Questions 33 to **38** are about the following information.

A market research company interviewed 299 randomly selected car owners in Auckland. Each car owner filled out a questionnaire and this information was used to classify each person as cautious conservative (CC), middle-of-the-road (MR), or confident explorer (CE). At the same time, each person was asked to give an overall opinion of small cars.

	Self-Perception			
Opinion of Small Cars	CC	MR	CE	Totals
Favourable	79	58	49	186
Neutral	10	8	9	27
Unfavourable	10	34	42	86
Totals	99	100	100	299

Table 10: Opinion of small cars by self-perception

The market research company would like to investigate whether a person's opinion of small cars was the same regardless of their self-perception as a driver.

33. The **correct hypotheses** for the market research company's investigation are:

(1) H_0: Self-perception and opinion of small cars are related.
H_1: Self-perception and opinion of small cars are independent.

(2) H_0: For every level of opinion of small cars the distribution of self-perception is the same.
H_1: The distribution of self-perception differs for some levels of opinion of small cars.

(3) H_0: The distribution of self-perception differs for some levels of opinion of small cars.
H_1: For every level of opinion of small cars the distribution of self-perception is the same.

(4) H_0: Self-perception and opinion of small cars are associated.
H_1: Self-perception and opinion of small cars are not associated.

(5) H_0: All treatment means are the same.
H_1: There is at least one treatment mean that differs from the remaining population means.

Suppose that the market research company was also interested in investigating if there was an association between the opinion of small cars and a driver's self-perception. A Chi-square test was conducted and the results obtained from SPSS are given in Table 11 below. Some values have been replaced with ++.

			Self-Perception			Total
			CC	MR	CE	
Opinion of Small Cars	Favourable	Count	79	58	49	186
		Expected Count	61.6	62.2	62.2	186.0
		Cell contribution	4.924	0.285	++	
	Neutral	Count	10	8	9	27
		Expected Count	++	++	++	27.0
		Cell contribution	0.126	++	0.000	
	Unfavourable	Count	10	34	42	86
		Expected Count	++	++	++	86.0
		Cell contribution	++	0.954	6.092	
Total		Count	99	100	100	299
		Expected Count	99.0	100.0	100.0	299.0

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	27.289[a]	++	.000
Likelihood Ratio	30.327	++	.000
Linear-by-Linear Association	24.391	++	.000
N of Valid Cases	299		

Table 11: Opinion of small cars by self-perception

34. The **expected count** for ***middle-of-the-roaders*** who have an ***unfavourable*** *opinion of small cars* is:

(1) 285.2

(2) 27.4

(3) 0.95

(4) 28.8

(5) 34

35. Suppose the expected count for *cautious conservatives* who have an *unfavourable opinion of small cars* is 28.5. The **cell contribution** for ***cautious conservatives*** who have an ***unfavourable*** *opinion of small cars* is:

(1) 12.0
(2) 0.42
(3) −0.65
(4) 34.1
(5) 1.14

36. The **degrees of freedom** for the above test would be:

(1) 9
(2) 6
(3) 4
(4) 5
(5) 1

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

(1) pr($X^2 \geq 27.289$) where $X^2 \sim$ Chi-square(df)
(2) $2 \times$ pr($X^2 \leq 27.289$) where $X^2 \sim$ Chi-square(df)
(3) pr($X^2 \leq 27.289$) where $X^2 \sim$ Chi-square(df)
(4) $2 \times$ pr($X^2 \geq 27.289$) where $X^2 \sim$ Chi-square(df)
(5) pr($0 \leq X^2 \leq 27.289$) where $X^2 \sim$ Chi-square(df)

38. From the results of the above test the **best** interpretation would be:

(1) There is very strong evidence that the distribution of self-perception of a driver and their opinion of a small car are the same.

(2) There is no evidence that the distribution of self-perception of a driver and their opinion of a small car are not the same.

(3) There is very strong evidence that the self-perception of a driver and their opinion of a small car are not related.

(4) There is no evidence that the self-perception of a driver and their opinion of a small car are not related.

(5) There is very strong evidence that the self-perception of a driver and their opinion of a small car are related.

39. Which **one** of the following statements is **true**:

(1) If, for all cells in a table of counts, there are relatively small differences between the observed counts and the expected counts under the null hypothesis, then the data provides evidence against the null hypothesis.

(2) The greater the value of the Chi-square test statistic, the larger the *P-value*.

(3) For a Chi-square test to be valid the total count in the table, *n*, is required to be small.

(4) The *P-value* in a Chi-square test is the probability, given that the null hypothesis is true, of obtaining a test statistic as extreme, or less so, as that observed.

(5) If the *P-value* is small, then the cells with the largest contributions to the test statistic show which cells have observed counts that are far different (relatively) from those expected under the null hypothesis.

40. The **most appropriate** plot to use for analysing proportions that are to be tested using Chi-square tests is:

(1) Scatterplot

(2) Side-by-side dot plot

(3) Side-by-side box plot

(4) Table of counts

(5) Histogram

5Qs (≈15mins)

Questions 41 to **45** are about the following information.

The data in the following table came from a study of predictors of social distress among 245 American third and fifth grade children by Crick and Ladd reported in *Developmental Psychology*, 1993. One aim of the study was to determine whether there was a relationship between the level of social distress and the peer status of the child.

Level of Social Distress by Peer Status

Level of Social Distress	Peer Status: Popular	Average	Neglected	Rejected	Controversial	Total
High	8	19	10	26	2	65
Low	41	57	32	33	17	180
Total	49	76	42	59	19	245

41. Suppose it is appropriate to conduct a Chi-square test for independence (**Note**: This may **not** be correct). The null and alternative hypotheses for this test are:

(1) *H₀*: The level of social distress for the child is not independent of the peer status of the child.
H₁: The level of social distress for the child is independent of the peer status of the child.

(2) H_0: The level of social distress for the child is independent of the peer status of the child.
H_1: The level of social distress for the child is not independent of the peer status of the child.

(3) H_0: The means of the social distress levels for the child are the same for each peer status factor.
H_1: The means of the social distress levels for the child are not the same for each peer status factor.

(4) H_0: The level of social distress for the child is related to the peer status of the child.
H_1: The level of social distress for the child is not related to the peer status of the child.

(5) H_0: The distribution of the social distress levels for the child are not the same for each peer status factor.
H_1: The distribution of the social distress levels for the child are the same for each peer status factor.

The Chi-square test for independence is conducted and SPSS output is shown below. Some values have been replaced with **.

Level of Social Distress * Peer Status Crosstabulation

			Peer Status					Total
			Popular	Average	Neglected	Rejected	Controversial	
Level of Social Distress	High	Count	8	19	10	26	2	65
		Expected Count	13.0	20.2	11.1	**	5.0	65.0
	Low	Count	41	57	32	33	17	180
		Expected Count	36.0	55.8	30.9	43.3	14.0	180.0
Total		Count	49	76	42	59	19	245
		Expected Count	49.0	76.0	42.0	59.0	19.0	245.0

Q43

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14.674[a]	**	.005
Likelihood Ratio	14.541	**	.006
N of Valid Cases	245		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.04.

Q42

all exp. counts are > 5 ∴ test is valid!

The cell contributions to the Chi-square test statistic are shown in the table below:

Level of Social Distress	Peer Status				
	Popular	Average	Neglected	Rejected	Controversial
High	1.923	0.071	0.109	6.757	1.800
Low	0.694	***	0.039	2.450	0.643

Cell contributions to the Chi-square test statistic

Q44

42. The **degrees of freedom** for this test are:
   (1) 5
   (2) 10
   (3) 8
   (4) 4
   (5) 2

$df = (I-1)(J-1)$
$= (2-1)(5-1)$
$= 1 \times 4 = 4$

43. The **expected count** for those children who had a **high** level of social distress in relation to a **rejected** peer status is:

(1) 14.0
(2) 15.7
(3) 5.0
(4) 11.1
(5) 43.3

65 x 59 / 245

44. The **cell contribution** for those children who had a **low** level of social distress in relation to an **average** peer status is:

(1) 0.109
(2) 0.694
(3) 0.039
(4) 0.026
(5) 0.071

$(57 - 55.8)^2 / 55.8$

St. ev. against Ho

45. The *P-value* for the test described in Question 41 is 0.005. Which one of the following statements gives the **best** interpretation of this *P-value*?

T (1) There is strong evidence that a child's level of social distress is not independent of their peer status.

F (2) There is strong evidence that a child's level of social distress is not independent of their peer status.

F (3) There is evidence that a child's level of social distress is not independent of their peer status.

F (4) There is weak evidence that a child's level of social distress is not independent of their peer status.

F (5) There is no evidence that a child's level of social distress is not independent of their peer status.

46. Which **one** of the following statements is **false**:

(1) If one or more of the expected counts in a table is less than 1 then we would have concerns with the validity of a Chi-square test carried out on these data.

(2) If, for several 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 *P-value* for a Chi-square test will be small.

(3) The greater the value of the Chi-square test statistic, the weaker the evidence against the null hypothesis.

(4) A Chi-square test for independence is used to carry out a formal analysis on data presented in a two-way table of counts.

(5) The Chi-square test statistic is a measure of the difference, over all cells in the table, between the counts observed from the sample and the counts that would have been expected under the null hypothesis.

47. When is it **not** appropriate to conduct a Chi-square test?

(1) 1 sample cross-classified by two factors of interest.

(2) Only 20% of expected counts are greater than 5.

(3) 2 samples classified on the factor of interest.

(4) No expected counts are less than 1.

(5) Testing proportions from tables of counts.

48. Which one of the following statements about data in tables of counts is **false**?

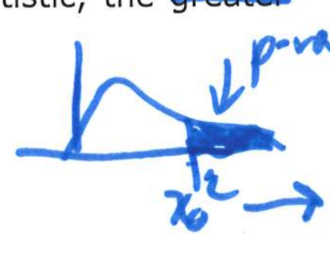
(1) The *P-value* in a Chi-square test is the probability, given that the null hypothesis is true, of obtaining a test statistic as extreme, or more so, as that observed.

(2) A Chi-square test of homogeneity on the column distributions can be used on a single random sample cross-classified by two response factors.

(3) A Chi-square test of homogeneity on the row distributions can be used on a single random sample cross-classified by two response factors.

(4) A Chi-square test of independence can be used on a single random sample cross-classified by two response factors.

(5) The greater the value of the Chi-square test statistic, the greater the *P-value*.



Answers

1. **(5)**	2. **(2)**	3. **(3)**	4. **(3)**	5. **(5)**	6. **(2)**
7. **(5)**	8. **(3)**	9. **(4)**	10. **(1)**	11. **(1)**	12. **(4)**
13. **(5)**	14. **(3)**	15. **(5)**	16. **(4)**	17. **(3)**	18. **(4)**
19. **(1)**	20. **(2)**	21. **(1)**	22. **(3)**	23. **(4)**	24. **(5)**
25. **(5)**	26. **(4)**	27. **(5)**	28. **(4)**	29. **(5)**	30. **(5)**
31. **(4)**	32. **(3)**	33. **(2)**	34. **(4)**	35. **(1)**	36. **(3)**
37. **(1)**	38. **(5)**	39. **(5)**	40. **(4)**	41. **(2)**	42. **(4)**
43. **(2)**	44. **(4)**	45. **(1)**	46. **(3)**	47. **(2)**	48. **(5)**

What should I do next?

- Do all the problems in this workshop handout and mark them. If you get a question wrong, have a look at the working on Leila's scanned slides at www.tinyURL.com/stats-CST to see how she did it.
- Go through the Chapter 9 blue pages. This includes:
  - the *notes* on page 13,
  - the *glossary* on page 14,
  - the *true/false statements* on page 15,
  - the *Sample Exam Questions* on pages 16-18, and
  - the *tutorial* material on pages 19 & 20. MCQs only!
- Attend the optional Chapter 9 tutorial.
- Try the **PRACTICE Ch9 Quiz**.
- Do three attempts of the **Chapter 9 Quiz** (due at 11pm on Wednesday 28 October 2020).
- Do the Chapter 9 parts of the Revision Assignment. Get this from Canvas under *Assignments*. **Note that this assignment is not one of the formal assessments for your final mark & grade so it is not to be handed in!**
- Try Chapter 9 questions from three of the past five exams on Canvas (get them from *Modules → Past Tests and Exams (with answers)* and use the *Exam questions index* document from there to identify the Chapter 9 questions!)
- If you get anything wrong and don't know why, get some help. You can post a question on Piazza (search first as it may have already been asked!), or talk to someone about it (your lecturer, an Assistance Room tutor or Leila).

FORMULAE

Ch 6-8, 10

Confidence intervals and t-tests

Confidence interval: $estimate \pm t \times \text{se}(estimate)$

Ch 7, 8, 10

t-test statistic: $t_0 = \dfrac{estimate - hypothesised\ value}{standard\ error}$

Applications:

Ch 6-8

1. Single mean μ: $estimate = \overline{x}$; $df = n - 1$

Ch 6, 7

2. Single proportion p: $estimate = \widehat{p}$; $df = \infty$

Ch 6-8

3. Difference between two means $\mu_1 - \mu_2$: (independent samples)

   $estimate = \overline{x}_1 - \overline{x}_2$; $df = \min(n_1 - 1, n_2 - 1)$

Ch 6, 7

4. Difference between two proportions $p_1 - p_2$:

   $estimate = \widehat{p}_1 - \widehat{p}_2$; $df = \infty$

   Situation (a): *Proportions from two independent samples*

   Situation (b): *One sample of size n, several response categories*

   Situation (c): *One sample of size n, many yes/no items*

Ch 8

The F-test (ANOVA)

F-test statistic: $f_0 = \dfrac{s_B^2}{s_W^2}$; $df_1 = k - 1$, $df_2 = n_{\text{tot}} - k$

Ch 9

The Chi-square test

"sum" or "total"

Chi-square test statistic: $\chi_0^2 = \displaystyle\sum_{\text{all cells in the table}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

→ cell contribution bit!

Expected count in cell $(i, j) = \dfrac{R_i C_j}{n}$

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

Ch 10

Regression

Fitted least-squares regression line: $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$

Inference about the intercept, β_0, and the slope, β_1: $df = n - 2$