Essential Mathematics & Statistics for Science: Second Edition by Graham Currell & Antony Dowman (John Wiley & Sons)

Answers to 'Q' Questions

14 Frequency and Proportion

To navigate, use the PDF Bookmarks

Q14.1

Reading from the table of 1-tailed critical values for χ^2_{CRIT}:

i) χ^2_{CRIT} (α = 0.05 and df = 4) = 9.49

ii) χ^2_{CRIT} (α = 0.01 and df = 1) = 6.63

Q14.2

Reading from the table of 1-tailed critical values for χ^2_{CRIT}:
χ^2_{CRIT} (α = 0.05 and df = 3) = 7.81

We would only accept the Alternative Hypothesis that there was a significant difference between observed and expected frequencies if $\chi^2 > \chi^2_{CRIT}$

However, we find that $\chi^2 = 3.6$, and thus $\chi^2 < \chi^2_{CRIT}$

Hence we accept that the differences between observed and expected frequencies could have occurred just by chance.

Q14.3

Observed frequencies:

	Drug	**Placebo**	**Row totals**
Much Improvement	26	13	$R_1 = 39$
Some Improvement	56	36	$R_2 = 92$
No Improvement	32	37	$R_3 = 69$
Column totals:	$C_1 = 114$	$C_2 = 86$	$T = 200$

Expected frequencies

	Drug	**Placebo**	Totals
Much Improvement	22.23	16.77	R_1 = 39
Some Improvement	52.44	39.56	R_2 = 92
No Improvement	39.33	29.67	R_3 = 69
Totals	C_1 = 114	C_2 = 86	T = 200

Using [14.1] $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$

$$\chi^2 = \frac{(26-22.23)^2}{22.33} + \frac{(13-16.77)^2}{16.77} + \frac{(56-52.44)^2}{52.44} + \frac{(36-39.56)^2}{39.56} + \frac{(32-39.33)^2}{39.33} + \frac{(37-29.67)^2}{29.67}$$

= 5.23

The degrees of freedom for a Contingency Table with r rows and c columns:

$df = (r - 1) \times (c - 1) = (3\text{-}1) \times (2\text{-}1) - = 2 \times 1 = 2$

The critical χ^2 value for 2 df is 5.99

As $\chi^2 < \chi^2_{CRIT}$: - Do Not Accept the Proposed Hypothesis. The frequencies could have occurred by chance.

Q14.4

Observed Values:

		Set A					Set B		
	Soil 1	Soil 2	Soil 3	*Totals*		Soil 1	Soil 2	Soil 3	*Totals*
Green	63	79	60	202		56	79	87	222
Yellow/green	20	25	19	64		11	25	19	55
Yellow	4	11	19	34		6	10	8	24
Totals	87	115	98	300	*Totals*	73	114	114	301

Expected Values:

	58.58	77.43	65.99			53.84	84.08	84.08	
	18.56	24.53	20.91			13.34	20.83	20.83	
	9.86	13.03	11.11			5.82	9.09	9.09	

Values for $(O\text{-}E)^2 / E$

	0.33	0.03	0.54			0.09	0.30	0.09	
	0.11	0.01	0.17			0.42	0.80	0.14	
	3.48	0.32	5.61			0.01	0.09	0.12	
	Chi-squared sum		10.61			Chi-squared sum:		2.06	
	Significance Level:		0.05			Significance Level:		0.05	
	Degrees of Freedom:		4			Degrees of Freedom:		4	
	Critical value:		9.4877			Critical value:		9.4877	

Accept Alternative Hypothesis *Do NOT Reject Null Hypothesis*

Q14.5

Observed:

	Course A	Course B	Totals:
Male	32	28	60
Female	87	40	127
Totals:	119	68	187

Expected:

	Course A	Course B
Male	38.182	21.818
Female	80.818	46.182

Calculation with Yates' correction:

0.8455	1.4796
0.3995	0.699

Chi-squared sum:	3.4236
Probability:	0.05
Critical value:	3.8415
P-value (CHIDIST):	0.0643

Do NOT Reject Null Hypothesis

Calculation without Yates' correction:

1.0009	1.7515
0.4728	0.8275

Chi-squared sum:	4.0527
Probability:	0.05
Critical value:	3.8415

P-value (CHIDIST):	0.0441
P-value (CHITEST):	0.0441

Accept Proposed Hypothesis

Q14.6

Category, *i*	**Score**	**1**	**2**	**3**	**4**	**5**	**6**	**Sum**
Observed Frequency	*O*	15	22	20	14	18	31	120
Expected Frequency	*E*	20	20	20	20	20	20	120
O - E		-5	2	0	-6	-2	11	
$(O-E)^2/E$		1.25	0.2	0	1.8	0.2	6.05	
χ^2		9.5						
df = 6-1 = 5 and χ^2_{CRIT} = 11.07 at 95%. As $\chi^2 < \chi^2_{CRIT}$: - Do Not Accept the Proposed Hypothesis. The frequencies could have occurred by chance.								

Q14.7

We perform a chi-squared test with the following hypotheses:
H_O: The distribution of frequencies could have occurred by random allocation.
H_1: The distribution of frequencies could NOT have occurred by random allocation with a probability of more than 5%.
The chi-squared calculation is set out in the table below:

					Total	
Observed numbers, O_i:	44	32	56	28	160	
Expected numbers: E_i	40	40	40	40		
$Oi-Ei$	4	-8	16	-12		
$(Oi-Ei)^2$	16	64	256	144		
$(Oi-Ei)^2 / Ei$	0.4	1.6	6.4	3.6	12	

The total number of calls was 160, which, if spread evenly between the four operators, would give the Expected frequencies of 40 per operator. For each category (i.e. operator) the value of $(Oi\text{-}Ei)^2 / Ei$ is calculated. The sum of these gives the value of the χ^2 statistic:

$\chi^2 = 12$

Degrees of freedom for a one-way chi-squared test, $df = n - 1 = 4 - 1 = 3$

The critical value, χ^2_{CRIT}, for $df = 3$ and for a significance level of 0.05 is given by:

$\chi^2_{\mathrm{CRIT}} = 7.81$

As $\chi^2 > \chi^2_{\mathrm{CRIT}}$

we accept the Proposed Hypothesis - the calls were not randomly allocated.

Q14.8

We perform a chi-squared test with the following hypotheses:

H_O: The distribution of frequencies could have occurred by random allocation.

H_1: The distribution of frequencies could NOT have occurred by random allocation with a probability of more than 5%.

The chi-squared calculation is set out in the table below:

	Red	Pink	White	Total	
Observed numbers, O_i:	18	22	10	50	
Ratios:	1	2	1	4	
Expected numbers: E_i	12.5	25	12.5		
Oi-Ei	5.5	-3	-2.5		
$(Oi\text{-}Ei)^2$	30.25	9	6.25		
$(Oi\text{-}Ei)^2 / Ei$	2.42	0.36	0.5	3.28	

The total number of plants, $N = 50$.

The sum of the ratios, 1, 2 and 1, is 4.
Thus the probabilities of finding the different colours of plants are:
P(red) = 1/4, P(pink) = 2/4, P(white) = 1/4

Out of a total of 50 plants we would then expect to see the following *expected* numbers of each colour:
$E(\text{red}) = N \times P(\text{red}) = 50 \times 1/4 = 12.5$
$E(\text{pink}) = N \times P(\text{pink}) = 50 \times 2/4 = 25$
$E(\text{white}) = N \times P(\text{white}) = 50 \times 1/4 = 12.5$

For each category (i.e. operator) the value of $(Oi\text{-}Ei)^2 / Ei$ is calculated.
The sum of these gives the value of the χ^2 statistic:
$\chi^2 = 3.28$

Degrees of freedom for a one-way chi-squared test, $df = n - 1 = 3 - 1 = 2$
The critical value, χ^2_{CRIT}, for $df = 2$ and for a significance level of 0.05 is given by:
$\chi^2_{CRIT} = 5.99$
As $\chi^2 < \chi^2_{CRIT}$
Do not reject the Null Hypothesis - the calls were randomly allocated.

Q14.9

i) Probability of 60 and above = 0.0284

ii) 2 tail, need probability of 10 above 50 (>60) and 10 below 60 (>40) which is 0/0284 + 0/0284 = 0.0568

iii) The probability distribution is symmetrical about 50 so the 2 tails have the same areas.

Q14.10

i)

Using [14.6], $CI_{95\%} \approx P \pm 1.96 \times \sqrt{\dfrac{P \times (1-P)}{n}}$

$n = 100$, $P = 60/100 = 0.6$

$CI_{95\%} \approx 0.6 \pm 1.96 \times \sqrt{\dfrac{0.6 \times (1-0.6)}{100}}$ ~ 0.6 ± 0.096

~ 0.504 to 0.696

ii) Yes, 0.504>0.5

Q14.11

The results are the same if Yates' correction is used in Q14.5