Oppgaver brukt i STA100 Sannsynlighetsregning og statistikk

Oppgave 1

Kari skal gjøre en undersøkelse blant studentene ved Universitetet i Stavanger for å kartlegge hva studentene synes om ulike spørsmål knyttet til studiekvalitet, velferdstilbud, etc. Hun innser fort at for hennes formål vil det være altfor omfattende å spørre alle studentene hva de mener om disse spørsmålene og hun bestemmer seg derfor for å velge ut en gruppe på ca 100 studenter og basere sine konklusjoner på tilbakemeldingene disse gir.

Hun vurderer tre forskjellige måter å utføre undersøkelsen sin på:

1. Møte opp på en forelesinger der det er rundt 100 studenter og få alle disse til å svare på spørsmålene hennes.
2. Gå rundtomkring på Universitetsområdet og spørre tilfeldige studenter hun treffer inntil hun har fått spurt ca 100 studenter.
3. Trekke ut 100 tilfeldige navn fra en liste over alle registrerte studenter ved universitetet og så oppsøke disse (telefon e.l.) og spørre dem hva de mener.

a) Hva er populasjonen og hva er utvalget i denne undersøkelsen?

b) Kommenter fordeler og ulemper ved de tre forskjellige måtene å utføre undersøkelsen på. Hvilken måte er best?

Oppgave 2

12 målinger av reaksjonstiden for en kjemisk reaksjon (i minutter) ga resultatene:

17 22 37 36 22 29 42 23 22 28 36 33

For disse dataene:

a) Regn ut gjennomsnittet, medianen og modus.

b) Regn ut utvalgsvariansen og utvalgsstandardavviket.

c) Lag et histogram over dataene.

d) Regn ut nedre kvartil, øvre kvartil, kvartilbredden og variasjonsbredden.

e) Lag et boksplott over dataene.

Oppgave 3

Betrakt følgende fire datasett:

i) 1, 2, 3, 4, 5, 6, 7, 8

ii) 1, 1, 1, 1, 8, 8, 8, 8

iii) 1, 1, 4, 4, 5, 5, 8, 8

iv) -6, -3, 0, 3, 6, 9, 12, 15

Alle disse datasettene har samme gjennomsnitt. Prøv først,*uten å gjøre noen beregninger*, å ranger disse datasettene etter utvalgsvarians (fra minst til størst utvalgsvarians). Regn så ut utvalgsvariansene og sjekk om intuisjonen din stemte.

Oppgave 4

En produksjonsbedrift har utført en kartlegging av hvor lang tid ulike arbeidsoppgaver tar. For en bestemt monteringsjobb gav 8 målinger av tidsbruken i timer resultatet gitt under.

Måling i	1	2	3	4	5	6	7	8
Tidsbruk x_i	2.6	2.1	2.3	2.6	2.5	2.0	2.4	9.9

a) Regn ut gjennomsnittet og medianentil målingene av tidsbruken. Tegn inn dataene samt gjennomsnittet og medianen for dataene på tallinja. Kommenter forskjellen mellom gjennomsnittet og medianen.

Bedriften undret seg over den høye verdien i måling nummer 8, og ved en nærmere undersøkelse av dataene viste det seg at feil verdi var blitt notert i denne målingen. Det var heller ikke lenger kjent hva den korrekte verdien var, så i den videre bruken av dataene så man bort fra denne målingen og brukte kun de 7 første målingene.

Bruk kun de 7 første målingene når du svarer på spørsmålene under.

b) Regn ut gjennomsnittet og medianen til målingene av tidsbruken. Regn også ut utvalgsvarians og utvalgsstandardavvik. Anslå et intervall som vil kunne inneholde omtrent 95% av alle målinger av tidsbruk for denne jobben.

Oppgave 5

Bensinprisen på to bensinstasjoner i Sandnes er registrert for ulike dager. La X være bensinpris på den ene stasjonen og Y være bensinpris på den andre stasjonen en tilfeldig dag. Registreringer av X og Y på 10 tilfeldig valgte dager er vist i tabellen under.

dag i	1	2	3	4	5	6	7	8	9	10
x_i	14.89	14.39	13.20	15.35	15.10	14.39	14.96	15.15	14.69	13.57
y_i	14.99	14.39	13.65	15.25	14.99	14.09	14.66	15.25	14.36	13.57

Det oppgis at $\sum_{i=1}^{10} x_i = 145.69$, $\sum_{i=1}^{10} y_i = 145.20$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 4.436$, og $\sum_{i=1}^{10}(y_i - \bar{y})^2 = 3.414$.

a) Regn ut gjennomsnittene, $\bar{x}$ og $\bar{y}$, og utvalgsstandardavvikene, s_x og s_y, for bensinprisene på de to bensinstasjonene. Kommenter resultatene. (Hint: Du kan spare mye arbeid ved å bruke de oppgitte resultatene.)

b) Plott bensinprisene på de to stasjonene mot hverandre i et spredningsplott (scatterplott). Kommenter kort hva dette plottet viser.

Oppgave 6

To oljereservoar betraktes, 1 og 2. Sannsynligheten for å finne olje i reservoar 1 er 0.6, og sannsynligheten for å finne olje i reservoar 2 er 0.7. Sannsynligheten for å finne olje i begge reservoarene er 0.45.

a) Finn sannsyligheten for at man finner olje i minst ett av reservoarene.

b) Finn sannsyligheten for at man ikke finner olje i noen av reservoarene.

c) Finn sannsyligheten for at man finner olje kun i reservoar 1.

d) Finn sannsyligheten for at man finner olje i kun reservoar 1 eller kun reservoar 2 (ikke i begge.

Oppgave 7

I denne oppgaven skal vi se på utviklingen til en børsindeks for tre etterfølgende dager. La B_1 være hendelsen at børsindeksen stiger dag 1, la B_2 være hendelsen at børsindeksen stiger dag 2 og la B_3 være hendelsen at børsindeksen stiger dag 3. Basert på lang tids observasjon av indeksen har man følgende informasjon:

$$P(B_1) = P(B_2) = P(B_3) = 0.54, \qquad P(B_2|B_1) = P(B_3|B_2) = 0.65, \qquad P(B_3|B_1 \cap B_2) = 0.73$$

a) Vis at $P(B_1 \cap B_2) = 0.351$.
Finn $P(B_1 \cup B_2)$.

b) Si med ord hva $B_1 \cap B_2$ og $B_1 \cup B_2$ betyr.
Finn $P(B_1 \cap \overline{B}_2)$, dvs sannsynligheten for at indeksen stiger dag 1 og synker dag 2.

c) Vis at $P(B_1 \cap B_2 \cap B_3) = 0.25623$ og finn $P(\overline{B}_1 \cap B_2 \cap B_3)$.

Oppgave 8

En markedsundersøkelse i en by viser at i løpet av en uke har 18% av innbyggerne sett et TV-program om naturvitenskap eller teknologi, 12% har lest i et tidsskrift om emnet, og 10% har gjort begge deler.

La A være hendelsen at en tilfeldig innbygger har sett TV-program om naturvitenskap/teknologi og la B være hendelsen at en tilfeldig innbygger har lest et tidsskrift om emnet.

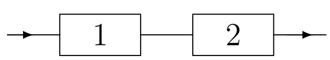
a) Formuler opplysningen gitt over som sannsynligheter knyttet til hendelsene A og B.
Forklar med ord hva $P(A|B)$ betyr, og regn ut sannsynligheten.
Er A og B uavhengige hendelser? Begrunn.

b) Hva er sannsynligheten for at en person som har lest et tiddskrift om naturvitenskap/teknologi ikke har sett et TV-program om emnet?
Hva er sannsynligheten for at en person som ikke har sett et TV-program om naturvitenskap/teknologi, har lest et tiddskrift om emnet?

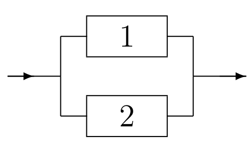
Oppgave 9

Et teknisk system har to komponenter som svikter uavhengig av hverandre. Anta at $P(\text{komponent 1 svikter}) = 0.1$ og $P(\text{komponent 2 svikter}) = 0.2$. Finn sannsynligheten for at systemet ikke svikter i de følgende situasjonene:

a) Komponetene er koplet i serie; dvs. systemet virker bare når begge komponentene virker.



b) Komponetene er koplet i parallell; dvs. systemet virker dersom minst en av komponentene virker.



Oppgave 10

I området der Per bor er det 100 husstander. En bedrift sender ut en vareprøve til 10 tilfeldig valgte husstander i dette området.

a) Hva er sannsynligheten for at Per mottar vareprøven?

Kari er nærmeste naboen til Per.

b) Hva er sannsynligheten for at både Per og Kari mottar vareprøven?

c) Hva er sannsynligheten for at minst en av de to naboene mottar vareprøven?

Oppgave 11

Spillet Lotto foregår ved at 7 av tallene fra 1 til 34 trekkes ut tilfeldig. De 7 tallene som trekkes ut kalles vinnerrekken.

a) Dersom man har tippet 5 ulike enkeltrekker i Lotto (dvs 5 ulike kombinasjoner av 7 tall), hva er sannsynligheten for å få 7 riktige?

Å tippe system i Lotto vil si å velge ut 8, 9, 10, 11 eller 12 av tallene 1 til 34.

b) Dersom man har tippet system med 8 tall, hva er sannsynligheten for å få 7 riktige? Hva er blir sannsynligheten for å få 7 riktige med system med 9 tall?

I tillegg til vinnerrekken på 7 tall trekkes det ett tilleggstall i Lotto. (Nest største gevinst går til de som har tippet 6 av de 7 vinnertallene og tilleggstallet.)

c) Dersom man har tippet system med 8 tall, hva er sannsynligheten for å få 6 riktige og tilleggstallet?

Oppgave 12

Et entreprenørfirma deltar i konkurransen om to store kommunale byggeoppdrag, henholdsvis bygging av en ny barnehage og bygging av eldreboliger. Vi definerer følgende hendelser:

B = firmaet får oppdraget med å bygge barnehagen.

E = firmaet får oppdraget med å bygge eldreboligene.

Ut fra opplysningene firmaet sitter med regner de med at sannsynligheten for at de får oppdraget med å bygge barnehagen er 0.5, sannsynligheten for at de får oppdraget med å bygge eldreboligene er 0.6, og sannsynligheten for at de får begge oppdragene er 0.4.
Dvs $P(B) = 0.5$, $P(E) = 0.6$ og $P(B \cap E) = 0.4$

a) Er hendelsene B og E uavhengige? Begrunn svaret.

b) Hva er sannsynligheten for at firmaet får minst ett av oppdragene?
Hva er sannsynligheten for at firmaet får ingen av oppdragene?

c) Hva er sannsynligheten for at firmaet også får oppdraget med å bygge barnehagen dersom de får oppdraget med å bygge eldreboligene?
Hva er sannsynligheten for at firmaet får oppdraget med å bygge barnehagen dersom de ikke får oppdraget med å bygge eldreboligene?

Firmaet regner med at oppdraget med å bygge barnehagen vil gi et overskudd på 0.25 millioner kroner dersom de får oppdraget, mens oppdraget med å bygge eldreboligene vil gi et overskudd på 0.3 millioner kroner dersom de får oppdraget.

La O = totalt overskudd fra de to mulige oppdragene.

d) Vis at fordelingen til O blir:

o	0	0.25	0.3	0.55
$P(O = o)$	0.3	0.1	0.2	0.4

e) Regn ut forventa totalt overskudd.
Regn også ut standardavviket til det totale overskuddet.

Oppgave 13

En satelitt i bane rundt jorda har tre panel med solceller. Alle må være aktive for at det skal produseres tilstrekkelig energi. Panelene fungerer uavhengig av hverandre. Sannsynligheten for at et panel svikter under et oppdrag er 0.02.

a) Hva er sannsynligheten for at det vil være tilstrekkelig energi under hele oppdraget?

Oppgave 14

Et bilutleiefirma i en by har to filialer. Den ene filialen har har fire biler og fordelingen til X=antall biler utleid fra denne filialen er:

x	0	1	2	3	4
$P(X = x)$	0.15	0.20	0.25	0.20	0.20

Den andre filialen har tre biler og fordelingen til Y=antall biler utleid fra denne filialen er

y	0	1	2	3
$P(Y = y)$	0.15	0.40	0.35	0.10

a) Finn $P(X = 3)$, $P(X < 3)$, $P(X \leq 3)$, $P(Y \geq 2)$ og $P(Y > 2)$.

b) Finn $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{Var}(X)$ og $\mathrm{Var}(Y)$. Gi en pratisk tolkning av $\mathrm{E}(X)$ og $\mathrm{E}(Y)$.

De faste utgiftene per dag for å drive den første filialen er 1500,-, mens de faste utgiftene per dag for å drive den andre filialen er 1100,-. For begge filialene er inntektene 750,- per døgn per bil de har utleid.

La F_X være fortjenesten (inntekt minus utgifter) per dag fra den første filialen og F_Y være fortjenesten per dag fra den andre filialen.

c) Finn $\mathrm{E}(F_X)$, $\mathrm{E}(F_Y)$, $\mathrm{Var}(F_X)$ og $\mathrm{Var}(F_Y)$. Hva forteller disse størrelsene oss noe om? Hvilken av filialene går best?

Oppgave 15

La X være en stokastisk variabel med $\mathrm{E}(X) = 0$ og $\mathrm{SD}(X) = 2$, og la Y være en stokastisk variabel med $\mathrm{E}(Y) = -1$ og $\mathrm{SD}(Y) = 4$.

a) Finn $\mathrm{E}(X - Y)$ og $\mathrm{E}(X + Y)$.

b) Finn $\mathrm{Var}(X - Y)$ og $\mathrm{Var}(X + Y)$ når X og Y er uavhengige.

c) Finn $\mathrm{E}(\frac{1}{2}X + \frac{1}{2}Y)$ og $\mathrm{Var}(\frac{1}{2}X + \frac{1}{2}Y)$ når X og Y er uavhengige.

d) Gjenta b) dersom vi istedenfor uavhengighet har at $\mathrm{Cov}(X, Y) = 1$.

Oppgave 16

La X være antall biler som blir solgt hos en bilforhandler i løpet av én dag. Sannsynlighetsfordelingen til X er gitt i tabellen under. Vi kan anta at salget på ulike dager er uavhengig av hverandre.

x	0	1	2
$P(X = x)$	0.65	0.25	0.1

a) Finn forventningen og standardavviket til X.

La X_1 og X_2 være antall solgte biler på henholdsvis dag 1 og dag 2, og la $Y = X_1 + X_2$ være totalt antall solgte biler i løpet av de to dagene.

b) Finn sannsynlighetsfordelingen til Y.
Finn $P(Y > 2)$. Regn ut forventet totalt salg $\mathrm{E}(Y)$ på to måter.

Bilforhandleren har åpent fem dager i uken. Ukentlig salg blir da $Z = X_1 + X_2 + X_3 + X_4 + X_5$.

c) Finn $P(Z = 0)$ og $P(Z = 1)$.

Oppgave 17

En bedrift produserer ventiler til bruk på offshore-installasjoner. Alle produserte ventiler trykktestes, og ventiler som ikke består trykktesten kan ikke brukes. Fra erfaring vet man at 7% av ventilene ikke består trykktesten, og at det er uavhengig fra ventil til ventil om de består trykktesten.

Nå skal 26 slike ventiler testes. La X være antallet av disse ventilene som ikke består trykktesten.

a) Forklar hvorfor X er binomisk fordelt.

b) Finn:
- Sannsynligheten for at akkurat 2 ventiler ikke består testen.
- Sannsynligheten for at færre enn 2 ventiler ikke består testen.
- Sannsynligheten for at minst en ventil ikke består testen.
- Sannsynligheten for at mellom 1 og 4 ventiler ikke består testen (fra og med 1 til og med 4).
- Forventet antall ventiler som ikke består testen.
- Forventet antall ventiler som består testen.
- Sannsynligheten for at all ventilene består testen.

Oppgave 18

En studentavis hevder at 80% av studentene støtter avisens syn i en bestemt sak. Et tilfeldig utvalg på 10 studenter blir tatt, og det viser seg at 4 av disse er enige med avisen.

a) Regn ut hva sannsynligheten for å få 4 eller færre som er enige med avisen i et tilfeldig utvalg på 10 blir dersom avisen sin påstand er rett. Kommenter ut fra dette troverdigheten til avisens påstand.

Oppgave 19

Levetiden til en LED-lysdiode kan antas eksponentialfordelt med forventning 10000 timer.

a) Ta utgangspunkt i sannsynlighetstettheten, og regn ut sannsynligheten for at en slik diode fungerer i: i) Mindre enn 5000 timer, ii) Mer enn 20000 timer, iii) Mellom 5000 og 10000 timer.

b) Ta utgangspunkt i den kumulative fordelingsfunksjonen og regn ut de samme sannsynlighetene som i punkt b).

I en lampe brukes to lysdioder av denne typen. Lampen gir lys så lenge minst en av de to diodene fungerer. Anta at diodene fungerer uavhengige av hverandre.

c) Finn sannsynligheten for at lampen gir lys i mer enn 20000 timer.

Oppgave 20

Funksjonstiden til en bestemt type elektroniske styreenheter er eksponentialfordelt med forventning 5 år.

a) Regn ut sannsynligheten for at en slik styreenhet fungerer i mer enn 2 år.

b) En bedrift tar i bruk 10 slike styreenheter. Hva er sannsynligheten for at minst 8 av enhetene fremdeles fungerer etter 2 år? (Anta at enhetene fungerer uavhengig av hverandre.)

Oppgave 21

Dersom X er en normalfordelt stokastisk variabel med forventning $\mu = 100$ og varians $\sigma^2 = 64$, finn:

a) $P(X < 107)$	b) $P(X < 97)$	c) $P(X > 110)$
d) $P(95 < X < 106)$	e) $P(60 < X < 108)$	f) b slik at $P(X > b) = 0.8708$

Oppgave 22

Et bryggeri har en tappemaskin som tapper øl på bokser som skal inneholde 0.5 liter øl. Tappemaskinen er imidlertid ikke helt presis, så mengden øl som blir tappet vil variere litt fra boks til boks. Vi antar at ølmengden som blir tappet på en boks er normalfordelt med forventning $\mu = 0.5$ liter og standardavvik $\sigma = 0.015$ liter. Vi antar også at mengden øl tappet på ulike bokser er uavhengig.

a) Vis at sannsynligheten for at det tappes mindre enn 0.49 liter øl på en ølboks er 0.25.

b) Regn ut sannsynligheten for at det tappes mellom 0.47 og 0.53 liter øl på en ølboks.

c) Finn den verdien k som er slik at bryggeriet kan garantere at 99% av alle ølbokser inneholder minst k liter øl.

d) Finn sannsynligheten for at minst 2 av 6 ølbokser inneholder mindre enn 0.49 liter øl.

e) Finn sannsynligheten for at minst 20 av 60 ølbokser inneholder mindre enn 0.49 liter øl.

f) Regn ut sannsynligheten for at gjennomsnittlig mengde øl i 6 bokser er mindre enn 0.49 liter.

g) Regn ut sannsynligheten for at gjennomsnittlig mengde øl i 60 bokser er mindre enn 0.49 liter.

h) Sammenlign svarene i a), f) og g) og kommenter.

Oppgave 23

En bedrift produserer en type pumper som brukes i offshoreinstallasjoner. Alle ferdigproduserte pumper funksjonstestes, og på noen av pumpene finner man en eller flere feil. Basert på lang tids erfaring vet man at antall feil man finner per pumpe, X, har sannsynlighetsfordelingen gitt i tabellen under.

x	0	1	2	3
$P(X = x)$	0.93	0.04	0.02	0.01

a) Regn ut $\mathrm{E}(X)$ og $\mathrm{Var}(X)$.

Nå skal 100 uavhengige pumper funksjonstestes.

b) Hva er forventet totalt antall feil i 100 pumper?
Hva er sannsynligheten for at totalt antall feil i 100 pumper er mindre enn 10?

Oppgave 24

Antall bakterier i en volumenhet av en vannprøve er Poisson-fordelt med forventning 2.5.

a) I en slik vannprøve (på en volumenhet), hva er sannsynligheten for:
i) Nøyaktig to bakterier, ii) Færre enn tre bakterier, iii) Minst fire bakterier

b) I en vannprøve på tre volumenheter, hva er sannsynligheten for:
i) Nøyaktig to bakterier, ii) Færre enn tre bakterier, iii) Minst fire bakterier

c) I en vannprøve på ti volumenheter, hva er sannsynligheten for mer enn 20 bakterier? (Hint: Tilnærming til normalfordeling)

Oppgave 25

En basketballspiller skårer på 80 % av skuddene. Anta at det er uavhengig fra skudd til skudd om det blir skåring. På 120 skudd, hva sannsynligheten for at hun skårer

a) Mindre enn 90 ganger? b) 105 eller flere ganger? c) Nøyaktig 100 ganger?

Oppgave 26

Det er gjort flere prøver av smeltepunktet for en ny legering. Måleapparatets presisjon oppgis ved at standardavviket til målinger er 7 grader. (Dvs.: en måling kan oppfattes som utfall av en stokastisk variabel med standardavvik 7 og forventning lik det virkelige smeltepunktet.)

a) Dersom vi skal ta 80 målinger, hva er sannsynligheten for at gjennomsnittet av målingene ikke vil avvike fra det virkelige smeltepunktet med mer enn 1.54 grader ?

Oppgave 27

La X være høyden til en tilfeldig mann og la Y være høyden til en tilfeldig kvinne. Anta at $X \sim \mathrm{N}(180, 6)$ og at $Y \sim \mathrm{N}(167, 6)$ målt i cm.

a) Regn ut $P(X < Y)$ og forklar hva dette i praksis betyr.

b) Regn ut $P(X + Y > 350)$ og forklar hva dette i praksis betyr.

c) Regn ut $P((X + Y)/2 < 170)$ og forklar hva dette i praksis betyr.

Oppgave 28

a) Forklar med ord hva μ, $\bar{x}$, σ^2 og s^2 er.
Hva er forskjellen på μ og $\bar{x}$?
Hva er sammenhengen mellom μ og $\bar{x}$?
Forklar med ord hva det i praksis betyr at en estimator er forventningsrett.

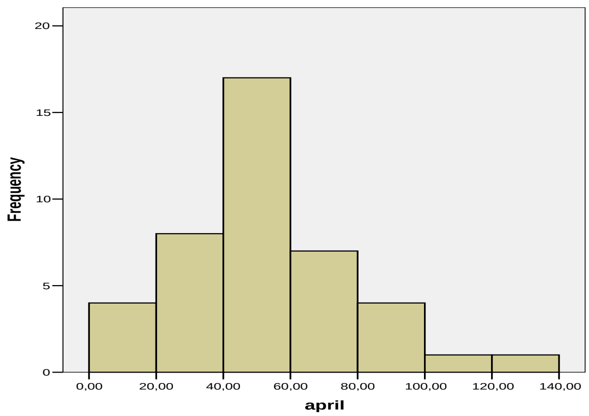
Oppgave 29

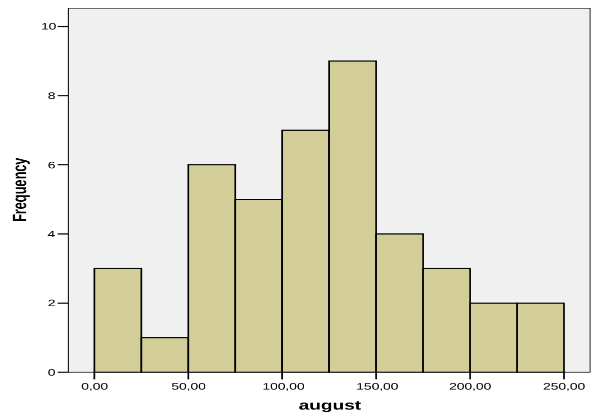
Total nedbørsmengde (i mm) på Sola for april og august måned er registrert i de 41 årene 1958 - 1998. Følende størrelser er beregnet for datasettene:

måned	gj.snitt	utv.stand.avvik	min.	1. kvartil	median	3. kvartil	maks.
april	52.7	26.4	3.0	40.0	48.9	64.1	126.3
august	119.1	57.4	4.9	77.5	118.2	149.5	241.2

a) Hva er estimert forventet nedbørsmengde i april og i august?
Hva er estimert standardavvik for nedbørsmengden i april og i august?

Histogram over dataene:





b) Synes det rimelig å bruke normalantakelse (antagelse om normalfordeling) for disse dataene? Hvilke av konfidensintervallene det spørres om i punkt c) og d) under krever normalantagelse?

c) Finn 95% konfidensintervall for forventet nedbørsmengde i april og i august. Sammenlign de to intervallene og kommenter kort.

d) Finn 95% konfidensintervall for standardavviket for nedbørsmengde i april og i august.

Oppgave 30

Man vil undersøke oppslutningen i befolkningen om en bestemt sak. For å finne ut hvor stor andel som er for saken, skal det gjennomføres en meningsmåling. Det skal benyttes $n = 1200$ tilfeldig utvalgte. La p være andelen av befolkningen som er for saken, og la X være antall av de 1200 i utvalget som er for.

a) Gi en forklaring på hvorfor vi kan bruke at X er binomisk fordelt.

b) Dersom $p = 0.52$, hva er sannsynligheten for at undersøkelsen vil indikere at det ikke er flertall for saken i befolkningen?

I realiteten er p ukjent, det er derfor undersøkelsen utføres. Resultat av undersøkelsen ble at 740 av de 1200 i utvalget var for.

c) Estimer andelen, p, som er for i befolkningen, og beregn et tilnærmet 90 % konfidensintervall for andelen.

d) Dersom man gjør en slik meningsmåling hver måned og hver gang beregner et slik 90 % konfidensintervall, hvor stor andel av intervallene vil ikke dekke den virkelige verdien av p ?

Oppgave 31

Tappemaskinen omhandlet i oppgave 22 gjennomgår en overhaling, og man er usikker på om maskinen etterpå er korrekt innstilt. Vi antar derfor nå at forventet mengde øl per boks, μ, er ukjent. Vi antar som før at standardavviket er $\sigma = 0.015$ liter, og at mengden øl tappet på ulike bokser er uavhengig og normalfordelt.

For å kartlegge hvordan maskinen er innstilt gjøres en undersøkelse hvor man måler mengden øl i 24 bokser. Resultatet av denne undersøkelsen ble at gjennomsnittlig mengde øl i de 24 boksene var 0.497 liter.

a) Utled et 95% konfidensintervall for μ, og regn ut et tallsvar fra informasjonen gitt over.

I en ny undersøkelsen ble mengden øl i 12 bokser målt, og man fant her at gjennomsnittlig mengde øl i de 12 boksene var 0.494 liter.

Vi ønsker nå å kombinere informasjonen fra de to undersøkelsene til en best mulig estimator for μ. La $X_1, \ldots, X_{24}$ betegne målt mengde øl i hver av de 24 boksene i den første undersøkelsen og la $Y_1, \ldots, Y_{12}$ betegne målt mengde øl i hver av de 12 boksene fra den andre undersøkelsen. La videre $\bar{X} = \frac{1}{24}\sum_{i=1}^{24} X_i$ og $\bar{Y} = \frac{1}{12}\sum_{i=1}^{12} Y_i$. Følgende to estimatorer som bruker informasjonen fra begge de to undersøkelsene er foreslått:

$$\hat{\mu}_1 = \frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y} \qquad \text{og} \qquad \hat{\mu}_2 = \frac{2}{3}\bar{X} + \frac{1}{3}\bar{Y}$$

b) Vis at begge estimatorene $\hat{\mu}_1$ og $\hat{\mu}_2$ er forventningsrette.
Regn ut estimatorenes varians. Hvilken estimator er best? Begrunn svaret.

c) Bruk den estimatoren du mener er best og regn ut et estimat for μ. Finn også et tilhørende 95% konfidensintervall.

Oppgave 32

Vekten av ulike brød fra et bakeri vil variere litt (avhengig av porsjonering av deig, etc). For fire tilfeldig utvalgte brød fikk man følgende resultater: 782, 758, 775, 769 (gram).

a) Beregn gjennomsnitt, utvalgsstandardavvik og median for datasettet.

Anta at vekten av et tilfeldig valgt brød er normalfordelt med forventning μ og standardavvik σ. Anta videre at vektene av ulike brød er uavhengige av hverandre. Anta i punkt b) at $\mu = 780$ og $\sigma = 10$.

b) Hva er sannsynligheten for at den samlede vekten av to brød er mindre enn 1540?
Hva er sannsynligheten for at den gjennomsnittlige vekten av to slike brød er mindre enn 775?

Anta i resten av oppgaven at μ og σ er ukjente.

c) Beregn et 95% konfidensintervall for μ i denne situasjonen.

Det blir påstått at bakeriet lager brød som har lavere forventet vektet enn en oppgitte forventet vekt på $\mu = 780$ gram.

d) Formuler påstanden som en hypotesetest og utfør testen. Bruk 5% nivå.
Forklar hva resultatet av testen i praksis betyr.

Oppgave 33

I en undersøkelse av forurensing i en innsjø er blykonsentrasjonen i øvre sedimentlag på innsjøbunnen registrert i 25 prøver (hver på 1 liter).
Resultatene ble: gjennomsnitt 0.38 og utvalgsstandardavvik 0.06.

a) Hvilke antagelser må vi her gjøre for å kunne lage konfidensintervall eller gjøre hypotesetester for forventet blykonsetrasjon? Anta i punkt b) og c) under at disse antagelsene holder.

b) Finn et 98% konfidensintervall for forventet blykonsetrasjon.

c) Er det grunn til å konkludere at forventet blykonsentrasjon er høyere enn 0.35?
Formuler spørsmålet som et hypotesetestingsproblem og gjennomfør testen med nivå 1%.

Oppgave 34

Gå tilbake til oppgave 16. Vi fant der at antall biler solgt per dag hadde forventningsverdi 0.45 og varians 0.4475. Anta at bilforhandleren holder åpent 300 dager i løpet av ett år.

a) Hva er sannsynligheten for at bilforhandleren selger minst 125 biler i løpet av ett år?

Oppgave 35

En veiingeniør studerer størkningstiden til en ny sementblanding til bruk for å utbedre hull i veibanen. Registreringer fra 100 gangers bruk gav størkningstider med gjennomsnitt 32 minutt og utvalgsstandardavvik 4 minutt.
Anta at størkningstidene er uavhengige og identisk fordelt (dvs alle størkningstidene har samme fordeling).

a) Hvorfor trenger vi i denne situasjonen ikke å anta normalfordeling for målingene for å kunne lage konfidensintervall eller gjøre hypotesetester om forventningsverdien?

b) Finn et tilnærmet 95% konfidensintervall for forventet størkningstid.

c) Man regnet tidligere med en størkningstid på 30 minutter for en lignende sementblanding. Kan man si at den nye sementblandingen har en forventet størkningstid forskjellig fra 30 minutt?
Formuler spørsmålet som et hypotesetestingsproblem og gjennomfør testen med signifikansnivå 1%.

Oppgave 36

En bedrift produserer pumper til bruk på offshoreinstallasjoner. Over tid har det vist seg at 7% av alle produserte pumper har en feil. Det gjøres en del endringer i produksjonsprosessen for å prøve å redusere andelen produserte pumper med feil. Etter endringene i produksjonsprosessen er innarbeidet teller man opp hvor mange pumper med feil man finner i en produksjon på n uavhengige pumper. Man ønsker å bruke denne informasjonen til å avgjøre om andelen pumper med feil er redusert.

a) Formuler problemstillingen som en hypotesetest. Utfør testen når resultatet ble at man i $n = 100$ pumper fant 3 pumper med feil. Bruk 5% nivå. Regn også ut testens p-verdi.

Oppgave 37

Antall bakterier Y i t volumenheter av en bestemt bakteriekultur er Poissonfordelt med forventning λt, der λ er en parameter som angir gjennomsnittlig antall bakterier per volumenhet i kulturen.

a) Vi antar i dette punktet at $\lambda = 70$. Hva er forventet antall bakterier i en prøve på 3 volumenheter?
Finn også sannsynligheten for at det er mer enn 200 bakterier i en slik prøve.

Vi antar nå at λ er ukjent, og for å estimere λ skal det tas 5 prøver fra kulturen, hver på 3 volumenheter. Bakterieantallene i prøvene, Y_1, Y_2, Y_3, Y_4, Y_5 kan antas å være uavhengige variabler.

b) Forklar kort hvorfor $\hat{\lambda} = (Y_1 + Y_2 + Y_3 + Y_4 + Y_5)/15$ er en rimelig estimtor for λ.
Vis at estimatoren er forventningsrett og finn estimatorens varians.

I de 5 prøvene ble det totalt funnet 1084 bakterier.

c) Finn et (tilnærmet) 95% konfidensintervall for λ.

Oppgave 38

I oppgave 5 så vi på data for bensinprisen på to bensinstasjoner i Sandnes på ulike dager. La X være bensinpris på den ene stasjonen og Y være bensinpris på den andre stasjonen en tilfeldig dag. Registreringer av X og Y på 10 tilfeldig valgte dager gav dataene under.

dag i	1	2	3	4	5	6	7	8	9	10
x_i	14.89	14.39	13.20	15.35	15.10	14.39	14.96	15.15	14.69	13.57
y_i	14.99	14.39	13.65	15.25	14.99	14.09	14.66	15.25	14.36	13.57

Det oppgis at $\sum_{i=1}^{10} x_i = 145.69$, $\sum_{i=1}^{10} y_i = 145.20$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 4.436$,
$\sum_{i=1}^{10}(y_i - \bar{y})^2 = 3.414$ og $\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 3.670$.

a) Regn ut korrelasjonen mellom bensinprisene og kommenter resultatet.

b) Regn ut en estimert regresjonslinje for sammenhengen mellom bensinprisene ved de to bensinstatsjonene.
Lag et spredningsplott for dataene og tegn regresjonslinja inn i spredningsplottet.

Oppgave 39

Et bestemt stoff tilsatt bensinen antas å kunne redusere utslipp av nitrogenoksid (NOX) fra bilmotorer. For å undersøke dette nærmere, ble mengde tilsetningsstoff, x, og reduksjon i NOX-utslipp, Y, målt for $n = 12$ biler. Dataene er gitt i tabellen under.

bil i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	1	1	2	3	4	4	5	6	6	7	8	8
y_i	2.1	2.5	3.1	3	3.8	3.2	4.3	3.9	4.4	4.8	4.3	4.5

Det oppgis at $\sum_{i=1}^{12} x_i = 55$, $\sum_{i=1}^{12} y_i = 43.9$, $\sum_{i=1}^{12}(x_i - \bar{x})^2 = 68.9$,
$\sum_{i=1}^{12}(y_i - \bar{y})^2 = 8.19$ og $\sum_{i=1}^{12}(x_i - \bar{x})(y_i - \bar{y}) = 21.89$.

Vi vil gjøre en regresjonsanalyse av dataene for å studere forventet reduksjon i NOX-utslipp som funksjon av x.

a) Lag et spredningsplott av dataene. Regn ut korrelasjonen mellom mengden tilsetningsstoff og reduksjon i NOX-utslipp og kommenter.

b) Still opp modellen for enkel regresjonsanalyse. Hva er vanlige antakelser? Gi en praktisk fortolkning av stigningstallet i modellen.

c) Regn ut estimert regresjonslinje for sammenhengen mellom mengden tilsetningsstoff og reduksjon i NOX-utslipp. Tegn regresjonslinja inn i spredningsplottet i punkt a).

d) Det oppgis her at SS_E er 1.235. Hva blir estimatet av variansen rundt regresjonslinja?
Er det sammenheng mellom mengde tilsetningsstoff og reduksjon i NOX-utslipp? Besvar spørsmålet ved hjelp av en passende hypotesetest. Gjør klart hva H_0 og H_1 er. Bruk 5% signifikansnivå.

Oppgave 40

I denne oppgaven skal vi se litt mer på egenskapene til estimatoren for β i den enkle lineære regresjonsmodellen.

a) Vis at $\hat{\beta}$ er en forventningsrett estimator for β.
(Hint: Bruk resultatet på side 15 i notatene som sier at $\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.)

b) Vis at $\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

Oppgave 41

Tabellen under viser salgsprisen (i tusen kroner) og boligarealet (i kvadrameter) for leiligheter solgt i Stavanger.

areal, x_i	75	73	93	91	64	69	49	110	79	85	101	58
pris, y_i	2952	2723	3500	4005	2060	2560	1910	5500	3150	3535	3390	2120
areal, x_i	66	72	95	120	102	82	65	76	82	74	84	
pris, y_i	2455	2795	3360	4010	4625	3430	2300	3150	3450	2750	3690	

Det oppgis at $\sum_{i=1}^{23} x_i = 1865$ og $\sum_{i=1}^{23} y_i = 73420$. Datautskriften vi får når vi bruker R til å tilpasse en regresjonsmodell til dataene er vist under.

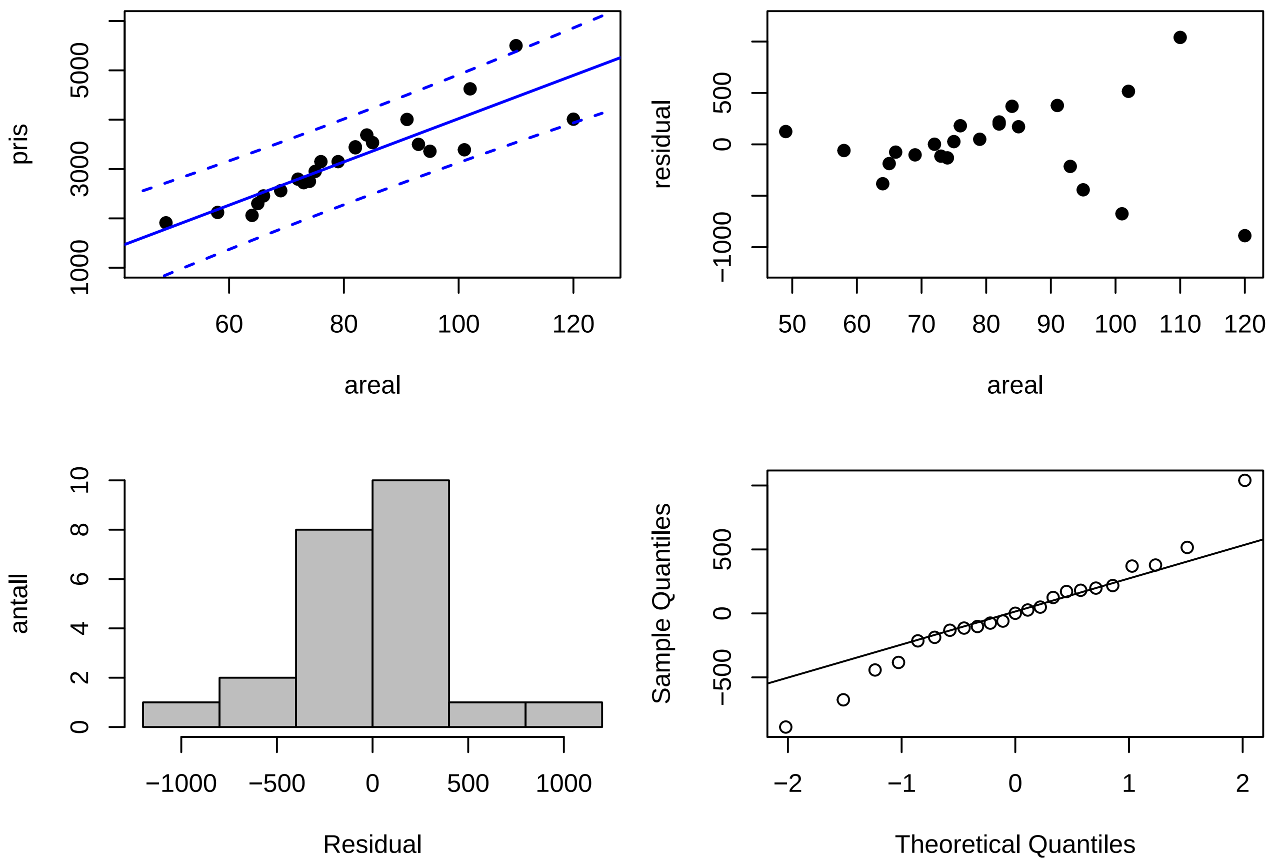
```
> regmod = lm(pris~areal)
> summary(regmod)

Call:
lm(formula = pris ~ areal)

Residuals:
    Min      1Q  Median      3Q     Max
-888.25 -159.16    1.23  189.33 1040.18

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -362.953    423.420  -0.857    0.401
areal         43.843      5.115   8.572 2.68e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 409 on 21 degrees of freedom
Multiple R-squared:  0.7777,Adjusted R-squared:  0.7671
F-statistic: 73.48 on 1 and 21 DF,  p-value: 2.675e-08

> confint(regmod)
                  2.5 %    97.5 %
(Intercept) -1243.50319 517.59805
areal          33.20665  54.48012
```

a) Skriv ned den enkle lineære regresjonsmodellen. Hvorfor er det rimelig å sette kvadrameter som x-variabelen og salgspris som Y-variabelen når vi tilpasser en regresjonsmodell til disse dataene?
Gi en praktisk tolkning av β-parameteren i regresjonsmodellen i denne situasjonen. Har α-parameteren en praktisk tolkning i denne situasjonen?
Hva er gjennomsnittlig kvadratmeterpris for de solgte leilighetene?

b) Skriv ned den estimerte regresjonslinja.
Hvor mye stiger estimert forventet pris når størrelsen øker 10 kvadrameter?
Hva er estimert forventet salgspris for en leilighet på 75 kvadratmeter?

c) Forklar kort hvordan vi raskt kan fastslå at det er reell sammenheng mellom areal og pris.
Lag både 95% konfidensintervall og 95% prediksjonsintervall for salgsprisen for en leilighet på 75 kvadratmeter. Forklar hva de to intervallene forteller oss.

I figuren under er det vist fire forskjellige plott knyttet til regresjonsmodellen.



d) Forklar kort hva hvert av plottene over viser.
Hvordan er residualene til en regresjonsmodell definert? Regn ut tallverdien til residualet for den første observasjonen.
Skriv ned de fire antagelsene en lineær regresjonsmodell bygger på, og for tre av antagelsene bruk plottene over til å avgjøre om antagelsene er oppfylte.

e) Hvor stor andel av den totale variasjonen i dataene er forklart av regresjonslinja? Forklar kort.
Nevn noen faktorer som i denne situasjonen bidrar til variasjonen rundt regresjonslinja.

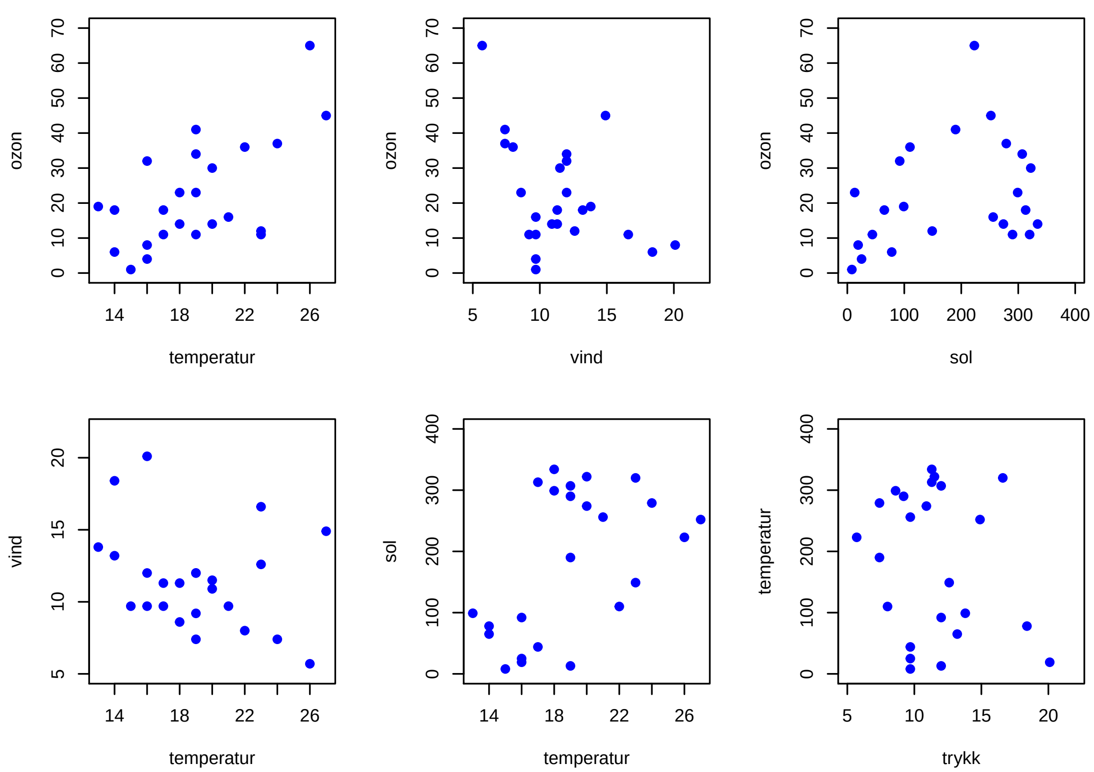
Oppgave 42

Ozon i lufta ved bakkenivå kan være et alvorlig forurensingsproblem dersom konsentrasjonen blir for stor. I denne oppgaven skal vi se på sammenhengen mellom ozonmengde og tre værvariable som antas å kunne ha innvirkning på ozonmengden i lufta, temperatur, vindhastighet og solstråling.

Data og plott av dataene fra 24 uavhengige målinger tatt på ulike tidspunkt ved en målestasjon like utenfor en amerikansk storby er gitt nedenfor.

Måling i	1	2	3	4	5	6	7	8	9	10	11	12
Ozon y_i	41	36	12	18	23	19	8	16	11	14	18	14
Temperatur x_{1i}	19	22	23	17	18	13	16	21	19	20	14	18
Vind x_{2i}	7.4	8.0	12.6	11.3	8.6	13.8	20.1	9.7	9.2	10.9	13.2	11.3
Sol x_{3i}	190	110	149	313	299	99	19	256	290	274	65	334

Måling i	13	14	15	16	17	18	19	20	21	22	23	24
Ozon y_i	34	6	30	11	1	11	4	32	23	45	37	65
Temperatur x_{1i}	19	14	20	17	15	23	16	16	19	27	24	26
Vind x_{2i}	12.0	18.4	11.5	9.7	9.7	16.6	9.7	12.0	12.0	14.9	7.4	5.7
Sol x_{3i}	307	78	322	44	8	320	25	92	13	252	279	223



For å undersøke hvordan ozonmengden avhenger av de tre værvariablene tilpasser vi først regresjonsmodellen

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

der $e \sim N(0, \sigma)$ og vi antar at for ulike målinger så er $e_1, \ldots, e_n$ uavhengige.

Når vi bruker et R til å estimere denne modellen fra dataene over får vi datautskriften gitt under.

```
> regmod = lm(ozon~temp+vind+sol, data=Ozondata)
> summary(regmod)

Call:
lm(formula = ozon ~ temp + vind + sol, data = Ozondata)

Residuals:
     Min       1Q   Median       3Q      Max
-17.6601 -11.1203   0.8171   8.9985  20.4124

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.373029  18.335857  -0.293   0.7725
temp         2.220528   0.800885   2.773   0.0117 *
vind        -1.233617   0.767297  -1.608   0.1236
sol         -0.003325   0.024774  -0.134   0.8946
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.13 on 20 degrees of freedom
Multiple R-squared:  0.4502,Adjusted R-squared:  0.3677
F-statistic: 5.458 on 3 and 20 DF,  p-value: 0.006603
```

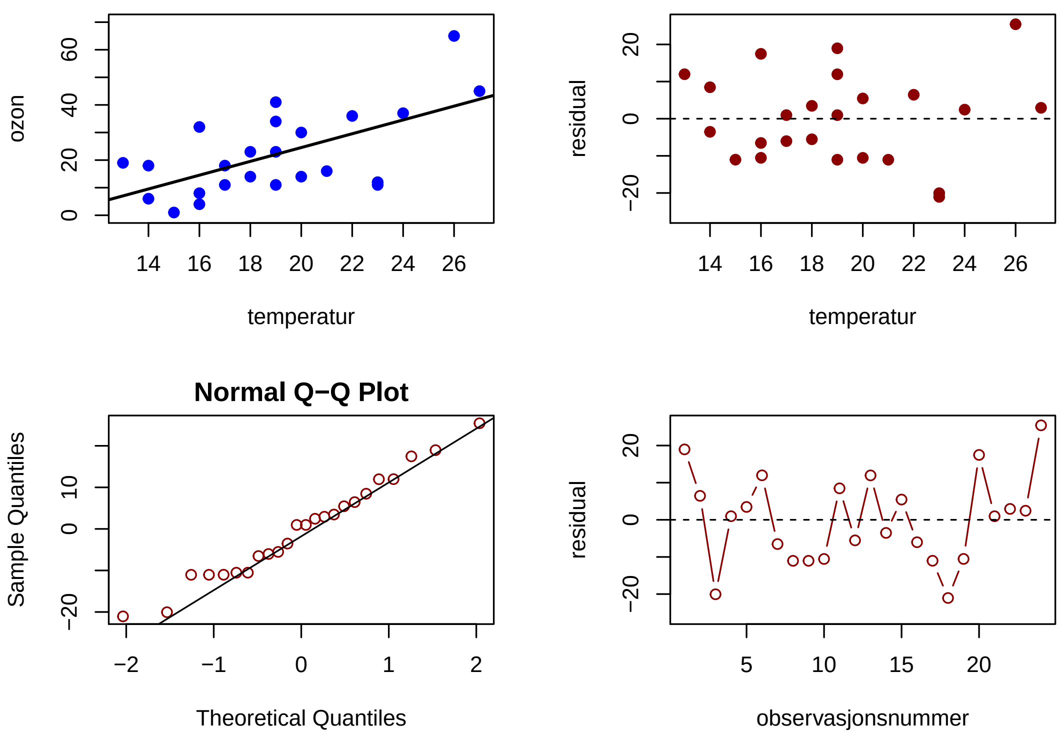
a) Skriv ned den estimerte regresjonslinja.
   Regn ut predikert mengde ozon en dag med temperatur 20, vind 10 og sol 100.
   Gi en praktisk tolkning av hver av de estimerte parameterverdiene $\hat{\beta}_1$, $\hat{\beta}_2$ og $\hat{\beta}_3$.

b) Har værvariablene i modellen samlet sett innflytelse på ozonmengden? Formuler dette som en hypotesetest og utfør testen på 5% signifikansnivå.
   Tyder informasjonen fra datautskriften på at hver enkelt av variablene i modellen har betydning eller kan noen av variablene utelates? Forklar kort

Ved å stegvis ta ut værvariabler som ikke har signifikant betydning blir man stående igjen med en modell med bare temperatur som x-variabel. I resten av oppgaven skal vi bruke denne modellen. Utskriften vi får når vi bruker R til å estimere denne modellen fra dataene over er gitt under.

```
> regmod = lm(ozon~temp, data=Ozondata)
> summary(regmod)

Call:
lm(formula = ozon ~ temp, data = Ozondata)

Residuals:
     Min       1Q   Median       3Q      Max
-21.0540 -10.5355   0.9614   6.9552  25.4367

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -25.517     13.217  -1.931  0.06652 .
temp           2.503      0.683   3.665  0.00136 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.29 on 22 degrees of freedom
Multiple R-squared:  0.3791,Adjusted R-squared:  0.3509
F-statistic: 13.43 on 1 and 22 DF,  p-value: 0.001361
```

Plott av data og residualer:



c) Hvordan er residualet til den $i$te observasjonen i en regresjonsmodell definert?
Regn ut residualet til observasjon nummer 7 (i=7).
Hvilke modellantagelser kan vi sjekke ved ulike plott av residualene? Hvordan bør disse plottene se ut dersom modellantagelsene holder?
Tyder residualplottene på forrige side på at den tilpassede modellen er en god modell?

d) Ved målestasjoner i andre områder har man erfart at en økning i temperaturen på 1 grad gir en økning i forventet ozonmengde på 2. Er det grunn til å påstå at dette forholdet er anderledes i området våre målinger er hentet fra? Formuler denne påstanden som en hypotesetest og utfør testen på 5% signifikansnivå.

e) Finn et 95% prediksjonsintervall for mengde ozon en dag med temperatur 25.
Det oppgis her at $\sum_{i=1}^{24} x_i = 456$, der x_i er temperatur dag i.

Oppgave 43

I en kjemisk prosess for fremstilling av bestemte gasser forbrukes koks. Koksforbruket avhenger av blandingsforholdet mellom luft og damp i en luft/damp-blanding som tilføres prosessen. Vi skal se på sammenhengen mellom luft/damp-forholdet og koksforbruk. Data for 27 målinger av samhørende verdier for luft/damp-forhold og koksforbruk er gitt under.

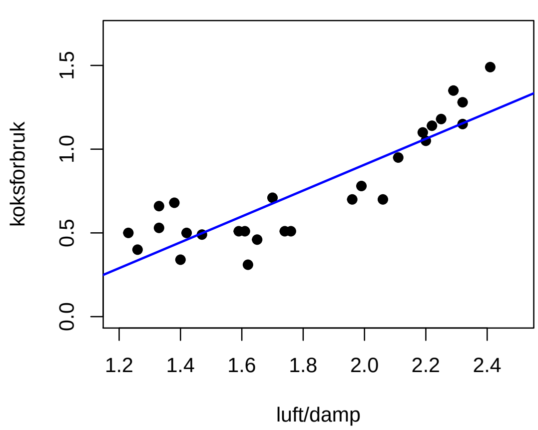
Måling i	1	2	3	4	5	6	7	8	9
Koksforbruk y_i	0.95	1.35	1.28	1.15	1.18	1.14	1.05	1.49	1.10
Luft/damp-forhold x_i	2.11	2.29	2.32	2.32	2.25	2.22	2.20	2.41	2.19

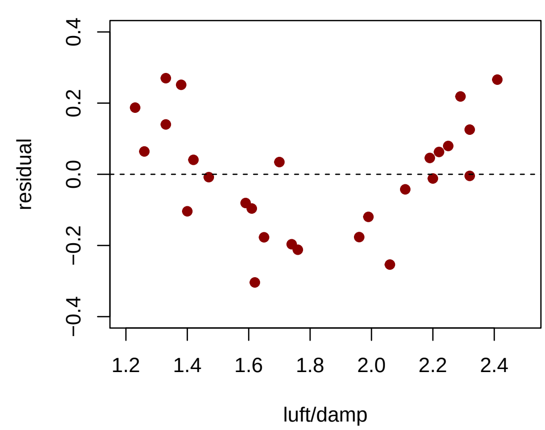
Måling i	10	11	12	13	14	15	16	17	18
Koksforbruk y_i	0.70	0.78	0.31	0.51	0.71	0.51	0.53	0.50	0.34
Luft/damp-forhold x_i	2.06	1.99	1.62	1.59	1.70	1.76	1.33	1.23	1.40

Måling i	19	20	21	22	23	24	25	26	27
Koksforbruk y_i	0.68	0.70	0.49	0.50	0.66	0.46	0.40	0.51	0.51
Luft/damp-forhold x_i	1.38	1.96	1.47	1.42	1.33	1.65	1.26	1.61	1.74

Datautskriften vi får når vi bruker R til å tilpasse en enkel lineær regresjonsmodell til dataene, et plott av dataene med regresjonslinjen tegnet inn og et plott av residualene er vist under.

```
> regmod = lm(koksforbruk~luft_damp, data=koksdata)
> summary(regmod)

Call:
lm(formula = koksforbruk ~ luft_damp, data = koksdata)

Residuals:
      Min        1Q    Median        3Q       Max
-0.303870 -0.111791 -0.004474  0.102556  0.270095

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.63724    0.15608  -4.083    4e-04 ***
luft_damp    0.77229    0.08451   9.139 1.91e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1662 on 25 degrees of freedom
Multiple R-squared:  0.7696,Adjusted R-squared:  0.7604
F-statistic: 83.52 on 1 and 25 DF,  p-value: 1.91e-09
```





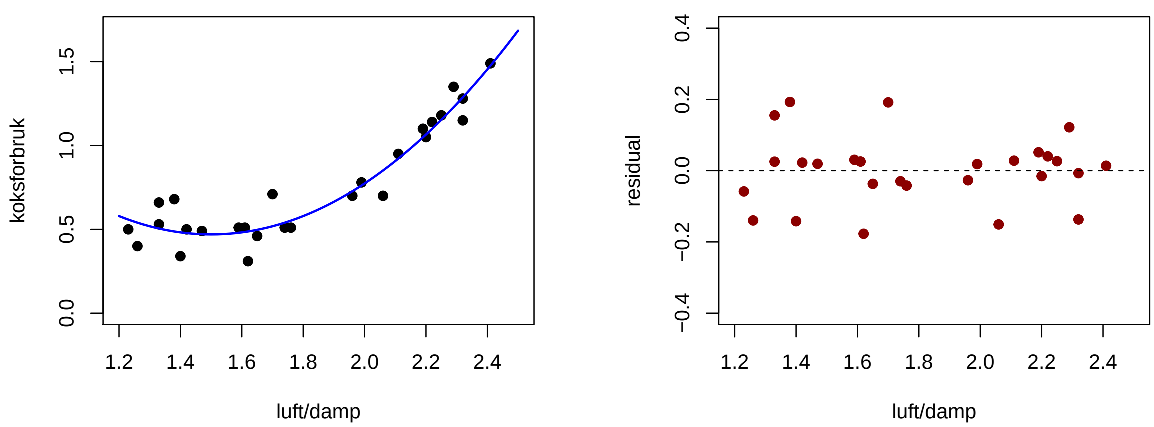
a) Skriv ned den enkle lineære regresjonsmodellen, og spesifiser alle antagelsene vi gjør når vi tilpasser en slik modell.
Hvordan er residualene i en enkel lineær regresjonsmodell definert? Regn ut residualet for den første og den siste målingen.
Hvilke modellantagelser kan vi sjekke ut fra residualplottet gitt på forrige side? Tyder plottet på at den tilpassede lineære modellen er god?

I stedet for den lineære modellen skal vi nå prøve den kvadratiske regresjonsmodellen

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$$

der $e \sim N(0, \sigma)$ og vi antar at $e_1, \ldots, e_n$ er uavhengige. Datautskrift og plott er vist under.

```
> regmod2 = lm(koksforbruk~luft_damp+I(luft_damp^2), data=koksdata)
> summary(regmod2)

Call:
lm(formula = koksforbruk ~ luft_damp + I(luft_damp^2), data = koksdata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.17722 -0.03948  0.01845  0.02922  0.19279 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      3.2043     0.5861   5.468 1.28e-05 ***
luft_damp       -3.6462     0.6672  -5.465 1.29e-05 ***
I(luft_damp^2)   1.2154     0.1830   6.642 7.19e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1007 on 24 degrees of freedom
Multiple R-squared:  0.9188,    Adjusted R-squared:  0.9121 
F-statistic: 135.8 on 2 and 24 DF,  p-value: 8.178e-14
```



b) Tyder residualplottet over på at den kvadratiske modellen er bedre enn den lineære modellen? Forklar. Hvilke andre plott av residualene ville du lage i tillegg til plottet over for å sjekke modellantagelsene?

c) Skriv ned den estimerte regresjonslinja. Regn ut estimert forventet koksforbruk når luft/damp-forholdet er 2.2.
Trenger vi å ha med kvadratleddet i modellen? Avgjør dette ved å utføre en hypotesetest. Skriv ned nullhypotese og alternativ hypotese og gi resultatet av testen. Bruk 5% nivå.

Til slutt prøver vi å tilpasse en modell med tredjegradspolynom:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + e$$

Datautskrift for denne modellen er vist under.

```
> regmod3 = lm(koksforbruk~luft_damp+I(luft_damp^2)+I(luft_damp^3), data=koksdata)
> summary(regmod3)

Call:
lm(formula = koksforbruk ~ luft_damp + I(luft_damp^2) + I(luft_damp^3),
    data = koksdata)

Residuals:
      Min        1Q    Median        3Q       Max
-0.187725 -0.039912  0.007701  0.037551  0.187705

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.3142     3.2749   0.401    0.692
luft_damp       -0.3472     5.6617  -0.061    0.952
I(luft_damp^2)  -0.6584     3.1982  -0.206    0.839
I(luft_damp^3)   0.3467     0.5907   0.587    0.563

Residual standard error: 0.1021 on 23 degrees of freedom
Multiple R-squared:   0.92,    Adjusted R-squared:  0.9096
F-statistic:  88.2 on 3 and 23 DF,  p-value: 9.198e-13
```

d) Er modellen med andregradspolynom eller modellen med tredjegradspolynom den beste modellen for dataene? Bruk informasjon i de to datautskriftene til å sammenligne de to modellen og til å avgjøre hvilken modell som er best.

Oppgave 44

I tabellen under er det vist data over prosentvis endring av Dow Jones indeksen de fem første dagene av året, x_i, og for hele året, y_i, for tretten ulike år.

x_i	1.5	0.2	-0.1	2.8	2.2	-1.6	-1.3	5.6	-1.4	1.4	1.5	-4.7	1.1
y_i	14.9	-9.2	19.6	20.3	-3.7	27.7	22.6	2.3	11.9	27	-4.3	20.3	4.2

Datautskriften vi får når vi bruker R til å tilpasse en regresjonsmodell til dataene er vist under.

```
> x = c(1.5, 0.2,-0.1,2.8 , 2.2,-1.6,-1.3,5.6,-1.4,1.4 ,1.5 ,-4.7,1.1)
> y = c(14.9,-9.2,19.6,20.3,-3.7,27.7,22.6,2.3 ,11.9, 27,-4.3,20.3, 4.2)
> rmod = lm(y~x)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-21.7352  -6.5044   0.7491   7.0137  16.9058

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.942      3.420   3.784  0.00302 **
x             -2.034      1.378  -1.476  0.16797
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.02 on 11 degrees of freedom
Multiple R-squared:  0.1653,Adjusted R-squared:  0.08944
F-statistic: 2.179 on 1 and 11 DF,  p-value: 0.168
```

a) Skriv ned den estimerte regresjonslinja. Hva betyr det i praksis at $\hat{\beta}$ er negativ? Lag et spredningsplott for dataene og tegn inn den estimerte regresjonslinja i plottet.

b) Skriv ned den observerte verdien på R^2 og kommenter kort.
Er det reell sammenheng mellom prosentvis endring av indeksen de fem første dagene og endringen for hele året? Formuler dette som en hypotesetest og gi konklusjonen av testen. Hva betyr dette i praksis? Kan vi bruke endringen av indeksen de første fem dagene til å predikere endringen for hele året?

Oppgave 45

I en undersøkelse av startlønn for nyutdannede ingeniører fant man for menn og kvinner i en bestemt bransje følgende tall (i 1000 kroner):

menn, x_i	434	399	450	439	440	459	444	468	488	455	490	425
kvinner, y_i	430	478	480	487	489	466	479	480	495			

Man har altså tall for $n_X = 12$ menn og $n_Y = 9$ kvinner. Det oppgis at $\sum_{i=1}^{12} x_i = 5391$, $\sum_{i=1}^{9} y_i = 4284$, $\sum_{i=1}^{12}(x_i - \bar{x})^2 = 7206.25$ og $\sum_{i=1}^{9}(y_i - \bar{y})^2 = 2912$

Anta at alle målinger er uavhengige og normalfordelte med samme ukjente varians σ^2 og med forventning μ_X for menn og μ_Y for kvinner.

a) Er det forskjell i forventet lønn mellom nyutdannede kvinner og menn i denne bransjen? Formuler problemstillingen som en hypotesetest og utfør testen. Bruk 5% signifikansnivå.

Oppgave 46

En bedrift tilbys et kurs som det påstås at vil gjøre arbeiderne mer effektive. Bedriften bestemmer seg for å først teste ut kurset på en liten gruppe ansatte. De velger ut 10 arbeidere som skal prøve kurset. Før kurset måler de hvor lang tid hver av disse arbeiderne bruker på å gjennomføre en arbeidsoppgave, og så måler de på nytt en stund etter kurset hvor lang tid de bruker på å gjøre den samme oppgaven. De målte tidene (i minutter) er gitt i tabellen under.

Person, i	1	2	3	4	5	6	7	8	9	10
Før kurs, x_i	48	53	52	57	43	53	59	71	40	61
Etter kurs, y_i	45	42	58	50	41	47	53	66	45	53
Differanse, d_i	3	11	-6	7	2	6	6	5	-5	8

Det oppgis at $\sum_{i=1}^{10} d_i = 37$ og $\sum_{i=1}^{10}(d_i - \bar{d})^2 = 268.1$

a) Blir arbeiderne mer effektive av kurset? Formuler dette som en hypotesetest og utfør testen.
Si kort hvilke antagelser testen bygger på.
Hva kalles den typen forsøk som er utført her? Nevn kort en annen fremgangsmåte bedriften kunne ha brukt for å teste effekten av kurset. Diskuter kort fordeler og ulemper med de to ulike fremgangsmåtene.

Oppgave 47

I en kartlegging av hvordan menn og kvinner stemte ved forrige valg ble et tilfeldig utvalg kvinner og menn spurt om hvordan de stemte ved forrige valg. Resultatet er summert opp i tabellen under.

	sosialistisk	borgerlig	stemte ikke
kvinner	203	211	96
menn	152	230	108

a) Bruk en kjikvadrattest til å avgjøre om undersøkelsen gir grunnlag for å konkludere at det er forskjell mellom menn og kvinner i hvordan de stemte. Bruk 5% nivå.
Er forutsetningene for testen oppfylte? Kommenter kort.

Oppgave 48

En bedrift som driver med matvareproduksjon er i sin produksjon bl.a. avhengige av at vannet som brukes holder god kvalitet. For å holde kontroll på vannkvaliteten tas det hver dag prøver der man bl.a. sjekker humusinnholdet i vannet. Basert på lang tids erfaring har man funnet ut at humusinnholdet i vannet en tilfeldig dag er normalfordelt med forventning $\mu = 10$ og varians $\sigma^2 = 2$ (på en skala for humusinnhold der 0 tilsvarer ingen humus). Vi skal også anta at humusinnholdet i vannet på ulike dager er uavhengig. (I praksis kan denne uavhengighetsantagelsen være tvilsom men det skal vi se bort fra i denne oppgaven.)

Dersom humusinnholdet en dag overskrider verdien 13 er vannkvaliteten så dårlig at produksjonen må stanses.

a) Vis at sannsynligheten for at produksjonen må stanses en tilfeldig dag på grunn av for høyt humusinnhold er 0.017.
Regn også ut sannsynligheten for at humusinnholdet en tilfeldig dag er mellom 8 og 12?

b) Finn den verdien a som er slik at det er 95% sannsynlighet for at humusinnholdet en tilfeldig dag er mellom $10 - a$ og $10 + a$.

c) Forklar hvorfor antall arbeidsdager der produksjonen må stanses i løpet av en måned med 20 arbeidsdager er binomisk fordelt med parametre $n = 20$ og $p = 0.017$.

d) Hva er sannsynligheten for at produksjonen må stanses på grunn av for høyt humusinnhold minst en arbeidsdag i løpet av en måned med 20 arbeidsdager?
Regn også ut forventa antall arbeidsdager i løpet av en måned med 20 arbeidsdager der produksjonen må stanses på grunn av for høyt humusinnhold.

Bedriften er ikke fornøyd med vannkvaliteten og bestemmer seg for å innstallere et vannrenseanlegg. Bedriften håper at dette medfører at forventa humusinnhold, μ, reduseres, mens man antar at variansen i humusinnhold, σ^2, fremdeles er 2. De 10 første dagene etter at renseanlegget er tatt i bruk registreres humusverdiene:

7.3	7.6	8.3	11.1	10.1	6.8	7.5	8.1	6.3	9.6

e) Still opp en forventningsrett estimator for forventa humusinnhold μ etter installasjon av renseanlegget, og vis at estimatoren er forventningsrett. Regn ut variansen til estimatoren.
Regn ut estimatet for μ fra dataene gitt over.

f) Utled et 95% konfidensintervall for μ, og regn ut et tallsvar fra informasjonen gitt over.

g) Tyder målingene fra de 10 første dagene på at forventa humusinnhold virkelig er redusert? Formuler dette som et hypotesetestingsproblem og utfør testen med 5% signifikansnivå.
Regn også ut p-verdien til testen.

h) Et ønske med renseanlegget er at forventet humusinnhold skulle reduserer til 9 eller mindre. Hvor stor styrke har testen i punkt g) dersom $\mu = 9$? Forklar kort hva dette betyr i praksis?
Hvor mange målinger må gjøres for å få en styrke på minst 90% dersom $\mu = 9$?

Til slutt skal vi anta at også variansen i humusinnhold etter innstallering av renseanlegget, σ^2, er ukjent.

i) Regn ut et estimat for σ^2 fra de oppgitte dataene.
Finn på nytt et 95% konfidensintervall for μ.

j) Finn et 95% konfidensintervall for σ.

k) Utfør testen i punkt g) på nytt.

Oppgave 49

Bremselengde for bil med to ulike dekktyper skal undersøkes. En bremseprøve utføres ved at man begynner å bremse når bilen kjører i $80 km/t$ og bremselengden måles. La $X_1, X_2, \ldots, X_{n_X}$ betegne bremselengdene i n_X prøver med dekktype 1. La tilsvarende $Y_1, Y_2, \ldots, Y_{n_Y}$ betegne bremselengdene i n_Y prøver med dekktype 2.

Anta at alle målinger er uavhengige og normalfordelte med samme ukjente varians σ^2. Resultatet av $n_X = n_Y = 10$ bremseprøver ble som gitt i følgende tabell:

i	1	2	3	4	5	6	7	8	9	10
x_i	33.0	30.8	28.0	28.7	28.9	26.6	27.9	28.9	27.8	27.4
y_i	23.4	25.3	25.0	28.9	26.7	25.9	24.4	26.8	28.8	25.5

Her er: $\sum_{i=1}^{10} x_i = 288.0$, $\sum_{i=1}^{10} y_i = 260.7$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 30.920$ og $\sum_{i=1}^{10}(y_i - \bar{y})^2 = 28.401$

a) Finn et 95% konfidensintervall for differansen i forventet bremselengde mellom de to dekktypene.

b) Avgjør med en hypotesetest om det er forskjell i forventet bremselengde mellom de to dekktypene. Bruk 5% nivå.
Forklar kort hvordan du alternativt kunne lese resultatet av testen ut av konfidensintervallet i a).

Oppgave 50

En landsdekkende butikk-kjede ønsker å test ut et nytt markedsføringstiltak. De prøver først tiltaket ut i en uke på åtte av butikkene sine, og registrerer overskudd (i hele tusen) uken før markedsføringstiltaket ble prøvd ut og den uken det ble prøvd. Resultatet er gitt i tabellen under.

Butikk, i	1	2	3	4	5	6	7	8
Med tiltak, x_i	39	42	22	57	63	51	62	70
Uten tiltak, y_i	35	44	28	50	68	45	53	63
Differanse, d_i	4	-2	-6	7	-5	6	9	7

Det oppgis at $\sum_{i=1}^{8} d_i = 20$ og $\sum_{i=1}^{8}(d_i - \bar{d})^2 = 246$

La $\mu_D = \mu_X - \mu_Y$ betegne forventet differanse i overskudd mellom uker med tiltak versus uker uten tiltak.

a) Har markedsføringstiltaket en positiv effekt? Formuler dette som en hypotesetest og utfør testen. Bruk 5% nivå.
Hvilke antagelser bygger testen på?

Ledelsen for butikk-kjeden konkluderer med at de må ha mer informasjon for å avgjøre om markedsføringstiltaket har effekt, og at de bare må se på den første uttestingen som en pilotstudie. Før de gjør mer testing ønsker de å gjøre et overslag over hvor mange butikker de bør teste tiltaket ut på for å kunne få dokumentert en eventuell positiv effekt.

Et forslag til grunnlag for denne beregningen er at de ønsker 90% sannsynlighet for å forkaste hypotesetesten satt opp i punkt a) dersom reell effekt av tiltaket er minst $\mu_D = 5$. Som en tilnærming vil de anta at variansen til differansen i overskudd, σ_D^2, er kjent og lik utvalgsvariansen beregnet i punkt a). Differansene i overskudd antas uavhengige og normalfordelte.

b) Regn ut hvor mange butikker de må teste ut markedsføringstiltaket på under betingelsene satt opp over.

Et annet forslag som kommer opp er å i stedet for å registrere størrelsene på overskuddene, bare registrere for hver butikk om overskuddet går opp eller ned fra uken uten tiltakt til uken med tiltak. Dersom man lar p betegne sannsynligheten for at overskuddet øker med markedsføringstiltaket kan man teste om tiltaket har effekt ved å teste om $p > 0.5$. Et forslag til dimensjonering av studien ut fra denne betrakningen er at de ønsker å ha 90% sannsynlighet for å med en hypotestest på 5% nivå kunne konkludere at $p > 0.5$ dersom virkelig verdi er $p = 0.7$ eller mer.

c) Hvor mange butikker må tas med i undersøkelsen dersom man bruker den siste fremgangsmåten foreslått over?
Diskuter kort fordeler/ulemper med den to fremgangsmåtene. Hvilken er best?

Oppgave 51

Ta utgangspunkt i dataene og problemstillingen gitt i oppgave 10.1 i læreboka. Dataene er også gjengitt under.

a) Lag et $\bar{x}$-diagram og et s-diagram for dataene. Kommenter resultatene.

b) Gjør oppgave 10.5 i læreboka. Dvs beregn kapabilitetsindeksen og avgjør om produksjonen holder god nok kvalitet når kravet som stilles er at volumet skal være mellom 0.48 og 0.52.

Tabell: Måling av volum i ølflasker

Stikkprøve nummer	Måling nummer 1	2	3	4	5	Gjennom-snitt	Standard-avvik	Varians
1	0,502	0,510	0,508	0,529	0,523	0,514	0,0112	0,00012
2	0,497	0,512	0,469	0,515	0,473	0,493	0,0216	0,00047
3	0,509	0,503	0,507	0,516	0,487	0,504	0,0107	0,00011
4	0,489	0,499	0,497	0,490	0,483	0,492	0,0064	0,00004
5	0,502	0,479	0,518	0,517	0,483	0,500	0,0185	0,00034
6	0,491	0,491	0,508	0,516	0,500	0,501	0,0107	0,00011
7	0,480	0,512	0,529	0,497	0,501	0,504	0,0184	0,00034
8	0,491	0,509	0,526	0,536	0,507	0,514	0,0173	0,00030
9	0,484	0,501	0,495	0,452	0,502	0,487	0,0210	0,00044
10	0,473	0,526	0,514	0,510	0,511	0,507	0,0199	0,00040
11	0,507	0,532	0,470	0,490	0,490	0,498	0,0232	0,00054
12	0,528	0,490	0,515	0,464	0,486	0,497	0,0254	0,00064

Oppgave 52

En bedrift som produserer bildeler er opptatt av kvalitet i alle ledd av bedriftens virksomhet. Bedriften mottar daglig et stort antall ordre på bildeler fra kunder over hele verden. For å sørge for at kundene er fornøyde fører bedriften blant annet kontroll med at ordrene blir korrekt ekspedert, dvs at kundene får tilsendt nøyaktig de typer og antall varer de spør etter. Denne kontrollen utføres ved at bedriften hver uke plukker ut $n = 90$ tilfeldige ordreforsendelser på vei ut fra bedriften og sjekker om forsendelser er i samsvar med ordren. Dersom forsendelsen ikke er i samsvar med ordren sier vi at ordren er feilaktig ekspedert. Andelen feilaktig ekspederte ordre i stikkprøver på 90 forsendelser tatt i 20 etterfølgende uker er gjengitt i tabellen under.

Uke	Andel feilaktig ekspederte	Uke	Andel feilaktig ekspederte
1	0.133	11	0.133
2	0.067	12	0.156
3	0.122	13	0.122
4	0.089	14	0.184
5	0.144	15	0.089
6	0.156	16	0.033
7	0.133	17	0.267
8	0.067	18	0.256
9	0.111	19	0.056
10	0.144	20	0.133

a) Bruk dataene fra de 10 første ukene som pilotdata og fra disse beregn senterlinje, øvre og nedre kontrollinje for et kontrolldiagram ($\hat{p}$-diagram) for andel feilaktig ekspederte ordre.
Tegn kontrolldiagramet for de 10 siste ukene. Kommenter resultatet.