Chapter Review 4

1 By the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(5, \frac{1}{100}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}(5, 0.01)$

$\mathrm{P}(\bar{X} > 5.2) = 1 - \mathrm{P}(\bar{X} < 5.2) \approx 1 - 0.9772 = 0.0228$ (4 d.p.)

2 $\mathrm{E}(X) = \frac{1}{6}(1+2+4+5+7+8) = \frac{27}{6} = 4.5$

$\mathrm{Var}(X) = \frac{1}{6}(1+2^2+4^2+5^2+7^2+8^2) - 4.5^2$

$= \frac{159}{6} - \frac{729}{36} = \frac{225}{36} = \frac{25}{4} = 6.25$

By the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(4.5, \frac{6.25}{20}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}(4.5, 0.3125)$

$\mathrm{P}(\bar{X} < 4) \approx 0.1855$ (4 d.p.)

3 $X \sim \mathrm{N}(1,1)$ and by the central limit theorem $\bar{X} \sim \mathrm{N}\left(1, \frac{1}{\sqrt{n}}\right)$

Standardise the sample mean.

$\mathrm{P}(\bar{X} < 0) = \mathrm{P}\left(Z < -\sqrt{n}\right)$ and so require $\mathrm{P}\left(Z > -\sqrt{n}\right) < 0.05$

Using the table for the percentage points of the normal distribution:

$\mathrm{P}(Z = -1.645) = 0.05$

$\Rightarrow -\sqrt{n} < -1.645$

$\Rightarrow n > 2.706$

So minimum sample size is $n = 3$ for the probability of a negative sample mean being less than 5%

4 Let the random variable X denote the number of sixes thrown by a student in 10 rolls of the dice, so $X \sim \mathrm{B}\left(10, \frac{1}{6}\right)$

$\mathrm{E}(X) = np = 10 \times \frac{1}{6} = \frac{5}{3}$

$\mathrm{Var}(X) = np(1-p) = \frac{5}{3} \times \frac{5}{6} = \frac{25}{18}$

By the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(\frac{5}{3}, \frac{25}{18 \times 20}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}\left(\frac{5}{3}, \frac{5}{72}\right)$

$\mathrm{P}(\bar{X} > 2) = 1 - \mathrm{P}(\bar{X} < 2) \approx 1 - 0.8970 = 0.1030$ (4 d.p.)

5 a Let X be the number of buses that arrive in a 10-minute period, then $X \sim \mathrm{Po}(2)$

$\mathrm{P}(X = 3) = \frac{\mathrm{e}^{-2}\, 2^3}{3!} = 0.1804$ (4 d.p.)

5 b Let T be the number of buses that arrive in a two-hour period, so $T = 12\bar{X}$

By the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(2, \frac{2}{12}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}\left(2, \frac{1}{6}\right)$

$\mathrm{P}(T \geqslant 25) = \mathrm{P}\left(\bar{X} \geqslant \frac{25}{12}\right)$

$\mathrm{P}\left(\bar{X} \geqslant \frac{25}{12}\right) = 1 - \mathrm{P}\left(\bar{X} < \frac{25}{12}\right) \approx 1 - 0.5809 = 0.4191$ (4 d.p.)

6 a Let the random variable X be the mass of an egg, then $X \sim \mathrm{N}(60, 25)$ and $\bar{X} \sim \mathrm{N}\left(60, \frac{25}{48}\right)$

$\mathrm{P}(\bar{X} > 59) = 1 - \mathrm{P}(\bar{X} < 59) = 1 - 0.0829 = 0.9171$ (4 d.p.)

b The answer in part **a** is not an estimate because the sample is taken from a population that is normally distributed.

c Let the random variable Y is the number of double yolk eggs in a crate of 48 eggs, so $Y \sim \mathrm{B}(48, 0.1)$

$\mathrm{E}(Y) = np = 48 \times 0.1 = 4.8$

$\mathrm{Var}(Y) = np(1-p) = 4.8 \times 0.9 = 4.32$

By the central limit theorem $\bar{Y} \approx\sim \mathrm{N}\left(4.8, \frac{4.32}{30}\right)$, i.e. $\bar{Y} \approx\sim \mathrm{N}(4.8, 0.144)$

The probability that the sample of 30 crates will contain fewer than 150 double-yolk eggs is $\mathrm{P}(\bar{Y} < 5)$ as $30 \times 5 = 150$

$\mathrm{P}(\bar{Y} < 5) \approx 0.7009$ (4 d.p.)

7 Consider a sample of 100 cups of coffee, so $\bar{S} \sim \mathrm{N}(4.9, 0.0064)$. One pack of milk powder will be sufficient, if $100\bar{S} < 500$, i.e. $\bar{S} < 5$

$\mathrm{P}(\bar{S} < 5) = 0.8944$ (4 d.p.)

8 Let the random variable be X, so by the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(40, \frac{9}{n}\right)$

Required to find minimum n such that $\mathrm{P}(\bar{X} > 42) < 0.05$

Standardise the sample mean using $Z = \frac{\bar{X} - \mu}{\sigma}$, $\mu = 40$ and $\sigma = \frac{3}{\sqrt{n}}$

So for $\bar{X} = 42$, $Z = \frac{(42-40)\sqrt{n}}{3} = \frac{2\sqrt{n}}{3}$ and $\mathrm{P}(\bar{X} > 42) = \mathrm{P}\left(Z > \frac{2\sqrt{n}}{3}\right)$

Using the table for the percentage points of the normal distribution;
$\mathrm{P}(Z > 1.6449) = 0.05$

So $\frac{2\sqrt{n}}{3} > 1.6449$

$\Rightarrow \sqrt{n} > 2.46735$

$\Rightarrow n > 6.0878\ldots$

So n = 7 is the minimum sample size required for $\mathrm{P}(\bar{X} > 42) < 0.05$

9 Let the random variable be X, so by the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(35, \frac{9}{20}\right)$

$\mathrm{P}(\bar{X} > 37) = 1 - \mathrm{P}(\bar{X} < 37) \approx 1 - 0.9986 = 0.0014$ (4 d.p.)

10 a The table describes the distribution of X

x	0	1
P(X = x)	0.4	0.6

E(X) = 0.6, Var(X) = 0.6 – 0.6^2 = 0.24

b By the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(0.6, \frac{0.24}{500}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}(0.6, 0.00048)$

$$\mathrm{P}(\bar{X} > 0.63) + \mathrm{P}(\bar{X} < 0.57) = 1 - \mathrm{P}(\bar{X} < 0.63) + \mathrm{P}(\bar{X} < 0.57)$$
$$\approx 1 - 0.91454 + 0.08545 = 0.1709 \text{ (4 d.p.)}$$

10 c Required to find minimum n such that $\mathrm{P}(0.57 < \bar{X} < 0.63) > 0.95$

Standardise the sample mean using $Z = \dfrac{\bar{X} - \mu}{\sigma}$, $\mu = 0.6$ and $\sigma = \sqrt{\dfrac{0.24}{n}}$

So for $\bar{X} = 42$, $Z = \dfrac{(42-40)\sqrt{n}}{3} = \dfrac{2\sqrt{n}}{3}$ and $\mathrm{P}(\bar{X} > 42) = \mathrm{P}\left(Z > \dfrac{2\sqrt{n}}{3}\right)$

So require $\mathrm{P}\left(-\dfrac{0.03\sqrt{n}}{\sqrt{0.24}} < Z < \dfrac{0.03\sqrt{n}}{\sqrt{0.24}}\right) > 0.95$

$\Rightarrow 1 - 2\mathrm{P}\left(Z < -\dfrac{0.03\sqrt{n}}{\sqrt{0.24}}\right) > 0.95$ (by the symmetry of the normal distribution)

$\Rightarrow \mathrm{P}\left(Z < -\dfrac{0.03\sqrt{n}}{\sqrt{0.24}}\right) < 0.025$

Using the table for the percentage points of the normal distribution

$\mathrm{P}(Z < -1.960) = 0.025$

$\Rightarrow -\dfrac{0.03\sqrt{n}}{\sqrt{0.24}} < -1.960$

$\Rightarrow \sqrt{n} > \dfrac{1.960 \times \sqrt{0.24}}{0.03} \Rightarrow \sqrt{n} > 32.0066\ldots$

$\Rightarrow n > 1024.42...$

So $n = 1025$

11 The sample of bands from the new supplier has

$$\bar{x} = \frac{\sum X}{100} = \frac{4715}{100} = 47.15$$

and

$$s^2 = \frac{\sum X^2}{100} - \left(\frac{\sum X}{100}\right)^2 = \frac{222910}{100} - 47.15^2 = 5.978$$

$$s = 2.445$$

H_0: $\mu = 46.5$
H_1: $\mu > 46.5$

Using the estimator above for s, under the null hypothesis the distribution of the mean breaking stress of a sample of 100 bands can be modelled as $\mathrm{N}\left(46.5, \frac{(2.445)^2}{100}\right)$

The probability a sample of 100 bands having a mean breaking stress of at least 47.15 is

$$Z = \frac{47.15 - 46.5}{\frac{2.445}{10}} = 2.659$$

$2.659 > 1.645$

Therefore, there is evidence that the new manufacturer's bands are stronger.

12

	x	x^2	f	fx	fx^2
	12	144	5	60	720
	13	169	42	546	7098
	14	196	210	290	41160
	15	225	31	465	6975
	16	256	12	192	3072
Totals			**300**	**4205**	**59025**

$$\bar{x} = \frac{\sum fx}{\sum f}$$
$$= \frac{4203}{300}$$
$$= 14.01 \text{ mg}$$

$$s^2 = \frac{\sum X^2}{100} - \left(\frac{\sum X}{100}\right)^2$$
$$= \frac{59025}{300} - \left(\frac{4203}{300}\right)^2$$
$$= 196.75 - 196.2801$$
$$= 0.4699$$

The variance of the mean is therefore $\frac{0.4699}{300}$

And the standard error is $\sqrt{\frac{0.4699}{300}} = 0.0396$

13 $\mathrm{N}\left(\mu_A - \mu_B, \frac{8.0^2}{25} + \frac{8.0^2}{20}\right)$

$\mathrm{H}_0 : \mu_A - \mu_B = 0$

$\mathrm{H}_0 : \mu_A - \mu_B > 0$

$$Z_{\text{Test}} = \frac{\overline{X}_A - \overline{X}_B - 0}{8.0\sqrt{\frac{1}{25} + \frac{1}{20}}}$$

$$= \frac{44.2 - 40.9}{8.0\sqrt{\frac{1}{25} + \frac{1}{20}}}$$

$$= 1.375$$

$Z_{\text{Test}} = 1.375 < Z_{\text{Crit}} = 1.645$

There is no evidence for a difference in means between the two schools.

14 H_0: μ_{2010} = £9.10

H_1: μ_{2010} > £9.10

For the random sample of 100 individual sales of unleaded fuel in 2010, the mean is 9.71 and the standard deviation is $\frac{3.25}{\sqrt{100}}$

$$Z_{\text{Test}} = \frac{9.71 - 9.10}{\left(\frac{3.25}{\sqrt{100}}\right)}$$

$$= 1.8769$$

$Z_{\text{Test}} = 1.8769 > Z_{\text{Crit}} = 1.645$

Therefore, there is evidence that the 2010 sales of unleaded are higher than the 2009 sales of unleaded.

15 a H_0: $\mu_{Drive} - \mu_{Walk} = 0$
H_1: $\mu_{Drive} - \mu_{Walk} = 0$

$$Z_{\text{Test}} = \frac{\bar{X}_{\text{Drive}} - \bar{X}_{\text{Walk}}}{\sqrt{\frac{\text{Var}_{\text{Drive}}}{n_{\text{Drive}}} + \frac{\text{Var}_{\text{Walk}}}{n_{\text{Walk}}}}}$$

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

$$= 2.651$$

$Z_{Test} = 2.651 > Z_{Crit} = 1.645$
Therefore, there is evidence that those who drive to work have a higher heart rate.

b Assume normal distribution or assume sample sizes large enough to use the central limit theorem; assume individual results are independent; assume $\sigma_2 = s_2$ for both populations.

Challenge

$X_1 + \ldots + X_n \sim \text{N}(n\mu, n\sigma^2)$ and so

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) \sim \text{N}\left(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}\right) = \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$$