(Det tas forbehold om feil i løsningsforslaget.)

Oppgave 1

a) $\mathrm{E}(X) = 3.55$
$\mathrm{E}(X^2) = 14.65$
$\mathrm{Var}(X) = 14.65 - 3.55^2 = 2.0475$

b) $P(X \geq 4) = 0.3 + 0.15 + 0.1 = 0.55$
$P(X \leq 2) = 0.1 + 0.15 = 0.25$

La A være begivenheten $\{X \geq 4\}$, og la B være begivenheten partall i et kast, $B = \{X = 2, X = 4, X = 6\}$. Vi skal finne $P(A \cup B)$. Vi har:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
$P(A) = 0.55$, $P(B) = 0.15 + 0.3 + 0.1 = 0.55$, $A \cap B$ er begivenheten $\{X = 4, X = 6\}$ som har sannsynlighet $P(A \cap B) = P(\{X = 4, X = 6\}) = 0.3 + 0.1 = 0.4$. Dermed får vi:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.55 + (0.15 + 0.3 + 0.1) - (0.3 + 0.1) = 0.7$$

c) $S = X_1 + X_2 + \cdots + X_{40}$

$$\mathrm{E}(S) = \mathrm{E}(X_1 + X_2 + \cdots + X_{40}) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_{40}) = 40 \cdot 3.55 = 142$$

$$\mathrm{Var}(S) = \mathrm{Var}(X_1 + X_2 + \cdots + X_{40}) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_{40}) = 40 \cdot 2.0475 = 81.9$$
$$\mathrm{SD}(S) = \sqrt{40 \cdot 2.0475} = \sqrt{81.9} = 9.05$$

$$P(S > 150) = 1 - P(S \leq 150) \approx 1 - P\left(Z \leq \frac{150 + 0.5 - 142}{9.05}\right)$$
$$= 1 - P(Z \leq 0.94) = 1 - 0.8264 = 0.1736$$

$$P(130 \leq S \leq 150) = P(S \leq 150) - P(S \leq 129)$$
$$\approx P(Z \leq 0.94) - P(Z \leq -1.38) = 0.8264 - 0.0838 = 0.7426$$

Oppgave 2

a) $P(X < 5.5) = P\Big(Z < \dfrac{5.5-6}{0.2}\Big) = P(Z < -2.5) = 0.0062$

$P(X > 6.4) = 1 - P\Big(Z < \dfrac{6.4-6}{0.2}\Big) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$

$$P(5.8 < X < 6.2) = P\Big(Z < \frac{6.2-6}{0.2}\Big) - P\Big(Z < \frac{5.8-6}{0.2}\Big)$$

$$= P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826$$

b) $\mathrm{E}(\overline{X}) = \mathrm{E}\Big\{\dfrac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)\Big\} = \cdots = \mu_X = 6$

$\mathrm{Var}(\overline{X}) = \mathrm{Var}\Big\{\dfrac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)\Big\} = \cdots = \dfrac{\sigma_X^2}{5} = \dfrac{0.2^2}{5}$; $\mathrm{SD}(\overline{X}) = \sqrt{\dfrac{0.2^2}{5}} = 0.0894$

$P(\overline{X} < 5.8) = P\Big(Z < \dfrac{5.8-6}{0.0894}\Big) = P(Z < -2.24) = 0.0125$

c) $X_i \sim N(\mu_X, 0.2)$

Test: forkast $H_0 : \mu_X = 5.5$, dersom $\dfrac{\overline{X} - 5.5}{\sqrt{\sigma_X^2/n}} = \dfrac{\overline{X} - 5.5}{\sqrt{0.2^2/11}} \geq z_{0.05} = 1.645.$

Utfall: $\dfrac{5.64 - 5.5}{\sqrt{0.2^2/11}} = 2.32$. Konklusjon: forkast H_0 siden utfallet $2.32 > z_{0.05} = 1.645$. Dataene gir grunnlag for å hevde at virkelig pH er høyere enn 5.5.

d) Styrke i alternativet μ: $\gamma(\mu) = P\Big(\dfrac{\overline{X} - 5.5}{\sqrt{\sigma_X^2/n}} \geq 1.645 \Big| \; \mu\Big)$, med $n = 11$ og $\sigma_X = 0.2$,

$= \cdots = 1 - P(Z \leq 1.645 + \dfrac{5.5 - \mu}{\sqrt{\sigma_X^2/n}}) = 1 - P(Z \leq -0.01) = 0.5053$, når $\mu = 5.6$ er satt inn.

Dimensjonering:

Test med n målinger: forkast H_0, dersom $\dfrac{\overline{X} - 5.5}{\sqrt{0.2^2/n}} \geq z_{0.05} = 1.645.$

Denne testen har styrke: $\gamma(\mu) = P\Big(\dfrac{\overline{X} - 5.5}{\sqrt{0.2^2/n}} \geq 1.645 \Big| \; \mu\Big) = 1 - P\Big(Z \leq 1.645 + \dfrac{5.5 - \mu}{\sqrt{\sigma_X^2/n}}\Big)$

Styrke lik 0.9 når i virkeligheten pH er 5.6, $\mu_X = 5.6$, betyr: $\gamma(5.6) = 0.9$, som igjen betyr $P\left(Z \leq 1.645 + \frac{5.5 - 5.6}{\sqrt{0.2^2/n}}\right) = 0.1$, og da må vi ha at

$$1.645 + \frac{5.5 - 5.6}{\sqrt{0.2^2/n}} = -z_{0.1} = -1.282 \quad \Rightarrow \quad n = \left\{\frac{(-1.282 - 1.645)0.2}{(5.5 - 5.6)}\right\}^2 = 34.3,$$ dvs. vi må ha $n = 35$ målinger.

Oppgave 3

a) $Y_i \sim N(\mu_Y, \sigma_Y)$; ukjent σ_Y (og ukjent μ_Y).

Test (t-test): forkast $H_0 : \mu_Y = 6.0$ dersom $\dfrac{\overline{Y} - 6}{\sqrt{S_Y^2/n_Y}} \leq -t_{0.05,n_Y-1} = -1.812$.

Utfall av teststørrelse: $\dfrac{5.98 - 6}{\sqrt{0.12^2/11}} = -0.55$

Konklusjon: Behold H_0 siden utfallet ikke er i forkastningsområdet ($-0.55 > -t_{0.05,10} = -1.812$). Dataene gir ikke grunnlag for å hevde at virkelig pH er lavere enn 6.0.

b) 90% konfidensintervall for μ_Y: $\left(\overline{Y} - t_{0.05,10}\sqrt{\frac{S_Y^2}{n_Y}}, \quad \overline{Y} + t_{0.05,10}\sqrt{\frac{S_Y^2}{n_Y}}\right)$

$$\left(5.98 - 1.812\sqrt{\frac{0.12^2}{11}}, \quad 5.98 + 1.812\sqrt{\frac{0.12^2}{11}}\right) = (5.91, \quad 6.05)$$

Fortolkning: Vi kan ha god grunn til å tro at den virkelige pH'en er i intervallet (5.91, 6.05). Dersom vi tenker oss 11 slike målinger gjort mange ganger, og hver gang beregnet et 90% konfidensintervall, så ville ca. 90% av disse intervallene dekket den virkelige pH'en, μ_Y. Derfor har vi god grunn til å tro at den virkelige pH'en er inneholdt i det intervallet vi har med våre data, (5.91, 6.05).

Oppgave 4

a) Antakelsene oppsummert: $X_i \sim N(\mu_X, \sigma_X)$, $Y_i \sim N(\mu_Y, \sigma_Y)$, alle X_i'ene og Y_i'ene uavhengige (spesielt X_i uavhengig av Y_i i samme år i) og $\sigma_X = \sigma_Y$. Dette er da en situasjon der vi skal benytte toutvalgs t-test av

$H_0 : \mu_X - \mu_Y = 0$ mot $H_1 : \mu_X - \mu_Y \neq 0$

Teststørrelse: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{S^2_{\text{pooled}}(\frac{1}{n_X} + \frac{1}{n_Y})}}$, der $S^2_{\text{pooled}} = \dfrac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$.

Nullfordeling: $t(n_X + n_Y - 2)$.

Forkastningsområde (for test med 5% sign.nivå): $(-\infty, -t_{0.025,20}] \cup [t_{0.025,20}, \infty)$,
$t_{0.025,20} = 2.086$

Utfall: S^2_{pooled} : 0.02925; Teststørrelse: $\dfrac{5.64 - 5.98}{\sqrt{0.02925(\frac{1}{11} + \frac{1}{11})}} = -4.66$;

Konklusjon: forkast H_0 siden $-4.66 < -t_{0.025,20} = -2.086$ er i forkastningsområdet. Dataene gir grunnlag for å hehvde at det er forskjellig pH ved de to stedene i elven.

b) Det kan være tvil om

1. X_i'ene er uavhengige av Y_i'ene. Av øverste figur ser pH-målinger ved Fidjeland og Nesset samme år ut til å være positivt korrelerte; og
2. prikkdiagrammet (og estimatene av σ_X og σ_Y) kan tyde på ulike varianser for X_i'ene og Y_i'ene.)

Begge punktene over er forutsetninger som er antatt innledningsvis (og for toutvalgstesten gjennomført i i a)).
Dersom man gjør en test etter parplanen vil ikke dette være noe problem fordi den forutsetter ikke uavhengighet mellom X_i'ene og Y_i'ene, OG
det betyr heller ikke noe om $\text{Var}(X_i) \neq \text{Var}(Y_i)$ siden vi da ser på $D_i = X_i - Y_i$; (alle $D_1, \ldots, D_{11}$ vil ha samme varians, $\text{Var}(D_i) = \text{Var}(X_i) + \text{Var}(Y_i) - 2\text{Cov}(X_i, Y_i)$).

Oppgave 5

a) La X være antall blant de 50 mennene som er for, og la p_X være andel av alle mannlige studenter på campus som er for. Vi kan da bruke at $X \sim B(n, p_X)$, med $n = 50$. Et tilnærmet 90% konfidensintervall for p_X er gitt ved:

$$\left(\widehat{p}_X - z_{0.05}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{50}}, \quad \widehat{p}_X + z_{0.05}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{50}}\right), \text{ der } \widehat{p}_X = \frac{X}{50} \text{ (utfall: } 30/50 = 0.6\text{)},$$
og $z_{0.05} = 1.645$.

Innsatt data: (0.486, 0.714)

b) Toutvalgs test av $H_0 : p_X - p_Y = 0$ mot $H_1 : p_X - p_Y \neq 0$

Teststørrelse: $\dfrac{\widehat{p}_X - \widehat{p}_Y}{\sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_X} + \frac{1}{n_Y})}}$,
her er $\widehat{p} = \dfrac{X+Y}{n_X + n_Y}$ (utfall: $50/100 = 0.5$), og $\widehat{p}_Y = \dfrac{Y}{50}$ (utfall: $20/50 = 0.4$).

Nullfordeling: tilnærmet $N(0,1)$.

Forkastningsområde (for test med 5% sign.nivå): $(-\infty, -z_{0.025}] \cup [z_{0.025}, \infty)$,
$z_{0.025} = 1.96$

Utfall av teststørrelse: 2; konklusjon: forkast H_0 siden utfallet er i forkastningsområdet ($2 > z_{0.025}$). Undersøkelsen gir grunnlag for å hevde at andel som er for saken er ulik for kvinner og for menn blant alle UiS-studentene.