CIMT Further Statistics P34 Ex 2.5

1. X = time taken to shave

$n = 7$

$\bar{x} = 231.571$

$s_{n-1} = 15.0206$

we assume $X \sim N(\mu, \sigma^2)$

so $\bar{X} \sim N(\mu, \frac{\sigma^2}{7})$ where $\bar{X}$ = mean shaving time

so $\dfrac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{7}}} \sim N(0, 1^2)$

we estimate σ with s_{n-1} and n is small so we use t_6 distribution

$\dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{7}}} \sim t_6$

so 95% CI for μ is $\bar{x} \pm t_{6,0.975} \times \sqrt{\frac{s^2}{7}}$

$= 231.571 \pm 2.44691 \times \sqrt{\frac{15.0206^2}{7}}$

$= (217.68, 245.463)$

$\approx (217.7, 245.5)$ to 1dp.

3.

proportion that fail $= 0.03$

$X =$ working life of component, in hours

$E(X) = 2400$

$Var(X) = 650^2$

Supplier improves design so that $Var(X) = 300^2$

a) $F =$ no. components that fail

$F \sim B(310, p)$ $\quad \hat{p} = \frac{12}{310}$

we approximate F with $G \sim N(310p, 310pq)$, acceptable as $310\hat{p} > 5$ and $310\hat{q} > 5$ ✓

let $\frac{G}{310} =$ proportion of components that fail

$\frac{G}{310} \sim N\left(p, \frac{pq}{310}\right)$

so 95% CI for p is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{310}}$

$= \frac{12}{310} \pm 1.96\sqrt{\frac{\frac{12}{310} \times \frac{298}{310}}{310}}$

$= (0.017236, 0.060183)$

$\approx \underline{(0.017, 0.060)}$ to 3dp.

b) $n = 3$

$\bar{x} = \frac{8150}{3} = 2716.66\ldots$

we assume that $X \sim N(\mu, 300^2)$

so $\bar{X} \sim N\left(\mu, \frac{300^2}{3}\right)$

so 90% CI for μ is $2716.666\ldots \pm 1.645 \times \sqrt{\frac{300^2}{3}}$

$= (2431.77, 3001.56)$

$\approx \underline{(2431.8, 3001.6)}$ to 1dp

we seek n so that $1.645 \times \sqrt{\frac{300^2}{n}} < 50$

$\frac{300^2}{n} < 924.029$

$97.3996 < n$

so we would $\underline{n \text{ to be } 98}$.

c) for b) i), with a sample of size 3, we would be very uncertain of our CI. However, for b) ii), with such a large sample of 98, we would be much more confident of our CI as normality can be justified through the use of the Central Limit Theorem.

d) The new design has a longer working life, as $(2431.8, 3001.6)$ is larger than 2400, and they have reduced variability as $300 < 650$, so the 2nd design is preferred.

5 X = weight of baking powder

$X \sim N(\mu, 7^2)$

nominal value for $\mu = 200$.

So i) $\mu > 200$

ii) $P(X < 191) < 0.025$

iii) $P(X < 182) < 0.001$

if $X \sim N(\mu, 7^2)$

$\Rightarrow \bar{X} \sim N(\mu, \frac{7^2}{30})$

so 95% CI for $\mu = \bar{x} \pm 1.96\sqrt{\frac{7^2}{30}}$

$= 204.933 \pm 1.96\sqrt{\frac{7^2}{30}}$

$= (202.428, 207.438)$

$\approx (202.4, 207.4)$ to 1dp.

b) in sample, 6 out of 30 were less than 191g.

let Y = no. bags less than 191

$Y \sim B(30, p)$

approx Y with normal $W \sim N(30p, 30pq)$

so $\frac{Y}{30}$ = proportion bags less than 191

so $\frac{Y}{30} \sim N(p, \frac{pq}{30})$

so 95% CI for $p = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{30}}$

$= \frac{6}{30} \pm 1.96\sqrt{\frac{6}{30} \cdot \frac{24}{30} \cdot \frac{1}{30}}$

$= (0.056864, 0.343136)$

$\approx (0.057, 0.343)$ to 3dp.

c) if $\mu = 202.4$, then $X \sim N(202.4, 7^2)$ [using exact value for $\mu = 202.428...$]

$\Rightarrow P(X < 182) = P(Z < \frac{182 - 202.4}{7})$

$= P(Z < -2.91835)$

$= 0.00176$

Hence, proportion less than 182g is $0.00176 > 0.001$ ☹

d) The last calculation points to the machine not meeting the requirement for 182g bags, but only by a small amount. I would recommend that they increase the mean weight of the filling machine to reduce this shortcoming, as it will likely be cheaper than reducing the standard deviation of 7.

8. a) P(refuse to answer questionnaire) = 0.2

X = no. of miners who refuse

$X \sim B(12, 0.2)$

i) $P(X \leq 3) = 0.794569$ from binomcdf (12, 0.2, 0, 3)

≈ 0.7946 (4dp)

ii) $P(X = 3) = 0.236223$

≈ 0.2362 (4dp)

iii) $P(X \leq 2) = 0.558346$ [note it's 10 to agree to answer]

≈ 0.5583 (4dp)

b) 26 out of 70 were returned

let X = no. questionnaires returned

$X \sim B(70, p)$ $\quad \hat{p} = \frac{26}{70}$

approx X with $Y \sim N(70p, 70pq)$ — ok as $70\hat{p} > 5$ and $70\hat{q} > 5$

so $\frac{Y}{70}$ = proportion who returned questionnare

so $\frac{Y}{70} \sim N(p, \frac{pq}{70})$

so 95% CI for $p = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{70}}$

$= (0.258237, 0.48462)$

$\approx (0.2582, 0.4846)$ to 4dp

c) 10 out of 26 gained employment so $\hat{p} = \frac{10}{26}$

let G = no. gained employment

$G \sim B(n, p)$

approximated G with $H \sim N(np, npq)$

let $\frac{H}{n}$ = proportion who have gained employment

so $\frac{H}{n} \sim N(p, \frac{pq}{n})$

we want 95% CI to have width of 0.1 $\Rightarrow 1.96\sqrt{\frac{pq}{n}} < 0.05$

$\sqrt{\frac{pq}{n}} < 0.025511$

$\frac{pq}{n} < 0.000651$

if we estimate p with $\hat{p} = \frac{10}{26}$, then $n > 363.688$

so we would need 364 replies to have a 95% CI of width 0.1 for p. — that's a lot!!

d) if we need 364 replies and we only received $\frac{26}{70}$ return rate, then a total of $364 \times \frac{70}{26} = 980$ questionnaires need to be posted in order to receive back sufficient information to make accurate estimates of the proportion of minors now back in employment.