

Denne kolonne er forbeholdt sensor

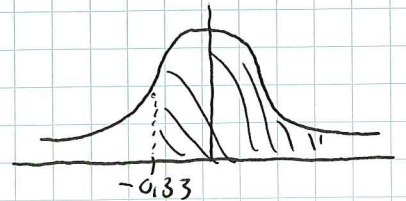
This column is for external examiner

Eksamen statistikk for økonomer høsten 2018
Oppgave 1:

 a) La X være normalfordelt med $\mu = 3$ og $\sigma = 3$

 Hva er $P(X > 2)$?

$$X \sim N(3, 3^2) \Rightarrow Z = \frac{2-3}{3}$$



$$\begin{aligned} P(Z > -0,33) &= 0,5 + P(0 < Z < 0,33) \\ &= 0,5 + 0,1293 = 0,6293 \approx 0,63 \end{aligned}$$

c) 0,63

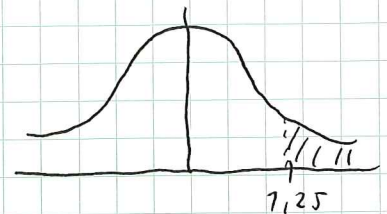
b) Mobilabonnement som er gunstig dersom total ringetid i løpet av en måned er under 275 min.

En potensiell kunde har ringetid

$$X \sim N(250, 20^2), \quad \mu = 250, \quad \sigma = 20$$

 Hva er $P(X > 275)$?

$$\begin{aligned} P(Z > \frac{275-250}{20}) &= P(Z > 1,25) \\ &= 0,5 - P(0 < Z < 1,25) \\ &= 0,5 - 0,3944 = 0,1056 \approx 0,11 \end{aligned}$$


A) 0,11

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 1- fortsettelse:

c) 18% strøk en eksamen. Av de som strøk var 40% kvinner. Av alle opppe til eksamen var 53% kvinner. Trekk ut et tilfeldig utvalg student fra denne eksamenen, som viser seg å være en mann. Hva er sannsynligheten for at han strøk på eksamen gitt at han er mann?

M = Mann S = strøk

$$P(S) = 0,18 \quad P(M) = 1 - 0,53 = 0,47$$

$$P(\bar{M}|S) = 0,4 \quad \Rightarrow \quad P(M|S) = 0,6$$

Skal finne $P(S|M)$.

$$P(S|M) = \frac{P(M|S) \cdot P(S)}{P(M)} = \frac{0,6 \cdot 0,18}{0,47} \approx 0,23$$

B) 0,23

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 1 - fortsettelse:

d) Hva er sannsynligheten for at Summen av to terningkast er 6, hvis det er gitt at ingen av terningene viser 1?

Kast 1	2	4	3
Kast 2	4	2	3
Sum	6	6	6

$$3 \cdot \frac{1}{5} \cdot \frac{1}{5} = 3 \cdot \frac{1}{25} = 0,12$$

D) 0,12

e) X = antall feilleveringer på en dag av Dagsposten
Dagsposten har 100 avisbud

x	0	1	2	3
$P(x)$	0,155	0,25	0,175	0,105

Hva er forventet antall feilleveringer av Dagsposten på en tilfeldig dag?

$$E(x) = \sum x_i \cdot P(x_i) = 0 \cdot 0,155 + 1 \cdot 0,25 + 2 \cdot 0,175 + 3 \cdot 0,105 = 0,7$$

$$0,7 \cdot 100 \text{ bud} = 70$$

A) 70

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 2:

Erfaring viser at 12,4% av de personene som reserverer en flyplass på et bestemt fly ikke kommer. Anta at flyet har plass til 280 passasjerer og at 300 har reservert plass. Hva er sannsynligheten for at alle som har reservert og kommer får plass?

Binomisk sannsynlighetsfordeling som kan tilnærmas normal fordeling:

- To utfall
- Konstant sannsynlighet for suksess
- Uavhengige delforsøk

$$n = 300, \quad \bar{p} = 1 - 0,124 = 0,876$$

$$n\bar{p} = 300 \cdot 0,876 = 262,8 > 5$$

$$n(1-\bar{p}) = 300 \cdot 0,124 = 37,2 > 5$$

$$X \sim N(\mu, \sigma), \quad X = \text{antall reservert og kommer opp}$$

$$\mu = n\bar{p} = 300 \cdot 0,876 = 262,8$$

$$\sigma = \sqrt{\bar{p}n(1-\bar{p})} = \sqrt{300 \cdot 0,876 \cdot (1 - 0,876)} = 5,709$$

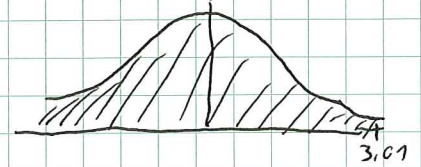
$$X \sim N(262,8, 5,709)$$

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 2 - fortsettelse:

Skal finne



$$P(X \leq 280) \Rightarrow P\left(Z < \frac{280 - 262,8 + 0,5}{5,709}\right)$$

$$= P(Z < 3,1) = 0,5 + P(0 < Z < 3,1)$$

$$\approx 0,5 + 0,5 \approx \underline{1} \quad (P(0 < Z < 3,1) \approx 0,5)$$

Det er tilnærmet 100% sannsynlig at alle de som har reservert plass og kommer vil få en plass.

Oppgave 3:

Samler inn pris på 41 tilfeldige pensumbøker.

X = prisen på pensumboka

$$n = 41, \quad \sum_{i=1}^n x_i = 24\,969, \quad \sum_{i=1}^n (x_i - \bar{x})^2 = 475\,240$$

a) Bruk utvalget til å finne forventningsrette estimatorene for gjennomsnitt μ og varians σ^2 til prisen på pensumbøker.

$$\underline{\underline{\mu}} = \frac{\sum x_i}{n} = \frac{24\,969}{41} = \underline{\underline{609}} = \bar{X}$$

$$\underline{\underline{\sigma^2}} = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{475\,240}{41-1} = \underline{\underline{11\,881}} = S^2$$

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 3 - fortsettelse:

a) Gjennomsnittsprisen på en persimmon er allerede estimert til å være 609 kr med en varians på 11 881 kr² (standardavvik 109 kr)

b) Bruk utvalget til å lage et 95% KI for den samme populasjonsgjennomsnittet, μ .

95% KI er gitt ved:

$$\bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

$$\bar{X} = 609, S = \sqrt{S^2} = \sqrt{11881} = 109, n = 41$$

$$Z_{\alpha/2} = Z_{0,025} = 1,96$$

$$609 \pm 1,96 \cdot \frac{109}{\sqrt{41}}$$

$$609 \pm 1,96 \cdot 17,0229$$

$$609 \pm 33,365$$

95% KI: [575,635, 642,365]

[575,64, 642,37]

95% KI: lager vi uendelig mange KI basert på samme populasjon med $n=41$, vil 95% av de inneholde den samme μ .

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 3 - fortsettelse:

c) Ifølge NTNU er gj. prisen på pensumbøker 585 kr. Velferdstinget mener at NTNU undrestimerer gj. prisen. Bruk utvalget til å teste dette med $\alpha = 0,05$.

Steg 1 - oppsett:

μ = gjennomsnittlig pris pensum bøker NTNU

$$H_0: \mu = 585 \text{ kr}$$

$$H_A: \mu > 585 \text{ kr}$$

Steg 2 - testobservator:

$n < 50$ og ukjent σ

$$\Rightarrow TS = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$

$$\alpha = 0,05$$

Steg 3 - data:

$$\bar{x} = 609, s = 109, n = 41, \mu_0 = 585$$

$$TS = \frac{609 - 585}{\frac{109}{\sqrt{41}}} = 1,4099 \sim t(41-1)$$

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 3 - fortsettelse:
Steg 4 - testkriterie:

 Forhaster H_0 dersom $TS > t_{\alpha}(n-1)$

 Skal finne $t_{0,05}(40)$, bruker tabell A.2

$$t_{0,05}(40) = 1,684$$

Steg 5 - konklusjon:

$$TS = 1,4099 < 1,684 = t_{0,05}(40)$$

$$TS < t_{\alpha}(n-1)$$

 \Rightarrow H_0 kan ikke forkastes på 5% signifikansnivå
Med 5% signifikansnivå stemmer ikke påstanden til
Velferdstinget om at ledelsen undrestimerer
gjennomsnittsprisen på pensumbøker.

d) Hva er testens p-verdi? Hva er tolkingen av denne?

$$\begin{aligned} \text{p-verdi} &= P(Z > TS) = P(Z > 1,41) \\ &= P(Z > 1,41) = 0,5 - P(0 < Z < 1,41) \\ &= 0,5 - 0,4207 = \underline{0,0793} \end{aligned}$$


Testens p-verdi er 0,0793 som tilsvarer signifikansnivå 7,93%.
Den sier at med de gitte data (TS) er laveste signifikansnivå der H_0 kan forkastes 7,93%. (Hvis velferdstinget skulle hatt rett i sin påstand måtte de ha brukt a større eller enn 0,0793)

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4:
 Y = antall timer studietid per uke

 X = antall timer i lønnet arbeid per uke

Tilfeldig utvalg på 10 studenter

Y	40	35	34	36	40	32	44	46	48	45
X	7.5	15	7.5	4	0	10	0	4	2	0

$$\bar{x} = 5$$

$$\bar{y} = 40$$

$$\sum (x_i - \bar{x})^2 = 223,5$$

$$\sum (y_i - \bar{y})^2 = 282$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = -176$$

a) Hva er korrelasjonen mellom antall timer studietid og antall timer i lønnet arbeid?

Utvalgskorrelasjon er gitt ved

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}}$$

$$\underline{\underline{R}} = \frac{-176}{\sqrt{223,5} \cdot \sqrt{282}} = \underline{\underline{-0,701}}$$

Korrelasjonen mellom antall timer studietid og lønnet arbeid er -0,701

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4 - fortsettelse:

b) Er korrelasjonen statistisk signifikant?

Velg et 5% signifikansnivå.

Steg 1 - oppsett:

ρ = korrelasjon

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

Steg 2 - testobservator:

Hypotesetest om korrelasjon

$$\Rightarrow TS = \frac{R \sqrt{n-2}}{\sqrt{1-R^2}} \sim t(n-2)$$

$$\alpha = 0,05$$

Steg 3 - data:

R har vi fra oppgave a

$$R = -0,701, \quad n = 10$$

$$TS = \frac{-0,701 \cdot \sqrt{10-2}}{\sqrt{1 - (-0,701)^2}} = -2,78 \sim t(10-2)$$

Steg 4 - forkastningskriterie:

Forkaster H_0 dersom $|TS| > t_{\alpha/2}(n-2)$

$$t_{0,05/2}(10-2) = t_{0,025}(8), \text{ bruk tabell A.2}$$

$$t_{0,025}(8) = 2,306$$

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 41 - fortsettelse:

Steg 5 - Konklusjon:

$$|T_S| = | -2,78 | = 2,78 > 2,306 = t_{\alpha/2}(n-2)$$

$$|T_S| > t_{\alpha/2}(n-2)$$

$\Rightarrow H_0$ forkastes

Korrelasjonen på 0, som satt i H_0 , er statistisk usignifikant korrelasjon. Siden H_0 forkastes på 5% signifikansnivå kan vi dermed si at korrelasjonen er statistisk signifikant.

Antar at sammenhengen mellom Y og x er lineær:

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

c) Hva er tolkningen av β ?

β er stigningskoeffisienten i den lineære funksjonen.

Den angir hvor mange enheter Y endres med (øker/redukes)

når x øker (eller minner) en enhet.

Negativ β gir reduksjonen i Y når x øker

en enhet

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4 - fortsettelse:

d) Bruk utvalget og OLS til å estimere α og β . Illustrer utvalgets regresjonslinje i et diagram.

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

$$\underline{b} = \frac{-176}{223,5} = \underline{\underline{-0,787}}$$

$$\alpha: a = \bar{y} - b\bar{x}$$

$$\underline{a} = 40 - (-0,787) \cdot 5 = 43,935 \approx \underline{\underline{43,94}}$$

$$\underline{\hat{y}} = 40 - 0,787x$$

x	0	10	20	30	40	50,83
\hat{y}	40	32,13	24,26	16,39	8,52	0

\hat{y} - verdene er estimat/prednot med

$$\hat{y} = 40 - 0,787x$$

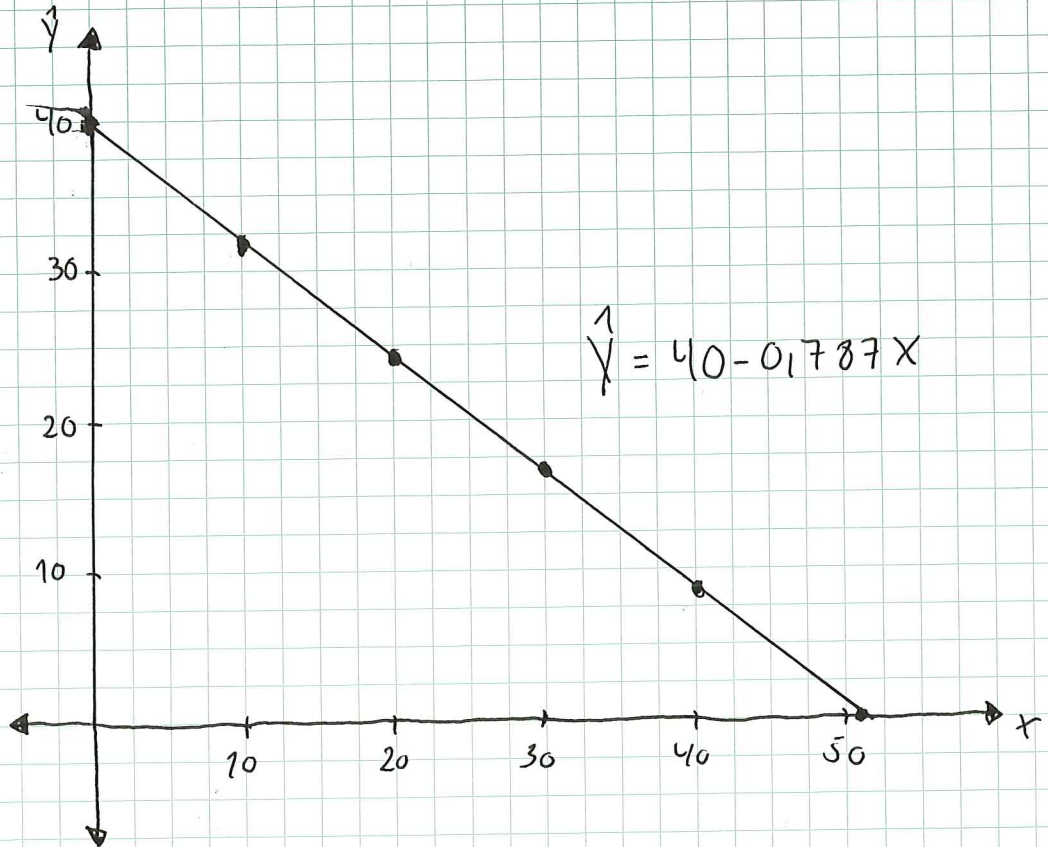
a = skjæring med y-aksen

b = Når x øker en enhet synker y med 0,787

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4 - fortsettelse:



e) Hva er modellens forklaringskraft?

Hva er tolkningen av denne?

Modellens forklaringskraft er gitt ved R^2 , R har vi ifra a

$$\underline{R^2} = (-0.701)^2 = \underline{0.491}$$

Modellens forklaringskraft er $0.491 = 49.1\%$

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4 - fortsettelse:

e) Modellens forklaringskraft angir hvor mye av variasjonen i Y som skyldes variasjon i X .

Her er $R^2 = 49,1\%$, som vil si at variasjonen i variabelen X (antall timer i lønnet arbeid) forklarer $49,1\%$ av variasjonen i Y (antall studie timer).

f) Hva er de tre årsakene til at korrelasjon ikke alltid innebærer kausalitet? Hva tror du er potensielle årsaker til at effekten av X på Y ikke nødvendigvis er kausal?

En av årsakene har med veien påvirkning gjør. Gjør påvirkning fra X til Y ($X \rightarrow Y$) eller fra Y til X ($Y \rightarrow X$), eller gjør den begge veier.

Noen ganger får man korrelasjon pga. tilfeldigheter. Variable X og Y kan ha ingenting med hverandre å gjøre, men det kan finnes en underliggende variabel Z som gir korrelasjonen.

Et eksempel er med salg av is (X) og haarsreg (Y)

De kan typisk virke sterkt korrelert, men det betyr ikke at det er kausalitet som regel går issalg opp når det er varmt, og flere folk badar når det er varmt. Så svaren for haarsreg går opp. Der er temperaturen en underliggende variabel Z som påvirker korrelasjonen.

Denne kolonne er forbeholdt sensor

This column is for external examiner

Oppgave 4 - fortsettelse:

I denne oppgaven har vi sett på sammenhengen mellom antall timer i lønnet arbeid og antall timer studiebld. Til tross for korrelasjon er ikke nødvendigvis det ikke nødvendigvis kausalitet mellom disse variablene.

Det kan f.eks. være en annen variabel Z som påvirker det kan f.eks. være timer med frivillig, ulønnet, arbeid for uke, eller deltakelse i et idrettslag. For alt vi vet kan det være dette som er utslagsgivende for hvor mye tid en person bruker på studier hver uke, og hvor mye personen jobber.

I tillegg vet vi ikke om studentene studerer det samme, gir samme arbeidsinn, og har samme fag. Noen kan hende har "lette" fag som de ikke behøver å studere mye (og da har mer tid til arbeid), mens andre kan hende sløter og har arbeidskrevende fag, og mer derfor studerer mye.

(Vet ikke om det er $X \rightarrow Y$ eller $Y \rightarrow X$
 Avgjør studietimer hver mye man jobber, eller omvendt?)