ANSWERS CHAPTER 9

THINK IT OVER

think it over

TIO 9.1: $\chi^2 = \sum_{i=0}^{k} \frac{(f_i - e_i)}{e_i}$

Breaking the equation down: the test statistic for the chi-squared distribution is equal to the sum over all categories of the expected frequency for category *i* subtracted from the observed frequency for category *i*, all divided by the expected frequency for category *i*.

TIO 9.2: If the values are the same, the difference is 0, therefore the null hypothesis cannot be rejected.

TIO 9.3: Fred can't sit down and run a marathon at the same time, he has to choose one or the other, therefore the events are mutually exclusive. His fitness level will have a huge impact on his performance, therefore his result is dependent on his fitness level.

TIO 9.4:

$$\sum_{y=1}^{5}\sum_{x=1}^{4} x + y$$

Writing this out 'long hand': set *y* to have the value 1: (1 + 1) + (2 + 1) + (3 + 1) + (4 + 1) = 14.
Next set *x* =1: (1 + 1) + (1 + 2) + (1 + 3) + (1 + 4) + (1 + 5) = 20.
Therefore, $\sum_{y=1}^{5}\sum_{x=1}^{4} x + y = \sum_{y=1}^{5}\sum_{x=1}^{4} 14 + 20 = 34$

TIO 9.5: Each distribution has certain assumptions associated with it – Binomial: discrete random variable, events are independent with the possibility of two outcomes, e.g. coin tossing. Normal: the random variable is continuous, e.g. test scores. Poisson: discrete random variable, and the probability of events is the same for any two intervals of equal length, events are independent, e.g. number of arrivals at a café.

TIO 9.6: It would mean we would accept the alternative H_1: the number of cars entering the drive-in line during 10-minute intervals does not have a Poisson distribution. Which just tells us to think again about the arrival patterns!

TIO 9.7: Think back to a standardised normal distribution. We know that a *z*-score of 1 corresponds to 1 standard deviation from the mean. If you look up a *z*-score of 1 in a table of areas under the normal distribution you will see it corresponds to an area of 0.3413, or approximately 34%. In our case we want an area of 10% so we look for a value close to 0.1 in the table, then read off the corresponding *z*-score. In this case it is 0.25. Remember we are working with one half of the distribution and the *z*-value we read off is close to the mean not in the tail. We do the same for 20%, which is 0.53, 30% is 0.84, 40% is 1.28.

TIO 9.8: The Central Limit Theorem states that: the distribution of a mean tends to be normal, even when the distribution from which the mean is computed is decidedly non-normal.

You would be testing the mean of the mean.

TIO 9.9: SPSS assumes the data is from the population and therefore only subtracts 1 in the degrees of freedom calculation. In the example this means the number of categories minus 1, whereas the hand calculation used the number of categories minus the two estimated parameters, giving a value of 7 for the degrees of freedom.

It could potentially cause a problem since if you look at the chi-squared table and compare the values for 7 and 9 degrees of freedom at a significance of 0.05, you will see different values. With 7 degrees of freedom the value is 2.17 and for 9 it is 3.33 so if the result from the test fell between these values, how would a decision be reached? This is where you must use your judgement, because if the value of the test statistic is this close to the value given in the table, you will need to look at other measures. This is the major advantage of using SPSS since you can ask it to display the mean and standard deviation and instigate other descriptive tests. You should remember for data to be normally distributed, the standard deviation should be small compared to the mean. You could also check the kurtosis and skewness and use these values to help you reach a decision. After all that, in the example just done the test statistic of 8.4 was sufficiently large to enable us to reach a reasonably robust decision.

EXERCISES

1. If the confidence level is set to 0.05, any value smaller than this would be in the tail of the distribution, meaning there is sufficient evidence to reject the null hypothesis.
2. It means if one event occurs another cannot. For example, rolling a die: if you roll a 6 you cannot roll a 1 at the same time.
3. $\sum_{j=1}^{5}\sum_{i=1}^{4} x_{ij} = \text{set } j = 1 \sum_{i=1}^{4} x_i = 2 + 4 + 5 + 7 = 18$

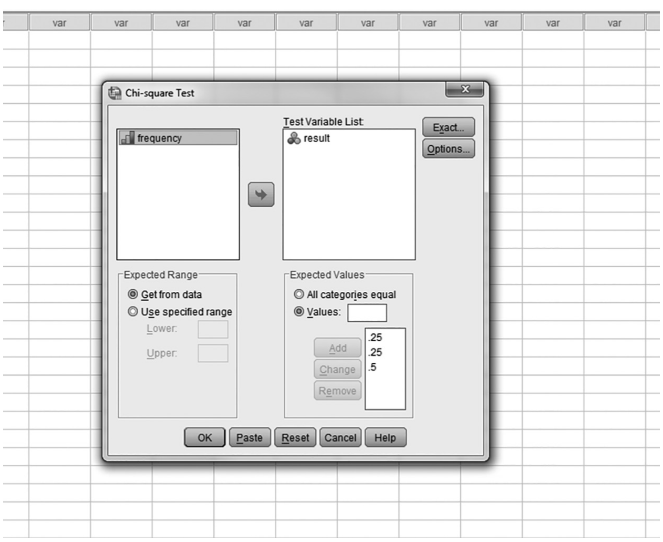
   $j = 2 \sum_{i=1}^{4} x_i = 3 + 0 + 1 + 5 = 9$

   Continue to set j up to 5. The answer is then $18 + 9 + 20 + 17 + 16 = 80$.
4. 0, i.e. rubbish!
5. The sample data is a random sampling from a fixed distribution or population where every collection of members of the population of the given sample size has an equal probability of selection.

   A sample with a sufficiently large size is assumed.

   Adequate expected cell counts. Some require 5 or more, and others require 10 or more. A common rule is 5 or more in all cells of a 2 × 2 table, and 5 or more in 80% of cells in larger tables, but no cells with zero expected count.

   The observations are always assumed to be independent of each other.
6.



result

	Observed N	Expected N	Residual
red	84	83.3	.8
blue	92	83.3	8.8
purple	157	166.5	-9.5
Total	333		

Test Statistics

	result
Chi-Square	1.468[a]
df	2
Asymp. Sig.	.480

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 83.3.

A chi-squared value of 1.468 is less than the 95% level (5.99), the null hypothesis cannot be rejected and therefore the results are consistent with the theory. Also, SPSS reports a significance value of 0.48 which is far greater than the 0.05 value of 0.103.

7. Goodness of fit refers to how close the expected frequency is to the observed frequency. In other words how well your model fits with reality!

8. Null hypothesis: students show no preference.
   Alternative hypothesis: students do show a preference.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Answer * Bar	489	100.0%	0	0.0%	489	100.0%

Answer * Bar Crosstabulation

			Bar		
			Plate of spuds	Bucket of Frogs	Total
Answer	Yes	Count	163	154	317
		Expected Count	147.2	169.8	317.0
	No	Count	64	108	172
		Expected Count	79.8	92.2	172.0
Total		Count	227	262	489
		Expected Count	227.0	262.0	489.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.053[a]	1	.003		
Continuity Correction[b]	8.490	1	.004		
Likelihood Ratio	9.130	1	.003		
Fisher's Exact Test				.003	.002
Linear-by-Linear Association	9.034	1	.003		
N of Valid Cases	489				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 79.84.

b. Computed only for a 2x2 table

$\chi^2 = 9.053$, if $p < 0.05$ then we do not have sufficient grounds to reject the null hypothesis, i.e. there is little evidence to suggest students show a preference.

9. Degrees of freedom = $(n - 1)(m - 1) = 2$ where n is the number of rows and m the number of columns.

10.

Reason * bar Crosstabulation

			bar			Total
			Spuds	Frogs	Olives	
Reason	drink cost	Count	23	7	37	67
		Expected Count	31.5	11.8	23.6	67.0
	location	Count	39	13	8	60
		Expected Count	28.2	10.6	21.2	60.0
	state	Count	13	5	13	31
		Expected Count	14.6	5.5	10.9	31.0
	other	Count	13	8	8	29
		Expected Count	13.6	5.1	10.2	29.0
Total		Count	88	33	66	187
		Expected Count	88.0	33.0	66.0	187.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.410[a]	6	.000
Likelihood Ratio	28.762	6	.000
Linear-by-Linear Association	2.664	1	.103
N of Valid Cases	187		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.12.

Depending on the wording of the null hypothesis, the test statistic indicates that if the null hypothesis was ‘there is no relationship between the primary reason for not using a bar and the bar’, then it cannot be rejected. In English, this means that students went to the different bars even though they complained about them!

11. No answer required.

12. No answer required.