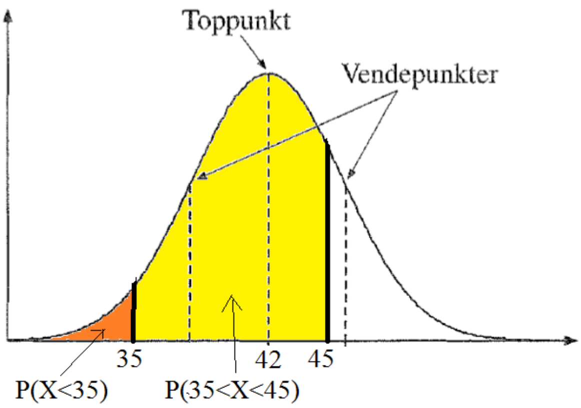
| | |
|---|---|
| Norges teknisk–naturvitenskapelige universitet<br>Institutt for matematiske fag | ST0103<br>Brukerkurs i statistikk<br>Løsningsskisse<br>eksamen 3.12.2016 |

$\boxed{1}$ **a)** $X$ er $N(42, 4)$.

$$P(X < 35) = \Phi\left(\frac{35-42}{4}\right) = \Phi(-1.75) = 0.0401$$

$$\begin{aligned} P(35 < X < 45) &= P(X < 45) - P(X < 35) \\ &= \Phi\left(\frac{45-42}{4}\right) - 0.0401 \\ &= \Phi(0.75) - 0.0401 = 0.7734 - 0.0401 \\ &= 0.7333 \end{aligned}$$




**b)**

$$\bar{X} \sim N(42, \frac{4}{\sqrt{4}}) = N(42, 2)$$

$$P(\bar{X} < 35) = \Phi\left(\frac{35-42}{2}\right) = \Phi(-3.5) = 0.0002$$

La antallet prøver som er mellom 35 og 45 være $Y$. Da er $Y \sim$ binomisk(4,0.7333).

$$\begin{aligned} P(Y \geq 3) &= P(Y = 3) + P(Y = 4) \\ &= \binom{4}{3} 0.7333^3(1 - 0.7333) + \binom{4}{4} 0.7333^4(1 - 0.7333)^0 \\ &= 4 \cdot 0.7333^3 \cdot 0.2667 + 0.7333^4 = 0.710 \end{aligned}$$

$\boxed{2}$ **a)**

$$\hat{\mu} = \bar{X} = \frac{879.41}{20} = 43.97$$

$$SE(\hat{\mu}) = \frac{\sigma}{\sqrt{n}} = \frac{4.0}{\sqrt{20}} = 0.8944$$

95% konfidensintervall for $\mu$:

$$\begin{array}{rcl} \bar{X} & \pm & 1.96 \cdot \frac{\sigma}{\sqrt{n}} \\ 43.97 & \pm & 1.96 \cdot \frac{4.0}{\sqrt{20}} \\ 43.97 & \pm & 1.75 \\ (42.22 & , & 45.72) \end{array}$$

**b)** Botanikeren ønsker å finne ut om $\mu \neq 42$, så dette velges som $H_1$. Dermed testes

$$H_0 : \mu = 42.0 \text{ mot } H_1 : \mu \neq 42.0.$$

Siden $\sigma$ er kjent, er testobservatoren

$$Z = \frac{\bar{X} - 42}{4/\sqrt{20}} \tag{1}$$

som er $N(0,1)$ hvis $H_0$ gjelder.
Utregning gir $Z = \frac{43.97-42.0}{4/\sqrt{20}} = 2.20$.
Med signifikansnivå 0.05 skal vi forkaste $H_0$ hvis

$$Z \geq z_{0.025} = 1.96 \text{ eller } Z \leq -z_{0.025} = -1.96$$

Vi forkaster altså $H_0$ med nivå 0.05 siden $Z = 2.20$ med våre data.
$p$-verdien blir $2 \cdot P(Z > 2.20) = 2 \cdot (1 - P(Z \leq 2.20)) = 2 \cdot (1 - 0.9861) = 0.0278$.
$p$-verdien er minste signifikansnivå som ville gi forkastning av $H_0$. Den kan også fortolkes som sannsynligheten for, hvis $H_0$ gjelder, å få et resultat likt det vi har fått eller mer ekstremt i forhold til nullhypotesen.

**c)** Type-I feil: Forkaste $H_0$ hvis $H_0$ er sann.
Type II-feil: Ikke forkaste $H_0$ hvis $H_1$ er sann.
Sannsynligheten for type I-feil er lik 0.05 (lik det valgte sannsynlighetsnivået).
Sannsynligheten for type II-feil dersom $\mu = 45.0$ er lik $P(-1.96 < Z < 1.96)$ der $Z$ er som definert i (1). Men vi trenger nå fordelingen til denne $Z$ når $\mu = 45.0$. Vi har da

$$\begin{array}{rcl} E(Z) & = & E\left(\frac{\bar{X} - 42.0}{4/\sqrt{20}}\right) = \frac{45 - 42}{4/\sqrt{20}} = 3.354 \\ Var(Z) & = & 1 \end{array}$$

dvs. $Z \sim N(3.354, 1)$.
Fra dette får vi, når $\mu = 45$ og med den aktuelle fordelingen for $Z$:

$$\begin{array}{rcl} P(\text{ikke forkaste}) & = & P(-1.96 < Z < 1.96) \\ & = & P(Z < 1.96) - P(Z < -1.96) \\ & = & \Phi(1.96 - 3.354) - \Phi(-1.96 - 3.354) \\ & = & \Phi(-1.39) - \Phi(-5.314) = 0.0823 - 0.000 \\ & = & 0.0823 \end{array}$$

$\boxed{3}$ **a)** Ved å ta logaritmen i den første ligningen nedenfor får vi

$$\begin{aligned} d &= \theta(1-c)^\beta \\ \ln d &= \ln\theta + \beta\ln(1-c) \\ y &= \alpha + \beta x \end{aligned}$$

Antagelser for modellen (1) er at $x_i$-ene er konstanter, mens $e_1, e_2, \ldots, e_{10}$ er uavhengige og $N(0,\sigma)$. $\alpha, \beta, \sigma$ er de ukjente parametrene i modellen.

Diskusjon: Observasjonene er basert på den samme kjerneprøven, så uavhengighet kan være diskutabelt. Variansene til $e_i$-ene skal være de samme, men disse variansene kan ofte avhenge av $x_i$-ene. Den lineære sammenhengen mellom $x_i$ og $y_i$ baserer seg på en antatt teoretisk sammenheng mellom $d$ og $c$.

**b)** Bruk formlene i vedlegget:

$$\begin{aligned} \hat\beta &= \frac{10\cdot(-10.108)-(-4.747)\cdot 32.924}{10\cdot 3.440-(-4.747)^2} = \frac{-101.08+156.2902}{34.40-22.53401} = 4.65 \\ \hat\alpha &= \frac{32.924-4.65\cdot(-4.747)}{10} = 5.50 \\ \hat\theta &= e^{\hat\alpha} = e^{5.50} = 244.58 \\ S &= \sqrt{\frac{SS_E}{n-2}} = \sqrt{\frac{0.1898}{8}} = 0.154 \\ SE(\hat\beta) &= \sqrt{\frac{S^2}{\sum(x_i-\bar x)^2}} = \frac{0.154}{\sqrt{1.187}} = 0.141 \end{aligned}$$

95% konfidensintervall for $\beta$:

$$\begin{aligned} \hat\beta &\pm t_{0.025,8}\cdot SE(\hat\beta) \\ 4.65 &\pm 2.306\cdot 0.141 \\ 4.65 &\pm 0.33 \\ (4.32 &\ ,\ \ 4.98) \end{aligned}$$

Andel forklart variasjon:

$$R^2 = \frac{SS_T - SS_E}{SS_T}$$

Vi trenger da

$$\begin{aligned} SS_T &= \sum y_i^2 - (1/10)(\sum y_i)^2 \\ &= 134.272 - 0.1\cdot 32.924^2 = 25.8730 \end{aligned}$$

Dermed er

$$R^2 = \frac{25.8730-0.1898}{25.8730} = 0.993$$

Dette er en høy $R^2$, noe som indikeres av spredningsplottet ved at punktene ligger meget nær den rette linja.

$\boxed{4}$ **a)** $X$ er Poisson-fordelt med forventning $\mu$ der

$$\mu = \lambda\cdot\ \text{tid} = 0.01\cdot 168 = 1.68.$$

$$P(X=0) = e^{-1.68} = 0.186$$

$$\begin{aligned}
P(X \geq 3) &= 1 - P(X=0) - P(X=1) - P(X=2) \\
&= 1 - e^{-\mu} - \frac{\mu^1}{1!}e^{-\mu} - \frac{\mu^2}{2!}e^{-\mu} \\
&= 1 - e^{-\mu}(1 + \mu + \mu^2/2) \\
&= 1 - e^{-1.68}(1 + 1.68 + 1.68^2/2) = 0.2375
\end{aligned}$$

La $Y$ være antall insekter fanget i uke 2. Vi skal finne $P(X = 0|X + Y = 3)$ (sjekk at det er dette som menes i teksten). Vi har da at $X+Y$ er Poisson-fordelt med forventning $2 \cdot 1.68 = 3.36$. Vi får:

$$\begin{aligned}
P(X = 0|X + Y = 3) &= \frac{P(X = 0 \cap X + Y \geq 3)}{P(X + Y \geq 3)} \\
&= \frac{P(X = 0 \cap Y \geq 3)}{P(X + Y \geq 3)} \\
&= \frac{P(X = 0) \cdot P(Y \geq 3)}{P(X + Y \geq 3)} \\
&= \frac{0.186 \cdot 0.2375}{1 - e^{-3.36}(1 + 3.36 + 3.36^2/2)} = 0.0677
\end{aligned}$$

**b)** $T$ er eksponentialfordelt med $\lambda = 0.01$.

$$P(T > 48) = e^{-0.01 \cdot 48} = e^{-0.48} = 0.619$$

$$P(T > 96|T > 48) = \frac{P(T > 96)}{P(T > 48)} = \frac{e^{-0.01 \cdot 96}}{e^{-0.01 \cdot 48}} = e^{-0.01(96-48)} = e^{-0.48} = 0.619$$

Gitt at ingen fanges i de 48 første timene, er dette sannsynligheten for at ingen fanges i de neste 48 timene. Svaret er det samme som sannsynligheten for at ingen insekter fanges i de første 48 timene. Dette er en egenskap ved eksponetialfordelingen: Hvis tiden til en hendelse er eksponentialfordelt, og det ikke har skjedd noen hendelse i løpet av de første $s$ tidsenheter, har tiden til neste hendelse samme fordeling som tiden til første hendelse beregnet fra start. Mer presist, $P(T > s+t|T > s) = P(T > t)$. Dette fortolkes som at eksponentialfordelingen er uten hukommelse.

**c)** $Y$ er Poisson-fordelt med forventning $\mu = \lambda \cdot 168 \cdot 5 = 840\lambda$.
Dermed er $E(Y) = 840\lambda$, så $\lambda = E(Y)/840$.
Momentmetoden gir da estimatoren

$$\hat{\lambda} = \frac{Y}{840}$$

Denne er forventningsrett da $E(\hat{\lambda}) = \frac{E(Y)}{840} = \lambda$.
Estimatet blir $\hat{\lambda} = 15/840 = 0.018$.

$$Var(\hat{\lambda}) = Var\left(\frac{Y}{840}\right) = \frac{1}{840^2} \cdot Var(Y) = \frac{1}{840^2} \cdot 840\lambda = \frac{\lambda}{840}$$

Standardfeilen blir da:

$$SE(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{840}} = \sqrt{\frac{15}{840^2}} = 0.0046$$