Løsningsforslag til obligatorisk innlevering 3.

Oppgave 1

* Vi skal sammenlikne to sensoere A og B. Begge har rettet den samme oppgaven. Hvis populasjonen er eksamensoppgavene, har vi altså brukt to metoder på hvert element i populasjonen og vi velger en paret test.
* Vi velger en t-test siden standardavviket er kjent.
* Vi bestemmer et 95% konfidensintervall for differansen d. Dersom 0 ikke er inkludert i intervallet er det grunnlag for å si at det er forskjell på sensorene.

Beregningene:

$$\bar{d} = 0.76 / 13 = 0.0584$$

$$s_d^2 = \frac{1}{n-1}\left(\sum_{i=1}^{13} d_i^2 - n\,\bar{d}^2\right) = 0.0816$$

$$s_d = 0.0904$$

Konfidensintervallet blir:

$$\left[\bar{d} \pm t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}\right] \quad \text{der } t_{\alpha/2} = 2.56 \ (12 \text{ frihetsgrader})$$

$$[0.0583 \pm 0.0642] \Rightarrow [-0.006,\ 0.123]$$

Konklusjonen blir:
Siden 0 er inkludert i konfidensintervallet, er det ikke grunnlag for å si at det er forskjell i sensorenes bedømmelse, når vi krever en konfidensgrad på 95%.

Oppgave 2

■ a)

Vi velger en binomisk modell. Sannsynligheten for at er frø skal spire er p = 0.6 og den er den samme for alle frøene. Sannsynligheten for at ingen av de 4 frøene spirer er:

$$P(x = 0) = \binom{0}{4} p^0 (1 - p)^4 = 0.0256$$

$$P(x \geq 2) = 1 - P(x \leq 1) = 1 - 0.179 = 0.821$$

■ b)

Vi har fortsatt en binomisk fordeling. Enten spirer noen av frøene eller de gjør det ikke. Sannsynligheten for at 4 før ikke spirer fant vi i oppgave a, og den er p = 0.0256. Siden vi her har 20 elever (n = 20) kan vi her gå over til en Poisson fordeling med λ= n p.

$$\lambda = 0.512$$
$$P(Y = 1) = 0.307$$
$$E(Y) = n\,p = 0.512$$
$$\text{Var}(Y) = n\,p\,(1 - p) = 0.499$$

■ c)

Sannsynligheten for å gå med overskudd når μ = 17 og σ = 3. Vi har her en normalfordeling og skal finne P (W >12)

$$P(W > 12) = 1 - P(W < 12) = 1 - G\left(\frac{12 - 15}{3}\right)$$
$$= 1 - 0.0478 = 0.952$$

■ d)

Vi har her en sum av normalfordelte størrelser. Vi kan da ut fra sentralgrenseteoremet bruke en normalfordeling på summen.

$$\mathbf{E(V) = E(W_1) + \;.....\; + E(W_5) = 75}$$
$$\mathbf{Var(V) = Var(W_1) + \;.....\; + Var(W_5) = 47}$$

$$P(v > 85) = 1 - P(W < 85) = 1 - G\left(\frac{85 - 75}{\sqrt{47}}\right)$$
$$= 1 - 0.928 = 0.072$$

Oppgave 3

a)

Median er middelverdien av de 2 mitterste verdiene 4.9 og 5.1 $\tilde{x} = 5.0$

Middelverdien $\bar{x} = 5.1$, standardavviket $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{6} (x_i - \bar{x})^2} = 0.341$

b)

Et 95% konfidensintervall for forventningsverdien er

$\left[\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}\right] = [4.74, 5.457]$ her er $t_{\alpha/2} = 2.571$

c)

Med et nytt forsøk med 8 prøver der forventningsverdi og standardavvik på pH- verdiene er henholdsvis 5.6 og 0.38 finner vi et 95% konfidensintervall for differansen: $\left[\mu_1 - \mu_2 \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_2} + \frac{1}{n_2}}\right]$

her er $S_p = \sqrt{\frac{(n_1-1) S_1^2 + (n_1-1) S_2^2}{n_1+n_2-2}}$ S_p^2 kalles for interpolert varians.

Antall frihetsgrader for $t_{\alpha/2}$ er $n_1 + n_2 - 2$.

Utregnet får vi et konfidensintervall på: [0.0716, 0.928]

$t_{\alpha/2} = 2.179$

$S_p = 0.364$

$\sqrt{\frac{1}{n_2} + \frac{1}{n_2}} = 0.54$

Vi ser at 0 ikke er inkludert i konfidensintervallet og kan konkludere med at det er forskjell på pH verdiene i de to blandingene.

■ d)

Vi har her et sett med kategoriske data og må analysere den med kategorisk kryss tabell.

$$X_{i,j} = \begin{pmatrix} 17 & 9 \\ 25 & 8 \\ 12 & 6 \end{pmatrix}$$

Forventningsverdiene blir:

$$E_{i,j} = \begin{pmatrix} 18.23 & 7.77 \\ 23.14 & 9.86 \\ 12.62 & 5.38 \end{pmatrix}$$

Avvikskvadratene delt på forventningsverdiene

$$Q_{i,j} = \begin{pmatrix} 0.083 & 0.150 \\ 0.195 & 0.351 \\ 0.030 & 0.071 \end{pmatrix}$$

Summen av alle $Q_{i,j}$ verdiene $Q = 0.88$

Hvis vi sammenlikner denne Q verdien med $Q_{kritisk}$ fra Kji -kvadrattabellen finner vi at $Q < Q_{kritisk}$ Hvis vi velger et 5% signifikansnivå for en hypotesetest der vi kontrollerer om det er forskjell på testene.

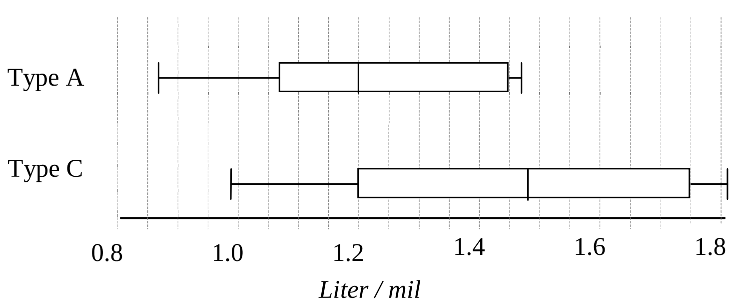
$Q_{kritisk} = 5.99$ med 2 frihetsgrader og $\alpha = 5\%$

Oppgave 4

■ a)

Nullhypotesen for bensinforbruket for de forskjellige bensintypene:

H_0: Det er ingen forskjell på bensintypene.



Middelverdien til forbruket av bensintype A er minst og bensintype C er størst.

Vi ser at midtre del av boksplottene over beninforbruket for beste og dårligste bensin overlapper hverandre godt. Boksplottene gir ikke et grunnlag for å si at den ene bensinen er vesentlig bedre enn den andre.

■ b)

En-faktor variansanalysen gir ikke et grunnlag for å forkaste nullhypotesen. F-verdien er lavere enn $F_{kritisk}$.

To-faktor analysen gir en F-verdi er høyere enn $F_{kritisk}$, og vi kan forkaste nullhypotesen. Vi ser at analysen skiller både på biltype og bensintype. Vi ser av analysen at biltypen er mer vesentlig for bensinforbruket en selve bensintypen.
Når vi forkaster nullhypotesen så kan vi gjøre det med en konfidensgrad gitt ved 1-p verdien. Vi ser at det her er tilnærmet 100%.

■ c)

Et 98% symmetrisk konfidensintervall for det midlere bensinforbruket er gitt ved:

$$\left[\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right]$$

der s^2 er variansen til bensinforbruket i bil 8

Data hentet fra variansanalysen $\bar{x} = 1.31$ og $s = \sqrt{0.01489} = 0.122$

Et 98% tosidig konfidensintervall for det midlere bensinforbruket til bil 8 er:

$$\left[1.31 - 4.541 \cdot \frac{0.122}{\sqrt{4}}, 1.31 + 4.541 \cdot \frac{0.122}{\sqrt{4}}\right] = \underline{\underline{[1.03, 1.59]}}$$

Et 98% tosidig konfidensintervall for standardavviket på bensinforbruket til bil 8 er:

$$\left[\frac{(n-1)s^2}{\chi_{\alpha/2}}, \frac{(n-1)s^2}{\chi_{(1-\alpha/2)}}\right] = \left[\frac{3 \cdot 0.0149}{11.34}, \frac{3 \cdot 0.0149}{0.11}\right]$$

=> Konfidensintervallet til standardavviket blir:

$$\left[\sqrt{0.0039}, \sqrt{0.406}\right] = \underline{\underline{[0,062, 0.637]}}$$

Oppgave 5

▪ a)

Kan firmaet ut fra de 14 problemene velge ut en av studentene? I tilfelle hvilken student.
I denne oppgaven bruker vi en paret test, slik at det er uaktuelt å studere F- verdien i Fisher-testen Vi skal altså bruke en t-test for parvise utvalg.

Vi ser at T kritisk, både ensidig og tosidig er mindre enn t-statistikk og vi kan konkludere med at det er en statistisk signifikant forskjell på student Y og X. Firmaet bør velge student Y.

T-Test: Gjennomsnitt for to parvise utvalg

	CPU tid student X	*CPU tid student Y*
Gjennomsnitt	2.137	1.647
Varians	1.149	0.704
Observasjoner	14	14
Pearson-korrelasjon	0.807	
Antatt avvik mellom gjennomsnittene	0	
fg	13	
t-Stat	2.894	
P(T<=t) ensidig	0.006	
T-kritisk, ensidig	1.771	
P(T<=t) tosidig	0.013	
T-kritisk, tosidig	2.160	

▪ b)

Testens p-verdi leses ut fra tabellen.Her er det naturlig å velge p-ensidig og p-verdien er 0.006.p-verdien forteller hvor stor sannsynlighet vi har for å begå en feil,når vi sier at det er forskjell på X og Y.