Høgskolen i Molde

Eksamen i

MAT110 Statistikk 1

Eksamensdag	**:**	**Tirsdag 22. mai 2018**
Tid	**:**	**09:00 – 13:00 (4 timer)**
Faglærer/telefonnummer	**:**	**Molde + Kristiansund:** **Per Kristian Rekdal / 924 97 051**
Hjelpemidler	**:**	**KD + formelsamling** **+ generell ordbok** **(morsmål/norsk/engelsk i papirformat)**
Antall sider inkl. forsiden	**:**	**21**
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.

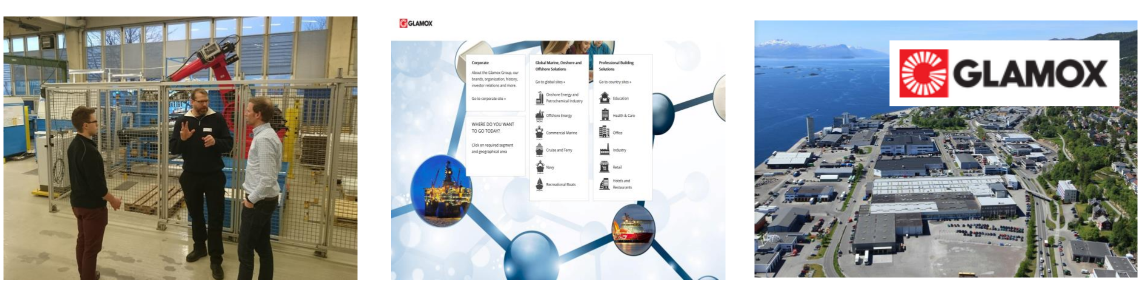
1) **Selgere**:

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 ($\overline{h}$) 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 ($\overline{a}$).
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 $(1, \overline{h}, a)$ inntraff 40 ganger osv.

Sannsynlighetene p_1, p_2, ..., p_8 for de 8 mulige utfallene i utfallsrommet er dermed:

$$\begin{aligned} \Omega \quad = \quad \Big\{ \; & \overbrace{(1, h, a)}^{p_1 \, = \, \frac{5}{350}} , \; \overbrace{(1, \overline{h}, a)}^{p_2 \, = \, \frac{40}{350}} \; , \; \overbrace{(1, h, \overline{a})}^{p_3 \, = \, \frac{25}{350}} , \; \overbrace{(1, \overline{h}, \overline{a})}^{p_4 \, = \, \frac{80}{350}} \\ & \overbrace{(2, h, a)}^{p_5 \, = \, \frac{8}{350}} , \; \overbrace{(2, \overline{h}, a)}^{p_6 \, = \, \frac{68}{350}} \; , \; \overbrace{(2, h, \overline{a})}^{p_7 \, = \, \frac{22}{350}} , \; \overbrace{(2, \overline{h}, \overline{a})}^{p_8 \, = \, \frac{102}{350}} \; \Big\} \end{aligned} \tag{1}$$

hvor utfallet $(1, h, a)$ betyr at selger 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 1 og $N_2 = 200$ gitt av selger 2.
Totalt er da: $N = N_1 + N_2 = 350$.

a) Vis at sannsynlighetfordelingen p_1, p_2, ..., p_8 i lign.(1) er en gyldig fordeling.

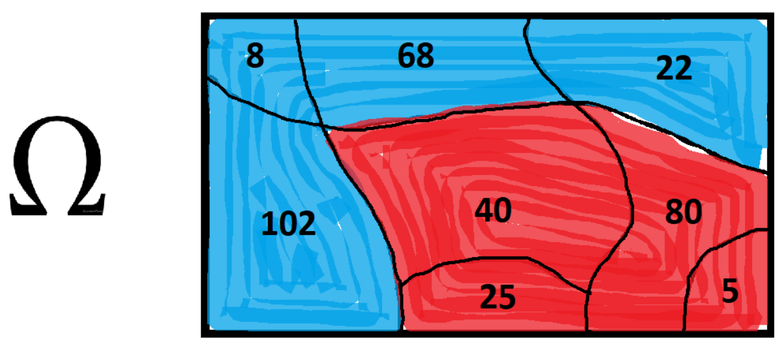
La oss definere følgende begivenhene:

$$S_1 = \Big\{ (1, h, a) \;,\; (1, \overline{h}, a) \;\;,\;\; (1, h, \overline{a}) \;,\; (1, \overline{h}, \overline{a}) \Big\} \tag{2}$$

$$S_2 = \Big\{ (2, h, a) \;,\; (2, \overline{h}, a) \;\;,\;\; (2, h, \overline{a}) \;,\; (2, \overline{h}, \overline{a}) \Big\} \tag{3}$$

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).
Den røde og den blå delen av utfallsrommet definerer begivenhetene S_1 og S_2, henholdsvis.

Er begivenhetene S_1 og S_2 disjunkte? Gi en *kort* begrunnelse. [1]



Figur 2: Utfallsrom Ω.

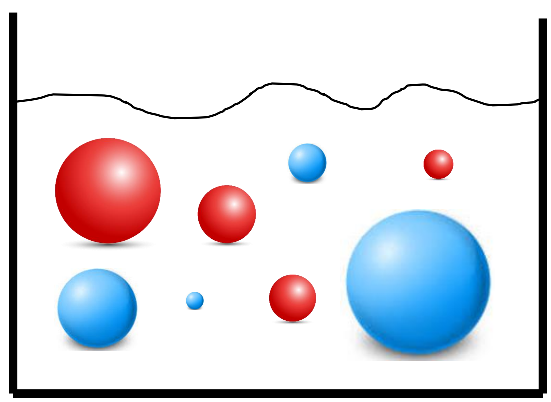
[1] Én setning er nok.

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, røde kuler og blå kuler.
De røde kulene er assosiert med utfallene til selger 1,
og de blå kulene er assosiert med utfallene til selger 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 $P(S_2)$ for at et tilbud er gitt av selger 2? [2]



Figur 3: Urne.

[2]Svar 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?

Man kan altså ikke bruke urnemodellen, men man kan likevel finne sannsynligheten $P(S_2)$ på følgende måte, via lign.(1):

$$\begin{aligned} \underline{P(S_2)} &= p_5 + p_6 + p_7 + p_8 \\ &= \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + \frac{102}{350} = \frac{200}{350} = \underline{0.5714} \end{aligned} \tag{4}$$

Sannsynligheter for andre begivenheter finner man på tilsvarende måte.

La oss også definere følgende begivenheter:

$$A = \Big\{ \ (1, h, a) \ , \ (1, \overline{h}, a) \quad , \quad (2, h, a) \ , \ (2, \overline{h}, a) \ \Big\}$$

og

$$H = \Big\{ \ (1, h, a) \ , \ (1, h, \overline{a}) \quad , \quad (2, h, a) \ , \ (2, h, \overline{a}) \ \Big\} \tag{5}$$

La oss nå se på et tilbud med høy (h) verdi laget av selger 2.
Den betingede sannsynligheten for at et slikt tilbud blir akseptert er gitt ved den generelle multiplikasjonssetningen:

$$\boxed{P\big(A \,|\, \overbrace{S_2 \cap H}^{\text{vet}} \big) = \frac{P(S_2 \cap H \cap A)}{P(S_2 \cap H)} \qquad (6)}$$

d) Bruk lign.(6) og vis at $P\big(A \,|\, S_2 \cap H\big)$ er: [3]

$$\boxed{P(A|S_2 \cap H) \;=\; \frac{8}{30} \qquad (= 0.2667) \qquad (7)}$$

e) Gi en tolkning av $P\big(\,A \,|\, S_2 \cap H\,\big)$,
dvs. forklar *kort* hva det betyr på "godt norsk".

■

[3]Tips: Finn først $P(S_2 \cap H \cap A)$ og $P(S_2 \cap H)$ ved å bruke tilsvarende metode som i lign.(4).

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:

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

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

Aksept-sannsynligheten for et gitt tilbud laget av selger 2 med høy (h) verdi fant vi i oppgave **1d**. For enkelhets skyld, la oss kalle denne sannsynligheten for p, dvs. $p = P(A|S_2 \cap H)$, altså:

$$p \;=\; \frac{8}{30} \;=\; 0.2667 \tag{9}$$

Den stokastiske variabelen Y definert ved:

$$\boxed{Y \;=\; Y_1 + Y_2 + Y_3 + Y_4 \qquad (10)}$$

beskriver da antall tilbud som aksepteres.

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

a) Forklar hvorfor Y er binomisk fordelt,
dvs. forklar hvorfor $Y \sim \text{Bin}\left[\,n=4, p=0.2667\,\right]$. [4]

b) Hva er sannsynligheten for at to eller flere tilbud blir akseptert,
dvs. hva er $P(Y \geq 2)$? [5]

c) Finn tallverdien for $E[Y]$
og gi en *kort* tolkning på hva $E[Y]$ betyr på godt norsk". [6]

---

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

[5]Bruk gjerne at: $P(Y \geq 2) = 1 - P(Y=0) - P(Y=1)$.

[6]Én setning er nok.

Summen av verdien til de aksepterte tilbudene er gitt ved:

$$\boxed{V \;=\; v_1 Y_1 + v_2 Y_2 + v_3 Y_3 + v_4 Y_4 \qquad (11)}$$

hvor v_i er verdien på tilbud nr. i.

d) Vis at forventningsverdien til verdien til de akspeterte tilbudene er: [7]

$$\boxed{E[V] \;=\; \big(v_1 + v_2 + v_3 + v_4 \big)\, p \qquad (13)}$$

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

f) Vis da at variansen til verdien til de akspeterte tilbudene er: [9]

$$\boxed{Var[V] \;=\; \big(v_1^2 + v_2^2 + v_3^2 + v_4^2 \big)\, p(1-p) \qquad (15)}$$

---

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

$$E[Y_1] \;=\; E[Y_2] \;=\; E[Y_3] \;=\; E[Y_4] \;=\; p \tag{12}$$

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

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

$$Var[Y_1] \;=\; Var[Y_2] \;=\; Var[Y_3] \;=\; Var[Y_4] \;=\; p(1-p) \tag{14}$$

g) Dersom selger 2 hadde lagd 200 tilbud istedet for bare 4 så er

$$V = v_1 Y_1 + v_2 Y_2 + ... + v_{200} Y_{200} \tag{16}$$

verdien på de aksepterte tilbudene.

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 $i = 1, 2, 3, ..., 200$ 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 $n = 4$ Bunnpris-butikker i Molde.

Alle disse butikkene selger jordbær når det er sesong.
Anta at etterspørselen er den samme hver dag, men at etterspørselen i de 4 butikkene er forskjellige.

Anta videre at etterspørselen i all hovedsak skjer på dagtid, ikke kveldstid.
Dermed er det hensiktsmessig å introdusere en stokastisk variabel D_i ("*demand*"), hvor:

$$\boxed{D_i \;=\; \text{etterspørsel av antall kurver med friske jordbær per dag på } \underline{\text{dagtid}} \text{ for butikk nr. } i \qquad (17)}$$

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

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

$$\boxed{D_{\text{tot}} \;=\; D_1 + D_2 + D_3 + D_4 \qquad (18)}$$

Figur 5: Bunnpris-butikker og jordbær.

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

i	1	2	3	4
E[D$_i$]	20	50	30	10
Var[D$_i$]	16	36	88	4

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

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

$$\boxed{E[D_{\text{tot}}] = 110 \qquad (19)}$$

kurver med jordbær.

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

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

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

$$\boxed{D_i \;\sim\; N\big[\, E[D_i]\,,\, Var[D_i]\,\big] \qquad (20)}$$

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]

[10]Se side 108 i formelsamlingen fra 2018. Definisjonen at D_{tot} er gitt i lign.(18).

Jordbærene kastes

La oss anta at jordbærene kun har holdbarhet én dag.
Det betyr at dersom Bunnpris ikke får solgt jordbærene på dagtid så må de kaste dem på kvelden når de stenger butikken. Fortjenesten F per dag blir da:

$$\boxed{F \; = \; \underbrace{p\,S_{\text{dag}}}_{\text{inntekt}} \; - \; \underbrace{c\,q}_{\text{kostnad}} \; , \qquad (21)}$$

hvor

$$\begin{aligned} S_{\text{dag}} &= \min(D_{\text{tot}}, q) \\ &= \text{antall kurver med jordbær som \underline{selges} i de 4 butikkene per dag på \underline{dagtid}} \qquad (22) \end{aligned}$$

$$\begin{aligned} q &= \text{antall kurver med jordbær som innkjøperne hos Bunnpris \underline{bestiller}} \qquad (23) \\ &\quad \text{per dag for de 4 butikkene} \quad (\textit{ order quantity }) \end{aligned}$$

$$\begin{aligned} c &= 30 \;\; \text{NOK} \quad (\underline{\text{kostnaden}} \text{ for en kurv med jordbær for Bunnpris}) \qquad (24) \\ p &= 40 \;\; \text{NOK} \quad (\underline{\text{pris}} \text{ som Bunnpris selger en kurv med jordbær for}) \qquad (25) \end{aligned}$$

Det optimale bestillingskvantumet q^* som gir <u>maksimal</u> forventet <u>fortjeneste</u> av lign.(21) er da bestemt av:

$$\boxed{P(D_{\text{tot}} \leq q^*) \; = \; 1 - \frac{c}{p} \qquad (26)}$$

hvor

$$\begin{aligned} q^* &= \text{antall kurver med jordbær som Bunnpris må bestille per dag} \qquad (27) \\ &\quad \text{for å \underline{maksimere} forventet fortjeneste} \end{aligned}$$

d) Finn q^*. [11]

---

[11] Tips 1: Bruk de oppgitte verdiene for c og p samt lign.(26).
Tips 2: Argumentet z_0 for sannsynligheten $P(Z \leq z_0) = 0.25$ finnes ikke direkte i tabellen i formelsamlingen.
Sannsynlighetene i formelsamlingen "stopper" ved 0.5000.
Men av symmetrigrunner så er $z_0 = -0.675$, altså negativ z_0.
Dermed <u>slipper</u> du å gjøre tabelloppslag i denne deloppgaven.

Redusert pris

Istedet for å kaste jordbærene så velger Bunnpris heller å selge jordbærene til redusert pris på kvelden, et par timer før stengetid.

Antall kurver med jordbær som selges på kvelden til redusert pris er da beskrevet av den stokastiske variabelen:

$$S_{\text{kveld}} = \begin{cases} q - D_{\text{tot}} & , \quad q \geq D_{\text{tot}} \\ 0 & , \quad q < D_{\text{tot}} \end{cases} \tag{28}$$

Fortjenesten F per dag blir da en utvidelse av lign.(21), nemlig:

$$\boxed{F = \underbrace{p\,S_{\text{dag}} + p_{\text{red}} S_{\text{kveld}}}_{\text{inntekt}} - \underbrace{c\,q}_{\text{kostnad}} \qquad (29)}$$

Det nye, optimale bestillingskvantumet q^* som gir maksimal forventet fortjeneste av lign.(29) er da bestemt av: [12]

$$\boxed{P(D_{\text{tot}} \leq q^*) = \frac{p - c}{p - p_{\text{red}}} \qquad (30)}$$

e) Anta at Bunnpris kjøper inn $q^* = 116$ kurver med jordbær per dag, hvilken redusert pris p_{red} må da Bunnpris sette på en kurv med jordbær for å maksimere den forventede fortjenesten? [13]

■

[12] Du skal ikke vise lign.(30). Bare ta den for gitt.

[13] Dette er en litt vanskelig oppgave. Hopper over den dersom du har dårlig tid.

Oppgave 4: (økonomi)

I kapittel 1 i kurset "*MAT110 Statistikk 1*" så vi på statistiske størrelser med både én og to variabler. For statistiske størrelser med to variabler så vi blant annet på graden av samvariasjon mellom variablene.

a) For observasjonene x_1, x_2, ..., x_n har vi tilhørende observasjoner y_1, y_2, ..., y_n. I denne sammenheng kan man snakke om korrelasjonskoeffisienten R_{xy} til disse obervasjonene.

i) Hva slags mulige verdier kan R_{xy} ha?

ii) Hva er R_{xy} et mål på?

iii) Hva slags enhet har R_{xy}?

Eiendomsmegleren *Notar* ønsker å se nærmere på sammenhengen mellom kvadratmeterpris y på leiligheter og årsinntekten til den som kjøper leiligheten.

$$x \quad = \quad \text{årsinntekt} \qquad (\text{ i antall 1000 NOK }) \tag{31}$$

$$y \quad = \quad \text{kvadratmeterpris} \qquad (\text{ NOK per } m^2 \text{ }) \tag{32}$$

Observasjonene x måles i antall 1000 NOK per år, brutto årsinntekt.
Observasjonene y er kvadratmeterpris målt i NOK per m^2 for den aktuelle leiligheten.

Foto: Colourbox

Figur 6: Leiligheter. Lønn.

Basert på de $n = 8$ salgene av leiligheter så langt i år fant *Notar* følgende resultat:

x (årsinntekt , 1000 NOK)	**320**	**450**	**480**	**560**	**610**	**750**	**810**	**880**
y (kvadratmeterpris , NOK / m²)	**23 500**	**36 000**	**33 000**	**48 000**	**45 000**	**52 000**	**57 000**	**53 000**

Tabell 3: Årsinntekt x og kvadratmeterpris y.

Ut fra tabell 3 kan man regne ut gjennomsnittene $\overline{x}$ og $\overline{y}$:

$$\overline{x} = 607.5 \qquad , \qquad \overline{y} = 43\,437.5 \tag{33}$$

Man kan også regne ut den empiriske variansen til x, dvs. S_x^2, og den empiriske variansen til y, dvs. S_y^2. Resultatene er:

$$S_x^2 \;=\; 37\,364 \qquad , \qquad S_y^2 \;=\; 133\,388\,393 \tag{34}$$

Den empiriske **ko**variansen mellom x og y er: [14]

$$S_{xy} \;=\; 2\,084\,821 \tag{35}$$

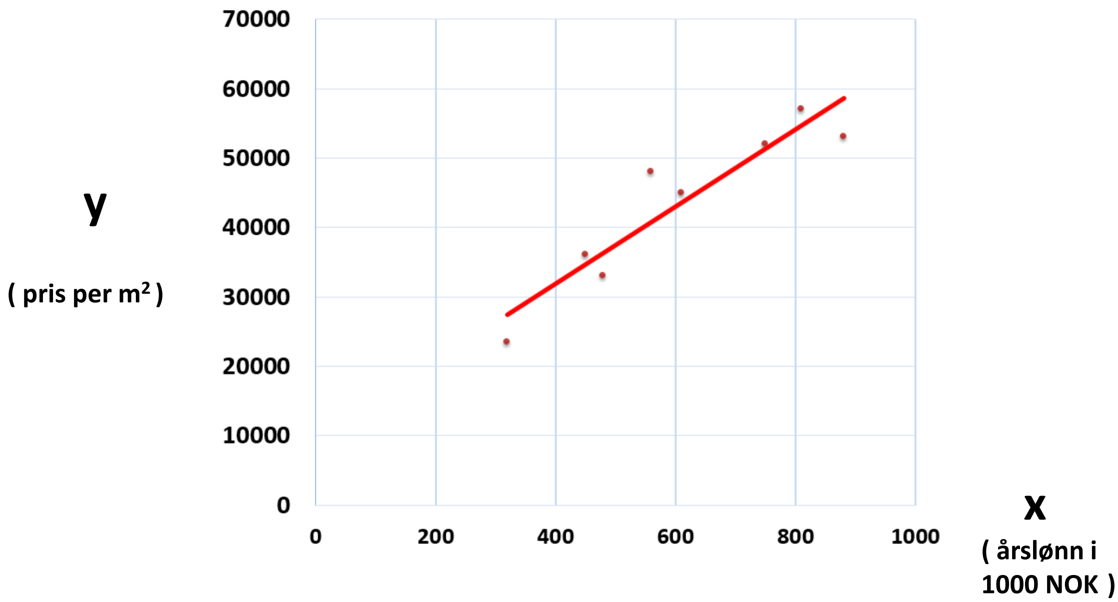
b) Regn ut korrelasjonskoeffisienten R_{xy} for observasjonene i tabell 3.

c) Tolk svaret du fikk for R_{xy} i oppgave **4b**. [15]

[14] Størrelsene i lign.(33), (34) og (35) trenger du å ikke regne ut. Bare ta dem for gitt. Benevningen, altså enheten, til disse størrelsene er utelatt.

[15] Hva sier den numeriske (tallmessige) verdien av R_{xy} fra oppgave **4b** om graden av korrelasjon mellom observasjonene x og y?

Pga. resultatene fra oppgave 4a og 4b konkluderer *Notar* med at det kan være hensiktsmessig å finne en eksplisitt lineær sammenhengen mellom x og y. Eiendomsmegleren bestemmer seg for å bruke lineær regresjonsanalyse.



Figur 7: Plott av dataene fra tabell 3.

d) Den rette linjen i figur 7 viser minste kvadraters regresjonslinje for observasjonene x og y. Finn et analytisk uttrykk for denne lineære **regresjonslinjen**. [16]

e) La oss anta at en person med en årinntekt på 1 millon NOK skal kjøpe seg en leilighet. Dersom vi antar at personen kjøper en leilighet med en kvadratmeterpris ihenhold til regresjonslinjen, hva blir da kvadratmeterpris på leiligheten? [17]

[16] Dvs. finn formelen for linjen via setningen for minste kvadraters lineære regresjonslinje. Se gjerne formelsamling.
[17] Bruk regresjonslinjen som du fant i oppgave 4d for å løse denne oppgaven.

f) Anta at en person har kjøpt ei leilighet som har en kvadratmeterpris på er 30 000 NOK/m^2.

Ifølge regresjonslinjen, hva slags årsinntekt har denne kjøperen?

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

g) Finn forklaringskraften R^2 uten å gjøre noe regning "for hånd".
Bare les av fra Excel-utskriften i figur 8 nedenfor.
(Bruk 4 desimalers nøyaktighet).

h) Kommenter svaret i oppgave 4g. [18]

1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0,9339							
5	R Square	0,8721							
6	Adjusted R Square	0,8508							
7	Standard Error	4461,4794							
8	Observations	8							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	814289960,1	814289960	40,90922938	0,000687936			
13	Residual	6	119428789,9	19904798,3					
14	Total	7	933718750						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
17	Intercept	9540,7188	5529,418628	1,725447	0,135202977	-3989,281179	23070,71876	-3989,281179	23070,71876
18	X Variable 1	55,7972	8,723716198	6,39603232	0,000687936	34,45100616	77,14333526	34,45100616	77,14333526

Figur 8: Utskrift fra Excel.

■

[18] For en person med en gitt årslønn som skal kjøpe leilighet, vi du si at regresjonslinjen predikerer kvadratmeterprisen til leiligheten i stor eller liten grad? Med stort eller lite presisjonsnivå?