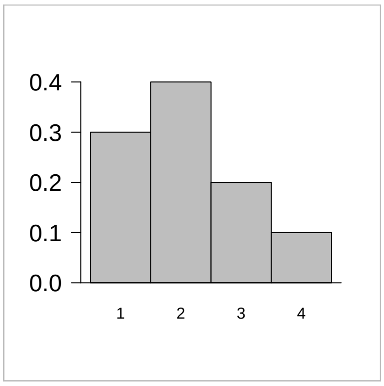
(Det tas forbehold om feil i løsningsforslaget.)

Oppgave 1

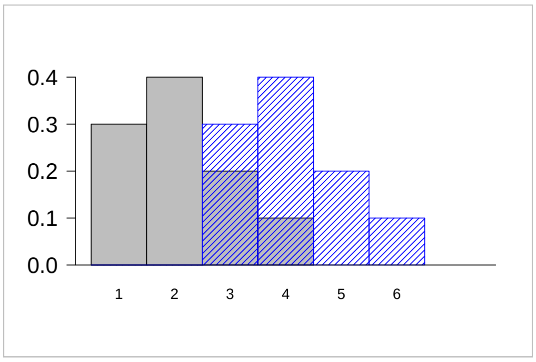
a)

$$P(Y \leq 1) = P(Y = 1) = 0.3,$$
$$P(Y > 2) = P(Y = 3) + P(Y = 4) = 0.2 + 0.1 = 0.3$$
$$\mathrm{E}(Y) = \sum_{y=1}^{4} yP(Y = y) = 1 \cdot 0.3 + \cdots + 4 \cdot 0.1 = 2.1$$
$$\mathrm{E}(Y^2) = \sum_{y=1}^{4} y^2 P(Y = y) = 1^2 \cdot 0.3 + \cdots + 4^2 \cdot 0.1 = 5.3$$
$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \{\mathrm{E}(Y)\}^2 = 5.3 - 2.1^2 = 0.89.$$



b)

$$\mathrm{E}(X) = \mathrm{E}(Y + 2) = \mathrm{E}(Y) + 2 = 4.1;$$
$$\mathrm{Var}(X) = \mathrm{Var}(Y + 2) = \mathrm{Var}(Y) = 0.89$$



Histogrammet til X (blåskravert i figuren til høyre, tegnet sammen med histogrammet til Y) er av <u>samme form</u> som histogrammet til Y; det er bare <u>forkjøvet to enheter</u> til høyre.

c) $$P(Y = 1 | Y \leq 2) = \frac{P(\{Y = 1\} \cap \{Y \leq 2\})}{P(Y \leq 2)} = \frac{P(\{Y = 1\})}{P(Y \leq 2)} = \frac{0.3}{0.7} = 0.4286$$

$$P(Y = 4 | Y \geq 3) = \frac{P(\{Y = 4\} \cap \{Y \geq 3\})}{P(Y \geq 3)} = \frac{P(\{Y = 4\})}{P(Y \geq 3)} = \frac{0.1}{0.3} = 0.33$$

Oppgave 2

a) $X \sim N(2.5,\ 0.3.)$.

$$\begin{aligned}
P(X < 1.9) &= P(\frac{X - 2.5}{0.3} < \frac{1.9 - 2.5}{0.3}) \\
&= P(Z < -2) = 0.0228 \\
P(X > 2.8) &= P(\frac{X - 2.5}{0.3} > \frac{2.8 - 2.5}{0.3}) \\
&= P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587
\end{aligned}$$

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

$$\begin{aligned}
P(2.2 < X < 2.8) &= P(X < 2.8) - P(X < 2.2) \\
&= P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826
\end{aligned}$$

b) $\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{4}\sum_{i=1}^{4} X_i) = \frac{1}{4}(\mu + \mu + \mu + \mu) = \mu = 2.5$

$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{1}{4}\sum_{i=1}^{4} X_i) = (\frac{1}{4})^2 \sum_{i=1}^{4} \mathrm{Var}(X_i) = (\frac{1}{4})^2(\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2) = \frac{4\sigma^2}{4^2} = \frac{\sigma^2}{4} = \frac{0.3^2}{4}$

$\overline{X} = \sum_{i=1}^{4} \frac{1}{4} X_i$ er en lineærkombinasjon av X_i'ene som er uavhengige og normalfordelte.

Derfor er også $\overline{X}$ normalfordelt. Vi får da at: $\overline{X} \sim N(2.5, \sqrt{\frac{0.3^2}{4}})$, og:

$$P(\overline{X} > 2.8) = P(\frac{\overline{X} - 2.5}{\sqrt{\frac{0.3^2}{4}}} > \frac{2.8 - 2.5}{\sqrt{\frac{0.3^2}{4}}}) = P(Z > 2) = 0.0228$$

Oppgave 3

a) Vi må anta at $X_1, \ldots, X_9$, er uavhengige og normalfordelte med $\mathrm{E}(X_i) = \mu_X$ og $\mathrm{SD}(X_i) = \sigma_X$. Siden $\mathrm{SD}(X_i) = \sigma_X$ er ukjent, må vi lage et t-intervall. Da er et 95% konfidensintervall for μ_X gitt ved:

$$\Big(\overline{X} - t_{0.025,8}\sqrt{\frac{S^2}{9}}, \quad \overline{X} + t_{0.025,8}\sqrt{\frac{S^2}{9}}\Big)$$

Vi har at $t_{0.025,8} = 2.306$. Innsatt data blir utregnet intervall:

$$\Big(97.6 - 2.306\sqrt{\frac{16.4^2}{9}}, \quad 97.6 + 2.306\sqrt{\frac{16.4^2}{9}}\Big) = (84.99, \quad 110.21)$$

b) Vi vil teste: $H_0 : \mu_X - \mu_Y = 0$ mot $H_1 : \mu_X - \mu_Y \neq 0$

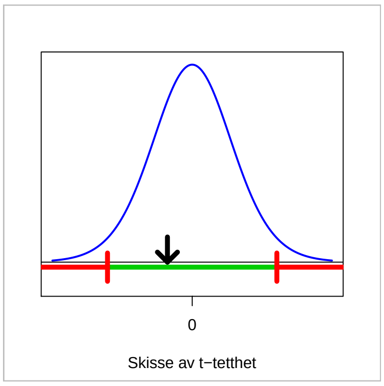
Teststørrelse: $\frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2(\frac{1}{9} + \frac{1}{9})}}$, der $S_p^2 = \frac{8S_X^2 + 8S_Y^2}{8 + 8}$. Fra dataene får vi S_p^2 : $\frac{8 \cdot 16.4^2 + 8 \cdot 17.8^2}{16} = 292.9$

Dersom H_0 er korrekt er

$$\underbrace{\frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2(\frac{1}{9} + \frac{1}{9})}}}_{\textit{teststørrelse}} \sim \underbrace{t(16)}_{\textit{nullfordeling}}$$

Med signifikansnivå 5%, dvs. $\alpha = 0.05$, er *forkastnings-området* $(\infty, -t_{0.025,16}] \cup [t_{0.025,16}, \infty)$.
Fra tabell: $t_{0.025,16} = 2.12$

Utfall av teststørrelsen: $\dfrac{92.6 - 97.6}{\sqrt{292.9(\frac{1}{9} + \frac{1}{9})}} = -0.62.$



Siden utfallet -0.62 ikke er i forkastnignsområdet, blir konklusjonen behold H_0. Dataene gir ikke grunnlag for å påstå at forventet puls er ulik ved 0 og 30 minutter.

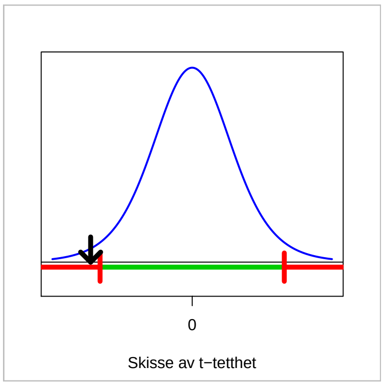
c) Test etter parplanen: Vi betrakter differansene $D_i = Y_i - X_i$. Vi har at $\mathrm{E}(D_i) = \mu_D = \mu_Y - \mu_X$. Derfor vil test av $H_0 : \mu_D = 0$ mot $H_1 : \mu_D \neq 0$ være å test de samme hypotesene som i punkt b).
La $\overline{D} = \frac{1}{9}\sum_{i=1}^{9} D_i$ og $S_D^2 = \frac{1}{8}\sum_{i=1}^{9}(D_i - \overline{D})^2$ (emiprisk varians for differansene).

Teststørrelse og nullfordeling:

$$\underbrace{\frac{\overline{D}}{\sqrt{S_D^2/9}}}_{\textit{teststørrelse}} \sim \underbrace{t(8)}_{\textit{nullfordeling}}$$

Med signifikansnivå 5%, dvs. $\alpha = 0.05$, er *forkastnings-området* $(\infty, -t_{0.025,8}] \cup [t_{0.025,8}, \infty)$.
Fra tabell: $t_{0.025,8} = 2.306$

Utfall av teststørelse: $\dfrac{-5}{\sqrt{5.9^2/9}} = -2.54 < -t_{0.025,8} = 2.306;$
vi får konklusjon forkast H_0 siden utfallet er i forkastningsområdet.



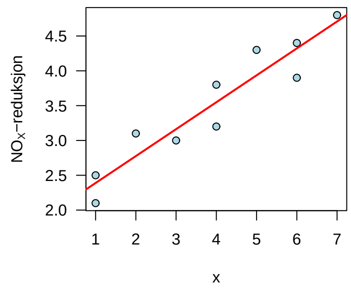
Siden vi har par av målinger på samme individ, vil vi kunne ha korrelasjon mellom X_i og Y_i. Dette bryter med forutsetningene for toutvalgstesten som bl.a. er at X_i og Y_i skal være uavhengige.

Når dataene er målinger på samme enhet (individ i dette tilfellet), vil det typisk kunne være relativt sterk positiv korrelasjon mellom X_i og Y_i. I slike tilfeller får vi redusert usikerheten i $\overline{X} - \overline{Y} = \overline{D}$ ved å bruke parplanen fordi vi da vil ha at $\mathrm{Var}(\overline{D}) < \mathrm{Var}(\overline{X}) + \mathrm{Var}(\overline{Y})$.

Oppgave 4

a)

$Y_i = \alpha + \beta x_i + e_i$, $e_i \sim N(0, \sigma)$, u.i.f.
β: forventet økt i reduksjon i NO_X-utslipp pr. enhet tilsettingsstoff.



b) $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{15.81}{40.90} = 0.3866$; $a = \overline{y} - b\overline{x} = 3.51 - 0.3866 \cdot 3.9 = 2.002$

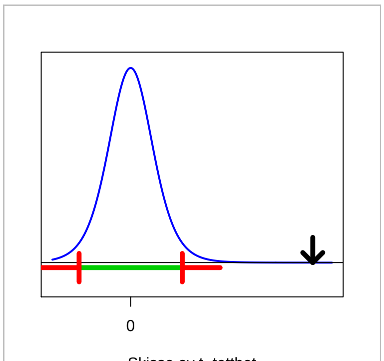
Estimat av σ^2: $S^2 = \dfrac{\text{SSE}}{n-2} : \dfrac{0.7376}{8} = 0.0922$

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

Teststørrelse: $\dfrac{\widehat{\beta}}{\sqrt{S^2/S_{xx}}}$, som er $t(n-2)$-fordelt under H_0.

Forkastningsområde: $(\infty, -t_{0.025,8}] \cup [t_{0.025,8}, \infty)$.
Fra tabell: $t_{0.025,8} = 2.306$

Utfall av teststørrelse $\dfrac{\widehat{\beta}}{\sqrt{S^2/S_{xx}}} : \dfrac{0.3866}{\sqrt{0.0922/40.90}} = 8.14$;
konklusjon: forkast H_0.



Skisse av t−tetthet

c)

i	residual: $y_i - (a + bx_i)$
1	$2.1 - (2.002 + 0.3866 \cdot 1) = -0.2886$
2	$2.5 - (2.002 + 0.3866 \cdot 1) = 0.1114$
3	$3.1 - (2.002 + 0.3866 \cdot 2) = 0.3248$

$R^2 = \dfrac{SSR}{SST}$, $SST = S_{yy}$ og $SSR = SST - SSE$; $R^2 = \dfrac{SSR}{SST} = \dfrac{SST - SSE}{SST} : \dfrac{6.85 - 0.7376}{6.85} = 0.89$

Dvs.: regresjonen forklarer ca. 89% av variasjonen.