

FORELESNING 24

MET 343/

Eivind Eriksen

26/04 2012

STATISTIKK

PLAN:

Gjennomgang: { Elsauen 06/2010 Oppg. 7
- 11 - 11/2010 Oppg. 2

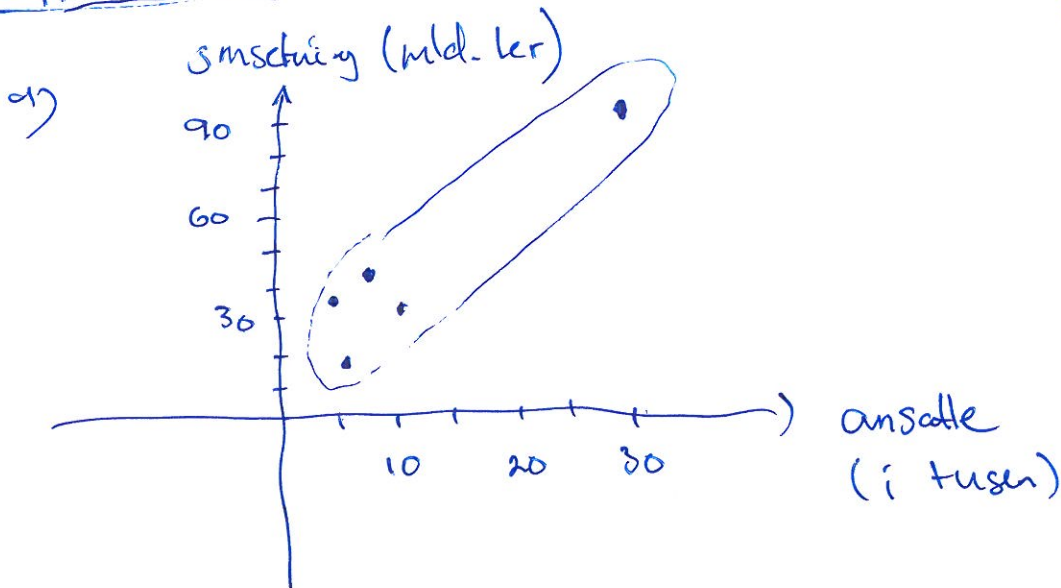
Husk: { Arbeidskrav 8 (det siste)
Frist: Mandag 30/04

Elsauen / Man 30/05

Veiledning 08/08

Elsauen 06/2010

Oppgave 7



ser ut til å være
positiv korrelasjon

$$b) r = 0.935 \Rightarrow r^2 = 0.935^2 \approx \underline{0.874}$$

Tolkning: 87.4% av variansen i ansettning skyldes / kan förklaras med Varians i arb. ansette.

c) Forutsetning: Cellene i krysstabelen ber ha forventet frekvens minst 5.

Forventet frekvens:

	M	K	
For	(2.4)	○	16
Imot	(2.6)	○	17
	5	28	33

$$\begin{array}{l} \text{For, M:} \\ \hline E = \frac{5 \cdot 16}{33} \\ \approx 2.4 \\ \\ \text{Imot, M:} \\ \hline E = \frac{5 \cdot 17}{33} \\ \approx 2.6 \end{array}$$

Minst 50% av cellene har forventet frekvens < 5 .
Dermed kan vi ikke bruke χ^2 -test.

- (c) I denne oppgaven skal du ikke beregne noe. Men tenk deg at du lager et 99 % konfidensintervall og et 90 % konfidensintervall basert på den samme stikkprøven. Hvilket intervall er lengst? Begrunn svaret kort.
- (d) Forklar kort med egne ord hva et konfidensintervall er for noe, og hvordan man skal tolke det.

Oppgave 6

Anta at månedslønn i by 1 er normalfordelt. Vi tar en stikkprøve med $n_1 = 17$ personer med gjennomsnittslønn $\bar{x}_1 = 23198$ kroner og standardavvik $s_1 = 14000$ kroner.

- (a) Test påstanden om at i by 1 er gjennomsnittslønnen lik 30000 kroner. Bruk signifikansnivå $\alpha = 0.05$. Skriv opp hypotesene på symbolsk form, skisser kritisk verdi og testobservator i en graf, og formuler konklusjonen i et lettfattelig språk.
- (b) Også i by 2 er månedslønnen normalfordelt. Der tar vi en stikkprøve med $n_2 = 14$ personer med gjennomsnittslønn $\bar{x}_2 = 26899$ kroner og standardavvik $s_2 = 13000$ kroner. Er stikkprøvene fra by 1 og 2 relaterte eller uavhengige? Test påstanden om at gjennomsnittslønnen i by 2 er høyere enn i by 1. Bruk signifikansnivå $\alpha = 0.05$. Skriv opp hypotesene på symbolsk form og formuler konklusjonen i et lettfattelig språk.

Oppgave 7

Ernst & Young og SR-Bank gjorde i 2009 en undersøkelse om norske oljeleverandører. De fant følgende tall for antall ansatte (i tusen) og omsetning (i milliarder kroner) for ulike regioner:

	Rogaland	Møre	Hordaland	Agder	Kongsberg
Ansatte	31	10	12	5	6
Omsetning	90	45	32	36	18

- (a) Lag et spredningsdiagram (scatterplott) for dataene med ansatte på x-aksen og omsetning på y-aksen. Ser det ut til å være en lineær korrelasjon mellom antall ansatte og omsetningen og er den evt. positiv eller negativ?
- (b) Korrelasjonskoeffisienten er $r = 0.935$. Beregn r^2 og gi en tolkning av dette tallet.
- (c) Du ønsker å bruke kji-kvadrattesten for å teste om det er en sammenheng mellom kjønn og holdning (for/mot) til bompenger. Dataene er oppsummert i følgende krystabell:

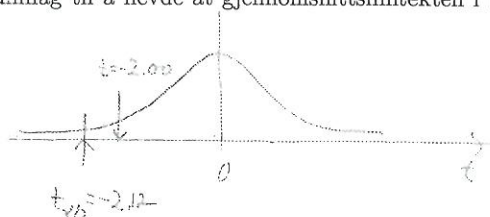
	Mann	Kvinne	
For	3	13	16
Mot	2	15	17
	5	28	33

Hvilken viktig forutsetning for å kunne bruke kji-kvadrattesten er ikke tilstede?

- (d) Et konfidensintervall er et intervall som er beregnet ifra en stikkprøve. Det inneholder en populasjonsparameter med en viss sikkerhet, kalt konfidensnivået. Hvis f.eks. konfidensnivået er 95 %, så er vi 95 % sikre på at intervallet inneholder populasjonsparameteren.

Oppgave 6 (4 poeng)

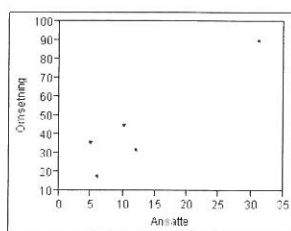
- (a) $H_0 : \mu_1 = 30000$ vs $H_1 : \mu_1 \neq 30000$. Testobservator $t = \frac{23198 - 30000}{14000/\sqrt{17}} = -2.00$. Kritisk verdi er $t_{0.025,16} = 2.12$. Vi kan ikke forkaste H_0 . Det er ikke tilstrekkelig grunnlag til å hevde at gjennomsnittsinntekten i by 1 er



forskjellig fra 30000.

- (b) Uavhengige stikkprøver, siden de to stikkprøvene har forskjellig størrelse. $H_0 : \mu_1 = \mu_2$ og $H_1 : \mu_1 < \mu_2$. Testobservator $t = \frac{26899 - 23198}{\sqrt{13000^2/14 + 14000^2/17}} = 0.76$. Kritisk verdi er $t_{0.05,13} = 1.771$. Vi kan ikke forkaste H_0 . Det er ikke tilstrekkelig grunnlag i dataene til å hevde at by 2 har høyere gjennomsnittsinntekt enn by 1.

Oppgave 7 (4 poeng)



- (a) Det er en ganske klar positiv korrelasjon.
- (b) $r^2 = 0.874$. 87.4 % av variasjonen i omsetning skyldes variasjon i antall ansatte.
- (c) Den forventede verdien E skal være mer enn 5 i mer enn 80 % av cellene. Her er $E < 5$ i to av fire celler. For eksempel er den forventede verdien av menn som er for lik $E = 16 \cdot 5/33 \simeq 2.4$.

Ekamen 11/2010

Oppgave 2

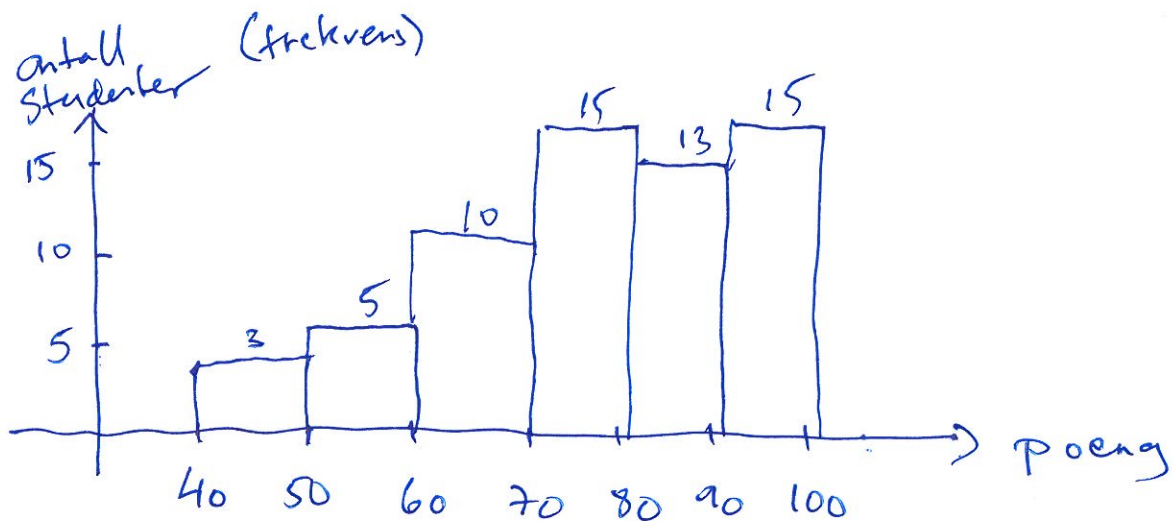
$[40, 50)$ betyr $x \geq 40$

$x < 50$

altså 40, 41, ..., 49

a) Antall studenter:

$$3 + 5 + 10 + 15 + 13 + 15 = 61$$



Ser ikke ut som om antall poeng er normalfordelt (~~normalfordelt~~ skjeiv)

b) Median: $\tilde{x}$ (median) er poengsum 31
når poengsumene ordnes i
stigende rekkefølge.

Fra frekvenstabell: $\tilde{x}$ er i intervallet 70-80

1-3	4-8	9-18	19-33
40-50	50-60	60-70	70-80

$$\tilde{x} = \frac{70 + 80}{2} = 75$$

Gjennomsnitt:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_{61}}{61} =$$

$$\approx \frac{3 \cdot 45 + 5 \cdot 55 + 10 \cdot 65 + 15 \cdot 75 + 13 \cdot 85 + 15 \cdot 95}{61}$$

$$\approx \underline{\underline{77.3}}$$

c) Stekkeprøve: 80, 89, 85, 82, 67, 50, 42

42, 50, 67, 80, 82, 85, 89

Øvre kvartil:

Nummer i rekken: $75\% \cdot 7$
 ≈ 5.25

Runder opp til 6 ($L=6$)

Øvre kvartil: 85

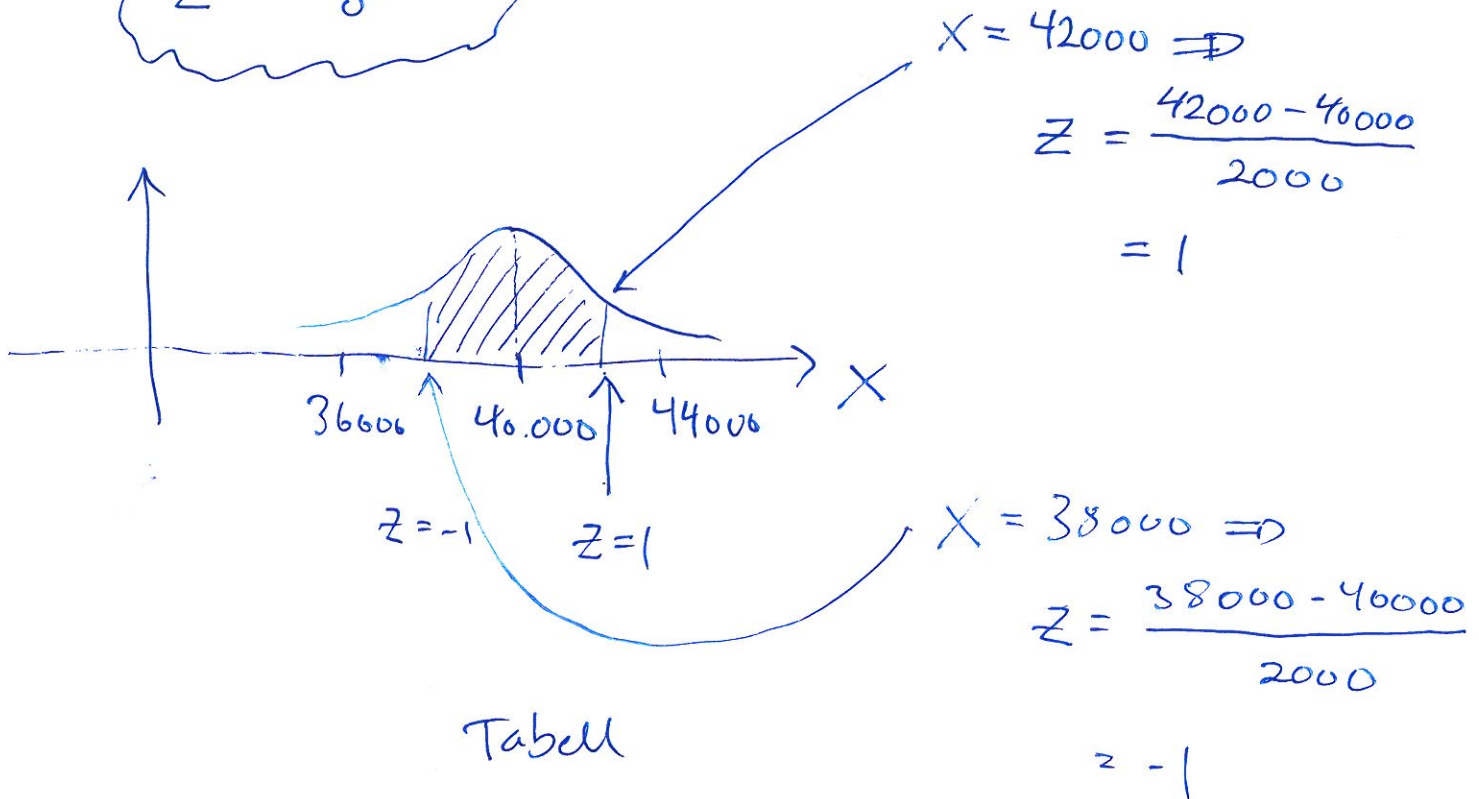
d) Gutt: 80, 89, 85 } std. avvik for
Jente: 82, 67, 50, 42 } jentene er størst

Oppgave 3:

$$X \sim \text{normal}(\mu = 40.000, \sigma = 2000)$$

a) $P(38.000 \leq X \leq 42.000) = P(-1 \leq Z \leq 1)$

$$Z = \frac{X - \mu}{\sigma}$$



Tabell

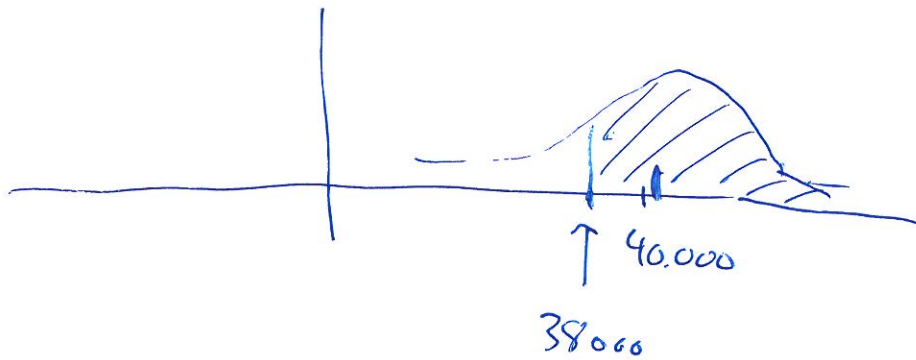
$$P(-1 \leq Z \leq 1) = 0.8413 - 0.1587 = \underline{\underline{0.6826}}$$

$\uparrow \qquad \qquad \uparrow$
 $P(Z \leq 1) \qquad P(Z \leq -1)$

b) $Y = \frac{X_1 + X_2 + X_3 + X_4}{4}$

$$Y \sim \text{normal}(\mu = 40.000, \sigma = \frac{2000}{\sqrt{4}} = 1000)$$

$$P(Y \geq 38.000) = ?$$



$$P(Y \leq 38000) = P\left(\frac{38000 - 40000}{1000}\right) = P(Z \leq -2) \\ = 0.0228$$

$$P(Y \geq 38.000) = 1 - 0.0228 = \underline{\underline{0.9772}}$$

$$P(X \geq 38.000) = 1 - P(X \leq 38000) \\ = 1 - P\left(Z \leq \frac{38000 - 40000}{2000}\right) \\ = 1 - P(Z \leq -1) \\ = 1 - 0.1587 = \underline{\underline{0.8413}}$$

↑
Sannsynlighet
for at ett dekk
lever minst 38.000 km

$$P(\text{Alle 4 dekk lever mer} \\ \text{enn 38000 km}) = 0.8413 \cdot 0.8413 \cdot \\ 0.8413 \cdot 0.8413$$

$$= 0.8413^4 = \underline{\underline{0.5010}}$$

c) Sannsynligheten for at ett dekk varer mindre enn 35000km:

$$p(x < 35000) = p\left(z \leq \frac{35000 - 40000}{2000}\right) = p(z \leq -2.5) \\ = 0.0062$$

Forventet utfall: Binomisk fordeling med $\begin{cases} n=3000 \\ p=0.0062 \end{cases}$

$$np = 3000 \cdot 0.0062 = \underline{\underline{18.6}}$$

d) Første prosentil: z-verdi A slik at $p(z \leq A) = 0.01$ (=1%)
Tabell: $p(z \leq -2.32) = 0.0102$
 $p(z \leq -2.33) = 0.0099$

Dermed er første prosentil ≈ -2.33
(velger heller 0.0099 enn 0.0102 for å være på den sikre siden; 0.0099 gir litt mindre enn 1% feil)

Dermed: $z = -2.33 \Rightarrow x = \mu + \sigma \cdot (-2.33)$
 $= 40.000 + 2000 \cdot (-2.33)$
 $\approx \underline{\underline{35.340 \text{ km}}}$

Skriftlig eksamen i: MET 34311 Statistikk

Eksamensdato: 23.11.10, kl. 14.00-19.00

Tillatte hjelpemidler: Alle hjelpemidler + eksamenskalkulator
TEXAS INSTRUMENTS BA II Plus TM

Innføringsark: Ruter

Totalt antall sider: 4

Oppgave 1

- (a) Forklar kort begrepene populasjon og stikkprøve. Gi deretter et eksempel på en populasjon og en tilhørende stikkprøve.
- (b) En student utarbeider et spørreskjema som hun skal dele ut til en stikkprøve på 120 medstudenter. Gi et eksempel på at hun foretar et utvalg som gjør stikkprøven stratifisert.
- (c) I stikkprøven i oppgave 1b) var det 15 % som oppgav at de røykte. Er dette tallet en observator eller en parameter? Forklar kort. Hvor mange av studentene i stikkprøven røykte ikke?
- (d) På spørreskjemaet til studenten var det et spørsmål der du skulle krysse av for nasjonalitet. Antallet for hver nasjonalitet ble

Norsk	Svensk	Engelsk	Dansk	Annet
75	10	15	5	15

På hvilket målenivå er nasjonalitet? Nevn to typer diagrammer som passer til å presentere disse dataene grafisk. Velg en av typene og lag et diagram.

Oppgave 2

En gruppe studenter tok en matematikkprøve og hver student fikk etterpå vite hvor mange poeng de klarte. Følgende frekvenstabell viser resultatene for hele gruppen:

Poeng	Antall
[40, 50 >	3
[50, 60 >	5
[60, 70 >	10
[70, 80 >	15
[80, 90 >	13
[90, 100 >	15

- (a) Hvor mange studenter var det i gruppen? Lag et histogram over resultatene. Ser poeng ut til å være en normalfordelt variabel? Begrunn svaret kort.
- (b) Bruk frekvenstabellen til å beregne medianen og gjennomsnittet av poeng.
- (c) Du kjenner tre gutter og fire jenter som tok prøven. De fikk 80, 89, 85, 82, 67, 50 og 42 poeng. Regn ut øvre kvartil i denne stikkprøven.
- (d) De tre guttene fikk 80, 89 og 85 poeng, mens de fire jentene fikk 82, 67, 50 og 42 poeng. Er standardavviket til guttenes poenger større eller mindre enn standardavviket til jentenes poenger? Du trenger ikke å regne ut standardavvikene for å svare på dette spørsmålet.

Oppgave 3

En spesiell type bildekk har gjennomsnittlig kjørelengde $\mu = 40000$ km før det må skiftes ut. Standardavviket på gjennomsnittlig kjørelengde er $\sigma = 2000$ km. Vi antar at kjørelengde er en normalfordelt variabel X .

- (a) Du kjøper et tilfeldig valgt bildekk. Hva er sannsynligheten for at det kan kjøres mellom 38000 og 42000 km? Skisser denne sannsynligheten som areal under en tetthetskurve.
- (b) Naboen kjøper fire tilfeldig valgte bildekk. Hva er sannsynligheten for at gjennomsnittlig kjørelengde på de fire dekkene blir mer enn 38000 km? Hva er sannsynligheten for at alle fire dekkene klarer mer enn 38000 km?
- (c) Et firma kjøper 3000 bildekk. Vi antar at denne stikkprøven er tilfeldig. Hvor mange av dekkene vil du forvente har en kjørelengde mindre enn 35000 km?
- (d) Produsenten av dekkene ønsker å innføre en garanti-ordning: Hvis dekket ikke klarer mer enn et visst antall (K) kilometer, så får du det erstattet med et nytt dekk uten omkostninger. Hva skal K være for at bedriften på lang sikt bare trenger å erstatte 1% av dekkene?

Oppgave 4

En bedrift har en support-telefon for sine produkter. Lederen antar at lengden på telefonsamtalene er en normalfordelt variabel. Han forsøker å bestemme gjennomsnittlig lengde på samtalene, μ , ved å tilfeldig velge 16 samtaler. Disse har gjennomsnittlig lengde $\bar{x}_1 = 24$ minutter med standardavvik $s_1 = 4.2$ minutter.

- (a) Lag et 95 % konfidensintervall for μ . Forklar kort hvordan du vil tolke intervallet.
- (b) Ett år senere lurte lederen på om gjennomsnittlig lengde på samtalene har gått ned i forhold til ifjor. Han tar en ny tilfeldig stikkprøve av 18 samtaler. Nå er $\bar{x}_2 = 22$ minutter med standardavvik $s_2 = 3.9$ minutter. Er dette et godt nok grunnlag til å hevde at samtaletiden i gjennomsnitt er redusert? Foreta en hypotesetest med 10% signifikansnivå. Skriv opp H_0 og H_1 og marker testobservator og kritisk verdi i en tetthetskurve.

Sensorveiledning i: MET 34311 Statistikk

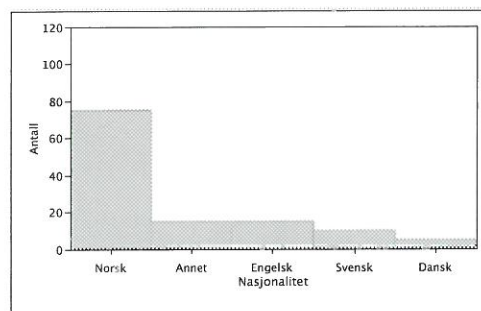
Eksamensdato: 23.11.10, kl. 14.00-19.00

Tillatte hjelpemidler: Alle hjelpemidler + eksamenskalkulator
TEXAS INSTRUMENTS BA II Plus™

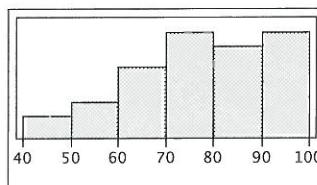
Totalt antall sider: 4

Oppgave 1 (4 poeng)

- (a) Populasjon: Hele mengden av objekter/personer som du er interessert i.
Stikkprøve: Et utvalg av populasjonen.
- (b) Del populasjonen i to undergrupper: mannlige og kvinnelige studenter. Velg så 60 menn og 60 kvinner tilfeldig.
- (c) Observator. 102 røykte ikke.
- (d) Nominalnivå. Kakediagram og Paretodiagram.

**Oppgave 2 (4 poeng)**

- (a) 61. Ikke normalfordelt, siden fordelingen ikke er symmetrisk, men venstre-skjev.



- (b) Siden vi ikke har rådataene, så antar vi at alle observasjonene ligger midt i klasseintervallet. Dvs. vi antar at det var tre personer som alle fikk 45 poeng. Osv. Da blir gjennomsnittet

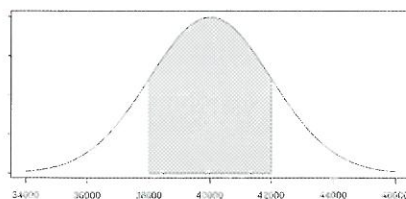
$$\bar{x} = \frac{3 \cdot 45 + 5 \cdot 55 + 10 \cdot 65 + 15 \cdot 75 + 13 \cdot 85 + 15 \cdot 95}{61} = 77.3$$

Medianen er observasjon nr 31 i sortert rekkefølge. Den ligger i klassen $[70, 80 >$, så medianen er 75.

- (c) Vi runder posisjonen $L = \frac{75}{100} \cdot 7 = 5.25$ opp til $L = 6$. Øvre kvartil er da i posisjon 6, altså 85. Andre svaralternativ er også mulige. F.eks. interpolering mellom 82 og 85 gir en øvre kvartil på 82.75.
- (d) Standardavviket er størst blant jentene, siden det er mye større variasjon i oppnådde poeng blant jentene enn guttene.

Oppgave 3 (5 poeng)

- (a) $P(38000 < X < 42000) = P(-1 < z < 1)$. Tabell gir at $P(z < 1) = 0.8413$. Dermed er $P(38000 < X < 42000) = 0.8413 - (1 - 0.8413) = 0.6826$



- (b) Gjennomsnittet $\bar{x}$ er normalfordelt med $\mu_{\bar{x}} = 40000$ og $\sigma_{\bar{x}} = 2000/\sqrt{4} = 1000$, slik at $P(38000 < \bar{x}) = P(-2 < z) = P(z < 2)$. Tabell gir at $P(z < 2) = 0.9772$.
Sannsynligheten for at ett dekk klarer mer enn 38000 er $P(z > -1) = 0.8413$. Sannsynligheten for at alle fire klarer mer enn 38000 lik $0.8413^4 = 0.501$.
- (c) $P(x < 35000) = P(z < -2.5) = 0.0062$. Så vi forventer at $3000 \cdot 0.0062 = 18.6$ dekk klarer mindre enn 35000 km.
- (d) Første persentil for standard normalfordelingen er -2.33 , dvs. $P(z < -2.33) = 0.01$. Så $K = 40000 - 2.33 \cdot 2000 = 35340$ km.

Oppgave 4 (4 poeng)

- (a) Feilmarginen $E = 2.131 \cdot \frac{4.2}{\sqrt{16}} = 2.24$. Konfidensintervallet blir da 24 ± 2.24 .
Tolkning: Vi er 95% sikre på at μ ligger i dette intervallet.