

Answers

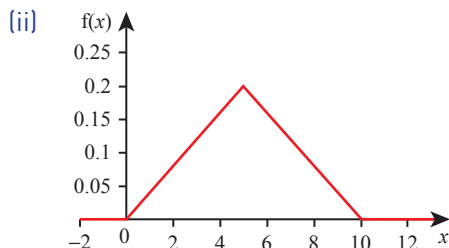
The questions, with the exception of those from past question papers, and all example answers that appear in this book were written by the authors. Cambridge Assessment International Education bears no responsibility for the example answers to questions taken from its past question papers, which are contained in this publication.

Non-exact numerical answers should be given correct to three significant figures (or one decimal place for angles in degrees) unless a different level of accuracy is specified in the question. You should avoid rounding figures until reaching your final answer.

Chapter 1

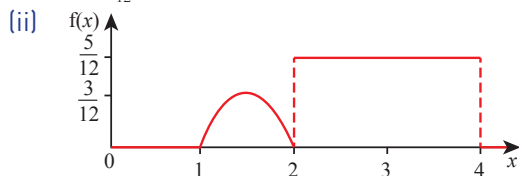
Exercise 1A (Page 7)

- 1 (i) $k = 0.04$



- (iii) 0.36

- 2 (i) $a = \frac{5}{12}$



- (iii) 0.292

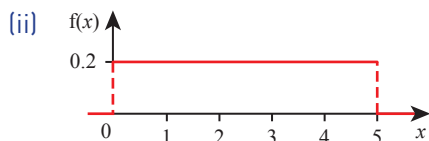
- (iv) $\frac{7}{12}$

- 3 (i) $a = 10$

- (ii) 0.12

- (iii) 0.64

- 4 (i) $k = 0.2$



- (iii) 2.5

- (iv) 7

- 5 (i) 0.8

- (ii) 0.6

- (iii) 0.026

- 6 0.8, 0.16, \$8

- 7 (i) 1.5

- (ii) 2.7

- (iii) 0.45

- (iv) 13.9

- (v) 0.45; both are the variance of Y

- 8 (i) 0.6

- (ii) -3.4

- (iii) 0.2

- (iv) 0.64

- 9 (i) $3\frac{2}{3}$

- (ii) $66\frac{1}{6}$

- (iii) $14\frac{5}{6}$, $66\frac{1}{6}$

- 10 (i) $f(x) = \frac{1}{6}$ for $2 \leq x \leq 8$

- (ii) $a = \frac{x^2\sqrt{3}}{4}$

- (iii) 0.352

- (iv) 12.12, 57.6

? (Page 13)

\$150 is quite a large prize, but probably small in comparison to the total entry fees, so it may be worth it if the model is very good. The model should work for any number of runners and should be adjustable if the nature of the entry changes.

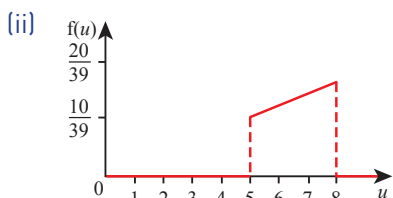
Exercise 1B (Page 23)

- 1 (i) 2.5

- (ii)
- $$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{5} & \text{for } 0 \leq x \leq 5 \\ 1 & \text{for } x > 5 \end{cases}$$

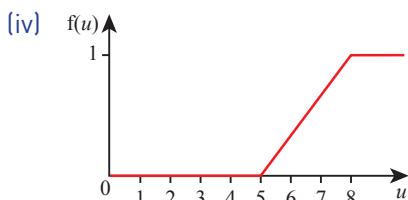
- (iii) 0.4

- 2 (i) $k = \frac{2}{39}$



(iii)

$$F(u) = \begin{cases} 0 & \text{for } u < 5 \\ \frac{u^2}{39} - \frac{25}{39} & \text{for } 5 \leq u \leq 8 \\ 1 & \text{for } u > 8 \end{cases}$$



3 (i) $c = \frac{1}{21}$

(ii)

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{x^3}{63} - \frac{1}{63} & \text{for } 1 \leq x \leq 4 \\ 1 & \text{for } x > 4 \end{cases}$$

(iii) 3.19

(iv) 4

4 (i)

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \frac{1}{(1+x)^3} & \text{for } x \geq 0 \end{cases}$$

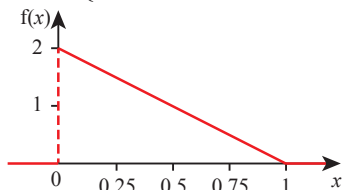
(ii) $x = 1$

5 (i) $\frac{1}{4}$

(ii) 0.134

(iii)

$$F(x) = \begin{cases} 2 - 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



6 $E(X) = \frac{9}{8}, \text{Var}(X) = \frac{171}{320}$

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{2} - \frac{x^3}{54} & \text{for } 0 \leq x \leq 3 \\ 1 & \text{for } x > 3 \end{cases}$$

7 0.5, 0.177

8 (i)

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^4 & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

(ii)

$$H(y) = \begin{cases} 0 & \text{for } y < 0 \\ y^{\frac{4}{3}} & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

(iii)

$$h(y) = \begin{cases} \frac{4}{3}y^{\frac{1}{3}} & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(iv) $\frac{1}{16} = 0.0625$

(v) 0.397

9 (i)

$$h(y) = \begin{cases} 1 & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Uniform distribution on $[0, 1]$

(ii) 0.99

(iii) 0.9

10 (i)

$$H(y) = \begin{cases} 0 & \text{for } y < 0 \\ 0.1y^{\frac{1}{4}} & \text{for } 0 \leq y \leq 10000 \\ 1 & \text{for } y > 10000 \end{cases}$$

(ii)

$$h(y) = \begin{cases} 0.025y^{-\frac{3}{4}} & \text{for } 0 \leq y \leq 10000 \\ 0 & \text{otherwise} \end{cases}$$

(iii) 0.5

(iv) 0.841

(v)

$$J(z) = \begin{cases} 0 & \text{for } z < 0 \\ 0.1z^{\frac{1}{p}} & \text{for } 0 \leq z \leq 10^p \\ 1 & \text{for } z > 10^p \end{cases}$$

(vi)

$$j(z) = \begin{cases} \frac{0.1}{p} z^{\left(\frac{1}{p}-1\right)} & \text{for } 0 \leq z \leq 10^p \\ 0 & \text{otherwise} \end{cases}$$

(vii) 0.5

11 (i)

$$H(y) = \begin{cases} 0 & \text{for } y < 45 \\ \sqrt{\frac{y}{125}} - 0.6 & \text{for } 45 \leq y \leq 320 \\ 1 & \text{for } y > 320 \end{cases}$$

- (ii)
$$h(y) = \begin{cases} (500y)^{-\frac{1}{2}} & \text{for } 45 \leq y \leq 320 \\ 0 & \text{otherwise} \end{cases}$$
- (iii) 0.2944 (to 4 d.p.)
- (iv) 5.5
- (v) $\sqrt{\frac{y}{125}} - 0.6 = 0.5$
 $y = 1.1^2 \times 125 = 151.25$
- (vi) $5 \times 5.5^2 = 151.25$
- (vii) $E(X) = 5.5$,
 $E(Y) = 162$. No since $161.7 \neq 5 \times 5.5^2$.

Chapter 2

Exercise 2A (Page 37)

- 1 (i) $\sqrt{\frac{s^2}{n}} = \sqrt{\frac{148.84}{7}} = 4.61$
- 2 (i) 11.25
- 3 (i) 322.9, 79.5
 (ii) 278.8–366.9
- 4 (i) 66, 17.2
 (ii) 51.7–80.3
 (iii) The distribution of the yield of all the fruit farmer's trees is normal.
 (iv) Number all the trees with different consecutive integers. Copy these integers on to separate pieces of paper; put these in a hat and pick out eight at random. The numbers chosen will identify the trees to be picked for the sample.
- 5 (i) 18.25, 3.72
 (ii) 16.32–20.18
 (iii) The distribution of lengths of sentences written by the accused man is normal and the text represented by the sample sentences is representative of the general length of sentences he writes.
 (iv) The sample mean lies just outside the particular 90% confidence limits provided by this sample but could well be inside those provided by other samples. However it is outside a 99% confidence interval. So it seems unlikely that the man wrote the letter.
- 6 (i) Monday
 (ii) The 23 weekdays because these are the days when the office is open and when it receives high numbers of calls.
- (iii) 630–660
- (iv) True, this distribution is not a normal one, but it may still be accurately modelled by one. $s = 37.24$ so the step size is small compared with the standard deviation. It is very common in statistics to make a normal approximation to a discrete distribution and the results are usually very reliable.
- 7 (i) 63.6–72.0
 (ii) Statistical: the distribution of tyre condemnation mileages is normal and the 12 tyres tested in the sample are representative of the distribution. Practical: the tyres are tested under genuine working conditions.
- 8 (i) 0.633–0.647
 (ii) 0.78, 0.160
 (iii) 0.616–0.944
 (iv) Large sample; no need to assume underlying normality.
- 9 (i) $-1.52 < \mu < 0.085$
 (ii) No, since the interval contains zero.
 (iii) Assumptions: random sample of areas – could well have been; normally distributed differences in unemployment rates – not unreasonable.
- 10 (i) $-0.279 < \mu < 0.186$
 (ii) Yes, since the interval contains zero.
 (iii) Assumptions: random sample of races – apparently not since all the sprints on just one afternoon are taken; normally distributed differences in timings – not unreasonable.
- 11 (i) $1.84 < \mu < 14.16$
 (ii) We do not know if the experimenter takes a random sample of the relevant population of rats. It seems plausible that the differences in the rats' times in the two conditions might be normally distributed.
 (iii) So that the effect of having learned to run the maze on the second trial affects each condition in the same way.

- 12** (i) The distribution of the contents of cola cans is normal.
 (ii) $H_0: \mu = 330, H_1: \mu < 330, -1.576 > -1.833$ so accept H_0 . There is insufficient evidence to suggest that the average content is less than 330 ml.
 (iii) No, because the sample would not be random.
- 13** (i) $t = 2.35, \nu = 11$, 1-tailed. Reject H_0 at 5% level: less.
 (ii) Assumptions: random sample of pairs; normally distributed differences between male and female heights.
- 14** (i) The distribution of the javelin speeds in the population is normal. Interval is [104.8–117.4]
 (ii) Since t and s are both smaller and n is larger $t \times \sqrt{(s/n)}$ will be smaller so the width will be smaller.
- 15** Interval is [29.38–40.92], smallest sample size is 20 ($n = 19.3$).
- 16** $H_0: \mu = 4.5, H_1: \mu > 4.5, 1.54 < 1.895$ so accept H_0 . There is insufficient evidence to suggest that the mean is greater than 4.5. Interval is [4.07, 6.56].
- 17** $H_0: \mu = 7.5, H_1: \mu < 7.5, 1.49 > 1.38$ so reject H_0 . There is sufficient evidence to suggest that the mean is less than 7.5.
- 18** Sample mean = 1.6 so $\Sigma x = 12.8, s^2 = 0.2645$ so $\Sigma x^2 = 22.3$

Investigation (Page 43)

Students' own results.

- 1** An advantage is that the after-lunch group have not done the experiment before so on both occasions the subjects are not 'in practice'. A disadvantage is that there may be differences between the two groups, particularly if they are small in number.
- 2** An advantage is that if the same people are used then any differences due to the general ability of people to estimate time will be eliminated and so the experiment will compare the before-lunch and after-lunch times without involving any other factors. A disadvantage is that the

after-lunch group have done the experiment before so the subjects will be 'in practice'.

Investigation (Page 46)

Students' own hypotheses.

? (Page 48)

- » Did you take two independent random samples of the appropriate population in the two conditions?
- » Is the variable normally distributed in each condition? The differences should be continuous rather than discrete and you could also look at a stem-and-leaf diagram for each condition – they should be unimodal and roughly symmetrical if the assumption of normality in the population is plausible.
- » Does the variable have the same variance in each condition? You could look at a stem-and-leaf diagram for each condition – they should have roughly the same spread if the assumption of equal variances in the population is plausible.

Exercise 2B (Page 55)

- 1** (i) $s = 2.646, t = 1.192, \nu = 17$, 1-tailed test H_0 even at 5% level: not heavier
 (ii) Assumptions: independent random samples on the two islands – it is hard to sample animal populations randomly – some individuals are more catchable than others; normally distributed weights on each island – seems reasonable; same variance of weights on each island – at first glance the Daphne Major variance seems larger but the samples are small.
- 2** (i) $s = 1.568, t = 1.989, \nu = 40$, 1-tailed test
 Accept H_0 at 2½% level, reject at 5% level
 (ii) Assumptions: independent random samples for the two groups – volunteer samples can be biased; normally distributed ratings with the same variance in each group – probably not a reasonable assumption with data taking only a small number of discrete values.
 (iii) Difficult because it is hard to pay two different rewards to the same person without alerting them to the point of the experiment.

- 3** (i) $s = 25.27$, $t = 1.264$, $\nu = 63$, 1-tailed test
Critical value at 5% level is more than 1.660
Assumptions: independent random samples for the two groups – hard to tell, but seems unlikely; normally distributed incomes with the same variance in each group – not reasonable, as income distribution is likely to be skewed and the variance to be different in the two groups.
- (ii) Those who have stayed on will have had fewer years in work. It would be better to investigate income eight years after completing education or income over time from age 24 to 30, say.
- 4** $t = -1.166$, $\nu = 22$, 1-tailed
Accept H_0 even at 5% level: not less
Assumptions: independent random samples of males and females; normally distributed lengths with the same variance in each group.
- 5** (i) $t = 2.35$, $\nu = 11$, 1-tailed
Reject H_0 at $1\frac{1}{2}\%$ level and above: less
- (ii) Assumptions: random sample of pairs; normally distributed differences between male and female lengths in mantis pairs.
- (iii) The paired test is better able to discriminate the difference due to sex in mantis size when other effects (e.g. age, diet, habitat) that affect male and female lengths similarly are eliminated by pairing.
- 6** $t = 1.005$, $\nu = 22$, 1-tailed
Accept H_0 even at 5% level: not more
Assumptions: independent random samples of patients for treatment or non-treatment; normally distributed post-operative stays with the same variance in each group.
- 7** (i) 1.1698–7.7192
- (ii) Assumptions: independent random samples of hen and duck eggs; normally distributed masses with the same variance for each type of egg.
- 8** Mean difference = 17.111, $s = 12.186$, $\nu = 14$, $z_a = 1.761$, upper limit 27.93
- 9** Mean difference = 241.48, $s = 136.13$, $\nu = 31$, $z_a = 1.696$ (interpolation) interval 241.48 ± 81.31
- 10** Interval is [11.92–14.58]
Distribution of fish from B is also normal with the same population variance.

$H_0: \mu_A = \mu_B$, $H_1: \mu_A > \mu_B$, $0.712 < 1.782$ so accept H_0 . There is insufficient evidence to suggest that the mean of A is greater than that of B .
Interval is $[-1.24, 2.44]$

Exercise 2C (Page 68)

- 1** 0.1161–0.2543 z_a ; $z_a = 1.96$ for 95% interval
Assumptions: random sample of birds; appropriate to use the central limit theorem to justify a normal approximation to the distribution of sample means; variance of population well approximated by its sample estimator.
- 2** $z = 1.287$, 1-tailed test
Accept H_0 at 5% level, reject at 10% level
Assumptions: independent random samples of waiting times; appropriate to use the central limit theorem to justify a normal approximation to the distribution of sample means; variances of populations well approximated by their sample estimators.
- 3** $z = 2.012$, 2-tailed test
Reject H_0 at 5% level, accept at 2% level
- 4** 221.85–281.49
 $z_a = 1.96$ for 95% interval
- 5** $z = 1.043$
Accept H_0 even at 10% level
Assumptions: independent random samples of boxes from the two machines; appropriate to use the central limit theorem to justify a normal approximation to the distributions of sample mean contents; variances of contents well approximated by their sample estimators.
- 6** (i) (a) $z = 1.699$, 1-tailed test
Accept H_0 even at 2% level
- (b) $z = 2.303$, 1-tailed test
Reject H_0 even at 2% level
- (ii) Either the variances in the two populations are, approximately, the same, in which case this is an erratic pair of samples (since the sample variances differ), possibly leading to a type II error in conclusion (a); or the variances are significantly different in the two populations and conclusion (a) is using an incorrect assumption and is hence unreliable.
- 7** (i) $z = 7.985$, Reject H_0
Assumptions: independent random samples of subjects; appropriate to use the central limit theorem to justify a normal approximation to the distribution of mean

- numbers or words recalled; variances of numbers of words recalled well approximated by their sample estimators.
- (ii) Prepare two lists of words, A and B: select a single set of subjects, divided into four equal subsets and have each subset learn and recall one of: list A randomly followed by list B alphabetically, list B randomly followed by list A alphabetically, list A alphabetically followed by list B randomly, list B alphabetically followed by list A randomly.
- 8** (i) $z_a = 5.061$, $\nu = 9$, 2-tailed test: critical value 2.262
Reject H_0
- (ii) $z_a = 1.974$, $\nu = 18$, 2-tailed test: critical value 2.101
Accept H_0 . The assumption is that the underlying population variances of the two distributions are equal.
- (iii) The first is a more discriminating analysis – a type II error is less likely because the paired design eliminates variation due to differences in resistance amongst components which might swamp the effect of the adjustment.
- 9** (i) 19.08–21.28
(ii) $z = 2.439$, 2-tailed test: critical value 1.96
Reject H_0
- 10** (i) H_0 : Mean performance difference = 0.
 H_1 : Mean performance difference $\neq 0$.
(ii) Over the population of jobs, performance differences are normally distributed.
(iii) $t = 1.090$, $\nu = 7$, 2-tailed test: critical value 3.499. Accept H_0
(iv) 3.25 ± 7.05
- 11** (i) H_0 : Mean arrivals at the two stations are equal.
 H_1 : Mean arrivals at the two stations are not equal.
(ii) Over the population of days, arrival numbers are independently and normally distributed at each station.
(iii) The variance of the arrival numbers is the same for each station.
(iv) $t = 2.507$, $\nu = 18$, 2-tailed test: critical value 2.101
Reject H_0
(v) Use a normal test (see pages 61–63)
- 12** (i) Filling volumes before and after overhaul are normally distributed with a common variance.
(ii) H_0 : Mean volumes delivered before and after overhaul are equal.
 H_1 : Mean volumes delivered before and after overhaul are not equal.
(iii) $t = 0.942$, $\nu = 15$, 2-tailed test: critical value 2.131
Accept H_0
(iv) No reason why an overhaul should not affect the variance of the volume delivered, so common variance assumption doubtful.
- 13** (ii) $E(T) = 32$. $\text{Var}(T) = 64$
(iii) The central limit theorem states that distribution of the sum of many identical random variables is normal even if the underlying distribution is not: $n = 8$ is not large, but the underlying distribution is continuous and unimodal so you can expect a reasonable approximation.
(iv) $T \approx N(8\theta, 16\theta)$
(v) The other solution represents the upper limit of a one-sided 95% confidence interval.
- 14** (i) A single sample of tasks has been taken and the times taken by the two employees are not therefore independent samples as required by an unpaired design.
(ii) Differences in times on population of tasks is normally distributed.
 $t = 2.274$, $\nu = 9$, 2-tailed test: critical value 2.262
Reject H_0
Confidence interval: 0.5–4.7

Chapter 3

? (Page 78)

They should ideally choose a random sample of customers. However, this will be difficult, so they could instead ask, for example, 1 in every 10 customers to fill in a questionnaire. They should do this over a period so that they sample customers who are watching different types of films.

? (Page 88)

- » the expected frequencies –
the first figure in each cell in the table
- » the contributions to the X^2 statistic –
the second figure in each cell in the table
- » the degrees of freedom –
next to df below the table
- » the value of the X^2 statistic –
next to X^2 below the table
- » the critical value for the test, denoted by p ;
this is the equivalent of $1 - p$ in the tables –
Next to p below the table

The observed frequencies, which is the third figure in each cell in the table.

Exercise 3A (Page 88)

- 1 (i) Walk 78, Cycle 66, Bus 113, Car 73, Age 13 172, Age 16 158
- (iv) No, because, for instance, younger students might be less likely to be allowed by their parents to cycle.

2 (i)

	Pass	Fail
Less than 10 hours	21.31	9.69
At least 10 hours	33.69	15.31

(ii)

	Pass	Fail
Less than 10 hours	3.24	7.13
At least 10 hours	2.05	4.51

- 3 $X^2 = 35.87$
 $\nu = 12$
Reject H_0 at 5% level or above; association
- 4 $X^2 = 2.886$
 $\nu = 1$
Accept H_0 at 5% level or below; independent
- 5 $X^2 = 0.955$
 $\nu = 1$
Accept H_0 at 10% level or below; no association
- 6 $X^2 = 5.36$
 $\nu = 2$
Accept H_0 at 5% level; no association
- 7 $X^2 = 10.38$
 $\nu = 1$
Reject H_0 at 1% level (c.v. = 6.635): related

- 8 $X^2 = 13.27$
 $\nu = 6$
Reject H_0 at 5% level or above: not independent

- 9 $X^2 = 1.64$
 $\nu = 2$
Accept H_0 at 5% level; no association
Minimum value of $n = 4$

- 10 H_0 : no association

H_1 : association

$$X^2 = 22.48$$

$$\nu = 9$$

Reject H_0 at 5% level (c.v. = 16.92): association

Considering the values of $\frac{(f_o - f_e)^2}{f_e}$ for each cell, shows that rural areas seem to be associated with more reasonable and excellent and less poor or good air quality than expected.

- 11 $X^2 = 2.45$
 $\nu = 2$
Accept H_0 at 5% level; no association
Minimum value of $N = 4$

- 12 (i) Some of the expected frequencies may be less than 5.

(ii) H_0 : no association, H_1 : association

(iii) $X^2 = 1.445$, $\nu = 2$, c.v. = 5.991.
Accept H_0 not enough evidence to suggest association.

(iv) Best to combine small and large businesses as these have more similarities than any other group.

H_0 : no association, H_1 : association

$$X^2 = 0.0384, \nu = 1, \text{c.v.} = 3.841.$$

Accept H_0 not enough evidence to suggest association.

- 13 (i) 4.136, 5.091
- (ii) Because some of the expected frequencies are less than 5.
- (iii) Combine 'Under 20' and '20–39' and combine '40–59' and '60 or over'.
- (iv) H_0 : no association, H_1 : association.
 $X^2 = 13.22$, $\nu = 2$, c.v. = 5.991.
Reject H_0 there is enough evidence to suggest association.

(v)

Contribution to test statistic	Pop	Classical	Jazz
Under 40	4.060	0.077	3.375
40 and over	3.086	0.058	2.565

The values of 4.06 and 3.09 show that under 40s have a strong positive association with pop, whereas 40 and over have a strong negative association with pop. The values of 3.38 and 2.57 show that under 40s have a strong negative association with jazz, whereas 40 and over have a strong positive association with jazz. The observed frequencies for classical are much as expected.

- 14 $X^2 = 0.168(A - 27)^2$
 Min $A = 23$
 Max $A = 31$
 Maximum value of $N = 4$

? (Page 93)

The data are real. The model is just your theory.

? (Page 94)

See text that follows.

? (Page 94)

See text that follows.

Activity 3.1 (Page 102)

H_0 : The binomial distribution is an appropriate model for the number of people who carry this defect.

H_1 : The binomial distribution is not an appropriate model for the number of people who carry this defect.
 $X^2 = 0.55$
 $\nu = 4$

Not significant. The data indicate that the binomial distribution is an appropriate model for the number of people who carry this defect.

Accept H_0 at 5% significance level; no association.

? (Page 108)

You can use the central limit theorem as the sample size of 60 is not small.

? (Page 112)

The fit looks suspiciously good but see the text that follows.

Exercise 3B (Page 113)

1

x	0	1	2	3	≥ 4
Exp	12.13	24.00	22.55	13.38	7.95

2

x	0	1	2	3	4	≥ 5
Exp	12.11	19.38	15.51	8.27	3.31	1.42

3

- (i) $\bar{x} = 1.725$
 H_0 : the number of mistakes on a page can be modelled by the Poisson distribution.
 H_1 : the number of mistakes on a page cannot be modelled by the Poisson distribution.
 $X^2 = 36.3$, $\nu = 4$, significant

- (ii) The mean rate may not be constant, for example she may make more mistakes when she is tired. The mistakes might not be independent if, for example, some sections are about things she cannot spell.

4

- (i) Binomial, $B(5, \frac{1}{4})$
 (ii) H_0 : the number of white flowers in each tray can be modelled by a binomial distribution, $B(5, \frac{1}{4})$
 H_1 : the number of white flowers in each tray cannot be modelled by this distribution.
 $X^2 = 0.342$, $\nu = 2$, not significant

5

- H_0 : The students guessed the answers at random.
 H_1 : The students did not guess the answers at random.
 $X^2 = 18.8$, $\nu = 4$, significant

6

- (i) H_0 : the size of rocks is distributed evenly on the scree slope.
 H_1 : the size of rocks is not distributed evenly on the scree slope.
 $X^2 = 7.82$, $\nu = 2$, significant
 (ii) The test shows that rocks of different sizes are not evenly distributed. Another, different test will be needed to determine whether the larger rocks are nearer the bottom of the slope.

7

- (i) Mean = 0.933, variance = 0.929
 (ii) $p = 0.1333$

- H_0 : these data can be modelled by the binomial distribution.
 H_1 : these data cannot be modelled by the binomial distribution.
 $X^2 = 0.28$, $\nu = 1$, not significant
- (iii) Poisson, because of general spread of data in table and because mean $\approx$ variance.
- 8** (i) Several observed frequencies are too small. In order to have $f_e \geq 5$ in each class there would be only two classes. There are two constraints and so no degrees of freedom, therefore the χ^2 test cannot be used.
- (ii) $\bar{x} = 0.92$
 H_0 : the occurrence of Li's spelling mistakes may be modelled by the Poisson distribution.
 H_1 : the occurrence Li's spelling mistakes may not be modelled by the Poisson distribution.
 $X^2 = 2.46$, $\nu = 1$, not significant
- (iii) Spelling mistakes occur singly, randomly and independently. This could be realistic.
- 9** Reject H_0 because the test statistic is much larger than the critical value, 9.488, at the 5% significance level when $\nu = 4$.
- 10** $\sum x = 5851$, $\sum x^2 = 635829 \Rightarrow S = 15.75$
 Taking 6 cells with boundaries 79.5, 89.5, ..., 119.5 gives $X^2 = 14.75$. Critical value for $\nu = 4$ at the 5% significance level is 9.488. Since $14.75 > 9.488$, H_0 is rejected; $N(100, \sigma^2)$ is not a good model.
- 11** Expected frequencies: 11.080 24.377 26.814 19.664 10.815 4.759 2.491
 $\lambda = 2.2$, $X^2 = 7.99$, $\nu = 4$
 Accept H_0 at 5% significance level. The Poisson distribution is an appropriate model.
- 12** Expected frequencies: 0.20 2.34 10.55 21.09 15.82
 $p = \frac{3}{4}$, $X^2 = 1.50$, $\nu = 1$
 Accept H_0 at 10% significance level. The binomial distribution is an appropriate model.
- 13** Expected frequencies: 10.540 23.715 26.679 20.009 11.255 5.065 1.899 0.6105 0.2275
 $\lambda = 2.25$, $X^2 = 10.8$, $\nu = 4$
 Accept H_0 at 2.5% significance level. The Poisson distribution is an appropriate model.
- 14** $X^2 = 9.13$, $\nu = 5$
 Accept H_0 at 5% significance level. The binomial distribution $B(6, 0.6)$ is an appropriate model.
- 15** (i) Mean = 4, variance = 3.57. Does support the suggestion, as the mean and variance are similar.
 (ii) 4.40, 11.7
 (iii) $\lambda = 4$, $X^2 = 5.29$, $\nu = 5$
 Accept H_0 at 10% significance level. The Poisson distribution is an appropriate model.
- 16** Geometric
 Expected frequencies: 120 48 19.2 7.68 3.072 1.229 0.819
 $p = 0.6$, $X^2 = 4.23$, $\nu = 4$
 Accept H_0 at 5% significance level. The geometric distribution is an appropriate model. Probability = 0.0126
- 17** (ii) $k = \frac{1}{72}$ $b = 37$ $c = 61$
 $X^2 = 6.69$, $\nu = 4$, $6.69 < 7.779$
 Accept H_0 at 10% significance level. The distribution is an appropriate model.
- 18** $q = 21.46$
 $X^2 = 11.3$, $\nu = 6$
 Accept H_0 at 5% significance level. The distribution is an appropriate model.

Chapter 4

2 (Page 129)

The first name is to differentiate this test from the Wilcoxon two-sample test also known as the Wilcoxon rank-sum test (which you will meet later in this chapter). The second name explains roughly how the test is carried out. The data are ranked and then the positively ranked items are added up and the negatively ranked items are added up.

Exercise 4A (Page 136)

- 1** H_0 : median = 120
 H_1 : median $>$ 120
 Critical value = 10
 $10 \geq 10$ so reject H_0 .
 The evidence suggests there has been an increase.
- 2** H_0 : median = 29
 H_1 : median $<$ 29
 Critical value = 11
 $9 < 11$ so accept H_0 .
 There is insufficient evidence to suggest that there has been a decrease.

- 3** H_0 : median = 184
 H_1 : median < 184
 Critical value = 37
 $49 > 37$ so reject H_0 .
 The evidence suggests there has been a decrease.
- 4** (i) 25
 (ii) $36 > 25$ so accept H_0 .
- 5** (i) $n = 9$
 (ii) 8
 (iii) $8 \leq 8$ so reject H_0 .
 There is sufficient evidence to suggest that the median is greater than 25.
- 6** (i) $W_- = 44$, $W_+ = 11$
 (ii) because $n = 10$ and $\frac{1}{2} \times 10 \times 11 = 55$
 (iii) $11 > 10$ so H_0 should be accepted. There is insufficient evidence to suggest that the median times have reduced.
- 7** $T = 19$, critical value = 10
 Accept H_0 ; no difference
- 8** $T = 13$; critical value = 13. Reject H_0 ; scores are better.
 She must have made the claim before she saw that the scores were better than the area average.
- 9** $T = 33$; critical value = 27. Accept H_0 ;
 insufficient evidence to suggest that inhalers are not delivering the correct amount.
- 10** (i) $T = 16$, critical value = 13
 Accept H_0 ; not greater
 The percentage of moisture in samples of grain is symmetrically distributed about its median level.
 (ii) $t = 1.933$, critical value = 1.812
 Reject H_0 ; greater
 The percentage of moisture in samples of grain is normally distributed.
 (iii) The half of the data that is between the upper and lower quartiles is closely grouped around 3.4; there are long upper and lower quartile tails and the data is positively skewed. It is this odd sample distribution that produces a significant t -statistic and an insignificant Wilcoxon statistic. It could be an erratic sample (only 11 items) or, if it is representative of the population, both normal distribution and symmetry look unreasonable as assumptions.
- 11** There are 25 sets of ranks with sums less than or equal to 8 and $\frac{25}{512} = 0.0488 < 0.05$, but 33 sets of ranks with sums less than or equal to 9 and $\frac{33}{512} = 0.0645 > 0.05$.
- 12** Assume that the population distribution is symmetrical.
 H_0 : Median = 500
 H_1 : Median > 500
 Using normal approximation
 $N(637.5, 10731.25)$
 Test statistic = 1.322
 Critical value = 1.282
 $1.322 > 1.282$ so reject H_0 .
 The evidence suggests there the new model does generate more on average.
- 13** H_0 : Median decrease is zero
 H_1 : Median decrease is positive
 (i) Critical value = 15
 $15 \leq 15$ so reject H_0 .
 The evidence suggests there has been a decrease.
 (ii) Critical value = 52
 $45 < 52$ so reject H_0 .
 The evidence suggests there has been a decrease.
- 14** H_0 : Median decrease is zero
 H_1 : Median decrease is positive
 (i) Critical value = 10
 $9 < 10$ so accept H_0
 There is insufficient evidence to suggest that there has been a decrease.
 (ii) Critical value = 21
 $18 < 21$ so reject H_0 .
 The evidence suggests there has been a decrease.
- 15** H_0 : Medians are equal
 H_1 : Medians are not equal
 (i) Critical value = 8
 $9 > 8$ so accept H_0 .
 There is insufficient evidence to suggest that one of the equation editing systems is better than the other.
 (ii) Critical value = 35
 $35 \leq 35$ so reject H_0 .
 There is sufficient evidence to suggest that one of the equation editing systems is better than the other.

Exercise 4B (Page 148)

- 1 (i) $W = 51$, critical value = 45
Accept H_0 : no difference.
- (ii) Unlikely to be representative of whole population as only from one school.
- 2 $W = 897$, critical value = 873
Accept H_0 : no difference.
A stem-and-leaf diagram of the data suggests that the men's data are much more strongly bimodal than the women's, so the distributions differ not only in their location but also in their shape.
- 3 $W = 87$, critical value = 86
Reject H_0 : harder.
- 4 (i) (a) Both histograms show approximate normality.
(b) The histograms are approximately the same shape.
- (ii) $W = 30$, critical value = 31
Reject H_0 : different.
- 5 (i) $W_{\text{to}} = 94$, $W_{\text{from}} = 116$
Critical value 1-tailed $m = n = 10$ is 82.
Since $94 > 82$, H_0 is accepted: times are not longer.
- (ii) Using paired Wilcoxon signed-rank test
 $W_+ = 50$ and $W_- = 5$.
Critical value 1-tailed $n = 10$ is 10.
Since $5 < 10$, H_0 is rejected: times are longer.
- (iii) The two tests give different results, illustrating that the paired test (when you can do it) is more sensitive than the 2-sample test.
- 6 (i) (b) $\{1, 2, 3, 4\}$, $\{1, 2, 3, 5\}$, $\{1, 2, 3, 6\}$,
 $\{1, 2, 4, 5\}$
(c) ${}^9C_4 = 126$
(e) 0.0331
- (ii) $W = 36$, critical value = 29
Accept H_0 .
- 7 (i) $W = 35$
(ii) $W = 39$
(iii) Rejection
(iv) Reject H_0

Calculate $\frac{\bar{x}_A - \bar{x}_B}{s\sqrt{\frac{1}{7} + \frac{1}{7}}}$, where $\bar{x}_A$, $\bar{x}_B$ are the sample means and s^2 is the unbiased pooled-sample estimate of the population

variance; compare this with the 1-tailed 5% critical value for the t -distribution with 12 degrees of freedom.

- 8 (ii) $\frac{1}{2}m(m+1) + mn$
(iii) $W = 52$, critical value = 53
Reject H_0 .
(iv) $N(76, 126.67)$
Test statistic = -2.13
Reject H_0 .
- 9 (i) Standard deviations are not known so normal tests should not be used. No reason to suppose the underlying distributions are normal, so t -tests should not be used.
- (ii) Leaner:
Wilcoxon (paired-sample) signed-rank test on the difference in the weights.
1-tailed test, $n = 10$, critical value = 10
 $W_- = 8$, $W_+ = 47$
Since $8 < 10$, H_0 (that there is no weight loss) is rejected. The claim that they are leaner is justified.
Faster (fitter):
Wilcoxon rank-sum test.
Critical value level with $m = n = 10$ is 82.
 $W_S = 109$, $W_E = 101$
Since $101 > 82$, the H_0 is accepted and so the organiser's claim that the children leave fitter is not accepted.
- (iii) Yes, individual times should have been recorded for the fitness test so that a paired sample test could have been used. This might well have given a different result.

Chapter 5

Exercise 5A (Page 161)

- 1 $G(t) = \frac{1}{36}t^2 + \frac{2}{36}t^3 + \frac{3}{36}t^4 + \frac{4}{36}t^5$
 $+ \frac{5}{36}t^6 + \frac{6}{36}t^7 + \frac{5}{36}t^8 + \frac{4}{36}t^9$
 $+ \frac{3}{36}t^{10} + \frac{2}{36}t^{11} + \frac{1}{36}t^{12}$
- 2 $G(t) = \frac{1}{6} + \frac{5}{18}t + \frac{2}{9}t^2 + \frac{1}{6}t^3 + \frac{1}{9}t^4 + \frac{1}{18}t^5$
- 3 $G(t) = \frac{1}{8} + \frac{3}{8}t + \frac{3}{8}t^2 + \frac{1}{8}t^3$
- 4 $G(t) = \frac{3}{5}t + \frac{3}{10}t^2 + \frac{1}{10}t^3$

$$5 \quad G(t) = \frac{3}{5}t + \frac{2}{5} \times \frac{3}{5}t^2 + \left(\frac{2}{5}\right)^2 \times \frac{3}{5}t^3 + \left(\frac{2}{5}\right)^3 \times \frac{3}{5}t^4 + \dots$$

$$+ \left(\frac{2}{5}\right)^{n-1} \times \frac{3}{5}t^n + \dots = \frac{3t}{5-2t}$$

$$6 \quad k = \frac{1}{153}; G(t) = \frac{1}{153}t + \frac{2}{153}t^2 + \frac{6}{153}t^3$$

$$+ \frac{24}{153}t^4 + \frac{120}{153}t^5$$

$$7 \quad G(t) = \frac{1}{3} + \frac{1}{2}t + \frac{1}{6}t^3$$

$$8 \quad G(t) = \frac{12}{35} + \frac{18}{35}t + \frac{1}{7}t^2$$

Exercise 5B (Page 165)

$$1 \quad E(X) = 2\frac{1}{6}; \text{Var}(X) = 2\frac{23}{36}$$

$$2 \quad E(X) = 2\frac{2}{3}; \text{Var}(X) = \frac{8}{9}$$

$$3 \quad E(X) = 3\frac{1}{2}; \text{Var}(X) = 2\frac{11}{12}$$

$$4 \quad E(X) = 1\frac{1}{2}; \text{Var}(X) = \frac{3}{4}$$

$$5 \quad \text{Var}(X) = nk(1-k)$$

$$6 \quad G(t) = \left(\frac{12}{13} + \frac{1}{13}t\right)^4; E(X) = \frac{4}{13}; \text{Var}(X) = \frac{48}{169}$$

$$8 \quad a = \frac{1}{3}, b = \frac{1}{6}, c = \frac{1}{2}$$

$$9 \quad G(t) = \frac{1}{36}t + \frac{3}{36}t^2 + \frac{5}{36}t^3 + \frac{7}{36}t^4 + \frac{9}{36}t^5$$

$$+ \frac{11}{36}t^6; E(X) = 4\frac{17}{36}; \text{Var}(X) = 1.97$$

Investigation (Page 166)

$$1 \quad G(t) = \frac{1}{3} + \frac{1}{2}t + \frac{1}{6}t^3$$

$$G'(t) = \frac{1}{2} + \frac{1}{2}t^2 \quad G''(t) = t$$

$$E(X) = \frac{1}{2} + \frac{1}{2} = 1 \quad \text{Var}(X) = 1 + 1 - 1^2 = 1$$

2 (i) Because there are 4! different possible arrangements of 4 objects. Only one is complete because all four have to match.

3 (ii) Because there is only way in which all can match. If $n-1$ match, then the final one must match too, so $P(X = n-1) = 0$.

$$4 \quad P(X = 5) = \frac{1}{120} \quad P(X = 4) = 0$$

$$P(X = 3) = \frac{1}{12} \quad P(X = 2) = \frac{1}{6}$$

$$G(t) = a + bt + \frac{1}{6}t^2 + \frac{1}{12}t^3 + \frac{1}{120}t^5$$

$$G'(t) = b + \frac{1}{3}t + \frac{1}{4}t^2 + \frac{1}{24}t^4$$

$$b = 1 - \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{24}\right) = \frac{3}{8}$$

$$a + \frac{3}{8} + \frac{1}{6} + \frac{1}{12} + \frac{1}{120} = 1$$

$$a = \frac{11}{30}$$

$$G(t) = \frac{11}{30} + \frac{3}{8}t + \frac{1}{6}t^2 + \frac{1}{12}t^3 + \frac{1}{120}t^5$$

$$5 \quad P(X = 6) = \frac{1}{720} \quad P(X = 5) = 0$$

$$P(X = 4) = \frac{1}{48} \quad P(X = 3) = \frac{1}{18}$$

$$G(t) = a + bt + ct^2 + \frac{1}{18}t^3 + \frac{1}{48}t^4 + \frac{1}{720}t^6$$

$$G'(t) = b + 2ct + \frac{1}{6}t^2 + \frac{1}{12}t^3 + \frac{1}{120}t^5$$

$$G''(t) = 2c + \frac{1}{3}t + \frac{1}{4}t^2 + \frac{1}{24}t^4$$

$$2c + \frac{1}{3} + \frac{1}{4} + \frac{1}{24} = 1 \quad c = \frac{3}{16}$$

$$= b + \frac{3}{8} + \frac{1}{6} + \frac{1}{12} + \frac{1}{120} = 1 \quad b = \frac{11}{30}$$

$$a + \frac{11}{30} + \frac{3}{16} + \frac{1}{18} + \frac{1}{48} + \frac{1}{720} = 1$$

$$a = \frac{53}{144}$$

$$G(t) = \frac{53}{144} + \frac{11}{30}t + \frac{3}{16}t^2 + \frac{1}{18}t^3$$

$$+ \frac{1}{48}t^4 + \frac{1}{720}t^6$$

6 (i) The constant term for $n-1$ is equal to the t coefficient for n .

The t coefficient for $n-1$ is equal to 2 times the t^2 coefficient for n .

The t^2 coefficient for $n-1$ is equal to 3 times the t^3 coefficient for n .

The t^3 coefficient for $n-1$ is equal to 4 times the t^4 coefficient for n .

Etc.

(ii) Generalise the above to the t^k coefficient for $n-1$ is equal to $(k+1)$ times the t^{k+1} coefficient for n and then use the fact that the coefficient of t^{n-1} is zero and the coefficient of t^n is $\frac{1}{n!}$.

Exercise 5C (Page 175)

$$1 \quad \text{(ii)} \quad G(t) = \left(\frac{t}{6-5t}\right)^2$$

$$\text{(iii)} \quad E(X) = 12; \text{Var}(X) = 60$$

- 2 (i) $G_X(t) = t(0.6 + 0.4t)$;
 $G_Y(t) = t(0.2 + 0.5t + 0.3t^2)$;
 $G(t) = G_X(t) \times G_Y(t)$
 $= t(0.6 + 0.4t) \times t(0.2 + 0.5t + 0.3t^2)$
(ii) $E(X + Y) = E(X) + E(Y) = 3.5$;
 $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 0.73$

- 3 (i) $G_X(t) = e^{3.4(t-1)}$, $G_Y(t) = e^{4.8(t-1)}$
(iii) $G_{X+Y}(t) = e^{8.2(t-1)}$

4 $G_Z(t) = G_X(t) \times G_Y(t) = \frac{t(1-t^6)}{6(1-t)} \times \frac{t(1-t^4)}{4(1-t)}$

- 5 (i) **Tip:** Let $G(t) = p_1t + p_2t^2 + p_3t + \dots$
and find $G(1) + G(-1)$.

(ii) (a) $G(t) = \frac{1}{36}(t^2 + 2t^3 + 3t^4 + 4t^5$
 $+ 5t^6 + 6t^7 + 5t^8 + 4t^9 + 3t^{10}$
 $+ 2t^{11} + t^{12}); 0.5$

(b) $G(t) = \frac{2t}{3-t}; 0.25$

6 (i) $\frac{3}{4} \times \frac{1}{5} = \frac{3}{20}$

(ii) $\frac{3}{4} \times \frac{4}{5} \times \frac{1}{4} = \frac{3}{20}$

(iii) $\frac{1}{4} + \frac{3}{5} \times \frac{1}{4} + \left(\frac{3}{5}\right)^2 \times \frac{1}{4} + \dots$
 $= \frac{1}{4} \times \left(\frac{1}{1-\frac{3}{5}}\right) = \frac{1}{4} \times \frac{5}{2} = \frac{5}{8}$

(iv) $G(t) = \frac{1}{4}t + \frac{3}{4} \times \frac{1}{5}t^2 + \frac{3}{5} \times \frac{1}{4}t^3 + \frac{3}{5}$
 $\times \frac{3}{4} \times \frac{1}{5}t^4 + \dots = \frac{1}{4}t \times \frac{1}{\left(1-\frac{3t^2}{5}\right)}$
 $+ \frac{3}{4} \times \frac{1}{5}t^2 \times \frac{1}{\left(1-\frac{3t^2}{5}\right)}$
 $= \frac{\left(\frac{1}{4}t + \frac{3}{4} \times \frac{1}{5}t^2\right)}{\left(1-\frac{3t^2}{5}\right)} = \frac{(5t + 3t^2)}{4(5 - 3t^2)}$

$G'(t) = \frac{1}{4} \times \frac{(5 - 3t^2) \times (5 + 6t) - (5t + 3t^2) \times -6t}{(5 - 3t^2)^2}$

$G'(1) = \frac{1}{4} \times \frac{2 \times 11 - (8) \times -6}{(5 - 3)^2} = \frac{1}{4} \times \frac{70}{4} = \frac{35}{8}$

$E(R) = \frac{35}{8}$

7 (i) $P(X = r) = e^{-\lambda} \frac{\lambda^r}{r!}$ $G(t) = e^{\lambda(t-1)}$

(ii) $k = \frac{1}{1 - e^{-\lambda}(\lambda + 1)}$

(iii) $G(t) = \frac{e^{-\lambda}(e^{\lambda t} - \lambda t - 1)}{1 - e^{-\lambda}(\lambda + 1)} = \frac{e^{\lambda t} - \lambda t - 1}{e^{\lambda} - \lambda - 1}$

(iv) $G'(t) = \frac{\lambda e^{\lambda t} - \lambda}{e^{\lambda} - \lambda - 1}$ so

$E(X) = \frac{\lambda e^{\lambda} - \lambda}{e^{\lambda} - \lambda - 1} = \frac{\lambda(e^{\lambda} - 1)}{e^{\lambda} - \lambda - 1}$

$G''(t) = \frac{\lambda^2 e^{\lambda t}}{e^{\lambda} - \lambda - 1}$ $\text{Var}(X) = \frac{\lambda^2 e^{\lambda}}{e^{\lambda} - \lambda - 1}$
 $+ \mu - \mu^2 = \mu \left(\frac{\lambda^2 e^{\lambda}}{\lambda(e^{\lambda} - 1)} \right) + \mu - \mu^2$
 $= \left(\frac{\mu}{\lambda(e^{\lambda} - 1)} \right) (\lambda^2 e^{\lambda} + \lambda(e^{\lambda} - 1)$
 $- \mu\lambda(e^{\lambda} - 1)) = \left(\frac{\mu}{\gamma} \right) (\lambda^2 e^{\lambda} + \gamma - \mu\gamma)$

8 $3\frac{6}{7}$

9 (i) $P(X = x) = pq^x$ $G(t) = p + pqt + pq^2t^2$
 $+ \dots = \frac{p}{1-qt}$

(ii) $G(t) = p(1-qt)^{-1}$ $G'(t) = pq(1-qt)^{-2}$

$G''(t) = 2pq^2(1-qt)^{-3}$

So $E(X) = \frac{pq}{p^2} = \frac{q}{p}$

$\text{Var}(X) = \frac{2pq^2}{p^3} + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{2pq^2 + p^2q - pq^2}{p^3}$
 $= \frac{pq^2 + p^2q}{p^3} = \frac{q^2 + pq}{p^2}$
 $= \frac{q(q+p)}{p^2} = \frac{q}{p^2}$

10 (i) $P(W = w) = e^{-\lambda_1} \frac{\lambda_1^w}{w!}$

$P(X = x) = e^{-\lambda_2} \frac{\lambda_2^x}{x!}$

$P(Y = y) = e^{-\lambda_3} \frac{\lambda_3^y}{y!}$

$$\begin{aligned}
 \text{(ii)} \quad G_W(t) &= e^{\lambda_1(t-1)} & G_X(t) &= e^{\lambda_2(t-1)} \\
 G_Y(t) &= e^{\lambda_3(t-1)} \\
 G_Z(t) &= G_{W+X+Y}(t) = e^{\lambda_1(t-1)} e^{\lambda_2(t-1)} e^{\lambda_3(t-1)} \\
 &= e^{(\lambda_1 + \lambda_2 + \lambda_3)(t-1)}
 \end{aligned}$$

$$\begin{aligned}
 \text{(iii)} \quad \text{[a]} \quad P(W + X + Y = z) \\
 &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \frac{(\lambda_1 + \lambda_2 + \lambda_3)^z}{z!}
 \end{aligned}$$

$$\text{[b]} \quad \lambda_1 + \lambda_2 + \lambda_3$$

$$\text{[c]} \quad \lambda_1 + \lambda_2 + \lambda_3$$