Universitetet i Stavanger

EKSAMEN I: STA100 SANNSYNLIGHETSREGNING OG STATISTIKK

VARIGHET: 4 TIMER DATO: 14. MAI 2019

HJELPEMIDLER: Godkjent enkel kalkulator.

EKSAMEN BESTÅR AV 4 OPPGAVER PÅ 4 SIDER OG 4 SIDER VEDLEGG, TOTALT 8 SIDER.

EMNEANSVARLIG: Jan Terje Kvaløy TELEFON: 51 83 22 55

## Oppgave 1

Et overnattingssted leier ut 4 rom. La $X$ = antall rom som er belagt et tilfeldig døgn. Vi skal anta at $X$ har fordelingen:

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X = x)$ | 0.05 | 0.15 | 0.20 | 0.25 | 0.35 |

Overnattingsstedet har en inntekt på kr 750,- per utleid rom per døgn. For å drive stedet er det en fast utgift på kr 1150,- per døgn. La $Y$ være fortjenesten (inntekt minus utgift) et tilfeldig døgn.

a) Finn forventning og varians til $X$.

Finn forventning og varians til $Y$.

Anta at overnattingsstedet har åpent 300 dager i året, og la $W = Y_1 + Y_2 + \cdots + Y_{300}$, der $Y_i$ er fortjeneste døgn $i$. Anta at $Y_1, Y_2, \ldots, Y_{300}$ er uavhengige.

b) Finn $\mathrm{E}(W)$ og $\mathrm{SD}(W)$.

Gi en praktisk tolkning av hva $\mathrm{E}(W)$ forteller oss.

Finn sannsynligheten for at fortjenesten i løpet av ett år overstiger kr 250 000,-.

(Hint: Dersom du ikke fikk til punkt a) kan du bruke verdiene $\mathrm{E}(Y) = 900$ og $\mathrm{Var}(Y) = 900000$ når du løser b). Dette er ikke riktige svar for $\mathrm{E}(Y)$ og $\mathrm{Var}(Y)$, men verdier du kan bruke for å løse b) dersom du ikke fikk til punkt a).)

Oppgave 2

I denne oppgaven skal vi se på utslipp av fosfor fra et kommunalt kloakkrenseanlegg. Det hevdes fra de som driver renseanlegget at mengden fosfor i utslippsvannet varierer rundt et gjennomsnittsnivå på 0.18 mg/l. De hevder videre at variasjonen rundt dette nivået er slik at mengdene fosfor i like store prøver tatt på ulike dager vil være uavhengige og normalfordelte med forventning $\mu = 0.18$ og standardavvik $\sigma = 0.02$. La $X$ være mengden fosfor i en tilfeldig prøve av utslippsvannet.

a) Regn ut $P(X < 0.15)$ og $P(0.15 < X < 0.20)$.

Bestem verdien $k$ som er slik at $P(X < k) = 0.99$.

Regn ut sannsynligheten for at gjennomsnittet av 5 uavhengige målinger er mindre enn 0.15. Sammenlign svaret med svaret for $P(X < 0.15)$ og kommenter.

Det blir stilt nye og strengere miljøkrav til renseanlegget, og for å imøtekomme disse blir nytt renseutstyr installert. Man regner med at forventet mengde fosfor i utslippsvannet, $\mu$, vil endre seg når det nye renseutstyret tas i bruk, men man antar at standardavviket fremdeles vil være $\sigma = 0.02$.

b) Finn et 99% konfidensintervall for $\mu$ når man har observert at gjennomsnittet av 12 uavhengige prøver av utslippsvannet etter installering av nytt renseutstyr ble 0.142.

Hvor mange målinger måtte man i denne situasjonen minst gjøre dersom man ønsket et 99% konfidensintervall for $\mu$ med lengde på maksimalt 0.01?

Et av de nye kravene til renseanlegget er at gjennomsnittlig mengde fosfor i utslippsvannet i det lange løp skal være mindre enn 0.15 mg/l, dvs at $\mu < 0.15$. Vi skal bruke informasjonen fra målingene av fosforinnhold som er gjort så langt til å utføre en hypotesetest der vi tester om vi har grunnlag for å konkludere at $\mu < 0.15$.

c) Formuler nullhypotese og alternativ hypotese. Utfør testen og konkluder når observasjonene er som i punkt b). Bruk 5% signifikansnivå.

Regn også ut testens $p$-verdi.

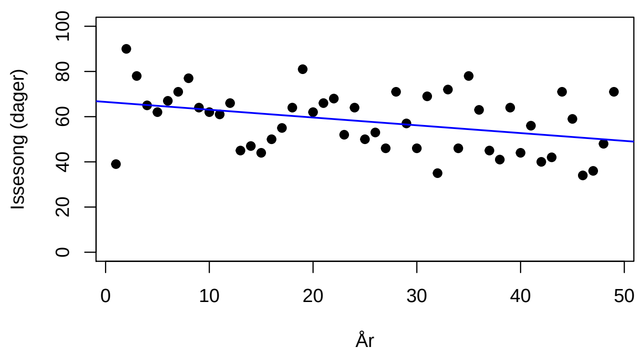
De som driver renseanlegget hevder at resultatet av testen i forrige punkt ble som det ble på grunn av at det var gjort for få målinger.

d) Hva forteller styrkefunksjonen oss? Regn ut styrken til testen for alternativene $\mu = 0.14$ og $\mu = 0.13$.

Hvor mange målinger må man gjøre for at styrken til testen for alternativet $\mu = 0.14$ skal bli minst 0.95?

Oppgave 3

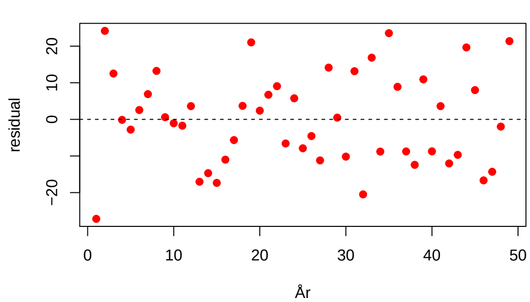
Rideu-kanalen i Ottawa i Canada blir om vinteren brukt som skøytebane så lenge isen er trygg. I denne oppgaven skal vi se på et datasett over lengde på issesongen (antall dager fra åpning til stenging) i perioden 1971 til 2019. Et plott av dataene med en lineær regresjonslinje tegnet inn er vist under. År langs $x$-aksen går fra 1 (=1971) til 49 (=2019). Regresjonslinjen tegnet inn i plottet er estimert fra regresjonsmodellen $Y = \alpha + \beta x + e$ der vi antar at $e \sim N(0, \sigma)$ og vi antar at feilleddene $e_1, \ldots, e_n$ for ulike målinger er uavhengige.

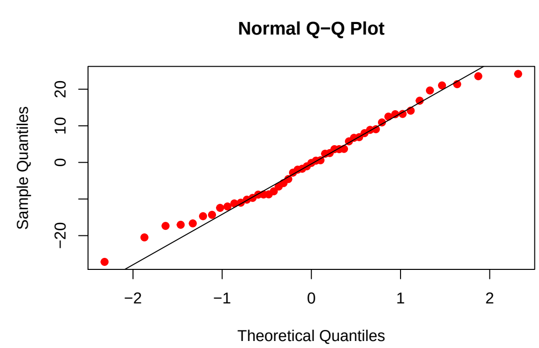



Deler av en datautskrift fra R og to plott av residualer for denne regresjonsmodellen er vist under.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  66.5153     3.6730  18.109  < 2e-16 ***
aar          -0.3447     0.1279  -2.695  0.00972 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.66 on 47 degrees of freedom
Multiple R-squared:  0.1339,Adjusted R-squared:  0.1155
F-statistic: 7.266 on 1 and 47 DF,  p-value: 0.009721
```



a) Skriv ned den estimerte regresjonslinja, og regn ut estimert forventet lengde på issesongen i 2020.

Forklar kort hvilke modellantagelser vi kan sjekke fra residualplottene på forrige side og kommenter om disse antagelsene ser ut til å holde.

Har forventet lengde på issesongen endret seg over tidsperioden vi har data fra? Formuler dette som en hypotesetest og gi resultatet av testen. Forklar hvilken retning en eventuell endring går i.

Regn ut estimert endring i forventet lengde på issesongen fra 1971 til 2019.

## Oppgave 4

Levetiden til en bestemt type lysrør er eksponentialfordelt med forventning 2.5 år. La $T$ være levetiden til et tilfeldig lysrør. Anta at levetidene til ulike lysrør er uavhengige.

a) Vis at $P(T < 1) = 0.33$.

Finn $P(1 < T < 2)$.

Finn median levetid til lysrørene, dvs finn tiden $t_{0.5}$ som er slik at $P(T < t_{0.5}) = 0.5$.

b) I et rom installeres 10 lysrør av den aktuelle typen. Finn sannsynligheten for at mer enn 3 lysrør feiler i løpet av ett år.

I en bygning installeres 100 lysrør av den samme typen. Finn sannsynligheten for at mer enn 30 lysrør feiler i løpet av ett år.

Alternativt til formuleringen brukt på formelarket kan sannsynlighetstettheten til eksponentialfordelingen skrives

$$f(t) = (1/\beta)e^{-t/\beta}, \text{ for } t \geq 0,$$

Sammenhengen med formuleringen på formelarket er at $\beta = 1/\lambda$ og vi får dermed at $\mathrm{E}(T) = \beta$ og $\mathrm{Var}(T) = \beta^2$. Dvs parameteren $\beta$ er lik forventningsverdien. For en ny variant av lysrørene er forventet levetid $\beta$ ukjent. For å estimere $\beta$ måles levetidene $T_1, \ldots, T_n$ til $n$ uavhengige lysrør.

c) Vis at $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} T_i = \overline{T}$ er en forventningsrett estimator for $\beta$. Finn også et uttrykk for variansen til estimatoren.

Bruk resultatene over som utgangspunkt, og finn et tilnærmet 95% konfidensintervall for $\beta$ under antagelsen at $n > 30$. (Hint: Du kan bruke tilnærmelsen $\beta^2/n \approx \hat{\beta}^2/n$.)

Regn ut et tallsvar for konfidensintervallet når måling av $n = 50$ lysrør gav gjennomsnittlig levetid $\bar{t} = 2.38$.