7

Nonparametric tests

Chapter 7 has a similar overall structure to the previous chapter, but in this case the focus is on nonparametric tests. These are examined in relation to the numbers of samples and/or attributes (variables) and the types of summary measure under scrutiny. It mainly deals with frequency counts of nominal and ordinal attributes as opposed to scalar (ratio/interval) variables. Grid-based spatial statistics, which present an alternative to distance measures for examining the patterns in location of spatial entities, are also examined. The statistical procedures covered in this chapter are some of the most widely used by students and researchers in Geography, Earth and Environmental Science and related disciplines.

Learning outcomes

This chapter will enable readers to:

- apply nonparametric tests to nominal and ordinal attributes;
- decide whether to accept or reject the Null and Alternative Hypotheses in the case of nonparametric tests;
- explore spatial patterns using grid-based techniques;
- continue planning an independent research investigation in Geography, Earth Science and related disciplines making use on nonparametric procedures.

Practical Statistics for Geographers and Earth Scientists Nigel Walford

7.1 Introduction to nonparametric tests

The assumptions and requirements of many of the parametric techniques examined in the previous chapter often means that they are difficult to apply in Geography, Earth and Environmental Sciences. This is especially the case where the intention is to investigate the connections between people and institutions, and the physical, environmental context in which they act. Interest in these connections has become an increasingly important focus of research in recent years. Sustainability is currently a major area for research and has become central to policy at international, national and local scales. It provides a 'classic' example of an issue relating to the interactions between human and physical environments to which geographers, Earth and environmental scientists are making important contributions. The difficulties of applying parametric procedures are particularly common where the focus of attention is on the actions and behaviour of humans in different geographical contexts. It is rare for some let alone all of the attributes and variables relevant in such investigations to conform to the assumptions of normality associated with parametric statistical tests.

Nonparametric tests do not impose the requirement that the variables under scrutiny in a sample's parent population should conform to the Normal Distribution, and are sometimes referred to as 'distribution-free'. This assumption is required in parametric tests so that a valid comparison may be made between the test statistic calculated from the sample data and the corresponding value in the Normal Distribution or at least its 'close relative' the t distribution, which have known probabilities associated with them. We have seen that parametric procedures are often applied to samples of data values in order to estimate a population parameter (e.g. the mean), hence the name. However, most nonparametric procedures examine the frequency counts of data values produced when samples of observations are distributed between categories or classes. When the data values of scalar (interval or ratio) variables are divided between classes, it is possible to estimate the mean value even when the individual values are unknown as explained in Chapter 5. Thus, although nonparametric techniques do not directly estimate the value of a population parameter with a known probability expressed as a confidence interval, they do help to decide whether the distribution of observed values is significantly different from a random one and whether the approximated mean or an other measure of central tendency is in the correct range. In general terms, the Null Hypothesis in most nonparametric tests may be stated as examining whether the observed distribution of frequency counts is significantly different from what would have occurred if the entities had 'fallen' randomly into the finite number of available categories.

This chapter parallels the structure of Chapter 6 by grouping nonparametric tests according the number of samples and attributes (variables) under examination. Most of the procedures are used with either nominal or ordinal frequency distributions and the following sections progress through those dealing with one, two or three plus samples. The Pearson's Chi Square test features in several sections since it may be employed in different circumstances. In contrast, some tests, such as the enticingly named Kolmogorov–Smirnov test that deals with a single sample of observations from which a frequency distribution of a ranked or sequenced attribute can be produced, only feature once. Wilcoxon's signed ranks test for a single sample of observations measured twice on a variable, which is the nonparametric equivalent of the paired-sample t test with less stringent assumptions, is also included in this chapter.

Some spatial statistics are also examined, in particular those focusing on the distribution of spatial phenomena within a regular square grid, which provide an alternative way of exploring whether they possess a clustered, random or dispersed pattern in location. There are close links between these spatial statistics and those used for nonspatial distributions, although in the former case the categories or classes are defined within the framework of grid coordinates and in the latter with respect to the values of the attributes or variables. The location of each grid square within this framework and its count of observations are jointly of crucial importance in spatial statistics, since these define the spatial pattern and density of the entities. In some nonspatial nonparametric tests the order of the categories is also important when these represent a numerical sequence or ranking, such as occurs when they are based on measurements of distance, time, elevation, currency, etc.: in these situations where each class occurs in the sequence (its relative position) matters when investigating the frequency distribution.

7.2 One variable and one sample

Most investigations carried out as a student project or full-scale research enquiry will involve the analysis of more than one variable, although attention may quite legitimately focus on a single sample of observations. In parallel with our examination of parametric tests we begin looking at nonparametric procedures by considering the relatively simple question of whether the frequency distribution produced by placing the data values of a single attribute for a sample of observations into a set of categories is significantly different from what would have occurred if they had been spread randomly across the classes. The difference to be examined in this situation, as in most nonparametric tests, does not relate to a single value but to the contrasts between the observed and expected frequencies. Just as with parametric tests it is necessary to discover the probability of a difference having arisen by chance. If the test results indicate that the difference is too large to have occurred through chance or sampling error, then the researcher needs to explain it in terms of some human or physical process.

This raises the now familiar question of how to determine the probabilities associated with what would occur if a random process was operating, although in this case such a process would produce a distribution of counts within a series of categories. The framework of rows and/or columns into which a particular set of entities are distributed provides three types of constraint – the numbers of categories, attributes (variables) and observations – that control the probabilities. The choice between

Table 7.1 Alternative ways of obtaining expected frequency counts.

			Expected count for sample of 100 farms based on:		
Area size class (ha)	Count from previous census	Proportion	Empirical proportions	Equal proportions	Randomness
0.00–99.9	20 923	0.867	87	20	16
100.0–299.9	2389	0.099	10	20	24
300.0–499.9	523	0.022	2	20	25
500.0–699.9	155	0.006	1	20	17
700.0 & over	118	0.005	0	20	18
	24 108	1.000	100	100	100

Note: Expected counts rounded to integers.
Source: Department of the Environment, Food and Rural Affairs (2003).

a posteriori and *a priori* probabilities exists, although the frequency distribution derived from population data with which to compare a sample is often difficult to determine. In some circumstances there may be frequency distributions of the same type of observations categorized with respect to identical attributes that were produced from data collected previously or in respect of a different locality. For example, a sample-based frequency distribution may be compared with tables of statistics from an earlier national census or survey. These statistics can be used to give the empirical probabilities or proportions of the total number of observations that previously occurred. An alternative approach is to argue that the probability of an observation occurring in a particular category is in proportion to the number of categories or cells available. Table 7.1 illustrates three ways of obtaining the expected counts for a sample of 100 farms in five size categories. The first is based on empirical probabilities, where the proportions of farms in the size groups that occurred in a previous census dictate the expected number in the sample. The second approach simply assigns an equal probability (0.2) to each of the cells. Finally, a random number generator was used to obtain 100 integers between 1 and 5 corresponding to the five categories, which has resulted in some clustering in the second and third categories and rather less in the first.

7.2.1 Comparing a sample mean with a population mean

The basis of the distinction between parametric and nonparametric statistical tests is not simply that the former are concerned with summary quantities (e.g. mean and variance) and the latter with frequency counts. It is more a question that parametric tests assume variables follow the Normal Distribution, whereas nonparametric ones do not. Thus there are some nonparametric tests that make less severe assumptions than their parametric counterparts when focusing on similar statistical measures. The

Wilcoxon Signed Ranks test illustrates this point since it is used to investigate whether the difference in mean for a set of paired measurements with respect to a scalar variable for one sample of observations are significantly different from each other. It is the nonparametric equivalent of the paired-sample t test.

Box 7.1a: The Wilcoxon Signed Ranks W test.

W test statistic: $= \frac{|\bar{d} - 0|}{s_d / \sqrt{n}}$

Box 7.1b: Application of the Wilcoxon Signed Ranks W test.

The Wilcoxon Signed Ranks test hypotheses that if there is no difference between the means of the paired sets of data values for a single variable in a sample of observations (i.e. subtracting one mean from the other equals zero), the sum of the ranks for the paired measurements for each observation also equals zero. Calculations for the test statistic W exclude pairs of measurements where the difference is zero. The probabilities associated with possible values of *W* vary according to number of nonzero differences. The purpose of the W test is to discover the probability of obtaining the observed difference in the ranks as a result of chance and to decide whether the difference is significantly different from zero.

The calculations for the test statistic are relatively simple from a computational point of view, although they do involve a number of steps. First, calculate the differences between the paired measurements and then, excluding any zero values, sort these from smallest to largest ignoring the positive and negative signs (i.e. +0.04 counts as smaller than −1.25). Now assign the rank score of 1 to the smallest difference, 2 to the next and so on, applying average rank scores to equal differences. Now assign the + or − signs associated with the difference values to the rank scores and separately sum the positive and negative ranks: the smaller of these totals is the required test statistic, *W*. Depending on the nature of differences under scrutiny a one- or two-tailed W test is needed, with the former involving the probability associated with the test statistic being halved. A decision on the fate of the Null and Alternative Hypotheses may be made according to the chosen level of significance either by reference to the probability of obtaining the W test statistic itself or by transforming *W* into *Z* and then determining its probability. Transformation of *W* to *Z* is preferable when there are more than 20 nonzero differences, since the W distribution approximates the normal distribution in these circumstances.

Information obtained from the samples of residents on the Isle of Wight who were asked about their residential preferences using maps of counties in England and Wales with and without names is used to illustrate the application of the Wilcoxon Signed Ranks Test. The first two data columns in Box 7.1d below represent the mean score of each county (excluding those not favoured by any survey respondent) using the named (x_n) and unnamed maps (x_u). Having calculated the differences between these mean scores, the remaining columns represent the different steps involved in calculating *W*. The sums of the positive and negative signed ranks are respectively 438.5 and −464.5 leading to *W* equalling 438.5. Since there are more than 20 differences this has been transformed in *Z* in order to obtain the test statistic's probability. The probability of obtaining a *Z* (or *W*) test value of this size is extremely low and thus at the 0.05 level of significance the Null Hypothesis can be rejected. There does not appear to be a genuine difference between people's residential preferences when looking at maps with and without names.

The key stages in carrying out a Wilcoxon Signed Ranks test are:

State Null Hypothesis and significance level: the difference between the mean rank scores of the two sets of Isle of Wight residents and zero is the result of sampling error and is not significant at the 0.05 level of significance.

Calculate test statistic (W): the results of the calculations are given below.

Select whether to apply a one- or two-tailed test: in this case there is no sound reason to argue that the difference should be in one direction from zero or the other since people's residential preferences should not change simply because they have looked at a map with county names as opposed to one without.

Determine the probability of the calculated W (Z): the probability of obtaining $Z = 0.149$ is 0.881.

Accept or reject the Null Hypothesis: the probability of *Z* is >0.05, therefore accept the Null Hypothesis and reject the Alternative Hypothesis.

Box 7.1c: Assumptions of the Wilcoxon Signed Ranks W test.

There are three main assumptions of the W test:

The differences are independent of each other.

Zero-value differences arise because rounding values to a certain number of decimal places creates equal values, thus their exclusion is arbitrary since they would disappear if more decimal places were used.

The test is less powerful than the equivalent parametric paired t-test, although the decision in respect of the Null and Alternative Hypotheses may be the same.

7.2.2 Comparing a sample's nominal counts with a population

It is now time to introduce one of the most useful and versatile nonparametric tests generally known as the Chi-square test, or more correctly as **Pearson's Chi-square test**. It is used in different guises with nominal (categorical) counts, although the chi-square (χ^2) probability distribution that provides the means for assessing prob-

Box 7.1d: Calculation of the Wilcoxon Signed Rank W test statistic.

County	x_n	x_u	$x_n - x_u = d$	Sorted counties	Sorted absoluted	Rank	Signed rank
Avon	0.154	0.423	0.269	Hants.	0.038	4.5	4.5
Beds	0.077	0.000	−0.077	North Yorks.	0.038	4.5	4.5
Berks.	0.077	0.000	−0.077	N'hants.	0.038	4.5	4.5
Bucks.	0.154	0.077	−0.077	Northum.	0.038	4.5	−4.5
Ches.	0.000	0.154	0.154	Notts.	0.038	4.5	−4.5
Clwyd	0.000	0.192	0.192	Shrops.	0.038	4.5	−4.5
Cornwall	1.885	1.462	−0.423	Staffs.	0.038	4.5	−4.5
Cumbria	0.231	0.308	0.077	Somerset	0.038	4.5	4.5
Derbs.	0.500	0.000	−0.500	Beds.	0.077	12	−12
Devon	1.115	1.769	0.654	Berks.	0.077	12	−12
Dorset	0.846	1.269	0.423	Bucks.	0.077	12	−12
Dyfed	0.192	0.077	−0.115	Cumbria	0.077	12	12
East Sussex	0.385	0.115	−0.269	Herts.	0.077	12	−12
Essex	0.154	0.269	0.115	Powys	0.077	12	−12
Glos.	0.000	0.538	0.538	South Yorks.	0.077	12	−12
Greater London	0.269	0.808	0.538	Essex	0.115	17.5	17.5
Greater Manch.	0.154	0.000	−0.154	Dyfed	0.115	17.5	−17.5
Gwynedd	0.038	0.154	0.115	Gwynedd	0.115	17.5	17.5
Hants.	0.692	0.731	0.038	Heref & Worcs.	0.115	17.5	17.5
Heref. & Worcs.	0.000	0.115	0.115	Ches.	0.154	21	21
Herts.	0.192	0.115	−0.077	Greater Manch.	0.154	21	−21
Hum'side	0.192	0.000	−0.192	Lancs.	0.154	21	−21
Kent	0.115	0.731	0.615	Clwyd	0.192	24.5	24.5

County	x_n	x_u	$x_n - x_u = d$	Sorted counties	Sorted absoluted	Rank	Signed rank
Lancs.	0.154	0.000	−0.154	Hum'side	0.192	24.5	−24.5
Leics.	0.269	0.000	−0.269	Lincs.	0.192	24.5	−24.5
Lincs.	0.192	0.000	−0.192	Warks.	0.192	24.5	24.5
Norfolk	0.500	0.000	−0.500	Suffolk	0.231	27	−27
North Yorks.	0.115	0.154	0.038	Avon	0.269	29	29
N'hants.	0.000	0.038	0.038	Leics.	0.269	29	−29
Northum.	0.077	0.038	−0.038	East Sussex	0.269	29	−29
Notts.	0.038	0.000	−0.038	Oxfords.	0.308	31	31
Oxfords.	0.115	0.423	0.308	Dorset	0.423	32.5	32.5
Powys	0.077	0.000	−0.077	Cornwall	0.423	32.5	−32.5
Shrops.	0.154	0.115	−0.038	Derbs.	0.5	34.5	−34.5
Somerset	0.346	0.385	0.038	Norfolk	0.5	34.5	−34.5
South Yorks.	0.077	0.000	−0.077	Glos.	0.538	36.5	36.5
Staffs.	0.077	0.038	−0.038	Greater London	0.538	36.5	36.5
Suffolk	0.308	0.077	−0.231	Kent	0.615	38.5	38.5
Surrey	0.192	1.038	0.846	Wilts.	0.615	38.5	−38.5
Warks.	0.000	0.192	0.192	Devon	0.654	40	40
West Sussex	0.923	0.115	−0.808	West Sussex	0.808	41	−41
Wilts.	0.808	0.192	−0.615	Surrey	0.846	42	42
Sum of the negative ranks				−464.5			
Sum of the negative ranks				438.5			
W statistic				438.5			
Z statistic				$z = \frac{\lvert W - \mu_W \rvert}{\sigma_W} = \left\lvert 438.5 - \frac{42(42+1)}{4} \right\rvert - 0.5 \Big/ \sqrt{\frac{42(42+1)((2\times 42)+1)}{24}}$			
					$z = 0.147$		
Probability					$p = 0.881$		

abilities relates to measurements rather than frequencies. We have seen that the data values of a normally distributed variable can be transformed or standardized into Z scores (see Chapter 5). A series of n such Z scores (Z_1, Z_2, Z_3, … Z_n) can be squared and summed to produce a value known as chi-square (χ^2), thus:

$$\chi^2 = (Z_1)^2 + (Z_2)^2 + (Z_3)^2 + \ldots + (Z_n)^2$$

The value of chi-square depends on the amount of variation in the series of Z scores, if they are all identical the statistic would equal zero whereas an infinite amount of variation would produce a maximum value of plus infinity. The number of data values (the sample size) provides the degrees of freedom, which influences the form and shape of the χ^2 distribution. Small sample sizes (degrees of freedom) produce a distribution curve that is highly skewed to the left (negative), whereas once the number of cases approaches 40 or more the shape and form is similar to that produced by the Normal Distribution (Figure 7.1). Given that the form and shape of the curve varies with sample size, but that individual values of χ^2 can be obtained from different sets of data values, it follows that the probabilities associated with the test statistic are related to the degrees of freedom. The probability of obtaining a χ^2 value of 11.34 is 0.50 when there are 12 degrees of freedom (nonsignificant at 0.05) and 0.01 if $df = 3$ (significant at 0.05). The implication is that it is more difficult to obtain a larger chi-square with a relatively small number of observations. Since the minimum value χ^2 is always zero and the curve extends towards plus infinity at the extreme right of the chart, the probabilities relate to the chance of obtaining the particular χ^2 value or a larger one. Thus, if 100 samples with 3 observations were selected randomly from the same population, five samples should produce χ^2 with a value of at least 11.34; similarly 100 samples with 12 observations in each should result in 50 with a χ^2 of at least 11.34.

The chi-square statistic can be used to test for a difference between variances, but much more often the closely related Pearson's chi-squared distribution is used with counts of (n) observations distributed across (j) groups or categories. The purpose of the Pearson's Chi-square test is to examine whether the difference between the observed and expected frequency distributions of counts is more or less than would be expected to occur by chance. Pearson's Chi-square test statistic, commonly but incorrectly referred to as χ^2, is also calculated from the summation of a series of discrete elements:

$$\text{Pearson's chi-square } \chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \ldots + \frac{(O_j - E_j)^2}{E_j}$$

where O and E, respectively, refer to the observed and expected frequency counts in each of j groups or categories. The result of summing the elements in this series is to produce a statistic that very closely approximates the χ^2 distribution and its associated probabilities with $(j - 1)$ degrees of freedom.

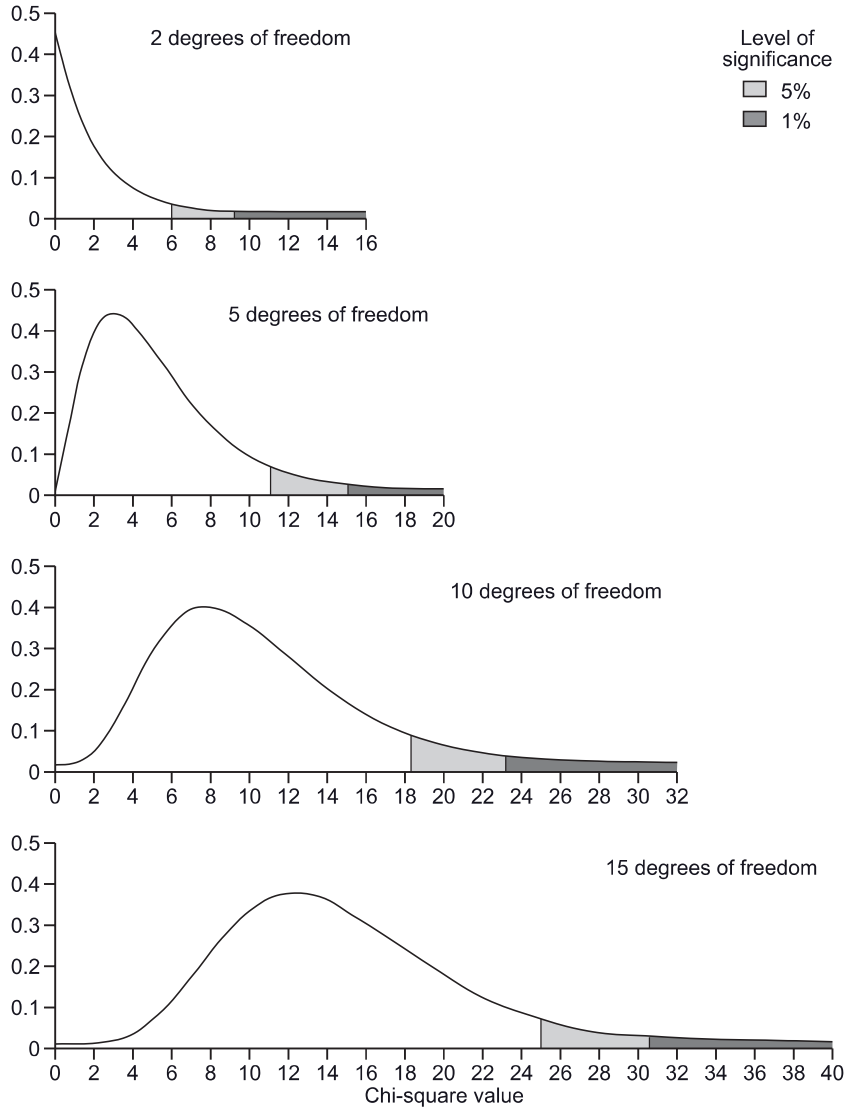


Figure 7.1 Chi-square distribution for selected degrees of freedom.

7.2.3 Comparing a sample's ordinal counts with a population

The requirement that the Pearson's Chi-square test should only be used with attributes or variables whose categories are simple nominal groups and not ordered or ranked in any way might initially seem unproblematic. After all, there must be lots of attributes and variables that simply differentiate between observations and do not

Box 7.2a: The Pearson's Chi-square (χ^2) test.

$$\text{Pearson's Chi-square } (\chi^2) \text{ test statistic: } = \sum_{j}^{1} \frac{(O_j - E_j)^2}{E_j}$$

Box 7.2b: Application of the Pearson's Chi-square (χ^2) test.

The Pearson's Chi-square test adopts the now familiar cautious stance (hypothesis) that any difference between the observed (O) and expected (E) frequency counts has arisen as a result of sampling error or chance. It works on the difference between the observed and expected counts for the entire set of observations in aggregate across the groups. Calculations for the test statistic do not involve complex mathematics, but, as with most tests, it can readily be produced using statistical software. The degrees of freedom are defined as the number of groups or cells in the frequency distribution minus 1 ($j - 1$). The probabilities of the test statistic depend on the degrees of freedom. The calculations involve dividing the squared difference between the observed and expected frequency count in each cell by the expected frequency and then summing these values to produce the test statistic. Comparison of the probability of having obtained this test statistic value in relation to the chosen level of significance enables a decision to be made regarding whether to accept or reject the Null Hypothesis, and by implication the Alternative Hypothesis.

The Pearson's Chi-square test has been applied to households in four villages in mid-Wales that have been pooled into a single sample in respect of the frequency distribution of the mode of transport used when doing their main household shopping for groceries. This comprises the univariate frequency distribution reproduced in Box 7.2d below. The expected frequencies have been calculated in two ways: on the basis that households would be distributed in equal proportions between the four modes of transport; and alternatively that they would occur in the unequal proportions obtained from the results of a national survey of households' shopping behaviour. The test statistics for these two applications of the procedure are, respectively, 9.12 and 14.80, whose corresponding probabilities are 0.023 and 0.002. These are both lower than the 0.05 significance level and so the Alternative Hypothesis is accepted.

The key stages in carrying out a Pearson's Chi-square test are:

State Null Hypothesis and significance level: the difference between the observed and expected frequencies has arisen as a result of sampling error and is not significant at the 0.05 level of significance. The sample frequency distribution is not significantly different from either an equal allocation between the categories or what would be expected according to a previous national survey.

Calculate test statistic (χ^2): the results of the calculations are given below.

Determine the degrees of freedom: in this example there are $4 - 1 = 3$ degrees of freedom.

Determine the probability of the calculated χ^2: the probability of obtaining $\chi^2 = 9.12$ and $\chi^2 = 14.80$ each with 3 degrees of freedom are, respectively, 0.023 and 0.002.

Accept or reject the Null Hypothesis: the probability of both χ^2 values is <0.05, therefore reject the Null Hypotheses and accept the Alternative Hypothesis.

Box 7.2c: Assumptions of the Pearson's Chi-square (χ^2) test.

There are five main assumptions of the Pearson's Chi-square test:

Simple or stratified random sampling should be used to select observations that should be independent of each other.

The number of categories or classes has a major effect on the size of the test statistic. In the case of nominal categories (as in this transport to shop example) these may be predetermined, although they can be collapsed but not expanded (e.g. family car and other car could be merged). However, there are many possibilities for classifying continuous interval or ratio scale variables (see Chapter 5), although this should not produce ordered classes.

Frequencies expressed as percentages (relative frequencies) will produce an inaccurate probability for Pearson's Chi-square statistic: if the sample size is <100 it will be too low and if >100 to high. If, as in the worked example, the sample size happens to equal 100 then absolute and relative frequencies are the same and the probability is accurate.

The minimum expected frequency count per cell is 5 and overall no more than 20% of the cells should be at or below this level. The number of cells with low expected frequencies may be adjusted by altering classes or categories.

The categories or classes should not be ordered or ranked (e.g. a sequence of distance groups or ordered social classes) because the same χ^2 will be obtained irrespective of the order in which categories are tabulated.

Box 7.2d: Calculation of the Pearson's Chi-square (χ^2) test.

Transport mode	Observed sample frequencies	Exp. freq. (equal props)	$\frac{(O-E)^2}{E}$	Exp. freq. (props from national survey)	$\frac{(O-E)^2}{E}$
Family car	36	25	4.84	40	0.40
Other car	18	25	1.96	10	6.40
Public transport	18	25	1.96	30	4.80
Walk or cycle	28	25	0.36	20	3.20
Total	100	100	9.12	100	14.80
Pearson's chi-square statistic	Equal proportions		$\chi^2 = 9.12$	Proportions from national survey	$\chi^2 = 14.80$
Probability	Equal proportions		$p = 0.023$	Proportions from national survey	$p = 0.002$

put them in order. However, the more you think about it the more classified variables measuring such things as distance, area, temperature, acidity/alkalinity, volume and population as well as various financial quantities (land value, income, expenditure, rental value, etc.) do in fact summarize a sequence of data values. For example, parcels of land classified according to different types of use and assigned to zones from a city centre constitute an ordered sequence of distance zones. Perhaps there are only relatively few truly nominal categories such as gender, household tenure, mode of transport, ethnicity, aspect, etc. Nevertheless, some form of statistical test is required that enables investigators to get around the assumption that the Pearson's Chi-square test should not be used with ordered categories. One such test is the **Kolmogorov–Smirnov D test**.

One of the most typical applications of this test involves comparing an observed set of frequency counts for a ranked attribute or variable with the frequencies expected to arise by applying a probability distribution, such as the Poisson distribution. This test focuses on the cumulative observed and expected probabilities of the number of observations in each of the series of ordered classes. The D statistic is simply the maximum absolute difference between the cumulative observed and expected probabilities. Because the observations are distributed between an **ordered** series of classes, the cumulative probabilities for each successive class represent the proportion of observations that have occurred or would be expected to occur according to the Poisson distribution up to that point. Figure 7.2 illustrates how the two cumulative

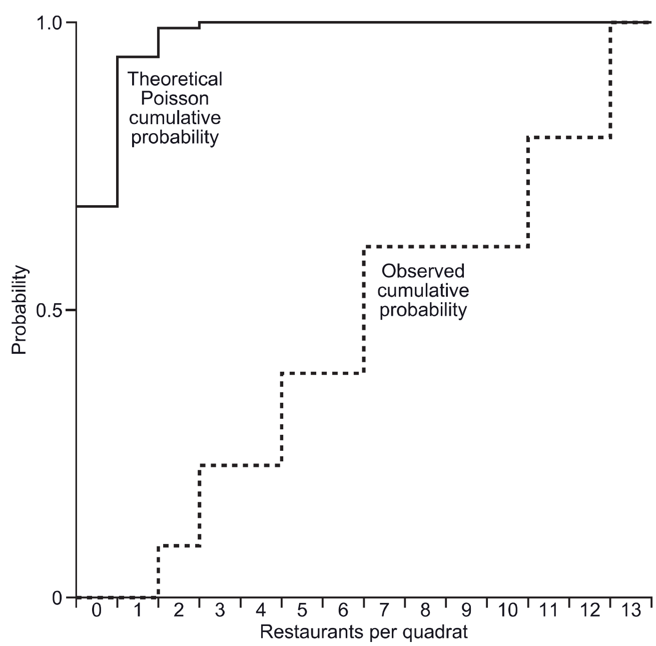


Figure 7.2 Comparison of observed and Poisson expected cumulative probability distributions.

probabilities increment in a stepwise fashion and that the maximum vertical difference between them indicates the goodness of fit between the two distributions. Application of the Kolmogorov–Smirnov is these circumstances is not universally accepted as legitimate by statisticians, since calculation of the Poisson expected frequencies relies on using the mean or average density of the sampled observations, which contravenes the expectation that they are obtained independently of the sample data. Despite these reservations, provided that a sufficiently rigorous level of significance is applied and caution is exercised in marginal cases, then the D test can prove to be a useful way of testing a ranked frequency distribution.

Box 7.3 illustrates use of the test to examine whether the distance travelled by households in four villages in mid-Wales to the settlement where they do their main shopping is similar to what might occur if the process of choosing where to go was random or otherwise. The distances have been classified into seven groups with the

Box 7.3a: The Kolmogorov–Smirnov D test.

Kolmogorov–Smirnov D test statistic: = MAX Abs(Cum P_{Obs} – Cum P_{Exp})

Box 7.3b: Application of the Kolmogorov–Smirnov D test.

The starting point for the Kolmogorov–Smirnov test is that any difference between the cumulative observed and expected frequency distributions of an ordered (ranked) attribute or classified variable for a single sample of observations is a consequence of sampling error. Application of the test proceeds by calculating the difference between the cumulative observed (Cum P_{Obs}) and expected (Cum P_{Exp}) probabilities across the ordered sequence of classes and the largest difference provides the D test statistic. The probability of having obtained the particular value of D, which varies according to sample size, n, helps with deciding on whether to accept or reject the Null Hypothesis at a chosen level of significance.

The Kolmogorov–Smirnov test has been applied to the distance travelled to the main settlement for shopping by households in the group of mid-Wales villages, with the original continuous variable classified into bands (groups) 10 km wide. Having tabulated the survey data into the observed frequency counts (f_{Obs}), the observed and expected probabilities are required. Box 5.7 showed how the Poisson probability distribution could be applied to produce such probabilities and has been used in this case. The observed probabilities are obtained by dividing the count of households in each class by the total (e.g. $64/35 = 0.55$). There are seven classes and 64 households producing an overall mean density of households per ranked class (λ) of 0.11 which has been inserted into the Poisson distribution equation to produce the expected probabilities. Each set of probabilities is accumulated across the classes and taking the absolute difference between these cumulative probabilities yields a column of differences. The D statistic can be identified as the maximum difference, which is 0.77 in this case. Since the probability of having obtained a maximum difference of this size is 0.17, which is larger than the standard significance level (0.05), the Null Hypothesis is accepted.

The key stages in carrying out a Kolmogorov–Smirnov test are:

State Null Hypothesis and significance level: sampling error accounts for the difference between the cumulative observed and expected frequencies and a difference this size is likely to have occurred more often than 95 times in 100. The observed ordered frequency distribution from the sample of observations is not significantly different from what would be expected according to the Poisson probability distribution

Calculate test statistic (D): the tabulated ordered frequency distributions and calculations are given below showing the maximum difference is 0.77.

Determine the probability of the calculated D: the probability of the test statistic varies according to sample size and in this case with a sample of 64 observations (households) it is 0.17.

Accept or reject the Null Hypothesis: the probability of D is >0.05, so the Null Hypothesis should be accepted with the consequence that the Alternative Hypothesis is rejected.

Box 7.3c: Assumptions of the Kolmogorov–Smirnov (D) test.

There is one main assumption of the Kolmogorov–Smirnov test:

Observations should be selected by means of simple random sampling and be independent of each other.

Box 7.3d: Calculation of Kolmogorov–Smirnov D statistic.

Distance (km)	f_{Obs}	P_{Obs}	P_{Exp}	Cum P_{Obs}	Cum P_{Exp}	Abs (Cum P_{Obs} – Cum P_{Exp})
0.0–0.5	8	0.13	0.90	0.13	0.90	**0.77**
0.5–10.4	35	0.56	0.10	0.67	0.99	0.32
10.5–20.4	17	0.27	0.01	0.94	1.00	0.06
20.5–30.4	2	0.03	0.00	0.97	1.00	0.03
30.5–40.4	0	0.00	0.00	0.97	1.00	0.03
40.5–50.4	1	0.02	0.00	0.98	1.00	0.02
50.5–60.4	1	0.02	0.00	1.00	1.00	0.00
	64	1.00	1.00			

Kolmogorov–Smirnov statistic: $D = \text{MAX Abs(Cum } P_{Obs} - \text{Cum } P_{Exp}) = 0.77$

Probability: $p = 0.17$

first containing the count of households travelling less than 0.5 km, regarded for this purpose as zero up to the seventh with just one household that journeyed in the range 50.5–60.4 km. The test results in this case indicate there is a moderate probability of obtaining *D* equal to 0.77 with 64 observations and so, despite some statisticians' reservations about the procedure, it may be considered reasonably reliable in this case.

The Kolmogorov–Smirnov test is used with an ordered series of classes in a frequency distribution, in other words the observations have been allocated to a set of integer classes but the test pays no attention to the order in which their outcomes occurred. This characteristic can be examined by the **Runs (or Wald–Walfowitz) test,** which focuses on the sequence of individual dichotomous outcomes through time and across space and, as the name suggests, whether 'runs' of the same outcome occur or if they are jumbled up. The test can be used to examine whether the outcomes for a sample of observations can be regarded as independent of each other. The test is founded on the notion that the outcomes of a certain set of events can only be ordered in a finite series of combinations, since the number of events is not open-ended. If the number of runs observed in a particular sample exceeds or falls short of what might be expected then the independence of the observations is brought into question. A run is defined as either one or an uninterrupted series of identical outcomes. Suppose an event involves tossing a coin four times in succession to find out how many combinations or sequences of outcomes there are that include a total of two heads and two tails: the answer is that there are six:

HHTT; HTHT; HTTH; TTHH; THTH; and THHT

How many runs of the same outcome are there in each of these sequences?

Box 7.4 applies the Runs test to a rather more geographical problem. Box 7.4b shows a 250 m stretch of road in a residential area along which there a series of 50 segments. If there was at least one car or other vehicle parked at the kerb in a 5 m segment it was counted as occupied, otherwise it was unoccupied. There were respectively 19 and 31 occupied and unoccupied segments on the day when the data were recorded, indicating that 38% had vehicles parked in 25 runs. Driving along the road from one segment to the next would mean that a new run started when going from a segment with parked vehicles to one without or vice versa, whereas adjacent segments with the same outcome would be part of the same run. The Runs test is applied in this case to determine whether there is sufficient evidence to conclude that the number of runs along the stretch of road on this occasion is random. The test seeks to establish whether the observed number of runs is more or less than the expected mean given that the overall sequence of dichotomous outcomes is finite. The comparison is made using the Z statistic and its associated probability distribution. Figure 7.3 relates to the example in Box 7.4 where there are a total 50 outcomes in the sequence and illustrates the distinctive shape of probability distribution. There are two areas where the combinations of dichotomous outcomes (e.g. total segments with and without

Box 7.4a: The Runs (Wald–Walfowitz) Test.

R = number of changes of dichotomous outcomes in complete sequence

Conversion of R to Z test statistic: $= \frac{R - \mu_R}{\sqrt{\sigma_R^2}}$

Box 7.4b: Application of the Runs (Wald–Walfowitz) Test.

The Runs test focuses on the sequence of individual dichotomous data values and compares the observed number of runs (R) with the average (μ_R) that would be expected if changes in the sequence of outcomes had occurred randomly. The expected mean (μ_R) and variance (σ_R^2) are calculated empirically from the sample data and are used to convert the observed number of runs into a Z statistic. The Null Hypothesis is evaluated by comparing the probability of Z in relation to the chosen level of significance, typically 0.05.

Cars and other vehicles parked at kerbs along roads going through residential areas are commonplace: from the motorists' perspective they represent an obstacle impeding and slowing progress, for residents they constitute a convenient parking place and for pedestrians they present a potential hazard when attempting to cross a road. If we consider any given stretch of road capable of being divided into segments of equal length (say 5 m), the Runs test can be used to examine the hypothesis that the presence or absence of one or more vehicles in a sequence of segments are mutually independent of each other. The Runs test has been applied to test such a hypothesis in respect of the stretch of road shown above together with a photo to illustrate kerbside parking. The sequence of segments occupied and unoccupied by cars or other vehicles is tabulated in Box 7.4d. The 25 runs were of various lengths ranging from single segments up to a run of five adjacent segments without any parked vehicles. The expected number of runs was 24.56 and the variance 10.85, which produce a Z statistic with high probability and so the Null Hypothesis of a random number of runs is upheld. So, it would seem that the outcome in each segment of the road is independent of the others.

The key stages in performing a runs test and calculating the associated Z statistic are:

State Null Hypothesis and significance level: the number of runs is no more or less than would be expected to have occurred randomly and the sequence of outcomes indicates the observations are independent of each other at the 0.05 level of significance.

Determine the number of runs, R: the sequence of road segments with and with parked vehicles is given below and the alternate shaded and unshaded sections indicate there are 25 runs.

Calculate Z statistic: the Z statistic requires that the mean and variance of the sequence be obtained before calculating the test statistic itself, which equals 0.134 in this case.

Determine the probability of the Z statistic): this probability is 0.897.

Accept or reject the Null Hypothesis: the probability of *Z* is much greater than 0.05 and so the Null Hypothesis should be accepted.

Box 7.4c: Assumptions of the Runs test.

There is one assumption of the Runs test:

Observations are assumed independent of each other (the test examines whether this is likely to be the case).

Box 7.4d: Calculation of the Runs test.

Segment	1	2	3	4	5	6	7	8	9	10	11	12
Car	NC	NC	NC	NC	C	C	NC	NC	C	C	NC	C
Runs	1	1	1	1	2	2	3	3	4	4	5	6
Seg. (cont.)	13	14	15	16	17	18	19	20	21	22	23	24
Car (cont.)	NC	C	NC	NC	C	NC	NC	NC	C	C	C	C
Runs (cont.)	7	8	9	9	10	11	11	11	12	12	12	12
Seg. (cont.)	25	26	27	28	29	30	31	32	33	34	35	36
Car (cont.)	NC	NC	C	NC	C	NC	NC	NC	NC	C	C	NC
Runs (cont.)	13	13	14	15	16	17	17	17	17	18	18	19
Seg. (cont.)	37	38	39	40	41	42	43	44	45	46	47	48
Car (cont.)	NC	NC	NC	C	NC	NC	NC	NC	NC	C	C	NC
Runs (cont.)	19	19	19	20	21	21	21	21	21	22	22	23
Seg. (cont.)	49	50										
Car (cont.)	C	NC		Segs. with cars $N_C = 19$						Segs. without cars $N_{NC} = 31$		
Runs (cont.)	24	25		Total runs $R = 25$								

Mean

$$\mu_R = \left(\frac{2(N_C \times N_{NC})}{N}\right) + 1 = \left(\frac{2(19 \times 31)}{50}\right) + 1 = 24.56$$

Variance

$$\sigma_R^2 = \frac{(2(N_C N_{NC})(2(N_C N_{NC} - N_C - N_{NC}))}{(N)^2(N-1)} = \frac{2(1178(1178-19-31))}{250(49)} = 10.85$$

Z

$$Z = \frac{R-\mu_R}{\sqrt{\sigma_R^2}} = \frac{25-24.56}{\sqrt{10.85}} = 0.134$$

Probability

$p = 0.897$

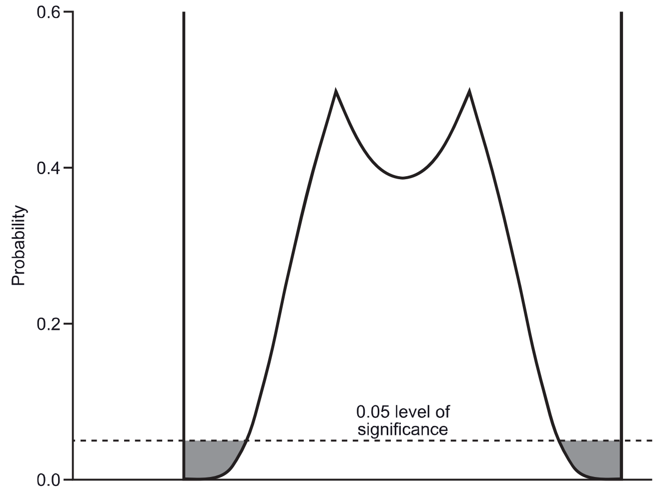


Figure 7.3 Probability distribution associated with the Runs test.

parked vehicles) fall below the 0.05 significance level (9 and 41; 10 and 40; 11 and 39; 12 and 38; and 13 and 37), where the Null Hypothesis would be rejected. The example of 19 and 31 segments in the two types of outcome is not one of these and so the Null Hypothesis is accepted. It is worth noting before moving on from the Runs test that the sequence of occupied and unoccupied road segments in this example may display **spatial autocorrelation**, a tendency for similar data values to occur close together in space, a topic that we will look at in Chapter 10. However, for the time being it is useful to consider the idea that in a sequence of 50 dichotomous outcomes a combination of 19 and 31 runs of the two types could produce a number of different spatial patterns in addition to the one shown in Box 7.4b.

7.2.4 Analysing patterns of spatial entities

Chapter 6 included an introduction to spatial pattern analysis and explored neighbourhood techniques that use measurements of the distances between observations. Quadrat or grid-based techniques offer another approach to analysing such patterns, particularly in respect of phenomena represented as points. These involve overlaying the entire set of points with a regular, but not necessarily rectangular, grid and counting the number of entities falling within each of its cells. In certain respects, this is analogous to crosstabulating a set of observations in respect of two attributes or categorized variables, and assigning them to a set of cells. The main difference is that the grid lines represent spatial boundaries and thus summarize the overall pattern of the points' location. The frequency distribution of the number of point entities per grid cell can be assessed in relation to what would be expected according to the Poisson probability distribution (see Chapter 5). In general, if the observed and expected frequency distributions are sufficiently similar, then it might be concluded that the entities are randomly distributed.

A useful feature of the Poisson distribution is that its mean and variance are equal, which enables us to formulate a Null Hypothesis stating that if the observed pattern of points is random then according to the Poisson distribution the **Variance Mean Ratio (VMR)** will equal 1.0. In contrast, a statistically significant difference from 1.0 suggests some nonrandom process at least partially accounts for the observed pattern, which must necessarily tend towards either clustering or dispersion. Chapter 6 introduced the idea that any particular set of points, even if it includes all known instances of the entity, should be considered as but one of a potentially infinite number of such sets of the phenomena in question: it may therefore be regarded as a sample. A finite number of points distributed across a fixed number of grid cells will have a mean that does not vary (λ) irrespective of whether they are clustered, random or dispersed. The three grids in Figure 7.4 represent situations where a set of 18 points distributed across nine cells is either clustered, random or dispersed and there is an overall average of two points per cell in each case. The nature of the pattern is therefore indicated by differences in the variances of the point patterns. A variance to mean ratio significantly greater than 1.0 suggests clustering and conversely less than 1.0 uniformity or dispersion. Since any one set of points constitutes a sample, the VMR should be tested for its significance, which can be achieved by conversion into a t or χ^2 test statistic.

We have already come across the illustration in Box 7.5b, which shows a nine-cell square grid overlain on the pattern of Burger King's restaurants in Pittsburgh. The numbers in the quadrats represent the observed frequency count of these restaurants, which seem to display some evidence of clustering in the pattern of their locations. The question is whether such visual evidence is supported by spatial statistics. The outcome of applying the Variance Mean Ratio procedure in Box 7.5 suggests that the pattern is not significantly different from what might be expected to occur if a random process was operating. However, there remains an issue over whether this result is an artefact of the number of quadrats and where the grid is located. The size of the

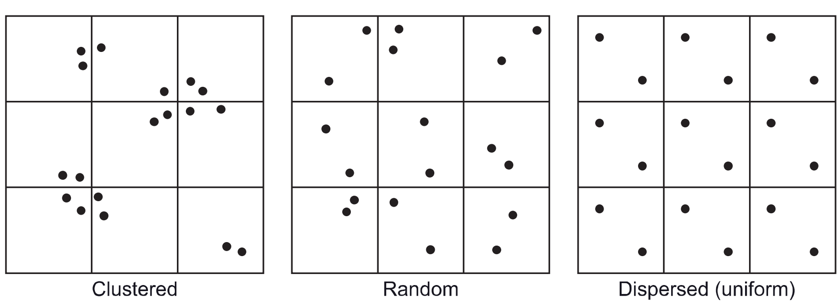


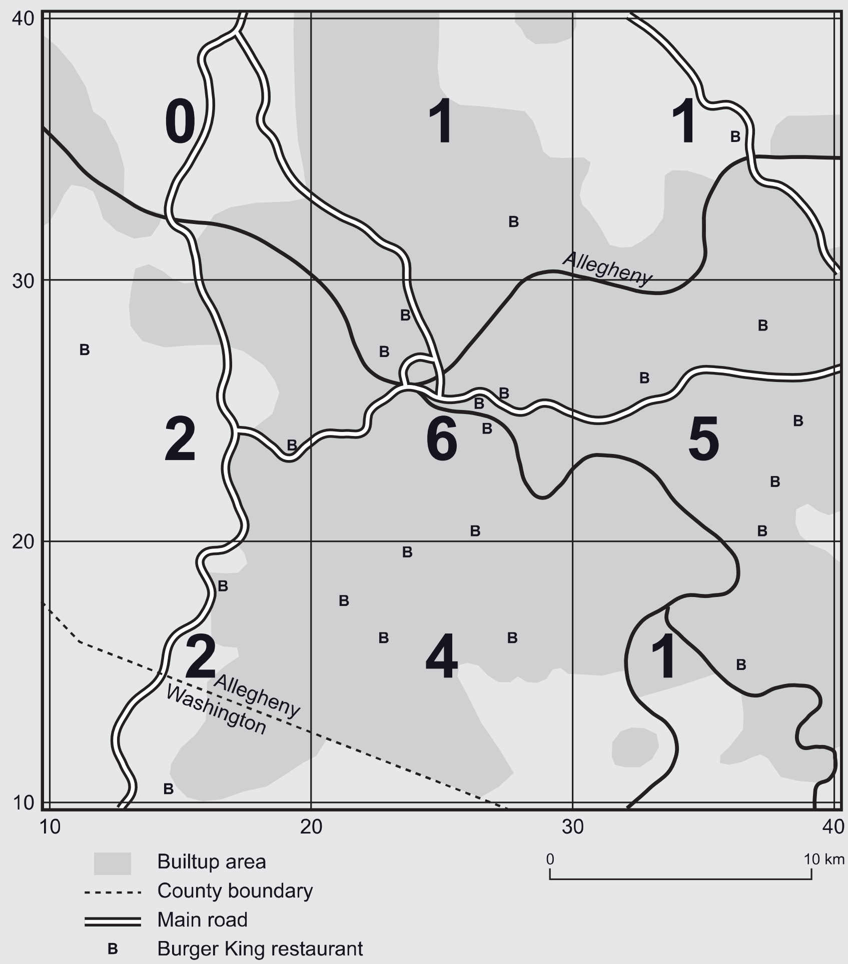
Figure 7.4 Clustered, random and dispersed spatial distributions of point features.

Box 7.5a: The Variance to Mean Ratio.

$$\mathrm{VMR} = \frac{s^2}{\lambda} = s^2 = \frac{1}{n-1}\left(\sum f_i x_i^{\,2} - \frac{\left(\sum f_i x_i\right)^2}{n}\right) : \lambda = \frac{n}{n_q}$$

Conversion of VMR to t test statistic: $= \dfrac{\mathrm{VMR}-1}{s_{\mathrm{VMR}}}$

Box 7.5b: Application and Testing of the Variance to Mean Ratio.



A regular grid of quadrats is overlain on a map showing the location of the point features and a count of points per quadrat is made. However, before the Variance to Mean Ratio (VMR) can be computed, it is necessary to calculate the variance and mean of the points in the quadrats. Calculating the mean or average (λ) number of points per quadrat is relatively easy, it is the number of points (n) divided by the total quadrats (k). The equation for calculating the variance is not unlike the one used to estimate the variance for a nonspatial frequency distribution in Chapter 5. The number of points per quadrat in the variance formula is denoted by x in each of i frequency classes going from 0, 1, 2, 3 … i and the frequency count of quadrats in each class is f. There are two alternative test statistics that can be calculated in order to examine the significance of the VMR – the t statistic and the continuous, χ^2, (not Pearson's Chi-square) in both cases with $k-1$ degrees of freedom. The probabilities associated with these can then be used in the usual way to decide whether to accept or reject the Null Hypothesis.

A regular grid containing nine cells or quadrats has been superimposed over the map of Pittsburgh showing the location of the Burger King restaurants and a count of restaurants per quadrat has been made. The observed counts of restaurants in the centre and centre-right quadrats are relatively high (6 and 5, respectively) whereas the remainder are less well served. But is this any different from what would be expected if the restaurants were randomly distributed? Or are they, as it might appear, mostly located in the main built-up area of the city? The mean number of restaurants per quadrat is 2.444 (22/9) and the variance 1.630 resulting in a VMR equal to 0.667, which suggests the Burger King restaurants tend towards being dispersed in a fairly uniform pattern. Calculating both types of test statistic with respect to this VMR value and their associated probabilities and with 8 degrees of freedom indicates that we really cannot conclude that this marginally dispersed pattern is any more likely than would be expected if a random locational process was operating.

The key stages in applying the variance to mean ratio and associated t and χ^2 tests are:

State Null Hypothesis and significance level: the restaurants are distributed at random in the city and any suggestion from the VMR ratio that they are either clustered or dispersed (uniform) is the result of sampling error and its value is not significant at the 0.05 level.

Calculate VMR ratio: the tabulated ordered frequency distribution of Burger King's restaurants per quadrat and calculations to obtain the variance and mean, prerequisites of the VMR are given below (VMR = 0.667).

Calculate t and/or χ^2 test statistic(s): the t statistic requires that the standard error of the mean be obtained before calculating the test statistic itself, which equals −1.333 in this case. The χ^2 statistic is simply VMR multiplied by the degrees of freedom and here equals 5.336.

Determine the probability of t and/or χ^2 test statistic(s): these probabilities are, respectively, 0.220 and 0.721 with 8 degrees of freedom.

Accept or reject the Null Hypothesis: both probabilities are such as to clearly indicate that the Null Hypothesis should be accepted at the 0.05 level of significance.

Box 7.5c: Calculation of Variance to Mean Ratio and associated t statistic.

Number of Burger King restaurants (x)	Frequency count (f)	fx	fx^2
0	1	0	0
1	3	3	3
2	2	4	8
3	0	0	0
4	1	4	16
5	1	5	25
6	1	6	36
	9	22	88

Sample variance $s^2 = \frac{1}{n-1}\left(\sum f_i x_i^{\,2} - \frac{\left(\sum f_i x_i\right)^2}{k}\right) = \frac{1}{22-1}\left(88 - \frac{(22)^2}{9}\right) = 1.630$

Sample mean $\lambda = \frac{n}{k} = \frac{22}{9} = 2.444$

Variance to Mean Ratio $\text{VMR} = \frac{s^2}{\lambda} = \frac{1.630}{2.444} = 0.667$

Standard error of the mean $\text{VMR}_s = \frac{2^2}{(k-1)} = \frac{2^2}{(9-1)} = 0.500$

t statistic $t = \frac{\text{VMR} - 1}{s_{\text{VMR}}} = \frac{0.667 - 1}{0.5} = -1.333$

Chi-square statistic $\chi^2 = (k-1)\text{VMR}$ $(9-1)0.667 = 5.336$

Probability t statistic $p = 0.220$ Chi-square statistic $p = 0.721$

quadrats, 10 km, has been determined in an arbitrary fashion, as have the orientation and origin of the grid. A grid with smaller or larger quadrats and/or rotated clockwise 45° is likely to have produced entirely different observed and therefore expected frequency distributions, although the underlying location of the points (restaurants) is unaffected. Help with deciding quadrat size is provided by the rule-of-thumb that these should be in the range D to $2D$, where D is the study area divided by the number of points. In this example there are 22 points in a 900 km^2 study area $D = 40.9$), which suggests the quadrats should have been in the range 40.9 to 8.18 km^2: the cells in the grid in Box 7.4b are slightly larger than this upper guideline at 100 km^2.

An alternative grid-based way of examining the spatial pattern of a set of point observations to see if they are randomly distributed across a set of quadrats is by carrying out a Pearson's Chi-square test using the Poisson distribution to generate the

Box 7.6: Calculation of Pearson's Chi-square statistic.

Number of Burger King restaurants per grid square (x)	Frequency count O	Expected frequency count E	$(O - E)$	$\frac{(O-E)^2}{E}$
0	1	0.78	−0.78	0.78
1	3	1.91	1.09	0.62
2	2	2.33	1.67	1.19
3	0	1.90	−1.90	1.90
4	1	1.16	2.84	6.93
5	1	0.57	4.43	34.58
6	1	0.23	5.77	143.80
	9	8.89		189.81

Pearson's chi-square statistic $\chi^2 = 189.81$
Probability $p < 0.000$

expected frequencies. Thus, the expected frequencies reflect what would have occurred if a random Poisson process was at work. The Pearson's Chi-square test for a univariate frequency distribution has already been examined and Box 7.6 simply illustrates the calculations involved with applying the procedure to the counts of Burger King's restaurants in Pittsburgh. The observed and expected frequency counts relate to the quadrats with different numbers of restaurants (i.e. 0, 1, 2, 3, 4, 5 and 6) not to the numbers of restaurants themselves. So, with a total of 22 restaurants and nine quadrats the average number per cell is 2.444. Although this example deals with fewer points than would normally be the case, it is realistic to assume that the probability of an observation occurring at a specific location is small. There are two factors reducing the degrees of freedom in this type of application of the test, not only is the total count known but also calculating the expected frequencies using the Poisson distribution employs the sample mean. Thus, the degrees of freedom are the number of classes minus 2 and equal 5 in this case. With a Null Hypothesis that the observed distribution of points is random and referring to the results shown in Box 7.6, we can see that there is an extremely low probability (<0.000) of having obtained a χ^2 test statistic as high as 189.81 with 5 degrees of freedom by chance and so the Alternative Hypothesis can be accepted that restaurants are not randomly distributed.

7.3 Two samples and one (or more) variable(s)

Many investigations are concerned with examining the differences and similarities between two populations through separate random or stratified random samples rather than the more limited case of focusing on a single set of observations. Our exploration of nonparametric tests now moves on to such situations as samples drawn from populations of arable and dairy farms, of acidic and alkaline soils, of sedimentary and igneous rocks, and of coastal and inland cities. The majority of questions to be examined in these situations involve nominal or ordinal frequency counts rather than measurements for continuous variables and are essentially concerned with deciding the fate of Null Hypotheses that claim the observations from the two samples are randomly distributed amongst the categories of an attribute or classified variable. Such tests seek to establish if the samples are genuinely from two significantly different populations or whether they should really be viewed as just one. The tests work by comparing the observed frequency counts with those that would be expected if they were distributed randomly or proportionately between the cells of a crosstabulation.

Some of the procedures examined in this section can also be applied in situations where there is only one sample. In these circumstances, rather than concentrating on the differences between two samples, the objective is to examine the distribution of observations between two attributes or categorized variables. For example, the data values in a set of water samples collected at different locations along a river might be crosstabulated in respect of an attribute recording the land use at the sample point and the acidity/alkalinity of the water with the aim of investigating any association between these characteristics. Most investigations using the nonparametric procedures examined here will carry out a series of such analyses in order to explore the data values in the sample(s) so as to reach conclusions about the research questions.

7.3.1 Comparing two attributes for one sample (or one attribute for two or more samples)

Section 7.2.2 outlined the theoretical background to Pearson's Chi-square test and illustrated how it could be applied to a univariate frequency distribution. The same test can be used in two further ways:

- to examine whether a sample of observations crosstabulated on two (or more) attributes or classified variables are independent or associated with each other;
- to discover whether the frequency distributions of observations from two (or more) samples in respect of the same attribute or categorized variable are significantly different from each other.

When used in the second situation Pearson's Chi-square test is the nonparametric equivalent of one-way analysis of variance (see Chapter 6). Although conceptually different the observations in each situation are normally displayed for the purpose of statistical analysis in the form of a crosstabulation comprising rows (r) and columns (*c*), which define the size of the table in terms of total cells. The simplest such table is a two-by-two crosstabulation where the total observations have been distributed across four cells. There is no reason why the two dimensions of a table must be equal and it is entirely feasible for the observations in a sample to be tabulated in respect of one attribute with three categories and another with four to produce a three-by-four table with 12 cells.

How many cells would there be in a crosstabulation of five samples by an attribute with three categories?

The total count of observations in the cells of a table is finite in any particular dataset and so the more cells there are the greater the chance that some will have few or no observations. In many cases the number of cells is also finite, being determined by the number of attribute/variable categories and/or samples. However, there are occasions when it is justifiable to introduce some flexibility by combining or collapsing the categories of an attribute or variable in order to increase counts in at least some of the cells. Farms in the parishes stretching across the South Downs in East and West Sussex have been allocated to two sets of size classes in Table 7.2 according to their land areas as recorded in the 1941 Agricultural Census. The first classification has 13 groups with eight of these referring to farms under 100 ha, whereas the second is more general with only two classes below this size threshold. The total number of farms in each classification is the same and so the

Table 7.2 The effect of adjusting attribute or variable categories on frequency distribution counts.

Detailed classification	East Sussex	West Sussex	Total	General classification	East Sussex	West Sussex	Total
0.0–1.9	3	1	4	0.00–49.9	174	177	351
2.0–4.9	54	54	108				
5.0–9.9	39	38	77				
10.0–19.9	30	27	57				
20.0–29.9	23	23	46				
30.0–39.9	17	18	35				
40.0–49.9	8	16	24				
50.0–99.9	40	42	82	50–99.9	40	42	82
100.0–199.9	48	49	97	100.0–299.9	61	68	129
200.0–299.9	13	19	32				
300.0–499.9	4	17	21	300.0 and over	8	20	28
500.0–699.9	3	1	4				
700.0 and over	1	2	3				
Total	283	307	590	Total	283	307	590

Source: MAFF Agricultural Statistics.

low counts in some cells have been overcome by collapsing some of them together. The interaction between the number of rows and columns and the total count of observations not only defines the number of cells in a table, but also constrains how the bivariate distribution is free to vary, in other words they control the degrees of freedom. Referring to Table 7.2, only one cell is free to vary within each column, because total farms in the South Downs parishes in the two counties are fixed. However, when distributing the observations in a single sample across the intersecting rows and columns defined by two categorized attributes/variables, the degrees of freedom are defined as $(r - 1)(c - 1)$.

How many degrees of freedom are there in this crosstabulation?		A	B	C	D	E
	1					
	2					
	3					

The data collected for investigations in geography, Earth and environmental sciences together with many of the social sciences will often includes variables that are unsuited to analysis by parametric procedures. Pearson's Chi-square test can offer a flexible and conceptually straightforward alternative means of addressing research questions in these subject areas. Pearson's Chi-square test is broadly concerned with testing a Null Hypothesis that the distribution of the observations in respect of two attributes or two or more samples and one attribute is random and they are unconnected with each other. In other words, observations are independently allocated to the cells in the crosstabulation. Thus, in the case of a two-by-two table, the fact that an observation is assigned to the top row has no influence on whether it is assigned to left or right column. Pearson's Chi-square test is reasonably robust, although since the probabilities associated with the test statistic are influenced by sample size its use is generally not recommended if there are fewer than 30 observations. However, even this number is problematic if the dimensions of the table result in more than 20 per cent of cells having an expected count of less than 5. Box 7.7 illustrates how to apply Pearson's Chi-square test when there are two or more samples and one attribute using data from the survey of households in four settlements in mid-Wales. The application examines whether people in the four settlements use different modes of transport to carry out their major shopping trip, or equivalently whether different types of transport are favoured by residents in the sampled settlements.

Box 7.7a: The Bivariate Pearson's Chi-square (χ^2) test.

Pearson's Chi-square (χ^2) test statistic: $= \sum_{j}^{1} \frac{(O_j - E_j)^2}{E_j}$

Box 7.7b: Application of the Bivariate Pearson's Chi-square (χ^2) test.

The Pearson's Chi-square tests focuses on the overall difference between the observed (O) and expected (E) frequencies. The test statistic formula calculates the sum of the squared differences between the observed and expected counts divided by the expected for each cell. The first step in carrying out the test is to distribute the observations into a crosstabulation and then to calculate the expected frequencies. Calculating the expected frequencies for any individual cell involves multiplying the corresponding marginal row and column totals by the total number of observations (see the upper part of Box 7.7d) for each cell in the crosstabulation. There is an observed count of 8 households in the case of the top left cell (travel by family car and Barmouth sample) and the expected value is 7.2 $\left(\frac{(36 \times 20)}{100}\right)$: the difference is very small (0.8) in this cell, which contrasts with the lower-right cell where the difference is 4.8. The second step involves squaring the difference in each cell and dividing by the expected count (lower part of Box 7.7d) and then summing these to obtain χ^2. The probability of obtaining the χ^2 value varies according to the degrees of freedom, which equals 9 in this example. A decision on the Null Hypothesis is reached by comparing this probability with the chosen level of significance. If the probability of having obtained the particular test value is equal to or less than 0.05 (or another selected significance level), the Null Hypothesis should be rejected.

The samples of households in the four settlements in mid-Wales have been tested with Pearson's Chi-square to see if there is a significant difference in the mode of transport used by households when doing their main shopping trip. The χ^2 statistic is 10.440, which has a probability of 0.316 with 9 degrees of freedom. This implies that the Null Hypothesis should be accepted since for a four-by-four crosstabulation (i.e. 9 degrees of freedom) the test statistic value would have had to reach at least 16.619 for a significant difference at the 0.05 level. The test points to the conclusion that households in the four settlements tend to be divided between the different modes of transport in a random fashion.

Application of Pearson's Chi-square test has six main stages:

Declare the Null Hypothesis and significance level: any differences between the observed and expected frequencies are a consequence of sampling error and they are not significant at the 0.05 level.

Compute the expected frequencies (E): these usually come from dividing the product of the marginal totals corresponding to each cell by the overall total number of observations.

Calculate test statistic (χ^2): this is the sum of the squared differences between O and E divided by E.

Determine the degrees of freedom: in this example $df = (4 - 1)(4 - 1)$ and so there are 9 degrees of freedom.

Determine the probability of the calculated χ^2: the test statistic equals 10.440, which has a probability of 0.316 with 9 degrees of freedom.

Accept or reject the Null Hypothesis: the Null Hypothesis should be accepted since 0.316 is >0.05 and would be obtained in approximately 32 samples out of a 100, if households in the four settlements were sampled on that number of occasions.

Box 7.7c: Assumptions of the Pearson's Chi-square (χ^2) test.

There are five main assumptions of the Pearson's Chi-square test:

Observations selected for inclusion in the sample(s) should be selected by simple or stratified random sample and be independent of each other.

Adjustment of attribute/variable classes may be carried out, although care should be taken to ensure the categories are still meaningful.

The total count of observations in the four samples from the mid-Wales settlements happen to equal 100 and the cell counts are percentages of the overall total. However, the test should not be applied to data recorded as relative frequencies (percentages or proportions) and these should be converted back to absolute counts before carrying out a Pearson's Chi-square test.

Low expected frequencies, usually taken to mean <5 counts in no more than 20% of cells, are likely to mean that the χ^2 test statistic is unreliable. The problem may be overcome by redefining attribute/variable classes.

The test should not be applied if any of the attributes or grouped variables are ordered or ranked at all, since the calculated test statistic will be the same regardless of the ranking.

7.3.2 Comparing the medians of an ordinal variable for one or two samples

The basis of the **Mann–Whitney U test** is that a set of ordered data values for a variable can be partitioned into two groups. This can arise in two ways:

- there are two samples that have been measured in respect of an ordinal variable;
- there is one sample containing a dichotomous attribute, which may be spatial or temporal, enabling the observations to be split into two groups across an ordinal variable.

The test works by pooling the ordinal measurements across the complete set of observations and then compares the medians for each group or sample. The test establishes whether the quantitative difference between the medians is small enough simply to be the result of sampling error. If the test is applied to variables that were originally collected on the interval or ratio scales of measurement, then some loss of information is implied by sorting into ascending order then assigning rank scores. However, where the assumption of normality of the variables cannot realistically be believed, the Mann–Whitney test provides a robust alternative to the (unpaired) two sample t test. Box 7.8 applies the test to the sample of road segments with or without vehicles analysed previously using the Runs test.

Box 7.7d: Calculation of the Pearson's Chi-square (χ^2) test.

		Sample settlement				
Transport mode		Barmouth	Brecon	Newtown	Talgarth	Row totals
Family car	O	8	6	16	6	36
	E	$\frac{(36\times20)}{100}=7.2$	$\frac{(36\times26)}{100}=9.4$	$\frac{(36\times35)}{100}=12.6$	$\frac{(36\times19)}{100}=6.8$	36
Other car	O	4	3	6	5	18
	E	$\frac{(18\times20)}{100}=3.6$	$\frac{(18\times26)}{100}=4.7$	$\frac{(18\times35)}{100}=6.3$	$\frac{(18\times19)}{100}=3.4$	18
Public transport	O	3	5	8	2	18
	E	$\frac{(18\times20)}{100}=3.6$	$\frac{(18\times26)}{100}=4.7$	$\frac{(18\times35)}{100}=6.3$	$\frac{(18\times19)}{100}=3.4$	18
Walk or cycle	O	5	12	5	6	28
	E	$\frac{(28\times20)}{100}=5.6$	$\frac{(28\times26)}{100}=7.3$	$\frac{(28\times35)}{100}=9.8$	$\frac{(28\times19)}{100}=5.3$	28
Column totals		20	26	35	19	100
Family car		$\frac{(8-7.2)^2}{7.2}=0.09$	$\frac{(6-9.4)^2}{9.4}=1.21$	$\frac{(16-12.6)^2}{12.6}=0.92$	$\frac{(6-6.8)^2}{6.8}=0.10$	
Other car		$\frac{(4-3.6)^2}{3.6}=0.04$	$\frac{(3-4.7)^2}{4.7}=0.60$	$\frac{(6-6.3)^2}{6.3}=0.01$	$\frac{(5-3.4)^2}{3.4}=0.73$	
Public transport		$\frac{(3-3.6)^2}{3.6}=0.10$	$\frac{(5-4.7)^2}{4.7}=0.02$	$\frac{(8-6.3)^2}{6.3}=0.46$	$\frac{(2-3.4)^2}{3.4}=0.59$	
Walk or cycle		$\frac{(5-5.6)^2}{5.6}=0.06$	$\frac{(8-7.2)^2}{7.2}=0.09$	$\frac{(5-9.8)^2}{9.8}=02.35$	$\frac{(6-5.3)^2}{5.3}=0.09$	
Pearson's chi-square statistic				$\chi^2 = 10.440$		
Degrees of freedom				$df = (r-1)(c-1) = (4-1)(4-1) = 9$		
Probability				$p = 0.316$		

Box 7.8a: Mann–Whitney U test.

Mann–Whitney U test statistic: $= \sum r - \frac{n_r(n_r+1)}{2}$

Conversion of U to Z test statistic: $= \frac{|U - \mu_U|}{\sigma_U}$

Box 7.8b: Application of Mann–Whitney U test.

The Mann–Whitney U test is commonly used to examine the difference between the medians of an ordinal variable for two samples or two groups of observations from one sample. The objective in both instances is to investigate whether the difference is of such a size that it is likely to have arisen from sampling error, in other words that the true difference is zero. In general terms, the Null Hypothesis is that the rank scores and therefore the medians of observations in the two samples or groups from a sample are such that they balance each other and so they really originate from one undifferentiated population in respect of the ordinal variable being examined. The calculations for the test statistic are relatively straightforward. First, the observations are sorted in ascending order in respect of the variable and their ranks (R) are assigned (smallest value assigned rank 1, second rank 2 and so on). The second step involves sorting the observations' rank scores according to the sample or group to which they belong. The rank scores are summed for these two groups of observations and the smaller total is designated $\sum r$. The value of this quantity is inserted into the equation for the test statistic together with n_r – the number of observations in the group that provides $\sum r$.

The test statistic's probability is determined in different ways depending on the number of observations in each group, although using statistical software to apply the test usually removes the need to choose between the methods. The probability distribution of U becomes approximately Normal when there are more than 20 observations in each sample or group. The probability of the test statistic is obtained by transforming U into Z having calculated a population mean and standard deviation of U given the known, fixed number of observations in each sample or group (see calculations below). The fate of the Null and Alternative Hypotheses is decided by reference to the specified significance level, typically whether ≤ 0.05.

The Mann–Whitney test has been applied to the 250 m stretch of road divided into 5 m segments used in Box 7.4. The distance from the start of the sequence of segments increments by 5 m when progressing from one segment to the next and they are categorized into two groups by means of the dichotomous attribute denoting whether or not there are parked vehicles on at least one side of the road: these are denoted by the subscripts C and NC (car and no car), respectively. The raw data (distance measurements and rank scores) are given below. The Null Hypothesis is that the median distance of segments in the two groups from the start of the stretch of road is the same and that sampling error accounts for the observed difference. Although strictly speaking there are just too few observations in the group with parked vehicles ($n_C = 19$), the U test statistic has been converted to Z to illustrate the procedure. The Z value 0.290 has a probability of 0.772, which clearly results in the Null Hypothesis being accepted at the 0.05 significance level.

Application of the Mann–Whitney test has seven main stages:

State the Null Hypothesis and significance level: any difference two groups of road segments is the result of sampling error and it is not significant at the 0.05 level.

Sort and assign ranks (R): since the segments are numbered from the start of the length of road and the distances increment in 5 m units from this point, the sorted rank scores are the same as the segment identification numbers. Sort the segments again according to dichotomous grouping attribute or sample membership.

Sum the rank scores of each group to identify $\sum r$*:* the results of these additions are shown below.

Calculate test statistic (U): insert $\sum r$ into the equation and calculate the test statistic, which equals 470.

Determine the probability of the calculated U: in this example, since the number of observations in each group is very nearly >20, U as been converted into Z (0.290), which has a probability of 0.772.

Select one- or two-tailed probability: if there is any prior, legitimate reason for arguing that one group of segments be nearer to the start of the road than the other, apply a one-tailed test by halving the probability associated with Z before comparing with the significance level.

Accept or reject the Null Hypothesis: the Null Hypothesis should be accepted since there is an extremely high probability of having obtained the test statistic by chance; division of the road segments according to the presence or absence of parked vehicles seems to be random along the entire length of the road (0.772 > 0.050).

Box 7.8c: Assumptions of the Mann–Whitney U test.

There are three main assumptions of the Mann–Whitney U test:

Observations should be selected by means of simple random sampling.

The observations should be independent of each other in respect of the ordinal, ranked variable. (Note: it is arguable whether this is the case in this example.)

The test is robust with small and large sample sizes, and is almost as powerful as the two sample t test with large numbers of observations.

Box 7.8d: Calculation of the Mann–Whitney test.

Segment	Vehicle parked	Distance (x) in metres	Overall rank (R)	Segment ID sorted	Sorted rank
1	N	5	1	1	1
2	N	10	2	2	2
3	N	15	3	3	3
4	N	20	4	4	4
5	Y	25	5	7	7
6	Y	30	6	8	8
7	N	35	7	11	11
8	N	40	8	13	13
9	Y	45	9	15	15
10	Y	50	10	16	16
11	N	55	11	18	18
12	Y	60	12	19	19
13	N	65	13	20	20
14	Y	70	14	25	25
15	N	75	15	26	26
16	N	80	16	28	28
17	Y	85	17	30	30
18	N	90	18	31	31
19	N	95	19	32	32
20	N	100	20	33	33
21	Y	105	21	36	36
22	Y	110	22	37	37
23	Y	115	23	38	38
24	Y	120	24	39	39
25	N	125	25	41	41
26	N	130	26	42	42
27	Y	135	27	43	43
28	N	140	28	44	44
29	Y	145	29	45	45
30	N	150	30	48	48
31	N	155	31	50	50
32	N	160	32	5	5
33	N	165	33	6	6
34	Y	170	34	9	9
35	Y	175	35	10	10
36	N	180	36	12	12
37	N	185	37	14	14
38	N	190	38	17	17
39	N	195	39	21	21
40	Y	200	40	22	22
41	N	205	41	23	23
42	N	210	42	24	24

Segment	Vehicle parked	Distance (x) in metres	Overall rank (R)	Segment ID sorted	Sorted rank
43	N	215	43	27	27
44	N	220	44	29	29
45	N	225	45	34	34
46	Y	230	46	35	35
47	Y	235	47	40	40
48	N	240	48	46	46
49	Y	245	49	47	47
50	N	250	50	49	49
Sum of the ranks of segments with parked vehicles					$\sum R_C = 470$
Sum of the ranks of segments without parked vehicles					$\sum R_{NC} = 805$
Mann–Whitney U statistic		$U = \sum r - \frac{n_r(n_r+1)}{2} = 470 - \frac{19(19+1)}{2} = 280$			
Population mean U		$\mu_U = \frac{n_C n_{NC}}{2} = \frac{19(31)}{2} = 294.50$			
Standard deviation U		$\sigma_U = \sqrt{\frac{n_C n_{NC}(n_C + n_{NC} + 1)}{12}} = \sqrt{\frac{19 \times 31(19+31+1)}{12}} = 50.03$			
Z test statistic		$Z = \frac{\lvert U - \mu_U \rvert}{\sigma_U} = \frac{\lvert 280 - 294.50 \rvert}{50.03} = 0.290$			
Probability		$p = 0.772$			

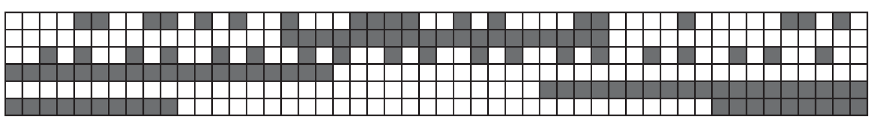


Figure 7.5 Combinations of road segments in groups with 19 and 31 observations.

The results of the Mann–Whitney test in Box 7.8 support the previous conclusion that the occurrence of the two types of segments is not significantly different from what might be expected if they were to have occurred randomly along the length of the road. However, there are many different ways in which 50 observations, in this case road segments, could be allocated to two groups containing 19 and 31 cases. The top row of Figure 7.5 reproduces the situation shown in Box 7.4, while the others present five rather more distinctive patterns. The clearest difference in medians occurs in the fourth row where all the segments without parked vehicles are at the start and the probability of such extreme differentiation occurring by chance is very small. Even the arrangement in rows 2, 5 and 6 with all the occupied segments clumped in one or two groups have relatively high probabilities of occurring randomly (>0.750). The regular, almost systematic spacing of vehicles along the stretch of road represented in row 3 is the one most likely to occur through a random process ($p = 0.928$). These examples help to reveal that the results of the test are dependent not only on the degree to which the ordinal variable separates observations between the two groups or samples, but also on the size of the groups. Rows 4 and 5 as just as differentiated as each other, but in one case the occupied segments are at the start and in the other at the end.

7.4 Multiple samples and/or multiple variables

The procedures examined in this section progress our exploration of statistical testing to those situations where there are two or more samples and two or more attributes, or one sample and three or more attribute or classified variables. These nonparametric techniques are less demanding in terms of the assumptions made about the underlying distribution of the data values than their parametric counterparts, but, nevertheless, they can provide important insights into whether the observed data are significantly different from what might have occurred if random processes were operating. The data values are either measurements recorded on the ordinal scale (i.e. by placing the observations in rank order, but not quantifying the size of the differences between them) or categorical attributes or classified variables that allow the observations to be tabulated as counts. The crosstabulations of data considered here have at least three dimensions. For example, a single sample of households might be categorized in respect of their housing tenure, the economic status of the oldest member of the household and whether or not anyone in the household had an overseas holiday during the previous 12 months. Supposing there were four tenure categories, six types of economic status and by definition the holiday attribute is dichotomous, then there would be 48 cells in the crosstabulation ($4 \times 6 \times 2$). Such a table is difficult to visualize and would become even more so with additional dimensions. Nevertheless, some statistical tests exist to cope with these situations. The Pearson's Chi-square test examined in Section 7.3.1 can be extended in this way and the Kruskal–Wallis H test is an extension of the Mann–Whitney test.

7.4.1 Comparing three attributes for one sample (or two attributes for two or more samples)

Extension of the Pearson's Chi-square test to those situations where the crosstabulation of counts has three or more dimensions can occur in two situations where there is (are):

- one sample and three (or possibly more) categorized attributes;
- three or more samples and two (or more) categorical variables.

Both situations are represented in the three-dimensional crosstabulation shown in Table 7.3. Attributes 1 and 2 are categorized into five (I, II, III, IV and V) and three (1, 2 and 3) classes, respectively. The third dimension is shown as jointly representing either three samples or three classes a third attribute (A, B, and C). Overall, Table 7.3 represents a five-by-three-by-three crosstabulation. The counts in the crosstabulation are for illustrative purposes only and each sample (or group for attribute 3) conveniently contains 100 observations, thus the individual cells and the marginal totals in each part of the table sum to 100. The calculation of Pearson's Chi-square test statistic proceeds in the same way in this situation as with a two-dimensional crosstabulation. Once the expected frequencies have been computed, the squared difference between the observed and expected count divided by the expected $\left(\frac{(O-E)^2}{E}\right)$ is calculated for each cell. The sum of these gives the test statistic. The degrees of freedom are defined as $(X_1-1)(X_2-1)\ \ldots\ (X_j-1)$, where X represents the number of categories in attributes 1 to j. In the case of the crosstabulation in Table 7.3 there are 16 degrees of freedom $(5-1)(3-1)(3-1)$. The probability associated with the test statistic with these degrees of freedom enables a decision to be reached regarding the Null Hypothesis. Since most multidimensional crosstabulations will have more cells than the simpler two dimensional ones, there is a greater likelihood that low expected frequencies might occur. It is therefore important to ensure that sample sizes are sufficiently large to reduce the chance of this problem occurring during the planning stage of an investigation.

Table 7.3 Three-dimensional crosstabulation with three categorized attributed or three samples with two categorized attributes.

			Attribute 1					
Attribute 3, group A	Attribute 2		I	II	III	IV	V	Total
OR		1	15	8	4	3	5	35
Sample A		2	7	6	2	5	0	18
		3	6	10	4	11	13	44
		Total	28	24	10	19	18	100
			Attribute 1					
Attribute 3, group B	Attribute 2		I	II	III	IV	V	Total
OR		1	6	5	3	10	4	28
Sample B		2	5	7	11	9	4	36
		3	9	2	1	14	10	36
		Total	20	14	15	35	18	100
			Attribute 1					
Attribute 3, group C	Attribute 2		I	II	III	IV	V	Total
OR		1	11	6	10	15	0	42
Sample C		2	7	7	3	5	6	28
		3	7	3	0	4	16	30
		Total	25	16	13	24	22	100

7.4.2 Comparing the medians of an ordinal variable for three or more samples (or one sample separated into three or more groups)

The **Kruskal–Wallis H** test represents an extension or generalization of the Mann–Whitney test that enables rather more complex types of question to be investigated. The test can be used in two situations where the focus of attention is on differences in the ranks of an ordinal variable in respect of:

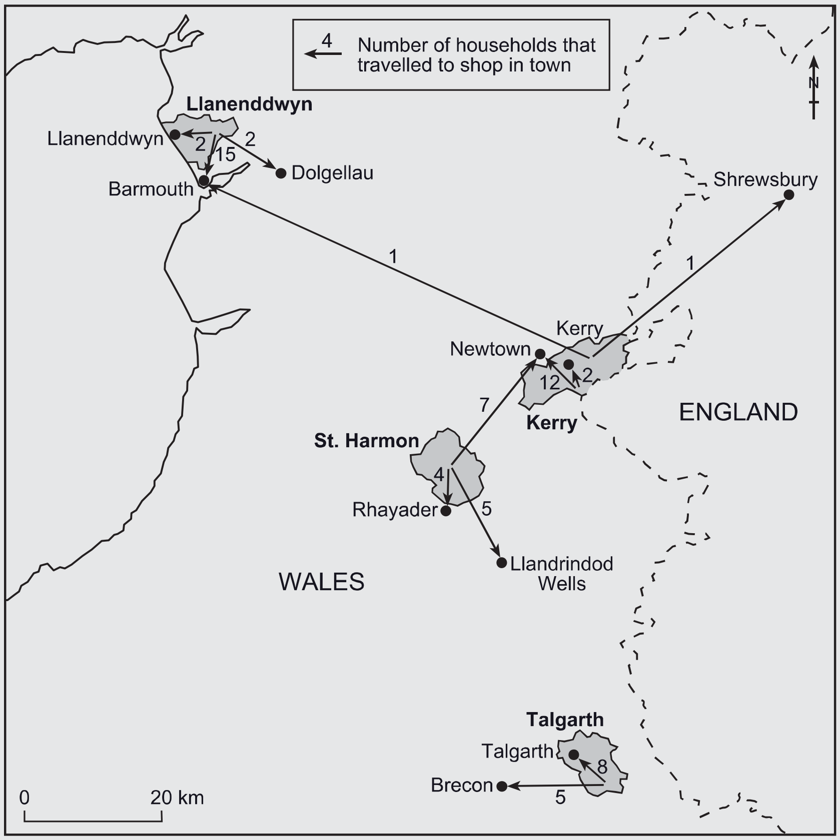
- one sample and three (or possibly more) ordered or ranked variables;
- three or more samples and one ordered variable.

The H test is essentially the nonparametric equivalent of the one-way ANOVA procedure examined in Section 6.3.1. One-way ANOVA explores whether the differences in the means of a continuous variable for three or more samples are statistically significant, the Kruskal–Wallis test performs an equivalent task in respect of differences in the medians. The ordinal (or sorted continuous) variable provides a unifying thread to each of these situations. The data values for this variable are pooled, just as with the Mann–Whitney test, across the three or more samples or groups from a single sample and rank scores are assigned with average ranks used for any tied observations. The general form of the Alternative and Null Hypotheses for the Kruskal–Wallis test, respectively, assert that the difference in medians between the three or more samples or groups from a sample in respect of the ordinal variable under investigation either are or are not significant. Just as with the Mann–Whitney test, the Kruskal–Wallis test explores the extent to which the complete set of observations is interleaved or interspersed with regard to their data values on the ordinal variable(s).

The Kruskal–Wallis test has been applied to the samples of households in mid-Wales and the length of their major shopping journeys to investigate whether there are differences between the populations of households in the four sampled settlements. Box 7.9 examines this application of the test procedure and reaches the conclusion that there is a significant difference between the median length of the journeys (in km) made by households in the four settlements where the sample surveys were carried out. Examination of the sample data reveals that the median distance for Talgarth households is greater than the other three settlements, however, from the perspective of the test results it cannot be concluded that this necessarily indicates that households in this settlement stand out from the rest. Just as the one-way ANOVA technique focused on the overall difference in the means of three or more samples in respect of a ratio or interval scale variable, so the Kruskal–Wallis test examines the overall, not specific, differences in medians between the samples.

Box 7.9a: Kruskal–Wallis (H) test.

Kruskal–Wallis H test statistic: $= \dfrac{12\sum N(\bar{R}_i - \bar{R})^2}{N(N+1)}$



Box 7.9b: Application of Kruskal–Wallis (H) test.

The Kruskal–Wallis test is used in two ways, either where one sample has been subdivided into three or more groups or where there are three or more samples. In both cases there is an ordinal variable that has been measured across the groups or samples. The calculations arising from the equation for Kruskal–Wallis test statistic (H) are relatively simple summation and division, although performing these by hand can be tedious, especially if the numbers of observations in the samples or groups is fairly large, because of the amount of data sorting required. The complete set of observations are pooled and sorted in respect of the ordinal variable, and then ranks are determined. Ties are dealt with by inserting the average rank, thus four observations with the same value for the ordinal variable in the 4^{th}, 5^{th}, 6^{th} and 7^{th} positions would each be assigned a rank score of 5.5. The sums of the rank scores for each sample or group are computed and designated scores R_1, R_2, R_3 … R_i where i denotes the number of samples. The means of the rank sum for each sample or group are calculated as

$\overline{R}_1$, $\overline{R}_2$, $\overline{R}_3$, … $\overline{R}_i$ with the number of observations in each represented as n_1, n_2, n_3, … n_i: both of these sets of values are inserted in the equation used to calculate H.

Examination of the probability associated with the calculated H test statistic in conjunction with the chosen level of significance indicates whether the Null Hypothesis should be accepted or rejected. Provided that there are more than 5 observations in each of the samples or groups, the probability distribution of the test statistic approximates Chi-square with $k - 1$, where k corresponds to the number of samples or groups. The Kruskal–Wallis test has been applied to the four samples of households in mid-Wales settlements to examine whether any differences in the median distance travelled to carry out their main shopping trip are significantly greater than might have occurred randomly or by chance.

Application of the Kruskal–Wallis test can be separated into six main steps:

State the Null Hypothesis and significance level: any differences between the observations in the samples (or groups from one sample) have arisen as a result of sampling error and are no greater than might be expected to occur randomly at least 6 times in 100 (i.e. using a 0.05 significance level).

Sort and assign ranks (R): pool all the observations, sort in respect of the ordinal variable (distance to 'shopping town'), assign rank scores to each, averaging any tied values.

Sum the rank scores of each sample (group): the sum of the ranks for each sample or group are calculated and identified as R_1, R_2, R_3 … R_i, where the subscript distinguishes the groups.

Calculate test statistic (H) and apply correction factor if >25 per cent of observations are tied: calculations using the equation for the test statistic are given below, which equals 21.90 in this example. Corrected H equals 31.43.

Determine the probability of the calculated H: the number of observations in each sample is >5 and so the probability of having obtained the test statistic by chance can be determined from the Chi-square distribution with 3 degrees of freedom and is <0.001.

Accept or reject the Null Hypothesis: the test statistic's probability is lower than the chosen significance level (0.05) and therefore the Null Hypothesis should be rejected. It is very unlikely that the difference in medians between the samples (groups) has arisen by chance.

Box 7.9c: Assumptions of the Kruskal–Wallis (H) test.

There are three main assumptions of the Kruskal–Wallis test:

Observations should be selected by means of simple random sampling.

Sampled entities should be independent of each other and selected from populations of observations that have approximately the same shape in respect of the ordinal variable.

If the number of ties is large (more than 25 per cent of the ordinal values), then the calculated H statistic should be divided by a correction factor $C = 1 - \frac{(T^3 - T)}{n_s^3 - n_s}$, where T and n_s are, respectively, the number of tied and total number of observations.

Box 7.9d: Calculation of the Kruskal–Wallis test.

Settlement	Distance (km)	Rank (R)	Settlement (sorted)	Rank (R) (sorted)
Kerry	0.36	3	Kerry	3
Kerry	0.36	3	Kerry	3
Kerry	0.36	3	Kerry	3
Kerry	0.36	3	Kerry	3
Kerry	0.36	3	Kerry	3
Kerry	0.42	7	Kerry	7
Kerry	0.42	7	Kerry	7
Kerry	0.42	7	Kerry	7
St Harmon	0.45	9	Kerry	44
Llanenddwyn	0.58	10	Kerry	46
Llanenddwyn	0.99	11	Kerry	46
St Harmon	1.30	12	Kerry	46
St Harmon	1.92	13.5	Kerry	48
St Harmon	1.92	13.5	Llanenddwyn	10
St Harmon	3.67	16.5	Llanenddwyn	11
St Harmon	3.67	16.5	Llanenddwyn	26
St Harmon	3.67	16.5	Llanenddwyn	27
St Harmon	3.67	16.5	Llanenddwyn	29
St Harmon	4.43	20	Llanenddwyn	29
St Harmon	4.43	20	Llanenddwyn	29
St Harmon	4.43	20	Llanenddwyn	32
St Harmon	4.62	22.5	Llanenddwyn	32
St Harmon	4.62	22.5	Llanenddwyn	32
Talgarth	5.36	24	Llanenddwyn	35
Talgarth	5.41	25	Llanenddwyn	35
Llanenddwyn	7.41	26	Llanenddwyn	35
Llanenddwyn	7.68	27	Llanenddwyn	39
Llanenddwyn	7.74	29	Llanenddwyn	40
Llanenddwyn	7.74	29	Llanenddwyn	41.5

Settlement	Distance (km)	Rank (*R*)	Settlement (sorted)	Rank (*R*) (sorted)
Llanenddwyn	7.74	29	Llanenddwyn	41.5
Llanenddwyn	7.75	32	Llanenddwyn	54
Llanenddwyn	7.75	32	Llanenddwyn	55
Llanenddwyn	7.75	32	St Harmon	9
Llanenddwyn	8.07	35	St Harmon	12
Llanenddwyn	8.07	35	St Harmon	13.5
Llanenddwyn	8.07	35	St Harmon	13.5
Talgarth	8.42	37.5	St Harmon	16.5
Talgarth	8.42	37.5	St Harmon	16.5
Llanenddwyn	8.45	39	St Harmon	16.5
Llanenddwyn	8.60	40	St Harmon	16.5
Llanenddwyn	8.80	41.5	St Harmon	20
Llanenddwyn	8.80	41.5	St Harmon	20
St Harmon	9.11	43	St Harmon	20
Kerry	11.55	44	St Harmon	22.5
Kerry	11.96	46	St Harmon	22.5
Kerry	11.96	46	St Harmon	43
Kerry	11.96	46	St Harmon	63
Kerry	12.41	48	St Harmon	64
Talgarth	13.21	49	Talgarth	24
Talgarth	13.63	50	Talgarth	25
Talgarth	13.96	52	Talgarth	37.5
Talgarth	13.96	52	Talgarth	37.5
Talgarth	13.96	52	Talgarth	49
Llanenddwyn	14.89	54	Talgarth	50
Llanenddwyn	15.33	55	Talgarth	52
Talgarth	19.68	58	Talgarth	52
Talgarth	19.68	58	Talgarth	52

Talgarth	19.68	58	Talgarth	58
Talgarth	19.68	58	Talgarth	58
Talgarth	19.68	58	Talgarth	58
Talgarth	21.61	61	Talgarth	58
Talgarth	21.80	62	Talgarth	58
St Harmon	41.10	63	Talgarth	61
St Harmon	55.05	64	Talgarth	62

	Kerry	Llanenddwyn	St Harmon	Talgarth
Sum of the ranks $\sum R_K, \sum R_L, \sum R_{SH}, \sum R_T$	266	633	389	792
Number of households n_K, n_L, n_{SH}, n_T	13	19	16	16
Mean of the ranks $\bar{R}_K, \bar{R}_L, \bar{R}_{SH}, \bar{R}_{T_i}$	20.5	33.3	24.3	49.5

Kruskal–Wallis H statistic

$$H = \frac{12\sum n_i(\bar{R}_i - \bar{R})^2}{N(N+1)} = \frac{12(13(20.5-32.5)^2 + 13(33.3-32.5)^2 + 13(24.3-32.5)^2 + 13(49.5-32.5)^2}{64(64+1)} = 21.90$$

Correction factor

$$C = 1 - \frac{(T^3 - T)}{n_s^3 - n_s} = 1 - \frac{(75\,907 - 43)}{262\,144 - 64} = 0.697$$

Corrected H

$$H_{Cor} = \frac{H}{C} = \frac{21.90}{0.697} = 31.43$$

Degrees of freedom

$$df = k - 1 = 4 - 1 = 3$$

Probability

$$p < 0.001$$

7.5 Closing comments

This chapter has examined a set of nonparametric statistical tests that parallel the parametric ones covered in Chapter 6. The main differences between the two groups of techniques relate to whether the data values for the variable(s) support the assumption of normality and whether they are continuous measurements on the ratio or interval scales in the case of parametric procedures. These techniques are concerned with testing numerical quantities, such as the mean, variance and standard deviation with respect to a population or between samples. The equivalent nonparametric tests either focus on ordinal measurements, in which case the main concern is with the median, or on frequency counts. Techniques concerned with frequencies generally look at whether the difference between observed and expected counts is more than might have occurred by chance through sampling error. Sometimes, these tests are used to compare samples and populations, although the role of the population data is generally to calculate the expected frequencies. For example, if an earlier population census had recorded certain proportions of households in five tenure categories, the expected frequencies for the sample of observations could be calculated using these.

Grid-based techniques provide another way of examining the pattern in location of spatial phenomena, especially those that can be represented as points. The process of counting points in a set of quadrats is similar to that of distributing nonspatial observations to set of cells in a crosstabulation. However, the key difference is that the distribution of points captures their pattern in two-dimensional geographical space, whereas the categories or classes used with a nonspatial crosstabulation by definition do not represent the pattern in location of the observations. The next chapter moves our exploration of statistical analysis in respect of spatial and nonspatial data into the realms of examining relationships between variables or attributes. However, this does not mean that we have finished with statistical testing, since the techniques used to calculate quantities to summarize the relationships between variables or attributes may well need to be tested to discover whether the value might have arisen by chance. Given that such quantities are often only produced using sampled data, it may be necessary to generate confidence limits within which the population parameter lies.

A new dataset was introduced in this chapter depicting a sequence of 5 m road segments with or without parked vehicles along a 250 m stretch of road. By way of an introduction to the topic correlation and its application in respect of spatial data examined in Chapter 8, it is worth thinking about whether the values held by each segment (i.e. vehicle or no vehicle) in this sequence are really independent of each other. Is it not possible that vehicles may well be parked outside residential properties and these are distributed with a certain regularity along the road? Are there less likely to be vehicles parked where there are parking restrictions, such as near a school or pedestrian crossing? What do these factors says about the possible lack of independence in the data?

Section 3
Forming relationships