosent. Presiser forutsetningene bak testen.
- c) Inflasjonen fra januar 2012 til januar 2013 var 3 prosent. Gir utvalgsundersøkelsen grunnlag for å si at realverdien av boligformuen har økt fra januar 2012 til januar 2013?
- d) I januar 2013 gjennomføres det en tilsvarende spørreundersøkelse blant 400 husholdninger i Bergen. Gjennomsnittlig boligformue i utvalget var 3 millioner kroner og standardavviket var 1,5 millioner kroner. Gir de to utvalgsundersøkelsene grunnlag for å si at gjennomsnittlig boligformue er høyere blant husholdningene i Bergen enn blant husholdningene i Trondheim? Bruk et signifikansnivå på 1 prosent. Presiser forutsetningene bak testen.

Oppgave 3 (20 %)

I en empirisk analyse av sammenhengen mellom lærerdekning og arbeidsledighet rapporteres følgende estimerte modell

$$(1) \bar{Y}_t = 4,66 - 0,757X_t, R^2 = 0,48$$

(0,46) (0,148)

Der Y_t er andelen lærere med formell utdanning i prosent av alle ansatte lærer i år t , X_t er arbeidsledighetsraten i prosent. Tall i parenteser under estimerte parametre er estimerte standardavvik og R^2 er multipl determinasjonskoeffisient. Modellen er estimert ved minste kvadraters metode basert på data for 30 år.

- a) Presiser og test hypotesen: «Arbeidsledigheten har ingen betydning for lærerdekningen».

I en alternativ spesifisering rapporteres den estimerte likningen:

$$(2) \bar{Y}_t = 4,51 - 2,231 \ln X_t$$

- b) Finn predikert endring i Y når arbeidsledigheten øker fra 2 prosent til 3 prosent ved bruk av resultatene rapportert i likning (1) og (2). Finn de tilsvarende predikerte endringene av en økning i ledighetsraten fra 5 prosent til 6 prosent.

Oppgave 4 (20 %)

I en empirisk analyse av etterspørselen etter kulturtjenester estimeres følgende modell:

$$(1) \ln Y_i = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 X_{i2} + \delta_0 D_i + \delta_1 D_i \ln X_{i1} + u_i$$

Der Y_i er utgifter til kultur for individ i , X_{i1} er individets inntekt, X_{i2} er individets utdanning målt i antall år. D_i er en dummyvariabel lik 1 hvis individet bor i en by, null hvis individet bor på landet. u_i er et stokastisk restledd.

- a) Tolk parametrene som inngår i (1). Forklar spesielt hvordan modellformuleringen tar hensyn til mulige forskjeller i etterspørselen mellom individer som bor i by og individer som bor på landet.

Modellen estimeres med minste kvadraters metode ved bruk av et datasett med observasjoner for 100 individer. Resultatene er gitt i likning (2):

$$(2) \widehat{\ln Y}_i = 0,266 + 1,109 \ln X_{i1} + 0,051 X_{i2} + 0,253 D_i + 0,145 D_i \ln X_{i1}$$

$$(0,086) \quad (0,039) \quad (0,054) \quad (0,0458) \quad (0,118)$$

$$R^2=0,788, \quad SSR = 418,07$$

Tall i parentes under estimerte parametre er estimerte standardavvik, R^2 er multiplisert determinasjonskoeffisient og SSR er summen av kvadrerte residualer.

- b) Forklar hvordan du kan teste om parameteren $\beta=1$ og test hypotesen ved bruk av resultatene gitt i likning (2).
- c) Du vil teste om etterspørselsfunksjonen er den samme for individer som bor i by og individer som bor på landet. Formuler nullhypotesen og forklar hvordan du vil teste denne. Bruk resultatene i likning (2) og i likning (3) gitt under til å teste hypotesen:

$$(3) \widehat{\ln Y}_i = 0,173 + 1,220 \ln X_{i1} + 0,071 X_{i2}$$

$$(0,085) \quad (0,033) \quad (0,051)$$

$$R^2=0,771, \quad SSR = 451,20$$

Oppgave 5 (10 %)

En kommentator kommer med følgende påstand om resultatene fra en analyse du har gjennomført: «Det er så store målefeil i de variablene som er brukt i analysen at koeffisientestimatene er systematisk feil». Kommenter dette utsagnet.

Nynorsk*Oppgåve 1 (25 %)*

Ein fotballklubb prøver ut ulike billettprisar for å undersøke samanhengen mellom billettpris og talet på tilskodarar. Observasjonane for dei 10 heimekampane i sesongen 2012 var som i tabellen under.

Billettpris (kroner), X_i	Talet på tilskodarar, Y_i
150	12400
200	11100
100	13200
175	12000
125	14300
150	11900
175	11000
250	10000
225	10500
175	11500

Basert på disse observasjonane kan vi berekne følgjande:

$$\sum X_i = 1725, \sum Y_i = 117900, \sum (X_i - \bar{X})^2 = 18062.5, \sum (Y_i - \bar{Y})^2 = 14769000, \sum (X_i - \bar{X})(Y_i - \bar{Y}) = -465250$$

- Bruk informasjonen over til å berekne gjennomsnitt, median, varians og standardavvik for billettpris og talet på tilskodarar.
- Berekn kovarians og korrelasjon mellom billettpris og talet på tilskodarar. Kva gjer kovarians og korrelasjon uttrykk for?
- Økonomisjefen i fotballklubben bruker informasjonen over til å estimere regresjonsmodellen $Y = a + bX + \varepsilon$, der a og b er parameterar og ε er eit tilfeldig restledd. Estimer a og b ved bruk av minste kvadraters metode og finn modellens forklaringskraft.
- Økonomisjefen ynskjer å undersøke påstanden om at økt billettpris reduserer talet på tilskodarar. Formuler og gjennomfør ein test av denne påstanden. Presiser føresetnadane du baserer testen på og bruk eit signifikansnivå på 5 prosent.

Oppgåve 2 (25 %)

I januar 2012 var gjennomsnittlig bustadformue for alle hushald i Trondheim 2,5 millionar kroner. For å undersøke om bustadformua hadde økt ble det i januar 2013 gjennomført ei utvalsundersøking blant 300 hushald.

- a) Gjer greie for korleis du vil estimere populasjonsgjennomsnittet basert på informasjon om bustadformue frå utvalsundersøkinga. Kva for eigenskaper har den føreslåtte estimatoren?
- b) Du får vite at gjennomsnittlig bustadformue i utvalet var 2,75 millionar kronar og at standardavviket var 1,25 millionar kronar. Gir utvalsundersøkinga grunnlag for å seie at gjennomsnittleg bustadformue økte frå januar 2012 til januar 2103? Bruk eit signifikansnivå på 1 prosent. Presiser føresetnadene bak testen.
- c) Inflasjonen frå januar 2012 til januar 2013 var 3 prosent. Gir utvalsundersøkinga grunnlag for å seie at realverdien av bustadformua økte frå januar 2012 til januar 2013?
- d) I januar 2013 utføres det ei tilsvarande undersøking blant 400 hushald i Bergen. Gjennomsnittleg bustadformue i utvalet var 3 millionar kronar og standardavviket var 1,5 millionar kronar. Gir de to utvalsundersøkingane grunnlag for å seie at gjennomsnittleg bustadformue er høgare blant hushalda i Bergen enn blant hushalda i Trondheim? Bruk eit signifikansnivå på 1 prosent. Presiser føresetnadane bak testen.

Oppgave 3 (20 %)

I ein empirisk analyse av samanhengen mellom lærardekning og arbeidsløyse blir fylgjande modell estimert:

$$(1) \hat{Y}_t = 4,66 - 0,757X_t, R^2 = 0,48$$

(0,46) (0,148)

Der Y_t er andelen lærarar med formell utdanning i prosent av alle tilsette lærarar i år t , X_t er arbeidsløysa i prosent. Tall i parentesar under estimerte parametrar er estimerte standardavvik og R^2 er multippel determinasjonskoeffisient. Modellen er estimert ved minste kvadraters metode basert på data for t alt 30 år.

- a) Presiser og test hypotesen: «Arbeidsløysa har ingen betydning for lærardekninga».

I ein alternativ spesifikasjon blir fylgjande modell estimert:

$$(2) \hat{Y}_t = 4,51 - 2,231 \ln X_t,$$

- b) Finn predikert endring i Y når arbeidsløysa aukar frå 2 prosent til 3 prosent ved bruk av resultat i likning (1) og (2). Finn tilsvarande predikerte endringar av ei auke i arbeidsløysa frå 5 prosent til 6 prosent.

Oppgave 4 (20 %)

I ein empirisk analyse av etterspurnaden etter kulturtenestar blir fylgjande modell estimert:

$$(1) \ln Y_i = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 X_{i2} + \delta_0 D_i + \delta_1 D_i \ln X_{i1} + u_i$$

Der Y_i er utgifter til kultur for individ i , X_{i1} er individets inntekt, X_{i2} er individets utdanning målt i år. D_i er en dummyvariabel lik 1 dersom individet bor i ein by, null dersom individet bor på landet. u_i er et stokastisk restledd.

a) Tolk parametrane som inngår i (1). Forklar spesielt korelis modellformuleringa tar omsyn til mulige skinader i etterspurnaden mellom individ som bor i by og individ som bor på landet.

Modellen estimeres med minste kvadraters metode ved bruk av et datasett med observasjonar for 100 individ. Resultata er vist i likning (2):

$$(2) \widehat{\ln Y_i} = 0,266 + 1,109 \ln X_{i1} + 0,051 X_{i2} + 0,253 D_i + 0,145 D_i \ln X_{i1}$$

(0,086) (0,039) (0,054) (0,0458) (0,118)

$$R^2=0,788, \text{ SSR} = 418,07$$

Tal i parentes under estimerte parametrar er estimerte standardavvik, R^2 er multiplert determinasjonskoeffisient og SSR er summen av kvadrerte residualer.

- b) Forklar korleis du kan teste om parameteren $\beta=1$ og test hypotesen ved bruk av resultata i likning (2).
- c) Du vil teste om etterspurnadsfunksjonen er den same for individ som bor i by og individ som bor på landet. Formulert nullhypotesen og forklar korleis du vil teste denne. Bruk resultata i likning (2) og i likning (3) gitt under til å teste hypotesen:

$$(3) \widehat{\ln Y_i} = 0,173 + 1,220 \ln X_{i1} + 0,071 X_{i2}$$

(0,085) (0,033) (0,051)

$$R^2=0,771, \text{ SSR} = 451,20$$

Oppgåve 5 (10 %)

Ein kommentator kommer med fylgjande påstand om resultata frå ein analyse du har utført: «Det er så store målefeil i variablane som er nytta i analysen at koeffisientestimata er systematisk feil». Kommenter påstanden.

English*Exercise 1 (25 %)*

A football club is experimenting with different ticket prices in order to investigate the relationship between and attendance. The observations for the 10 home games in the 2012 are reported in the table below.

Ticket price (kroner), X_i	Attendance, Y_i
150	12400
200	11100
100	13200
175	12000
125	14300
150	11900
175	11000
250	10000
225	10500
175	11500

From these observations the following can be calculated:

$$\sum X_i = 1725, \sum Y_i = 117900, \sum (X_i - \bar{X})^2 = 18062.5, \sum (Y_i - \bar{Y})^2 = 14769000, \sum (X_i - \bar{X})(Y_i - \bar{Y}) = -465250$$

- Use the information above to calculate mean, median, variance and standard deviation for ticket price and attendance.
- Calculate the covariance and the correlation coefficient between ticket price and attendance. What do covariance and coefficient of correlation measure?
- The financial director of the club uses the data above to estimate the regression model $Y = a + bX + \varepsilon$, where a and b are parameters and ε is an error term (disturbance). Estimate a and b using ordinary least squares (OLS) and calculate the coefficient of determination (goodness of fit).
- The financial director of the clubs wants to investigate whether a higher ticket price reduces attendance. Formulate and conduct a test of this claim. State the assumptions and use 5 percent as level of significance.

Exercise 2 (25 %)

In January 2012 the average housing wealth of households in Trondheim was 2.5 million kroner. In January 2013 a survey of 300 households was conducted to investigate whether the housing wealth had increased.

- a) Describe how the population average may be estimated based on information about housing wealth from the survey. What are the properties of the suggested estimator?
- b) The average housing wealth in the 2013 survey was 2.75 million kroner with a standard error of 1.25 million kroner. Can you conclude that average housing wealth in the population increased from January 2012 to January 2013? Use 1 percent as level of significance and state the assumptions for the test.
- c) The rate of inflation from January 2012 to January 2013 was 3 percent. Can you (based on the survey) conclude that real housing wealth increased from January 2012 to January 2013?
- d) In January 2013 a similar survey was conducted among 400 households in Bergen. In the sample average housing wealth was 3 million kroner with a standard deviation of 1.5 million kroner. Can you (based on the two surveys) conclude that average housing wealth is higher among households in Bergen than among households in Trondheim? Use 1 percent as level of significance and state the assumptions for the test.

Exercise 3 (20 %)

As part of a study of the relationship between teacher shortage and the rate of unemployment the following equation is estimated:

$$(1) \hat{Y}_t = 4,66 - 0,757X_t, R^2 = 0,48$$

(0,46) (0,148)

Y_t is the share of certified teachers in percent of all teachers in year t , while X_t is the rate of unemployment measured in percent. The figures in parentheses below the estimated parameters are estimated and standard errors and R^2 is the coefficient of multiple determination. The model estimated by OLS using data for 30 years.

- a) Specify and test the hypothesis: «The rate of unemployment has no effect on teacher shortage».

An alternative specification is the following:

$$(2) \hat{Y}_t = 4,51 - 2,231 \ln X_t$$

- b) Calculate the predicted change in Y when the rate of unemployment increases from 2 percent to 3 percent using the results from equation (1) and (2). Also calculate the predicted increase in Y when the rate of unemployment increases from 5 percent to 6 percent.

Exercise 4 (20 %)

In an empirical analysis of demand for cultural services the following equation is to be estimated:

$$(1) \ln Y_i = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 X_{i2} + \delta_0 D_i + \delta_1 D_i \ln X_{i1} + u_i$$

Y_i is cultural expenses for individual i , X_{i1} is income, X_{i2} is the level of education (in years). D_i is a dummy variable that equals 1 if the individual lives in a city and 0 if he or she lives on the countryside. u_i is an error term.

- a) Interpret the parameters in (1). Pay particular attention to how the model allows for differences in demand depending on whether the individual lives in a city or on the countryside.

The model is estimated by OLS using data for 100 individuals. The results are shown in equation (2):

$$(2) \widehat{\ln Y_i} = 0,266 + 1,109 \ln X_{i1} + 0,051 X_{i2} + 0,253 D_i + 0,145 D_i \ln X_{i1}$$

(0,086) (0,039) (0,054) (0,0458) (0,118)

$$R^2 = 0,788, \text{ SSR} = 418,07$$

The figures in parentheses below the estimated parameters are estimated and standard errors, R^2 is the coefficient of multiple determination, and SSR is the residual sum of squares.

- b) Explain how you can test whether $\beta=1$. Perform the test using the information in (2).
 c) You want to test whether the demand for cultural services is the same for individuals living in a city as for individuals living on the countryside. Formulate the null hypothesis and explain how you will go about to test it. Perform the test using the information in (2) and (3).

$$(3) \widehat{\ln Y_i} = 0,173 + 1,220 \ln X_{i1} + 0,071 X_{i2}$$

(0,085) (0,033) (0,051)

$$R^2 = 0,771, \text{ SSR} = 451,20$$

Exercise 5 (10 %)

A commentator is interested in an empirical analysis you have conducted and writes the following: «The measurement errors for the variables used in the analysis are so substantial that the estimated parameters are systematically biased». Make a comment to this claim.