# QED 1–7

## Matematikk for grunnskolelærerutdanningen

*Bind 2*

## Fasit kapittel 4 – Statistikk og kvantitativ metode

CAPPELEN DAMM AKADEMISK

# Kapittel 4

### Oppgave 1

La $x$ være antall øyne på terningen.

a) Vi får følgende sannsynlighetsfordeling

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(x)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

b) $\mu = 3{,}5$
c) $\sigma^2 = 2{,}92$
d) $\sigma = 1{,}71$

### Oppgave 2

La $x$ være verdien på kortet.

a) Vi får følgende sannsynlighetsfordeling

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $P(x)$ | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 |

b) $\mu = 7$
c) $\sigma^2 = 14$
d) $\sigma = 3{,}74$

### Oppgave 3

Det betyr at $Y$ er normalfordelt med forventning på 10 og standardavvik på 4.

### Oppgave 4

Det betyr at $U$ er normalfordelt med forventning på 12 og standardavvik på 3.

### Oppgave 5

Det betyr at $Z$ er normalfordelt med forventning på 0 og standardavvik på 1.

### Oppgave 6

a) $P = 0{,}6628$
b) $P = 0{,}9981$
c) $P = 0{,}3372$
d) $P = 0{,}0019$
e) $P = 0{,}3372$
f) $P = 0{,}0019$

**Oppgave 7**

a) $P = 0{,}0344$
b) $P = 0{,}2946$
c) $P = 0{,}9156$
d) $P = 0{,}6054$
e) $P = 0{,}2946$

**Oppgave 8**

$-z_{\alpha/2} = -2{,}58$ , $z_{\alpha/2} = 2{,}58$

**Oppgave 9**

$P(X \geq 137{,}5) = 0{,}0062$, $P(X \leq 85) = 0{,}1587$

**Oppgave 10**

$P(X \leq 80) = 0{,}8413$, $P(X < 80) = 0{,}8413$

**Oppgave 11**

$P(X \geq 66) = 0{,}0668$

**Oppgave 12**

$P(X \geq 96) = 0{,}9599$, $P(X > 96) = 0{,}9599$

**Oppgave 13**

$P(\bar{X} < 90) = 0{,}0680$, $P(X < 90) = 0{,}2525$

**Oppgave 14**

$P(90 < \bar{X} < 110) = 0{,}9902$, $P(90 < X < 110) = 0{,}4950$

**Oppgave 15**

$P(\bar{X} < 45) = 0{,}0001$, $P(\bar{X} > 50) = 0{,}5000$

**Oppgave 16**

$P(45 < \bar{X} < 50) = 0{,}4999$

**Oppgave 17**

a) $P(\bar{X} > 53) = 0{,}000020$
b) $P(\bar{X} < 47) = 0{,}000020$
c) $P(47 < \bar{X} < 53) = 0{,}999960$

**Oppgave 18**

Trekning med tilbakelegging er knyttet til den binomiske sannsynlighetsfordelingen og trekning uten tilbakelegging er knyttet til den hypergeometriske sannsynlighetsfordelingen. Dersom vi har en hypergeometrisk fordeling der populasjonen er stor i forhold til utvalget, kan en tilnærme den hypergeometriske fordelingen med den binomiske fordelingen.

**Oppgave 19**

$X$ er binomisk fordelt fordi spørsmålene er uavhengig av hverandre og fordi at det er samme sannsynlighet for suksess på hvert enkelt spørsmål.

$P(X \leq 3) = 0{,}2252$ med eksakte metoden
$P(X \leq 3) \approx 0{,}1515$ med normaltilnærmingen uten Yates korreksjon
$P(X \leq 3) \approx 0{,}2206$ med normaltilnærmingen med Yates korreksjon

**Oppgave 20**

$P(X \leq 3) = 0{,}0111$ med eksakte metoden
$P(X \leq 3) \approx 0{,}0104$ med normaltilnærmingen uten Yates korreksjon
$P(X \leq 3) \approx 0{,}0143$ med normaltilnærmingen med Yates korreksjon

**Oppgave 21**

En punktestimator inneholder en tilfeldig variabel. I det øyeblikk vi har gjennomført en undersøkelse som f. eks å kaste med fiskestangen 50 ganger så kan vi lage et punktestimat for sannsynligheten for å få fisk.

**Oppgave 22**

a) Det er ikke urimelig å tenke at antall treff kan beskrives som en binomisk variabel, da vi kan anta at sjansen for treff på et skudd er uavhengig av om vi treffer på de andre. Dette er imidlertid ikke helt riktig. Det er mange forhold i skiskyting som gjør at sjansen for treff kan påvirkes av andre forhold. Det er f. eks mang en skiskytter som har kollapset på siste stående skyting på 20 km normalprogram når de ser de har gullet i siktet.

b) $\hat{p} = \frac{27}{30} = 0{,}9$

c) $\hat{p} = \frac{95}{100} = 0{,}95$

d) I spørsmål c) har vi lagt til grunn 100 skudd. Jo flere observasjoner vi legger til grunn jo mer presist vil estimatet vanligvis være.

**Oppgave 23**

a) Andelen $p$ følger en binomisk fordeling, siden hvert kast er uavhengig av hverandre.
b)-e) Den teoretiske sannsynligheten for å få krone på begge kastene er $p = 0{,}25$. Når en øker antall kast så bør estimatet nærmere seg den teoretiske verdien mer og mer.

**Oppgave 24**

Undersøkelsen viser at blant de spurte så er 49% av de med Facebook profil innom nettstedet daglig. Vi kan betrakte dette som et estimat over hele befolkingen (forutsatt at utvalget er representativt). Med andre ord så vil nær halvparten av Norges befolkning med Facebook profil være innom nettstedet daglig.

**Oppgave 25**

| **Observasjon** | $\boldsymbol{x_i}$ | $\boldsymbol{x_i - \bar{x}}$ | $\boldsymbol{(x_i - \bar{x})^2}$ |
|---|---|---|---|
| 1 | 255 | 105 | 11025 |
| 2 | 85 | -65 | 4225 |
| 3 | 170 | 20 | 400 |
| 4 | 204 | 54 | 2916 |
| 5 | 153 | 3 | 9 |
| 6 | 0 | -150 | 22500 |
| 7 | 152 | 2 | 4 |
| 8 | 119 | -31 | 961 |
| 9 | 204 | 54 | 2916 |
| 10 | 158 | 8 | 64 |
| **Sum** | **1500** | **0** | **45020** |

$\bar{x} = 150$
$s^2 = 67{,}10$
$\hat{\sigma}^2 = 70{,}73$

Populasjonen kan i dette tilfelle være alle studentene som er på skolen.

**Oppgave 26**

a) Populasjonen kan f. eks være alle studentene på skolen.
b) $\hat{\mu} = 2, \hat{\sigma}^2 = 2$
c) En benytter 29 frihetsgrader
d) $\left[2 - 2{,}05\frac{1{,}41}{\sqrt{30},}, 2 + 2{,}05\frac{1{,}41}{\sqrt{30},}\right] = [1.47\,, 2.53]$

e) Dersom observasjonene ikke er normalt fordelt bør en ha minst 30 observasjoner for å bruke sentralgrenseteoremet. Her har vi akkurat 30 observasjoner, så populasjonen bør være normalfordelt for at vi skal få et best mulig resultat.

**Oppgave 27**

a) Populasjonen kan f. eks være alle jentene på en bestemt skole.
b) $\hat{\mu} = 2780$, $\hat{\sigma}^2 = 1863250$,

c) $\left[2780 - 2{,}78\frac{1365}{\sqrt{5},}, 2780 + 2{,}78\frac{1365}{\sqrt{5},}\right] = [1083, 4470]$
d) Her er antall observasjoner mye mindre enn 30. Vi må derfor forutsette at populasjonen er normalfordelt for at resultatet skal være gyldig.

**Oppgave 28**

a) $1 - \alpha = 0{,}9$ $\quad \alpha = 0{,}1$ $\quad \frac{\alpha}{2} = 0{,}05$
b) $t_{\propto/2} = 1{,}72$ $\quad -t_{\propto/2} = -1{,}72$

**Oppgave 29**

Her vil $\hat{p}$ og konfidensintervallet avhenge av andelen du fikk med to kron i oppgave 23.

**Oppgave 30**

$$\left[0{,}7 - 1{,}645\sqrt{\frac{0{,}7 \cdot 0{,}3}{10}}\ ,\ 0{,}7 + 1{,}645\sqrt{\frac{0{,}7 \cdot 0{,}3}{10}}\right] = [0.56\ , 0.84]$$

Her er $n$ mye mindre enn 30 så det er litt tvilsomt å bruke normaltilnærmelsen her.

**Oppgave 31**

$$\left[0{,}39 - 2{,}58\sqrt{\frac{0{,}39 \cdot 0{,}61}{85}}\ ,\ 0{,}39 + 2{,}58\sqrt{\frac{0{,}39 \cdot 0{,}61}{85}}\right] = [0.25\ , 0.53]$$

**Oppgave 32**

$$\left[0{,}54 - 1{,}96\sqrt{\frac{0{,}54 \cdot 0{,}46}{1200}}\ ,\ 0{,}54 + 1{,}96\sqrt{\frac{0{,}54 \cdot 0{,}46}{1200}}\right] = [0.51\ , 0.57]$$

**Oppgave 33**

$$\left[0{,}49 - 2{,}58\sqrt{\frac{0{,}49 \cdot 0{,}51}{1021}}\ ,\ 0{,}49 + 2{,}58\sqrt{\frac{0{,}49 \cdot 0{,}51}{1021}}\right] = [0.45\ , 0.53]$$

**Oppgave 34**

a) $-0{,}15$
b) $0{,}0089$
c) $[-0.3\ , 0]$
d) Forskjellen mellom $\hat{p}_1$og $\hat{p}_2$ vil med 90% sikkerhet ligge i intervallet vi har funnet i c)
e) Nå har vi fått øvre grensen til å bli akkurat lik 0. Med mer nøyaktig regning ville vi fått $[-0.3051\ , 0.0051]$ Her ser vi at $0$ ligger i intervallet og vi har ikke belegg for å påstå at det er forskjell på $p_1$ og $p_2$
f) Dette kan f. eks være elver fra videregående som ønsker å begynne på høyere utdanning.
g) Ja, her er antallet observasjoner så høyt at vi oppfyller kravet med at $n\hat{p}_1 > 5$ og $n\hat{p}_2 > 5$

**Oppgave 35**

a) $0{,}05$
b) $0{,}0004$
c) $[0.01\ , 0.09]$
d) Forskjellen i andelen av EU motstandere ligger med 95% sikkerhet mellom 0,01 og 0,09
e) Nei. Vi kan her gå ut fra at det er forskjell på $p_1$ og $p_2$
f) Populasjonen kan f. eks være Norges befolkning over 18 år.

**Oppgave 36**

a) $0{,}09$
b) $0{,}00098$
c) $[0.04\ , 0.14]$
d) Forskjellen i andelen av kvinner og menn som er innom Facebook daglig ligger med 90% sikkerhet mellom 0,04 og 0,14
e) Nei. Vi kan her gå ut fra at det er forskjell mellom kvinner og menns Facebook vaner.
f) Populasjonen kan f. eks være Norges befolkning over 18 år.

**Oppgave 37**

a) $\hat{\mu}_1 = 3{,}2 \quad \hat{\mu}_2 = 4{,}3 \qquad \hat{\sigma}_1 = 1{,}2 \qquad \hat{\sigma}_2 = 1{,}8$
b) $\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 0{,}33$
c) Det er 83 frihetsgrader etter denne metoden. $t_{\propto/2} = 1{,}9890 \approx 1{,}99$. Denne er funnet i Excel siden tabellen ikke har med verdiene for 83 frihetsgrader. Vi ser imidlertid at det ikke er store avvikene om vi legger til grunn 100 frihetsgrader som vi finner i tabellen.
d) $[-1{,}1 - 1{,}99 \cdot 0{,}33\ , -1{,}1 + 1{,}99 \cdot 0{,}33] = [-1.76\ , -0.44]$
e) Undersøkelsen viser ganske klart at det er forskjell mellom Facebookvanene til disse gruppene. Utvalg som er spurt er imidlertid lite.

**Oppgave 38**

a) $\hat{\mu}_1 = 2{,}7$ $\hat{\mu}_2 = 1{,}2$ $\hat{\sigma}_1 = 1{,}3$ $\hat{\sigma}_2 = 0{,}5$
b) $\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 0{,}22$
c) Antall frihetsgrader er 49. $t_{\propto/2} = 2{,}6799 \approx 2{,}68$. Denne er funnet i Excel siden tabellen ikke har med verdiene for 49 frihetsgrader.
d) $[1{,}5 - 2{,}68 \cdot 0{,}22\ ,\ 1{,}5 + 2{,}68 \cdot 0{,}22] = [0.91\, , 2.09]$
e) Forskjellene i antall ganger disse to gruppene bytter varer i løpet av et år ligger mellom 0,91 og 2,09.

**Oppgave 39**

a) $H_0$ uttrykker at 20% ikke gjør leksene sine. $H_1$ uttrykker at flere enn 20% ikke gjør leksene sine. Verdien $p$ er den virkelige andelen som ikke gjør leksene.
b) Han kan gjøre type I og type II feil. Type I feil i dette tilfelle vil si at vi forkaster $H_0$ feilaktig. Dette vil si at vi konkluderer med at flere enn 20% ikke gjør leksene sine til tross for at det bare er 20% som ikke gjør dem. Type II feil i dette tilfelle vil si at vi ikke forkaster $H_0$ når vi burde gjøre det. I vår situasjon vil det si at vi konkluderer med at 20% ikke gjør leksene sine til tross for at det er flere enn 20% som ikke gjør dem. Sjansen for å begå type I feil er 5%. Sannsynligheten for å begå type II feil kjenner vi ikke.
c) $X$ følgen en binomisk fordeling.
d) Excel viser at $P(X \leq 2|p = 0{,}2) = 0{,}9421$ og at $P(X \leq 3|p = 0{,}2) = 0{,}9933$. Det betyr at minst 3 av elevene må ha svart at de ikke gjør lekser før vi kan forkaste $H_0$.
e) $P(X \geq 4|p = 0{,}2) = 0{,}0067$. Denne forteller oss at sannsynligheten for at flere enn 4 av de 5 elevene ikke har gjort leksene sine er på 0,0067 dersom andelen elever som ikke gjør lekser i hele klassen er på 0,2.

**Oppgave 40**

a) Punktestimator: $\hat{p} = \frac{X}{900}$. Punkestimat: $\hat{p} = \frac{405}{900} = 0{,}45$
b) Antall spurte i stikkprøven som her er 900 er mye mindre enn hele populasjonen som er Norges befolkning. Da kan vi si at $X$ tilnærmer følger en binomisk fordeling.
c) $E(X) = 405.\ Var(X) = 222{,}75.$
d) $\left[0{,}45 - 1{,}96\sqrt{\frac{0{,}45 \cdot 0{,}55}{900}}\ ,\ \ 0{,}45 + 1{,}96\sqrt{\frac{0{,}45 \cdot 0{,}55}{900}}\right] = [0.42\, , 0.48]$

e) $H_0$ uttrykker at 50% er i mot EU medlemskap. $H_1$ uttrykker alternativet, at det ikke er 50% som er imot EU medlemskap.

$g_1 = 0{,}5 - 1{,}96 \cdot 0{,}0166 = 0{,}4675$
$g_2 = 0{,}5 + 1{,}96 \cdot 0{,}0166 = 0{,}5325$

Vi ser at resultatet fra stikkprøven er under den nedre grensen. Altså forkaster vi $H_0$

**Oppgave 41**

a) Sannsynligheten for at et frø skal spire er uavhengig av om de andre spirer. Sjansen for et enkelt frø skal spire er 0,8 og derfor kan vi betrakte $X$ som en binomisk fordelt variabel.
b) $E(X) = 16$. $Var(X) = 3{,}2$. $E(X)$ forteller oss at dersom vi planter 20 frø mange ganger så vil i gjennomsnitt 16 frø spire hver gang. $Var(X)$ forteller oss noe om hvor stor variasjonen vil være fra gang til gang om vi gjør mange slike forsøk.
c) $H_0$ uttrykker at 80% av frøene vil spire. $H_1$ uttrykker alternativet, at det er flere enn 80% som vil spire.
d) $g = 0{,}8 + 1{,}645 \cdot 0{,}0894 = 0{,}9471$. Dette vil si at minst 19 frø må spire før vi forkaster $H_0$.
e) $P(X \geq 19 | p = 0{,}9) = 0{,}3917$ (Eksakt regning)

**Oppgave 42**

a) $H_0$ uttrykker at elevene brukte 50 kroner hver i gjennomsnitt på godter forrige lørdag. $H_1$ uttrykker alternativet, at de brukte mer enn 50 kroner i gjennomsnitt på godter.
b) Det er hvor mye de 25 spurte i gjennomsnitt brukte på godter forrige lørdag.
c) $g = 50 + 1{,}71 \cdot \frac{9}{\sqrt{25}} = 53{,}08$.
d) Siden $\bar{X} = 56$ forkaster vi $H_0$
e) $P(\bar{X} \geq 56 | \mu = 50) = 0{,}0014$. Denne er funnet ved hjelp av Excel siden tabellen ikke angir verdien for 24 frihetsgrader.
f) $g_1 = 50 - 2{,}06 \cdot \frac{9}{\sqrt{25}} = 46{,}29$
$g_2 = 50 + 2{,}06 \cdot \frac{9}{\sqrt{25}} = 53{,}71$
g) Konklusjonen blir den samme som for ensidige alternativet. Vi forkaster $H_0$
h) $P(|t| \geq 3{,}33) = 2 \cdot 0{,}0014 = 0{,}0028$
i) Tore har nok hatt mistanke om at elevene bruker mer enn 50 kroner på lørdagsgodter og har derfor valgt det ensidige alternativet.

**Oppgave 43**

a) $H_0: \mu = 800$ mot $H_1: \mu > 800$
b) Her er dessverre estimert standardavvik $\hat{\sigma}$ uteglemt fra oppgaveteksten. Dersom vi hadde kjent det kunne vi satt opp følgende uttrykk for grensen $g = 800 + 2{,}60 \cdot \frac{\hat{\sigma}}{\sqrt{16}} =$ Vi måtte deretter vurdert om 890 kroner er over eller under denne grensen. Legger vi til grunn at $\hat{\sigma} = 110$ finner vi at $g = 871{,}5$ og vi forkaster dermed $H_0$
c) Her er utvalget mindre enn 30 personer og vi må derfor forutsette at populasjonen er normalfordelt.

**Oppgave 44**

a) $H_0$ uttrykker at leilighetene koster 15000 dollar. $H_1$ uttrykker alternativet, at de koster enten mer eller mindre enn 15000 dollar|

b) $g_1 = 150 - 2{,}14 \cdot \frac{34{,}21}{\sqrt{15}} = 131{,}10$

$g_1 = 150 + 2{,}14 \cdot \frac{34{,}21}{\sqrt{15}} = 168{,}90$

Vi beholder $H_0$.

c) $\left[145{,}05 - 2{,}14 \cdot \frac{34{,}21}{\sqrt{15}}\ ,\ \ 145{,}05 + 2{,}14 \cdot \frac{34{,}21}{\sqrt{15}}\right] = [126.15\ , 163.95]$

d) Vi ser at konfidensintervallet omslutter den påståtte verdien i $H_0$. Dette samsvarer med at vi beholdt $H_0$ i hypotesetesten.

e) $P(|t| \geq 0{,}56) = 2 \cdot 0{,}29 = 0{,}58$

**Oppgave 45**

a) $H_0$ uttrykker at antall søsken i hver familie i gjennomsnitt er 1,90. $H_1$ uttrykker alternativet, at antall søsken i familiene i gjennomsnitt er flere eller færre enn 1,90.

b) $g_1 = 1{,}90 - 2{,}09 \cdot \frac{1{,}44}{\sqrt{20}} = 1{,}23$

$g_1 = 1{,}90 + 2{,}09 \cdot \frac{1{,}44}{\sqrt{20}} = 2{,}57$

Vi beholder $H_0$ siden studentene i snitt er 2,20 søsken i de spurte familiene. Dette ligger godt innenfor grensene vi har beregnet.

**Oppgave 46**

a) $\bar{X} = 2{,}53 \quad \hat{\sigma} = 1{,}57$

b) $H_0$: $\mu = 2{,}5$ mot $H_1$: $\mu > 2{,}5$

c) $g = 2{,}5 + 2{,}04 \cdot \frac{1{,}57}{\sqrt{32}} = 3{,}07$

Vi beholder $H_0$. Vi har ikke noe grunnlag for å påstå at ungdommene bruker mer enn 2,5 time på sosiale medier.

d) $P(t \geq 0{,}11) = 0{,}46$

e) Se spørsmål c)

**Oppgave 47**

a) $H_0$ betyr at det ikke er forskjell i andelen som jobber for pengene på 6. og 7. trinn. $H_1$ uttrykker at det er forskjell.

b) $\hat{p}_1 = \frac{19}{54} = 0{,}35 \qquad \hat{p}_2 = \frac{21}{48} = 0{,}44$

Det virker som elevene i 7 klassene er flinkere til å jobbe hjemme enn elevene i 6. klasse.

c) Vi finner at $\hat{\sigma}_0 = 0{,}10$. Vi finner videre at $\hat{p}_1 - \hat{p}_2 = -0{,}11 > -1{,}28 \cdot 0{,}1 = -0{,}128$.

Dermed beholder vi $H_0$

d) Vi beholder $H_0$ også med 5% signifikansnivå. $\hat{p}_1 - \hat{p}_2 = -0{,}11 > -1{,}645 \cdot 0{,}1 = -0{,}1645$

e) Vi forkaster $H_0$ om $|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \cdot \hat{\sigma}_0$. Vi får at $|\hat{p}_1 - \hat{p}_2| = 0{,}11$ og $z_{\alpha/2} \cdot \hat{\sigma}_0 = 0{,}1645$, altså beholder vi $H_0$.

f) Her er kravet som stilles for å benytte normaltilnærmelsen oppfylt.

**Oppgave 48**

a) $H_0$ betyr at det ikke er forskjell på andelen av menn og kvinner som har hørt om Rent tøy. $H_1$ uttrykker at det er forskjell.

b) $\hat{p}_1 = \frac{79}{250} = 0{,}32 \quad \hat{p}_2 = \frac{57}{150} = 0{,}38$

Det virker som at det er flere kvinner enn menn som har hørt om det nye vaskepulveret.

c) Vi finner at $\hat{\sigma}_0 = 0{,}05$. Vi finner videre at $\hat{p}_1 - \hat{p}_2 = -0{,}06 > -1{,}645 \cdot 0{,}05 = -0{,}08$. Dermed beholder vi $H_0$. Det er ikke signifikant forskjell.

d) Vi beholder $H_0$ også med 1% signifikansnivå. $\hat{p}_1 - \hat{p}_2 = -0{,}06 > -2{,}33 \cdot 0{,}05 = -0{,}12$

e) Vi forkaster $H_0$ om $|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \cdot \hat{\sigma}_0$. Vi får at $|\hat{p}_1 - \hat{p}_2| = 0{,}06$ og $z_{\alpha/2} \cdot \hat{\sigma}_0 = 0{,}10$, altså beholder vi $H_0$.

**Oppgave 49**

a) $\hat{p}_1 - \hat{p}_2 = -0{,}15$

b) Vi finner at $\hat{\sigma}_0^2 = 0{,}009$.

c) Vi forkaster $H_0$ om $|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \cdot \hat{\sigma}_0$. Vi får at $|\hat{p}_1 - \hat{p}_2| = 0{,}15$ og $z_{\alpha/2} \cdot \hat{\sigma}_0 = 0{,}157$, altså beholder vi $H_0$.

d) Ja forutsetningene for å bruke normalfordelingen er oppfylt.

**Oppgave 50**

a) $\hat{p}_1 - \hat{p}_2 = 0{,}05$

b) Vi finner at $\hat{\sigma}_0^2 = 0{,}00045$.

c) Vi forkaster $H_0$ om $|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \cdot \hat{\sigma}_0$. Vi får at $|\hat{p}_1 - \hat{p}_2| = 0{,}05$ og $z_{\alpha/2} \cdot \hat{\sigma}_0 = 0{,}04$, altså forkaster vi $H_0$.

d)Populasjonen kan være Norges befolkning over 18 år.

**Oppgave 51**

a) $\hat{p}_1 - \hat{p}_2 = 0{,}09$

b) Vi finner at $\hat{\sigma}_0^2 = 0{,}00099$.

c) Vi forkaster $H_0$ om $|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \cdot \hat{\sigma}_0$. Vi får at $|\hat{p}_1 - \hat{p}_2| = 0{,}09$ og $z_{\alpha/2} \cdot \hat{\sigma}_0 = 0{,}062$, altså forkaster vi $H_0$.

**Oppgave 52**

a) $H_0$ uttrykker at det ikke er forskjell mellom karakterene på kurs 1 og kurs 2. $H_1$ uttrykker alternativet, at det er forskjell.
b) $\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 0{,}1121$
c) Antall frihetsgrader er 117. $t_{\propto/2} = 1{,}66$. (Denne er funnet ved hjelp av Excel)
d) $|t| = \left|\frac{2{,}5-3{,}1}{0{,}1121}\right| = 5{,}35 > 1{,}66$. Dermed forkaster vi hypotesen $H_0$

**Oppgave 53**

a) $H_0$ uttrykker at det ikke er forskjell mellom Facebook vanene til tenåringene og de eldre ungdommene. $H_1$ uttrykker alternativet, at det er forskjell.
b) $\hat{\mu}_1 = 14{,}2$ $\hat{\mu}_2 = 9{,}5$ $\hat{\sigma}_1 = 3{,}2$ $\hat{\sigma}_2 = 2{,}8$
c) $\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 0{,}6559$
d) Antall frihetsgrader er 83. $t_{\propto/2} = 1{,}99$. (Denne er funnet ved hjelp av Excel). NB. Signifikansnivå er ikke oppgitt, men vi legger til grunn infoen i spørsmål e)
e) $|t| = \left|\frac{14{,}2-9{,}4}{0{,}6559}\right| = 7{,}32 > 1{,}99$. Dermed forkaster vi hypotesen $H_0$. Med andre ord betyr det at vi har grunnlag for å påstå at det er forskjell på Facebook vanene til tenåringene og de eldre ungdommene.
f) Utvalget vi bruker er lite og et relevant spørsmål å stille er om dette er et representativt utvalg av tenåringer og ungdommer siden antall spurte er såpass få.

**Oppgave 54**

a) $H_0$ uttrykker at det ikke er forskjell mellom antall ganger tenåringene og de eldre ungdommene bytter klær. $H_1$ uttrykker alternativet, at det er forskjell.
b) $\hat{\mu}_1 = 2{,}7$ $\hat{\mu}_2 = 1{,}2$ $\hat{\sigma}_1 = 1{,}3$ $\hat{\sigma}_2 = 0{,}5$
c) $\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 0{,}2186$
d) Antall frihetsgrader er 49,17 som vi runder ned til 49. $t_{\propto/2} = 2{,}01$. (Denne er funnet ved hjelp av Excel). NB. Signifikansnivå i oppgaven er ikke oppgitt. Vi har lagt til grunn at det er 5% i fasiten. Også på spørsmål e) har vi brukt 5% signifikansnivå.
e) $|t| = \left|\frac{2{,}7-1{,}2}{0{,}2186}\right| = 6{,}86 > 2{,}01$. Dermed forkaster vi hypotesen $H_0$.

**Oppgave 55**

a) Nedfor er tabell over differansene

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{X_1}$ | 2,6 | 2,3 | 2,5 | 2,4 | 2,4 | 2,4 | 2,1 | 2,7 | 2,9 | 2,2 |
| $\boldsymbol{X_2}$ | 2,3 | 1,9 | 2,4 | 2,2 | 1,9 | 2,5 | 1,7 | 2,5 | 2,6 | 2,3 |
| $\boldsymbol{D_i}$ | 0,3 | 0,4 | 0,1 | 0,2 | 0,5 | -0,1 | 0,4 | 0,2 | 0,3 | -0,1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 2,1 | 2,2 | 1,7 | 2,0 | 2,2 | 2,2 | 2,9 | 2,4 | 2,9 | 2,2 |
| $X_2$ | 2,2 | 2,0 | 1,4 | 1,9 | 1,7 | 1,7 | 2,5 | 2,4 | 2,3 | 1,8 |
| $D_i$ | -0,1 | 0,2 | 0,3 | 0,1 | 0,5 | 0,5 | 0,4 | 0,0 | 0,6 | 0,4 |

$\overline{D} = 0{,}255$

b) $\hat{\sigma}_D = 0{,}216$

c) $|t| = \left|\frac{0{,}255}{\frac{0{,}216}{\sqrt{20}}}\right| = 5{,}28 > t_{\propto/2} = 2{,}09$. Vi forkaster hypotesen.

d) Dette betyr at det er signifikant forskjell mellom prestasjonene på grunnkurset og påbyggingskurset.

e) $\left[0{,}255 - 2{,}09 \cdot \frac{0{,}216}{\sqrt{20}}\ ,\ \ 0{,}255 + 2{,}09 \cdot \frac{0{,}216}{\sqrt{20}}\right] = [0.15\ , 0.36]$

Dersom vi skulle beholdt $H_0$ måtte 0 ligge i konfidensintervallet. Det gjør det ikke og det er dermed konsistent med konklusjonen i sted om at vi forkaster hypotesen

**Oppgave 56**

a) Nedfor er tabell over differansene

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Før** | 29 | 22 | 25 | 29 | 26 | 24 | 31 | 46 | 34 | 28 |
| **Etter** | 30 | 26 | 25 | 35 | 33 | 36 | 32 | 54 | 50 | 43 |
| $D_i$ | 1 | 4 | 0 | 6 | 7 | 12 | 1 | 8 | 16 | 15 |

$\overline{D} = 7$

b) $\hat{\sigma}_D = 5{,}793$

c) $t = \frac{7}{\frac{5{,}793}{\sqrt{10}}} = 3{,}82 \geq t_{\propto/2} = 2{,}82$. Vi forkaster hypotesen.

d) Dette betyr at det er signifikant forskjell mellom prestasjonene til elevene før og etter lesekurset.

e) $P(\overline{D} \geq 7) = 0{,}002$

**Oppgave 57**

$|t| = \frac{0{,}6}{\sqrt{\frac{1-0{,}36}{30}}} = 4{,}1 > t_{\propto/2} = 1{,}7$ . Vi forkaster dermed $H_0$

**Oppgave 58**

Her må først korrelasjonskoeffisienten beregnes. Vi kan vise at $r = 0{,}87$. Antall elever er 15 slik at $n = 15$

$|t| = \frac{0{,}87}{\sqrt{\frac{1-0{,}7569}{13}}} = 6{,}4 > t_{\propto/2} = 2{,}16$ . Vi forkaster dermed $H_0$

**Oppgave 59**

Fra definisjonen av $Cov(X,Y)$, $s_x$ og $s_y$ har vi

$$Cov(X,Y) = \frac{\sum(x_i - \bar{X})\sum(y_i - \bar{Y})}{n}$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}}$$

$$s_y = \sqrt{\frac{\sum(y_i - \bar{Y})^2}{n}}$$

Dette gir oss

$$\frac{Cov(X,Y)}{s_x \cdot s_y} = \frac{\frac{\sum(x_i - \bar{X})\sum(y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} \cdot \sqrt{\frac{\sum(y_i - \bar{Y})^2}{n}}} = \frac{\frac{1}{n} \cdot \sum(x_i - \bar{X})\sum(y_i - \bar{Y})}{\sqrt{\frac{1}{n}} \cdot \sqrt{\sum(x_i - \bar{X})^2} \cdot \sqrt{\frac{1}{n}} \cdot \sqrt{\sum(y_i - \bar{Y})^2}} =$$

$$= \frac{\sum(x_i - \bar{X})\sum(y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2} \cdot \sqrt{\sum(y_i - \bar{Y})^2}}$$

**Oppgave 60**

$$s_x = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} = \sqrt{\frac{1304}{8}} = 12{,}77$$

$$s_y = \sqrt{\frac{\sum(y_i - \bar{Y})^2}{n}} = \sqrt{\frac{836}{8}} = 10{,}22$$

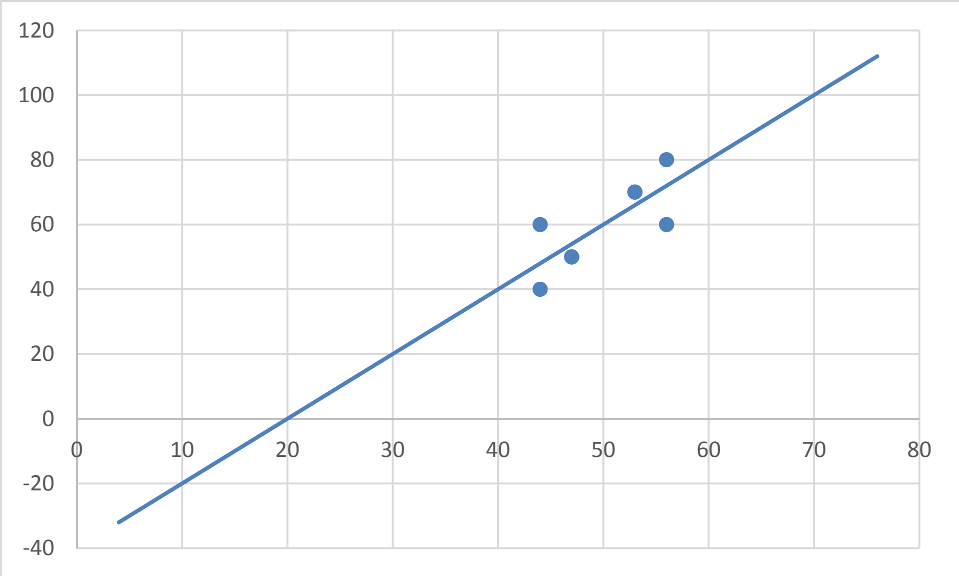
**Oppgave 61**

a) $Y = -40 + 2X$
b) $\hat{\beta}_1$ er stigningstallet til linjen. Den forteller oss at avlingen i gjennomsnitt økte med to enheter når vi øker gjødselmengden med en enhet.
c) $Y = -40 + 2 \cdot 47 = 54$. Det betyr at om vi bruker 47 enheter med gjødsel kan vi forvente en avling på 54 enher.
d) $Y = -40 + 2 \cdot 56 = 72$. Med $X = 56$ ser vi at de to bøndene har fått avling på henholdsvis 60 og 80 enheter. Vårt estimat ligger (nesten) midt mellom disse verdiene, noe som ikke er urimelig.
e) $Y = -40 + 2 \cdot 0 = -40$. I følge modellen sier den at om vi ikke bruker gjødsel så får vi en avling på -40 enheter. Dette resultatet gir ingen mening. Vi kan ikke ha en negativ avling.

**Oppgave 62**

a) $Y = -40 + 2X$

b) Figuren under viser punktene og regresjonslinjen




**Oppgave 63**

a) Det er svært tett sammenheng mellom antall senger og antall liggedøgn på sykehusene. En korrelasjon på 0.974 er høyt. Det er liten sammenheng mellom antall senger i forhold til de ulike regionene og det samme gjelder også for antall liggedøgn i forhold til regionene.

b) Her ligger punktene nokså konsentrert om en (tenkt) rett linje og det virker fornuftig å bruke en lineær regresjonsmodell her.

c) Det er $\beta_1$ som er den mest interessante her. Den forteller oss at antall liggedøgn vil øke med i gjennomsnitt $\beta_1$ dersom vi øker antall senger med en. $\beta_0$ forteller oss hvor regresjonslinjen skjærer $y$ aksen. Verdien $\varepsilon$ er feilleddet. Denne betegner det individuelle avvik fra trenden.

d) $Y = 144 + 31{,}5X$. Regresjonslikningen viser at en økning på en seng vil generere 31,5 flere liggedøgn i den gitte perioden. Det er vanskelig å gi noe god tolkning av 144. Det er verdien der regresjonslinjen skjærer $y$ – aksen. Men å si vi har 144 liggedøgn med 0 senger gir ikke mening. Modell har ikke noe gyldighet for et så lavt antall senger. Verdien 1,564 er det estimerte standardavviket til $\hat{\beta}_1$

**Oppgave 64**

a) La oss se på første kolonnen. Vi ser først på antall deltakere og totaltilfredshet. Her ser vi at korrelasjonskoeffisienten er negativ. Det betyr at jo færre deltakere det er jo mer tilfreds er dem. Når deltakerantallet øker synker tilfredsheten. Når det gjelder de to andre forholdene mat/total og lokalitet/total ser vi at vi har en positivt korrelasjonskoeffisient. Det betyr at det er sammenheng mellom hvordan de opplever maten/lokalitetene og hvor tilfreds dem er totalt sett.

Når det gjelder siste del av spørsmålet ser vi at vi har en negativ korrelasjonskoeffisient for antall og mat. Vår undersøkelse tyder derfor på at påstanden at jo flere deltakere det er jo færre er det som er tilfredshet med mattilbudet.

b) Det er $\beta_1$ som er den mest interessante her. Den forteller oss at totaltilfredsheten vil øke med i gjennomsnitt $\beta_1$ dersom vi øker verdien for tilfredshet med maten med en. $\beta_0$ forteller oss hvor regresjonslinjen skjærer $y$ aksen. Verdien $\varepsilon$ er feilleddet. Denne betegner det individuelle avvik fra trenden.

c) Dataene ligger litt spredt, men vi ser likevel en trend med at totaltilfredsheten øker i forhold til hvordan de opplever mattilbudet, så det er mulig å bruke en lineær regresjonsmodell her.

d) $Y = 2{,}22 + 0{,}58X$. Verdien $0{,}58$ forteller oss at dersom en tilfredsheten med mat øker med en verdi på en vil totaltilfredsheten øke med $0{,}58$. Det er vanskelig å gi noe god tolkning av 2,22. Det er verdien der regresjonslinjen skjærer $y$ – aksen. Hvis vi velger $X = 1$ får vi at $Y = 2{,}8$. Dette viser at selv om gjest er svært misfornøyd med mattilbudet, så vurderes likevel ikke totalopplevelsen som like dårlig.

e) Korrelasjonskoeffisienten finner vi fra tabellen. Den er $r = 0{,}6472$.