Eksamen, høsten 20 deleksamen Statistikk

Løsningsforslag og sensorveilednig.

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) Kim spiller en fotballkamp. La A, B og C være de tre hendelsene:

A: Kim får rødt kort i andre omgang.

B: Kim scorer mål i første omgang.

C: Kim forlater banen med skade i andre omgang.

Følgende sannsynligheter er oppgitt:

$$\begin{aligned}
&P(A) = 0{,}2\\
&P(B) = 0{,}4\\
&P(C) = 0{,}1\\
&P(A \cap B) = 0{,}1\\
&P(A \cap C) = 0\\
&P(B \cap C) = 0{,}04
\end{aligned}$$

Regn ut sannsynlighetene $P(A \cup B)$, $P(\overline{C})$ og $P(B|C)$. Er noen av de tre hendelsene A, B og C uavhengige av hverandre?

b) Fotballaget til Kim har en stall på 22 utespillere. Av disse er 7 venstrebente. Dersom treneren stiller opp et tilfeldig lag med 10 utespillere, hva er sannsynligheten for at nøyaktig 3 av utespillerne er venstrebente?

c) Sannsynligheten for at en tilfeldig person har lest en bok av Yuval Noah Harari er 0,1. Andelen fotballspillere blant Hararis lesere er 0,45. Sannsynligheten for at en tilfeldig person, blant de som ikke leser bøker av Harari, er fotballspiller er 0,05. Hva er sannsynligheten for at en tilfeldig fotballspiller har lest en bok av Harari?

Løsning Versjon 1:

a) Ved regnereglene for sannsynligheter har vi:

$$\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) = 0{,}2 + 0{,}4 - 0{,}1 = \underline{\underline{0{,}5}}\\
P(\overline{C}) &= 1 - P(C) = 1 - 0{,}1 = \underline{\underline{0{,}9}}\\
P(B|C) &= \frac{P(B \cap C)}{P(C)} = \frac{0{,}04}{0{,}1} = \underline{\underline{0{,}4}}
\end{aligned}$$

Å være *uavhengige hendelser* betyr at de ikke påvirker hverandres sannsynligheter, med formler er kriteriet $P(H_1 \cap H_2) = P(H_1) \cdot P(H_2)$.

Vi ser i denne oppgaven at B og C er uavhengige, men ikke A og B, og heller ikke A og C.

b) Dette er en *hypergeometrisk fordeling* med en populasjon på $N = 22$ utespillere i stallen, hvorav $M = 7$ har den spesielle egenskapen å være venstrebente. Treneren gjør et utvalg på $n = 10$ utespillere i en lagoppstilling. Sannsynligheten for at nøyaktig $x = 3$ i utvalget er venstrebente en punktsannsynlighet fra hypergeometrisk fordeling. Vi benytter den kjente formelen:

$$P(X = x) \;=\; \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Innsatt tall fra teksten:

$$\text{Sannsynligheten} = \frac{\binom{7}{3} \cdot \binom{22-7}{10-3}}{\binom{22}{10}} = \frac{35 \cdot 6435}{646646} \approx \underline{\underline{0{,}3483}}$$

På denne deloppgaven kan man gjerne finne svaret ved å bruke hypergeometrisk fordeling i Geogebra, fylle inn tall og avlese svaret som en punktsannsynlighet. I så fall forventes det at det ligger ved et skjermbilde eller lignende som dokumentasjon.

c) Denne deloppgaven løses ved å bruke Bayes formel. Vi har

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B})}.$$

Her er nevneren den totale sannsynligheten for $P(A)$.
Vi identifiserer de to hendelsene A og B fra teksten:

A: En tilfeldig person er fotballspiller.

B: En tilfeldig person har lest en bok av Harari.

Så avleser vi opplysningene fra teksten om sannsynligheter. Vi har

$$\begin{aligned} P(B) &= 0{,}1 \\ P(A|B) &= 0{,}45 \\ P(A|\overline{B}) &= 0{,}05 \end{aligned}$$

Vi regner ut sannnsynligheten for ikke B:

$$P(\overline{B}) = 1 - P(B) = 1 - 0{,}1 = 0{,}9$$

Så finner vi total sannsynlighet for A:

$$P(A) = P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B}) = 0{,}45 \cdot 0{,}1 + 0{,}05 \cdot 0{,}9 = 0{,}09$$

Innsatt i Bayes formel får vi:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B})} = \frac{0{,}45 \cdot 0{,}1}{0{,}09} = \underline{\underline{0{,}5}}$$

<u>Sannsynligheten for at en tilfeldig fotballspiller har lest en bok av Harari er 0,5.</u>

Oppgavetekst versjon 2:

a) Kim sykler til jobben. La A, B og C være de tre hendelsene:

A: Kim blir forbikjørt av en elsparkesykkel.

B: Kim punkterer.

C: Kim sin oppladbare frontlykt går tom for batteri.

Følgende sannsynligheter er oppgitt:

$$P(A) = 0{,}3$$
$$P(B) = 0{,}2$$
$$P(C) = 0{,}1$$
$$P(A \cap B) = 0{,}08$$
$$P(A \cap C) = 0{,}03$$
$$P(B \cap C) = 0{,}02$$

Regn ut sannsynlighetene $P(A \cup B)$, $P(\overline{C})$ og $P(B|C)$. Er noen av de tre hendelsene A, B og C uavhengige av hverandre?

b) I byen der Kim bor er det 30 personer som sykler til jobb. Av disse har 22 riktig lys på sykkelen. En tidlig morgen er det politikontroll av myke trafikanter. I denne kontrollen stoppes 10 tilfeldige syklister. Hva er sannsynligheten for at politiet stoppet nøyaktig 7 syklister med lys i orden?

c) Sannsynligheten for at en tilfeldig person har sett en Børning-film er 0,3. Andelen syklister blant de som har sett en Børning-film er 0,35. Sannsynligheten for at en tilfeldig person, blant de som ikke har sett noen Børning-film, er syklist er 0,15. Hva er sannsynligheten for at en tilfeldig syklist har sett en Børning-film?

Fasit versjon 2:

a) $P(A \cup B) = 0{,}42$, $P(\overline{C}) = 0{,}9$, $P(B|C) = 0{,}2$. Hendelsene A og C er uavhengige, hendelsene B og C er uavheninge, men hendelsene A og B er ikke uavhengige.

b) $P(X = 7) = \frac{\binom{22}{7} \cdot \binom{8}{3}}{\binom{30}{10}} \approx 0{,}3179$.

c) Hendelse A: En tilfeldig person er syklist. Hendelse B: En tilfeldig person har sett en Børning-film. $P(B|A) = 0{,}5$.

Oppgavetekst versjon 3:

a) Kim eier en hytte på fjellet. La A, B og C være de tre hendelsene:

A: Kim må måke ned snø fra taket.

B: Det er sol.

C: Det er ikke skiføre.

Følgende sannsynligheter er oppgitt:

$$P(A) = 0{,}1$$
$$P(B) = 0{,}4$$
$$P(C) = 0{,}2$$
$$P(A \cap B) = 0{,}05$$
$$P(A \cap C) = 0$$
$$P(B \cap C) = 0{,}08$$

Regn ut sannsynlighetene $P(A \cup B)$, $P(\overline{C})$ og $P(B|C)$. Er noen av de tre hendelsene A, B og C uavhengige av hverandre?

b) Kim har 18 gode venner. Av disse er det 7 som disponerer en bil. Kim treffer 6 av de atten gode vennene ved en tilfeldighet, og kan fortelle at hytta er ledig den siste helga i november. Hva er sannsynligheten for at nøyaktig 2 av de seks vennene disponerer bil?

c) Sannsynligheten for at en tilfeldig person liker å gå på ski er 0,4. Andelen hytteeiere blant de som liker å gå på ski er 0,1. Sannsynligheten for at en tilfeldig person, blant de som ikke liker å gå på ski, eier hytte er 0,04. Hva er sannsynligheten for at en tilfeldig hytteeier liker å gå på ski?

Fasit versjon 3:

a) $P(A \cup B) = 0{,}45$, $P(\overline{C}) = 0{,}8$, $P(B|C) = 0{,}4$. Hendelsene B og C er uavhengige, men hendelsene A og B er ikke uavheninge, og hendelsene A og C er ikke uavhengige.

b) $P(X = 2) = \frac{\binom{7}{2} \cdot \binom{11}{4}}{\binom{18}{6}} \approx 0{,}3733$.

c) Hendelse A: En tilfeldig person er hytteeier. Hendelse B: En tilfeldig person liker å gå på ski. $P(B|A) = 0{,}625$.

Oppgavetekst versjon 4:

a) Kim liker å jogge. La A, B og C være de tre hendelsene:

A: Kim ser et rådyr i løpet av joggeturen.
B: Kim jogger i byen.
C: Det regner i løpet av joggeturen.

Følgende sannsynligheter er oppgitt:

$$P(A) = 0{,}1$$
$$P(B) = 0{,}5$$
$$P(C) = 0{,}3$$
$$P(A \cap B) = 0$$
$$P(A \cap C) = 0{,}04$$
$$P(B \cap C) = 0{,}2$$

Regn ut sannsynlighetene $P(A \cup B)$, $P(\overline{C})$ og $P(B|C)$. Er noen av de tre hendelsene A, B og C uavhengige av hverandre?

b) I et mosjonsløp deltar 29 løpere. Av disse er det 17 som har pulsklokke. I mosjonsløpet deles det ut premier til 12 av løperne. Hva er sannsynligheten for at nøyaktig 2 av de premierte har pulsklokke?

c) Sannsynligheten for at en tilfeldig person noen gang har prøvd å lage Tiramisu er 0,2. Andelen joggere blant de som har prøvd å lage Tiramisu er 0,6. Sannsynligheten for at en tilfeldig person, blant de som ikke har prøvd å lage Tiramisu, er en jogger er 0,3. Hva er sannsynligheten for at en tilfeldig jogger har prøvd å lage Tiramisu?

Fasit versjon 4:

a) $P(A \cup B) = 0{,}6$, $P(\overline{C}) = 0{,}7$, $P(B|C) = \frac{2}{3} \approx 0{,}6667$. Ingen av hendelsene er uavhengige.

b) $P(X = 2) = \frac{\binom{17}{2} \cdot \binom{12}{10}}{\binom{29}{12}} \approx 1{,}73 \cdot 10^{-4}$.

c) Hendelse A: En tilfeldig person er jogger. Hendelse B: En tilfeldig person har noen gang prøvd å lage Tiramisu. $P(B|A) = \frac{1}{3} \approx 0{,}3333$.

Oppgave 2. **Oppgavetekst versjon 1:**

La X være antall kjærestepar som dannes i løpet av en måned på en videregående skole.

a) Anta at X er Poissonfordelt. Høsten 2019 var raten $\lambda = 4{,}2$ pr mnd. Finn $E(X)$, $\text{Var}(X)$ og $P(X > 2)$.

b) Hvilke forutsetninger må man anta for at X skal være Poissonfordelt? I hvilken grad mener du at disse forutsetningene gjelder i eksempelet over? Gi minst to konkrete innvendinger som viser at forutsetningene ikke gjelder eksakt.

c) Høsten 2020 ble det dannet 12 kjærestepar i løpet av 4 måneder. Bruk dette til å estimere en ny rate, og lage et 95%-konfidensintervall. Har raten endret seg signifikant fra høsten 2019 til høsten 2020?

Løsningsforslag versjon 1:

a) Siden X er antallet nye kjærestepar i løpet av en måned, så har vi en tid $t = 1$ måned. Sammen med raten $\lambda = 4{,}2$ får man følgende fra formlene til Poissonfordelingen:

$$E(X) = \lambda t = 4{,}2 \cdot 1 = \underline{\underline{4{,}2}}$$
$$\text{Var}(X) = \lambda t = 4{,}2 \cdot 1 = \underline{\underline{4{,}2}}$$

For å regne $P(X > 2)$ finner vi punktsannsynlighetene $P(X = 0)$, $P(X = 1)$ og $P(X = 2)$, og benytter deretter formelen $P(X > 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$. Formelen for punktsannsynlighet er:

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

Vi får:

$$P(X=0) = \frac{(4{,}2)^0}{0!}e^{-4{,}2} \approx 0{,}014995577$$
$$P(X=1) = \frac{(4{,}2)^1}{1!}e^{-4{,}2} \approx 0{,}062981423$$
$$P(X=2) = \frac{(4{,}2)^2}{2!}e^{-4{,}2} \approx 0{,}132260988$$

Dermed har vi

$$P(X > 2) \approx 1 - 0{,}014995577 - 0{,}062981423 - 0{,}132260988 \approx \underline{\underline{0{,}7898}}$$

b) Antall forekomster av hendelsen at det dannes et nytt kjærestepar på en videregående skole er poissonfordelt dersom:

(1) Dannelser av nye kjærestepar innenfor disjunkte tidsrom skjer uavhengig av hverandre.

(2) Forventet antall dannelser har en konstant rate.

(3) To ulike dannelser av kjærestepar kan ikke skje samtidig.

For disse betingelsene gjelder nok (1) og (2) i mindre grad, mens (3) er temmelig riktig. Eksempler på innvendinger er:

Mot (1): Dannelse av ett kjærestepar kan inspirere andre til å bli sammen like etterpå.

Mot (1): Dannelse av ett kjærestepar kan gjør at det er færre signle på skolen etterpå og det reduserer derfor mulighetene til nye par.

Mot (2): Raten for nye par kan være større når det er godt vær i august, enn senere på høsten.

Mot (2): Raten for dannelse av par kan være større i helgene enn på hverdagene.

Mot (3): Det kan være vanskelig å avgjøre eksakt tidspunkt for dannelsen av et nytt par. Det kan være at man teller tidspunktet der det blir gjort kjent for andre, og i så fall kan det være at to venninner snakker sammen og i samme samtale forteller at de begge har fått kjæreste.

c) Et konfidensintervall for Poisson-raten λ er gitt ved

$$\left[\hat{\lambda} - z_{\alpha/2} \cdot \sqrt{\hat{\lambda}/t}\,,\ \hat{\lambda} + z_{\alpha/2} \cdot \sqrt{\hat{\lambda}/t}\right]$$

under forutsetning av at normaltilnærming gjelder. I dette tilfellet observeres $x = 12$ nye kjærestepar i løpet av $t = 4$ måneder. Vi bruker signifikanssannsynlighet $\alpha = 0{,}05$, som tilsvarer et 95%-konfidensintervall. Det tilsvarende kvantilet er $z_{\alpha/2} = 1{,}960$. Den estimerte raten blir

$$\hat{\lambda} = \frac{x}{t} = \frac{12}{4} = \underline{\underline{3}}.$$

Øvre og nedre grense for konfidensintervallet regnes ut ved:

$$\hat{\lambda} + z_{\alpha/2} \cdot \sqrt{\hat{\lambda}/t} \approx 3 + 1{,}960 \cdot \sqrt{3/4} \approx \underline{\underline{4{,}6973}}$$
$$\hat{\lambda} - z_{\alpha/2} \cdot \sqrt{\hat{\lambda}/t} \approx 3 - 1{,}960 \cdot \sqrt{3/4} \approx \underline{\underline{1{,}3026}}$$

Den gamle raten på 4,2 fra høsten 2019 er innfor dette konfidensintervallet.
Derfor har ikke raten endret seg signifikant.

Kravet om normaltilnærming er innfridd om variansen er større enn 10. Dersom vi estimerer $\lambda \approx \hat{\lambda} = 3$ vil $\lambda t \approx 3 \cdot 4 = 12 > 10$, og kravet er i orden. Men dersom vi legger nedre grense for konfidensintervallet til grunn, altså $\lambda \approx 1{,}3$, så holder ikke normaltilnærmingen så godt. Vi bruker likevel konfidensintervallet vi har funnet.

Oppgavetekst versjon 2:
La X være antall bilulykker på en veistrekning i løpet av en måned.

a) Anta at X er Poissonfordelt. Vinterhalvåret 2018–19 var raten $\lambda = 3{,}5$ pr mnd. Finn $E(X)$, $\text{Var}(X)$ og $P(X > 2)$.

b) Hvilke forutsetninger må man anta for at X skal være Poissonfordelt? I hvilken grad mener du at disse forutsetningene gjelder i eksempelet over? Gi minst to konkrete innvendinger som viser at forutsetningene ikke gjelder eksakt.

c) Vinterhalvåret 2019–20 ble det var registret 13 bilulykker på veistrekningen i løpet av 6 måneder. Bruk dette til å estimere en ny rate, og lage et 95%-konfidensintervall. Har raten endret seg signifikant fra vinteren 2018–19 til vinteren 2019–20?

Fasit versjon 2:

a) $E(X) = 3{,}5$, $\text{Var}(X) = 3{,}5$, $P(X > 2) \approx 0{,}6792$

b) Forutsetningene (1) og (2) gjelder i noen grad, mens (3) i større grad. Eksempel på innvendinger:

Mot (2): Været kan påvirke raten for bilulykker.

Mot (1): Mange bilulykker kan gjøre at det igangsettes kampanjer for å redusere flere ulykker.

Mot (3): Skal man telle en kollisjon mellom to kjøretøyer som 2 ulykker, og en kjedekollisjon som mer enn 2 ulykker? Inntreffer disse hendelsene i så fall på samme tidspunkt?

Mot (2): Risikoen for ulykke varierer alt etter om det er dag eller natt.

c) $\hat{\lambda} \approx 2{,}1667$ Konfidensintervall $[0{,}9889\,,\ 3{,}3445]$. Endringen er signifikant.

Oppgavetekst versjon 3:
Une spiller i breddefotballen. La X være antall mål Une scorer pr kamp.

a) Anta at X er Poissonfordelt. Sesongen 2018 var raten $\lambda = 1{,}3$ pr kamp. Finn $E(X)$, $\text{Var}(X)$ og $P(X > 2)$.

b) Hvilke forutsetninger må man anta for at X skal være Poissonfordelt? I hvilken grad mener du at disse forutsetningene gjelder i eksempelet over? Gi minst to konkrete innvendinger som viser at forutsetningene ikke gjelder eksakt.

c) Sesongen 2019 scorte Une 28 mål i løpet av en spilletid på 14 kamper. Bruk dette til å estimere en ny rate, og lage et 95%-konfidensintervall. Har raten endret seg signifikant fra sesongen 2018 til sesongen 2019?

Fasit versjon 3:

a) $E(X) = 1{,}3$, $\text{Var}(X) = 1{,}3$, $P(X > 2) \approx 0{,}1429$

b) Forutsetningene (1) gjelder i liten grad, mens (2) gjelder i noen grad. Forutsetning (3) gjelder. Eksempel på innvendinger:

Mot (1): Å score mål gir selvtillit, og høy selvtillit gir flere mål.

Mot (1): I en kamp er sjansen for et nytt mål umiddelbart etter scoring noe redusert. Spillerne feirer målet, og det tar noe tid før kampen er i gang igjen.

Mot (2): Hvilket lag som er motstander påvirker sannsynligheten for scoring.

c) $\hat{\lambda} = 2$ Konfidensintervall $[1{,}2592\,,\ 2{,}7408]$. Endringen er ikke signifikant.

Oppgavetekst versjon 4:
En bilforhandler selger hybridbiler. La X være antall hybridbiler som selges pr måned.

a) Anta at X er Poissonfordelt. Høsten 2019 var raten $\lambda = 5{,}5$ pr måned. Finn $E(X)$, $\text{Var}(X)$ og $P(X > 2)$.

b) Hvilke forutsetninger må man anta for at X skal være Poissonfordelt? I hvilken grad mener du at disse forutsetningene gjelder i eksempelet over? Gi minst to konkrete innvendinger som viser at forutsetningene ikke gjelder eksakt.

c) Høsten 2020 solgte forhandleren 37 hybridbiler i løpet av 5 måneder. Bruk dette til å estimere en ny rate, og lage et 95%-konfidensintervall. Har raten endret seg signifikant fra høsten 2019 til høsten 2020?

Fasit versjon 4:

a) $E(X) = 5{,}5$, $\text{Var}(X) = 5{,}5$, $P(X > 2) \approx 0{,}9116$

b) Forutsetningene (1) og (2) gjelder i noen grad. Forutsetning (3) gjelder i mindre grad. Eksempel på innvendinger:

Mot (3): En kunde kan kjøpe flere hybridbiler samtidig.

Mot (2): Lansering av nye bilmodeller påvirker salgsraten.

Mot (1): Dersom en selger selger over forventning mange biler kan han kanskje ta seg avspassering en tid, og dermed ikke selge flere biler på en stund.

c) $\hat{\lambda} = 7{,}4$ Konfidensintervall $[5{,}0156\,,\ 9{,}7844]$. Endringen er ikke signifikant.

Oppgave 3. **Oppgavetekst versjon 1:**
Russiske forskere ønsker raskt å utvikle en ny vaksine mot en smittsom virussykdom. Den russiske hæren stiller soldater til disposisjon for forskerne. Ved å eksponere vaksinerte og uvaksinerte soldater for smitte kan forskerne finne ut om vaksinen virker eller ikke.

a) Når en uvaksinert russisk soldat utsettes for smitte er sannsynligheten $p_0 = 0{,}8$ for at vedkommende smittes og blir syk. I et forsøk blir $n = 100$ uvaksinerte soldater utsatt for smitte. La X være antall av disse som blir syke.

Hvilken fordeling har X? Regn ut forventningsverdi og varians. Er X tilnærmet normalfordelt? Finn sannsynligheten for at mindre enn 75 soldater blir syke.

b) De russiske forskerne gjør et nytt forsøk der $n = 200$ soldater vaksineres og deretter utsettes for smitte. Av disse blir $x = 68$ syke.

Utfør en ensidig hypotesetest av sannsynligheten p der alternativ hypotese er

$$H_1 : \text{Vaksinen reduserer sannsynligheten for å bli syk.}$$

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

c) Norske forskere vurderer å gjøre egne forsøk med den russiske vaksinen. De ønsker derfor å anslå hvor stort utvalg av norske forsøkspersoner man trenger for å oppnå gode resultater. De norske forsøkspersonene skal ikke utsettes for smitte på kunstig vis, men leve som normalt.

De norske forskerne ønsker å beregne et 95% konfidensintervall for sannsynligheten p for at en norsk forsøksperson blir smittet selv om vedkommende er vaksinert. Dette konfidensintervallet skal deretter sammenlignes med smittetrykket blant alle uvaksinerte forsøkspersoner. Det er forventet at smittetrykket i løpet av forsøket vil være $p_1 = 0{,}01$. Her skal man forstå *smittetrykk* som sannsynligheten for at en uvaksinert forsøksperson pådrar seg smitte i løpet av forsøksperioden.

De norske forskerne mener resultatene er gode dersom konfidensintervallet for p får en lengde L på 0,002 eller mindre.

Kravet til utvalgsstørrelse er gitt ved ulikheten

$$n \geq 4\hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{L}\right)^2.$$

Anslå antall norske forsøkspersoner som minst må vaksineres i forsøket med den russiske vaksinen. Begrunn de antagelsene du gjør underveis, og kommenter eventuelle forutsetninger.

Løsningsforslag versjon 1:

a) Om vi antar at sykdom hos ulike soldater forekommer uavhengig, så vil X være $\underline{\underline{\text{binomisk fordelt}}}$. Forventningsverdi og varians regnes ut ved kjente formler:

$$E(X) = np = 100 \cdot 0{,}8 = \underline{\underline{80}}$$
$$\text{Var}(X) = np(1-p) = 100 \cdot 0{,}8 \cdot (1 - 0{,}8) = \underline{\underline{16}}$$

Siden variansen er større enn 5 blir den binomisk fordelingen tilnærmet normalfordelt. Det betyr at X er tilnærmet lik en normalfordelt Y med $\mu = 80$ og $\sigma = \sqrt{16}$. Vi kan bruke Y til å finne sannsynligheter for X. Vi har

$$P(X < 75) \approx P(Y < 75) = G\left(\frac{75-\mu}{\sigma}\right) = G\left(\frac{75-80}{4}\right) = G(-1{,}25) = \underline{\underline{0{,}1056}}$$

Enda mer nøyaktig blir svaret om man bruker heltallskorreksjon. Da får man

$$P(X < 75) \approx P(Y \leq 74{,}5) \approx \underline{\underline{0{,}0846}}$$

$\underline{\underline{\text{Sannsynligheten for at mindre enn 75 soldater blir syke er 0,1056.}}}$ (eller 0,0846 mer nøyaktig)

b) Vi utfører en ensidig hypotesetest av sannsynligheten p med følgende hypoteser:

$$H_0 : \text{Vaksinen reduserer ikke sannsynligheten for å bli syk.} \quad p \geq 0{,}8$$
$$H_1 : \text{Vaksinen reduserer sannsynligheten for å bli syk.} \quad p < 0{,}8$$

Her kommer $p_0 = 0{,}8$ fra opplysningene i deloppgave a). Denne testen forutsetter at antallet som smittes er tilnærmet normalfordelt, og dette er i orden om variansen til antallet er større enn 5.

Testobservator er gitt ved

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}},$$

og vi skal forkaste nullhypotesen om antallet blir tilstrekkelig lite, altså når $Z < -z_\alpha$.

Det er oppgitt signifikansnivå $\alpha = 0{,}01$.

Vi finner kritisk verdi som z-kvantilet $z_{0,01} = 2{,}326$.

Observerte verdier er $x = 68$ og $n = 200$. Det gir følgende verdi for testobservatoren

$$z = \frac{68 - 200 \cdot 0{,}8}{\sqrt{200 \cdot 0{,}8 \cdot (1 - 0{,}8)}} \approx -16{,}263$$

Vi forkaster derfor nullhypotesen. Den russiske vaksinen virker.

La oss se etter om forutsetningen om varians holder. Vi estimerer $\hat{p} = \frac{x}{n} = 0{,}34$. Det gir varians $n\hat{p}(1-\hat{p}) \approx 44{,}88$, som er større enn 5. Altså holder forutsetningene for testen.

c) Når de norske forskerne vi ha et 95% konfidensintervall med lengde $L = 0{,}002$ eller mindre, gir følgende ulikhet et krav til utvalgsstørrelsen:

$$n \geq 4\hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{L}\right)^2$$

Siden de norske forskerne ikke har gjennomført vaksineforsøket vet de ikke verdien av estimatoren $\hat{p}$. Men man kan få en tilnærmet grense for utvalgsstørrelsen ved å bruke smittetrykket

$$\hat{p} \approx p_1 = 0{,}01.$$

Vi setter også inn kvantilet $z_{0,025} = 1{,}960$ og får

$$n \geq 4 \cdot 0{,}01(1 - 0{,}01) \left(\frac{1{,}960}{0{,}002}\right)^2 \approx 38030.$$

De norske forskerne bør teste den russiske vaksinen på minst 38030 personer.

Forutsetningene for utregningen er varians større enn 5, og dette holder når $n = 38030$ og $p_1 = 0{,}01$.

Oppgavetekst versjon 2:

Kinesiske forskere ønsker raskt å utvikle en ny vaksine mot en smittsom virussykdom. Det kinesiske folks frigjøringshær stiller soldater til disposisjon for forskerne. Ved å eksponere vaksinerte og uvaksinerte soldater for smitte kan forskerne finne ut om vaksinen virker eller ikke.

a) Når en uvaksinert kinesisk soldat utsettes for smitte er sannsynligheten $p_0 = 0{,}7$ for at vedkommende smittes og blir syk. I et forsøk blir $n = 120$ uvaksinerte soldater utsatt for smitte. La X være antall av disse som blir syke.

   Hvilken fordeling har X? Regn ut forventningsverdi og varians. Er X tilnærmet normalfordelt? Finn sannsynligheten for at mindre enn 85 soldater blir syke.

b) De kinesiske forskerne gjør et nytt forsøk der $n = 150$ soldater vaksineres og deretter utsettes for smitte. Av disse blir $x = 101$ syke.

   Utfør en ensidig hypotesetest av sannsynligheten p der alternativ hypotese er

$$H_1 : \text{Vaksinen reduserer sannsynligheten for å bli syk.}$$

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

c) Norske forskere vurderer å gjøre egne forsøk med den kinesiske vaksinen. De ønsker derfor å anslå hvor stort utvalg av norske forsøkspersoner man trenger for å oppnå gode resultater. De norske forsøkspersonene skal ikke utsettes for smitte på kunstig vis, men leve som normalt.

   De norske forskerne ønsker å beregne et 99% konfidensintervall for sannsynligheten p for at en norsk forsøksperson blir smittet selv om vedkommende er vaksinert. Dette konfidensintervallet skal deretter sammenlignes med smittetrykket blant alle uvaksinerte forsøkspersoner. Det er forventet at smittetrykket i løpet av forsøket vil være $p_1 = 0{,}02$. Her skal man forstå *smittetrykk* som sannsynligheten for at en uvaksinert forsøksperson pådrar seg smitte i løpet av forsøksperioden.

   De norske forskerne mener resultatene er gode dersom konfidensintervallet for p får en lengde L på 0,005 eller mindre.

   Kravet til utvalgsstørrelse er gitt ved ulikheten

$$n \geq 4\hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{L} \right)^2 .$$

   Anslå antall norske forsøkspersoner som minst må vaksineres i forsøket med den kinesiske vaksinen. Begrunn de antagelsene du gjør underveis, og kommenter eventuelle forutsetninger.

Fasit versjon 2:

a) $E(X) = 84$, $\text{Var}(X) = 25{,}2$, tilnærmet normalfordelt, $P(X < 85) \approx 0{,}5789$, (eller bedre $P(X < 85) = 0{,}5397$ med heltallskorreksjon)

b) Kritisk verdi er $z_\alpha = 1{,}645$. Testobservator blir $z = -0{,}7127$. Beholder nullhypotese. Dette forsøket viser ikke at den kinesiske vaksinen virker.

c) Utvalgsstørrelse minst 20807

Oppgavetekst versjon 3:

Amerikanske forskere ønsker raskt å utvikle en ny vaksine mot en smittsom virussykdom. U.S. Army stiller soldater til disposisjon for forskerne. Ved å eksponere vaksinerte og uvaksinerte soldater for smitte kan forskerne finne ut om vaksinen virker eller ikke.

a) Når en uvaksinert amerikansk soldat utsettes for smitte er sannsynligheten $p_0 = 0{,}9$ for at vedkommende smittes og blir syk. I et forsøk blir $n = 80$ uvaksinerte soldater utsatt for smitte. La X være antall av disse som blir syke.

   Hvilken fordeling har X? Regn ut forventningsverdi og varians. Er X tilnærmet normalfordelt? Finn sannsynligheten for at mindre enn 70 soldater blir syke.

b) De amerikanske forskerne gjør et nytt forsøk der $n = 100$ soldater vaksineres og deretter utsettes for smitte. Av disse blir $x = 85$ syke.

   Utfør en ensidig hypotesetest av sannsynligheten p der alternativ hypotese er

$$H_1 : \text{Vaksinen reduserer sannsynligheten for å bli syk.}$$

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

c) Norske forskere vurderer å gjøre egne forsøk med den amerikanske vaksinen. De ønsker derfor å anslå hvor stort utvalg av norske forsøkspersoner man trenger for å oppnå gode resultater. De norske forsøkspersonene skal ikke utsettes for smitte på kunstig vis, men leve som normalt.

   De norske forskerne ønsker å beregne et 95% konfidensintervall for sannsynligheten p for at en norsk forsøksperson blir smittet selv om vedkommende er vaksinert. Dette konfidensintervallet skal deretter sammenlignes med smittetrykket blant alle uvaksinerte forsøkspersoner. Det er forventet at smittetrykket i løpet av forsøket vil være $p_1 = 0{,}05$. Her skal man forstå *smittetrykk* som sannsynligheten for at en uvaksinert forsøksperson pådrar seg smitte i løpet av forsøksperioden.

   De norske forskerne mener resultatene er gode dersom konfidensintervallet for p får en lengde L på 0,01 eller mindre.

   Kravet til utvalgsstørrelse er gitt ved ulikheten

$$n \geq 4\hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{L}\right)^2.$$

   Anslå antall norske forsøkspersoner som minst må vaksineres i forsøket med den amerikanske vaksinen. Begrunn de antagelsene du gjør underveis, og kommenter eventuelle forutsetninger.

Fasit versjon 3:

a) $E(X) = 72$, $\mathrm{Var}(X) = 7{,}2$, tilnærmet normalfordelt, $P(X < 70) \approx 0{,}2280$, (eller bedre $P(X < 70) = 0{,}1757$ med heltallskorreksjon)

b) Kritisk verdi er $z_\alpha = 1{,}645$. Testobservator blir $z = -1{,}667$. Forkast nullhypotese. Den amerikanske vaksinen virker. Likevel er det ikke noe overbevisende resultat gitt at observert z-verdi så nær kritisk verdi. Det bør derfor gjøres flere forsøk.

c) Utvalgsstørrelse minst 7299

Oppgavetekst versjon 4:

Tyske forskere ønsker raskt å utvikle en ny vaksine mot en smittsom virussykdom. Bundeswehr stiller soldater til disposisjon for forskerne. Ved å eksponere vaksinerte og uvaksinerte soldater for smitte kan forskerne finne ut om vaksinen virker eller ikke.

a) Når en uvaksinert tysk soldat utsettes for smitte er sannsynligheten $p_0 = 0{,}6$ for at vedkommende smittes og blir syk. I et forsøk blir $n = 150$ uvaksinerte soldater utsatt for smitte. La X være antall av disse som blir syke.

   Hvilken fordeling har X? Regn ut forventningsverdi og varians. Er X tilnærmet normalfordelt? Finn sannsynligheten for at mindre enn 100 soldater blir syke.

b) De tyske forskerne gjør et nytt forsøk der $n = 250$ soldater vaksineres og deretter utsettes for smitte. Av disse blir $x = 94$ syke.

   Utfør en ensidig hypotesetest av sannsynligheten p der alternativ hypotese er

$$H_1 : \text{Vaksinen reduserer sannsynligheten for å bli syk.}$$

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

c) Norske forskere vurderer å gjøre egne forsøk med den tyske vaksinen. De ønsker derfor å anslå hvor stort utvalg av norske forsøkspersoner man trenger for å oppnå gode resultater. De norske forsøkspersonene skal ikke utsettes for smitte på kunstig vis, men leve som normalt.

   De norske forskerne ønsker å beregne et 99% konfidensintervall for sannsynligheten p for at en norsk forsøksperson blir smittet selv om vedkommende er vaksinert. Dette konfidensintervallet skal deretter sammenlignes med smittetrykket blant alle uvaksinerte forsøkspersoner. Det er forventet at smittetrykket i løpet av forsøket vil være $p_1 = 0{,}04$. Her skal man forstå *smittetrykk* som sannsynligheten for at en uvaksinert forsøksperson pådrar seg smitte i løpet av forsøksperioden.

   De norske forskerne mener resultatene er gode dersom konfidensintervallet for p får en lengde L på 0,01 eller mindre.

   Kravet til utvalgsstørrelse er gitt ved ulikheten

$$n \geq 4\hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{L}\right)^2 .$$

   Anslå antall norske forsøkspersoner som minst må vaksineres i forsøket med den tyske vaksinen. Begrunn de antagelsene du gjør underveis, og kommenter eventuelle forutsetninger.

Fasit versjon 4:

a) $E(X) = 90$, $\mathrm{Var}(X) = 36$, tilnærmet normalfordelt, $P(X < 100) \approx 0{,}9522$, (eller bedre $P(X < 100) = 0{,}9433$ med heltallskorreksjon)

b) Kritisk verdi er $z_\alpha = 1{,}645$. Testobservator blir $z = -7{,}2296$. Forkaster nullhypotese. Den tyske vaksinen virker.

c) Utvalgsstørrelse minst 10191

Oppgave 4. Oppgavetekst versjon 1:

Henrik undersøker endringer i priser på dagligvarer etter stans i svenskehandelen. Før nedstengningen var prisen på smågodt i ulike butikker normalfordelt med en forventningsverdi på 13,85 kr pr hg og et standardavvik på $\sigma = 7{,}20$. Henrik hentet inn disse prisene på smågodt fra 5 forskjellige butikker:

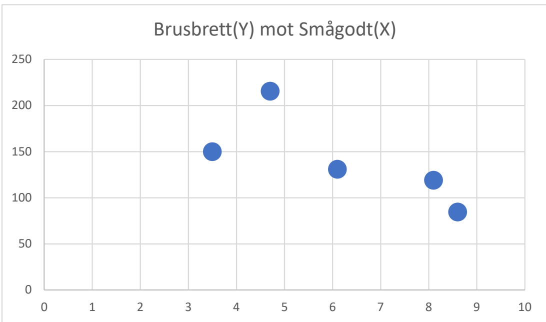
Smågodt X	3.50	6.10	8.10	8.60	4.70

a) Henrik vil undersøke om forventningsverdien til prisen på smågodt har endret seg etter stans i svenskehandelen. Hvilken hypotesetest bør Henrik velge? Begrunn valget. Formuler nullhypotese og alternativ hypotese. Du trenger ikke å utføre testen.

Henrik samlet samtidig inn prisene på brett med brus på boks i de samme butikkene. Han har en hypotese om at det finnes en sammenheng i hvordan prisene på disse varene settes. For å undersøke dette vil han bruke lineær regresjon. Data på prisene på de to varegruppene er som følger:

Butikk nr	1	2	3	4	5
Smågodt X	3.5	6.1	8.1	8.6	4.7
Brusbrett Y	150	131	119	84.5	215.5

Dette plottes i en graf:



Til slutt gir lineær regresjon i Excel følgende utskrift:

SAMMENDRAG (UTDATA)

Regresjonsstatistikk	
Multippel R	0.74079543
R-kvadrat	0.54877786
Justert R-kva	0.39837048
Standardfeil	37.6059222
Observasjone	5

Variansanalyse

	fg	*SK*	*GK*	*F*	*Signifkans-F*
Regresjon	1	5159.88385	5159.88385	3.64860997	0.15210601
Residualer	3	4242.61615	1414.20538		
Totalt	4	9402.5			

	Koeffisienter	*Standardfeil*	*t-Stat*	*P-verdi*	*Nederste 95%*	*Øverste 95%*	*Nedre 95.0%*	*Øverste 95.0%*
Skjæringspun	242.388478	56.1791758	4.31456095	0.02293288	63.6012674	421.175688	63.6012674	421.175688
Smågodt	-16.514271	8.64561072	-1.9101335	0.15210601	-44.028462	10.9999213	-44.028462	10.9999213

b) Hvor stor er den empiriske korrelasjonen? Hva er formelen for regresjonslinja? Er det en signifikant lineær sammenheng mellom prisene på signifikansnivå $\alpha = 0{,}05$?

Løsningsforslag versjon 1:

a) *For å oppnå full score på denne deloppgaven må man påpeke at* ***Henrik har for få målinger***. Både Z-test og T-test har som forutsetning at målingene er *normalfordelte* eller man bruker *setralgrenseteoremet*, altså at man har mange målinger (over 30 er bra). *Derfor må man gjøre* ***egne tilleggsantagelser*** *for å få full score:* (1) enten at målingene er normalfordelte, (2) eller at målingene er normalfordelte og at standardavviket er som før nedstengning. *Dersom man ikke gjør tilleggsantagelser er det* ***koblingen*** *mellom standardavvik og valg av test som er viktig:* (1) Dersom man sier at det oppgitte standardavviket ikke gjelder de målte dataene skal man velge T-test. (2) Dersom man sier at man går ut fra at det oppgitte standardavviket også gjelder nye data kan man bruke Z-test. (3) Dersom man ikke begrunner valg av test ut fra standardavviket gis det **ingen score** for valg av test.

Løsning med T-test:
Henrik har for få målinger til å kunne bruke sentralgrenseteoremet. Han vet heller ikke at de nye målingene av prisen på smågodt er *normalfordelte.* Derfor må vi anta at målingene er normalfordelte for å kunne gjøre en hypotesetest. Standardavviket oppgitt i teksten gjelder fordelingen av smågodtprisene *før* nedstengning i svenskehandelen. Derfor er standardavviket ukjent, og Henrik bør velge en T-test. Han ønsker å påvise endring i prisen, derfor blir dette alternativ hypotese. Han bruker en tosidig test. Hypotesene blir:

$$H_0 : \text{Forventningsverdien til smågodtprisen er som før.} \quad \mu = 13{,}85$$
$$H_1 : \text{Forventningsverdien er endret.} \quad \mu \neq 13{,}85$$

Henrik bør være åpen på at han har lite data når han trekker konklusjonen på denne testen. Hvis nullhypotesen forkastes kan han konkludere at forventningsverdien til prise er endret eller at fordelingen ikke lengre er normalfordelt.

Løsning med Z-test:
Henrik har for få målinger til å kunne bruke sentralgrenseteoremet. Han vet heller ikke at de nye målingene av prisen på smågodt er *normalfordelte.* Derfor må vi anta at målingene er normalfordelte for å kunne gjøre en hypotesetest. Standardavviket oppgitt i teksten gjelder fordelingen av smågodtprisene *før* nedstengning i svenskehandelen. Siden Henrik allerede gjør tilleggsantagelser kan han også anta at *standardavviket ikke endres etter nedstengning.* Da kan han si at han antar at standardavviket har den kjente verdien $\sigma = 7{,}20$ og bruke en Z-test. Han ønsker å påvise endring i prisen, derfor blir dette alternativ hypotese. Han bruker en tosidig test. Hypotesene blir:

$$H_0 : \text{Forventningsverdien til smågodtprisen er som før.} \quad \mu = 13{,}85$$
$$H_1 : \text{Forventningsverdien er endret.} \quad \mu \neq 13{,}85$$

Henrik bør være åpen på at han har lite data når han trekker konklusjonen på denne testen. Hvis nullhypotesen forkastes kan han konkludere at forventningsverdien til prise er endret, at standardavviket er endret eller at fordelingen ikke lengre er normalfordelt.

b) Den empiriske korrelasjonen finner vi som *Multippel R* i Excelutskriften, men vi ser at fortegnet mangler. Avlesning gir

$$\underline{\underline{R = -0{,}74079543}}.$$

Det er negativt fortegn siden regresjonslinja er avtagende.
Vi leser også av koeffisientene for regresjonslinja. De er:

$$\hat{\alpha} \approx 242{,}388$$
$$\hat{\beta} \approx -16{,}514$$

Dermed har regresjonslinja formelen

$$\underline{\underline{\hat{y} \approx 242{,}388 - 16{,}514 \cdot x}}.$$

Vi avleser p-verdien 0,15210601. Denne er større enn signifikansnivået $\alpha = 0{,}05$. Derfor har Henrik <u>ikke påvist en signifikant lineær sammenheng</u> mellom prisen på smågodt og prisen på brusbrett.

Oppgavetekst versjon 2:
Henrik undersøker endringer i priser på dagligvarer etter stans i svenskehandelen. Før nedstengningen var prisen på kjøttdeig i ulike butikker normalfordelt med en forventningsverdi på 140,00 kr pr kg og et standardavvik på $\sigma = 25{,}00$. Henrik hentet inn disse prisene på kjøttdeig fra 5 forskjellige butikker:

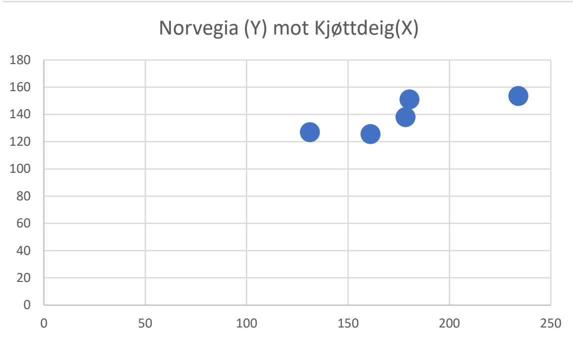
Kjøttdeig X	234,00	161,10	178,40	131,20	180,30

a) Henrik vil undersøke om forventningsverdien til prisen på kjøttdeig har endret seg etter stans i svenskehandelen. Hvilken hypotesetest bør Henrik velge? Begrunn valget. Formuler nullhypotese og alternativ hypotese. Du trenger ikke å utføre testen.

Henrik samlet samtidig inn prisene på Norvegiaost i de samme butikkene. Han har en hypotese om at det finnes en sammenheng i hvordan prisene på disse varene settes. For å undersøke dette vil han bruke lineær regresjon. Data på prisene på de to varegruppene er som følger:

Butikk nr	1	2	3	4	5
Kjøttdeig X	234,00	161,10	178,40	131,20	180,30
Norvegia Y	153,50	125,50	138,00	127,00	151,00

Dette plottes i en graf:



Til slutt gir lineær regresjon i Excel følgende utskrift:

SAMMENDRAG (UTDATA)

Regresjonsstatistikk	
Multippel R	0.832922099
R-kvadrat	0.693759223
Justert R-kvadrat	0.591678965
Standardfeil	8.340725172
Observasjoner	5

Variansanalyse

	fg	*SK*	*GK*	*F*	*Signifkans-F*
Regresjon	1	472.7969108	472.7969108	6.796213405	0.079894574
Residualer	3	208.7030892	69.5676964		
Totalt	4	681.5			

	Koeffisienter	*Standardfeil*	*t-Stat*	*P-verdi*	*Nederste 95%*	*Øverste 95%*
Skjæringspunkt	87.62638312	20.05628502	4.369023627	0.022178162	23.79833296	151.4544333
Kjøttdeig	0.290246423	0.111335425	2.606954814	0.079894574	-0.06407259	0.644565436

b) Hvor stor er den empiriske korrelasjonen? Hva er formelen for regresjonslinja? Er det en signifikant lineær sammenheng mellom prisene på signifikansnivå $\alpha = 0{,}05$?

Fasit versjon 2:

a) Se løsningsforslag over: For få målinger. Gjør tilleggsantagelser om normalfordelte måliunger. Enten uten kjent standardavvik velg tosidig T-test. Eller med antatt kjent standardavvik velg tosidig Z-test.

b) $R = 0{,}8329$, $\hat{y} = 87{,}626 + 0{,}290 \cdot x$, $p = 0{,}079$, ikke signifikant.

Oppgavetekst versjon 3:

Henrik undersøker endringer i priser på dagligvarer etter stans i svenskehandelen. Før nedstengningen var prisen på knekkebørd i ulike butikker normalfordelt med en forventningsverdi på 112,00 kr pr kg og et standardavvik på $\sigma = 34{,}00$.

Henrik hentet inn disse prisene på knekkebrød fra 5 forskjellige butikker:

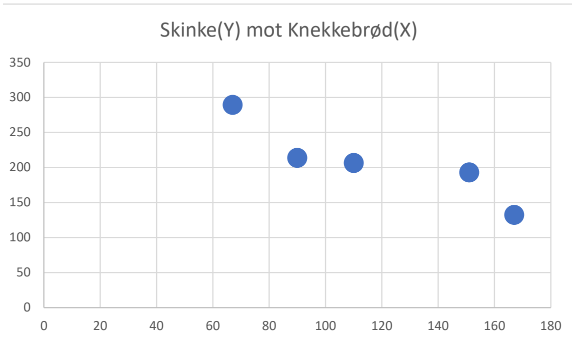
Knekkebrød X	151,00	67,00	110,00	90,00	167,00

a) Henrik vil undersøke om forventningsverdien til prisen på knekkebrød har endret seg etter stans i svenskehandelen. Hvilken hypotesetest bør Henrik velge? Begrunn valget. Formuler nullhypotese og alternativ hypotese. Du trenger ikke å utføre testen.

Henrik samlet samtidig inn prisene på skinke i de samme butikkene. Han har en hypotese om at det finnes en sammenheng i hvordan prisene på disse varene settes. For å undersøke dette vil han bruke lineær regresjon. Data på prisene på de to varegruppene er som følger:

Butikk nr	1	2	3	4	5
Knekkebrød X	151,00	67,00	110,00	90,00	167,00
Skinke Y	192,90	289,50	206,50	213,60	132,50

Dette plottes i en graf:



Til slutt gir lineær regresjon i Excel følgende utskrift:

SAMMENDRAG (UTDATA)

Regresjonsstatistikk	
Multippel R	0.909842739
R-kvadrat	0.82781381
Justert R-kvadrat	0.770418413
Standardfeil	26.8911043
Observasjoner	5

Variansanalyse

	fg	*SK*	*GK*	*F*	*Signifkans-F*
Regresjon	1	10429.72553	10429.72553	14.42300005	0.032053307
Residualer	3	2169.394472	723.1314907		
Totalt	4	12599.12			

	Koeffisienter	*Standardfeil*	*t-Stat*	*P-verdi*	*Nederste 95%*	*Øverste 95%*
Skjæringspunkt	350.4928901	39.65125176	8.839390297	0.00305174	224.3049104	476.6808698
Knekkebrød	-1.226434958	0.322936191	-3.797762506	0.032053307	-2.254162046	-0.19870787

b) Hvor stor er den empiriske korrelasjonen? Hva er formelen for regresjonslinja? Er det en signifikant lineær sammenheng mellom prisene på signifikansnivå $\alpha = 0{,}05$?

Fasit versjon 3:

a) Se løsningsforslag over: For få målinger. Gjør tilleggsantagelser om normalfordelte måliunger. Enten uten kjent standardavvik velg tosidig T-test. Eller med antatt kjent standardavvik velg tosidig Z-test.

b) $R = -0{,}9098$, $\hat{y} = 350{,}493 - 1{,}226 \cdot x$, $p = 0{,}032$, signifikant lineær sammenheng.

Oppgavetekst versjon 4:

Henrik undersøker endringer i priser på dagligvarer etter stans i svenskehandelen. Før nedstengningen var prisen på kyllingfilet i ulike butikker normalfordelt med en forventningsverdi på 127,30 kr pr kg og et standardavvik på $\sigma = 18{,}90$.

Henrik hentet inn disse prisene på kyllingfilet fra 5 forskjellige butikker:

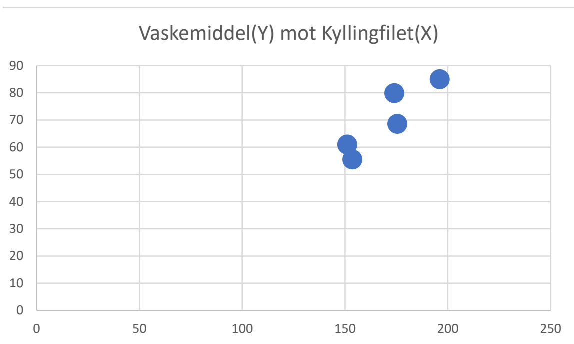
Kyllingfilet X	175,40	153,50	196,00	174,00	151,10

a) Henrik vil undersøke om forventningsverdien til prisen på kyllingfilet har endret seg etter stans i svenskehandelen. Hvilken hypotesetest bør Henrik velge? Begrunn valget. Formuler nullhypotese og alternativ hypotese. Du trenger ikke å utføre testen.

Henrik samlet samtidig inn prisene på vaskemiddel i de samme butikkene. Han har en hypotese om at det finnes en sammenheng i hvordan prisene på disse varene settes. For å undersøke dette vil han bruke lineær regresjon. Data på prisene på de to varegruppene er som følger:

Butikk nr	1	2	3	4	5
Kyllingfilet X	175,40	153,50	196,00	174,00	151,10
Vaskemiddel Y	68,60	55,50	85,00	79,00	61,00

Dette plottes i en graf:



Til slutt gir lineær regresjon i Excel følgende utskrift:

SAMMENDRAG (UTDATA)

Regresjonsstatistikk	
Multippel R	0.911318766
R-kvadrat	0.830501893
Justert R-kvadrat	0.774002524
Standardfeil	5.900511936
Observasjoner	5

Variansanalyse

	fg	*SK*	*GK*	*F*	*Signifkans-F*
Regresjon	1	511.7718767	511.7718767	14.69931274	0.031276521
Residualer	3	104.4481233	34.8160411		
Totalt	4	616.22			

	Koeffisienter	*Standardfeil*	*t-Stat*	*P-verdi*	*Nederste 95%*	*Øverste 95%*
Skjæringspunkt	-34.64549614	27.42156534	-1.263439768	0.295689329	-121.9131555	52.62216317
Kyllingfilet	0.615561742	0.16055473	3.833968276	0.031276521	0.104604934	1.12651855

b) Hvor stor er den empiriske korrelasjonen? Hva er formelen for regresjonslinja? Er det en signifikant lineær sammenheng mellom prisene på signifikansnivå $\alpha = 0{,}05$?

Fasit versjon 4:

a) Se løsningsforslag over: For få målinger. Gjør tilleggsantagelser om normalfordelte måliunger. Enten uten kjent standardavvik velg tosidig T-test. Eller med antatt kjent standardavvik velg tosidig Z-test.

b) $R = 0{,}9113$, $\hat{y} = -34{,}645 + 0{,}616 \cdot x$, $p = 0{,}031$, signifikant lineær sammenheng.