study unit eighteen 18

I n this study unit we shall be working with qualitative (categorical) data for the first time. So far we have dealt only with quantitative data. When industrial psychologists work with frequencies, they have to determine somehow whether there are significant differences between them. The test statistic used for this purpose is the chi-square χ^2 test. In this study unit we deal with two types of chi-square tests (χ^2) and you will learn how to interpret their results.

Why use the chi-square test?

The chi-square test helps us to analyse frequency or categorical data. Do you still remember the difference between categorical and measurement data?

activity

Go through the index at the back of Howell and trace all references to categorical and measurement data. Also consult the study guide. Then complete the table below by listing the properties of the two types of data.

Measurement data	Categorical data

What it boils down to is that categorical data represent the number of observations (scores) in each category, whereas measurement data are obtained by measuring objects.

What types of chi-square tests are there?

By way of orientation, look at the inside of the cover of Howell and follow the tree diagram for qualitative data. You will see that we can only work with two categorical variables and that we can use only two types of chi-square tests: goodness-of-fit and contingency tables.

activity

Howell explains the reason for using the chi-square test in his introduction to chapter 19.

- Go through the introduction to chapter 19 in Howell and make sure that you see the need for using this type of statistic. The difference between the two types of chi-square tests that we shall be using — one- and two-classification variables — is explained with reference to simple examples.
- Summarise the difference between a one-classification and a two-classification variable in the table below:

One-classification variable	Two-classification variable

Now that you know the difference between the two types of variables we shall explain how to compute them.

When to use the one-classification variable: the chi-square goodness-of-fit test

When we are working with only one variable we use the goodness-of-fit test. That is the name of this particular type of chi-square test. Here the assumption is that we compare only one set of observed data with a theoretical prediction of frequencies, that is, we are working with observed (O) and expected (E) scores.

activity

At this stage we are introduced to certain formulas and symbols. To familiarise yourself with them, study section 19.1 in Howell.

- Learn the symbols of the formulas and how to compute them.
- Check the computations with your pocket calculator.
- Make sure that you can read the value off table E.1 correctly.
- Then follow the steps that Howell describes under the heading "The Chi-square distribution", on page 375.

Does the process seem familiar? It should, since it is the same procedure that is followed in all hypothesis testing (eg when computing the *t*- and *F*-tests). The only difference is that the data look different, which is why we use a different test statistic — in this case the chi-square test.

What happens when we work with more than two categories?

It is possible to classify data in more than two categories. This is simply an extension of what we were doing above.

activity

Here is an example of data in more than two categories. Work through "Extension to the multicategory case" on page 377 in Howell. You have to be able to compute it, so follow the steps and check the computations with your pocket calculator. When you have finished, do the following exercise:

Write the formulas for both types of data and summarise the difference between them in your own words:

_____

_____

_____

_____

_____

The formulas look identical; so do the computations. The only difference is that in the case of more than two categories your summation covers more categories — all of them, in fact.

By now you have come across a lot of new concepts. Let us first summarise what you have done up to now. You should be able to

- know *when* to perform the chi-square test
- compute the chi-square test (goodness-of-fit) for two or more categories
- interpret the results on the basis of rejection or nonrejection of the null hypothesis

activity

Now attempt the following exercises to evaluate your proficiency in applying the chi-square test.

Exercises 19.5 and 19.1 (Yes, in that order — 19.5 is simpler for a first attempt!)

Remember to follow all the steps that you learned in study unit 13. So you should

start by formulating a null hypothesis, then ... (Do you still remember all the steps?).

Compare your answers with the ones in the *Student handbook*.

When to use the two-classification variables: contingency table chi-square

Often the data we are using cannot be classified as a single variable. Sometimes the data fall into two categories, and the variable (data) is classified as two variables. In such cases we speak of a contingency table.

activity

Here we have another new concept. Howell explains it very simply. Work through section 19.2 for a clear grasp of the concept.

● Use your pocket calculator to check the computations.

Write your own definition of a contingency table:

_____

_____

_____

Did your definition include the following? We work with a classification in which the observation is classified in respect of two variables simultaneously.

activity

1   Read section 19.3 on "Correction for continuity" attentively.
2   Study section 19.4 in Howell.
   ● Use your pocket calculator to check the computations.

You have learnt quite a lot of new things regarding contingency tables. Having learnt these concepts and skills, you should now be able to

- explain what a contingency table is
- know when to use a contingency table
- compute the chi-square test for two classification variables
- interpret the result on the basis of rejection or nonrejection of the null hypothesis

activity

Now try to do the following exercise to evaluate your proficiency at applying the chi-square test.

1 A television company wants to find out whether there is a difference between men and women in their preference for viewing times of favourite television programmes. A group of 300 persons are randomly selected to complete a questionnaire. They are asked to choose only one of the three times indicated as their preference. The three times are indicated as late afternoon (16:00 – 18:00), early evening (18:00 – 20:00) and late evening (20:00 – 22:00). Of the female respondents, 60 indicate late afternoon as their preferred time, 40 early evening and 50 late evening. Among the male respondents 60 indicate late evening as their preferred time, while 50 indicate early evening and 40 indicate late afternoon. In order to schedule their programming optimally, the television company needs to know whether males and females differ significantly in their preference of viewing times of favourite programmes.

1.1 Make a contingency table to represent the given information.

1.2 Do the necessary calculations to determine if there is a significant difference between the preferences of the two genders in their choice of preferred viewing time.

1.3 Work on a significance level of 0,05. What is your deduction? (Refer to the research question.)

2 Also do exercise 19.11 in Howell.

1.1

	16:00-18:00	18:00-20:00	20:00-22:00	Total
Men	40 (50)	50 (45)	60 (55)	150
Women	60 (50)	40 (45)	50 (55)	150
Total	100	90	110	300

The number in parentheses are the expected frequencies, computed according to the formula:

$$E_{ij} = \frac{R_i\, C_j}{N}$$

For our question:

$$E_{11} = \frac{150 \times 100}{300} = 50$$

$$E_{12} = \frac{150 \times 90}{300} = 45$$

$$E_{13} = \frac{150 \times 110}{300} = 55$$

$$E_{21} = \frac{150 \times 100}{300} = 50$$

$$E_{22} = \frac{150 \times 90}{300} = 45$$

$$E_{23} = \frac{150 \times 110}{300} = 55$$

1.2 $\chi^2 = \Sigma \frac{(O - E)^2}{E}$

$\quad = \frac{(40 - 50)^2}{50} + \frac{(60 - 50)^2}{50} + \frac{(50 - 45)^2}{45} + \frac{(40 - 45)^2}{45} + \frac{(60 - 55)^2}{55} + \frac{(50 - 55)^2}{55}$

$\quad = \frac{(-10)^2}{50} + \frac{10^2}{50} + \frac{5^2}{45} + \frac{(-5)^2}{45} + \frac{5^2}{55} + \frac{(-5)^2}{55}$

$\quad = 2 + 2 + 0,56 + 0,56 + 0,45 + 0,45$

$\quad = 6,02$

1.3 $\quad \alpha = 0,05$

$\quad df = (R - 1)(C - 1)$

$\quad\quad = (2 - 1)(3 - 1)$

$\quad\quad = 2$

$\chi^2_{\text{calc}} = 6,02 \quad\quad \chi^2_{\text{crit}} = 5,99$

$\quad\quad 6,02 > 5,99$

$\therefore$ Reject H_0

$\therefore$ There is significant difference between men and women with regard to their preferred viewing times of favourite programmes.

2 Compare your answer with the one in the *Student handbook*.

Do small frequencies play a role?

Yes, small frequencies do play a role. Even when they are small, they influence a test. Sometimes frequencies are expressed as proportions, but usually they are converted back into frequencies. It is also possible to make nonindependent observations. Read section 19.5 in Howell attentively.
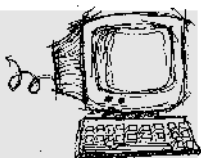
Work through section 19.8 and make sure that you can identify the χ^2-value on the printouts.

Data set exercise: Appendix D (Appenc.dta)

The output of exercise (appenc.dta) looks like this:

social problems in ninth grade	sex		
	male	female	Total
no	48	30	78
	61.54	38.46	1 00.00
	87.27	90.91	88.64
yes	7	3	10
	70.00	30.00	100.00
	12.73	9.09	11.36
Total	55	33	88
	62.50	37.50	100.00
	100.00	100.00	100.00

Pearson chi2(1) = 0.2708   Pr = 0.603

Chi square ⟶   ⟵ Probability

Thus: Do not reject H_o

study unit nineteen

multivariate statistical techniques

T he statistical techniques that you have learnt so far will enable you to understand a lot of research reports and articles, but you will also come across many studies involving multivariate statistical techniques (descriptive and inferential). You will only study these techniques at postgraduate level in Industrial Psychology, but we introduce you to them here so that you are not completely ignorant about them.

Definition of multivariate statistics

Multivariate statistics are all statistical techniques that analyse multiple (more than two) measurements of each individual or object simultaneously. These techniques are actually just extensions of the following:

- univariate distributions, such as
  — mean
  — variance and standard deviation
  — t-tests (related and independent)
  — one-way ANOVA (one dependent variable)

  and

- bivariate analyses, such as
  — correlation
  — regression (simple regression with one independent variable)

The purpose of multivariate statistical analysis is to measure, explain and predict relations between variates (weighted combinations of variables).

Every type of multivariate statistical technique requires a special measurement scale. Do you still remember the names of the four kinds of measurement scales? (We dealt with them in study unit 4.) Ranked from the simplest to the most complex in terms of their characteristics, the four scales are

- nominal
- ordinal
- interval
- ratio