



ECONnect

NTNU

Faktor

- en eksamensavis utgitt av ECONnect



Eksamensbesvarelse:

SØK1004 – Statistikk for økonomer

Eksamen:

Våren 2010

Antall sider:

7



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
Tone Hedvig Berg (Bedriftsansvarlig)	tone@econnect-ntnu.no
Elise Caspersen (Fagdagsansvarlig)	elise@econnect-ntnu.no
Tiril Toftdahl (Faktoransvarlig)	tiril@econnect-ntnu.no
Tormod Hagerup (Økonomi/Kandidattreffet)	tormod@econnect-ntnu.no
Louis Dieffenthaler	louis@econnect-ntnu.no
Daniel Johansson	daniel@econnect-ntnu.no
Georg Næsheim	georg@econnect-ntnu.no
Mariell Toven	mariell@econnect-ntnu.no
Ellen Normann	ellen@econnect-ntnu.no
Ragnhild Grøv	ragnhild@econnect-ntnu.no
Johan Berg Fossen	johan@econnect-ntnu.no
Ole Christian Grytten	ole@econnect-ntnu.no

<i>Post- og besøksadresse:</i>	<i>Organisasjonsnummer:</i>	<i>Hjemmeside:</i>
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.

Oppgave 1 (20%)

Lungeavdelingen ved et sykehus regner med at 7 % av den voksne befolkningen lider av KOLS. Lungeavdelingen regner også med at 90 % av de som lider av KOLS, røyker, mens av dem som ikke har KOLS, røyker 25 %.

Vi trekker ut tilfeldig en voksen person. K er utfallet at personen har en lungesykdom og R er utfallet at personen røyker.

- Formuler opplysningene i oppgaven som sannsynligheter. Vis at $Pr(R)=0,296$.
- Regn ut $Pr(K/R)$ og forklar kort hva denne sannsynligheten sier oss
- Hva er sannsynligheten for at en person som ikke røyker har KOLS?

Oppgave 2 (30%)

Farten til en tilfeldig bil X (km/t) på en bestemt veistrekning kan oppfattes som en normalfordelt stokastisk variabel. Tidligere målinger har vist at gjennomsnittsfarten på denne strekningen var 77 km/t.

En holdningskampanje er gjennomført, og en vil undersøke om fartsnivået på strekningen har avtatt. På en dag ble det målt fart for 12 biler. Målingene X_1, X_2, \dots, X_{12} antas å være uavhengige og normalfordelte variabler med forventning μ .

- Tyder resultatene på at fartsnivået på strekningen nå er lavere enn 77 km/t? Formuler dette spørsmålet som en hypotesetest. Bruk signifikansnivå 5 %. Gjennomfør testingen og angi konklusjonen når målt gjennomsnittsfart for de 12 bilene var 73,33 km/t og populasjonsstandardavvik er $\sigma=10,0$ km/t.
- Lag et 98 % konfidensintervall for utvalgsgjennomsnittet for de 12 målte bilene basert på tallene gitt i 2a. Kan du bruke dette for å gjennomføre en hypotesetest med et gitt signifikansnivå? Begrunn i så fall hva slags test og vis resultatet av testen.
- Anta nå at σ er ukjent, men at du kjenner standardavviket for utvalget som er 10,0 km/t. Gjennomfør den samme hypotesetesten som i oppgave 2a (samme gjennomsnittsfart og signifikansnivå) og angi konklusjonen.
- Hvis utvalgsstandardavviket hadde endret seg (uten at noe annet gjorde det), måtte det vært større eller mindre enn 10 for å endre konklusjonen i oppgave 2c? Forklar, med ord, hvorfor, og regn så ut hvor mye utvalgsstandardavviket måtte vært for å endre konklusjonen på testen. (Vi antar at gjennomsnittsfart, utvalgsstørrelse og signifikansnivå er som i oppgave 2c.)

Oppgave 3 (25%)

Følgende tabell angir antall epletrær solgt i begynnelsen av juni i en hagebutikk i Trondheim (D) og utetemperatur (T) (helgene er ikke inkludert).

D	92	70	72	61	75	50	84
T	22	23	19	22	24	18	26

Anta at $D = \alpha + \beta T + \varepsilon$, hvor ε er et restledd.

- Finn Minste kvadraters metode estimatorene av α og β . Tolk resultatene.
- I følge Meteorologisk institutt kan utetemperaturen i Trondheim variere mellom 4 og 28 grader i begynnelsen av juni. Prediker antall solgte epletrær for de to ytterpunktene. Hvor troverdige er resultatene?

Oppgave 4 (25%)

I følge forskning finnes det et bytteforhold (trade-off) mellom lengden på nattsøvn og tiden brukt i inntektsgivende arbeid. En gruppe på 17 kvinner rapporterte totalt antall timer de hadde sovnet (*Søvn*) og jobbet (*Arbeid*) i løpet av en uke. Forskerne brukte deretter data til å estimere en modell hvor lengden på søvn er responsvariabelen. De benyttet Minste kvadraters metode. Følgende er utdrag fra Excelutskrift:

Variansanalyse

	<i>fg</i>	<i>SK (SS)</i>
Regresjon	1	
Residualer	15	634,59
Totalt	16	813,42

	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>
Skjæringspunkt	62,865	4,864	12,923
Arbeid	-0,224	0,109	-2,056

AVVIK (UTDATA)

<i>Observasjon</i>	<i>Fremskrevet (estimert)</i>	
	<i>Søvn</i>	<i>Residualer</i>
1	53,3	-4,2
2	51,0	

I tillegg vet vi følgende for datasettet:

	Søvn	Arbeid
Minimumsverdi	38,9	6,2
Maksimumsverdi	66,5	72,1
Gjennomsnittet	53,4	42,2
Median	54,9	45

SØK1004 Statistikk for økonomer

- a) Skriv den estimerte regresjonslinjen på formatet $\hat{Y} = a + bX$.
- b) Hva forteller skjæringspunktet i denne ligningen?
- c) Hvis en person øker den ukentlige arbeidsinnsatsen med 2 timer, hvor mye kan vi forvente at det totale antall timer sovet i løpet av en uke forandrer seg basert på den estimerte modellen?
- d) For den første observasjonen i datasettet, hva er det rapporterte antall timer sovet per uke (Y_1)?
- e) Regn ut forklart variasjon og restvariasjon for den andre observasjonen som rapporterte 53 timer på jobb og søvn på 55 timer.
- f) Hvor mye av variasjon i *Søvn* er forklart av *Arbeid*? Begrunn svaret ved å bruke tallene gitt i Excelutskrift.

Løsningsforslag oppg 1 og 2, SØK1004 V2010

Oppgave 1

a)

$$\begin{aligned}\Pr(K) &= 0,07 \\ \Pr(R|K) &= 0,90 \\ \Pr(R|\bar{K}) &= 0,25\end{aligned}$$

$$\begin{aligned}\Pr(R) &= \Pr(R \cap K) + \Pr(R \cap \bar{K}) \\ &= \Pr(R|K) \cdot \Pr(K) + \Pr(R|\bar{K}) \cdot \Pr(\bar{K}) \\ &= \Pr(R|K) \cdot \Pr(K) + \Pr(R|\bar{K}) \cdot (1 - \Pr(K)) \\ &= 0,90 \cdot 0,07 + 0,25 \cdot (1 - 0,07) \\ &= 0,2955\end{aligned}$$

b)

$$\Pr(K|R) = \frac{\Pr(K \cap R)}{\Pr(R)} = \frac{\Pr(R|K) \cdot \Pr(K)}{\Pr(R)} = \frac{0,90 \cdot 0,07}{0,2955} = 0,2132$$

Sannsynligheten for at en tilfeldig utplukket røykende person har KOLS.

c)

Sannsynligheten for at en person som ikke røyker har KOLS:

$$\begin{aligned}\Pr(K|\bar{R}) &= \frac{\Pr(K \cap \bar{R})}{\Pr(\bar{R})} = \frac{\Pr(K) - \Pr(K \cap R)}{1 - \Pr(R)} \\ &= \frac{\Pr(K) - \Pr(R|K) \cdot \Pr(K)}{1 - \Pr(R)} = \frac{0,07 - 0,90 \cdot 0,07}{1 - 0,2955} \\ &= 0,0099\end{aligned}$$

Oppgave 2

a)

$$\begin{aligned}\mu &= 77 \\ n &= 12 \\ \sigma &= 10\end{aligned}$$

$$\begin{aligned}H_0 &: \mu \geq 77 \text{ eller } H_0 : \mu = 77 \text{ (begge godtas)} \\ H_A &: \mu < 77\end{aligned}$$

Kritisk verdi hentes fra normalfordelingen, sign.nivå 5% siden vi kjenner populasjonsstandardavviket $\sigma = 10$:

$$k = -1,645$$

$$TS = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{73,33 - 77}{10/\sqrt{12}} = -1,271$$

Siden $TS > k$, er den ikke i forkastningsområdet og vi beholder H_0 . Fartsnivået har ikke minket.

b)

98% konfidensintervall:

$$\begin{aligned}\bar{X} \pm 2,326 \cdot \sigma/\sqrt{n} \\ 73,33 \pm 2,326 \cdot 10/\sqrt{12} \\ 73,33 \pm 6,715 \\ [66,61 \pm 80,05]\end{aligned}$$

En hypotesetest kan være:

$$\begin{aligned}H_0 &: \mu = 77 \\ H_A &: \mu \neq 77\end{aligned}$$

- Et 98% konfidensintervall tilsvarer et 2% signifikansnivå.
- Konfidensintervallet inkluderer $\mu = 77$ slik at vi beholder H_0 ved et 2% signifikansnivå.
- Farten er uendret.

c)

Kritisk verdi hentes fra normalfordelingen, sign.nivå 5% siden vi kjenner utvalgsstandardavviket $\sigma = 10$:

$$k = t_{0,05}(11) = -1,7959$$

$$TS = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{73,33 - 77}{10/\sqrt{12}} = -1,271$$

Siden $TS > k$, er den ikke i forkastningsområdet og vi beholder H_0 . Fartsnivået har ikke minket.

d)

Hvis $TS < -1,7959$ vil konklusjonen endre seg.

Noe om at lavere standardavvik gjør at målingen er "sikrere" og at et lavere standardavvik derfor gir utvalgsgjennomsnittet mer "kraft".

$$\begin{aligned} \frac{73,33 - 77}{s/\sqrt{12}} &= -1,7959 \\ s/\sqrt{12} &= \frac{73,33 - 77}{-1,7959} \\ s &= \frac{73,33 - 77}{-1,7959} \sqrt{12} \\ s &= 7,08 \end{aligned}$$

Dersom utvalgsstandardavviket hadde vært $s = 7,08$ eller lavere, ville vi akseptert H_A

3

①

$$\textcircled{A} \quad b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}; \quad a = \bar{y} - b\bar{x}$$

hvor $X = T$
 $Y = D$

$$\bar{T} = 22 \quad \sum x_i y_i = 140$$

$$\bar{D} = 72 \quad \sum x_i^2 = 46$$

$$\Rightarrow b = \frac{140}{46} = 3.04 \sim 3$$

$$a = 72 - 3.04 \cdot 22 = 5.04 \sim 5$$

$$D = 5.04 + 3.04T$$

TOLKNING: FOR HVER GRAD T ØKER, ØKER
SALG VED 3 TRÆR

HVIS $T=0$, BLIR 5 TRÆR SOLGT

POSITIV RELASJON MELLOM D OG T

$$\textcircled{B} \quad T=4 \Rightarrow D = 5.04 + 3.04 \cdot 4 = 17.2$$

$$T=28 \Rightarrow D = 90.2$$

TROVERDIG? NOE OM AT T_4 OG T_{28} LIGGER
UTENFOR SAMPELET $T = [18, 26]$. FOR $T=28$
NOK UPROBLÆMATISK, FOR $T=4$ PROBLÆMATISK

D	T	gj. D	Gj. T	x_i	x_i^2	y_i	$x_i y_i$
92	22	72	22	0	0	20	0
70	23	72	22	1	1	-2	-2
72	19	72	22	-3	9	0	0
61	22	72	22	0	0	-11	0
75	24	72	22	2	4	3	6
50	18	72	22	-4	16	-22	88
84	26	72	22	4	16	12	48
SUM					46		140

b 3.04

a 5.04

4

A) KASUS

LESER DIREKTE FRA ARKET:

$$\hat{Y} = 62.86 - 0.22X \quad Y = \text{SØVN}$$

$$X = \text{ARBEID}$$

B) SKJÆRINGSPUNKTET.

BASERT PÅ DEN ESTIMERTE MODELLEN KAN VI FORVENTE AT PERSONEN SOM IKKE JOBBER (ALTSÅ $X=0$) SOVER 63 TIMER I LØPET AV UKEN

C) FORENDRING \Leftrightarrow 2. b

$$\Leftrightarrow 2 \cdot (-0.22) = \underline{\underline{-0.44}}$$

PERSONEN SOVER 0.44 TIMER ~~MINDRE~~ MINDRE

D) $Y_i = \hat{Y}_i + e_i$: OBSERVERT SØVN ER LIK
 $Y_1 = \hat{Y}_1 + e_1$ ESTIMERT + RESIDUAL

$$Y_1 = 53.3 - 4.2 = \underline{\underline{49.1}}$$

E) FORKLART VARIASJON : $\hat{Y}_i - \bar{Y}$ (ESTIMERT MINUS G.SNITT)

$$\Leftrightarrow \hat{Y}_2 - \bar{Y} = 51.0 - 53.4 = \underline{\underline{-2.4}}$$

RESTVARIASJON : $Y_2 - \hat{Y}_2$ (OBSERVERT MINUS ESTIMERT)

$$\Leftrightarrow 55 - 51.0 = \underline{\underline{4}}$$

⑤

$$R^2 = \frac{SSE}{SST} = \frac{SST - SSR}{SST}$$

$$= \frac{813 - 634}{813} = \underline{\underline{0.22}}$$

SVAR: ARBEID FORKLARER 22% AV
VARIASJON I SØVN

⑥