



# ECONnect

NTNU

## Faktor

- en eksamensavis utgitt av ECONnect



**Eksamensbesvarelse:**

**SØK1004 – Statistikk for økonomer**

Eksamen:

Våren 2009

Antall sider:

25



## 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](http://www.econnect-ntnu.no).

ECONnect består av følgende personer ved utgivelsestidspunkt:

Bjørn Bergholt (Leder)	<a href="mailto:bjorn@econnect-ntnu.no">bjorn@econnect-ntnu.no</a>
Elise Caspersen (Fagdagensvarlig)	<a href="mailto:elise@econnect-ntnu.no">elise@econnect-ntnu.no</a>
Pål Christian Vågbø (Bedriftansvarlig)	<a href="mailto:paal@econnect-ntnu.no">paal@econnect-ntnu.no</a>
Tormod Hagerup (Økonomi/Samfunnsøk.)	<a href="mailto:tormod@econnect-ntnu.no">tormod@econnect-ntnu.no</a>
Tiril Toftedahl	<a href="mailto:tiril@econnect-ntnu.no">tiril@econnect-ntnu.no</a>
Louis Dieffenthaler	<a href="mailto:louis@econnect-ntnu.no">louis@econnect-ntnu.no</a>
Tone Hedvig	<a href="mailto:tone@econnect-ntnu.no">tone@econnect-ntnu.no</a>
Ole Christian Grytten	<a href="mailto:ole@econnect-ntnu.no">ole@econnect-ntnu.no</a>

*Post- og besøksadresse:*

ECONnect, NTNU Dragvoll  
Institutt for samfunnsøkonomi  
Bygg 7, Nivå 5  
7491 Trondheim

*Organisasjonsnummer:*

NO 994 625 314

*Hjemmeside:*

[www.econnect-ntnu.no](http://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.*



**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.**

**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
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

$$\bar{X} = 148.25$$

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

- a) Bruk resultatene til å estimere forventningen  $\mu$  og standardavviket  $\sigma$ .  
b) Finn et 90% konfidensintervall for  $\mu$  når vi antar at  $\sigma = 3$  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  $\sigma = 3$  gram. Hva blir konklusjonen?
- d) Gjennomfør testen i oppgave c dersom du antar at  $\sigma$  er ukjent.
- e) Regn ut teststyrken for  $\mu=147.5$  gram og  $\mu=145$  gram. Forklar med ord hva den beregnede styrken betyr. Anta at  $\sigma = 3$  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:

$X \downarrow$	$Y \rightarrow$	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 aksene, og anvendt tid  $Y$  langs den vertikale aksene. Velg enheter slik at plottet blir lett leselig.

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

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

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

Estimeringen gir følgende resultater:

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

- d) Gi en praktisk tolkning av  $\beta$ . Er det mulig å gi en tilsvarende praktisk tolkning av  $\alpha$ ?
- e) Hvor god vil du si at modellen er?
- f) Beregn et 90% konfidensintervall for  $\beta$ . 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?

Denne kolonnen er  
forbeholdt sensor

Oppgave 1,

Fint vær	0,6	0,001
Regn vær	0,39	0,003
snøvær	0,01	0,009

$$P(\text{ulykke}) = 0,001 + 0,003 + 0,009$$

$$= \underline{\underline{0,013}}$$

Det er 1,3% sannsynlighet for å ha en bilulykke.

Gitt at det skjer en ulykke ( $U$ ), hva er sannsynligheten for at snøvær var medvirkende årsak?

$$P(U) = 0,013$$

$$P(S = \text{snøvær}) = 0,01$$

$$P(S \cap U) = P(\text{snøvær og ulykke samtidig}) = 0,009$$

$$P(S|U) = \frac{P(S \cap U)}{P(U)} = \frac{0,009}{0,013} = \underline{\underline{0,69}}$$

Det er 69% sannsynlighet for at ulykken ble forårsaket av snøvær (snøvær var en medvirkende faktor).

Denne kolonnen er  
forbeholdt sensor

Oppgave 2)

A = "Skjønnhetsfeil"

B = "Funksjonsfeil"

$$P(A) = 10\% = 0,1$$

$$P(B) = 8\% = 0,08$$

Feiltypene A og B er uavhengig av hverandre, dvs. at den ene hendelsen ikke påvirker sannsynligheten for at det andre skal hendre.

Antar videre at om en enhet har feil, påvirker ikke dette sannsynligheten for at andre enheter har feil.

cy

Sannsynligheten for at en enhet har både A og B feil:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$= 0,1 \cdot 0,08$$

$$= \underline{0,008}$$

Det er 0,8% sannsynlighet for at en enhet har begge feilene.

Denne kolonnen er  
forbeholdt sensor

Sannsynligheten for at en enhet bare har  
sløyenhetsteil:

$P(\bar{B})$ : dvs sannsynligheten for at B ikke  
sløyer.

$$P(\bar{B}) = 1 - P(B) = 1 - 0,08 = \underline{0,92}$$

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) \cdot P(\bar{B}) \\ &= 0,1 \cdot 0,92 \\ &= \underline{0,092} \end{aligned}$$

Det er 9,2% sannsynlighet for at bare  
A sløyer (dvs. A, men ikke B).

Sannsynligheten for at en enhet er feilfri  
er sannsynligheten for at ikke A ( $\bar{A}$ ) og  
ikke B ( $\bar{B}$ ) hender samtidig.

$$P(\bar{A}) = 1 - 0,1 = \underline{0,9}$$

$$P(\bar{B}) = 0,92$$

$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P(\bar{A}) \cdot P(\bar{B}) \\ &= 0,9 \cdot 0,92 \\ &= \underline{0,828} \end{aligned}$$

Det er 82,8% sannsynlighet for at en enhet  
er feilfri.



Denne kolonnen er  
forbeholdt sensor

by

Kaller en feilfri enhet for C.

$$P(C) = 0,828$$

Der som bedriften selger varene sine får den

80 kr. for C

60 kr. for A

B kan ikke selges (se dermed bort fra den).

Vet at Z er prisen som bedriften får for en tilfeldig valgt enhet.

$E(Z)$  er forventningsverdien til Z. Dvs. hvor mye forventer vi at prisen på ~~to~~ en tilfeldig vare vil være når vi har to ulike utfall (kan enten trekke ut en tilfeldig enhet som har feil eller en som er feilfri).

Forventningsverdier kalles også middelveier og gir gjennomsnittlig pris (her) når vi trekker ut mange enheter.

$$\begin{aligned} E(X) &= \sum x \cdot p(x) \\ &= \sum z \cdot p(z) \end{aligned}$$

$$\begin{aligned} E(Z) &= 80 \cdot 0,828 + 60 \cdot 0,1 \\ &= \underline{\underline{72,24}} \end{aligned}$$

Denne kolonnen er  
forbeholdt sensor

$$E(Z) = \mu_z = \underline{\underline{72,24}}$$

dvs. at vi forventer en pris på 72,24 kr.

et uttrykk for

Var(Z) gir variansen til prisen, dvs. spredningsmålet.

$$\begin{aligned} \text{Var}(Z) &= \sum_x x^2 p(x) - \mu_x^2 \\ &= \sum_z z^2 p(z) - \mu_z^2 \end{aligned}$$

$$\begin{aligned} \sum_z z^2 p(z) &= 80^2 \cdot 0,828 + 60^2 \cdot 0,1 \\ &= \underline{\underline{5659,2}} \end{aligned}$$

$$\begin{aligned} \text{Var}(Z) &= \sum_z z^2 p(z) - \mu_z^2 \\ &= 5659,2 - (72,24)^2 \\ &= 5659,2 - 5218,62 \\ &= \underline{\underline{440,6}} \end{aligned}$$

Variansen er 440,6.

Dette kan brukes til å finne standardavvik ( $\sqrt{\text{Var}(Z)}$ ) som gir faktisk spredningsmål, dvs. gjennomsnittlig avvik fra middelverdien.

Denne kolonnen er  
forbeholdt sensor

G) Enheter produsert =  $n = 800$

$$E(Z) = \mu = 72,24$$

$$\text{Var}(Z) = \sigma_x^2 = 440,6$$

$$\text{Std. avvik} = \sigma_z = 21$$

X - antall feilne enheter

X er binomisk fordelt fordi det er en sekvens med uavhengige (oppgitte) bernoulliforsøk. Bernoulliforsøk er et tilfeldig eksperiment med kun to utfall (suksess eller fiasko). Ettersom X er antall feilne, vil det "motsatte" være antall enheter med feil (her går både A og B sammen).

Benytter normalfordelingen til å finne en tilnærming av binominalsannsynligheten:

X - antall feilne enheter (dvs. suksess)

$$n = 800$$

$$\pi = \text{sannsynligheten for suksess (X) i hvert forsøk} \\ = 0,09035 \quad (\text{fra oppg. 2a})$$

Krav på å bruke denne tilnærmingen:

$$n\pi > 5 \quad \text{og} \quad n(1-\pi) > 5$$

Denne kolonnen er  
forbeholdt sensor

$$800 \cdot 0,828 = 662,4 > 5$$

$$800 \cdot (1 - 0,828) = 137,6 > 5$$

Dette gir at  $X$  er normalfordelt med  $\mu = n\pi = 662,4$   
og  $\text{var}(X) = \sigma^2 = 137,6$ .

$$X \sim N(662,4 \rightarrow 137,6)$$

Tester om vi kan benytte korrelasjonsfaktorer (dvs.  
utvide intervallet med 0,5 i hver retning).

Krav:

$$5 < n\pi(1-\pi) < 9$$

$$n\pi(1-\pi) = 662,4 \cdot 0,172 = 113,93$$

Benytter ikke korrelasjonsfaktorer.

$$P(X > 650)$$

$$P\left(\frac{X - \mu}{\sigma} > \frac{650 - \mu}{\sigma}\right)$$

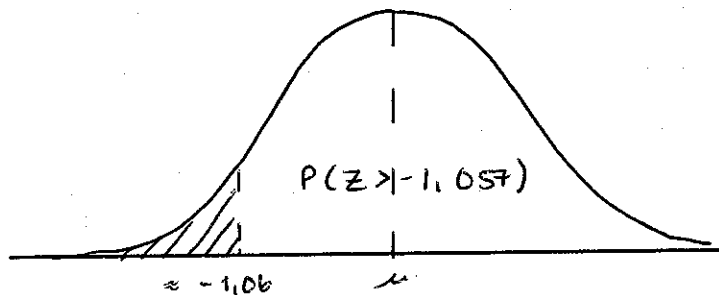
$$\text{Har at } \frac{X - \mu}{\sigma} = Z.$$

$$P\left(Z > \frac{650 - 662,4}{\sqrt{137,6}}\right)$$

Denne kolonnen er  
forbeholdt sensor

$$P(Z > -1,057)$$

Benyttar kumulativ standardnormalfordelings-  
tabell:



$$\begin{aligned} P(Z > -1,057) &= 1 - F(-1,057) \\ &= 1 - F(-1,06) \\ &= 1 - 0,1446 \\ &= \underline{0,8554} \end{aligned}$$

Det er 85,54 % sannsynlighet for at  $X$  er  
store enn 650 enheter.

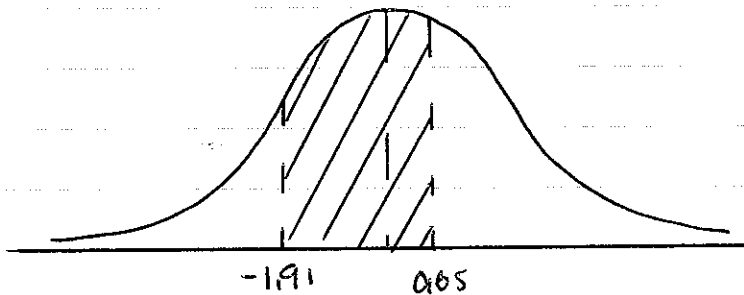
$$P(640 < X \leq 670)$$

$$P\left(\frac{640 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{670 - \mu}{\sigma}\right)$$

$$P\left(\frac{640 - 662,4}{\sqrt{1376}} < Z \leq \frac{670 - 662,4}{\sqrt{1376}}\right)$$

Denne kolonnen er  
forbeholdt sensor

$$P(-1,91 < Z \leq 0,65)$$



$$\begin{aligned} P(-1,91 < Z \leq 0,65) &= F(-1,91) - F(0,6) \\ &= F(0,65) - F(-1,91) \\ &= 0,7422 - 0,0281 \\ &= \underline{0,7141} \end{aligned}$$

Det er 71,41 % sannsynlighet for at  
X er mellom 640 og 670 enheter.

Denne kolonnen er  
forbeholdt sensor

### Oppgave 3j

$X$  er normalfordelt  
der  $X$  viser målt vekt

$$\mu = 150$$

$$\sigma = 3$$

Uttvalg:

$$n = 16$$

$$\bar{x} = 148,25$$

$$\sum (x_i - \bar{x})^2 = 177$$

og Antar at jeg skal bruke resultatene fra  
målingene til å finne  $\mu$  og  $\sigma$  for utvalget:

$$\begin{aligned} \mu &= E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \end{aligned}$$

$$= \frac{1}{16} (146 + 148 + 154 + 152 + 144 + 145 + 144 + 153 + 145 + 148 + 146 + 153 + 147 + 148 + 152 + 147)$$

$$= 2372 \cdot \frac{1}{16}$$

$$= \underline{148,25}$$

ser at  $E(\bar{x}) = \mu$

Dvs. at gjennomsnittet for utvalget er forventningsrett.

Denne kolonnen er  
forbeholdt sensor

~~i forhold til populasjonen~~

For å finne et forventningsrett standardavvik  
benytte 3:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Her oppgitt  $\sum (x_i - \bar{x})^2 = 177$

$$s^2 = \frac{177}{15} = 11,8$$

$$s = \sqrt{11,8} = \underline{3,435}$$

$$\underline{\underline{\mu = 148,25 \text{ og } \sigma = s = 3,44}}$$

by Konfidensintervall:

Ettersom  $n$  er liten og vi må estimere  $s^2$ , benytter  
jeg  $t$ -fordeling. Før da følgende konfidensintervall:

$$\bar{x} \pm t(16)_{0,1/2} \cdot \frac{s}{\sqrt{16}}$$

$$148,25 \pm 1,753 \cdot 0,75$$

$$148,25 \pm 1,315$$

$$\rightarrow \underline{\underline{[146,935 \text{ } \rightarrow \text{ } 149,565]}}$$



Denne kolonnen er forbeholdt sensor

$\mu$  vil med 90% sikkerhet ligge innenfor dette intervallet.

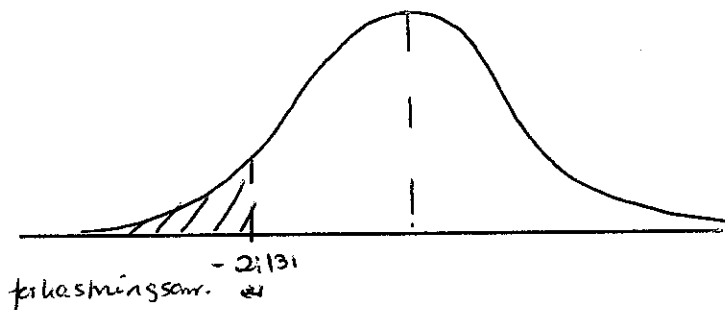
g) Hypotesetest:  $H_0: \mu = 150$   
 $H_A: \mu < 150$  (ensidig test).

$$\text{Sign. nivå} = \alpha = 0,025$$

Kritisk verdi:  $t(n-1)\alpha = t(15)_{0,025} = 2,131$   
 (pga.  $\mu < 150$  vil vi ha kritisk verdi like  $-2,131$ ).

Ferkesst  $H_0$  når  $TS < -2,131$ .

$$TS = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1) \text{ dersom vi har at } H_0 \text{ gjelder.}$$



$$TS = \frac{148,25 - 150}{3/\sqrt{16}} = -2,33.$$

Se at  $TS = -2,33 < -2,131$ . Dvs. at vi ferkesster.

Denne kolonnen er  
forbeholdt sensor

$H_0$  med 2,5% sannsynlighet for å ta feil.  
 Det ser ut som om posere på markedet  
 i gjennomsnitt inneholder mindre enn 150 gr.  
 Dette underbygges også av konfidensintervall,  
 eller som  $\mu = 150$  ikke inngår der med 90%  
 sikkerhet.

da dersom  $\sigma$  er ulikt benyttes  $s$ .

Benytter  $t$ -fordelingen, men har  $s = 3,44$   
 istedenfor  $\sigma = 3$ .

$$H_0: \mu = 150$$

$$H_A: \mu < 150$$

Forkaster  $H_0$  dersom  $T_S < -2,131$ .

$$T_S = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{148,25 - 150}{3,44/\sqrt{16}} = -2,03$$

Ser at  $T_S = -2,03 > -2,131$ .

Dvs. at vi ikke har forkastet  $H_0$ .

Det ser ikke ut til at populasjonsgjennomsnittet  
 er mindre enn 150 utifra dette utvalget.

Denne kolonnen er forbeholdt sensor

ej Test styrken viser sannsynligheten for at vi har tatt riktig avgjørelse.

$P(\text{forkaste } H_0 \mid H_0 \text{ er sann})$

~~= 1~~

Vet at vi forkaster  $H_0$  dersom  $T_5 < -2,131$ .

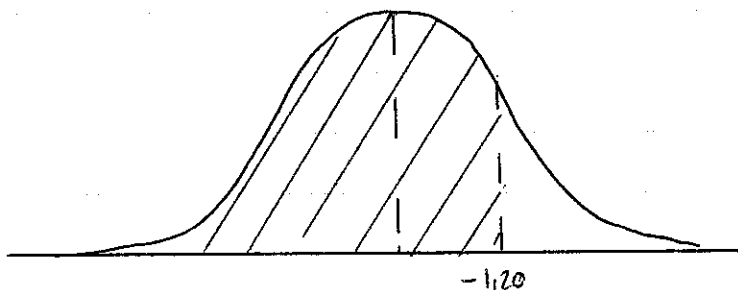
$$\begin{aligned}
 P(\text{forkaste } H_0 \mid H_0 \text{ er sann}) &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < -2,131 \mid \frac{\bar{X} - \mu^*}{\sigma/\sqrt{n}} \sim N(0,1)\right) \\
 &= P\left(\frac{\bar{X} - \mu - \mu^*}{\sigma/\sqrt{n}} < -2,131 - \frac{\mu^*}{\sigma/\sqrt{n}} \mid \frac{\bar{X} - \mu^*}{\sigma/\sqrt{n}} \sim N(0,1)\right) \\
 &= P\left(\frac{\bar{X} - \mu^*}{\sigma/\sqrt{n}} < -2,131 - \frac{\mu^* + \mu}{\sigma/\sqrt{n}} \mid \frac{\bar{X} - \mu^*}{\sigma/\sqrt{n}} \sim N(0,1)\right) \\
 &= P\left(Z < -2,131 + \frac{150 - \mu^*}{0,75}\right)
 \end{aligned}$$

Der  $\mu^*$  er den riktige  $\mu$  (dvs.  $H_0$  er rett).

Her at :

1)  $\mu^* = 147,5$

$$\begin{aligned}
 \text{Test styrken} &= P\left(Z < -2,131 + \frac{150 - 147,5}{0,75}\right) \\
 &= P(Z < 1,20)
 \end{aligned}$$



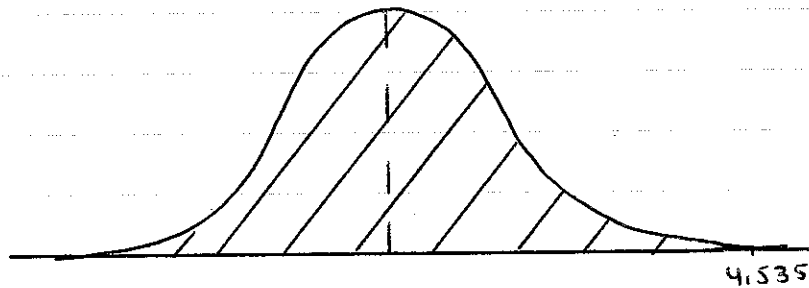
$$P(Z < 1,20) = F(1,20) = \underline{0,8849}$$

Denne kolonnen er forbeholdt sensor

Dette betyr at vi har 88,5% sannsynlighet for å forkaste  $H_0$  når  $H_A$  er rett (dvs. gjøre riktig).

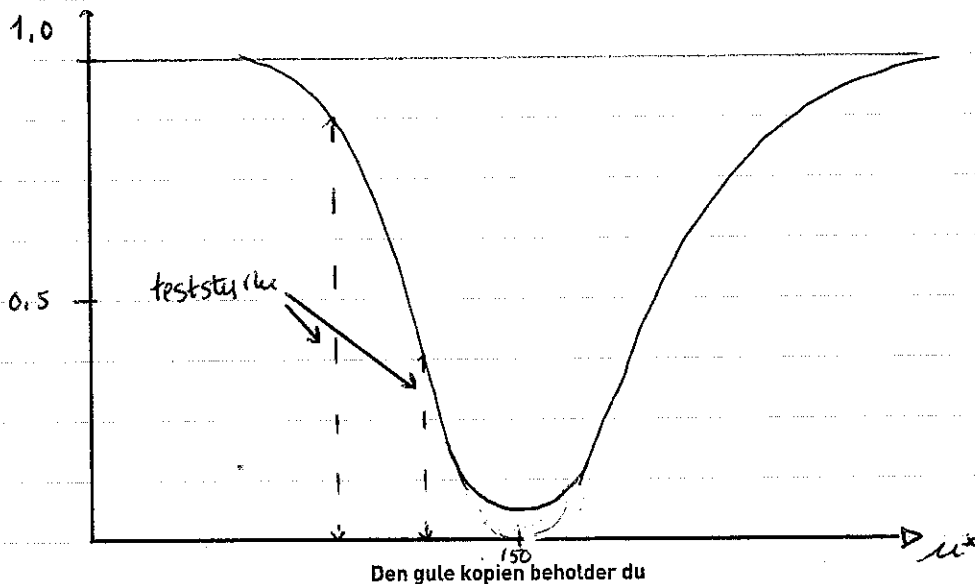
$$ii) \mu^* = 145$$

$$\begin{aligned} \text{Teststyrken} &= P\left(Z < -2,131 + \frac{150 - 145}{0,75}\right) \\ &= P(Z < 4,535) \end{aligned}$$



$$\begin{aligned} P(Z < 4,535) &= F(4,535) \\ &\approx 1 \end{aligned}$$

Det er 100% sannsynlighet for å forkaste  $H_0$  når den egentlige  $\mu^* = 145$ .



Denne kolonnen er  
forbeholdt sensor

ser av fig. at teststyrken (sannsynligheten for å velge riktig) øker jo lenger  $\mu^*$  (den egentlige  $\mu$ ) kommer fra den antatte  $\mu$  (her = 150).

Dette stemmer med resultatet fra oppgaven der vi ser at  $\mu^* = 145$  har en større teststyrke enn  $\mu^* = 147,5$ .

f) Skal teste kvadratt standardavviket ( $\sigma$ ) er mindre enn 4, dvs om variansen  $\sigma^2$  er mindre enn 16:

$$\text{Hypotese: } H_0: \sigma^2 = 16$$

$$H_A: \sigma^2 < 16 \text{ (ensidig test)}$$

Benyttar kjikvadratfordelinger:

Under  $H_0$  er

$$T_S = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2$$

Har sign.nivå  $\alpha = 0,025$

Benyttar kjikvadratfordelinger og finner kritiske verdi. Ettersom vi har  $H_A <$ , vil vi få:

~~$$\chi^2_{0,025} (16-1) = 27,1$$~~

$$\chi^2_{1-0,025} (16-1) = \chi^2_{0,975} (15) = 6,262$$

Denne kolonnen er  
forbeholdt sensor

Dvs. at vi forkaster  $H_0$  dersom  $T_3$  er mindre enn  
kritisk verdi:  $T_3 < 6,262$ .

$$T_3 = \frac{s^2(n-1)}{\sigma^2} = \frac{(3,44)^2 \cdot 15}{3^2}$$

$$= \frac{177,5}{9} = \underline{19,72}$$

Ser at  $T_3 = 19,72 > 6,262$ .

Dvs. at vi beholder  $H_0$ . Fra utvalget kan det  
se ut som om standardavviket er 4.

Denne kolonnen er forbeholdt sensor

### Oppgave 4,

g)

$X \backslash Y$	0	2	4	$f(x)$
1	0,04	0,04	0,02	0,1
2	0,17	0,31	0,12	0,6
3	0,09	0,15	0,06	0,3
$g(y)$	0,3	0,5	0,2	1

Kovarians er et mål på lineær sammenheng mellom variablene  $X$  og  $Y$ .

Dersom kovariansen er negativ vil sannsynligvis en høy verdi på  $X$  henge sammen med en lav verdi på  $Y$ .

Dersom kovariansen er positiv vil i sannsynligvis en høy verdi på  $Y$  sammen med en høy verdi på  $X$ .

Dersom variablene er uavhengig har vi at kovariansen er null. Det kommer av at

$$E(XY) = E(X) \cdot E(Y) \text{ for alle verdier av } X \text{ og } Y.$$

$$\text{COV}(X, Y) = E(XY) - \mu_X \mu_Y$$

$$\begin{aligned}
 E(XY) &= \sum_x \sum_y x \cdot y \cdot p(x, y) \\
 &= 1 \cdot 0 \cdot 0,04 + 1 \cdot 2 \cdot 0,04 + 1 \cdot 4 \cdot 0,02 + 2 \cdot 0 \cdot 0,17 + 2 \cdot 2 \cdot 0,31 + \\
 &\quad 2 \cdot 4 \cdot 0,12 + 3 \cdot 0 \cdot 0,09 + 3 \cdot 2 \cdot 0,15 + 3 \cdot 4 \cdot 0,06 \\
 &= \underline{\underline{3,36}} \quad \underline{\underline{3,98}}
 \end{aligned}$$

Denne kolonnen er  
forbeholdt sensor

$$\begin{aligned}\mu_x &= E(X) = \sum_i x_i \cdot f(x_i) \\ &= 1 \cdot 0,1 + 2 \cdot 0,6 + 3 \cdot 0,3 \\ &= \underline{2,2}\end{aligned}$$

$$\begin{aligned}\mu_y &= E(Y) = \sum_j y_j \cdot g(y_j) \\ &= 0 \cdot 0,3 + 2 \cdot 0,5 + 4 \cdot 0,2 \\ &= \underline{1,8}\end{aligned}$$

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - \mu_x \mu_y \\ &= 3,98 - (2,2 \cdot 1,8) \\ &= 3,98 - 3,96 \\ &= \underline{0,02}\end{aligned}$$

Vi har ikke uavhengige variabler eller som  
 $E(X) \cdot E(Y) \neq E(XY)$ , og  $\text{Cov}(X, Y) \neq 0$ .

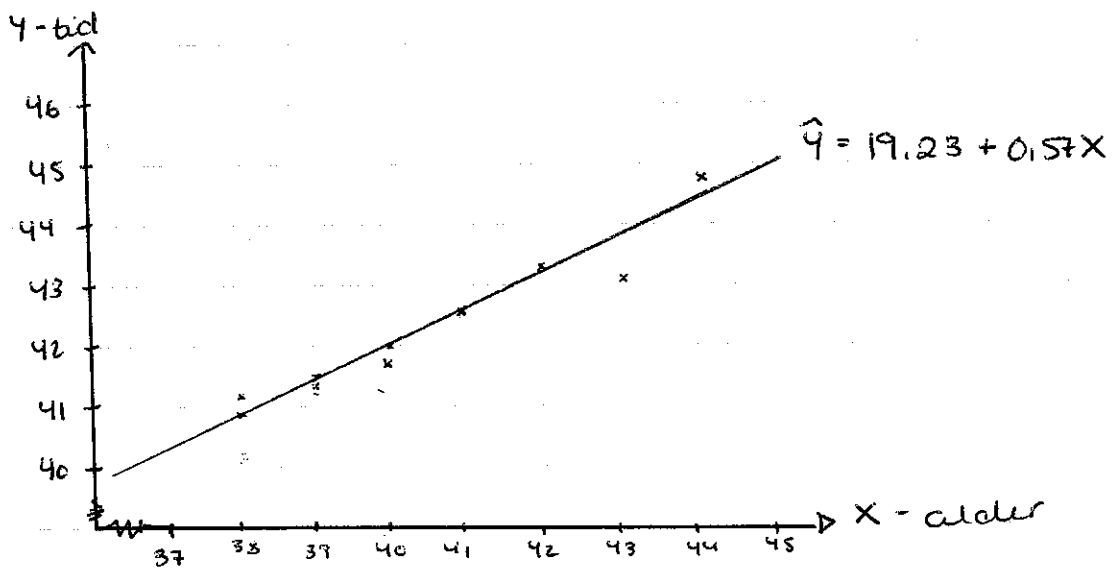
Ser at det er en positiv sammenheng  
mellom  $X$  og  $Y$ .



Denne kolonnen er forbeholdt sensor

## Oppgave 5)

X	38	39	40	41	42	43	44
Y	41,2	41,3	41,7	42,6	43,3	43,1	44,8


 b) Antar at dataene kan beskrives med :  $Y = \alpha + \beta X + \varepsilon$ 

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

utvalgskorrelasjonen R gis ved:

$$R = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}$$

$$= \frac{16}{\sqrt{28} \cdot \sqrt{10}} = \frac{16}{16,73} = \underline{0,96}$$

Denne kolonnen er  
forbeholdt sensor

Utrektskorrelasjonen = 0,96 sier at det er en sterkt positiv samvariansjon mellom 44-åringens alder og anvendt tid på løpere. Dette stemmer godt med plottene i diagrammet. Dvs. at jo høyere alder han får (blir eldre), jo lengre tid bruker han på løpet,  $X$  og  $Y$  stiger sammen.

g) Benytter minste kvadratsmetode (OLS):

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{16}{28} = \underline{0,57}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Finner  $\bar{y}$  og  $\bar{x}$ :

$$\bar{y} = \frac{\sum_{i=1}^n y}{n} \approx \underline{42,6}$$

$$\bar{x} = \frac{\sum_{i=1}^n x}{n} = \underline{41}$$

$$\hat{\alpha} = 42,6 - 0,57 \cdot 41$$

$$= \underline{19,23}$$

Får følgende estimerte regresjonslinje:

$$\hat{y} = 19,23 + 0,57x + \varepsilon$$

Denne kolonnen er forbeholdt sensor

Setter opp en tabell og tegner inn linjer i diagrammet :

X	38	39	40	41	42	43	44
$\hat{y} = 19,23 + 0,57x$	40,89	41,46	42,03	42,6	43,17	43,74	44,31

Ved å benytte OLS - metode finner jeg den linjen som passer best til spredningsplottet.

Det skjer ved å minimere  $\sum e^2$ , dvs. minimere de kvadrerte avvikene fra plottene vi har oppgitt.

Denne estimeringen gir :

$$\sum_{i=1}^7 (y_i - \hat{y}_i)^2 = 0,891, \quad \sum_{i=1}^7 (y_i - \bar{y})^2 = 10,034, \quad \sum_{i=1}^7 (\hat{y}_i - \bar{y})^2 = 9,143$$

der  $\beta$  er stigningstallet og gir "sammenheng" mellom  $X$  og  $Y$ . Dersom  $\beta > 0$  stiger  $Y$  med  $X$ , mens dersom  $\beta < 0$  synker  $Y$  når  $X$  øker.

$\alpha$  gir skjæringspunktet med  $Y$  akser.

Dersom  $X = 0$ , vil  $Y = \alpha$ .

I dette tilfellet ser vi at  $\beta > 0$ , og vi har en positiv sammenheng mellom  $X$  og  $Y$ . Ettersom  $\beta = 0,57$  vil  $Y$  øke med  $0,57$  når  $X$  øker med 1 enhet.

$\alpha$  er her  $19,23$ . Det vil si at når  $X = 0$  er  $Y = 19,23$ . Dette viser at modellen ikke kan

Denne kolonnen er  
forbeholdt sensor

benyttes for alle  $X$ . Det er tydelig at linjen vi fant er optimal for spredningsplottet, men antakelig ikke for resten (alle andre  $X$ -verdier).

ej For å si noe om hvor god modellen er benyttes  $R$ -kvadrat ( $R^2$ ).

Denne viser hvor stor del av plottene den estimerte regresjonslinjen fanger opp.

Der som:

$R^2 = 0$  har vi en svært dårlig modell, den fanger opp 0%.

$R^2 = 90$  har vi en perfekt modell som fanger opp 100% av plottene.

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

$SSE$  gir feilaktig variasjon =  $\sum_{i=1}^7 (\hat{y}_i - \bar{y})^2 = 9,143$

$SST$  gir total variasjon =  $\sum_{i=1}^7 (y_i - \bar{y})^2 = 10,034$

$$R^2 = \frac{9,143}{10,034} = \underline{\underline{0,91}}$$

Ser at dette er en god modell (regresjonslinjen fanger opp 91% av dataene / plottene).

Den klarer å forklare 91% av variasjonene.

Denne kolonnen er  
forbeholdt sensor

f) Skal beregne et 90% konfidensintervall for  $\beta$ :

$$\rightarrow \alpha = 0,1$$

$$\hat{\beta} \pm t_{\alpha/2} \cdot s_{\hat{\beta}}$$

$s_{\hat{\beta}}$  er standard feil til  $\hat{\beta}$ . For å finne dette benyttes:

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum (x_i - \bar{x})^2}$$

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{0,891}{5} = 0,1782$$

$$s_{\hat{\beta}}^2 = \frac{0,1782}{28} = 0,0064$$

$$s_{\hat{\beta}} = \sqrt{0,0064} = \underline{0,08}$$

Detta gir:

$$0,57 \pm t_{(7-2)0,05} \cdot 0,08$$

$$0,57 \pm 2,015 \cdot 0,08$$

$$0,57 \pm 0,1612$$

$$\underline{[0,4088 \rightarrow 0,7312]}$$

Denne kolonnen er  
forbeholdt sensor

Dette gir den virkelige  $\beta$  med 90% sikkerhet.  
 dvs. at  $\beta$  ligger mellom 0,4088 og 0,7312  
 med 90% sikkerhet.

Deresom vi istedet hadde benyttet et 95%  
 konfidensintervall ville intervallet vært større enn  
 det for 90%. Det er fordi prosentandelen  
 høyner, vi må dermed utvide intervallet for å  
 kunne fange opp en større sikkerhet.  
 Dvs. at for å øke sikkerheten for at  $\beta$  ligger  
 i intervallet, må vi utvide intervallet.  
 Dvs. at jo større / sikkerere % - nis konfidensintervall,  
 jo større må selve intervallet være.

g) Antar at modellen også brukes neste år.

Da er  $X = 45$ .

setter dette inn i  $\hat{Y} = 19,23 + 0,57X$

$$\begin{aligned}\hat{Y} &= 19,23 + 0,57 \cdot 45 \\ &= \underline{44,88}\end{aligned}$$

Forventer at mannen vil løpe på 44,88,  
 dvs. 44,53 minutter