

# Eksamen, høsten 20 deleksamen Statistikk

## Ingeniør-varianten

### Løsningsforslag og sensorveiledning.

Oppgavesettet har fire oppgaver, og på hver oppgave fikk studentene trukket en av fire mulige versjoner av oppgaven. Løsningsforslaget prøver å dekke alle versjonene, ofte gis fullstendig løsning for en versjon, så beskrives endringer i de andre versjonene.

#### Oppgave 1. Oppgavetekst Versjon 1:

- a) Une har et uavklart forhold til sin identitet som Senterpartivelger. Noen ganger føler hun at hun er Sp-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Une disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Une føler seg som Sp-velger er 75%.

Gitt at Une føler seg som SP-velger, så er det en sannsynlighet på 14% for at hun synes at partileder Vedum er et fugleskremsel.

I motsatt fall, når Une ikke føler seg som Sp-velger, så er det en sannsynlighet på 83% for at hun synes at partileder Vedum er et fugleskremsel.

Hva er den subjektive sannsynligheten for at Une føler at hun er Sp-velger gitt at partileder Vedum er et fugleskremsel?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 700 325 Sp-velgere. Anta at det gjøres et utvalg på 100 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Sp-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Sp-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,03$ . La  $Y$  være antall løgnere i utvalget på  $n = 100$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 2 eller færre løgnere i utvalget.

#### Løsning Versjon 1:

- a) Det er to hendelser i oppgaveteksten. La  $A$  være at "partileder Vedum er et fugleskremsel" og la  $B$  være "Une føler seg som Sp-velger". Avlest fra teksten har vi oppgitt

$$P(B) = 0,75$$

$$P(A|B) = 0,14$$

$$P(A|\bar{B}) = 0,83$$

Vi finner sannsynligheten for ikke  $B$ :

$$P(\bar{B}) = 1 - P(B) = 1 - 0,75 = 0,25$$

Total sannsynlighet for  $A$  er

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}) = 0,14 \cdot 0,75 + 0,83 \cdot 0,25 = 0,3125$$

Ved Bayes' regel har vi

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{0,14 \cdot 0,75}{0,3125} = \underline{\underline{0,336}}$$

Det er 33,6% sannsynlig at Une føler seg som Sp-velger gitt at Vedum er et fugleskremsel.

- b) Dette er oppsettet for en hypergeometrisk fordeling for  $X$ . Populasjonen er  $N = 3869200$ , hvorav  $M = 700325$  er spesielle. Utvalget er  $n = 100$ . Formlene for hypergeometrisk fordeling gir

$$p = \frac{M}{N} = \frac{700325}{3869200} = 0,1810$$

$$E(X) = np = 100 \cdot 0,1810 = \underline{\underline{18,10}}$$

$$\text{Var}(X) = np(1-p) \cdot \frac{N-n}{N-1} = 100 \cdot 0,1810 \cdot (1-0,1810) \cdot \frac{3869200-100}{3869200-1} = \underline{\underline{14,82}}$$

Siden variansen er større enn 10 og  $p$  ikke nær 0 eller 1, så er  $X$  tilnærmet normalfordelt. Grensene for et 95% spredningsintervall er gitt ved  $\mu \pm z_{\alpha/2} \cdot \sigma$  der  $\alpha = 0,05$ . Vi setter inn

$$\mu = E(X) = 18,10$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{14,82} = 3,850$$

$$z_{\alpha/2} = z_{0,025} = 1,960$$

Grensene blir

$$\text{nedre } \mu - z_{\alpha/2} \cdot \sigma = 18,10 - 1,960 \cdot 3,850 = 10,55$$

$$\text{øvre } \mu + z_{\alpha/2} \cdot \sigma = 18,10 + 1,960 \cdot 3,850 = 25,65$$

Et 95% spredningsintervall for antall Sp-velgere i utvalget er [10,55, 25,65].

- c) **Løsningen på denne deloppgaven i ulike oppgaveversjoner bruker forskjellige metoder. Det avgjørende er om  $Y$  er tilnærmet normalfordelt eller ikke. Metoden under er for tilfellene der  $Y$  ikke er tilnærmet normalfordelt.**

Antall løgnere  $Y$  i utvalget er binomisk fordelt med  $p = 0,03$  og  $n = 100$ . Variansen blir

$$\text{Var}(Y) = np = 100 \cdot 0,03 = 3$$

Siden variansen er under 5 sier vi ikke at  $Y$  er tilnærmet normalfordelt.

Vi finner punktsannsynligheter for binomisk fordeling ved formelen

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}.$$

Det gir:

$$P(Y = 0) = \binom{100}{0} \cdot 0,03^0 \cdot 0,97^{100} = 1 \cdot 1 \cdot 0,04755 = 0,04755$$

$$P(Y = 1) = \binom{100}{1} \cdot 0,03^1 \cdot 0,97^{99} = 100 \cdot 0,03 \cdot 0,04902 = 0,1471$$

$$P(Y = 2) = \binom{100}{2} \cdot 0,03^2 \cdot 0,97^{98} = 4950 \cdot 0,0009 \cdot 0,05054 = 0,2252$$

Sannsynligheten for 2 eller færre løgnere i utvalget blir

$$P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = 0,04755 + 0,1471 + 0,2252 = \underline{\underline{0,4198}}$$

## Oppgavetekst versjon 2:

- a) Fadi har et uavklart forhold til sin identitet som Arbeiderpartivelger. Noen ganger føler han at han er Ap-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Fadi disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Fadi føler seg som Ap-velger er 80%.

Gitt at Fadi føler seg som Ap-velger, så er det en sannsynlighet på 90% for at han irriterer seg fordi partileder Støre kommer fra overklassen.

I motsatt fall, når Fadi ikke føler seg som Ap-velger, så er det en sannsynlighet på 15% for at han irriterer seg fordi partileder Støre kommer fra overklassen.

Hva er den subjektive sannsynligheten for at Fadi føler at han er Ap-velger gitt at han allerede irriterer seg fordi partileder Støre kommer fra overklassen?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 831 880 Ap-velgere. Anta at det gjøres et utvalg på 1000 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Ap-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Ap-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,05$ . La  $Y$  være antall løgnere i utvalget på  $n = 1000$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 40 eller færre løgnere i utvalget.

## Fasit versjon 2:

- a) Hendelser  $A$  "Fadi irriterer seg fordi Støre kommer fra overklassen" og  $B$  "Fadi føler seg som Ap-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,90 \cdot 0,80}{0,90 \cdot 0,80 + 0,15 \cdot 0,20} = 0,96$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 215,0$   $\text{Var}(X) = 168,7$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[189,5, 240,5]$ .

- c) **Dette er metoden når  $Y$  er tilnærmet normalfordelt.**

$Y$  er binomisk fordelt med  $n = 1000$  og  $p = 0,05$ . Variansen er  $\text{Var}(Y) = np(1-p) = 47,5$ . Siden variansen er stor er  $Y$  tilnærmet normalfordelt. Vi kan finne  $P(Y \leq 40)$  ved normaltilnærming. Treger derfor forventning og standardavvik:

$$\mu = E(Y) = np = 50$$

$$\sigma = \sqrt{\text{Var}(Y)} = \sqrt{47,5} = 6,892$$

Vi får

$$P(Y \leq 40) \approx G\left(\frac{40 - \mu}{\sigma}\right) = G(-1,45) = \underline{\underline{0,0734}}.$$

Dersom man bruker heltallskorreksjon får man den mer nøyaktige tilnærmingen  $P(Y \leq 40) \approx 0,08404$ .

### Oppgavetekst versjon 3:

- a) Helge har et uavklart forhold til sin identitet som MDG-velger. Noen ganger føler han at han er MDG-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Helge disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Helge føler seg som MDG-velger er 70%. Gitt at Helge føler seg som MDG-velger, så er det en sannsynlighet på 30% for at han er opptatt av å elektrifisere oljeplattformene før de stenges.

I motsatt fall, når Helge ikke føler seg som MDG-velger, så er det en sannsynlighet på 55% for at han er opptatt av å elektrifisere oljeplattformene før de stenges.

Hva er den subjektive sannsynligheten for at Helge føler at han er MDG-velger gitt at han for tiden er opptatt av å elektrifisere oljeplattformene før de stenges?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 170 240 MDG-velgere. Anta at det gjøres et utvalg på 6000 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er MDG-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet MDG-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,04$ . La  $Y$  være antall løgnere i utvalget på  $n = 6000$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 250 eller færre løgnere i utvalget.

### Fasit versjon 3:

- a) Hendelser  $A$  "Helge er opptatt av å elektrifisere oljeplattformene før de stenges" og  $B$  "Helge føler seg som MDG-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,30 \cdot 0,70}{0,30 \cdot 0,70 + 0,55 \cdot 0,30} = 0,56$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 264,0$   $\text{Var}(X) = 252,0$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[232,9, 295,1]$ .
- c)  $Y$  er binomisk fordelt med  $n = 6000$  og  $p = 0,04$ .  $Y$  **tilnærmet normalfordelt**.  $P(Y \leq 250) \approx 0,7450$  eller med heltallskorreksjon:  $P(Y \leq 250) \approx 0,7555$ .

#### Oppgavetekst versjon 4:

- a) Ixi har et uavklart forhold til sin identitet som Kristelig Folkeparti-velger. Noen ganger føler hun at hun er KrF-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Ixi disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Ixi føler seg som KrF-velger er 85%.

Gitt at Ixi føler seg som KrF-velger, så er det en sannsynlighet på 30% for at hun ønsker å se Olaug Bollestad i neste sesong av Farmen kjendis.

I motsatt fall, når Ixi ikke føler seg som KrF-velger, så er det en sannsynlighet på 80% for at hun ønsker å se Olaug Bollestad i neste sesong av Farmen kjendis.

Hva er den subjektive sannsynligheten for at Ixi føler at hun er KrF-velger gitt at hun for tiden gjerne kunne ønske å se Olaug Bollestad i Farmen kjendis?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 135 666 KrF-velgere. Anta at det gjøres et utvalg på 600 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er KrF-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet KrF-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,01$ . La  $Y$  være antall løgnere i utvalget på  $n = 600$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 2 eller færre løgnere i utvalget.

#### Fasit versjon 4:

- a) Hendelser  $A$  "Ixi ønsker å se Olaug Bollestad i Farme kjendis" og  $B$  "Ixi føler seg som KrF-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,30 \cdot 0,85}{0,30 \cdot 0,85 + 0,80 \cdot 0,15} = 0,68$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 21,04$   $\text{Var}(X) = 20,30$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[12,2, 29,9]$ .
- c)  $Y$  er binomisk fordelt med  $n = 600$  og  $p = 0,01$ . Varians lik 5,94, så  $Y$  **tilnærmet normalfordelt**, men det er helt på grensa. I dette tilfellet kan begge metoder brukes. Med normaltilnærming:  $P(Y \leq 2) \approx 0,0504$  eller med heltallskorreksjon:  $P(Y \leq 2) \approx 0,0755$ . Med punktsannsynligheter:  $P(Y = 0) = 0,002405$ ,  $P(Y = 1) = 0,01458$ ,  $P(Y = 2) = 0,04410$ ,  $P(Y \leq 2) = 0,06108$ . I dette tilfellet er det relativt store forskjeller på de ulike tilnærmingene. Metoden med punktsannsynligheter gir det eksakt riktige svaret.

### Oppgavetekst Versjon 5:

- a) Hassen har et uavklart forhold til sin identitet som Venstre-velger. Noen ganger føler han at han er Venstre-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Hassen disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Hassen føler seg som Venstre-velger er 55%.

Gitt at Hassen føler seg som Venstre-velger, så er det en sannsynlighet på 9% for at han ønsker å se Abid Raja i neste sesong av Skal vi danse.

I motsatt fall, når Hassen ikke føler seg som Venstre-velger, så er det en sannsynlighet på 69% for at han ønsker å se Abid Raja i neste sesong av Skal vi danse.

Hva er den subjektive sannsynligheten for at Hassen føler at han er Venstre-velger gitt at Hassen en lørdagskveld intenst ønsker at Abid Raja var deltager i Skal vi danse?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 135 123 Venstre-velgere. Anta at det gjøres et utvalg på 2000 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Venstre-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Venstre-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,01$ . La  $Y$  være antall løgnere i utvalget på  $n = 2000$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 25 eller færre løgnere i utvalget.

### Fasit Versjon 5:

- a) Hendelser  $A$  "Hassen ønsker å se Abid Raja i Skal vi danse" og  $B$  "Hassen føler seg som Venstre-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,09 \cdot 0,55}{0,09 \cdot 0,55 + 0,69 \cdot 0,45} = 0,1375$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 69,8$   $\text{Var}(X) = 67,4$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[53,8, 85,9]$ .
- c)  $Y$  er binomisk fordelt med  $n = 2000$  og  $p = 0,01$ . Varians 19,8.  $Y$  **tilnærmet normalfordelt**.  $P(Y \leq 25) \approx 0,8694$  eller med heltallskorreksjon:  $P(Y \leq 25) \approx 0,8918$ .



### Oppgavetekst versjon 6:

- a) Idunn har et uavklart forhold til sin identitet som Fremskrittspartivelger. Noen ganger føler hun at hun er Frp-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Idunn disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Idunn føler seg som Frp-velger er 90%. Gitt at Idunn føler seg som Frp-velger, så er det en sannsynlighet på 12% for at hun ønsker at Black Box teater kunne sette opp et show der dragkinggruppa Gutta harselerer med bompenger og alkoholavgifter.

I motsatt fall, når Idunn ikke føler seg som Frp-velger, så er det en sannsynlighet på 72% for at hun ønsker at Black Box teater kunne sette opp et show der dragkinggruppa Gutta harselerer med bompenger og alkoholavgifter.

Hva er den subjektive sannsynligheten for at Idunn føler at hun er Frp-velger gitt at Idunn en helg ønsker å se Gutta ha show på Black Box om bompenger og alkoholavgifter?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 472 166 Frp-velgere. Anta at det gjøres et utvalg på 80 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Frp-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Frp-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,02$ . La  $Y$  være antall løgnere i utvalget på  $n = 80$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 3 eller færre løgnere i utvalget.

### Fasit versjon 6:

- a) Hendelser  $A$  "Idunn ønsker å se Gutta på Black Box" og  $B$  "Idunn føler seg som Frp-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,12 \cdot 0,90}{0,12 \cdot 0,90 + 0,72 \cdot 0,10} = 0,60$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 9,76$   $\text{Var}(X) = 8,57$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[4,02, 15,50]$ .

- c)  $Y$  er binomisk fordelt med  $n = 80$  og  $p = 0,02$ . Varians lik 1,568, så  $Y$  **ikke tilnærmet normalfordelt**. Med punktsannsynligheter:  $P(Y = 0) = 0,1986$ ,  $P(Y = 1) = 0,3243$ ,  $P(Y = 2) = 0,2614$ ,  $P(Y = 3) = 0,1387$   $P(Y \leq 3) = 0,9231$ .

### Oppgavetekst versjon 7:

- a) Nathalie har et uavklart forhold til sin identitet som Høyrevelger. Noen ganger føler hun at hun er Høyrevelger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Nathalie disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Nathalie føler seg som Høyrevelger er 60%.

Gitt at Nathalie føler seg som Høyrevelger, så er det en sannsynlighet på 90% for at hun tror at Erna Solberg ville gjort en god jobb som megler for en handelsavtale mellom Storbritannia og EU.

I motsatt fall, når Nathalie ikke føler seg som Høyrevelger, så er det en sannsynlighet på 27% for at hun tror at Erna Solberg ville gjort en god jobb som megler for en handelsavtale mellom Storbritannia og EU.

Hva er den subjektive sannsynligheten for at Nathalie føler at hun er Høyrevelger gitt at hun er sikker på at Erna Solberg ville gjort en god jobb som megler for en handelsavtale mellom Storbritannia og EU?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 888 222 Høyre-velgere. Anta at det gjøres et utvalg på 5000 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Høyre-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Høyre-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,10$ . La  $Y$  være antall løgnere i utvalget på  $n = 5000$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 450 eller færre løgnere i utvalget.

### Fasit versjon 7:

- a) Hendelser  $A$  "Nathalie tror at Erna Solberg ville være en god megler for en handelsavtale mellom Storbritannia og EU" og  $B$  "Nathalie føler seg som Høyre-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,90 \cdot 0,90}{0,60 \cdot 0,90 + 0,27 \cdot 0,10} = 0,8333$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 1147,8$   $\text{Var}(X) = 883,2$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[1089,6, 1206,1]$ .
- c)  $Y$  er binomisk fordelt med  $n = 5000$  og  $p = 0,10$ . Variansen er 450, så  $Y$  **tilnærmet normalfordelt**.  $P(Y \leq 450) \approx 0,0092$  eller med heltallskorreksjon:  $P(Y \leq 450) \approx 0,0098$ .

### Oppgavetekst versjon 8:

- a) Frank har et uavklart forhold til sin identitet som SV-velger. Noen ganger føler han at han er SV-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Frank disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Frank føler seg som SV-velger er 65%.

Gitt at Frank føler seg som SV-velger, så er det en sannsynlighet på 96% for at han synes at helte dyrkelsen av kongehuset i NRK-serien Atlantic Crossing er historieforfalskning.

I motsatt fall, når Frank ikke føler seg som SV-velger, så er det en sannsynlighet på 36% for at han synes at helte dyrkelsen av kongehuset i NRK-serien Atlantic Crossing er historieforfalskning.

Hva er den subjektive sannsynligheten for at Frank føler at han er SV-velger gitt at han har kommet til det standpunktet at helte dyrkelsen av kongehuset i NRK-serien Atlantic Crossing er historieforfalskning?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 278 900 SV-velgere. Anta at det gjøres et utvalg på 250 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er SV-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet SV-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,01$ . La  $Y$  være antall løgnere i utvalget på  $n = 250$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 2 eller færre løgnere i utvalget.

### Fasit versjon 8:

- a) Hendelser  $A$  "Frank synes at Atlantic Crossing er historieforfalskning" og  $B$  "Frank føler seg som SV-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,96 \cdot 0,65}{0,96 \cdot 0,65 + 0,36 \cdot 0,35} = 0,832$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 18,02$   $\text{Var}(X) = 16,72$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[10,01, 26,03]$ .

- c)  $Y$  er binomisk fordelt med  $n = 250$  og  $p = 0,01$ . Varians lik 2,475, så  $Y$  **ikke tilnærmet normalfordelt**. Med punktsannsynligheter:  $P(Y = 0) = 0,08106$ ,  $P(Y = 1) = 0,2047$ ,  $P(Y = 2) = 0,2574$ ,  $P(Y \leq 2) = 0,5432$ .

### Oppgavetekst versjon 9:

- a) Marianne har et uavklart forhold til sin identitet som Rødt-velger. Noen ganger føler hun at hun er Rødt-velger, andre ganger ikke. For å beskrive sine følelser om parti-identitet bruker Marianne disse subjektive sannsynlighetene:

Den subjektive sannsynligheten for at Marianne føler seg som Rødt-velger er 95%.

Gitt at Marianne føler seg som Rødt-velger, så er det en sannsynlighet på 96% for at hun synes at leksefri skole er en god idé.

I motsatt fall, nå Marianne ikke føler seg som Rødt-velger, så er det en sannsynlighet på 51% for at hun synes at leksefri skole er en god idé.

Hva er den subjektive sannsynligheten for at Marianne føler at hun er Rødt-velger gitt at hun har kommet fram til at leksefri skole er en god idé?

- b) Det er i dag 3 869 200 personer med stemmerett i Norge. Av disse er det 155 078 Rødt-velgere. Anta at det gjøres et utvalg på 1500 personer som intervjues til en spørreundersøkelse. La  $X$  være antallet i utvalget som er Rødt-velgere.

Hvilken fordeling har  $X$ ? Finn forventning  $E(X)$  og varians  $\text{Var}(X)$ . Er  $X$  tilnærmet normalfordelt? Angi et 95%-spredningsintervall for antallet Rødt-velgere i utvalget.

- c) Spørreundersøkelser blir aldri helt nøyaktige fordi endel personer i utvalget lyver under intervjuet. Anta at sannsynligheten for at en tilfeldig person lyver er  $p = 0,20$ . La  $Y$  være antall løgnere i utvalget på  $n = 1500$  personer.

Hvilken fordeling har den stokastiske variabelen  $Y$ ? Er  $Y$  tilnærmet normalfordelt? Finn en tilnærmet verdi for sannsynligheten for at det er 333 eller færre løgnere i utvalget.

### Fasit versjon 9:

- a) Hendelser  $A$  "Marianne synes leksefri skole er en god idé" og  $B$  "Marianne føler seg som Rødt-velger". Ved Bayes' regel og total sannsynlighet

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0,96 \cdot 0,95}{0,96 \cdot 0,95 + 0,51 \cdot 0,05} = 0,9728$$

- b)  $X$  er hypergeometrisk fordelt.  $E(X) = 60,12$   $\text{Var}(X) = 57,69$ . Ja,  $X$  er tilnærmet normalfordelt. Et 95% spredningsintervall er  $[45,23, 75,01]$ .
- c)  $Y$  er binomisk fordelt med  $n = 1500$  og  $p = 0,20$ . Variansen er 240, så  $Y$  **tilnærmet normalfordelt**.  $P(Y \leq 333) \approx 0,9834$  eller med heltallskorreksjon:  $P(Y \leq 333) \approx 0,9847$ .

## Oppgave 2. Oppgavetekst versjon 1:

Henrik gambler gjerne med sine venner. Denne gangen har han planlagt å gjøre to veddemål samtidig, for han mener at sannsynligheten for å ha uflaks to ganger er veldig liten, og at han dermed kan forvente en fortjeneste på veddemålene sett under ett.

Mot Une vedder Henrik på at David Beckham blir den neste James Bond skuespilleren. Mot Hassen vedder Henrik på at David Beckham er manager for Salford City i 2021. La  $X$  være fortjenesten i veddemålet mot Une og la  $Y$  være fortjenesten i veddemålet mot Hassen. Dersom Daniel Craig fremdeles spiller James Bond i neste film, er Une og Henrik enige om å dele den første potten.

- a) Henrik har innsideinformasjon via sine kontakter i Spice Girls og har funnet følgende sannsynligheter for utfallene av veddemålene:

	$x$	-100	0	100	
$y$					$P(Y = y)$
-100		?	0,02	0,09	0,21
100		?	0,08	0,01	?
	$P(X = x)$	0,8	0,1	0,1	1,0

Hjelp Henrik å fylle ut resten av tabellen. Hva sannsynligheten for at David Beckham er manager for Salford City i 2021 gitt at han er den neste som spiller James Bond? Finn forventningene  $E(X)$ ,  $E(Y)$  og  $E(X + Y)$ .

- b) Forklar hvordan man kan regne ut korrelasjonen mellom  $X$  og  $Y$ . Du trenger ikke gjøre utregningen av korrelasjonen, men du skal fortelle hvilke størrelser man regner ut i prosessen. Tror du korrelasjonen kommer til å være positiv, null eller negativ i dette tilfellet?

Henrik spekulerer også i verdipapirer. Han har investert 90 000 kroner i fotballaget Juventus og 20 000 kroner i aksjer i Netflix. Anta at verdien av disse investeringene ikke er uavhengig, men har en ukjent korrelasjon.

- c) Anta at standardavviket på investeringen i Juventus er  $\sigma_1 = 40\,000$  og at standardavviket på investeringen i Netflix er  $\sigma_2 = 5\,000$ . Hva er den laveste mulige verdien for standardavviket til summen av de to investeringene? Begrunn svaret.

## Løsningsforslag versjon 1:

- a) Tabellen er en simultanfordeling. Tallene i margin er summen av radene eller kolonnene. Vi finner  $P(Y = 100)$  først. Bruker at summen av kolonnen i margin blir 1. Da er  $0,21 + P(Y = 100) = 1,0$ , som gir  $\underline{P(Y = 100) = 0,79}$ . Punktsannsynlighetene  $P(-100, -100)$  og  $P(-100, 100)$  kan vi finne på tilsvarende måte. Vi ser på summen av radene  $P(-100, -100) + 0,02 + 0,09 = 0,21$  gir  $\underline{P(-100, -100) = 0,10}$ . Videre gir  $P(-100, 100) + 0,08 + 0,01 = 0,79$  at  $\underline{P(-100, 100) = 0,70}$ . Fullstendig tabell blir

	$x$	-100	0	100	
$y$					$P(Y = y)$
-100		0,10	0,02	0,09	0,21
100		0,70	0,08	0,01	0,79
	$P(X = x)$	0,8	0,1	0,1	1,0

Den betingede sansynligheten for at David Beckham er manager for Salford City i 2021 gitt at David Beckham blir neste James Bond er

$$P(Y = 100 | X = 100) = \frac{P(100, 100)}{P(X = 100)} = \frac{0,01}{0,1} = \underline{\underline{0,1}}.$$

Vi finner forventningene ved sum av verdi ganger sannsynlighet:

$$E(X) = -100 \cdot 0,8 + 0 \cdot 0,1 + 100 \cdot 0,1 = \underline{\underline{-60}}$$

$$E(Y) = -100 \cdot 0,21 + 100 \cdot 0,79 = \underline{\underline{58}}$$

Forventningen til summen av veddemålene er

$$E(X + Y) = E(X) + E(Y) = -60 + 58 = \underline{\underline{-2}}.$$

b) For å finne korrelasjonen  $\rho$  mellom  $X$  og  $Y$  gjelder definisjonen ved formelen

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

Til utregningen trenger man

- Forventningen  $\mu_X = E(X)$  til  $X$  regnet ut som vist over.
- Variansen til  $X$  regnet ut f.eks. som  $\text{Var}(X) = E(X^2) - \mu_X^2$ .
- Standardavviket  $\sigma_X$  til  $X$  regnet ut som kvadratroten av varians.
- Forventningen  $\mu_Y = E(Y)$  til  $Y$  regnet ut som vist over.
- Variansen til  $Y$  regnet ut f.eks. som  $\text{Var}(Y) = E(Y^2) - \mu_Y^2$ .
- Standardavviket  $\sigma_Y$  til  $Y$  regnet ut som kvadratroten av varians.
- Kovariansen mellom  $X$  og  $Y$  regnet ut f.eks. som  $\text{Cov}(X, Y) = E(X \cdot Y) - \mu_X \cdot \mu_Y$ .

Siden veddemålene i stor grad vil få **motsatte resultater** av hverandre; sannsynlighetene 0,70 og 0,09 utgjør mesteparten av 100%, så er det grunn til å tro at korrelasjonen vil være negativ.

c) Variansen til en sum av to variabler er

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y).$$

La  $X$  være investeringen i Juventus og la  $Y$  være investeringen i Netflix. Det minste standardavviket til summen av investeringene inntreffer når  $\text{Var}(X + Y)$  er minst mulig. Variansene til  $X$  og  $Y$  er konstante

$$\text{Var}(X) = \sigma_1^2 = (40\,000)^2 = 1,6 \cdot 10^9$$

$$\text{Var}(Y) = \sigma_2^2 = (5\,000)^2 = 2,5 \cdot 10^7$$

Den laveste mulige kovariansen innreffer når korrelasjonen er  $\rho = -1$ . Da blir

$$\text{Cov}(X, Y) = \rho \cdot \sigma_1 \cdot \sigma_2 = -1 \cdot 40\,000 \cdot 5\,000 = -2 \cdot 10^8.$$

Minst mulige varians til summen blir

$$\text{Var}(X + Y) = 1,6 \cdot 10^9 - 2 \cdot 2 \cdot 10^8 + 2,5 \cdot 10^7 = 1,225 \cdot 10^9.$$

Minste mulige standardavvik blir

$$\sigma_{\text{sum}} = \sqrt{\text{Var}(X + Y)} = \underline{\underline{35\,000}}.$$

Et mer uformelt resonement er at man har minimalt standardavvik på summen av investeringene når enhver fluktusjon i den ene investeringen gir en motsatt endring i den andre. Da vil standardavviket til total investering være differansen mellom standardavvikene, så

$$\sigma_{\text{sum}} = \sigma_1 - \sigma_2 = 40\,000 - 5\,000 = \underline{\underline{35\,000}}.$$

I dette argumentet er bruken av differase dårlig begrunnet. Om man skal begrunne denne formelen, så må man igjen se på formelen for varians av en sum, altså

$$\text{Var}(X + Y) = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2.$$

Poenget er igjen å bruke minimal korrelasjon på  $\rho = -1$ , så  $\text{Var}(X + Y) = (\sigma_1 - \sigma_2)^2$  ved andre kvadratsetning.

**Oppgavetekst versjon 2:**

Henrik gambler gjerne med sine venner. Denne gangen har han planlagt å gjøre to veddemål samtidig, for han mener at sannsynligheten for å ha uflaks to ganger er veldig liten, og at han dermed kan forvente en fortjeneste på veddemålene sett under ett.

Mot Fadi vedder Henrik på at Magnus Carlsen vinner neste sjakk-VM. Mot Ixi vedder Henrik på at Magnus Carlsen er Vikingen i Maskorama. La  $X$  være fortjenesten i veddemålet mot Fadi og la  $Y$  være fortjenesten i veddemålet mot Ixi. Dersom Magnus Carlsen trekker seg fra neste sjakk-VM, er Fadi og Henrik enige om å dele den første potten.

- a) Henrik har innsideinformasjon via sine kontakter i Fredriksstad Schakselskap og har funnet følgende sannsynligheter for utfallene av veddemålene:

$y$	$x$	-100	0	100	$P(Y = y)$
-100		0,12	0,06	0,42	?
100		?	0,04	?	0,4
	$P(X = x)$	0,2	0,1	0,7	1,0

Hjelp Henrik å fylle ut resten av tabellen. Hva sannsynligheten for at Magnus Carlsen trekker seg fra neste sjakk-VM gitt at han er Vikingen i Maskorama? Finn forventningene  $E(X)$ ,  $E(Y)$  og  $E(X + Y)$ .

- b) Forklar hvordan man kan regne ut korrelasjonen mellom  $X$  og  $Y$ . Du trenger ikke gjøre utregningen av korrelasjonen, men du skal fortelle hvilke størrelser man regner ut i prosessen. Tror du korrelasjonen kommer til å være positiv, null eller negativ i dette tilfellet?

Henrik spekulerer også i verdipapirer. Han har investert 50 000 kroner i Play Magnus appen og 120 000 kroner i flyselskapet Norwegian. Anta at verdien av disse investeringene ikke er uavhengig, men har en ukjent korrelasjon.

- c) Anta at standardavviket på investeringen i Play Magnus er  $\sigma_1 = 10\,000$  og at standardavviket på investeringen i Norwegian er  $\sigma_2 = 70\,000$ . Hva er den laveste mulige verdien for standardavviket til summen av de to investeringene? Begrunn svaret.

**Fasit versjon 2:**

- a) Tabell blir

$y$	$x$	-100	0	100	$P(Y = y)$
-100		0,12	0,06	0,42	0,6
100		0,08	0,04	0,28	0,4
	$P(X = x)$	0,2	0,1	0,7	1,0

Den betingede sannsynligheten er

$$P(X = 0 | Y = 100) = \frac{P(0, 100)}{P(Y = 100)} = \frac{0,04}{0,4} = \underline{\underline{0,1}}.$$



$$E(X) = 50, E(Y) = -20, E(X + Y) = 30$$

- b) Utregning av  $\rho$  ved forventning, varians og standardavvik for  $X$  og  $Y$ . Deretter kovarians mellom  $X$  og  $Y$  ved definisjonen og korrelasjon ved kovarians delt på standardavvikene. **I dette tilfellet er  $X$  og  $Y$  uavhengige.** Man ser at tallene i tabellen er produkter av tallene i margen. Dermed vil korrelasjonen være  $\rho = 0$ .
- c) Minimalt standardavvik er differansen:  $\sigma_{\text{minimal}} = \sigma_2 - \sigma_1 = 60\,000$ .

### Oppgavetekst versjon 3:

Henrik gambler gjerne med sine venner. Denne gangen har han planlagt å gjøre to veddemål samtidig, for han mener at sannsynligheten for å ha uflaks to ganger er veldig liten, og at han dermed kan forvente en fortjeneste på veddemålene sett under ett.

Mot Helge vedder Henrik på at Espen Nakstad blir utnevnt til helseminister etter Stortingsvalget høsten 2021. Mot Idunn vedder Henrik på at Espen Nakstad skal krysse Uralfjellene sammen med Lars Monsen etter koronavaksinasjonen er fullført. La  $X$  være fortjenesten i veddemålet mot Helge og la  $Y$  være fortjenesten i veddemålet mot Idunn. Dersom Bent Høie fremdeles er helseminister etter Stortingsvalget 2010, er Helge og Henrik enige om å dele den første potten.

- a) Henrik har innsideinformasjon via sine kontakter i NRK og har funnet følgende sannsynligheter for utfallene av veddemålene:

$y$	$x$	-100	0	100	$P(Y = y)$
-100		0,08	0,06	0,39	0,53
100		?	0,18	?	0,47
	$P(X = x)$	0,36	0,24	?	1,0

Hjelp Henrik å fylle ut resten av tabellen. Hva sannsynligheten for at Espen Nakstad blir med Lars Monsen over Uralfjellene gitt at han blir den neste helseministeren? Finn forventningene  $E(X)$ ,  $E(Y)$  og  $E(X + Y)$ .

- b) Forklar hvordan man kan regne ut korrelasjonen mellom  $X$  og  $Y$ . Du trenger ikke gjøre utregningen av korrelasjonen, men du skal fortelle hvilke størrelser man regner ut i prosessen. Tror du korrelasjonen kommer til å være positiv, null eller negativ i dette tilfellet?

Henrik spekulerer også i verdipapirer. Han har investert 200 000 kroner i BioNTech aksjer og 80 000 kroner i nettbutikken Love Norway. Anta at verdien av disse investeringene ikke er uavhengig, men har en ukjent korrelasjon.

- c) Anta at standardavviket på investeringen i BioNTech er  $\sigma_1 = 60\,000$  og at standardavviket på investeringen i Love Norway er  $\sigma_2 = 40\,000$ . Hva er den laveste mulige verdien for standardavviket til summen av de to investeringene? Begrunn svaret.

### Fasit versjon 3:

- a) Tabell blir

$y$	$x$	-100	0	100	$P(Y = y)$
-100		0,08	0,06	0,39	0,53
100		0,28	0,18	0,01	0,47
	$P(X = x)$	0,36	0,24	0,40	1,0

Den betingede sannsynligheten er

$$P(Y = 100 | X = 100) = \frac{P(100, 100)}{P(X = 100)} = \frac{0,01}{0,40} = \underline{\underline{0,025}}.$$

$$E(X) = 4, E(Y) = -6, E(X + Y) = -2$$

- b) Utregning av  $\rho$  ved forventning, varians og standardavvik for  $X$  og  $Y$ . Deretter kovarians mellom  $X$  og  $Y$  ved definisjonen og korrelasjon ved kovarians delt på standardavvikene. **Veddemålene utligner hverandre.** Man ser at tallene øverste høyre hjørne og nederste venstre hjørne er store. Dermed kan man gå ut fra at korrelasjonen vil være negativ  $\rho < 0$ .
- c) Minimalt standardavvik er differansen:  $\sigma_{\text{minimal}} = \sigma_1 - \sigma_2 = 20\,000$ .

**Oppgavetekst versjon 4:**

Henrik gambler gjerne med sine venner. Denne gangen har han planlagt å gjøre to veddemål samtidig, for han mener at sannsynligheten for å ha uflaks to ganger er veldig liten, og at han dermed kan forvente en fortjeneste på veddemålene sett under ett.

Mot Nathalie vedder Henrik på at Ståle Solbakken blir den neste treneren for Norges fotballandslag. Mot Frank vedder Henrik på at Ståle Solbakken skal delta i Mestrenes Mester 2022. La  $X$  være fortjenesten i veddemålet mot Nathalie og la  $Y$  være fortjenesten i veddemålet mot Frank. Dersom Lars Lagerbäck forlenger kontrakten med Norges fotballandslag, er Nathalie og Henrik enige om å dele den første potten.

- a) Henrik har innsideinformasjon fra Egil Drillo Olsen og har funnet følgende sannsynligheter for utfallene av veddemålene:

	$x$	-100	0	100	
$y$					$P(Y = y)$
-100		?	?	0,76	0,8
100		0,12	0,04	0,04	0,2
	$P(X = x)$	0,15	?	0,8	1,0

Hjelp Henrik å fylle ut resten av tabellen. Hva sannsynligheten for at Ståle Solbakken deltar i Mestrenes Mester 2022 gitt at han blir ny landslagstrener i fotball? Finn forventningene  $E(X)$ ,  $E(Y)$  og  $E(X + Y)$ .

- b) Forklar hvordan man kan regne ut korrelasjonen mellom  $X$  og  $Y$ . Du trenger ikke gjøre utregningen av korrelasjonen, men du skal fortelle hvilke størrelser man regner ut i prosessen. Tror du korrelasjonen kommer til å være positiv, null eller negativ i dette tilfellet?

Henrik spekulerer også i verdipapirer. Han har investert 50 000 kroner i restauranten Noma i København og 80 000 kroner i aksjer i sportsutstyrsjeden XXL. Anta at verdien av disse investeringene ikke er uavhengig, men har en ukjent korrelasjon.

- c) Anta at standardavviket på investeringen i Noma er  $\sigma_1 = 10\,000$  og at standardavviket på investeringen i XXL er  $\sigma_2 = 15\,000$ . Hva er den laveste mulige verdien for standardavviket til summen av de to investeringene? Begrunn svaret.

**Fasit versjon 4:**

- a) Tabell blir

	$x$	-100	0	100	
$y$					$P(Y = y)$
-100		0,03	0,01	0,76	0,8
100		0,12	0,04	0,04	0,2
	$P(X = x)$	0,15	0,05	0,8	1,0

Den betingede sannsynligheten er

$$P(Y = 100 | X = 100) = \frac{P(100, 100)}{P(X = 100)} = \frac{0,04}{0,8} = \underline{\underline{0,05}}.$$

$$E(X) = 65, E(Y) = -60, E(X + Y) = 5$$

- b) Utregning av  $\rho$  ved forventning, varians og standardavvik for  $X$  og  $Y$ . Deretter kovarians mellom  $X$  og  $Y$  ved definisjonen og korrelasjon ved kovarians delt på standardavvikene. **Veddemålene utligner hverandre.** Man ser at tallene øverste høyre hjørne og nederste venstre hjørne er store. Dermed kan man gå ut fra at korrelasjonen vil være negativ  $\rho < 0$ .
- c) Minimalt standardavvik er differansen:  $\sigma_{\text{minimal}} = \sigma_2 - \sigma_1 = 5\,000$ .

### Oppgave 3. Oppgavetekst versjon 1:

Per Erik har satt opp et eksperiment for å måle bølgelengden i emisjonsspektra til gasser. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Per Erik antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Per Erik en ukjent gass, som kanskje kan være klorgass. Det er kjent at klorgass har en sterk spektrallinje på bølgelengde 452,62 nm. Per Erik gjør 5 målinger av en sterk blå spektrallinje fra den ukjente gassen og får disse resultatene målt i nm:

453,58	452,85	453,07	451,50	499,70
--------	--------	--------	--------	--------

- Finne gjennomsnitt og standardavvik for målingene.
- Gjøre en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 452,62.

$H_1$  : Forventningsverdien til målingene er ulik 452,62.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Per Erik ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 99%-konfidensintervall for bølgelengden til den blå spektrallinja. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn 1 nm. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Per Erik bør gjøre.

### Løsningsforslag versjon 1:

- Vi finner gjennomsnitt ved utregning:

$$\bar{x} = \frac{1}{5} (453,58 + 452,85 + 453,07 + 451,50 + 499,70) = \underline{\underline{452,14}}$$

Variansen regnes ut som

$$s^2 = \frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2 = \frac{1}{4} ((1,44)^2 + (0,71)^2 + (0,93)^2 + (-0,64)^2 + (-2,44)^2) = 2,45145$$

Standardavviket finner vi som kvadratroten av variansen

$$\underline{\underline{s = 1,5657.}}$$

Disse utregningene kan man gjerne gjøre i Excel eller lignende, og da er det fint om man legger ved utklipp fra regnearket som dokumentasjon på utregningen.

- Per Erik har antatt at målingene er normalfordelte. Han vet ikke hva standardavviket er. Han sammenligner forventningen med en oppgitt verdi  $\mu_0 = 452,62$ . Derfor

bør vi velge en T-test. Hypotesene er  $H_0 : \mu = 452,62$  og  $H_1 : \mu \neq 452,62$ . Testen er derfor en tosidig T-test. Testobservator er

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Vi skal forkaste nullhypotesen om tallverdien til  $T$  overstiger kritisk verdi  $t_{\alpha/2}$ . Signifikansnivå er oppgitt til  $\alpha = 0,05$ . Vi finner kritisk verdi i tabellen for Student T-fordelingen. Vi har  $n - 1 = 4$  frihetsgrader og da er

$$t_{0,025} = 2,776.$$

Vi regner nå ut observert verdi av  $T$ . Setter inn  $\bar{x} = 452,14$ ,  $s = 1,5657$ ,  $\mu_0 = 452,62$  og  $n = 5$ . Det gir

$$t = \frac{452,14 - 452,62}{1,5657/\sqrt{5}} = -0,68551146$$

Tallverdien av observert testobservator er mindre enn kritisk verdi. Vi beholder nullhypotesen. Konklusjon: Hypotesetesten viste ikke signifikant avvik mellom Per Eriks målinger og den kjente blå spekrallinjen i klograss. Den kan være klograss.

- c) Per Erik er villig til å gjøre mer en 30 målinger. Da kan han ved sentralgrenseteoremet benytte et  $Z$ -intervall for å bestemme forventningsverdien mer presis. Man kan anslå antall målinger ved formelen

$$n \geq \left( \frac{2 \cdot z_{\alpha/2} \cdot \sigma}{L} \right)^2.$$

På signifikansnivå  $\alpha = 0,01$ , som tilsvarer 99%-konfidensintervall, bruker vi  $z_{\alpha/2} = 2,576$ . Vi har oppgitt  $L = 1$ . Videre antar vi at standardavviket er tilnærmet lik det målte standardavviket  $\sigma \approx 1,57$ . Det gir

$$n \geq \left( \frac{2 \cdot 2,576 \cdot 1,57}{1} \right)^2 = 65,4.$$

Det burde holde med 66 målinger for å få et 99%-konfidensintervall med lengde mindre enn 1 nm.

### Oppgavetekst versjon 2:

Mikjel har satt opp et eksperiment for å måle radioaktivitet i gjenstander. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Mikjel antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Mikjel en gammel ventil fra reaktoren i Halden, som kanskje kan være radioaktiv. Dersom den spesifikke radioaktiviteten er mer enn 100 Bq/g ønsker Mikjel og Anna-Lena snarest mulig å avhende seg med ventilen ved å levere den inn til et godkjent mottak for radioaktivt avfall. Mikjel gjør 5 målinger av den spesifikke radioaktiviteten fra ventilen målt i Bq/g:

101,1	97,7	98,9	95,3	97,5
-------	------	------	------	------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er over 100.

$H_1$  : Forventningsverdien til målingene er under 100.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Mikjel ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 95%-konfidensintervall for spesifikk radioaktivitet i ventilen. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn 2 Bq/g. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Mikjel bør gjøre.

### Fasit versjon 2:

- $\bar{x} = 98,1$  og  $s = 2,1213$ .
- Ensidig  $T$ -test med  $t_\alpha = 2,132$  og observert verdi for testobservator  $t = -2,003$   
Behold nullhypotese.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 2,12$  burde 17 målinger gi den ønskede sikkerheten. Men Mikjel bør minst bruke 30 målinger likevel for å kunne bruke sentralgrenseteoremet.



### Oppgavetekst versjon 3:

Jo har satt opp et eksperiment for å måle spesifikk varmekapasitet til metall. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Jo antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Jo en bit ukjent metall, som kanskje kan være titan. Det er kjent at titan har en spesifikk varmekapasitet på  $523 \text{ J}/(\text{kg} \cdot \text{K})$ . Jo gjør 5 målinger av den spesifikke varmekapasiteten til den ukjente metallbiten og får disse resultatene målt i  $\text{J}/(\text{kg} \cdot \text{K})$ :

523	504	520	510	506
-----	-----	-----	-----	-----

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 523.

$H_1$  : Forventningsverdien til målingene er ulik 523.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Jo ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 95%-konfidensintervall for spesifikk varmekapasitet til den ukjente metallbiten. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn  $2 \text{ J}/(\text{kg} \cdot \text{K})$ . Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Jo bør gjøre.

### Fasit versjon 3:

- $\bar{x} = 512,6$  og  $s = 8,4734$ .
- Tosidig  $T$ -test med  $t_{\alpha/2} = 2,776$  og observert verdi for testobservator  $t = -2,744$ . Behold nullhypotese.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 8,5$  burde 278 målinger gi den ønskede sikkerheten.

**Oppgavetekst versjon 4:**

Tore har laget et instrument for å måle diameteren til kuler. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Tore antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Tore en liten ball laget av et nyutviklet materiale. Kanskje denne ballen kan brukes til å spille golf? Ifølge golfens regler skal golfballer ha en diameter ikke mindre enn 42,67 mm. Tore gjør 5 målinger av diameteren til ballen og får disse resultatene målt i mm:

42,76	43,31	43,15	43,32	43,71
-------	-------	-------	-------	-------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er mindre enn 42,67.

$H_1$  : Forventningsverdien til målingene er ikke mindre enn 42,67.

Bruk signifikansnivå  $\alpha = 0,01$ . Hva blir konklusjonen?

- Tore ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 99%-konfidensintervall for diameteren til ballen. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn 0,1 mm. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Tore bør gjøre.

**Fasit versjon 4:**

- $\bar{x} = 43,25$  og  $s = 0,3429$ .
- Ensidig  $T$ -test med  $t_\alpha = 3,747$  og observert verdi for testobservator  $t = 3,783$ . Forkast nullhypotese. Golfballen har diameter signifikant større enn 42,67 mm.
- På signifikansnivå  $\alpha = 0,01$  og anslått verdi for standardavvik  $\sigma \approx 0,34$  burde 307 målinger gi den ønskede sikkerheten.

**Oppgavetekst versjon 5:**

Kent har satt opp et eksperiment for å måle frekvensen på lyd. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Kent antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Kent en stemmegaffel, som lager en klar tone. Det er vanlig å stemme instrumenter ut fra enstrøken A på 440,0 Hz. Kent gjør 5 målinger av en tonen fra stemmegaffelen og får disse resultatene målt i Hz:

440,04	440,47	440,20	440,44	440,25
--------	--------	--------	--------	--------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 440,0.

$H_1$  : Forventningsverdien til målingene er ulik 440,0.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Kent ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 95%-konfidensintervall for frekvensen fra stemmegaffelen. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn 0,1 Hz. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Kent bør gjøre.

**Fasit versjon 5:**

- $\bar{x} = 440,28$  og  $s = 0,1779$ .
- Tosidig  $T$ -test med  $t_{\alpha/2} = 2,776$  og observert verdi for testobservator  $t = 3,519$ . Forkast nullhypotese. Stemmegaffelen har en annen frekvens enn 440 Hz.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 0,18$  burde 50 målinger gi den ønskede sikkerheten.

**Oppgavetekst versjon 6:**

Magnus har laget sitt eget Voltmeter. Dessverre er hans selvlagde måleinstrument noe unøyaktig. Magnus antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har funnet et svært gammelt batteri, og hun vil gjerne at Magnus måler polspenningen. Miljøskadelige kvikksølvbatterier har en spenning på 1,35 V. Kanskje det er et slikt kvikksølvbatteri? Magnus gjør 5 målinger av polspenningen og får disse resultatene målt i V:

1,353	1,419	1,358	1,342	1,353
-------	-------	-------	-------	-------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 1,35.

$H_1$  : Forventningsverdien til målingene er ulik 1,35.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Magnus ønsker å anslå hvor mange målinger han bør gjøre for å oppnå et mer sikkert resultat. Han er villig til å gjøre mer enn 30 målinger. Han ønsker å finne et 95%-konfidensintervall for polspenningen på batteriet. Dette konfidensintervallet ønsker han at skal ha en lengde mindre enn 0,01 V. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Magnus bør gjøre.

**Fasit versjon 6:**

- $\bar{x} = 1,365$  og  $s = 0,030749$ .
- Tosidig  $T$ -test med  $t_{\alpha/2} = 2,776$  og observert verdi for testobservator  $t = 1,091$ . Behold nullhypotese.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 0,03$  burde 139 målinger gi den ønskede sikkerheten.

**Oppgavetekst versjon 7:**

Beathe har satt opp et eksperiment for å måle massen salt av oppløst i vann. Dessverre gir hennes selvlagde laboratorieoppsett noe unøyaktige avlesninger. Beathe antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Beathe en mettet oppløsning, og de ønsker å bestemme hvor mye salt oppløsningen inneholder. Det er kjent at en mettet natriumnitratoppløsning 25 °C inneholder 91,2 g pr desiliter. Beathe gjør 5 målinger av den mettede oppløsningen og får disse resultatene målt i g pr desiliter:

90,9	93,6	92,5	84,1	93,4
------	------	------	------	------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 91,2.

$H_1$  : Forventningsverdien til målingene er ulik 91,2.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Beathe ønsker å anslå hvor mange målinger hun bør gjøre for å oppnå et mer sikkert resultat. Hun er villig til å gjøre mer enn 30 målinger. Hun ønsker å finne et 95%-konfidensintervall for oppløst masse av salt. Dette konfidensintervallet ønsker hun at skal ha en lengde mindre enn 1,0 g pr desiliter. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Beathe bør gjøre.

**Fasit versjon 7:**

- $\bar{x} = 90,9$  og  $s = 3,9478$ .
- Tosidig  $T$ -test med  $t_{\alpha/2} = 2,776$  og observert verdi for testobservator  $t = -0,170$ . Behold nullhypotese.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 4$  burde 246 målinger gi den ønskede sikkerheten.

**Oppgavetekst versjon 8:**

Birte har satt opp et eksperiment for å måle molar masse til proteiner. Dessverre er hennes selvlagde måleoppsett noe unøyaktig. Birte antar at hver enkeltmåling har en feil som er tilfeldig og normalfordelt.

Anna-Lena har gitt Birte et ukjent protein, som kanskje kan være insulin. Det er kjent at insulin har en molar masse på 5793,6 g/mol. Birte gjør 5 målinger av en proteinet og får disse resultatene målt i g/mol:

5752,8	5732,8	5896,7	5778,7	5692,5
--------	--------	--------	--------	--------

- Finn gjennomsnitt og standardavvik for målingene.
- Gjør en hypotesetest der hypotesene er:

$H_0$  : Forventningsverdien til målingene er lik 5793,6.

$H_1$  : Forventningsverdien til målingene er ulik 5793,6.

Bruk signifikansnivå  $\alpha = 0,05$ . Hva blir konklusjonen?

- Birte ønsker å anslå hvor mange målinger hun bør gjøre for å oppnå et mer sikkert resultat. Hun er villig til å gjøre mer enn 30 målinger. Hun ønsker å finne et 95%-konfidensintervall for den molare massen. Dette konfidensintervallet ønsker hun at skal ha en lengde mindre enn 50 g/mol. Bruk de foreløpige målingene til å gjette en rimelig verdi for standardavviket, og anslå deretter hvor mange målinger Birte bør gjøre.

**Fasit versjon 8:**

- $\bar{x} = 5770,7$  og  $s = 77,156$ .
- Tosidig  $T$ -test med  $t_{\alpha/2} = 2,776$  og observert verdi for testobservator  $t = -0,664$ . Behold nullhypotese.
- På signifikansnivå  $\alpha = 0,05$  og anslått verdi for standardavvik  $\sigma \approx 80$  burde 40 målinger gi den ønskede sikkerheten.

#### Oppgave 4. Oppgavetekst versjon 1:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Eliteserien i Norge før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Eliteserien i Norge er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 16.juni til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	88
Uavgjort $X_1 = 1$	43
Borteseier $X_1 = 0$	61

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,5990$  og  $S_1 \approx 1,3423$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Eliteserien 2019. Dette datasettet inneholder alle kamper spilt fra og med 30.mars til og med 1.desember 2019.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	113
Uavgjort $X_2 = 1$	73
Borteseier $X_2 = 0$	54

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Eliteserien før	Eliteserien etter
Gjennomsnitt	1.716666667	1.598958333
Varians	1.6013947	1.801674302
Observasjoner	240	192
Gruppevarians	1.690356105	
Antatt avvik mellom gjennomsnittene	0	
fg	430	
t-Stat	0.935044907	
P(T<=t) ensidig	0.175144977	
T-kritisk, ensidig	1.648404969	
P(T<=t) tosidig	0.350289953	
T-kritisk, tosidig	1.965496192	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Eliteserien er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

### Løsningsforslag versjon 1:

- a) Datasettet består av 88 kopier av 3-ere, 43 kopier av 1-ere og 61 repeterte 0-ere. Gjennomsnittet blir

$$\bar{x} = \frac{1}{88 + 43 + 61} (88 \cdot 3 + 43 \cdot 1 + 61 \cdot 0) \approx 1,5990.$$

Variansen regnes også ved å gange opp kvadratavvikene etter antall kopier:

$$s^2 = \frac{1}{88 + 43 + 61 - 1} (88 \cdot (3 - 1,5990)^2 + 43 \cdot (1 - 1,5990)^2 + 61 \cdot (0 - 1,5990)^2) \approx 1,8017$$

Ved kvadratrot blir standardavviket  $s \approx 1,3423$ .

- b) I hypotesetesten vil man sammenligne forventningsverdien i to ulike datasett. Vi kjenner ikke standardavviket i noen av dem. Dessuten er dataene ikke parete. Det er mer enn 30 målinger i hver dataserie, så man trenger ikke anta at målingene er normalfordelt. Vi bruker en ensidig uparet T-test med antatt like varianser. Fra utskriften ser man at variansene er tilnærmet like store, så antagelsen om variansene er i orden. Det er også en forutsetning at dataene er *uavhengige*, altså at resultatet i en fotballkamp ikke påvirker resultatet i en annen kamp. Denne antagelsen virker jo umiddelbart gal, men kanskje det jevner seg ut med mange kamper i datasettet, så vi regner likevel.

Hypotesen  $H$  som er oppgitt er at hjemmebanefordelen har blitt mindre etter innføringen av koronatiltak. Dette er påstanden vi ønsker å bevise, altså er  $H$  vår alternative hypotese. Nullhypotesen blir at hjemmebanefordelen ikke har blitt mindre etter innføringen av koronatiltak.

Vi gjennomfører hypotesetesten ved å sammenligne signifikansnivået  $\alpha = 0,05$  med avlest  $p$ -verdi. For en ensidig test ser vi at  $p = 0,175144977$ . Dette er større enn  $\alpha$ . Vi beholder nullhypotesen. Kommentarer: Det trengs nok flere spilte kamper før man kan påvise en signifikant sammenheng.



### Oppgavetekst versjon 2:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Toppserien for kvinner i Norge før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Toppserien i Norge er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 3.juli til og med 15.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	39
Uavgjort $X_1 = 1$	20
Borteseier $X_1 = 0$	27

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,5930$  og  $S_1 \approx 1,3406$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Toppserien 2019. Dette datasettet inneholder alle kamper spilt fra og med 23.mars til og med 16.november 2019.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	54
Uavgjort $X_2 = 1$	31
Borteseier $X_2 = 0$	47

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	<i>Toppserien før</i>	<i>Toppserien etter</i>
Gjennomsnitt	1.462121212	1.593023256
Varians	1.792447375	1.797127223
Observasjoner	132	86
Gruppevarians	1.794288982	
Antatt avvik mellom gjennomsnittene	0	
fg	216	
t-Stat	-0.705193298	
P(T<=t) ensidig	0.240724851	
T-kritisk, ensidig	1.651938651	
P(T<=t) tosidig	0.481449703	
T-kritisk, tosidig	1.971007472	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Toppserien er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 2:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,240724851$  større enn  $\alpha$ . Merk her at fortegnet på testobservator er negativt. I observerte data har faktisk hjemmebanefordelen økt i Toppserie etter innføringen av koronatiltak. Økningen er ikke signifikant.

### Oppgavetekst versjon 3:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Premier League i England før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Premier League i England er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 17.juni til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	72
Uavgjort $X_1 = 1$	35
Borteseier $X_1 = 0$	63

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,4765$  og  $S_1 \approx 1,3596$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Premier League 2019/20. Dette datasettet inneholder alle kamper spilt fra og med 9.august 2019 til og med 9.mars 2020.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	130
Uavgjort $X_2 = 1$	73
Borteseier $X_2 = 0$	87

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Premier League før	Premier League etter
Gjennomsnitt	1.596551724	1.476470588
Varians	1.743240663	1.848555517
Observasjoner	290	170
Gruppevarians	1.782101384	
Antatt avvik mellom gjennomsnittene	0	
fg	458	
t-Stat	0.931221785	
P(T<=t) ensidig	0.176114723	
T-kritisk, ensidig	1.648187415	
P(T<=t) tosidig	0.352229446	
T-kritisk, tosidig	1.965157098	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Premier League er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 3:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,176114723$  større enn  $\alpha$ . Reduksjonen av hjemmebanefordel i Premier League er ikke signifikant.

#### Oppgavetekst versjon 4:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Bundesliga i Tyskland før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Bundesliga i Tyskland er hentet fra [nifs.no](https://www.nifs.no). Dette datasettet inneholder alle kamper spilt fra og med 16.mai til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	49
Uavgjort $X_1 = 1$	39
Borteseier $X_1 = 0$	57

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,2828$  og  $S_1 \approx 1,2948$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Bundesliga 19/20. Dette datasettet inneholder alle kamper spilt fra og med 16.august 2019 til og med 11.mars 2020.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	97
Uavgjort $X_2 = 1$	49
Borteseier $X_2 = 0$	78

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Bundesliga før	Bundesliga etter
Gjennomsnitt	1.517857143	1.282758621
Varians	1.820307495	1.676436782
Observasjoner	224	145
Gruppevarians	1.763856861	
Antatt avvik mellom gjennomsnittene	0	
fg	367	
t-Stat	1.660783636	
P(T<=t) ensidig	0.048805473	
T-kritisk, ensidig	1.649016151	
P(T<=t) tosidig	0.097610946	
T-kritisk, tosidig	1.966448946	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Bundesliga er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 4:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,048805473$  mindre enn  $\alpha$ . Forkast nullhypotese. Reduksjonen av hjemmebanefordel i Bundesliga er signifikant.

### Oppgavetekst versjon 5:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Primera División i Spania før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Primera División i Spania er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 11.juni til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	79
Uavgjort $X_1 = 1$	52
Borteseier $X_1 = 0$	62

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,4974$  og  $S_1 \approx 1,3115$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Primera División 19/20. Dette datasettet inneholder alle kamper spilt fra og med 16.august 2019 til og med 10.mars 2020.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	129
Uavgjort $X_2 = 1$	75
Borteseier $X_2 = 0$	66

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Primera División før	Primera División etter
Gjennomsnitt	1.711111111	1.497409326
Varians	1.656009913	1.720045337
Observasjoner	270	193
Gruppevarians	1.682679764	
Antatt avvik mellom gjennomsnittene	0	
fg	461	
t-Stat	1.747742528	
P(T<=t) ensidig	0.040587195	
T-kritisk, ensidig	1.648165676	
P(T<=t) tosidig	0.08117439	
T-kritisk, tosidig	1.965123216	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Primera División er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 5:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,040587195$  mindre enn  $\alpha$ . Forkast nullhypotese. Reduksjonen av hjemmebanefordel i Primera División er signifikant.



### Oppgavetekst versjon 6:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Eredivisie i Nederland før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Eredivisie i Nederland er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 12.september til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	30
Uavgjort $X_1 = 1$	19
Borteseier $X_1 = 0$	22

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,5352$  og  $S_1 \approx 1,3183$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Eredivisie 19/20. Dette datasettet inneholder alle kamper spilt fra og med 2.august 2019 til og med 26. serierunde, hvorpå serien ble avblåst.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	113
Uavgjort $X_2 = 1$	73
Borteseier $X_2 = 0$	54

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Eredivisie før	Eredivisie etter
Gjennomsnitt	1.762931034	1.535211268
Varians	1.766065831	1.738028169
Observasjoner	232	71
Gruppevarians	1.759545444	
Antatt avvik mellom gjennomsnittene	0	
fg	301	
t-Stat	1.265762833	
P(T<=t) ensidig	0.103288422	
T-kritisk, ensidig	1.649931694	
P(T<=t) tosidig	0.206576843	
T-kritisk, tosidig	1.967876531	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Eredivisie er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 6:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,103288422$  større enn  $\alpha$ . Reduksjonen av hjemmebanefordel i Eredivisie er ikke signifikant.

### Oppgavetekst versjon 7:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Allsvenskan i Sverige før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Allsvenskan i Sverige er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 14.juni til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	80
Uavgjort $X_1 = 1$	71
Borteseier $X_1 = 0$	65

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,4398$  og  $S_1 \approx 1,2635$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Allsvenskan 2019. Dette datasettet inneholder alle kamper spilt fra og med 31.mars til og med 2.november 2019.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	110
Uavgjort $X_2 = 1$	60
Borteseier $X_2 = 0$	70

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Allsvenskan før	Allsvenskan etter
Gjennomsnitt	1.625	1.439814815
Varians	1.741631799	1.596360896
Observasjoner	240	216
Gruppevarians	1.672836107	
Antatt avvik mellom gjennomsnittene	0	
fg	454	
t-Stat	1.526615327	
P(T<=t) ensidig	0.063776516	
T-kritisk, ensidig	1.648216847	
P(T<=t) tosidig	0.127553032	
T-kritisk, tosidig	1.965202973	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Allsvenskan er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 7:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,063776516$  større enn  $\alpha$ . Reduksjonen av hjemmebanefordel i Allsvenskan er ikke signifikant.

### Oppgavetekst versjon 8:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Damallsvenskan for kvinner i Sverige før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Damallsvenskan i Sverige er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 27.juni til og med 15.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	60
Uavgjort $X_1 = 1$	22
Borteseier $X_1 = 0$	50

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,5303$  og  $S_1 \approx 1,3894$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Damallsvenskan 2019. Dette datasettet inneholder alle kamper spilt fra og med 14.april til og med 26.oktober 2019.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	63
Uavgjort $X_2 = 1$	32
Borteseier $X_2 = 0$	37

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	<i>Damallsvenskan før</i>	<i>Damallsvenskan etter</i>
Gjennomsnitt	1.674242424	1.53030303
Varians	1.748033773	1.930372427
Observasjoner	132	132
Gruppevarians	1.8392031	
Antatt avvik mellom gjennomsnittene	0	
fg	262	
t-Stat	0.862257126	
P(T<=t) ensidig	0.194667284	
T-kritisk, ensidig	1.650690284	
P(T<=t) tosidig	0.389334568	
T-kritisk, tosidig	1.969059715	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Damallsvenskan er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 8:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,194667284$  større enn  $\alpha$ . Reduksjonen av hjemmebanefordel i Damallsvenskan er ikke signifikant.

### Oppgavetekst versjon 9:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Ligue 1 i Frankrike før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Ligue 1 i Frankrike er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 21.august til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	43
Uavgjort $X_1 = 1$	22
Borteseier $X_1 = 0$	33

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,5408$  og  $S_1 \approx 1,3483$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Ligue 1 19/20. Dette datasettet inneholder alle kamper spilt fra og med 9.august 2019 til og med 28. serierunde, hvorpå serien ble avblåst.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	134
Uavgjort $X_2 = 1$	70
Borteseier $X_2 = 0$	75

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	Ligue 1 før	Ligue 1 etter
Gjennomsnitt	1.691756272	1.540816327
Varians	1.717593667	1.817904481
Observasjoner	279	98
Gruppevarians	1.743540731	
Antatt avvik mellom gjennomsnittene	0	
fg	375	
t-Stat	0.97349249	
P(T<=t) ensidig	0.165468056	
T-kritisk, ensidig	1.648927129	
P(T<=t) tosidig	0.330936112	
T-kritisk, tosidig	1.966310161	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Ligue 1 er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 9:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,165468056$  større enn  $\alpha$ . Reduksjonen av hjemmebanefordel i Ligue 1 er ikke signifikant.



### Oppgavetekst versjon 10:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Serie A i Italia før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Serie A i Italia er hentet fra [nifs.no](http://nifs.no). Dette datasettet inneholder alle kamper spilt fra og med 20.juni til og med 8.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	81
Uavgjort $X_1 = 1$	45
Borteseier $X_1 = 0$	68

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,4845$  og  $S_1 \approx 1,3398$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Serie A i 19/20. Dette datasettet inneholder alle kamper spilt fra og med 24.august til og med 9.mars 2020.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	103
Uavgjort $X_2 = 1$	58
Borteseier $X_2 = 0$	95

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	<i>Serie A før</i>	<i>Serie A etter</i>
Gjennomsnitt	1.43359375	1.484536082
Varians	1.799494485	1.795096416
Observasjoner	256	194
Gruppevarians	1.797599781	
Antatt avvik mellom gjennomsnittene	0	
fg	448	
t-Stat	-0.399159908	
P(T<=t) ensidig	0.344982793	
T-kritisk, ensidig	1.648261984	
P(T<=t) tosidig	0.689965586	
T-kritisk, tosidig	1.965273324	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen er

$$H : \text{ I Serie A er hjemmebanefordelen mindre etter innføringen av koronatiltak. } (\mu_1 < \mu_2)$$

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 10:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,344982793$  større enn  $\alpha$ . Merk her at fortegnet på testobservator er negativt. I observerte data har faktisk hjemmebanefordelen økt i Serie A etter innføringen av koronatiltak. Økningen er ikke signifikant.

### Oppgavetekst versjon 11:

I fotball har man begrepet hjemmebanefordel. Etter 12.mars i 2020 har koronatiltakene gitt store begrensninger i fotballen. Blant annet har kampene blitt spilt uten publikum, eller med kun et begrenset publikum. Dette kan ha hatt innvirkning på hjemmebanefordelen. I denne oppgaven skal du se på hjemmebanefordelen i Division 1 Féminine for kvinner i Frankrike før og etter koronatiltakene, og sammenligne ved hypotesetesting.

- a) La  $X_1$  være antall poeng til hjemmelaget i fotballkamper spilt med koronarestriksjoner. Det gis 3 poeng ved hjemmeseier, 1 poeng ved uavgjort, og 0 poeng ved borteseier. Data fra Division 1 Féminine i Frankrike er hentet fra `nifs.no`. Dette datasettet inneholder alle kamper spilt fra og med 5.september til og med 14.november 2020:

Resultat	Antall kamper
Hjemmeseier $X_1 = 3$	25
Uavgjort $X_1 = 1$	6
Borteseier $X_1 = 0$	17

Vis utregning av gjennomsnitt og standardavvik for dette datasettet. Svarene skal bli  $\bar{X}_1 \approx 1,6875$  og  $S_1 \approx 1,4166$ .

- b) Til sammenligning la  $X_2$  være poeng til hjemmelaget i en fotballkamp spilt før innføringen av koronatiltakene. Vi har følgende data fra Division 1 Féminine 19/20. Dette datasettet inneholder alle kamper spilt fra og med 24.august 2019 til og med 16. serierunde, hvorpå serien ble avblåst.

Resultat	Antall kamper
Hjemmeseier $X_2 = 3$	37
Uavgjort $X_2 = 1$	16
Borteseier $X_2 = 0$	43

Vi legger disse to datasettene inn i Excel og får følgende utskrift:

t-Test: To utvalg med antatt like varianser

	<i>Division 1 Féminine før</i>	<i>Division 1 Féminine etter</i>
Gjennomsnitt	1.322916667	1.6875
Varians	1.905153509	2.006648936
Observasjoner	96	48
Gruppevarians	1.938747066	
Antatt avvik mellom gjennomsnittene	0	
fg	142	
t-Stat	-1.481191543	
P(T<=t) ensidig	0.070385543	
T-kritisk, ensidig	1.655655173	
P(T<=t) tosidig	0.140771087	
T-kritisk, tosidig	1.976810994	

Bruk utskriften fra Excel til å gjennomføre en hypotesetest der den ene hypotesen

er

$H$  : Division 1 Féminine er hjemmebanefordelen mindre  
etter innføringen av koronatiltak. ( $\mu_1 < \mu_2$ )

Benytt signifikansnivå  $\alpha = 0,05$ .

Hvilken hypotesetest kan du bruke i dette tilfellet? Hvilke forutsetninger krever denne testen? Gjelder forutsetningene for datasettet som er oppgitt? Hva er nullhypotese og alternativ hypotese? Hva blir konklusjonen? Gi din kommentar til resultatet.

**Fasit versjon 11:**

- a) Se løsningsforslag på versjon 1.
- b)  $p = 0,070385543$  større enn  $\alpha$ . Merk her at fortegnet på testobservator er negativt. I observerte data har faktisk hjemmebanefordelen økt i Division 1 Féminine etter innføringen av koronatiltak. Økningen er ikke signifikant.