(Det tas forbehold om feil i løsningsforslaget.)

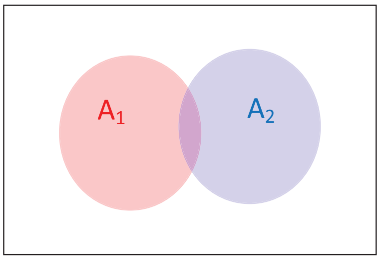
Oppgave 1

a) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 5/30 + 5/30 - (5/30)(5/30) = 275/900 = 0.306$
$P(A_1 \cap A_2) = P(A_1)P(A_2) = (5/30)(5/30) = 25/900 = 1/36$, siden begivenhetene A_1 og A_2 er antatt uavhengige.
$P(A_1 \cap \overline{A_3}) = P(A_1)(1 - P(A_3)) = (5/30)(1 - 5/30) = 5/36$

b) To begivenheter A og B er statistisk uavhengige dersom $P(A|B) = P(A)$.

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{2/30}{5/30} = 2/5 = 0.4.$$

$$P(A_2|\overline{A_1}) = \frac{P(A_2 \cap \overline{A_1})}{P(\overline{A_1})} = \frac{P(A_2 \cap \overline{A_1})}{1 - P(A_1)}.$$

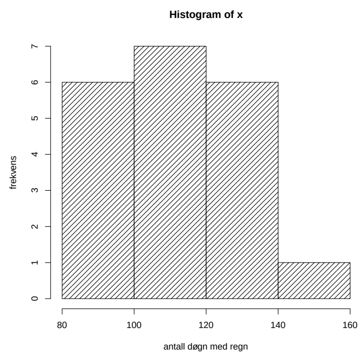


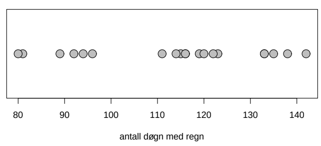
Vi trenger å finne $P(A_2 \cap \overline{A_1})$. Men vi har at $P(A_2) = P(A_2 \cap \overline{A_1}) + P(A_2 \cap A_1)$ (jf. figur tilhøyre), som gir at $P(A_2 \cap \overline{A_1}) = P(A_2) - P(A_2 \cap A_1) = 5/30 - 2/30 = 3/30 = 1/10$.

Derfor: $P(A_2|\overline{A_1}) = \dfrac{3/30}{25/30} = \dfrac{3}{25} = 0.12.$

Oppgave 2

a) Til høyre under vises histogram og prikkdiagram over dataene.





Første kvartil: $\frac{1}{2}(95 + 96) = 95$
Tredje kvartil: $\frac{1}{2}(123 + 133) = 128$
Kvartilavvik: $128 - 95 = 33$
Variasjonsbredde: $142 - 80 = 62$
Median: $\frac{1}{2}(116 + 116) = 116$

Det kan være stor forskjell mellom median og gjennomsnitt dersom dataene er usymmetrisk fordelt.

b) $X \sim N(100, 20)$

$$\begin{aligned} P(X < 80) &= P\left(\frac{X - 100}{20} < \frac{80 - 100}{20}\right) \\ &= P(Z < -1) = 0.1587 \end{aligned}$$

(Her er $Z \sim N(0,1)$.)

$$\begin{aligned} P(X > 140) &= P\Big(\frac{X-100}{20} > \frac{140-100}{20}\Big) \\ &= P(Z > 2.0) = 1 - P(Z < 2.0) = 1 - 0.9772 = 0.0228 \end{aligned}$$

$$\begin{aligned} P(90 < X < 120) &= P(X < 120) - P(X < 90) \\ &= P(Z < 1.0) - P(Z < -0.5) = 0.8413 - 0.3085 = 0.5328 \end{aligned}$$

$$\mathrm{Var}\Big(\frac{X_1+X_2}{2}\Big) = \cdots = \frac{20^2}{2} = 200$$

$$\begin{aligned} P\Big(\frac{X_1+X_2}{2} > 140\Big) &= P\Big(\frac{\frac{X_1+X_2}{2} - 100}{\sqrt{200}} > \frac{140-100}{\sqrt{200}}\Big) \\ &= P(Z > 2.83) = 1 - P(Z < 2.83) = 1 - 0.9977 = 0.0023 \end{aligned}$$

c*) Vi antar at dataene $x_1, \ldots, x_{20}$ er utfall av $n = 20$ u.i.f. tilfeldige variable $X_1, \ldots, X_{20}$, normalfordelte med $\mathrm{E}(X_i) = \mu_X$ og $\mathrm{SD}(X_i) = \sigma_X$.
Da er et 95% konfidensintervall for μ_X gitt ved:

$$\Big(\overline{X} - t_{0.025,19}\sqrt{\frac{S^2}{20}}, \quad \overline{X} + t_{0.025,19}\sqrt{\frac{S^2}{20}}\Big)$$

Vi har at $t_{0.025,19} = 2.093$. Innsatt data blir utregnet intervall:

$$\Big(113.45 - 2.093\sqrt{\frac{18.94^2}{20}}, \quad 113.45 + 2.093\sqrt{\frac{18.94^2}{20}}\Big) = (104.59, \quad 122.31)$$

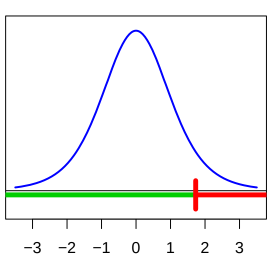
d) Vi vil teste: $H_0: \quad \mu_X = 100 \qquad \text{mot} \qquad H_1: \quad \mu_X > 100$

t-test:

Dersom H_0 er korrekt er

$$\underbrace{\frac{\overline{X} - 100}{\sqrt{S^2/20}}}_{\textit{teststørrelse}} \sim \underbrace{t(19)}_{\textit{nullfordeling}}$$

Nullfordeling: t(19)−tetthet.



Med signifikansnivå 5%, dvs $\alpha = 0.05$, er *forkastningsområdet* $[t_{0.05,19}, \infty)$. Fra tabell: $t_{0.05,19} = 1.729$

Utfall av teststørrelsen: $\dfrac{113.45 - 100}{\sqrt{18.94^2/20}} = 3.18$

Siden utfallet 3.18 er i forkastnignsområdet, blir konklusjonen forkast H_0. Dataene gir grunnlag for å påstå at forventet antall døgn med regn er høyere enn 100.

e) Test: forkast H_0 dersom $\dfrac{\overline{X} - 100}{\sqrt{20^2/20}} > z_{0.05} = 1.645$

Styrke: $\gamma(\mu_X) = P\Big(\dfrac{\bar{X} - 100}{\sqrt{20^2/20}} > 1.645 \Big| \mu_X\Big) = P\Big(\dfrac{\bar{X} - \mu_X}{\sqrt{20^2/20}} > 1.645 + \dfrac{100 - \mu_X}{\sqrt{20^2/20}} \Big| \mu_X\Big)$

$$= P\Big(Z > 1.645 + \frac{100 - \mu_X}{\sqrt{20^2/20}}\Big)$$

Styrke i alternativet $\mu_X = 110$:

$$\gamma(110) = P\Big(Z > 1.645 + \frac{100 - 110}{\sqrt{20^2/20}}\Big) = P(Z > -0.59) = 1 - 0.2776 = 0.7224$$

Oppgave 3

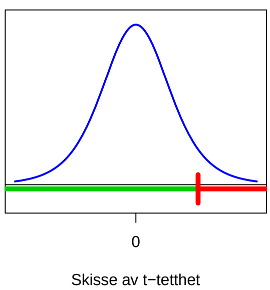
a*) Toutvalgs t-test:

Anta like varianser; estimator: $S^2_{\text{pooled}} = \dfrac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$ der $S_X^2 = \frac{1}{n_X - 1}\sum_{i=1}^{n_X}(X_i - \overline{X})^2$, og $S_Y^2 = \frac{1}{n_Y - 1}\sum_{i=1}^{n_Y}(Y_i - \overline{Y})^2$.

Test (sign.nivå α) for: $H_0 : \mu_X - \mu_Y = 0$ mot $H_1 : \mu_X - \mu_Y > 0$

Forkast H_0 dersom

$$\underbrace{\frac{\overline{X} - \overline{Y}}{S_{\text{pooled}}\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}}_{\textit{teststørrelse}} \geq t_{\alpha, n_X + n_Y - 2}$$



Skisse av t-fordeling og forkastningsområde.

Nullfordeling: $t(n_X + n_Y - 2) = t(48)$

Med signifikansnivå 5%, dvs $\alpha = 0.05$, er *forkastningsområdet* $[t_{0.05,48}, \infty)$. Fra tabell: $t_{0.05,48} = 1.677$

Utfall av S^2_{pooled}: $\dfrac{(20-1)18.94^2 + (30-1)17.07^2}{20+30-2} = 318.04$

Utfall av teststørrelse: $\dfrac{113.45 - 102.27}{\sqrt{318.04}\sqrt{\frac{1}{20}+\frac{1}{30}}} = 2.17$

Siden utfallet av teststørrelsen er i forkastningsområdet, kan vi forkaste H_0. Dvs.: det er grunnlag i dataene for å hevde at forventet antall regndager er høyere i den siste perioden.

b) Et 95% konfidensintervall for $\mu_X - \mu_Y$ er i denne situasjonen gitt ved:

$$\left(\overline{X} - \overline{Y} - t_{0.025,48}\sqrt{S^2_{\text{pooled}}(\frac{1}{20}+\frac{1}{30})},\quad \overline{X} - \overline{Y} + t_{0.025,48}\sqrt{S^2_{\text{pooled}}(\frac{1}{20}+\frac{1}{30})}\right)$$

Vi har at $t_{0.025,48} = 2.011$. Innsatt data blir utregnet intervall:

$$\left(113.45 - 102.27 - 2.011\sqrt{318.04(\frac{1}{20}+\frac{1}{30})},\quad 113.45 - 102.27 + 2.011\sqrt{318.04(\frac{1}{20}+\frac{1}{30})}\right) = (0.83, \quad 21.53)$$

c*) $Y_i = \alpha + \beta x_i + e_i, i = 1, \ldots, 50$. Her er $e_1, \ldots, e_{50}$ u.i.f. tilfeldige variable der $e_i \sim N(0, \sigma)$.

Estimat av β: $b = \dfrac{s_{xy}}{s_{xx}} = \dfrac{4774.5}{10412.5} = 0.459$

Estimat av α: $a = \overline{y} - b\overline{x} = 106.74 - 0.459 \cdot 1984.5 = -804.1$

Fortolkning av estimert stignigstall (β): forventet antall regndager øker med 0.459 for hvert år.

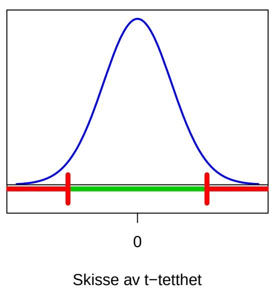
<u>Prediksjon av forventet antall</u>:
1970: $-804.1 + 0.459 \cdot 1970 = 100.13$
2005: $-804.1 + 0.459 \cdot 2005 = 116.2$

d)

Vi vil teste: $H_0 : \beta = 0$ mot $H_1 : \beta \neq 0$.

t-test:
Dersom H_0 er korrekt er

$$\underbrace{\frac{\widehat{\beta}}{\sqrt{S^2/s_{xx}}}}_{\textit{teststørrelse}} \sim \underbrace{t(48)}_{\textit{nullfordeling}}$$



Skisse av t-fordeling og forkastnings-område.

Med signifikansnivå 5%, dvs $\alpha = 0.05$, er *forkastningsområdet* $(\infty, -t_{0.025,48}] \cup [t_{0.025,48}, \infty)$ (tosidig test). Fra tabell: $t_{0.025,48} = 2.011$.

Utfall av teststørrelsen: $\dfrac{0.459}{\sqrt{303.7/10412.5}} = 2.69$.
(Fra oppgaveteksten: Estimat av $\mathrm{Var}(e_i) = \sigma^2$: $\frac{SSE}{50-2} = \frac{14578.3}{48} = 303.7$).

Siden utfallet 2.69 er i forkastnignsområdet, blir konklusjonen forkast H_0. Dataene gir grunnlag for å påstå at det er en trend over tid.

e) $R^2 = \dfrac{SSR}{SST}$, $SST = S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^{50}$ og $SSR = SST - SSE$.

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} : \frac{16767.62 - 14578.3}{16767.62} = 0.1306$$

Dvs.: regresjonen forklarer (kun) ca. 13% av variasjonen.