CHAPTER FIVE

ANALYSIS OF DISCRETE VARIABLES

Discrete variables are those which can only assume certain fixed values. Examples include outcome variables with results such as live vs die, pass vs fail, and extubated vs reintubated. Analysis of data obtained from discrete variables requires the use of specific statistical tests which are different from those used to assess continuous variables (such as cardiac output, blood pressure, or PaO_2) which can assume an infinite range of values. The analysis of continuous variables is discussed in the next chapter.

The two statistical tests which are most commonly used to analyze discrete variables are the **chi-square test** (including the chi-square test with **Yates' correction**) and **Fisher's exact test**. Both of these tests are based on the use of **2 x 2 contingency tables** (Figure 5-1) which classify patients as either true positives, true negatives, false positives, or false negatives with regard to their disease status and test outcome.

	Disease Present	Disease Absent
Test Positive	**True Positive**	**False Positive**
Test Negative	**False Negative**	**True Negative**

Figure 5-1: 2 x 2 Contingency Table

To use these two tests, we must first carefully define the disease being studied as well as the criteria which constitute a positive test, assigning each patient to one of the four possible outcomes. Having created a 2 x 2 contingency table of these results, the appropriate statistical test can be performed calculating the **critical value** of the test which identifies whether a statistically significant difference exists between the two groups of patients. The significance level associated with this critical value (more commonly referred to as the **p-value**) can then be obtained from a chi-square distribution table to quantitate the significance of the difference between the two groups.

CHI-SQUARE

The **chi-square test** is a statistical method for determining the approximate probability of whether the results of an experiment may arise by chance or not. The test is performed by first creating a 2 x 2 contingency table of the observed disease and test outcome frequencies.

	Disease	No Disease	
Test Positive	**a**	**b**	**(a + b)**
Test Negative	**c**	**d**	**(c + d)**
	(a + c)	**(b + d)**	**n**

where: a = true positives, b = false positives, c = false negatives, d = true negatives, n = total patients

If the null hypothesis is true (the test does not discriminate between patients with the disease and patients without the disease), we would expect the disease frequencies to be equally distributed based on the probabilities of a positive and a negative test result. Since the frequency of an event is given by the probability of the event multiplied by the number of events, the expected frequency of diseased patients with a positive test result (i.e., true positives or the frequency in cell "a") is:

expected true positives = probability of disease x probability of a positive test x n

Mathematically, this can be expressed as:

$$\text{expected true positives} = \frac{(a+c)}{n} \times \frac{(a+b)}{n} \times n$$

The <u>expected</u> frequencies for cells b, c, and d (i.e., false positives, false negatives, and true negatives respectively) can be calculated similarly. The chi-square (X^2) test compares the <u>observed</u> (O) frequencies (the actual patient data) with the <u>expected</u> (E) frequencies (those which are expected based on the probability of the disease) and determines how likely it is that their difference (O-E) occurred by chance. This results in the formula below

$$X^2 = \frac{(O_a - E_a)^2}{E_a} + \frac{(O_b - E_b)^2}{E_b} + \frac{(O_c - E_c)^2}{E_c} + \frac{(O_d - E_d)^2}{E_d}$$

which calculates the **critical value** for chi-square. If the critical value obtained is small, the <u>observed</u> frequencies are not very different from the <u>expected</u> frequencies and the two groups are likely to be similar. If the critical value is large, the observed and expected frequencies are very different and the probability that the two groups are different from one another is real and is not likely due to chance alone.

In actual usage, the chi-square test is calculated using the following approximation:

$$X^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

As previously stated, the critical value is that value of the test which must be obtained in order that the two groups can be considered significantly different. Chi-square has a known distribution from which the critical value for any significance level and contingency table can be obtained. Tables of critical values for commonly used significance levels can be found in most statistics books using one **degree of freedom** (df) or can be calculated for a particular critical value by a computer statistics package to obtain the exact level of significance. The critical value of chi-square for a significance level (or p-value) of 0.05, for example, is 3.84. The null hypothesis for the chi-square test is that there is no difference in test results between patients with and without the disease. Thus, if the critical value of chi-square is less than 3.84, we would accept the null hypothesis and state that the test does not discriminate between patients with and without the disease. If the value of chi-square obtained from the above equation is greater than 3.84, we would accept the alternate hypothesis and state that the test identifies patients with the disease at a statistically higher rate than those without the disease (with a 5% chance of having committed a Type I error). If we wished to use the smaller significance level of 0.01 instead of 0.05, for example, the critical value of chi-square would increase to 6.64.The critical values of chi-square for the most commonly used significance levels (using one degree of freedom) are listed below:

Critical values of chi-square (df=1)

Significance level	0.10	0.05	0.01	0.001
Critical value	2.70	3.84	6.64	10.83

Degrees of freedom are determined by sample size and are defined as the number of observations (n) minus 1. They arise from the fact that if a particular statistic is known, only n - 1 of the observations are free to vary if the statistic is to remain the same. For example, if we make 5 observations and calculate their mean, we are free to change the value of only 4 of the 5 observations as once we have done so, we will automatically know the value of the 5th observation if the mean is to remain the same. Contingency tables represent a special situation in which the degrees of freedom are given by: df = (rows - 1)(columns - 1). For a 2 x 2 table this results in df = (2-1)(2-1) = 1.

Consider a study in which we wish to evaluate a particular set of extubation criteria (the test) in predicting successful extubation from mechanical ventilation (the disease). Suppose we studied 123 patients and noted whether they passed or failed our extubation criteria and whether they remained extubated or required reintubation. We might find that 105 patients were successfully extubated while 18 patients required reintubation. Of these 123 patients, 72 patients passed our extubation criteria while 51 failed the criteria. We would set up the following 2 x 2 contingency table to analyze our data. Based on these test and disease outcomes, we would <u>expect</u> to see the frequencies listed below:

Expected Frequencies

	Extubated	Reintubated	
Pass criteria	**105 x $\frac{72}{123}$ = 61**	**18 x $\frac{72}{123}$ = 11**	**72**
Fail criteria	**105 x $\frac{51}{123}$ = 44**	**18 x $\frac{51}{123}$ = 7**	**51**
	105	**18**	**123**

Observed Frequencies

	Extubated	Reintubated	
Pass criteria	**66**	**6**	**72**
Fail criteria	**39**	**12**	**51**
	105	**18**	**123**

We would then use the chi-square test to compare our expected and observed frequencies to determine whether their difference is greater than that which we would expect to see by chance alone. Note that chi-square is a **one-tailed** test as we are only evaluating the difference in one direction (i.e., greater than by chance alone). We would define our study hypotheses in the following manner:

Null hypothesis: the criteria do not predict successful extubation
Alternate hypothesis: the criteria predict successful extubation

In practice, the expected frequencies are rarely calculated and the following equation is used to calculate the critical value of chi-square based on the observed frequencies:

$$X^2 = \frac{(123)(792-234)^2}{(72)(51)(105)(18)} = 5.52$$

Since 5.52 exceeds the critical value of 3.84 required for a significance level of 0.05, we would reject our null hypothesis and conclude that our criteria accurately predict successful extubation with a significance level of less than 0.05 (i.e., the probability that these results occurred by chance is less than 0.05 or 5%). Since 5.52 is less than 6.64 (the critical value for a significance of 0.01) our actual significance level is somewhere between 0.05 and 0.01. We could use a computer statistics package to calculate the exact value.

The standard chi-square test should be used only if the total number of observations (n) is greater than 40 and the expected frequency in each cell is at least 5. If n is between 20 and 40, and the expected frequency in each cell of the contingency table is at least 5, the **chi-square test with Yates' correction** should be used. Yates' correction takes into account the uncertainty introduced by small numbers of observations which might result in our concluding that a difference exists when it does not. It is a more conservative test which makes a Type I error less likely, but a Type II error more likely. If n is less than 20, or any of the expected frequencies are less than 5, the chi-square test, even with Yates' correction, is not appropriate and **Fisher's exact test** should be used. When the number of observations is small, therefore, the expected frequencies should be calculated to ensure that the appropriate statistical test is used. Calculation of the chi-square test with Yates' correction for n between 20 and 40 is:

$$X^2 = \frac{n\left[|ad-bc| - 0.5 \cdot n\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$

Suppose we wish to know whether patients who require reintubation are more likely to be older than 60 years of age. We might study 30 patients noting their age and whether they were successfully extubated or required reintubation. Our null hypothesis would be that increased age does not affect successful extubation while the alternate hypothesis would be that increased age does affect successful extubation. The expected and observed frequencies for our study might look like this:

Expected Frequencies

	Extubated	Reintubated	
Under 60	**20 x $\frac{15}{30}$ = 10**	**10 x $\frac{15}{30}$ = 5**	**15**
Over 60	**20 x $\frac{15}{30}$ = 10**	**10 x $\frac{15}{30}$ = 5**	**15**
	20	**10**	**30**

Observed Frequencies

	Extubated	Reintubated	
Under 60	**12**	**3**	**15**
Over 60	**8**	**7**	**15**
	20	**10**	**30**

Since our total n is only 30 (i.e., less than 40) and the expected frequency in each cell is at least 5, the chi-square test with Yates' correction is appropriate and is calculated as follows:

$$X^2 = \frac{\left[|84 - 24|(0.5)(30)\right]^2}{(15)(15)(20)(10)} = 1.35$$

As the critical value of chi-square does not exceed 3.84, we must accept the null hypothesis and conclude that age does not significantly affect successful extubation. In fact, the actual significance level associated with this critical value is 0.25. Note that there is a trend for patients over 60 years of age to require reintubation, but that the trend does not reach significance. It is possible that a true difference does exist, but that we have not studied enough patients yet to detect a significant difference and have committed a Type II error. The issue of adequate sample size will be addressed in Chapter Nine.

FISHER'S EXACT TEST

Whereas the chi-square test measures the approximate probability of an event's occurrence, **Fisher's exact test** calculates the exact probability of the observed frequencies in a 2 x 2 contingency table. Computationally, it can become quite involved, but it is easily calculated on most computers. It is most commonly used when the study population is small ($n < 20$) or when the expected frequency in one of the outcome groups is less than 5. It can, however, be used for any 2 x 2 contingency table regardless of the number of observations. Unlike chi-square, which by definition is one-tailed, Fisher's exact test can be calculated as both a one-tailed or a two-tailed test reflecting its ability to look at differences in both directions.

The probability (P) of observing the frequencies in a 2 x 2 contingency table using Fisher's exact test is given by:

$$P = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!\,a!\,b!\,c!\,d!}$$

In order to calculate the exact probability of an event's occurrence, however, we must also take into account the more extreme occurrences which, although more rare, would be even more likely to demonstrate a significant difference had they occurred. The exact probability is thus given by the sum of not only the probability of the observed frequencies, but also all of the more extreme occurrences. For example, if we use Fisher's exact test to calculate the probability of observing the frequencies from the previous example of successful extubation and patient age we obtain:

$$P_{obs} = \frac{(15)!\,(15)!\,(20)!\,(10)!}{30!\,12!\,3!\,8!\,7!} = 0.097$$

Taking into account the more extreme occurrences of reintubation which would give even more evidence for an effect of age on successful extubation, we obtain the following probabilities which we sum to obtain the exact probability of the observed frequencies:

	Extubated		**Reintubated**		
	Under 60	**Over 60**	**Under 60**	**Over 60**	**P**
Study Data	**12**	**3**	**8**	**7**	**0.097**
Extreme Occurrence	**13**	**2**	**7**	**8**	**0.022**
Extreme Occurrence	**14**	**1**	**6**	**9**	**0.0025**
Extreme Occurrence	**15**	**0**	**5**	**10**	**0.0001**
				sum =	**0.122**

Thus, the exact one-tailed probability of observing the study frequencies is 0.122 which is greater than the probability (significance level) of 0.05 which we would normally consider to indicate a significant difference. Fisher's exact test therefore confirms our conclusion that age does not affect successful extubation. Note that the exact probability calculated by Fisher's exact test is smaller than the approximate probability of 0.25 which was calculated using the chi-square test with Yates' correction (which tends to be conservative and is more likely to result in a Type II error).

If our alternate hypothesis had been "increased age either increases or decreases the incidence of successful extubation," we would have been asking a bidirectional question and would have needed a two-tailed test to appropriately answer our hypothesis. The chi-square test, by definition, is a one-tailed test and would therefore not have been appropriate. Fisher's exact test can be used as both a one- and two-tailed test. Some statisticians approximate the probability of a two-tailed Fisher's exact test by doubling the one-tailed probability. In the situation in which either the sum of the two rows or the sum of the two columns is the same, this is appropriate. When this is not the case, however, the calculation of the two-tailed Fisher's exact test becomes more involved and the reader is referred to Glantz (reference 8) for further details.

SUGGESTED READING

1. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. Baltimore: Williams and Wilkins, 1982:41-58.
2. Wassertheil-Smoller S. Biostatistics and epidemiology: a primer for health professionals. New York: Springer-Verlag, 1990:18-21.
3. Campbell MJ, Machin D. Medical statistics: a commonsense approach. New York: John Wiley and sons, 1990:134-137.
4. Dawson-Saunders B, Trapp RG. Basic and clinical biostatistics. Norwalk: Appleton and Lange, 1994:147-155.
5. Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. Arch Pathol Lab Med 1986;110:13-20.
6. Yates F. Tests of significance for 2 x 2 contingency tables. J R Statist. Soc A 1984;147:426-463.
7. O'Brien PC, Shampo MA. Statistics for clinicians: 8.comparing two proportions: the relative deviate test and chi-square equivalent. Mayo Clin Proc 1981;56:513-515.
8. Glantz SA. Primer of biostatistics (3rd Ed). New York: McGraw-Hill, 1991:144-148.