Chapter Review 3

1 a The null hypothesis is that $\mu = 0.48$. The alternative hypothesis is that $\mu \neq 0.48$.

$H_0 : \mu = 0.48$

$H_1 : \mu \neq 0.48$.

Using a significance level of 10%, the confidence interval is $(0.4533,\ 0.5227)$ and 0.48 lies within this interval so we accept H_0.

b First, we require the mean and standard deviation of the sample. The mean is given to be 0.482 kg. We calculate the standard deviation by using part **a** and working backwards.
Since the confidence interval (in part **a**) is $(0.4533,\ 0.5227)$, we know the first sample mean to be:

$$\mu_1 = \frac{0.4533 + 0.5227}{2} = 0.488$$

We also know that the form of a 90% confidence interval is $\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$

We take the higher value of

$$0.5227 = \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

$$0.5227 = 0.482 + 1.96 \times \frac{\sigma}{\sqrt{80}}$$

to rearrange for $\sigma = 0.189$.

Using the statistics tables, we find that the value that corresponds with a 95% confidence interval is 1.96.

This means our 95% confidence interval is found by substituting the values we have found into the equation:

$$\begin{aligned}\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}} &= 0.482 \pm 1.96 \times \frac{0.189}{\sqrt{120}} \\ &= 0.482 \pm 0.0338 \\ &= (0.4482,\ 0.5158)\end{aligned}$$

2 a i $\frac{2X_1 + X_{20}}{3}$ is a statistic since it only contains known data.

ii $\sum_{1}^{20} (X_i - \mu)^2$ is not a statistic since it contains unknown population parameters.

iii $\frac{\sum_{1}^{20} X_i^2}{n}$ is a statistic since it only contains known data.

2 b The mean of the statistic $\frac{4X_1 - X_{20}}{3}$ is $\frac{4\mu - \mu}{3} = \mu$ and the variance is:

$$\begin{aligned} Var\left(\frac{4X_1 - X_{20}}{3}\right) &= Var\left(\frac{4X_1}{3}\right) + Var\left(\frac{X_{20}}{3}\right) \\ &= \left(\frac{4}{3}\right)^2 Var(X_1) + \left(\frac{1}{3}\right)^2 Var(X_{20}) \\ &= \frac{16}{9}\sigma^2 + \frac{1}{9}\sigma^2 \\ &= \frac{17}{9}\sigma^2 \end{aligned}$$

3 a 95% C.I. is $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

$= 47.15 \pm 1.96 \times \frac{2.4572\ldots}{\sqrt{100}}$

$= (46.6683\ldots, 47.6316\ldots)$

$= (46.7, 47.6)$ (3 s.f.)

N.B. Since n is large, we have assumed $s = \sigma$.

b $\bar{x} = \frac{4715}{100} = 47.15$

$s = \sqrt{\frac{222\,910 - 100 \times 47.15^2}{99}} = 2.4572\ldots$

$H_0 : \mu = 46.50$ (no better) $H_1 : \mu > 46.50$

t.s. is $z = \frac{(47.15 - 46.50)}{\left(\frac{2.4572\ldots}{\sqrt{100}}\right)} = 2.645\ldots$

5% c.v. is $z = 1.6449$

t.s. is $2.645 > 1.6449$

Result is significant so reject H_0.

There is evidence that the new bands are better.

4 a $n = 100$, $\sum x = 453$, $\sum x^2 = 2391$

$\bar{x} = \frac{453}{100} = 4.53$

$s = \sqrt{\frac{2391 - 100 \times 4.53^2}{99}} = 1.85022\ldots$

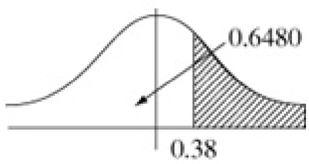
Standard error $= \frac{s}{\sqrt{n}} = 0.185$ (3 d.p.)

4 b 98% C.I. is $\bar{x} \pm 2.3263\dfrac{\sigma}{\sqrt{n}}$ Use $\dfrac{s}{\sqrt{n}}$

$= (4.0995\ldots, 4.9604)$

$= (4.10, 4.96) \quad (3\text{ s.f.})$

c



$\mathrm{P}(\bar{x} > 4.6) = \mathrm{P}\left(Z > \dfrac{4.6 - 4.53}{0.185\ldots}\right)$

$= \mathrm{P}(Z > 0.378\ldots)$ Use 0.38

$= 1 - 0.6480$

$= 0.3520$ (tables)

or $= 0.35259\ldots$ (calculator)

So accept awrt $0.352 \sim 0.353$

5 $n = 300, \sum x = 4203, \sum x^2 = 59\ 025$

$\bar{x} = \dfrac{4203}{300} = 14.01$

$s = \sqrt{\dfrac{59\ 025 - 300 \times 14.01^2}{299}} = 0.6866\ldots$

Standard error $= \dfrac{s}{\sqrt{n}} = 0.039643\ldots = 0.04\ (2\text{ d.p.})$

6 $\sigma = 0.43$

Width of 99% C.I. is $2 \times 2.5758\dfrac{\sigma}{\sqrt{n}}$

Require $\dfrac{2 \times 2.5758 \times 0.43}{\sqrt{n}} < 0.60$

$\therefore \sqrt{n} > \dfrac{2 \times 5.2758 \times 0.43}{0.6} = 3.691\ldots$

$n > 13.63\ldots$

So the smallest value of n is 14

7 a $\sigma = 8.0$

$n_A = 25 \quad \bar{x}_A = 44.2$

95 % C.I. is $44.2 \pm 1.96 \times \dfrac{8.0}{\sqrt{25}}$

$= (41.064, 47.336)$

$= (41.1, 47.3) \quad (3\text{ s.f.})$

7 b $n_B = 20 \quad \bar{x}_B = 40.9$

$\text{H}_0 : \mu_A = \mu_B \quad \text{H}_1 : \mu_B < \mu_A$ 5% c.v. is $z = -1.6449$

$$\text{t.s. is } z = \frac{(40.9 - 44.2) - 0}{\sqrt{\frac{8^2}{20} + \frac{8^2}{25}}} = -1.375 > -1.6449$$

Not significant so accept H_0.

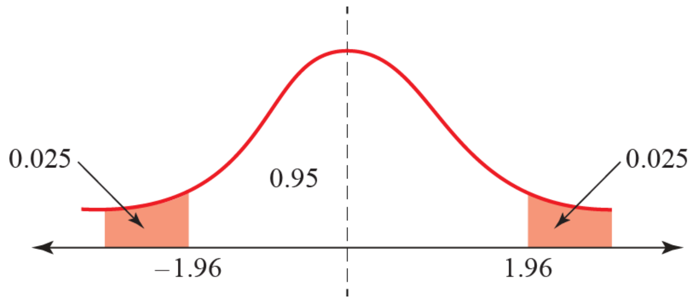
There is insufficient evidence to support the head-teacher's claim.

8 a $X \sim \text{N}\left(\mu,\ \sigma^2\right)$

$$\bar{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

b This is exact because X has a normal distribution.

c $$\text{P}(|\bar{X} - \mu| < 15) = \text{P}\left(|Z| < \frac{15}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)$$



$$\text{Require P}\left(|Z| < \frac{15\sqrt{n}}{\sigma}\right) > 0.95$$

$$\text{i.e.} \frac{15\sqrt{n}}{\sigma} > 1.96$$

$$\sigma = 40 \Rightarrow \sqrt{n} > \frac{40 \times 1.96}{15} = 5.2266\ldots$$

$$\therefore n > 27.318\ldots$$

So need $n = 28$ or more

9 a $$\bar{x} = \frac{\sum x}{n} = \frac{15.0}{20} = 0.75$$

$$s^2 = \frac{103.21 - 20 \times 0.75^2}{19} = 4.84$$

b $\sigma = 2.5$

i Assume that X has a normal distribution.

ii Assume that the sample was random.

c $$95\%\ \text{C.I. is } \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$= 0.75 \pm 1.96 \times \frac{2.5}{\sqrt{20}}$$

$$= (-0.34567\ldots, 1.8456\ldots)$$

$$= (-0.346, 1.85) \quad (3\text{s.f.})$$

d Since 0 is in the interval it is reasonable to assume that trains do arrive on time.

10 a $\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$

b 95% C.I. is an interval within in which we are 95% confident μ lies.

c L = sales of leaded petrol U = sales of unleaded petrol

$L \sim N(8.72, 3.25^2)$ $U \sim N(9.71, 3.25^2)$

90% of L between $8.72 \pm 1.6449 \times 3.25$

$= (3.3740\ldots, 14.0659\ldots)$

$= (3.37, 14.1)$ (3 s.f.)

d $n = 100$ $\bar{u} = 9.71$

95% C.I. for μ_u is : $9.71 \pm 1.96 \times \frac{3.25}{\sqrt{100}}$

$= (9.073, 10.347)$

$= (9.07, 10.35)$ (nearest penny)

e $\mathrm{H}_0 : \mu_u = 9.10$ (i.e. same as 2009) $\mathrm{H}_1 : \mu_u > 9.10$ (2010 sales > 2009 sales)

5% c.v. is $z = 1.6449$

t.s. is $z = \frac{(9.71 - 9.10)}{\left(\frac{3.25}{\sqrt{100}}\right)} = 1.8769\ldots > 1.6449$ Significant so reject H_0.

There is evidence that the mean sales of unleaded petrol in 2010 were greater than in 2009.

f We require that $P(|\bar{U} - \mu_u| < 0.50) > 0.95$

A 95% confidence interval for μ_u is given by $\bar{u} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$

We therefore require that $1.96 \times \frac{3.25}{\sqrt{n}} < 0.5$, or $\frac{0.5\sqrt{n}}{3.25} > 1.96$

i.e. $\sqrt{n} > 12.74$ or $n > 162.30\ldots$ $\therefore$ $n = 163$

11 a A 98% C.I. is an interval within which we are 98% sure the population mean will lie.

b L = length of willow tree leaves $L \sim \mathrm{N}(\mu, 1.33)$

$n = 40, \bar{L} = 10.20$

Standard error of the mean $= \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{1.33}}{\sqrt{40}} = 0.18234\ldots$

$= 0.182$ (3 d.p.)

c 95% C.I. is $10.20 \pm 1.96 \times 0.182\cdots$

$= (9.8426\ldots, 10.5573\ldots)$

$= (9.84, 10.56)$ (2 d.p.)

11 d Width of 98% C.I. is $2\times 2.3263\times\dfrac{\sigma}{\sqrt{n}}$

$\therefore$ Require $\dfrac{2\times 2.3263\times\sqrt{1.33}}{\sqrt{n}} < 1.50$

or $\sqrt{n} > 3.57\ldots$

i.e. $n > 12.79\ldots$

$\therefore$ need $n = 13$

12 a $\mathrm{E}(\bar{X}) = \mu \qquad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$

b i By Central Limit Theorem $\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

ii $\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

c $\sigma = 4.25 \qquad n = 100 \qquad \bar{x} = 18.50$

95% C.I. is $18.50 \pm 1.96\times\dfrac{4.25}{\sqrt{100}}$

$= (17.667, 19.333)$

$= (17.7, 19.3)$ (3 s.f.)

d $P(|\bar{X} - \mu| > 0.50) > 0.95$

i.e. $\dfrac{0.50\times\sqrt{n}}{4.25} > 1.96$

i.e. $\sqrt{n} > 16.66\ldots$ i.e. $\sqrt{n} > 277.55\ldots$

i.e. $n = 278$ or more

12 e Let μ_{2013} denote the mean of all the till receipts of the supermarket during 2013 (in £), we know this to be £19.40, and μ_{2014} denote the mean of all the till receipts of the supermarket during 2014 (also in £). The null hypothesis is that the means are the same. The alternative hypothesis is that the means are different.

$H_0 : \mu_{2014} - \mu_{2013} = 0$

$H_1 : \mu_{2014} - \mu_{2013} \neq 0.$

We have standard deviation and sample size of

$\sigma_{2014} = 4.25,$

$n_{2014} = 100.$

The value of the test statistic is

$$z = \frac{\overline{x}_{2014} - \overline{x}_{2013} - (\mu_{2014} - \mu_{2013})}{\sqrt{\frac{\sigma_{2014}{}^2}{n_{2014}}}}$$

$$= \frac{18.5 - 19.4 - (0)}{\sqrt{\frac{4.25^2}{100}}}$$

$$= -2.118(3dp)$$

The 5% (two-tailed) critical value for z is $z = \pm 1.96$

t.s. $= -2.118 < -1.96$ so our test statistic value is significant.

So we reject H_0 and conclude that the mean of all the till receipts of the supermarket during 2013 and the mean of all the till receipts of the supermarket during 2014 are different.

13 $D = \text{diameter} \sim \text{N}(0.824, 0.046^2)$

$n = 100$

a $\text{P}(\overline{D} < 0.823) = \text{P}\left(Z < \frac{(0.823 - 0.824)}{\frac{0.046}{\sqrt{100}}}\right)$

$= \text{P}(Z < -0.2173\ldots)$ Use -0.22

$= 1 - 0.5871$

$= 0.4129$ (tables)

or $= 0.41395\ldots$ (calculator)

i.e. accept awrt $0.413 \sim 0.414$

$0.41395 \times 200 = 82.79$

So out of 200 samples, approximately 83 will have mean < 0.823

b $n = 100, \overline{d} = 0.834$

98% C.I. is $0.834 \pm 2.3263 \times \frac{0.046}{\sqrt{100}}$

$= (0.82329\ldots, 0.84470\ldots)$

$= (0.823, 0.845)$ (3 s.f.)

13 c Since 0.824 is *in* the C.I. we can conclude that there is insufficient evidence of a malfunction.

14 a Let μ_d denote the mean heart rate of the people who drive to work and μ_w denote the mean heart rate of the people who walked to work.
The null hypothesis is that the means are the same.
The alternative hypothesis is that the mean heart rate of people who drive to work is greater than the mean heart rate of people who walk to work.

$H_0 : \mu_d = \mu_w$

$H_1 : \mu_d > \mu_w$.

We have standard deviations and sample sizes of

$\sigma_d = \sqrt{60.2} = 7.759,$

$n_d = 30,$

$\sigma_w = \sqrt{55.8} = 7.470,$

$n_w = 36.$

The value of the test statistic is

$$z = \frac{\bar{x}_d - \bar{x}_w - (\mu_d - \mu_w)}{\sqrt{\frac{\sigma_d^2}{n_d} + \frac{\sigma_w^2}{n_w}}}$$

$$= \frac{52 - 47 - (0)}{\sqrt{\frac{60.2}{30} + \frac{55.8}{36}}}$$

$$= 2.651 \text{ (3 d.p.)}$$

(Note that we used the null hypothesis of $H_0 : \mu_d - \mu_w = 0$ in this calculation).
The 5% (one-tailed) critical value for z is $z = 1.6449$ so our test statistic value is significant.

So we reject H_0 and conclude that the mean heart rate of people who drive to work is greater than the mean heart rate of people who walk to work.

b We have assumed normal distribution and that individual results are independent.
We have also assumed $\sigma^2 = s^2$ for both populations.

14 c An unbiased estimator for σ^2 is given by the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

Let $\bar{y}_d$ denote the mean heart rate of people who drive to work including the extra observation mentioned in the question. We calculate:

$$\bar{y}_d = \frac{30\times\bar{x}_d + 55}{30+1}$$
$$= \frac{1560+55}{31}$$
$$= 52.097 \text{ (5 s.f.)}.$$

For notation purposes, we will let $Y_{di} = \begin{cases} X_{di} \text{ for } 0 \leqslant i \leqslant 30 \\ 55 \text{ for } i = 31 \end{cases}$

where X_{di} is the i'th observation of the initial sample of heart rates of people who drive to work. Now we note that from part a, we have an expression for:

$$s_x^{\ 2} = 60.2$$
$$= \frac{1}{30-1}\sum_{i=1}^{30}\left(X_{di} - \bar{X}_d\right)^2$$
$$= \frac{1}{29}\sum_{i=1}^{30}\left(X_{di} - 52\right)^2$$
$$= \frac{1}{29}\left(\sum_{i=1}^{30}\left(X_{di}^{\ 2}\right) - \sum_{i=1}^{30}\left(2\times 52X_{di}\right) + \sum_{i=1}^{30}\left(52^2\right)\right)$$
$$= \frac{1}{29}\left(\sum_{i=1}^{30}\left(X_{di}^{\ 2}\right) - 60\times 52X_{di} + 30\times 52X_{di}\right)$$
$$= \frac{1}{29}\left(\sum_{i=1}^{30}X_{di}^{\ 2} - 30\times 52^2\right)$$
$$60.2 = \frac{\sum_{i=1}^{30}X_{di}^{\ 2} - 81120}{29}$$

From here we can rearrange in order to obtain the expression $\sum_{i=1}^{30}X_{di}^{\ 2} = 82865.8$.

$$S^2 = \frac{1}{31-1}\sum_{i=1}^{31}\left(Y_{di} - \bar{Y}_d\right)^2$$
$$= \frac{1}{30}\left(\sum_{i=1}^{31}\left(Y_{di}^{\ 2}\right) - 31\bar{Y}_d^{\ 2}\right)$$
$$= \frac{1}{30}\left(\sum_{i=1}^{30}\left(Y_{di}^{\ 2}\right) + Y_{d31}^{\ 2} - 31\times\left(52.097\right)^2\right)$$
$$= \frac{1}{30}\left(\sum_{i=1}^{30}\left(X_{di}^{\ 2}\right) + 55^2 - 84137\right)$$
$$= \frac{1}{30}\left(82865.8 + 3025 - 84137\right)$$
$$= 58.5 \text{ (3 s.f.)}.$$

Challenge

$$\mathrm{E}(\bar{X}) = \mu \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\mathrm{E}(\bar{Y}) = \mu \qquad \mathrm{Var}(\bar{Y}) = \frac{\sigma^2}{m}$$

a $T = r\bar{X} + s\bar{Y}$

$E(T) = r\mu + s\mu = (r+s)\mu$

So if T is unbiased, $r + s = 1$.

b $r + s = 1 \Rightarrow s = 1 - r$

$\therefore T = r\bar{X} + (1-r)\bar{Y}$

$$\mathrm{Var}(T) = r^2\mathrm{Var}(\bar{X}) + (1-r)^2\mathrm{Var}(\bar{Y}) = r^2\frac{\sigma^2}{n} + (1-r)^2\frac{\sigma^2}{m}$$

$$= \sigma^2\left[\frac{r^2}{n} + \frac{(1-r)^2}{m}\right]$$

c $$\frac{\mathrm{d}}{\mathrm{d}r}\mathrm{Var}(T) = \sigma^2\left[\frac{2r}{n} + \frac{2(1-r)\times(-1)}{m}\right]$$

$\therefore$ Var (T) is a quadratic function of r with positive r^2 terrm $\therefore$ min

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathrm{Var}(T) = 0 \Rightarrow rm = (1-r)n \quad \text{or} \quad r(m+n) = n \quad \text{or} \quad r = \frac{n}{m+n}$$

d Best estimator of T is

$$T = \frac{n}{m+n}\bar{X} + \frac{m}{m+n}\bar{Y} \quad \text{or} \quad \frac{n\bar{X} + m\bar{Y}}{m+n}$$