Review Exercise 1

1 a A census observes every member of a population.
A disadvantage of a census is it would be time-consuming to get opinions from all the employees.
OR It would be difficult/time-consuming to process the large amount of data from a census.

b Opportunity sampling.

c It is not a random sample. The sample only includes cleaners, there are no types of other employees such as managers.
The first 50 cleaners to leave may be in the same group/shift so may share the same views.

d i Allocate a number from 1–550 to all employees.
For a sample of 50, you need every eleventh person since 550 ÷ 50 = 11.
Select the first employee using a random number from 1 - 11, then select every eleventh person from the list; e.g. if person 8 is then the sample is 8, 19, 30, 41...

ii For this sample, you need $\frac{55}{550}\times 50 = 5$ managers and $\frac{495}{550}\times 50 = 45$ cleaners.
Label the managers 1–55 and the cleaners 1–495.
Use random numbers to select 5 managers and 45 cleaners.

2 a Advantage – a sampling frame is not required.
Disadvantage – sampling errors cannot be calculated.

b Advantage – very quick to administer.
Disadvantage – may not be representative of the population.

3 a 390 and 372

b The list is alphabetical and has not been sorted by gender.

c Stratified sampling

4 Label standard rooms between 1 and 180.
Use random numbers in the range 1–180 to select 18 rooms.
Label premier rooms between 1 and 100.
Use random numbers in the range 1–100 to select 10 rooms.
Label executive rooms between 1 and 40.
Use random numbers in the range 1–40 to select 4 rooms.

5 a $\text{P}(F < 3R) = \text{P}(F - 3R < 0)$
Let $X = F - 3R$
$\text{E}(X) = \text{E}(F) - 3\text{E}(R) = 238 - 3 \times 82 = -8$
$\text{Var}(X) = \text{Var}(F) + 3^2\text{Var}(R) = 7^2 + 3^2 \times 3^2 = 130$
Hence $X \sim \text{N}(-8, 130)$
$\text{P}(X < 0) = 0.7586$ (4 d.p.) (from calculator or tables)

b The assumption made is that the duration of the two rides are independent.
Validity: this is likely to be the case – two separate control panels operate each ride.

c $D = R_1 + R_2 + R_3$
$\text{E}(D) = \text{E}(R) + \text{E}(R) + \text{E}(R) = 3 \times 82 = 246$
$\text{Var}(D) = \text{Var}(R) + \text{Var}(R) + \text{Var}(R) = 3 \times 3^2 = 27$
Hence $D \sim \text{N}(246, 27)$

d $\text{P}(|F - D| < 10) = \text{P}(-10 < F - D < 10)$
Let $Y = F - D$
$\text{E}(Y) = 238 - 246 = -8$
$\text{Var}(Y) = 49 + 27 = 76$
Hence $Y \sim \text{N}(-8, 76)$
$\text{P}(-10 < Y < 10) = \text{P}(Y < 10) - \text{P}(Y < -10) = 0.5713$ (4 d.p.)

6 a $\text{E}(R) = \text{E}(X) + \text{E}(Y) = 20 + 10 = 30$

b $\text{Var}(R) = \text{Var}(X) + \text{Var}(Y) = 4 + 0.84 = 4.84$

c $R \sim \text{N}(30, 4.84)$ from parts **a** and **b**
$\text{P}(28.9 < R < 32.64) = \text{P}(R < 32.64) - \text{P}(R < 28.9) = 0.8849 - 0.3085 = 0.5764$ (4 d.p.)

7 $\bar{X} \sim \mathrm{N}\left(\mu, \sigma^2\right)$, $\bar{X} = \dfrac{X_1 + X_2 + X_3}{3}$ and $\mathrm{P}\left(\dfrac{X_1 + X_2}{2} > \bar{X} + k\sigma\right) = 0.2$

$$\mathrm{E}\left(\frac{X_1 + X_2}{2} - \bar{X}\right) = 0$$

$$\frac{X_1 + X_2}{2} - \bar{X} = \frac{X_1 + X_2}{2} - \frac{X_1 + X_2 + X_3}{3}$$

$$= \frac{1}{6}\left(X_1 + X_2 - 2X_3\right)$$

$$\mathrm{Var}\left(\frac{X_1 + X_2}{2} - \bar{X}\right) = \left(\frac{1}{6}\right)^2\left(\sigma^2 + \sigma^2 + (-2\sigma)^2\right)$$

$$= \frac{\sigma^2}{6}$$

$$\mathrm{P}\left(\frac{X_1 + X_2}{2} > \bar{X} + k\sigma\right) = 0.2$$

$$\mathrm{P}\left(Z > \frac{\left(\bar{X} + k\sigma\right) - \bar{X}}{\sqrt{\frac{\sigma^2}{6}}}\right) = 0.2$$

$$P\left(Z < \sqrt{6}k\right) = 0.8$$

$$\sqrt{6}k = 0.8416$$

$$k = 0.3436$$

8 Possible totals are shown in the table.

		Coin 1		
		0.1	**0.5**	**1**
Coin 2	**0.1**	0.2	0.6	1.1
	0.5	0.6	1	1.5
	1	1.1	1.5	2

Mean value	**Probability**
0.1	$\frac{2}{10} \times \frac{1}{9} = \frac{2}{90}$
0.3	$2\left(\frac{3}{10} \times \frac{2}{9}\right) = \frac{12}{90}$
0.55	$2\left(\frac{5}{10} \times \frac{2}{9}\right) = \frac{20}{90}$
0.5	$\frac{3}{10} \times \frac{2}{9} = \frac{6}{90}$
0.75	$2\left(\frac{5}{10} \times \frac{3}{9}\right) = \frac{30}{90}$
1	$\frac{5}{10} \times \frac{4}{9} = \frac{20}{90}$

9 $C = 2A + 5B$ where $A \sim \mathrm{N}(15, 1.5^2)$ and $B \sim \mathrm{N}(\mu, 2^2)$

$\mathrm{E}(C) = 2 \times 15 + 5 \times \mu$
$= 30 + 5\mu$

$\mathrm{Var}(C) = 2^2 \times 1.5^2 + 5^2 \times 2^2$
$= 109$

Therefore:

$C \sim \mathrm{N}(30 + 5\mu, 109)$

$\mathrm{P}(C < 83.5) = 0.9$

$$\mathrm{P}\left(Z < \frac{83.5 - (30 + 5\mu)}{\sqrt{109}}\right) = 0.9$$

$$\frac{83.5 - (30 + 5\mu)}{\sqrt{109}} = 1.282$$

$$\frac{53.5 - 5\mu}{\sqrt{109}} = 1.282$$

$$\mu = \frac{53.5 - 1.282\sqrt{109}}{5}$$

$\mu = 8.02$ (3 s.f.)

10 $A \sim \mathrm{N}(24, 4^2)$ and $X \sim \mathrm{N}(20, 3^2)$

$$B = 3X + \sum_{i=1}^{4} A_i$$
$$= 3X + A_1 + A_2 + A_3 + A_4$$

$\mathrm{E}(B) = 3\mathrm{E}(X) + \mathrm{E}(A_1) + \mathrm{E}(A_2) + \mathrm{E}(A_3) + \mathrm{E}(A_4)$
$= 3 \times 20 + 4 \times 24$
$= 156$

$\mathrm{Var}(B) = 3^2\,\mathrm{Var}(X) + \mathrm{Var}(A_1) + \mathrm{Var}(A_2) + \mathrm{Var}(A_3) + \mathrm{Var}(A_4)$
$= 3^2 \times 3^2 + 4 \times 4^2$
$= 145$

$\mathrm{P}(B > 156) = 0.5$

$$\mathrm{P}(B \le 170) = 1 - \mathrm{P}\left(Z < \frac{170 - 156}{\sqrt{145}}\right)$$
$= \mathrm{P}(Z < 1.1626)$
$= 0.8775$

$\mathrm{P}(156 < B \le 170) = 0.8775 - 0.5$
$= 0.3775$

$$\mathrm{P}(B \le 170 \mid B > 156) = \frac{0.3775}{0.5}$$
$= 0.755$ (3 s.f.)

11 a $\mathrm{E}(\bar{X}) = \dfrac{(5\alpha - 9) + (\alpha - 3)}{2}$

$= 3\alpha - 6$

$3\alpha - 6 \neq \alpha$

Therefore $\mathrm{E}(\bar{X})$ is a biased estimator.

The bias is $(3\alpha - 6) - \alpha = 2\alpha - 6$

b $\mathrm{E}(Y) = \mathrm{E}(k\bar{X} + 2)$

$= k(3\alpha - 6) + 2$

and if Y is an unbiased estimator for α then $\mathrm{E}(Y) - \alpha = 0$.
So it follows that:

$k(3\alpha - 6) + 2 - \alpha = 0$

$3k(\alpha - 2) = \alpha - 2$

$3k = 1$

$k = \dfrac{1}{3}$

c From the data:

$\bar{X} = 36.36$

Using

$Y = k\bar{X} + 2$

$= \dfrac{1}{3}\bar{X} + 2$

$= \dfrac{1}{3} \times 36.36 + 2$

$= 14.12$

From $X \sim \mathrm{U}(\alpha - 3, 5\alpha - 9)$, the max value that X can take is estimated as:

$5\alpha - 9 = 5 \times 14.12 - 9$

$= 61.6$

12 a Let X be the total weight of 4 randomly chosen adult men.

$X = M_1 + M_2 + M_3 + M_4$

$\mathrm{E}(X) = 4 \times 84 = 336$

$\mathrm{Var}(X) = 4 \times 11^2 = 484$

$X \sim \mathrm{N}(336, 484)$

$\mathrm{P}(X < 350) = 0.7377$ (4 d.p.)

b Let $M \sim \mathrm{N}(84,121)$ and $W \sim \mathrm{N}(62,100)$ and $Y = M - 1.5W$

$\mathrm{E}(Y) = 84 - 1.5 \times 62 = -9$

$\mathrm{Var}(Y) = \mathrm{Var}(M) + 1.5^2\,\mathrm{Var}(W) = 11^2 + 1.5^2 \times 10^2 = 346$

So $Y \sim \mathrm{N}(-9, 346)$, and $P(Y < 0) = 0.6858$ (4 d.p.)

13 a $\mathrm{E}(D) = \mathrm{E}(A) - 3\,\mathrm{E}(B) + 4\,\mathrm{E}(C) = 5 - 3 \times 7 + 4 \times 9 = 20$

$\mathrm{Var}(D) = \mathrm{Var}(A) + 3^2\ \mathrm{Var}(B) + 4^2\ \mathrm{Var}(C) = 2^2 + 9 \times 3^2 + 16 \times 4^2 = 341$

So $D \sim \mathrm{N}(20, 341)$, and $\mathrm{P}(D < 44) = 0.9031$ (4 d.p.)

b $\mathrm{E}(X) = \mathrm{E}(A) - 3\,\mathrm{E}(B) + 4\,\mathrm{E}(C) = 5 - 3 \times 7 + 4 \times 9 = 20$

$\mathrm{Var}(D) = \mathrm{Var}(A) + \mathrm{Var}(B) + \mathrm{Var}(B) + \mathrm{Var}(B) + 4^2\ \mathrm{Var}(C) = 2^2 + 3 \times 3^2 + 16 \times 4^2 = 287$

So $X \sim \mathrm{N}(20, 287)$, and $\mathrm{P}(X > 0) = 1 - \mathrm{P}(X \leqslant 0) = 1 - 0.1189 = 0.8811$ (4 d.p.)

14 a Let $W = C_1 - C_2$

$\mathrm{E}(W) = 350 - 350 = 0$

$\mathrm{Var}(W) = 8 + 8 = 16$

So $W \sim \mathrm{N}(0, 16)$

$\mathrm{P}(|W| > 6) = 1 - \mathrm{P}(W < 6) + \mathrm{P}(W < -6) = 0.0668 + 0.0668 = 0.1336$ (4 d.p.)

b Let $X = C - L$

$\mathrm{E}(X) = 350 - 345 = 5$

$\mathrm{Var}(X) = 8 + 17 = 25$

So $X \sim \mathrm{N}(5, 25)$

$\mathrm{P}(X > 0) = 1 - \mathrm{P}(X < 0) = 1 - 0.1587 = 0.8413$ (4 d.p.)

c Let $Y = \sum_{i=1}^{24} C_i + B$

$\mathrm{E}(Y) = 24 \times 350 + 100 = 8500$

$\mathrm{Var}(Y) = 24 \times 8 + 2^2 = 196$

So $Y \sim \mathrm{N}(8500, 196)$

$\mathrm{P}(8510 < Y < 8520) = \mathrm{P}(Y < 8520) - \mathrm{P}(Y < 8510) = 0.92343 - 0.76247 = 0.1610$ (4 d.p.)

d All random variables (each can of cola and the box) are independent and normally distributed.

15 a

$\bar{x} = \frac{361.6}{80} = 4.52$

$\hat{\sigma}^2 = s^2 = \frac{1753.95 - 80 \times \bar{x}^2}{79} = 1.5128$

or $\hat{\sigma}^2 = s^2 = \frac{80}{79} \times \left(\frac{1753.95}{80} - \bar{x}^2\right) = 1.5128$

Using $\frac{\sum x^2 - n\bar{x}^2}{n-1}$ or $\frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right)$

b $H_0 : \mu_A = \mu_B \quad H_1 : \mu_A > \mu_B$

This is a difference of means test. When stating hypotheses you must make it clear which mean is greater when it is a one-tailed test.

$z = \frac{4.52 - 4.06}{\sqrt{\frac{1.5128}{80} + \frac{2.50}{60}}} = \left(\frac{0.46}{\sqrt{0.060576}}\right)$

$= 1.8689$ or -1.8689 if $B - A$ was used.

Using $z = \frac{(\bar{A} - \bar{B}) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$

One tail c.v. is $z = 1.6449$
$1.87 > 1.6449$ so reject H_0.

Use the percentage point table and quote the figure in full.

There is evidence that diet A is better than diet B *or* evidence that (mean) weight lost in the first week using diet A is greater than using diet B.

State your conclusion in the context of the question.

c CLT enables you to assume that $\bar{A}$ and $\bar{B}$ are normally distributed since both samples are large.

d Assumed $\sigma_A^2 = s_A^2$ and $\sigma_B^2 = s_B^2$

Variance must be known to use the test. Remember, σ^2 is the population variance and s^2 is an unbiased estimator of the population variance.

16

$123.5 = \bar{x} - 2.5758 \times \frac{\sigma}{\sqrt{n}}$ **(1)**

$154.7 = \bar{x} - 2.5758 \times \frac{\sigma}{\sqrt{n}}$ **(2)**

99% confidence interval, so each tail is 0.05. Use the percentage point table and quote the figure in full. C.I. $\bar{x} \pm 2.5758 \times \frac{\sigma}{\sqrt{n}}$

$\bar{x} = \frac{1}{2}(123.5 + 154.7) = 139.1$

Add equations **(1)** and **(2)** to find $\bar{x}$ or calculate the mean of the given limits.

$2.5758 \times \frac{\sigma}{\sqrt{n}} = 154.7 - 139.1$

$= 15.6$

$\frac{\sigma}{\sqrt{n}} = \frac{15.6}{2.5758}$

Substitute $\bar{x}$ into equation **(1)** or **(2)** to find $\frac{\sigma}{\sqrt{n}}$

So 95% C.I. $= 139.1 \pm 1.9600 \times \frac{15.6}{2.5758}$

$= (127.22\ldots, 150.97\ldots)$

$= (127, 151)$

95% confidence interval, so each tail is 0.025.

Substitute in $\bar{x}$ and $\frac{\sigma}{\sqrt{n}}$

Answers should be given to at least 3 significant figures.

17 a

$\bar{X} = \frac{500}{10} = 50$

$s^2 = \frac{25\,001.74 - 10 \times 50^2}{9}$

$= 0.193$

Using $\frac{\sum x^2 - n\bar{x}^2}{n-1}$ or $\frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right)$

b i For 95% confidence interval, z value is 1.96.
Confidence interval is therefore:

$\left(50 - 1.96 \times \frac{0.5}{\sqrt{10}}, 50 + 1.96 \times \frac{0.5}{\sqrt{10}}\right)$

$= (49.690..., 50.309...)$

$= (49.7, 50.3)$

ii For 99% confidence interval, z value is 2.5758.
Confidence interval is therefore:

$\left(50 - 2.5758 \times \frac{0.5}{\sqrt{10}}, 50 + 2.5758 \times \frac{0.5}{\sqrt{10}}\right)$

$= (49.592..., 50.407...)$

$= (49.6, 50.4)$

18 a

$\hat{\mu} = \bar{x} = \frac{82 + 98 + 140 + 110 + 90 + 125 + 150 + 130 + 70 + 110}{10}$

$= 110.5$

$s^2 = \frac{128153 - 10 \times 110.5^2}{9}$

$= 672.28$

Using $\frac{\sum x^2 - n\bar{x}^2}{n-1}$ or $\frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right)$

b 95% confidence limits are:

$110.5 \pm 1.96 \times \frac{25}{\sqrt{10}}$

$= (95.005, 125.995)$

$= (95.0, 126)$

95% confidence interval, so each tail is 0.025. Use the percentage point table and quote the figure in full.

C.I.: $\bar{x} \pm 1.9600 \times \frac{\sigma}{\sqrt{n}}$

Answers should be given to at least 3 significant figures.

c $0.95^{15} = 0.46329\ldots = 0.4633$ (4 d.p.)

19 a

$\bar{x} = \left(\frac{6046}{36} = \right)167.94\ldots$

Using $\frac{\sum x^2 - n\bar{x}^2}{n-1}$ or $\frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right)$

$s^2 = \frac{1\,016\,338 - 36 \times \bar{x}^2}{35}$

$= 27.0$

b 99% confidence interval is: $\bar{x} \pm 2.5758 \times \frac{5.1}{\sqrt{36}}$

99% confidence interval, so each tail is 0.005. Use the percentage point table and quote the figure in full.

C.I.: $\bar{x} \pm 2.5758 \times \frac{\sigma}{\sqrt{n}}$

$= 167.94 \pm 2.5758 \times \frac{5.1}{\sqrt{36}}$

$= (167.75, 170.13)$

$= (166, 170)$

Answers should be given to at least 3 significant figures.

20 a Let X represent repair time

$\therefore \sum x = 1435 \therefore \bar{x} = \frac{1435}{5} = 287$

Using $\frac{\sum x^2 - n\bar{x}^2}{n-1}$ or $\frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right)$

$\sum x^2 = 442\,575$

$\therefore s^2 = \frac{442\,575 - 5 \times 287^2}{4}$

$= 7682.5$

b $\text{P}(|\mu - \hat{\mu}|) < 20 = 0.95$

The repair time is between 80 and 120. 95% confidence interval, so each tail is 0.025. Use the percentage point table and quote the figure in full.

C.I.: $\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$

$\therefore 1.96 \times \frac{\sigma}{\sqrt{n}} = 20$

$\therefore n = \frac{1.96^2\sigma^2}{20^2} = \frac{1.96^2\sigma^2}{400} = \frac{1.96^2 \times 100^2}{400} = 96.04$

$\therefore$ Sample size $(\geqslant)$ 97 required

21 a Let W_C be the weight of a slice of cheesecake.

$W_C \sim \mathrm{N}(135, 3^2)$

Let W_B be the weight of the box.

$W_B \sim \mathrm{N}(100, 6^2)$

Let W_T be the total weight of the box and 12 slices of cheesecake.

$$\begin{aligned}\mathrm{E}(W_T) &= \mathrm{E}(W_B) + 12\mathrm{E}(W_C)\\ &= 100 + 12 \times 135\\ &= 1720\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(W_T) &= \mathrm{Var}(W_B) + 12\mathrm{Var}(W_C)\\ &= 6^2 + 12 \times 3^2\\ &= 144\end{aligned}$$

$$\begin{aligned}\mathrm{P}(W_T > 1700) &= 1 - \mathrm{P}(W_T < 1700)\\ &= 1 - \mathrm{P}\left(Z_T < \frac{1700 - 1720}{\sqrt{144}}\right)\\ &= 1 - \mathrm{P}\left(Z_T < -\frac{5}{3}\right)\\ &= 1 - 0.04779\\ &= 1 - 0.04779\\ &= 0.9522\end{aligned}$$

b The weights of each slice are independent.

22 A 95% confidence interval is

$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + 1.96 \times \frac{1.3}{\sqrt{n}} = 15.7096$$

$$\bar{x} - 1.96 \times \frac{1.3}{\sqrt{n}} = 14.6904$$

$$2 \times 1.96 \times \frac{1.3}{\sqrt{n}} = 15.7096 - 14.6904$$

$$\frac{1}{\sqrt{n}} = \frac{15.7096 - 14.6904}{2 \times 1.96 \times 1.3}$$

$$\sqrt{n} = \frac{2 \times 1.96 \times 1.3}{15.7096 - 14.6904}$$

$$\sqrt{n} = 5$$

$$n = 25$$

23 a A c% confidence interval is

$$\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$$
$$\bar{x} + z \times \frac{3.6}{\sqrt{36}} = 16.2364 \Rightarrow \bar{x} + 0.6z = 16.2364$$
$$\bar{x} - z \times \frac{3.6}{\sqrt{36}} = 12.9636 \Rightarrow \bar{x} - 0.6z = 12.9636$$
$$2\bar{x} = 29.2$$
$$\bar{x} = 14.6$$

b Since

$$\bar{x} + 0.6z = 16.2364$$
$$14.6 + 0.6z = 16.2364$$
$$z = 2.7273$$
$$P(Z < 2.7273) = 0.997$$

By symmetry,

$$P(-2.7273 < Z < 2.7273) = 0.994$$

Therefore a 99.4% confidence interval.

Challenge

a $T = \frac{2X_1}{3} - \frac{X_2}{2} + \frac{5X_3}{6}$

$$E(T) = E\left(\frac{2X_1}{3} - \frac{X_2}{2} + \frac{5X_3}{6}\right) = E\left(\frac{2X_1}{3}\right) - E\left(\frac{X_2}{2}\right) + E\left(\frac{5X_3}{6}\right)$$
$$= \frac{2}{3}\mu - \frac{1}{2}\mu + \frac{5}{6}\mu = \mu$$

Therefore T is an unbiased estimator of μ

b $\hat{\mu} = aX_1 + bX_2$

$$\text{Var}(\hat{\mu}) = \text{Var}(aX_1 + bX_2) = a^2\text{Var}(X_1) + b^2\text{Var}(X_2)$$
$$a + b = 1, \text{ so } b = a - 1$$
$$\therefore \text{Var}(\hat{\mu}) = a^2\text{Var}(X_1) + (a-1)^2\text{Var}(X_2)$$
$$= (2a^2 - 2a + 1)\sigma^2$$

c Find the minimum of $\text{Var}(\hat{\mu})$ by differentiating $(2a^2-2a+1)$ and setting equal to zero

$$\frac{d}{da}(2a^2 - 2a + 1) = 4a - 2 = 0$$

$\therefore$ minimum when $a = b = 0.5$