(Det tas forbehold om feil i løsningsforslaget.)

Oppgave 1

a) Antall utvalg (uordnede) på $s = 2$ av totalt $N = 12$: $\binom{12}{2} = 66$.
Antall par der begge kan jobben: $\binom{9}{2} = 36$,
og antall par der 1 kan jobben (og 1 ikke kan): $\binom{9}{1} \cdot \binom{3}{1} = 27$ — dvs. det er $36 + 27 = 63$ par der en eller begge kan jobben (der høyest 1 ikke kan jobben).

To velges tilfeldig: $P(\text{ingen kan jobben}) = \dfrac{\text{antall utvalg der begge ikke kan}}{\text{totalt antall utvalg}} = \frac{3}{66} = \frac{1}{22}$.

Oppgave 2

La O_1 = finne olje i reservoar 1, og O_2 = finne olje i reservoar 2. Vi har da: $P(O_1) = 0.6$, $P(O_2) = 0.7$, $P(\text{finne olje i begge reservoarene}) = P(O_1 \cap O_2) = 0.45$.

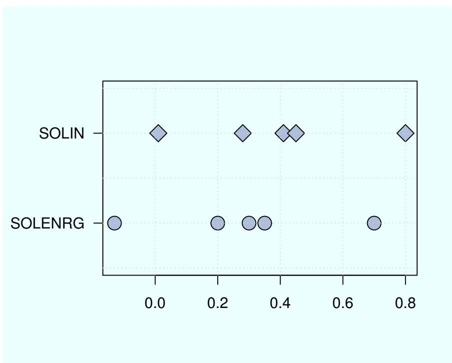
a) Finne olje i minst ett av reservoarene $= O_1 \cup O_2$; $P(O_1 \cup O_2) = P(O_1) + P(O_2) - P(O_1 \cap O_2) = 0.6 + 0.7 - 0.45 = 0.85$
Ikke finne olje i noen av de to $= (O_1 \cup O_2)^C$; $P((O_1 \cup O_2)^C) = 1 - P(O_1 \cup O_2) = 1 - 0.85 = 0.15$

b) Finne olje i kun reservoar 1 $= O_1 \cap O_2^C$; $P(O_1 \cap O_2^C) = P(O_1) - P(O_1 \cap O_2) = 0.6 - 0.45 = 0.15$ — bruk Venndiagram ! $P(O_2|O_1) = \dfrac{P(O_2 \cap O_1)}{P(O_1)} = \dfrac{0.45}{0.6} = 0.75$

c) Finne olje i kun reservoar 1 eller kun reservoar 2 $= (O_1 \cap O_2^C) \cup (O_1^C \cap O_2)$; $P\{(O_1 \cap O_2^C) \cup (O_1^C \cap O_2)\} = P(O_1) + P(O_2) - 2P(O_1 \cap O_2) = 0.6 + 0.7 - 2 \cdot 0.45 = 0.4$ — bruk Venndiagram !

Oppgave 3

Selskap	gj.sn.	datamed.	datastd.
SOLENRG	0.28	0.30	0.298
SOLIN	0.39	0.41	0.287



a) Merk: *datastd.* (for x_i'ene) $= \sqrt{\frac{1}{5-1}\sum_{i=1}^{5}(x_i - \overline{x})^2} = \sqrt{0.356/4} = 0.298$; tilsvarende for y_i'ene: $\sqrt{0.329/4} = 0.287$. Dette er estimat av $\sigma_X = SD(X_i)$ og $\sigma_Y = SD(Y_i)$, hhv.

Prikkdiagram til de to datasettene i figur til høyre.

b) Vi antar at målingene kommer fra en normalfordeling: $X_1, \dots, X_5$ er u.i.f. $N(\mu_X, \sigma_X)$. Et 95% konfidensintervall for μ_X er da gitt ved:

$$\left(\overline{x} - t_{0.025,4} \cdot s_X\sqrt{\frac{1}{5}}, \quad \overline{x} + t_{0.025,4} \cdot s_X\sqrt{\frac{1}{5}}\right)$$

Her er s_X estimat av σ_X: $s_X = \sqrt{\frac{1}{4}\sum_{i=1}^{5}(x_i - \overline{x})^2}$, og $t_{0.025,4} = 2.776$. Innsatt data:

$$\left(0.28 - 2.776 \cdot 0.30\sqrt{\frac{1}{5}}, \quad 0.28 + 2.776 \cdot 0.30\sqrt{\frac{1}{5}}\right) = (-0.09, \ \ 0.65)$$

c) Vi vil teste: $H_0 : \mu_X = 0$ mot $H_1 : \mu_X > 0$.

Teststørrelse: $T = \dfrac{\overline{X}}{s_X/\sqrt{5}}$; store verdier av T indikerer at H_1 er riktig.
Nullfordeling: $T \sim t(4)$. Forkastningsområde: $[t_{0.05,4}, \infty)$, $t_{0.05,4} = 2.132$.

Utfall av teststørrelse: $\dfrac{0.28}{0.30/\sqrt{5}} = 2.09 < 2.132$. Siden utfallet ikke er i forkastningsområdet, beholdes H_0. Dataene gir ikke grunnlag for å hevde at det er positiv forventet verdistigning for selskapet SOLENRG.

d) I tillegg til det nevnt i b), antar vi at også y_i-dataene kommer fra en normalfordeling, $Y_1, \dots, Y_5$ er u.i.f. $N(\mu_Y, \sigma_Y)$. Og vi antar at X_i'ene uavhengige av Y_i'ene. Videre kan vi anta at variansene er like: $\sigma_X^2 = \sigma_Y^2$ som ikke synes urimelig ut fra dataene.

Et 95% konfidensintervall for $\mu_X - \mu_Y$ er da gitt ved:

$$\left(\overline{x} - \overline{y} - t_{0.025,8} \cdot s_p\sqrt{\frac{1}{5} + \frac{1}{5}}, \quad \overline{x} - \overline{y} + t_{0.025,8} \cdot s_p\sqrt{\frac{1}{5} + \frac{1}{5}}\right) = (-0.54, \ \ 0.32)$$

Her er $s_p = \sqrt{\frac{\sum_{i=1}^{5}(x_i-\overline{x})^2 + \sum_{i=1}^{5}(y_i-\overline{y})^2}{5+5-2}} = 0.29$; $t_{0.025,8} = 2.306$.

e) Vi vil teste: $H_0 : \mu_X - \mu_Y = 0$ mot $H_1 : \mu_X - \mu_Y \neq 0$.
Det enkleste er å bruke konfidensintervallet fra d) for å gjennomføre testen. Siden 0 er inneholdt i konfidensintervallet beholdes H_0. Det er ikke grunnlag for å hevde at det er forskjell i virkeligheten. Dette er en test med 5% signifikansnivå, siden intervallet har konfidensgrad 95%.

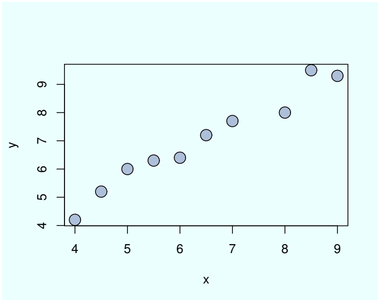
f) Vi antar at d_i'ene er utfall av fem tilfeldige variable $D_1, \dots, D_5$ som er u.i.f. $N(\mu_D, \sigma_D)$. Et 95% konfidensintervall for $\mu_D = \mu_X - \mu_Y$ er da gitt ved:

$$\left(\overline{d} - t_{0.025,4} \cdot s_D\sqrt{\frac{1}{5}}, \quad \overline{d} + t_{0.025,4} \cdot s_D\sqrt{\frac{1}{5}}\right) = (-0.24, \ \ 0.02)$$

Her er $s_D = \sqrt{\frac{\sum_{i=1}^{5}(d_i - \bar{d})^2}{5-1}} = 0.106$; $t_{0.025,4} = 2.776$.

Vi kan bruke dette intervallet til å teste: $H_0 : \mu_D = 0$ mot $H_1 : \mu_D \neq 0$.
Siden 0 er inneholdt i konfidensintervallet beholdes H_0. Det er ikke grunnlag for å hevde at det er forskjell i virkeligheten. Dette er en test med 5% signifikansnivå, siden intervallet har konfidensgrad 95%.
Siden $\mu_D = \mu_X - \mu_Y$ er dette er det samme som å teste: $H_0 : \mu_X - \mu_Y = 0$ mot $H_1 : \mu_X - \mu_Y \neq 0$.
Konklusjonen blir den samme som ved toutvalgstesten i punkt e).

Oppgave 4



Modell: $Y_i = \alpha + \beta x_i + e_i$ der $e_1, \ldots, e_n$ er u.i.f. $N(0, \sigma)$ $(n = 10)$.
(Fra utskriften: estimat av α: 0.75; estimat av β: 0.97.

a) Dvs. estimert regresjonslinje:
$\widehat{y} = 0.75 + 0.97x$.)
Prediksjoner: for dose $x_0 = 5$: $\widehat{y} = 0.75 + 0.97 \cdot 5 = 5.6$.
for dose $x_0 = 9$: $\widehat{y} = 0.75 + 0.97 \cdot 9 = 9.48$.

b) Dersom i virkeligheten $\beta = 0$, betyr det at gitt dose ikke har sammenheng med målt konsetrasjon; dersom $\beta \neq 0$, betyr det at det er sammenheng. Vi vil derfor teste:

$$H_0 : \quad \beta = 0 \qquad \text{mot} \qquad H_1 : \quad \beta \neq 0$$

Testobservator

$$T = \frac{\widehat{\beta}}{\sqrt{\frac{S^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim t_{n-2} \text{ under } H_0$$

Her er S^2 estimator for σ^2 (estimat: 0.3564^2). Vi forkaster H_0 dersom $T \leq -t_{\gamma/2,n-2} = -t_{0.025,8} = -2.306$ eller $T \geq t_{\gamma/2,n-2} = 2.306$ ($\gamma = 0.05$ er signifikansnivå). Fra utskrift: beregnet verdi av T: 14.022; Dvs. forkast H_0 — det er sammenheng mellom dose og forventet målt konsentrasjon i kroppen. (Vi kan også lese av p-verdien: 0.000000649 som er (mye!) mindre enn 0.05.)

c) Konstantleddet, α, i modellen $Y_i = \alpha + \beta x_i + e_i$ skal være null; da vil forventet målt konsentrasjon i kroppen være null dersom en pasient ikke har fått noen dose ($x_0 = 0$) (regresjonslinjen går gjennom origo).
Vi kan teste: $H_0 : \alpha = 0$ mot $H_1 : \alpha \neq 0$. Regresjonsanalysen viser p-verdi: 0.138; dvs. behold H_0; dataene gir ikke grunnlag får å hevde at $\alpha \neq 0$.

d) I denne situasjonen er summen av kvadrerte avvik:

$$\mathrm{SSE(b)} = \sum_{i=1}^{n} \Big(y_i - bx_i\Big)^2$$

Vi deriverer SSE mht. b:

$$\frac{\partial}{\partial b}\sum_{i=1}^{n}\Big(y_i - bx_i\Big)^2 = -2\sum_{i=1}^{n}\Big(y_i - bx_i\Big)x_i = -2\Big\{\sum_{i=1}^{n} x_iy_i - b\sum_{i=1}^{n} x_ix_i\Big\}$$

Vi setter den deriverte lik null og løser likningen mht. b: $\Rightarrow b = \frac{\sum_{i=1}^{n} x_iy_i}{\sum_{i=1}^{n} x_i^2}$ (løsningen er et minimumspunkt, fordi den andrederiverte er positiv). Dvs. minstekvadratersestimatoren av β i modellen $Y_i = \beta x_i + E_i$ er

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} x_iY_i}{\sum_{i=1}^{n} x_i^2}$$

Oppgave 5

Vi har: $P(-t_{0.025,9} \leq \frac{\widehat{\beta}-\beta}{\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}} \leq t_{0.025,9}) = 0.95$, der $\widehat{\beta}$ som i 4d); estimat av β: $472.4/436 = 1.08$. (Merk at frihetsgradene nå er $9 = n-1$; det er bare en forventningsparameter som estimeres.)

Vi har (denne utledningen under er strengt tatt ikke nødvendig for full skåre ;-) Resultatet under streken er.):

$$-t_{0.025,9} \leq \frac{\widehat{\beta}-\beta}{\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}} \leq t_{0.025,9}$$
$$\Updownarrow$$
$$-t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2} \leq \widehat{\beta}-\beta \leq t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}$$
$$\Updownarrow$$
$$-\widehat{\beta}-t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2} \leq -\beta \leq -\widehat{\beta}+t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}$$
$$\Updownarrow$$
$$\widehat{\beta}-t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2} \leq \beta \leq \widehat{\beta}+t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}$$

Dette betyr at: $P\left(\widehat{\beta}-t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2} \leq \beta \leq \widehat{\beta}+t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}\right) = 0.95$, og derfor er et 95% konfidensintervall for β gitt ved:

$$\left(\widehat{\beta}-t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2},\quad \widehat{\beta}+t_{0.025,9}\sqrt{S_E^2/\sum_{i=1}^{10} x_i^2}\right)$$

Innsatt data: $\left(1.08 - 2.262\sqrt{0.3889^2/436},\quad 1.08 + 2.262\sqrt{0.3889^2/436}\right) = (1.04, 1.12)$