

Institutt for samfunnsøkonomi

## **Eksamensoppgave i SØK2900 – Empirisk metode**

**Faglig kontakt under eksamen: Bjarne Strøm**

**Tlf.: 73 59 19 33**

**Eksamensdato:** 4. juni 2014

**Eksamenstid:** 4 timer

**Sensurdato:** 26. juni 2014

**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 Casio fx-82ES PLUS, Citizen SR-270x, HP 30S eller SR-270X College

**Målform/språk:** Bokmål og nynorsk

**Antall sider:** 8 inkl forside

**Antall sider vedlegg:** 11

## Bokmål

Eksamensoppgaven består av seks oppgaver og et vedlegg med fire statistiske tabeller. Alle oppgavene skal besvares, og vekting ved sensur er gitt i parentes.

*Oppgave 1 (20%)*

- a) Innbyggerne i en stor by har i gjennomsnitt en inntekt på 376 000 kr, med standardavvik på 102 000 kr. Anta at inntekt er normalfordelt i populasjonen. Finn sannsynligheten for at et tilfeldig valgt individs inntekt ligger under fattigdomsgrensen, definert ved halvparten av gjennomsnittsinntekten.
- b) En bedrift produserer en komponent som inngår i produksjonen av et gitt bilmerke. Over tid viser det seg at 5% av komponentene som blir produsert er defekte. Et utvalg på 1000 komponenter blir levert til bilprodusenten. La variabelen  $X$  representere antallet defekte komponenter i utvalget. Beregn sannsynligheten for at utvalget på 1000 komponenter inneholder mer enn 65 defekte enheter. Presiser forutsetningene du benytter.
- c) Anta at den stokastiske variabelen  $Z$  er standard normalfordelt. Finn verdien  $k$  slik at  $P(k < Z < 0.81) = 0.439$ .
- d) I et utvalg på 500 personer av den mannlige arbeidsstyrken, er 20 arbeidsledige. Konstruer 95% og 99% konfidensintervall for populasjonens mannlige arbeidsledighetsrate. Presiser forutsetningene du benytter og kommenter resultatene.

*Oppgave 2 (10%)*

Ledelsen ved et stort sykehus hevder at 85% av de ansatte sykepleierne er fornøyde med arbeidstiden sin. Mange sykepleiere ønsker imidlertid økt stillingsprosent og mener ledelsen overdriver. I en undersøkelse utført av sykepleiernes fagforening svarer 62 av 80 spurte sykepleiere at de er fornøyd med eksisterende arbeidstid. Bruk opplysningene fra undersøkelsen til å teste om sykepleierne har rett i sin påstand. Bruk 5% signifikansnivå. Beregn også testens  $p$ -verdi. Hva forteller denne?

*Oppgave 3 (20%)*

En stor bilprodusent vil teste ut en ny motor for å avklare om den tilfredsstillende ny maksimumsstandard for luftforurensing. Gjennomsnittlig utslipp av karbonpartikler for denne type motor skal ligge under 20 ppm (parts per million). La  $X$  være målt utslipp fra en motor.  $X$  er normalfordelt med forventning  $\mu$  og standardavvik  $\sigma$ .

10 motorer blir produsert til testformål. Gjennomsnitt og standardavvik for utslipp fra testen er på henholdsvis 17.17 og 2.98 ppm. Produsenten konkluderer med at den nye motoren tilfredsstillende den nye standarden.

a) Utfør hypotesetesten som ligger til grunn for bilprodusentens konklusjon.

Bilprodusenten (bilprodusent 1) setter den nye motoren i produksjon. En konkurrerende bilprodusent (bilprodusent 2) hevder at deres motor for tilsvarende type bil er mer miljøvennlig. I en forbrukerundersøkelse blir det i alt gjennomført utslippsmålinger av 61 kjøretøy fra hvert av de to bilmerkene. De 61 nye målingene av biler fra produsent 1 ga gjennomsnitt og standardavvik på henholdsvis 17.15 og 2.78, mens målingene av de 61 bilene fra produsent 2 ga gjennomsnitt og standardavvik på henholdsvis 17.05 og 2.20.

b) Bruk resultatene fra forbrukerundersøkelsen til å teste om bilprodusent 2 har rett i sin påstand. Benytt både 5% og 10% signifikansnivå.

#### Oppgave 4 (20%)

En forsker er interessert i å undersøke hvordan salget av nye biler påvirkes av pris, inntektsnivå og rentenivå og foreslår følgende modell for å undersøke dette:

$$(1) \ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \varepsilon_t \quad t=1,2,\dots,40$$

der Y er antall solgte nye biler per innbygger,  $X_2$  er en indeks for realprisen på nye biler,  $X_3$  er realinntekta pr innbygger, mens  $X_4$  er et mål på rentenivået.  $\varepsilon_t$  er et stokastisk restledd. Fotskrift t indikerer tidsperiode (kvartal). Forskeren estimerer modellen på kvartalsdata for 10 år, til sammen 40 observasjoner. Han estimerer modellen med minste kvadraters metode (OLS) og oppnår følgende estimerte ligning, der estimerte standardavvik er angitt i parentes og SSR angir summen av kvadrerte residualer:

$$(2) \hat{\ln Y}_t = -31.852 - 1.751 \ln X_2 + 4.732 \ln X_3 - 0.193 \ln X_4 \quad \text{SSR}=0.3104$$

(5.491) (0.230) (0.706) (0.058)

a) Tolk parametrene i ligning (1) og de estimerte koeffisientene i (2).

b) Test en hypotese om at elastisiteten av nybilsalget med hensyn på nybilprisen er lavere enn -1.

c) Forskeren estimerer to alternative varianter av modellen, der han tillater at nybilsalget varierer med årstidene og rapporterer følgende estimerte ligninger, (3) og (4), der Q2 er en dummyvariabel som tar verdien 1 for observasjoner for 2.kvartal og 0 ellers, Q3 er en dummyvariabel med verdien 1 for 3.kvartal og 0 ellers, mens Q4 er en dummyvariabel med verdien 1 for 4.kvartal og 0 ellers:

$$(3) \hat{\ln Y}_t = -33.320 - 1.774 \ln X_2 + 4.907 \ln X_3 - 0.197 \ln X_4 + 0.092Q2 - 0.051Q3 - 0.049Q4 \quad \text{SSR}=0.17699$$

(4.384) (0.181) (0.706) (0.046) (0.033) (0.032) (0.033)

$$(4) \hat{\ln Y}_t = -32.582 - 1.760 \ln X_2 + 4.814 \ln X_3 - 0.196 \ln X_4 + 0.125Q2 \quad \text{SSR}=0.19341$$

(4.403) (0.183) (0.706) (0.047) (0.027)

Hva er referansekvartaler i modellene som er estimert i (3) og (4)? Test en hypotese om at nivået på nybilsalget er uavhengig av årstidene med utgangspunkt i de rapporterte resultatene i relasjon (2), (3) og (4).

## Oppgave 5 (20%)

En forsker er interessert i å avdekke hvordan inflasjonen avhenger av nivået på arbeidsledighet og har tilgjengelig kvartalsdata for disse variablene i Norge for perioden 4.kvartal 1980 til 4.kvartal 2008, til sammen 113 observasjoner.

Han estimerer en lineær relasjon mellom Y og X der Y er prosentvis inflasjon og X er arbeidsledighetsnivå i perioden målt som prosentandelen av arbeidsstyrken som er uten arbeid. Han oppnår følgende resultater:

$$\hat{Y} = 8.361 - 1.335X$$

SST=1131.50615 angir total variasjon i Y

SSR=842.387005 angir summen av kvadrerte residualer

$\sum (X_i - \bar{X})^2 = 162.17389$  angir den totale variasjonen i X

a) Bruk estimeringsresultatene til å predikere hvor mye inflasjonen øker dersom arbeidsledigheten, X, reduseres fra i) 6 til 5 prosentpoeng og ii) 3 til 2 prosentpoeng.

Beregn modellens forklaringskraft og estimer standardavviket til regresjonskoeffisienten foran X på grunnlag av opplysningene du har fått.

Konstruer 90 og 95 prosent konfidensintervall rundt regresjonskoeffisienten foran X. En kommentator påstår på basis av tidligere forskning at en økning i arbeidsledigheten med ett prosentpoeng gir en reduksjon i inflasjonen på ett prosentpoeng. Formuler en hypotese knyttet til kommentatorens påstand og test hypotesen.

b) En kollega av forskeren presenterer to alternative modeller (Modell 1 og Modell 2) for inflasjonen der han i stedet for X inkluderer  $\ln X$  som forklaringsvariabel i Modell 1 og  $\frac{1}{X}$  som forklaringsvariabel i Modell 2. Han oppnår følgende resultater fra estimeringen av disse modellene, der tallene i parentes angir estimerte standardavvik til koeffisientestimatene og  $R^2$  angir nivået på determinasjonskoeffisienten:

$$\text{Modell 1: } \hat{Y} = 9.004 - 4.506 \ln X, \quad R^2 = 0.303$$

(0.752) (0.648)

$$\text{Modell 2: } \hat{Y} = -0.556 + 12.82 \frac{1}{X}, \quad R^2 = 0.335$$

(0.666) (1.716)

Bruk opplysningene fra estimeringen av Modell 1 og Modell 2 til å predikere hvor mye inflasjonen øker dersom arbeidsledigheten reduseres fra i) 6 til 5 prosentpoeng og ii) 3 til 2 prosentpoeng slik som i spørsmål a). Forklar kort hvorfor effektene du finner er forskjellig fra det du fant i spørsmål a).

Oppgave 6 (10%)

Fra og med 1976 ble det forbudt å reklamere for tobakksvarer (sigaretter, rulle- og pipetobakk) i Norge. Skisser hvordan du vil gå fram for å undersøke om reklameforbudet førte til redusert forbruk av tobakksvarer.

Nynorsk

Eksamensoppgåva inneheld seks oppgåver og eit vedlegg med fire statistiske tabellar. Alle oppgåvene skal svarast på, og vekting ved sensur er gitt i parentes.

Oppgave 1 (20%)

- a) Innbyggjarane i ein stor by har i gjennomsnitt ei inntekt på 376 000 kr, med standardavvik på 102 000 kr. Anta at inntekt er normalfordelt i populasjonen. Finn sannsynet for at eit tilfeldig valt individ si inntekt ligg under fattigdomsgrensa, definert ved halvparten av gjennomsnittsinntekta.
- b) Ei bedrift produserer ein komponent som inngår i produksjonen av eit gitt bilmerke. Over tid viser det seg at 5% av komponentane som blir produsert er defekte. Eit utval på 1000 komponentar blir levert til bilprodusenten. La variabelen  $X$  representere talet på defekte komponentar i utvalet. Berekn sannsynet for at utvalet på 1000 komponentar inneheld meir enn 65 defekte einingar. Presiser føresetnadane du brukar.
- c) Anta at den stokastiske variabelen  $Z$  er standard normalfordelt. Finn verdien  $k$  slik at  $P(k < Z < 0.81) = 0.439$ .
- d) I et utval på 500 personar av den mannlege arbeidsstyrken, er 20 arbeidsledige. Konstruer 95% og 99% konfidensintervall for populasjonen si mannlege arbeidsløyserate. Presiser føresetnadane du brukar og kommenter resultatata.

Oppgave 2 (10%)

Leiinga ved eit stort sjukehus hevder at 85% av de tilsette sjukepleiarane er fornøgde med arbeidstida si. Mange sjukepleiarar ynskjer imidlertid økt stillingsprosent og meiner leiinga overdriv. I ei undersøking utført av sjukepleiarane si fagforeining svarer 62 av 80 spurde sjukepleiarar at dei er fornøgd med eksisterande arbeidstid. Bruk opplysningane frå undersøkinga til å teste om sjukepleiarane har rett i sin påstand. Bruk 5% signifikansnivå. Berekn også testen sin p-verdi. Kva fortel denne?

*Oppgave 3 (20%)*

Ein stor bilprodusent vil teste ut ein ny motor for å avklare om den tilfredsstillar ny maksimumsstandard for luftforureining. Gjennomsnittlig utslepp av karbonpartiklar for denne type motor skal ligge under 20 ppm (parts per million). La  $X$  vere målt utslepp frå ein motor.  $X$  er normalfordelt med forventning  $\mu$  og standardavvik  $\sigma$ .

10 motorar blir produsert til testformål. Gjennomsnitt og standardavvik for utslepp frå testen er på henholdsvis 17.17 og 2.98 ppm. Produsenten konkluderer med at den nye motoren tilfredsstillar den nye standarden.

a) Utfør hypotesetesten som ligger til grunn for bilprodusenten sin konklusjon.

Bilprodusenten (bilprodusent 1) setter den nye motoren i produksjon. En konkurrerende bilprodusent (bilprodusent 2) hevder at deira motor for tilsvarende type bil er meir miljøvenleg. I ein forbrukarundersøking blir det i alt gjennomført utsleppsmålingar av 61 kjøretøy frå kvar av dei to bilmerka. Dei 61 nye målingane av biler frå produsent 1 ga gjennomsnitt og standardavvik på henholdsvis 17.15 og 2.78, mens målingane av dei 61 bilene frå produsent 2 ga gjennomsnitt og standardavvik på henholdsvis 17.05 og 2.20.

b) Bruk resultata frå forbrukarundersøkinga til å teste om bilprodusent 2 har rett i sin påstand. Bruk både 5% og 10% signifikansnivå.

*Oppgave 4 (20%)*

Ein forskar er interessert i å undersøke korleis salet av nye biler blir påverka av pris, inntektsnivå og rentenivå og foreslår fylgjande modell for å undersøke dette:

$$(1) \ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \varepsilon_t \quad t=1,2,\dots,40$$

der  $Y$  er talet på selde nye biler per innbyggjar,  $X_2$  er ein indeks for realprisen på nye biler,  $X_3$  er realinntekta pr innbyggjar, mens  $X_4$  er eit mål på rentenivået.  $\varepsilon_t$  er eit stokastisk restledd. Fotskrift  $t$  indikerer tidsperiode (kvartal). Forskaren estimerer modellen på kvartalsdata for 10 år, til saman 40 observasjonar. Han estimerer modellen med minste kvadraters metode (OLS) og oppnår fylgjande estimerte likning, der estimerte standardavvik er angitt i parentes og SSR angir summen av kvadrerte residualer:

$$(2) \ln Y_t = -31.852 - 1.751 \ln X_2 + 4.732 \ln X_3 - 0.193 \ln X_4 \quad \text{SSR}=0.3104$$

(5.491)   (0.230)   (0.706)   (0.058)

a) Tolk parametrane i likning (1) og dei estimerte koeffisientane i (2).

b) Test ein hypotese om at elastisiteten av nybilsalet med omsyn til nybilprisen er lågare enn -1.

c) Forskaren estimerer to alternative variantar av modellen, der han tillet at nybilsalget varierer med årstidene og rapporterer fylgjande estimerte likningar, (3) og (4), der  $Q2$  er ein dummyvariabel som

tar verdien 1 for observasjonar for 2.kvartal og 0 elles, Q3 er ein dummyvariabel med verdien 1 for 3.kvartal og 0 elles, mens Q4 er ein dummyvariabel med verdien 1 for 4.kvartal og 0 elles:

$$(3) \quad \hat{\ln Y}_t = -33.320 - 1.774 \ln X_2 + 4.907 \ln X_3 - 0.197 \ln X_4 + 0.092Q2 - 0.051Q3 - 0.049Q4 \quad SSR=0.17699$$

(4.384) (0.181) (0.706) (0.046) (0.033) (0.032) (0.033)

$$(4) \quad \hat{\ln Y}_t = -32.582 - 1.760 \ln X_2 + 4.814 \ln X_3 - 0.196 \ln X_4 + 0.125Q2 \quad SSR=0.19341$$

(4.403) (0.183) (0.706) (0.047) (0.027)

Kva er referansekvartaler i modellane som er estimert i (3) og (4)? Test ein hypotese om at nivået på nybilsalet er uavhengig av årstidene med utgangspunkt i dei rapporterte resultatane i relasjon (2), (3) og (4).

### Oppgåve 5 (20%)

Ein forskar er interessert i å avdekke korleis inflasjonen avheng av nivået på arbeidsløysing og har tilgjengelig kvartalsdata for desse variablane i Noreg for perioden 4.kvartal 1980 til 4.kvartal 2008, til saman 113 observasjonar.

Han estimerer ein lineær relasjon mellom Y og X der Y er prosentvis inflasjon og X er arbeidsløysingnivå i perioden målt som prosentdelen av arbeidsstyrken som er utan arbeid. Han oppnår fylgjande resultat:

$$\hat{Y} = 8.361 - 1.335X$$

SST=1131.50615 angir total variasjon i Y

SSR=842.387005 angir summen av kvadrerte residualer

$$\sum (X_i - \bar{X})^2 = 162.17389 \text{ angir den totale variasjonen i X}$$

a) Bruk estimeringsresultata til å predikere kor mykje inflasjonen aukar dersom arbeidsløysing, X, reduserast frå i) 6 til 5 prosentpoeng og ii) 3 til 2 prosentpoeng.

Berekn modellen si forklaringskraft og estimer standardavviket til regresjonskoeffisienten framfor X på grunnlag av opplysningane du har fått.

Konstruer 90 og 95 prosent konfidensintervall rundt regresjonskoeffisienten framfor X. Ein kommentator påstår på basis av tidlegare forskning at ei auke i arbeidsløysing med eitt prosentpoeng gir ein reduksjon i inflasjonen på eitt prosentpoeng. Formulér ein hypotese knytt til kommentatoren sin påstand og test hypotesen.

b) Ein kollega av forskaren presenterer to alternative modeller (Modell 1 og Modell 2) for inflasjonen der han i staden for X inkluderer  $\ln X$  som forklaringsvariabel i Modell 1 og  $\frac{1}{X}$  som forklaringsvariabel i Modell 2. Han oppnår fylgjande resultat frå estimeringen av disse modellane,

der talla i parentes angir estimerte standardavvik til koeffisientestimata og  $R^2$  angir nivået på determinasjonskoeffisienten:

Modell 1:  $\hat{Y} = 9.004 - 4.506 \ln X$ ,  $R^2 = 0.303$   
(0.752) (0.648)

Modell 2:  $\hat{Y} = -0.556 + 12.82 \frac{1}{X}$   $R^2 = 0.335$   
(0.666) (1.716)

Bruk opplysningane frå estimeringen av Modell 1 og Modell 2 til å predikere kor mykje inflasjonen aukar dersom arbeidsløysa reduserast frå i) 6 til 5 prosentpoeng og ii) 3 til 2 prosentpoeng slik som i spørsmål a). Forklar kort kvifor effektane du finn er forskjellig frå det du fann i spørsmål a).

### *Oppgave 6 (10%)*

Frå og med 1976 ble det forbode å reklamere for tobakksvarer (sigarettar, rulle- og pipetobakk) i Noreg. Skisser korleis du vil gå fram for å undersøke om reklameforbodet førte til redusert forbruk av tobakksvarer.