ECONnect NTNU

Faktor

- en eksamensavis utgitt av ECONnect

Løsningsforslag:
SØK1004 – Statistikk for økonomer

Eksamen: Våren 2009
Antall sider: 16

Om ECONnect:

ECONnect er en frivillig studentorganisasjon for studentene på samfunnsøkonomi- og finansøkonomistudiet ved NTNU. Vi arbeider for økt faglig kompetanse blant våre studenter samt tettere kontakt med næringslivet. Det gjør vi ved å arrangere fagdager, gjesteforelesninger, bedriftspresentasjoner m.m. I dag går det ca. 200 studenter på bachelornivå (1.-3. klasse) og ca. 70 studenter på masternivå (4.-5. klasse). Studentene på masternivå er fordelt på de to linjene samfunnsøkonomi (ca. 50 stk) og finansiell økonomi (ca. 20 stk). Mer om ECONnect og aktuelle arrangementer på www.econnect-ntnu.no.

ECONnect består av følgende personer ved utgivelsestidspunkt:

Bjørn Bergholt (Leder)	bjorn@econnect-ntnu.no
Elise Caspersen (Fagdagansvarlig)	elise@econnect-ntnu.no
Pål Christian Vågbø (Bedriftansvarlig)	paal@econnect-ntnu.no
Tormod Hagerup (Økonomi/Samfunnsøk.)	tormod@econnect-ntnu.no
Tiril Toftedahl	tiril@econnect-ntnu.no
Louis Dieffenthaler	louis@econnect-ntnu.no
Tone Hedvig	tone@econnect-ntnu.no
Ole Christian Grytten	ole@econnect-ntnu.no

Post- og besøksadresse:	*Organisasjonsnummer:*	*Hjemmeside:*
ECONnect, NTNU Dragvoll Institutt for samfunnsøkonomi Bygg 7, Nivå 5 7491 Trondheim	NO 994 625 314	www.econnect-ntnu.no

Merk: Eksamensbesvarelsene har i varierende grad feil og mangler, både oppsett og innhold. De vil også kun vise en av flere mulige fremgangsmåter. ECONnect står ikke ansvarlig for selve faginnholdet.

Norges teknisk-naturvitenskapelige universitet
Institutt for samfunnsøkonomi

EKSAMENSOPPGAVE I SØK1004

STATISTIKK FOR ØKONOMER

Faglig kontakt under eksamen: Bjørnar Kivedal Karlsen

Tlf.: 9 16 54

Eksamensdato: Tirsdag 2. juni 2009

Eksamenssted: Dragvoll

Eksamenstid: 4 timer

Studiepoeng: 7,5

Tillatte hjelpemidler: Alle trykte og håndskrevne hjelpemidler, samt godkjent kalkulator Citizen SR-270x el. HP 30S.

Sensur: 23. juni 2009.

Eksamen består av 5 oppgaver med delspørsmål som alle skal besvares.
Oppgaveteksten er skrevet på bokmål og nynorsk.

Merk! Studentene må primært gjøre seg kjent med sensur på stud.web eller ved å oppsøke sensuroppslagene. Evt telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

Oppgave 1

Du skal kjøre bil fra Oslo til Trondheim en sen høstkveld, og er engstelig for dårlig vær og hvordan dette påvirker risikoen for å havne i en ulykke. Det spås fint vær, regnvær og snøvær med sannsynlighet på henholdsvis 0.60, 0.39 og 0.01. Du vet også at sannsynlighet for en bilulykke på denne strekningen er 0.001, 0.003 og 0.009 i de tre nevnte værtypene. Hva er sannsynligheten for å være involvert i en ulykke? Gitt at det skjer en ulykke, hva er sannsynligheten for at snøvær var medvirkende årsak?

Oppgave 2

Produktene som kommer til sluttkontrollen i en bedrift kan ha to typer feil:
A = ”Skjønnhetsfeil”
B = ”Funksjonsfeil”
Erfaringsmessig vet bedriften at 10% av de produserte enhetene har skjønnhetsfeil, og 8% av enhetene har funksjonsfeil. De to feiltypene opptrer uavhengig av hverandre. Vi antar dessuten at om en enhet har feil eller ikke, er uavhengig av hva som skjer med de andre enhetene.

a) Hva er sannsynligheten for at en enhet fra produksjonen har begge typer feil?
Hva er sannsynligheten for at en enhet *bare* har skjønnhetsfeil?
Vis at sannsynligheten for at en enhet er feilfri, er 0.828.

Bedriften får kr 80 for en enhet som er feilfri og kr 60 for en enhet som har bare skjønnhetsfeil. Enheter som har funksjonsfeil, kan ikke selges. Vi lar Z være prisen som bedriften får for en tilfeldig enhet.

b) Beregn $E(Z)$ og $\text{Var}(Z)$.

Bedriften produserer 800 enheter på en dag. Vi lar X være antall av disse som er feilfrie.

c) Begrunn kort at X er binomisk fordelt. Finn sannsynligheten for at flere enn 650 enheter er feilfrie. Finn også $P(640<X\leq 670)$.

Oppgave 3

En sukkertøyprodusent leverer poser med lakrisbåter. Nettovekten som er angitt på posene, er 150 gram. Vi lar X være målt vekt av lakrisbåtene i en tilfeldig pose. Vi antar at X er normalfordelt med forventning $\mu=150$ gram og standardavvik $\sigma = 3$ gram.

En kunde har mistanke om at vekten av innholdet i posene er for lav. Hun har veid 16 poser og vil på bakgrunn av disse målingene ta stilling til om hun vil påstå at innholdet i posene er for lite i forhold til den angitte nettovekten på 150 gram. Resultatene for X er som følger:

146	148	154	152	144	145	144	153	145	148	146	153	147	148	152	147

$$\bar{X} = 148.25$$

$$\sum\left(X_i - \bar{X}\right)^2 = 177$$

a) Bruk resultatene til å estimere forventningen μ og standardavviket σ.
b) Finn et 90% konfidensintervall for μ når vi antar at σ = 3 gram.

c) Utfør en test for å undersøke om posene på markedet gjennomsnittlig inneholder mindre enn 150 gram. Bruk signifikansnivå 2.5% og anta at σ = 3 gram. Hva blir konklusjonen?
d) Gjennomfør testen i oppgave c dersom du antar at σ er ukjent.
e) Regn ut teststyrken for μ=147.5 gram og μ=145 gram. Forklar med ord hva den beregnede styrken betyr. Anta at σ = 3 gram.
f) Sukkertøyprodusenten i oppgave 3 ønsker å innføre et krav om at standardavviket for vekten på godteposene skal være mindre enn 4.0 gram. Test om dette kravet er oppfylt. Bruk 2.5% signifikansnivå.

Oppgave 4

Vi har to stokastiske variabler, *X* og *Y*, med følgende simultanfordeling:

X↓ *Y*→	0	2	4
1	0.04	0.04	0.02
2	0.17	0.31	0.12
3	0.09	0.15	0.06

a) Finn de marginale sannsynlighetsfordelingene til henholdsvis *X* og *Y*.
b) Finn kovariansen til *X* og *Y*. Er variablene uavhengige? Begrunn svaret.

Oppgave 5

En 44-åring har deltatt i et mosjonsløp de siste sju årene. Anvendt tid er oppgitt i kronologisk rekkefølge i tabellen nedenfor. Tallene er uttrykt som avrundede desimaltall (43 min og 20 sek skrives altså 43.3). Det første tallet gjelder som 38-åring, og det siste tallet gjelder som 44-åring.

41.2	41.3	41.7	42.6	43.3	43.1	44.8

a) Plott dataene i et koordinatsystem. La mosjonistens alder *X* gå langs den horisontale aksen, og anvendt tid *Y* langs den vertikale aksen. Velg enheter slik at plottet blir lett leselig.

Vi antar at dataene kan beskrives med regresjonsmodellen $Y = \alpha + \beta X + e$. Dataene gir også:

$$\sum_{i=1}^{7}\left(X_i - \bar{X}\right)^2 = 28, \quad \sum_{i=1}^{7}\left(Y_i - \bar{Y}\right)^2 = 10, \quad \sum_{i=1}^{7}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = 16$$

b) Regn ut utvalgskorrelasjonen mellom *X* og *Y*. Hva betyr korrelasjonskoeffisienten for samvariasjonen mellom *X* og *Y*?
c) Beregn estimater for de ukjente parameterne α og β som inngår i modellen. Tegn estimert regresjonslinje i samme koordinatsystem som i oppgave a.

Estimeringen gir følgende resultater:

$$\sum_{i=1}^{7}\left(Y_i - \hat{Y}_i\right)^2 = 0.891, \quad \sum_{i=1}^{7}\left(Y_i - \bar{Y}\right)^2 = 10.034, \quad \sum_{i=1}^{7}\left(\hat{Y}_i - \bar{Y}\right)^2 = 9.143$$

d) Gi en praktisk tolkning av β. Er det mulig å gi en tilsvarende praktisk tolkning av α?
e) Hvor god vil du si at modellen er?
f) Beregn et 90% konfidensintervall for β. Ville et 95% konfidensintervall vært større eller mindre enn 90%-konfidensintervallet du beregnet? Begrunn hvorfor.
g) Anta at modellen også kan brukes neste år. Hva blir forventet tid neste år?

Løsningsforslag eksamen SØK1004 02.06.2009

Oppgave 1

Vi lar F, R og S betegne hhv. fint vær, regn og snø. Kan da skrive de oppgitte sannsynlighetene som følger:

$$\begin{aligned} P(F) &= 0.60 \\ P(R) &= 0.39 \\ P(S) &= 0.01 \end{aligned}$$

$$\begin{aligned} P(U|F) &= 0.001 \\ P(U|R) &= 0.003 \\ P(U|S) &= 0.009 \end{aligned}$$

Sannsynligheten for å være involvert i en ulykke:

$$\begin{aligned} P(U) &= P(U \cap F) + P(U \cap R) + P(U \cap S) \\ &= P(U|F) \cdot P(F) + P(U|R) \cdot P(R) + P(U|S) \cdot P(S) \\ &= 0.001 \cdot 0.60 + 0.003 \cdot 0.39 + 0.009 \cdot 0.01 \\ &= 0.00186 \\ &= 0.19\% \end{aligned}$$

Sannsynligheten for at snøvær var medvirkende årsak, gitt at det skjer en ulykke:

$$\begin{aligned} P(S|U) &= \frac{P(S\cap U)}{P(U)} \\ &= \frac{P(U|S)\cdot P(S)}{P(U)} \\ &= \frac{0.009\cdot 0.01}{0.00186} \\ &= 0.0484 \\ &= 4.84\% \end{aligned}$$

Oppgave 2

$$\begin{aligned} P(A) &= 0.10 \\ P(B) &= 0.08 \end{aligned}$$

a)

Sannsynligheten for at en enhet har begge typer feil:

$$\begin{aligned} P(A\cap B) &= P(A)\cdot P(B) \\ &= 0.10\cdot 0.08 \\ &= 0.008 \end{aligned}$$

Sannsynligheten for at en enhet bare har skjønnhetsfeil:

$$\begin{aligned} P\left(A\cap \bar{B}\right) &= P(A)\cdot P\left(\bar{B}\right) \\ &= P(A)\cdot (1-P(B)) \\ &= 0.10\cdot (1-0.08) \\ &= 0.10\cdot 0.92 \\ &= 0.092 \end{aligned}$$

Vis at sannsynligheten for at en enhet er feilfri er 0.828:

$$\begin{aligned} P\left(\bar{A}\cap \bar{B}\right) &= P\left(\bar{A}\right)\cdot P\left(\bar{B}\right) \\ &= (1-P(A))\cdot \left(1-P\left(\bar{B}\right)\right) \\ &= (1-0.10)\cdot (1-0.08) \\ &= 0.90\cdot 0.92 \\ &= 0.828 \end{aligned}$$

b)

Sannsynlighetsfordeling:

	funksjonsfeil	skjønnhetsfeil	feilfri
Z	0	60	80
$P(Z=z)$	0.08	0.092	0.828

$$\begin{aligned} E(Z) &= \sum Z \cdot P(Z=z) \\ &= 0 \cdot 0.08 + 60 \cdot 0.092 + 80 \cdot 0.828 \\ &= 71.76 \end{aligned}$$

$$\begin{aligned} Var(Z) &= \sum (Z - E(Z))^2 \cdot P(Z=z) \\ &= (0-71.76)^2 \cdot 0.08 + (60-71.76)^2 \cdot 0.092 + (80-71.76)^2 \cdot 0.828 \\ &= 480.90 \end{aligned}$$

c)

X er binomisk fordelt fordi:

- To mulige utfall (feilfri/ikke feilfri)
- Konstant sannsynlighet
- Uavhengige utfall

Sannsynligheten for at flere enn 650 enheter er feilfrie:

$$\begin{aligned} P(X > 650) &= 1 - P(X \leq 650) \\ &= 1 - P\left(Z \leq \frac{\bar{X} - \mu}{\sigma}\right) \\ &= 1 - P\left(Z \leq \frac{650 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\ &= 1 - P\left(Z \leq \frac{650 - 800 \cdot 0.828}{\sqrt{800 \cdot 0.828 \cdot (1 - 0.828)}}\right) \\ &= 1 - P\left(Z \leq \frac{650 - 662.4}{\sqrt{113.93}}\right) \\ &= 1 - P(Z \leq -1.1617) \\ &= 1 - \Phi(-1.16) \\ &= 1 - 0.1230 \\ &= 0.877 \end{aligned}$$

$$
\begin{aligned}
P(640 < X \leq 670) &= P(X \leq 670) - P(X \leq 640) \\
&= \Phi\left(\frac{670 - 662.4}{\sqrt{113.93}}\right) - \Phi\left(\frac{640 - 662.4}{\sqrt{113.93}}\right) \\
&= \Phi(0.71) - \Phi(-2.10) \\
&= 0.7611 - 0.0179 \\
&= 0.7432
\end{aligned}
$$

Oppgave 3

a)

$$
\begin{aligned}
\hat{\mu} &= E(X) = \bar{X} = 148.25 \\
\hat{\sigma} &= s = \sqrt{\frac{\sum \left(X - \bar{X}\right)^2}{n-1}} = \sqrt{\frac{177}{15}} = 3.44
\end{aligned}
$$

b)

Et 90% konfidensintervall for μ når vi antar at $\sigma = 3.0$ gram:

$$
\begin{aligned}
& \bar{X} \pm 1.645 \cdot \frac{3}{\sqrt{16}} \\
= {} & 148.25 \pm 1.234 \\
= {} & [147.02, 149.48]
\end{aligned}
$$

c)

En test for å undersøke om posene på markedet gjennomsnittlig inneholder mindre enn 150 gram, signifikansnivå 2.5% og antar at $\sigma = 3.0$ gram:

$$
\begin{aligned}
H_0 &: \mu \geq 150 \\
H_A &: \mu < 150
\end{aligned}
$$

$$
\begin{aligned}
TS &= \frac{148.25 - 150}{3/\sqrt{16}} = -2.33 \\
k_{0.025} &= -1.96
\end{aligned}
$$

Konklusjon: $TS < k$, så vi aksepterer H_A. Posene på markedet inneholder gjennomsnittlig mindre enn 150 gram.

d)

Dersom σ er ukjent, må vi estimere den (noe som ble gjort i oppgave a). Siden vi har lite utvalg og ukjent varians, bruker vi t-fordelingen for å gjennomføre testen.

$$TS = \frac{148.25 - 150}{3.44/\sqrt{16}} = -2.035$$

$$k = -t_{0.025,15} = -2.131$$

Konklusjon: $TS > k$, så vi beholder H_0. Posene på markedet inneholder ikke gjennomsnittlig mindre enn 150 gram.

e)

$$\begin{aligned}
\mu^* &= 147.5 : \\
\gamma &= P\left(Z \leq -1.96 + \frac{150 - 147.5}{3/\sqrt{16}}\right) \\
&= P\left(Z \leq 1.3773\right) \\
&= \Phi\left(1.38\right) \\
&= 0.9162
\end{aligned}$$

Det er 91.62% sannsynlighet for at vi forkaster H_0 dersom det faktiske populasjonsgjennomsnittet er 147.5.

$$\begin{aligned}
\mu^* &= 145 : \\
\gamma &= P\left(Z \leq -1.96 + \frac{150 - 145}{3/\sqrt{16}}\right) \\
&= P\left(Z \leq 4.7067\right) \\
&= \Phi\left(4.71\right) \\
&= 1
\end{aligned}$$

Det er 100% sannsynlighet for at vi forkaster H_0 dersom det faktiske populasjonsgjennomsnittet er 145.

Styrkefunksjonen forteller sannsynligheten for å forkaste H_0 dersom H_0 faktisk er feil (og det er et annet populasjonsgjennomsnitt, gitt ved μ^*).

f)

Siden alternativhypotesen er den hypotesen vi ønsker å teste (den påstanden som krever bevis), setter vi $\sigma^2 > 16$ som alternativhypotese siden bedriften ønsker å teste om variansen er for stor til at den kan godkjennes. Tvilen skal komme nullhypotesen til gode, så vi velger å godkjenne variansen såfremt ikke noe annet er bevist.

$$\begin{aligned} H_0 &: \sigma^2 \leq 16 \ (\sigma \leq 4) \\ H_A &: \sigma^2 > 16 \ (\sigma > 4) \end{aligned}$$

Bruk av motsatt hypotese ($H_0 : \sigma^2 \geq 16$ mot $H_0 : \sigma^2 < 16$) vil gi ganske høy uttelling på denne oppgaven også.

$$TS = \frac{(n-1)\, s^2}{\sigma^2}$$

Vi har fra oppgave 3a at $s^2 = 3.44^2$

$$TS = \frac{(16-1)\, 3.44^2}{4^2} = 11.09$$

For 2.5% signifikansnivå har vi følgende kritiske verdi:

$$k = \chi^2_{0.025,15} = 27.49$$

Konklusjon: $TS < k$, så vi beholder H_0. Standardavviket ser ut til å være innenfor kravet.

Oppgave 4

$x \downarrow \quad y \rightarrow$	0	2	4	$f(x)$
1	0.04	0.04	0.02	0.10
2	0.17	0.31	0.12	0.60
3	0.09	0.15	0.06	0.30
$g(y)$	0.30	0.50	0.20	1

a)

$f(x)$ viser marginalsannsynligheten til X og $g(y)$ viser marginalsannsynligheten til Y.

b)

$$Cov\,(X,Y) = E\,(XY) - \mu_X \mu_Y$$

$$\begin{aligned} E\,(XY) &= 0 \cdot 1 \cdot 0.04 + 2 \cdot 1 \cdot 0.04 + 4 \cdot 1 \cdot 0.02 \\ &\quad +0 \cdot 2 \cdot 0.17 + 2 \cdot 2 \cdot 0.31 + 4 \cdot 2 \cdot 0.12 \\ &\quad +0 \cdot 3 \cdot 0.09 + 2 \cdot 3 \cdot 0.15 + 4 \cdot 3 \cdot 0.06 \\ &= 3.98 \end{aligned}$$

$$\begin{aligned} \mu_X &= E\,(X) = \sum X \cdot P\,(X = x) \\ &= 1 \cdot 0.10 + 2 \cdot 0.60 + 3 \cdot 0.30 \\ &= 2.2 \end{aligned}$$

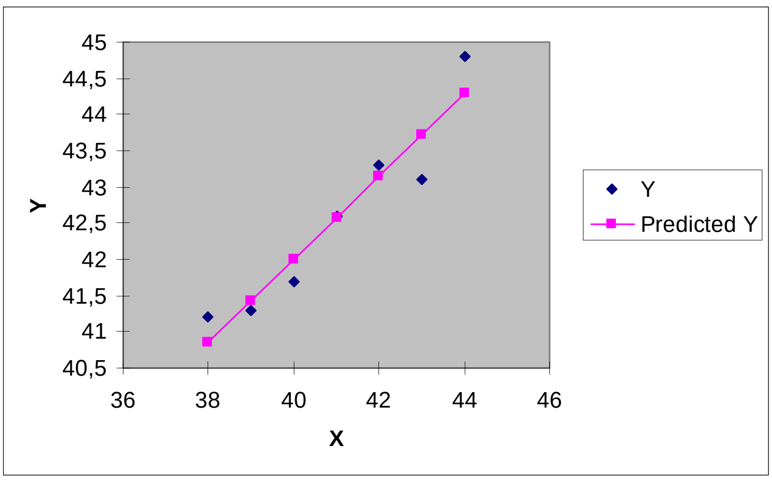
$$\begin{aligned} \mu_Y &= E\,(Y) = \sum Y \cdot P\,(Y = y) \\ &= 0 \cdot 0.30 + 2 \cdot 0.50 + 4 \cdot 0.20 \\ &= 1.8 \end{aligned}$$

$$\begin{aligned} Cov\,(X,Y) &= 3.98 - 2.2 \cdot 1.8 \\ &= 0.02 \end{aligned}$$

Siden $Cov\,(X,Y) \neq 0$, er variablene ikke uavhengige. Uavhengighet mellom to variabler gir kovarians på 0.

Oppgave 5

a)



b)

$$
\begin{aligned}
R &= \frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sqrt{\sum\left(X-\bar{X}\right)^2}\cdot\sqrt{\sum\left(Y-\bar{Y}\right)^2}} \\
&= \frac{16}{\sqrt{28}\cdot\sqrt{10}} \\
&= 0.955
\end{aligned}
$$

Dette vil si at det er kraftig positiv samvariasjon mellom X og Y. En høy verdi på Y (lang tid) henger sammen med en høy verdi på X (høy alder).

c)

$$
\hat{\beta} = \frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sum\left(X-\bar{X}\right)^2} = \frac{16}{28} = 0.571
$$

$$
\begin{aligned}
\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\
&= \frac{41.2+41.3+41.7+42.6+43.3+43.1+44.8}{7} \\
&\quad -0.571\frac{38+39+40+41+42+43+44}{7} \\
&= 19.16
\end{aligned}
$$

$$
\hat{Y} = 19.16 + 0.571\cdot X
$$

Beregner f.eks. $\hat{Y}_{38}$ og $\hat{Y}_{44}$ for å få to punkter for å tegne regresjonslinjen som en rett linje mellom disse punktene.

$$
\begin{aligned}
\hat{Y}_{38} &= 40.86 \\
\hat{Y}_{44} &= 44.28
\end{aligned}
$$

d)

β viser hvor mye Y endres dersom X øker med en enhet. Dette vil si at dersom personen blir 1 år eldre, vil tiden øke med 0.571 min, eller 34,26 sekunder.

Den praktiske tolkningen av α er hva Y predikeres til når $X = 0$. $X = 0$ vil si at alderen er 0 år. Dette gir ikke noen mening.

e)

$$R^2 = \frac{SSE}{SST} = \frac{\sum\left(\hat{Y}-\bar{Y}\right)^2}{\sum\left(Y-\bar{Y}\right)^2} = \frac{9.143}{10.034} = 0.9112$$

91% av variasjonen i Y kan forklares gjennom X i denne modellen.

f)

90% konfidensintervall for β :

$$0.571 \pm t_{0.05,5} \cdot s_{\hat{\beta}}$$

$$s_{\hat{\beta}}^2 = \frac{\sum\left(Y-\hat{Y}\right)^2/n-2}{\sum\left(X-\bar{X}\right)^2}$$

$$\begin{aligned} s_{\hat{\beta}}^2 &= \frac{0.891/5}{28} \\ &= 0.00636 \\ s_{\hat{\beta}} &= \sqrt{0.00636} \\ &= 0.07975 \end{aligned}$$

$$\begin{aligned} &0.571 \pm t_{0.05,5} \cdot s_{\hat{\beta}} \\ &0.571 \pm 2.015 \cdot 0.07975 \\ &0.571 \pm 0.1607 \\ &[0.41, 0.73] \end{aligned}$$

Et konfidensintervall viser sannsynligheten for at den faktiske verdien befinner seg i intervallet. Siden sannsynligheten for at den faktiske verdien er i intervallet er større ved et 95% konfidensintervall, så må et 95% konfidensintervall være større enn et 90% konfidensintervall.

g)

Neste år er personens alder 45 (X=45)

$$\begin{aligned} \hat{Y}_{45} &= 19.14 + 0.571 \cdot 45 \\ &= 44.84 \end{aligned}$$