

Eksamen i

MAT110 Statistikk 1

Eksamensdag	: Mandag 10. sept. 2018
Tid	: 13:00 – 17:00 (4 timer)
Faglærer/telefonnummer	: Molde + Kristiansund: Per Kristian Rekdal / 924 97 051
Hjelpemidler	: KD + formelsamling fra 2018 + generell ordbok (morsmål/norsk/engelsk i papirformat)
Antall sider inkl. forsiden	: 20
Målform	: Norsk (bokmål)

Noen generelle råd:

- **Skriv rett inn. Ikke bruk så mye tid på kladding.**
- **Det er totalt 4 oppgaver.**

Oppgave 1: (logistikk)

Glamox er en norsk produsent og leverandør av belysning.

Glamox har til enhver tid mange tilbud ute hos kunder som de venter tilbakemelding på. Hensikten med denne oppgaven er å finne relevante sannsynligheter for at tilbud aksepteres basert på to kriterier: **selger** som har laget tilbudet og **verdien** på tilbudet.

1) Selgere:

La oss kun se på 2 selgere hos Glamox, nemlig selgerne 1 og 2.

2) Verdi:

La oss for enkelhets skyld kun skille mellom store og små tilbud, dvs. tilbud som har høy (h) verdi eller lav ($\bar{h}$) verdi i kroner og øre. Det kan f.eks. være tilbud som er over eller under 10 000 NOK.

Et tilbud blir enten akseptert (a) eller ikke akseptert ($\bar{a}$). Utfallsrommet har da 8 utfall, se lign.(1) på neste side.



Figur 1: Glamox.

Glamox bruker 350 historiske data for å estimere relevante sannsynligheter.

Resultatene av de historiske dataene er at utfall $(1, h, a)$ inntraff 5 ganger, utfall $(1, \bar{h}, a)$ inntraff 40 ganger osv.

Sannsynlighetene $p_1, p_2, \dots, p_8$ for de 8 mulige utfallene i utfallsrommet er dermed:

$$\Omega = \left\{ \overbrace{(\textcolor{red}{1}, h, a)}^{p_1 = \frac{5}{350}}, \overbrace{(\textcolor{red}{1}, \bar{h}, a)}^{p_2 = \frac{40}{350}}, \overbrace{(\textcolor{red}{1}, h, \bar{a})}^{p_3 = \frac{25}{350}}, \overbrace{(\textcolor{red}{1}, \bar{h}, \bar{a})}^{p_4 = \frac{80}{350}} \right. \\ \left. \overbrace{(\textcolor{blue}{2}, h, a)}^{p_5 = \frac{8}{350}}, \overbrace{(\textcolor{blue}{2}, \bar{h}, a)}^{p_6 = \frac{68}{350}}, \overbrace{(\textcolor{blue}{2}, h, \bar{a})}^{p_7 = \frac{22}{350}}, \overbrace{(\textcolor{blue}{2}, \bar{h}, \bar{a})}^{p_8 = ?} \right\} \quad (1)$$

hvor utfallet $(1, h, a)$ betyr at selger $\textcolor{red}{1}$ lager et tilbud med høy (h) verdi som blir akseptert (a). Tilsvarende for de øvrige utfallene.

Videre er $N_1 = 150$ tilbud gitt av selger $\textcolor{red}{1}$ og $N_2 = 200$ gitt av selger $\textcolor{blue}{2}$.

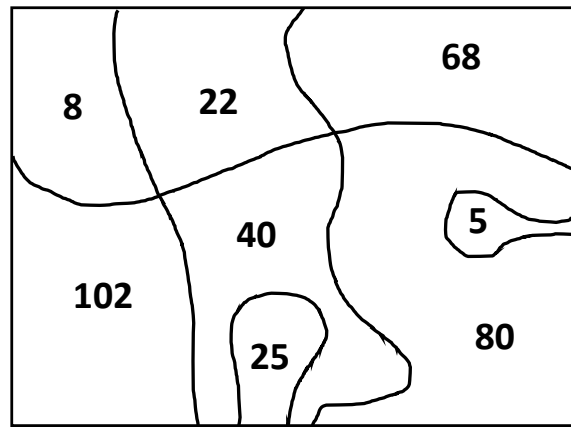
Totalt er da: $N = N_1 + N_2 = 350$ tilbud.

- a)** Sannsynlighetfordelingen $p_1, p_2, \dots, p_8$ i lign.(1) er en gyldig sannsynlighetsfordeling. Vis da at sannsynligheten p_8 er gitt ved:

$$\textcolor{red}{p_8} = \frac{\textcolor{red}{102}}{\textcolor{red}{350}} \quad (2)$$

- b) Figur 2 viser Venn-diagrammet for de 8 utfallene i utfallsrommet Ω . Tallene i figuren viser antall ganger de respektive utfallene inntreffer, se lign.(1).

Er utfallene i utfallsrommet disjunkte? Gi en *kort* begrunnelse. ¹



Figur 2: Utfallsrom Ω .

¹En setning er nok.

La oss definere følgende begivenheter:

$$S_1 = \left\{ (\textcolor{red}{1}, h, a), (\textcolor{red}{1}, \bar{h}, a), (\textcolor{red}{1}, h, \bar{a}), (\textcolor{red}{1}, \bar{h}, \bar{a}) \right\} \quad (3)$$

$$S_2 = \left\{ (\textcolor{blue}{2}, h, a), (\textcolor{blue}{2}, \bar{h}, a), (\textcolor{blue}{2}, h, \bar{a}), (\textcolor{blue}{2}, \bar{h}, \bar{a}) \right\} \quad (4)$$

Her er S_1 begivenheten at et tilbud er laget av selger $\textcolor{red}{1}$.

Her er S_2 begivenheten at et tilbud er laget av selger $\textcolor{blue}{2}$.

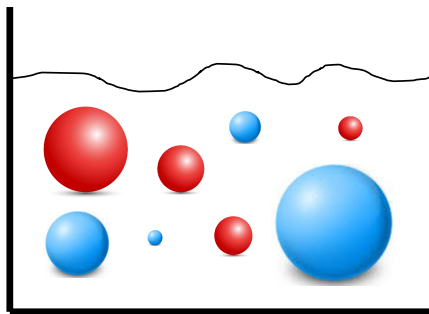
- $\textcolor{blue}{c}$) La oss si at hvert av de 8 utfallene i utfallsrommet Ω er assosiert med ei kule. Altså totalt 8 kuler, se figur 3.

La oss videre anta at vi har to typer kuler, $\textcolor{red}{r\ddot{o}de}$ kuler og $\textcolor{blue}{bl\ddot{a}}$ kuler som tilsvarer begivenhetene S_1 og S_2 henholdsvis.

De $\textcolor{red}{r\ddot{o}de}$ kulene er assosiert med utfallene til selger $\textcolor{red}{1}$,
og de $\textcolor{blue}{bl\ddot{a}}$ kulene er assosiert med utfallene til selger $\textcolor{blue}{2}$.

La oss anta at tilbudene laget av selger 1 er uavhengige av tilbudene laget av selger 2.

På tross av at tilbudene som de to selgerne har laget er uavhengige av hverandre, hvorfor kan vi ikke bruke urnemodellen til å f.eks. finne sannsynligheten $P(S_1)$ for at et tilbud er gitt av selger $\textcolor{red}{1}$? ²



Figur 3: Urne.

²Svar kort på denne oppgaven. Én setning er nok. Ingen utregninger behøves. Hva må være oppfylt, i tillegg til uavhengighet, for at urnemodellen skal gjelde? Er det oppfylt i vårt tilfelle?

Begivenhetene S_1 og S_2 definert i lign.(3) and (4).

Begivenhetene er delmengder av det totale utfallsrommet Ω i lign.(1)

Som nevnt på forrige side, S_1 er begivenheten at et tilbud er laget av selger 1.

Tilsvarende for selger 2.

- d) På tilsvarende måte som for begivenhetene S_1 og S_2 i lign.(3) and (4), skriv ned definisjonen av begivenheten for aksepterte tilbud A og begivenheten for høye tilbud H .³

Man kan altså ikke bruke urnemodellen, men man kan likevel finne sannsynligheten $P(A)$ på følgende måte, via lign.(1):

$$\begin{aligned}\underline{P(A)} &= p_1 + p_2 + p_5 + p_6 \\ &= \frac{5}{350} + \frac{40}{350} + \frac{8}{350} + \frac{68}{350} = \frac{121}{350} = \underline{0.3457}\end{aligned}\tag{5}$$

- e) Finn sannsynligheten $P(H)$ ved å bruke tilsvarende metode som i lign.(5).

³Her skal du altså bare skrive ned svaret. Ingen utregninger behøves.

La oss nå se på et tilbud med høy (h) verdi laget av selger 1.
 Den betingede sannsynligheten for at et slikt tilbud blir akseptert er gitt ved den generelle multiplikasjonssetningen:

$$P(A | \overbrace{S_1 \cap H}^{\text{vet}}) = \frac{P(S_1 \cap H \cap A)}{P(S_1 \cap H)} \quad (6)$$

f) Bruk lign.(6) og vis at: ⁴

$$P(A | S_1 \cap H) = \frac{5}{30} \quad (= 0.1667) \quad (7)$$

g) Gi en tolkning av $P(A | S_1 \cap H)$,
 dvs. forklar *kort* hva det betyr på ”godt norsk”.

■

⁴Tips: Finn først $P(S_1 \cap H \cap A)$ og $P(S_1 \cap H)$ ved å bruke tilsvarende metode som i lign.(5).

Oppgave 2: (logistikk - prognostisering)

La oss fortsette med Glamox-beskrivelsen fra oppgave 1.

For å unngå mye repeterende regning, la oss kun se på 4 nye tilbud.

For enkelhets skyld, la oss anta at alle disse 4 nye tilbudene er gitt av selger 2 og at alle har høy (h) verdi.

Anta at verdien til tilbud nr. i er v_i (i NOK), se tabell 1:

Tilbud nr. i ($i = 1,2,3,4$)	1	2	3	4
Høyt (h) tilbud	h	h	h	h
Selger S_2	2	2	2	2
Verdi v_i ($i=1,2,3,4$) (NOK)	40 000	33 000	105 000	75 000

Tabell 1: Nye tilbud.



Figur 4: Glamox.

La oss introdusere en stokastisk variabel som sier noe om et tilbud aksepteres eller ikke:

$$Y_i = \begin{cases} 1 & , \text{ dersom kunde aksepterer tilbud nr. } i \\ 0 & , \text{ dersom kunde ikke aksepterer tilbud nr. } i \end{cases} \quad (8)$$

hvor $i = 1, 2, 3, 4$ indikerer tilbud nr. i .

Aksept-sannsynligheten for et gitt tilbud er laget av selger 1 og har høy (h) verdi fant vi i oppgave [1f](#). For enkelhets skyld, la oss kalle denne sannsynligheten for p , dvs. $p = P(A|S_1 \cap H)$, altså:

$$p = \frac{5}{30} = 0.1667 \quad (9)$$

Den stokastiske variabelen Y definert ved:

$$Y = Y_1 + Y_2 + Y_3 + Y_4 \quad (10)$$

beskriver da antall tilbud som aksepteres.

Anta at tilbudene er uavhengige av hverandre, dvs. at Y_i 'ene er uavhengige.

- a) Forklar hvorfor Y er **binomisk** fordelt,
dvs. forklar hvorfor $Y \sim \text{Bin}\left[\underbrace{4}_{n=4}, \underbrace{0.1667}_{p=0.1667}\right]$.⁵
- b) Hva er sannsynligheten for at to tilbud blir akseptert,
dvs. hva er $P(Y = 2)$?
- c) Finn tallverdien for $E[Y]$
og gi en *kort* tolkning på hva $E[Y]$ betyr på godt norsk”.⁶

⁵Hvilke 4 kriterier må være oppfylt for et en forsøksserie skal være binomisk? Er disse oppfylt i vårt tilfelle?

⁶En setning er nok.

Summen av verdien til de aksepterte tilbudene er gitt ved:

$$V = v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4 \quad (11)$$

hvor v_i er verdien på tilbud nr. i .

d) Vis at forventningsverdien til verdien til de aksepterte tilbudene er: ⁷

$$E[V] = (v_1 + v_2 + v_3 + v_4) p \quad (13)$$

e) Bruk lign.(13) og regn ut tallverdien til $E[V]$. ⁸

f) Vis da at variansen til verdien til de aksepterte tilbudene er: ⁹

$$Var[V] = (v_1^2 + v_2^2 + v_3^2 + v_4^2) p(1 - p) \quad (15)$$

⁷Tips: Siden $Y_i \sim \text{Bin}[\overbrace{1}^{n_i=1}, p]$ så er:

$$E[Y_1] = E[Y_2] = E[Y_3] = E[Y_4] = p \quad (12)$$

⁸Bruk tallene for v_i i tabell 1 på side 8.

⁹Siden $Y_i \sim \text{Bin}[\overbrace{1}^{n_i=1}, p]$ så er:

$$Var[Y_1] = Var[Y_2] = Var[Y_3] = Var[Y_4] = p(1 - p) \quad (14)$$

g) Dersom selger 1 hadde lagd 200 tilbud istedet for bare 4 så er

$$V = V_1 + V_2 + \dots + V_{200} \quad (16)$$

$$= v_1 Y_1 + v_2 Y_2 + \dots + v_{200} Y_{200} \quad (17)$$

verdien på de aksepterte tilbudene, hvor $V_i = v_i Y_i$ og $i = 1, 2, 3, \dots, 200$.

La oss si at alle tilbudene har høy verdi (h) slik at aksept-sannsynligheten p til tilbudene er den samme, men at verdien v_i på de $i = 1, 2, 3, \dots, 200$ tilbudene er forskjellig.

Dersom vi i tillegg fortsatt antar at alle tilbudene er uavhengige, hvorfor kan vi ikke da anvende sentralgrenseteoremet og si at V er tilnærmet normalfordelt?

■

Oppgave 3: (logistikk og økonomi)

La oss anta at det er $n = 4$ Kiwi-butikker i Molde.

Alle disse butikkene selger bananer når det er sesong.

La oss introdusere en stokastisk variabel D_i (“demand”), hvor:

$$D_i = \text{etterspørsel av antall kg bananer per dag i butikk nr. } i \quad (18)$$

med $i = 1, 2, 3, 4$.

Den totale etterspørselen per dag for de 4 butikkene er da beskrevet av:

$$D_{\text{tot}} = D_1 + D_2 + D_3 + D_4 \quad (19)$$

Anta videre at de stokastiske variablene D_i har samme sannsynlighetsfordeling for alle dagene i sesongen, dvs. fordelingen til etterspørselen er den samme hver dag for en gitt butikk.



Figur 5: Kiwi-butikker og bananer.

Innkjøperne hos Kiwi har funnet ut at forventet etterspørsel og tilhørende varians per dag for de 4 aktuelle butikkene ($i = 1, 2, 3, 4$) er som oppgitt i tabell 2:

i	1	2	3	4
$E[D_i]$	40	100	60	20
$\text{Var}[D_i]$	64	144	352	16

Tabell 2: Forventning $E[D_i]$ og varians $\text{Var}[D_i]$ for $i = 1, 2, 3, 4$.

- a) Vis at den totale forventede etterspørselen $E[D_{\text{tot}}]$ per dag for de 4 butikkene er:

$$E[D_{\text{tot}}] = 220 \quad (20)$$

kg med bananer.

Anta at etterspørselen av bananer i de 4 butikkene er uavhengige.

- b) Regn ut den totale variansen $\text{Var}[D_{\text{tot}}]$ per dag for de 4 butikkene.

Anta videre at alle de fire stokastiske variablene D_i er normalfordelt, dvs. anta:

$$D_i \sim N[E[D_i], Var[D_i]] \quad (21)$$

hvor $E[D_i]$ og $Var[D_i]$ er som oppgitt i tabell 2 for $i = 1, 2, 3, 4$.

c) Hva slags fordeling har da D_{tot} ?¹⁰

¹⁰Se side 108 i formelsamlingen fra 2018. Definisjonen at D_{tot} er gitt i lign.(19).

Bananene kastes

La oss anta at bananene kun har holdbarhet én dag.

Det betyr at dersom Kiwi ikke får solgt bananene en gitt dag så må de kaste dem når de stenger butikken om kvelden.

La oss nå kun se på butikk nr. 3.

Fortjenesten F_3 per dag for denne butikken er:

$$F_3 = \underbrace{p S_3}_{\text{inntekt}} - \underbrace{c q_3}_{\text{kostnad}}, \quad (22)$$

hvor

$$\begin{aligned} S_3 &= \min(D_3, q_3) \\ &= \text{antall } kg \text{ med bananer som } \underline{\text{selges}} \text{ i butikk nr. 3 per dag} \end{aligned} \quad (23)$$

$$q_3 = \text{antall } kg \text{ bananer som innkjøperne hos Kiwi } \underline{\text{bestiller}} \text{ per dag for butikk nr. 3 } \quad (\text{order quantity}) \quad (24)$$

$$c = 8 \text{ NOK} \quad (\underline{\text{kostnaden}} \text{ per } kg \text{ bananer for Kiwi}) \quad (25)$$

$$p = 12.9 \text{ NOK} \quad (\underline{\text{pris}} \text{ per } kg \text{ som Kiwi selger bananene for}) \quad (26)$$

Det optimale bestillingskvantumet q_3^* som gir maksimal forventet fortjeneste av lign.(22) er da bestemt av:

$$P(D_3 \leq q_3^*) = 1 - \frac{c}{p} \quad (27)$$

hvor

$$q_3^* = \text{antall } kg \text{ med bananer som Kiwi må bestille per dag for å } \underline{\text{maksimere}} \text{ forventet fortjeneste til butikk nr. 3} \quad (28)$$

d) Finn q_3^* .¹¹

¹¹**Tips 1:** Bruk de oppgitte verdiene for c og p samt lign.(27).

Tips 2: I denne oppgaven slipper du å gjøre tabelloppslag. Bruk verdien $z_0 = -0.30$, altså negativ verdi.

Tips 3: Husk at $E[D_3]$ og $\sigma[D_3] = \sqrt{\text{Var}[D_3]}$ finnes i tabell 2 på side 14.

Redusert pris

I stedet for å kaste bananene kan Kiwi gjøre en avtale med et firma som kjøper bananene som Kiwi ikke får solgt en gitt dag. Dette firmaet får tilbud om å kjøpe bananene til redusert pris p_{red} .

Antall kg med bananer som ikke blir solgt en gitt dag for butikk nr. 3, dvs. usolgte bananer, er da beskrevet av følgende stokastiske variabel:

$$U_3 = \begin{cases} q_3 - D_3 & , \quad q_3 \geq D_3 \\ 0 & , \quad q_3 < D_3 \end{cases} \quad (29)$$

Fortjenesten F_3 per dag for butikk 3 blir da en utvidelse av lign.(22), nemlig:

$$F_3 = \underbrace{p S_3 + p_{\text{red}} U_3}_{\text{inntekt}} - \underbrace{c q_3}_{\text{kostnad}} \quad (30)$$

Det nye, optimale bestillingskvantumet q_3^* som gir maksimal forventet fortjeneste av lign.(30) er da bestemt av: ¹²

$$P(D_3 \leq q_3^*) = \frac{p - c}{p - p_{\text{red}}} \quad (31)$$

- e) Anta at Kiwi kjøper inn $q_3^* = 70 \text{ kg}$ bananer per dag til butikk nr. 3, hvilken redusert pris p_{red} må da Kiwi og dette firmaet bli enige om dersom Kiwi skal maksimere den forventede fortjenesten F_3 ? ¹³

■

¹²Du skal ikke vise lign.(31). Bare ta den for gitt.

¹³Dette er en litt vanskelig oppgave. Hopp over den dersom du har dårlig tid. Tips: Bruk lign.(31) og løs med hensyn på p_{red} .

Oppgave 4: (økonomi)

I fjor opplevde teleselskapet Telenor at $\overbrace{\text{aksjekursen}}^{\text{pris på aksje}}$ økte i en periode.
La oss se på 5 dager med omsetning i denne aktuelle perioden.

$$x = \text{dag nr. } x, \text{ hvor } x = 1, 2, 3, 4, 5 \quad (32)$$

$$y = \text{aksjekurs (i NOK)} \quad (33)$$

Aksjekurs, dvs. aksjepris, for dag nr. x ($x = 1, 2, 3, 4, 5$) er gitt ved følgende tabell:

x (dag nummer)	1	2	3	4	5
y (aksjekurs i NOK)	268	279	285	289	299

Figur 6: Omsetningsdag nr. x og aksjekurs (pris) y .

Ut fra tabellen i figur (6) kan man regne ut den empiriske variansen for aksjekursen y og den empiriske variansen for x . Resultatet er:

$$S_y^2 = 133 \text{ NOK}^2, \quad S_x^2 = 2.5 \quad (34)$$

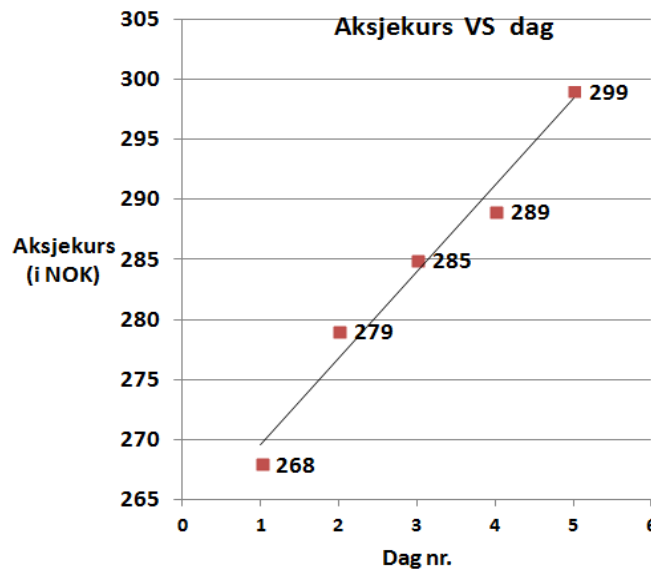
Den empiriske **kovariansen** mellom x og y er:

$$S_{xy} = 18 \text{ NOK} \quad (35)$$

(Størrelsene i lign.(34) og (35) trenger du å ikke regne ut. Bare ta dem for gitt).



Figur 7: Telenor.



Figur 8: Plott av dataene fra tabellen i figur (6).

- a) Grafen i figur (8) viser minste kvadraters regresjonslinje for x og y . Finn et analytisk uttrykk for denne **regresjonslinjen**.¹⁴
- b) i) Forklar *kort*, på generelt grunnlag, hva en regresjonslinje mellom variablene x og y beskriver.¹⁵
- ii) På generelt grunnlag, hvordan tolker du parameteren $\hat{\beta}$ i en *lineær* regresjonslinje?¹⁶
- iii) Spesielt, hvordan tolker du parameteren $\hat{\beta}$ i regresjonslinjen fra oppgave 4a?
- c) Anta trenden på aksjemarkedet for Telenor fortsetter på samme måte som de 5 dagene i en periode fremover. Dersom Telenor ønsker å *predikere* hvor mye aksjekursen blir for dag nummer 12 så kan de bruke modellen fra oppgave 4a, altså regresjonslinjen.

Hvor mye predikerer regresjonslinjen at aksjekursen vil være for dag nummer 12?

¹⁴Dvs. finn formelen for linjen. Se gjerne formelsamlingen for formel for minste kvadraters regresjonslinje.

¹⁵Tips: estimat / prediksjon.

¹⁶Tips: stigningstall.

d) Forklaringsstyrken R^2 blir ofte brukt i forbindelse med lineær regresjonsanalyse.

- i) Hva slags mulige verdier har R^2 ?
- ii) Hva slags benevning/enhet har R^2 ?
- iii) Hva er R^2 et mål på?

For å finne forklaringsstyrken R^2 kan man bruke et dataprogram, f.eks. Excel, i stedet for å regne ut R^2 “for hånd” via definisjonen.

e) Finn forklaringsstyrken uten å gjøre noe regning “for hånd”. Bare les av fra Excel-utskriften i figur (9) nedenfor. (Bruk 4 desimalers nøyaktighet).

f) Kommenter svaret i oppgave 4e.¹⁷

A	B	C	D	E	F	G	H
SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0,987135295						
R Square	0,97443609						
Adjusted R Square	-1,666666667						
Standard Error	2,12916259						
Observations	1						
ANOVA							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression	5	518,4	103,68	114,3529	#NUM!		
Residual	3	13,6	4,533333				
Total	8	532					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>
X Variable 4	262,4	2,233084563	117,5056	1,36E-06	255,2933283	269,5066717	255,2933283
X Variable 5	7,2	0,673300329	10,69359	0,001748	5,057257855	9,342742145	5,057257855

Figur 9: Utskrift fra Excel.

¹⁷For en gitt dag, vi du si at regresjonslinjen predikerer aksjekursen i stor eller liten grad? Med stort eller lite presisjonsnivå?