Practice Questions

Questions 1 and **2** are about the following information.

The table below was created from a random sample taken from the US population and then cross-classified by **Job Satisfaction** and **Income ($US)**:

	Job Satisfaction				
Income ($US)	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied	Totals
< 6,000	20	24	80	82	206
6,000 – 15,000	22	38	104	125	289
15,000 – 25,000	13	28	81	113	235
> 25,000	7	18	54	92	171
Totals	62	108	319	412	901

A Chi-square test was conducted and the SPSS output is given below.

			Age (years)				
			Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied	Total
Income ($US)	< 6,000	Count	20	24	80	82	206
		Expected Count	14.2	24.7	72.9	94.2	206.0
		Cell contribution	2.393	0.019	0.684	1.579	
	6,000 – 15,000	Count	22	38	104	125	289
		Expected Count	19.9	34.6	102.3	132.2	289.0
		Cell contribution	0.225	0.326	0.028	0.387	
	15,000 – 25,000	Count	13	28	81	113	235
		Expected Count	16.2	28.2	83.2	107.5	235.0
		Cell contribution	0.622	0.001	0.058	0.286	
	> 25,000	Count	7	18	54	92	171
		Expected Count	11.8	20.5	60.5	78.2	171.0
		Cell contribution	1.931	0.304	0.707	2.438	
Total		Count	41	143	186	130	500
		Expected Count	41.0	143.0	186.0	130.0	500.0

test stat!

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	11.989[a]	++	.214
Likelihood Ratio	12.037	++	.211
Linear-by-Linear Association	39.546	++	.002
N of Valid Cases	500		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.77.

Note: Some values have been replaced by ++.

p-val (no ev. against H_0)

$df = (I-1)(J-1)$

$= (4-1)(4-1)$

$= 3 \times 3 = 9$

ie. all exp. counts ≥ 5 so test is valid!

1. Which one of the following statements is **false**?

T (1) The Chi-square test can be used to test for independence between *Job Satisfaction* and *Income*.

T (2) The degrees of freedom for this Chi-square test are 9.

T (3) When testing for independence between *Job Satisfaction* and *Income*, the null hypothesis tested is that *Income* is independent of *Job Satisfaction*.

T (4) For Chi-square tests, a large value for the Chi-square statistic provides evidence against the null hypothesis.

F (5) When testing for independence between *Job Satisfaction* and *Income*, a large *P*-value provides evidence against independence.

no!

2. On the basis of the information from the Chi-square test, which one of the following statements is **false**?

T (1) Small departures from independence between *Income* and *Job Satisfaction* cannot reliably be detected with small samples.

F (2) The Chi-square test works best when there are ~~small~~ large counts in several of the cells.

T (3) The sampling situation for this data is "One sample cross-classified by two factors."

T (4) Large departures from independence between *Income* and *Job Satisfaction* can reliably be detected with large samples.

T (5) This data provides no evidence of a relationship between *Income* and *Job Satisfaction*. → p-val=.214

3. Which **one** of the following statements is **false**?

T (1) A Chi-square test for independence is used to carry out formal analyses on data presented in two-way tables of counts.

T (2) The Chi-square test statistic is a measure of the difference between what we see in the data and what we would expect to see if the null hypothesis was true. → O

F (3) If all of the expected counts in a table are ~~less~~ greater → E than 5 then we would have no concerns with the validity of a Chi-square test carried out on these data.

T (4) If, for one or more cells in a table of counts, there are relatively large differences between the observed counts and the expected counts under the null hypothesis, then the data provides evidence against the null hypothesis. → i.e. big cell contr.(s)

T (5) The *P-value* in a Chi-square test is the probability, assuming that the null hypothesis is true, of observing data at least as unusual as that obtained in the data on which the test is carried out.

(see pg 7!)

Questions 4 to **10** refer to the following **KiwiSaver Data** information.

KiwiSaver is a voluntary, work-based savings initiative that was set up by the New Zealand Government in July 2007. The main purpose of KiwiSaver is to assist New Zealanders with their long-term saving for retirement.

In March 2010 the research company *UMR* conducted a telephone poll of 750 New Zealanders aged 18 years and over on issues related to KiwiSaver membership and knowledge of KiwiSaver schemes. One of the questions asked was:

Are you a member of KiwiSaver? — Yes, No

Those who responded *'Yes'* to the question above were then asked this question:

How much would you say you know about the KiwiSaver scheme you are invested in? — a lot, a fair amount, not much or hardly anything?

The respondents were also categorised into age groups. The results are shown in Table 1.

	Knowledge				
Age Group (years)	A lot	A fair amount	Not much	Hardly anything	**Total**
Under 30	8	17	20	15	60
30–44	7	22	25	21	75
45–59	6	30	31	17	84
60 plus	3	16	13	2	34
Total	24	85	89	55	253

Table 1: Knowledge about KiwiSaver scheme

A Chi-square test was carried out to see if we can claim that the pattern in knowledge about KiwiSaver depends on age group. The results of the Chi-square test are shown in Table 2.

$\frac{(O-E)^2}{E} = \frac{(6-8)^2}{8} = 0.5$

$\frac{R_i C_j}{n} = \frac{75 \times 55}{253} = 16.3$ (1dp)

Age Group (years) * Knowledge Crosstabulation

			Knowledge				
			A lot	A fair amount	Not much	Hardly anything	Total
Age Group (years)	Under 30	Count	8	17	20	15	60
		Expected Count	5.7	20.2	21.1	13.0	60.0
	30-44	Count	7	22	25	21	75
		Expected Count	7.1	++	26.4	++	75.0
	45-59	Count	6	30	31	17	84
		Expected Count	8.0	++	29.5	++	84.0
	60 plus	Count	3	16	13	2	34
		Expected Count	3.2	11.4	12.0	7.4	34.0
Total		Count	24	85	89	55	253
		Expected Count	24.0	85.0	89.0	55.0	253.0

(Q6) $\frac{84 \times 55}{253} = 18.3$ (1dp)

χ^2 (test stat)

$df = (I-1)(J-1) = (4-1)(4-1) = 3 \times 3 = 9$

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	10.244	++	.331
Likelihood Ratio	11.518	++	.242
Linear-by-Linear Association	1.675	++	.196
N of Valid Cases	253		

p-val → no ev. against H_0!

Note: Some values have been replaced by ++.

Table 2: SPSS output for knowledge about KiwiSaver scheme

4. Which **one** of the following statements is a correct **null** hypothesis for the Chi-square test described above?

(1) The underlying distribution of knowledge about KiwiSaver is associated with age group. — not

(2) The underlying distribution of age group is not the same for each level of knowledge about KiwiSaver.

(3) The underlying distribution of knowledge about KiwiSaver is independent of age group.

(4) The underlying distribution of knowledge about KiwiSaver is related to age group. — not / opposite!

(5) The underlying distribution of knowledge about KiwiSaver is not the same for each age group.

same/similar!

5. Which **one** of the following statements gives the **best** justification concerning the validity of this Chi-square test?

(1) Because there are 5 cells out of 16 with an observed count of less than 10, the results of this test are **not** valid.

(2) Because there is 1 cell out of 16 with an expected count less than 5, the results of this test are **not** valid.

(3) Because there are no cells with an observed count of less than 1, there is **no** cause for any concern with the validity.

(4) Because there are only 2 cells out of 16 with an observed count of less than 5, there is **no** cause for any concern with the validity.

(5) Because there is only 1 cell out of 16 with an expected count less than 5 and none less than 1, there is **no** cause for any concern with the validity.

all exp count > 1 and at least 80% of exp count > 5

min exp count = 3.2

15/16 = 94%.

Questions 6 to **10** assume that the use of the Chi-square test is appropriate.

(Note: This assumption may not be true.)

6. Consider the cell in Table 2, page 12, for those **aged 45 to 59** who say they know **hardly anything** about the KiwiSaver scheme they are invested in. Under the null hypothesis, the expected count for this cell is approximately:

(1) 17.0

(2) 18.3

(3) 13.8

(4) 15.8

(5) 21.0

$\frac{84 \times 33}{253}$

7. Consider the cell in Table 2, page 12, for those **aged 45 to 59** who say they know **a lot** about the KiwiSaver scheme they are invested in. This cell's contribution to the Chi-square test statistic, to 2 decimal places, is:

(1) 0.25

(2) 0.67

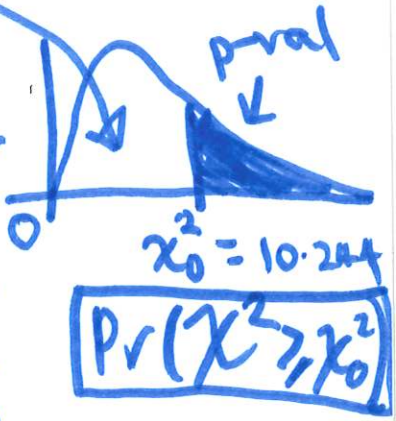
(3) 0.06

(4) 0.33

(5) 0.50

$\frac{(6-8)^2}{8}$

8. The *P-value* for this Chi-square test is calculated by:



(1) $pr(0 \leq \chi^2 \leq 10.244)$ where $\chi^2 \sim$ Chi-square(df = 9)

(2) $pr(\chi^2 \geq 10.244)$ where $\chi^2 \sim$ Chi-square(df = 16)

(3) $pr(\chi^2 \geq 10.244)$ where $\chi^2 \sim$ Chi-square(df = 9)

(4) $2 \times pr(\chi^2 \geq 10.244)$ where $\chi^2 \sim$ Chi-square(df = 16)

(5) $2 \times pr(\chi^2 \geq 10.244)$ where $\chi^2 \sim$ Chi-square(df = 9)

9. Based on the results of this Chi-square test, which **one** of the following statements is **true**?

(1) If the test statistic is at least as large as 10.244, then the probability that the null hypothesis is true is approximately 0.331.

(2) The probability that the alternative hypothesis is true is approximately 0.331.

(3) If the test statistic is at least as large as 10.244, then the probability that the alternative hypothesis is true is approximately 0.331.

(4) If the null hypothesis is true, then the probability of getting a test statistic at least as large as 10.244 is approximately 0.331.

(5) If the alternative hypothesis is true, then the probability of getting a test statistic at least as large as 10.244 is approximately 0.331.

pval = .331 (no ev. against H_0)

10. Based on the results of this Chi-square test, which **one** of the following statements is **true**?

(1) There is no evidence against the null hypothesis.

(2) There is very strong evidence for the alternative hypothesis.

(3) There is no evidence against the alternative hypothesis.

(4) There is no evidence for the null hypothesis.

(5) There is very strong evidence for the null hypothesis.

8Qs ≈ 24 mins ☕ have a break!

Questions 11 to **18** refer to the following information:

In October 2010 Research New Zealand conducted a telephone poll of 500 New Zealanders aged 15 years and over to find out whether they expected the outlook for the economy would get better or worse in the following 12 months and whether they felt their own financial circumstances would get better or worse over the same period. One of the questions asked was:

> *Do you think that your own personal financial circumstances will get better or worse over the next 12 months?*

The 500 respondents were asked to choose one of four possible responses: *Better, Stay the same, Worse* or *Don't know*. The 500 respondents were then cross-classified according to their response to the question and their age. The results are shown in Table 3.

	Age (years)				
Response	15–24	25–44	45–64	65+	**Total**
Better	8	42	48	7	105
Stay the same	18	58	76	78	230
Worse	12	37	58	39	146
Don't know	3	6	4	6	19
Total	41	143	186	130	500

Table 3: Response by age group

A Chi-square test is conducted to see whether the pattern of the responses changes with age. Note: Some values have been replaced by ++:

			Age (years)				
			15 – 24	25 – 44	45 – 64	65+	Total
Response	Better	Count	8	42	48	7	105
		Expected Count	8.61	30.03	39.06	27.30	105.0
		Cell contribution	0.043	4.771	2.046	15.095	
	Stay the same	Count	18	58	76	78	230
		Expected Count	+++	+++	85.56	59.80	230.0
		Cell contribution	0.039	0.920	1.068	5.539	
	Worse	Count	12	37	58	39	146
		Expected Count	+++	+++	54.31	37.96	146.0
		Cell contribution	0.000	0.542	+++	+++	
	Don't Know	Count	3	6	4	6	19
		Expected Count	1.56	5.43	7.07	4.94	19.0
		Cell contribution	1.335	0.059	1.332	0.227	
Total		Count	41	143	186	130	500
		Expected Count	41.0	143.0	186.0	130.0	500.0

Q14 Q15

Table 4: Chi-square test output for Response against Age

min exp. count

second smallest exp. count

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	33.296	++	.000
Likelihood Ratio	38.630	++	.000
Linear-by-Linear Association	3.880	++	.049
N of Valid Cases	500		

Table 5: Chi-square test output for Response against Age

11. Which **one** of following is a correct pair of hypotheses for this Chi-square test?

(1) H_0: The response is not related to age group.
H_1: The response is related to age group.

(2) H_0: The underlying distribution of the response is the same for all four age groups.
H_1: The underlying distribution of the response is different for all four age groups.

(3) H_0: The response does depend on age group.
H_1: The response does not depend on age group.

(4) H_0: The response is associated with age group.
H_1: The response is not associated with age group.

(5) H_0: The underlying distribution of the response is not the same for all four age groups.
H_1: The underlying distribution of the response is the same for all four age groups.

12. Which **one** of following statements gives the **best** justification for us to have no concerns about the validity of this Chi-square test?

(1) The sample size, n = 500, is greater than 30.

(2) Two cells have an observed count less than 5 and one of these cells has an expected count that is also less than 5.

(3) None of the cells has an expected count of less than 1.

(4) Only two cells have an expected count of less than 5, both of which are greater than 1.

(5) At least 80% of the cells have an expected count of 5 or more.

Questions 13 to **18** assume that the use of the Chi-square test is appropriate.

13. The number of degrees of freedom, *df*, for this Chi-square test is:

(1) 16
(2) 15
(3) 8
(4) 25
(5) 9

see pg 16

14. Consider the cell in Table 4, page 15, for 25–44 year-olds who thought that their personal financial circumstances will get worse over the following 12 months.

Under the null hypothesis, the expected count for this cell is approximately:

(1) 23.56
(2) 52.34
(3) 41.76
(4) 37
(5) 29.37

$\frac{R_i C_j}{n} = \frac{146 \times 143}{500} = 41.756$

15. Consider the cell in Table 4, page 15, for 45–64 year-olds who thought that their personal financial circumstances will get worse over the following 12 months.

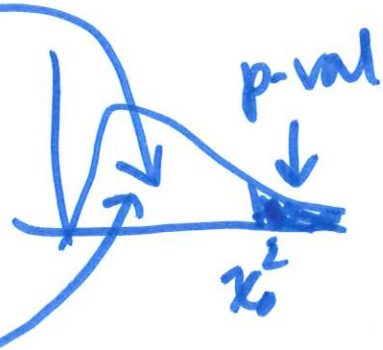
This cell's contribution to the Chi-square test statistic is:

(1) 0.235
(2) 0.213
(3) 0.068
(4) 0.064
(5) 0.251

$\frac{(O-E)^2}{E} = \frac{(58-54.31)^2}{54.31} = .251 \text{ (3dp)}$

16. The *P-value* for this Chi-square test is calculated by:

(1) $2 \times \text{pr}(\chi^2 \geq 33.296)$ where $\chi^2 \sim$ Chi-square(*df*)
(2) $\text{pr}(0 \leq \chi^2 \leq 33.296)$ where $\chi^2 \sim$ Chi-square(*df*)
(3) $2 \times \text{pr}(\chi^2 \leq 33.296)$ where $\chi^2 \sim$ Chi-square(*df*)
(4) $\text{pr}(\chi^2 \geq 33.296)$ where $\chi^2 \sim$ Chi-square(*df*)
(5) $\text{pr}(\chi^2 \leq 33.296)$ where $\chi^2 \sim$ Chi-square(*df*)



⊛ biggest contr. = 15.095 for (Better; 65+) which came about because O(7) < E(27.3)

17. Which **one** of the following statements gives the **best** explanation for the small *P-value* of 0.000?

F ✗ (1) There are far more people aged 15–24 who felt their personal financial circumstances will get better in the following 12 months than would have been expected when the null hypothesis is true. O=8

F ✗ (2) There are far fewer people aged 15–24 who felt their personal financial circumstances will get better in the following 12 months than would have been expected when the null hypothesis is true. O=8 E=8.61

O & E about the same!

T ✓ (3) There are far fewer people aged 65 years or older who felt their personal financial circumstances will get better in the following 12 months than would have been expected when the null hypothesis is true. O=7 E=8.61 E=27.30 Contr. = 15.095

T ✗ (4) The number of people aged 15–24 years who felt their personal financial circumstances will stay the same in the following 12 months and the number of people expected to be in this category when the null hypothesis is true, are quite similar. Contr. = .039!

T ✗ (5) There is almost no difference between the number of people aged 15–24 years who felt their personal financial circumstances will get worse in the following 12 months and the number of people expected to be in this category when the null hypothesis is true. Contr. = .000!

18. Which **one** of the following statements is a **valid** conclusion at the 5% level of significance? p-val = .000

✗ (1) The underlying pattern of response does not change for different age groups. ✗

✗ (2) It would not be surprising to obtain the observed differences in the patterns of response for different age groups if there is a difference in the underlying patterns of response.

✗ (3) The observed differences between age groups in the patterns of response could be just due to sampling variation alone. ✗ a diff in underlying distn. for at least 1 age group.

✓ (4) The observed differences between age groups in the patterns of response are more than just sampling variation alone.

✗ (5) It would not be surprising to obtain the observed differences in the patterns of response for different age groups if there is no difference in the underlying patterns of response. ✗ a