(Det tas forbehold om feil i løsningsforslaget.)

Oppgave 1

a) Vi har 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 ($2 > z_{0.025}$).

Oppgave 2

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) $P(\text{begge oddetall}) = P(\text{oddetall i kast en, og oddetall i kast to})$
$= P(\text{oddetall i kast en})P(\text{oddetall i kast to})$, siden resultatene i ulike kast kan antas å være uavhengige. Derfor: $P(\text{begge oddetall}) = P(\text{oddetall i kast en})P(\text{oddetall i kast to}) = 0.45 \cdot 0.45 = 0.2025$.
$P(\text{sum minst } 10) = P(X_1 = 6 \cap X_2 = 4) + P(X_1 = 4 \cap X_2 = 6) + P(X_1 = 5 \cap X_2 = 5) + P(X_1 = 5 \cap X_2 = 6) + P(X_1 = 6 \cap X_2 = 5) + P(X_1 = 6 \cap X_2 = 6) = 2 \cdot 0.3 \cdot 0.10 + 0.15 \cdot 0.15 + 2 \cdot 0.15 \cdot 0.10 + 0.10 \cdot 0.10 = 0.1225$.

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

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

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

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

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

Oppgave 3

a) $P(X < 6) = P\Big(Z < \dfrac{6 - 6.5}{0.2}\Big) = P(Z < -2.5) = 0.0062$

$$P(X > 6.9) = 1 - P\Big(Z < \frac{6.9 - 6.5}{0.2}\Big) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

$$P(6.3 < X < 6.5) = P\Big(Z < \frac{6.5 - 6.5}{0.2}\Big) - P\Big(Z < \frac{6.3 - 6.5}{0.2}\Big)$$
$$= P(Z < 0) - P(Z < -1) = 0.5 - 0.1587 = 0.3413$$

b) $\mathrm{E}(\overline{X}) = \mathrm{E}\Big\{\dfrac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)\Big\} = \dfrac{1}{5}5\mu_X = \mu_X = 6.5$

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

$$P(\overline{X} < 6.3) = P\Big(Z < \frac{6.3 - 6.5}{0.0894}\Big) = P(Z < -2.24) = 0.0125$$

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

Test: forkast $H_0 : \mu_X = 6.0$, dersom $\dfrac{\overline{X}-6}{\sqrt{\sigma_X^2/n}} = \dfrac{\overline{X}-6}{\sqrt{0.2^2/6}} \leq -z_{0.05} = -1.645$.

Utfall: $\dfrac{5.84-6}{\sqrt{0.2^2/6}} = -1.79$. Konklusjon: forkast H_0 siden utfallet $-1.79 < -z_{0.05} = -1.645$. Dataene gir grunnlag for å hevde at virkelig pH er lavere enn 6.

d) Styrke i alternativet μ: $\gamma(\mu) = P\Big(\dfrac{\overline{X}-6}{\sqrt{\sigma_X^2/n}} \leq -1.645 \Big|\ \mu\Big)$, med $n=6$ og $\sigma_X = 0.2$,

$$= \cdots = P(Z \leq -1.645 + \frac{6-\mu}{\sqrt{\sigma_X^2/n}}) = P(Z \leq -0.42) = 0.3372, \text{ når } \mu = 5.9 \text{ er satt inn.}$$

<u>Dimensjonering:</u>

Test med n målinger: forkast H_0, dersom $\dfrac{\overline{X}-6}{\sqrt{0.2^2/n}} \leq -z_{0.05} = -1.645$.

Denne testen har styrke: $\gamma(\mu) = P\Big(\dfrac{\overline{X}-6}{\sqrt{0.2^2/n}} \leq -1.645 \Big|\ \mu\Big) = P\Big(Z \leq -1.645 + \dfrac{6-\mu}{\sqrt{\sigma_X^2/n}}\Big)$

Styrke lik 0.8 når i virkeligheten pH er 5.9, $\mu_X = 5.9$, betyr: $\gamma(5.9) = 0.8$, som igjen betyr $P\Big(Z \leq -1.645 + \dfrac{6-5.9}{\sqrt{0.2^2/n}}\Big) = 0.8$, og da må vi ha at

$$-1.645 + \frac{6-5.9}{\sqrt{0.2^2/n}} = z_{0.2} = 0.84 \quad \Rightarrow \quad n = \left\{\frac{(0.84+1.645)0.2}{(6-5.9)}\right\}^2 = 24.7,$$ dvs. vi må ha $n = 25$ målinger.

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 <u>toutvalgs t-test</u> av null- og alternativhypotesene:

$H_0 : \mu_X - \mu_Y = 0$ 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): $(-\infty, -t_{0.025,10}] \cup [t_{0.025,10}, \infty)$,
$t_{0.025,10} = 2.228$

Utfall: $S^2_{\text{pooled}} : 0.03365$; Teststørrelse: $\dfrac{5.84-6.035}{\sqrt{0.03365(\frac{1}{6}+\frac{1}{6})}} = -1.84$;

Konklusjon: Behold H_0 siden utfallet av teststørrelsen ikke er i forkastningsområdet, $-t_{0.025,10} = -2.228 < -1.32 < t_{0.025,10} = 2.228$. Dataene gir ikke grunnlag for å hevde at det er forskjellig pH i de to vannene.

b) Det kan være tvil om

1. X_i'ene er uavhengige av Y_i'ene. Av figuren til venstre ser pH-målinger ved Mosvatnet og Breiavatnet samme år ut til å være positivt korrelerte; og
2. prikkdiagrammet (og estimatene av σ_X og σ_Y) kan (kaaanskje) 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 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 $D_1, \ldots, D_6$ vil ha samme varians, $\text{Var}(D_i) = \text{Var}(X_i) + \text{Var}(Y_i) - 2\text{Cov}(X_i, Y_i)$).

Test med parplanen, $H_0 : \mu_D = \mu_X - \mu_Y = 0$ mot $H_1 : \mu_D = \mu_X - \mu_Y \neq 0$
Teststørrelse: $\dfrac{\overline{D}}{\sqrt{S_D^2/6}}$
Test: forkast H_0 dersom $\dfrac{\overline{D}}{\sqrt{S_D^2/6}}$ er mindre enn $-t_{0.025,5} = -2.571$ eller større enn $t_{0.025,5} = 2.571$.

Utfall av teststørrelse: $\dfrac{-0.195}{\sqrt{0.02243/6}} = -3.189$; konklusjon: forkast H_0.

Vi har her:

Utregninger:

x_i:	5.63	6.09	6.15	5.66	5.71	5.80
y_i:	5.85	6.08	6.21	6.07	5.95	6.05
$d_i = x_i - y_i$:	-0.22	0.01	-0.06	-0.41	-0.24	-0.25

$$s_D = \sqrt{\sum_{i=1}^{6}(d_i - \overline{d})^2} = \sqrt{0.02243}.$$

Oppgave 5

Vi betrakter antall personer i forrige måling som sier de vil stemme AP, som utfall av $X \sim B(n, p_X)$, med $n = 1000$; tilsvarende for den siste målingen: $Y \sim B(n, p_Y)$, med $n = 1000$. Vi har da utfallene 290 av X og 244 av Y.

For å svare på spørsmålet om dette er et uttrykk for virkelig endring i Ap sin oppslutning, kan vi gjennomføre en statistisk test av hypotesene: $H_0 : p_X - p_Y = 0$ mot $H_1 : p_X - p_Y \neq 0$.

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: $(244 + 290)/2000 = 0.267$), $\widehat{p}_X = \dfrac{Y}{1000}$ (utfall: $290/1000 = 0.290$) og $\widehat{p}_Y = \dfrac{Y}{1000}$ (utfall: $244/1000 = 0.244$).

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: 2.325; konklusjon: forkast H_0 siden utfallet er i forkastningsområdet $(2.325 > z_{0.025})$.