Chapter Review 6

1 $P(Y < y) = 1 - P(Y > y)$

So $P(Y < y) = 0.99 \Rightarrow P(Y > y) = 0.01$

$\chi^2_{10}(1\%) = 23.209$, so $P\left(\chi^2_{10} > 23.209\right) = 0.01 \Rightarrow y = 23.209$

2 $\chi^2_8(5\%) = 15.507$, so $P(\chi^2_8 > 15.507) = 0.05 \Rightarrow x = 15.507$

3 Degrees of freedom $= (5-1)\times(3-1) = 8$

From the tables: $\chi^2_8(5\%) = 15.507$

Critical region is $\chi^2 > 15.507$

4 Amalgamation gives a 3×4 contingency table.

Degrees of freedom $= (4-1)\times(3-1) = 6$

Critical value is $\chi^2_6(5\%) = 12.592$

5 H_0: There is no association between catching a cold and taking the new medicine.
H_1: There is an association between catching a cold and taking the new medicine.

These are the observed frequencies (O_i) with totals for each row and column:

	Cold	**No cold**	**Total**
Medicine	34	66	100
Placebo	45	55	100
Total	79	121	200

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency 'Cold' and 'Taken medicine' $= \frac{100\times 79}{200} = 39.5$

The expected frequency and test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
34	39.5	0.766
66	60.5	0.5
45	39.5	0.766
55	60.5	0.5

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.53$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi^2_1(5\%) = 3.841$

As 2.53 is less than 3.841, there is insufficient evidence to reject H_0 at the 5% level. It appears taking the new medicine doesn't affect the chance of a person catching a cold.

6 H_0: The data can be modelled by a Poisson distribution.
H_1: The data cannot be modelled by Poisson distribution.

Total frequency $= 38 + 32 + 10 = 80$

$$\text{Mean} = \lambda = \frac{1\times 32 + 2\times 10}{80} = \frac{52}{80} = 0.65$$

Calculate the expected frequencies as follows:

$$E_0 = 80\times P(X=0) = 80\times\frac{e^{-0.65}\,0.65^0}{0!} = 41.764$$

$$E_1 = 80\times P(X=1) = 80\times\frac{e^{-0.65}\,0.65^1}{1!} = 27.146$$

$$E_2 = 80\times P(X=2) = 80\times\frac{e^{-0.65}\,0.65^2}{2!} = 8.823$$

$$E_{i>2} = 80 - (41.764 + 27.146 + 8.823) = 2.267$$

To get values for E greater than 5, combine the last two cells:

Number of breakdowns	**0**	**1**	**$\geqslant 2$**	**Total**
Observed (O_i)	38	32	10	80
Expected (E_i)	41.764	27.146	11.090	80
$\frac{(O_i - E_i)^2}{E_i}$	0.339	0.868	0.107	1.314

The number of degrees of freedom $\nu = 1$ (three data cells with two constraints as λ is estimated by calculation)
From the tables: $\chi_1^2(5\%) = 3.841$

As 1.314 is less than 3.841, there is insufficient evidence to reject H_0 at the 5% level. The data may be modelled by a Poisson distribution.

7 H_0: There is no association between gender and passing a driving test at the first attempt.
H_1: There is an association between gender and passing a driving test at the first attempt.

These are the observed frequencies (O_i) with totals for each row and column:

	Pass	**Fail**	**Total**
Male	23	27	50
Female	32	18	50
Total	55	45	100

The expected frequency and test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
23	27.5	0.736
27	22.5	0.9
32	27.5	0.736
18	22.5	0.9

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 3.272$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi_1^2(10\%) = 2.705$

As 3.27 is greater than 2.705, reject H_0 at the 10% level. Conclude there is evidence of an association between gender and passing a driving test at the first attempt.

8 a We would expect each box to have an equal chance of being opened, and so would expect each box to have been opened 20 times.

b H_0: The data can be modelled by a discrete uniform distribution.
H_1: The data cannot be modelled by a discrete uniform distribution.

The observed and expected results are:

Box number	**1**	**2**	**3**	**4**	**5**
Observed (O_i)	20	16	25	18	21
Expected (E_i)	20	20	20	20	20
$\frac{(O_i - E_i)^2}{E_i}$	0	0.8	1.25	0.2	0.05

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.3$$

Degrees of freedom $\nu = 4$ (five data cells with a single constraint); from the tables:
$\chi_4^2(5\%) = 9.488$
As 2.3 is less than 9.488, there is insufficient evidence to reject H_0 at the 5% level.
The data may be modelled by a discrete uniform distribution.

9 a Total number of dead flies $= 0\times1+1\times1+2\times5+3\times11+4\times24+5\times8=180$
Total number of flies sprayed $= 50\times5=250$

So $\text{P(fly dies when sprayed)} = \frac{180}{250} = 0.72$

b H_0: A B(5, 0.72) distribution is a suitable model for the data.
H_1: The data cannot be modelled by a B(5, 0.72) distribution.

Find the expected frequencies by multiplying the total frequency 50 samples by the probability $P(X=i)$ using the probability equation for a binomial random variable.

$E_0 = 50\times P(X=0) = 50\times\binom{5}{0}\times0.72^0\times0.28^5 = 0.086$

$E_1 = 50\times P(X=1) = 50\times\binom{5}{1}\times0.72^1\times0.28^4 = 1.1064$

$E_2 = 50\times P(X=2) = 50\times\binom{5}{2}\times0.72^2\times0.28^3 = 5.6900$

Combine to get all E values to be 5 or more

Similarly $E_3 = 14.6313,\ E_4 = 18.8117,\ E_5 = 9.6746$

After combining the relevant cells, this gives:

Number of dead flies	$\leqslant 2$	3	4	5	Total
Observed (O_i)	7	11	24	8	50
Expected (E_i)	6.8825	14.6313	18.8117	9.6476	50
$\frac{(O_i-E_i)^2}{E_i}$	0.0020	0.9012	1.4309	0.2905	2.62

The number of degrees of freedom $\nu = 2$ (four data cells with two constraints as p is estimated by calculation)
From the tables: $\chi_2^2(5\%) = 5.991$

As 2.62 is less than 5.991, there is insufficient evidence to reject H_0 at the 5% level.
The distribution B(5, 0.72) may be a suitable model for the data.

10 H_0: The data can be modelled by a Poisson distribution.
H_1: The data cannot be modelled by Poisson distribution.

Total frequency $= 112 + 56 + 40 = 208$

$$\text{Mean} = \lambda = \frac{1\times 56 + 2\times 40}{208} = \frac{136}{208} = 0.654 \text{ (3 d.p.)}$$

Calculate the expected frequencies as follows:

$$E_0 = 208\times \text{P}(X=0) = 208\times \frac{e^{-0.654}\,0.654^0}{0!} = 108.152$$

$$E_1 = 208\times \text{P}(X=1) = 208\times \frac{e^{-0.654}\,0.654^1}{1!} = 70.731$$

$$E_2 = 208\times \text{P}(X=2) = 208\times \frac{e^{-0.654}\,0.654^2}{2!} = 23.129$$

$$E_{i>2} = 208 - (108.152 + 70.731 + 23.129) = 5.988$$

This gives all E values of 5 or more:

Number of accidents	**0**	**1**	**2**	**⩾ 3**	**Total**
Observed (O_i)	112	56	40	0	208
Expected (E_i)	108.152	70.731	23.129	5.988	208
$\frac{(O_i - E_i)^2}{E_i}$	0.1369	3.0680	12.2062	5.988	21.499

Degrees of freedom $\nu = 2$ (four data cells with two constraints as λ is estimated by calculation)
From the tables: $\chi^2_2(5\%) = 5.991$

As 21.5 is greater than 5.991, reject H_0 at the 5% level. This suggests that the data cannot be modelled by Po(0.654)

11 H_0: Rocks in site *B* occur with the same distribution as seen in the sample from site *A*
H_1: Rocks in site *B* do not occur with the same distribution as seen in the sample from site *A*

In the sample from site *A*, Igneous : Sedimentary : Other = 6 : 11 : 3
Applying this to the total 60 stones collected in site *B* to obtain expected values:

Rock type	**Igneous**	**Sedimentary**	**Other**	**Total**
Observed (O_i)	10	35	15	60
Expected (E_i)	18	33	9	60
$\frac{(O_i - E_i)^2}{E_i}$	3.556	0.121	4	7.677

Degrees of freedom $\nu = 3 - 1 = 2$, and from the tables: $\chi^2_2(5\%) = 5.991$

As 7.677 is greater than 5.991, reject H_0 at the 5% level. The distribution of rocks at Site *B* does not match the distribution seen in the sample from site *A*.

12 a Mean $= \dfrac{1\times4+2\times7+3\times8+4\times10+5\times6+6\times7+7\times4+8\times4}{4+7+8+10+6+7+4+4} = \dfrac{214}{50} = 4.28$

b H_0: The data can be modelled by a Po(4.28) distribution.
H_1: The data cannot be modelled by Po(4.28) distribution.

Calculate the expected frequencies as follows:

$$E_0 = 50\times P(X=0) = 50\times\frac{e^{-4.28}\,4.28^0}{0!} = 0.6921$$

$$E_1 = 50\times P(X=1) = 50\times\frac{e^{-4.28}\,4.28^1}{1!} = 2.9623$$

$$E_2 = 50\times P(X=2) = 50\times\frac{e^{-4.28}\,4.28^2}{2!} = 6.3394$$

Combine to get all E values to be 5 or more.

$$E_3 = 50\times P(X=3) = 50\times\frac{e^{-4.28}\,4.28^3}{3!} = 9.0442$$

Similarly $E_4 = 9.6773$, $E_5 = 8.2838$, $E_6 = 5.9091$ and $E_{i\geq7} = 7.0918$
After combining cells to ensure all values of E are greater than 5, this gives:

Weekly sales	$\leqslant 2$	**3**	**4**	**5**	**6**	$\geqslant 7$	**Total**
Observed (O_i)	11	8	10	6	7	8	50
Expected (E_i)	9.9938	9.0442	9.6773	8.2838	5.9091	7.0918	50
$\dfrac{(O_i-E_i)^2}{E_i}$	0.1013	0.1206	0.0108	0.6296	0.2014	0.1163	1.18

Degrees of freedom $\nu = 4$ (six data cells with two constraints as λ is estimated by calculation)
From the tables: $\chi^2_4(5\%) = 9.488$

As 1.18 is less than 9.488, there is insufficient evidence to reject H_0 at the 5% level.
The distribution Po(4.28) may be a suitable model for the data.

13 H_0: There is no association between gender and left- and right-handedness.
H_1: There is an association between gender and left- and right-handedness.

These are the observed frequencies (O_i) with totals for each row and column:

	Left-handed	**Right-handed**	**Total**
Male	100	600	700
Female	80	800	880
Total	180	1400	1580

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency 'Male' and 'Left-handed' $= \dfrac{700 \times 180}{1580} = 79.747$

The expected frequency and test statistic (X^2) calculations are:

O_i	E_i	$\dfrac{(O_i - E_i)^2}{E_i}$
100	79.747	5.1436
600	620.253	0.6613
80	100.253	4.0915
800	779.747	0.5260

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 10.42$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi_1^2(5\%) = 3.841$

As 10.42 is greater than 3.841, reject H_0 at the 5% level. Conclude there is evidence of an association between gender and left- and right-handedness in this population.

14 a H_0: There is no association between gender and preferred science subject.
H_1: There is no association between gender and preferred science subject.

b Total females = 130; total biology = 68; total individuals sampled = 300

$$E_{F,\,Bio} = \frac{130 \times 68}{300} = 29.47 \text{ (2 d.p.)}$$

c The expected frequency and test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
74	$\frac{170 \times 119}{300} = 67.43$	0.6401
28	$\frac{170 \times 68}{300} = 38.53$	2.8778
68	$\frac{170 \times 113}{300} = 64.03$	0.2461
45	$\frac{130 \times 119}{300} = 51.57$	0.8370
40	$\frac{130 \times 68}{300} = 29.47$	3.7625
45	$\frac{130 \times 113}{300} = 48.97$	0.3218

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.685$$

d The number of degrees of freedom $\nu = (3-1)(2-1) = 2$; from the tables: $\chi_2^2(1\%) = 9.210$
As 8.685 is less than 9.210, there is insufficient evidence to reject H_0 at the 1% level.

e From the tables: $\chi_2^2(5\%) = 5.991$
As 8.685 is greater than 5.991, H_0 would be rejected at the 5% significance level.

15 a i $P(X=1)=\frac{e^{-2.15}\times 2.15^1}{1!}=0.2504$ (4 d.p.)

ii $P(X>2)=1-P(X\leqslant 2)=1-0.6361=0.3639$ (4 d.p.)

b Mean calls received $=\frac{\sum fx}{\sum f}=\frac{10\times 0+12\times 1+14\times 2+12\times 3+8\times 4+3\times 5+1\times 6}{60}=\frac{129}{60}=2.15$

c Expected frequency $E_x=60\times P(X=x)$

$a=60\times P(X=2)=60\times 0.2692=16.15$ (2 d.p.)

$b=60-(6.99+15.03+a+11.58+6.22+2.67)=1.36$

d H_0: The data is drawn from a Poisson distribution.
H_1: The data is not drawn from a Poisson distribution.

e From part **c**, the observed and expected frequencies are:

Number of calls	**0**	**1**	**2**	**3**	**4**	**5**	**⩾ 6**	**Total**
Observed (O_i)	10	12	14	12	8	3	1	60
Expected (E_i)	6.99	15.03	16.15	11.58	6.22	2.67	1.36	60

The final three cells should be combined so that the expected value in each cell is at least 5.

f The calculation of the test statistic is:

Number of calls	**0**	**1**	**2**	**3**	**⩾ 4**	**Total**
Observed (O_i)	10	12	14	12	12	60
Expected (E_i)	6.99	15.03	16.15	11.58	10.25	60
$\frac{(O_i-E_i)^2}{E_i}$	1.2962	0.6108	0.2862	0.1523	0.2988	2.507

Degrees of freedom $\nu=3$ (five data cells with two constraints as λ is estimated by calculation)
From the tables: $\chi^2_3(5\%)=7.815$

As 2.507 is less than 7.815, there is insufficient evidence to reject H_0 at the 5% level and to conclude that the data is not drawn from Poisson distribution.

16 Each interval is 10 units, therefore the probability of a randomly selected bar falling within a given interval is $\frac{10}{60} = \frac{1}{6}$

	Probability	E_i	O_i
$0 \le d < 10$	$\frac{1}{6}$	16.67	15
$10 \le d < 20$	$\frac{1}{6}$	16.67	17
$20 \le d < 30$	$\frac{1}{6}$	16.67	18
$30 \le d < 40$	$\frac{1}{6}$	16.67	20
$40 \le d < 50$	$\frac{1}{6}$	16.67	12
$50 \le d < 60$	$\frac{1}{6}$	16.67	18

$$\chi^2_{\text{test}} = \frac{(15-16.67)^2}{16.67} + \frac{(17-16.67)^2}{16.67} + \frac{(18-16.67)^2}{16.67} + \frac{(20-16.67)^2}{16.67} + \frac{(12-16.67)^2}{16.67} + \frac{(18-16.67)^2}{16.67}$$

$$= 2.36$$

There are 6 cells and 1 restriction, therefore, $v = 6 - 1 = 5$

$\chi^2_{\text{crit}}(5) = 11.070$

$\chi^2_{\text{test}}(5) = 2.36 < \chi^2_{\text{crit}}(5) = 11.070$

Therefore, the fracture distances can be modelled by a uniform distribution.

17

	Midpoint	f	Midpoint × f	(Midpoint)^2	(Midpoint)^2× f
	2.5	7	17.5	6.25	43.75
	7.5	63	472.5	56.25	3543.75
	12.5	221	2762.5	156.25	34531.25
	17.5	177	3097.5	306.25	54206.25
	22.5	32	720	306.25	26200.00
Totals			**7070**		**108525**

$$\bar{x} = \frac{7070}{500} = 14.14$$

$$s^2 = \frac{108525}{500} - \bar{x}^2 = 17.11$$

H_0: Call length can be modelled by a normal distribution.

H_1: Call length does not approximate a normal distribution.

	$Z = \left(\frac{l - \bar{X}}{s}\right)$	F(Z)	P(Z)	E$_i$	O$_i$
$l < 5$	−2.210	0.0136	0.0136	6.80	7
$5 \le l < 10$	−1.001	0.1584	0.1448	72.4	63
$10 \le l < 15$	0.208	0.5824	0.4240	212	221
$15 \le l < 20$	1.417	0.9218	0.3394	169.7	177
$l \ge 20$			0.0782	389.1	32

$$\chi^2_{\text{test}} = \frac{(7-6.8)^2}{6.8} + \frac{(63-72.4)^2}{72.4} + \frac{(221-212)^2}{212} + \frac{(177-169.7)^2}{169.7} + \frac{(32-39.1)^2}{39.1}$$

$$= 3.212$$

There are 5 cells and 3 restrictions, therefore, $v = 5 - 3 = 2$

$$\chi^2_{\text{crit}}(2) = 5.991$$

$$\chi^2_{\text{test}}(2) = 3.212 < \chi^2_{\text{crit}}(2) = 5.991$$

Not significant, therefore accept H_0.