ERIC Notebook

Second Edition

Common Statistical Tests and Applications in Epidemiological Literature

Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

Any individual in the medical field will, at some point, encounter instances when epidemiological methods and statistics will be valuable tools in addressing research questions of interest.

Examples of such questions might include:

- Will treatment with a new anti-hypertensive drug significantly lower mean systolic blood pressure?
- Is a visit with a social worker, in addition to regular medical visits, associated with greater satisfaction of care for cancer patients as compared to those who only have regular medical visits?

There are a number of steps in evaluating data before actually addressing the above questions. These steps include description of your data as well as determining what the appropriate tests are for your data.

Description of data

The type of data one has determines the statistical procedures that are utilized. Data are typically described in a number of ways: by type, distribution, location and variation.

There are three different types of data: nominal, ordinal, and continuous data. Nominal data do not have an established order or rank and contain a finite number of values. Gender and race are examples of nominal data. Ordinal data have a limited number of values between which no other possible values exist. Number of children and stage of disease are good examples of ordinal data. It should be noted that ordinal data do not have to have evenly spaced values as occurs with continuous data, however, there is an implied underlying order. Since both ordinal and nominal data have a finite number of possible values, they are also referred to as discrete data. The last type of data is continuous data which are characterized by having an infinite number of evenly spaced values. Blood pressure and age fall into this category. It should be noted for data collection and analysis that continuous, ordinal, or nominal values can be grouped. Grouped data are often referred to as categorical. Possible categories might include: low, medium, high, or those representing a numerical range.

A second characteristic of data description, distribution, refers to the frequencies or probabilities with which values occur within our population. Discrete data are often represented graphically with bar graphs like the one below (Figure 1).

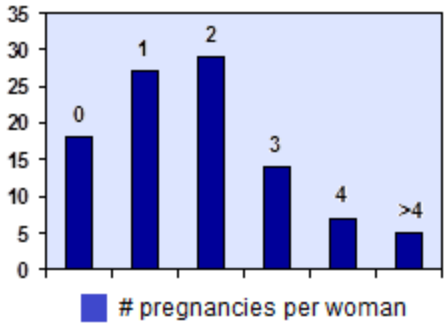

Figure 1. Bar graph

Continuous data are commonly assumed to have a symmetric, bell-shaped curve as shown below (Figure 2). This is known as a Gaussian distribution, the most commonly assumed distribution in statistical analysis.

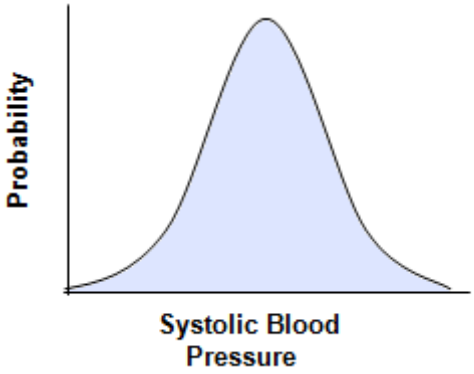

Figure 2. Gaussian distribution

Hypothesis testing

Hypothesis testing, also known as statistical inference or significance testing, involves testing a specified hypothesized condition for a population's parameter. This condition is best described as the null hypothesis. For example, in a clinical trial of a new anti-hypertensive drug, the null hypothesis would state that there is no difference in effect when comparing the new drug to the current standard treatment. Contrary to the null is the alternative hypothesis, which generally defines the possible values for a parameter of interest. For the previous example, the alternative hypothesis is that there is a difference in the mean blood pressure of the standard treatment and new drug group following therapy. The alternative hypothesis might also be described as your "best guess" as to what the values are.

However, in statistical analysis, the null hypothesis is the main interest, and is the one actually being tested. In statistical testing, we assume that the null hypothesis is correct and determine how likely we are to have obtained the sample (or values) we actually obtained in our study under the condition of the null. If we determine that the probability of obtaining the sample we observed is sufficiently small, then we can reject the null hypothesis. Since we are able to reject the null hypothesis, we have evidence that the alternative hypothesis may be true.

On the other hand, if the probability of obtaining our study results is not small, we fail to reject the assumption that the null hypothesis is true. It should be noted that we are not concluding that the null is true. This is a small, but important distinction. A test that fails to reject the null hypothesis should be considered inconclusive. An example will help to illustrate this point.

In a sealed bag, we have 100 blue marbles and 20 red marbles. (This bag is essentially representing the entire population). One individual formulates the null hypothesis that "all the marbles are blue", and the alternative which is "all the marbles are not blue". To test this hypothesis, 10 marbles are sampled from the bag. All ten marbles selected are indeed blue. Thus the individual has failed to reject the null that all the marbles in the bag are blue. However, because all of the marbles were not sampled, you cannot conclude that all the marbles in the bag are blue. (We happen to know this is not true, but it is impossible to know in the real world with populations too large to fully evaluate). If another individual selects 10 marbles from the bag and finds that 8 are blue and 2 are red, we can reject the null hypothesis that all the marbles are blue since we have selected at least one red marble.

Error in statistical testing

Earlier, we indicated that we can reject the null hypothesis if the probability of obtaining a sample like the one observed in our study is sufficiently small. You may ask "What is sufficiently small?" "How small" is determined by how willing we are to reject the null hypothesis when it accurately reflects the population from which it is sampled. This type of error is called a *Type I error*. This error is also commonly called alpha (α). Alpha is the probability of rejecting the null hypothesis when the null is true. This probability is selected by the researcher and is typically set at 0.05. It is important to remember that this is an arbitrary cut-point and should be taken into consideration when making conclusions about the results of the study.

There is a second type of error that can be made during statistical testing. It is known as *Type II error,* which is the probability of not rejecting the null when the alternative hypothesis is indeed true, or in other words, failing to reject the null when the null hypothesis is false. Type II error is commonly known as β. Beta relates to another important parameter in statistical testing which is *power*. Power is equal to (1-β) and is essentially the ability to avoid making a type II error. Like α, power is also defined by the researcher, and is typically set at 0.80. Below is a schematic of the relationships between α, β and power.

Decision	Truth	
	Null True	Null False
Reject Null	α	power
Accept Null		β

Students' T test

This test is most commonly used to test the difference between the means of the dependent variables of two groups. For example, this test would be appropriate if one wanted to evaluate whether or not a new anti-hypertensive drug reduces mean systolic blood pressure.

Example

To evaluate if drug Z reduces mean systolic blood pressure, a randomized clinical trial will be performed where 12 individuals receive drug Z and 8 receive a placebo. The null hypothesis to be tested is that there is no difference in the mean systolic blood pressure of the experimental and placebo groups. The alternative hypothesis is that there is a difference between the means of the two groups. The type I error for your trial will be 5%.

Results

Below is the group assignments and resulting systolic blood pressure (SBP)

Patient	Assignment	Systolic BP
1	Drug Z	100
3	Drug Z	110
5	Drug Z	122
7	Drug Z	109
9	Drug Z	108
11	Drug Z	111
13	Drug Z	118
15	Drug Z	105
17	Drug Z	115
18	Drug Z	119
19	Drug Z	106
20	Drug Z	109
2	Placebo	129
4	Placebo	125
6	Placebo	136
8	Placebo	129
10	Placebo	135
12	Placebo	134
14	Placebo	140
16	Placebo	128

$$\text{mean}_{drug} = \frac{100 + 110 + \ldots + 109}{12} = 111 \text{ mm Hg}$$

$$\text{mean}_{placebo} = \frac{129 + 125 + \ldots + 128}{8} = 132 \text{ mm Hg}$$

$$\text{mean}_{drug} - \text{mean}_{placebo} = -21 \text{ mm Hg}$$

Now that we have determined the difference between means, we need to determine the standard error for that difference which is calculated using the pooled estimate of the variance (σ^2).

The formula for the standard error of the drug Z group is:

$$\sigma^2 drug = \frac{\sum (SBP_{drug} - mean_{drug})^2}{n_{drug} - 1} =$$

$$\sigma^2 drug = \frac{[(100-111)^2 + (110-111)^2 + \ldots + (109-111)^2]}{12-1} = 40.9$$

The standard error for the placebo group is calculated in the same manner substituting the values for the placebo group.

$$\sigma^2 placebo = 25.1$$

Next, we would need to calculate a pooled estimate of the variance using the following equation:

$$\sigma^2_p = \frac{[(n_{drug} - 1)\ \sigma^2_{drug}] + (n_{placebo} - 1)\ \sigma^2_{placebo}]}{(n_{drug} - 1) + (n_{placebo} - 1)} =$$

$$\sigma^2_p = \frac{(11)(40.9) + (7)(25.1)}{11 + 7} = \frac{626}{18} = 34.8$$

The pooled estimate of the variance can then be utilized to calculate the standard error for the difference in means:

$$SE^2 (\text{mean}_{drug} - \text{mean}_{placebo}) = \frac{\sigma^2_p}{n_{drug}} + \frac{\sigma^2_p}{n_{placebo}}$$

$$SE^2 = \frac{34.8}{12} + \frac{34.8}{8} = 7.236$$

$$SE = 2.69$$

Now we are finally ready to test for significant differences in the mean blood pressure of our two groups: (*mean indicates the hypothesized values for the null-generally this quantity would = 0 when there is no difference expected between the drug and placebo groups).

$$t = \frac{(\text{mean}_{drug} - \text{mean}_{placebo}) - (*\text{mean}_{drug} - *\text{mean}_{placebo})}{SE\ (\text{mean}_{drug} - \text{mean}_{placebo})}$$

$$t = \frac{-21 - 0}{2.69} = -7.8 = |-7.8| = 7.8$$

We now compare our calculated value to a table of critical values for the Students' T distribution (found in most basic statistics books). The table also requires that we know the degrees of freedom and the value of a we have selected. Degrees of freedom (df) refers to the amount of information that a sample has in estimating the variance. It is generally the sample size minus one. The df for our calculation is 12 + 8 - 2 = 18 (the sample size for each group - 1). With a two tailed a of 0.05, our value |-7.8| is greater than the critical value from the table (2.101). Thus, we can reject the null hypothesis that there is no difference between mean blood pressure levels, and accept, by elimination, our alternative hypothesis.

Chi-square analysis

What happens if we don't have continuous data, and are faced with categorical data instead? We could turn to chi-square analysis to evaluate if there are significant associations between a given exposure and outcome (the row and column variables in a contingency table). 2 X 2 contingency tables are one of the most common ways to present categorical data, and we can see this in analyzing data that was collected to address the question presented in this notebook.

Is a visit with a social worker, in addition to regular medical visits, associated with greater satisfaction of care for cancer patients as compared to those who only have regular medical visits?

Below is a generic 2 X 2 table representing the data. It is important to note the set-up of the table, as cell "a" generally represents the group of interest (diseased and exposed) and cell d represents the referent group (no disease and unexposed).

	Row value (often disease or health outcome)		
Column Value (often Exposure)	1	0	**Total**
1	*a*	*b*	*a + b*
0	*c*	*d*	*c + d*
Total	*a + c*	*b + d*	*n*

Here we have the contingency table with data from our trial:

	Greater Satisfaction?		
Social Worker Visit?	Yes	No	**Total**
Yes	64	46	90
No	36	54	110
Total	100	100	200

In chi-square analysis we are testing the null hypothesis that there is no association between a social worker visit and a greater satisfaction with care.

Generally, in evaluating this type of data, it is important for each of the individual cells to have large values, (i.e. greater than 5 or 10 each), If these conditions are not met, a special type of chi-square analysis is conducted called the Fisher's exact test. This will not be discussed in this notebook.

To calculate the chi-square statistic (χ^2):

$$\chi^2 = \sum \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

with *i* representing the frequency in a particular cell of the 2 X 2 table. Below is the calculation for the frequencies that are **expected** in each cell.

	Row value		
Column Value	1	2	**Total**
1	$\frac{(a+b)(a+c)}{n}$	$\frac{(a+b)(b+d)}{n}$	$a + b$
2	$\frac{(c+d)(a+c)}{n}$	$\frac{(c+d)(b+c)}{n}$	$c + d$
Total	$a + c$	$b + d$	n

Thus, we now have a table that has both the actual and expected (in parentheses) values:

	Greater Satisfaction?		
Social Worker Visit?	Yes	No	**Total**
Yes	64 (55)	46 (55)	90
No	36 (45)	54 (45)	110
Total	100	100	200

With this information, we can now calculate the χ^2 statistic:

$$\chi^2 = \sum \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

$$\chi^2 = \frac{(64-55)^2}{55} + \frac{(46-55)^2}{55} + \frac{(36-45)^2}{45} + \frac{(54-45)^2}{45}$$

$$\chi^2 = 6.545$$

The chi-square statistic for these data has approximately 1 degree of freedom, an α of 0.05, and it is compared to the critical values on standard Chi-square table. Note that the degrees of freedom would increase as the number of rows and columns of our tables increases (for instance a 3 X 4 table). Since our calculated value ($\chi 2 = 6.545$) is greater than the critical value (3.841), we can once again reject the null hypothesis that there is no association between the exposure and the outcome of interest, and conclude that in this case seeing a social worker is significantly associated with a greater satisfaction with care.

Important notes

It is important to remember that the statistical tests and examples presented here are only an elementary presentation of the large scope of situations that can be addressed by these data. The intention of this notebook is to provide a basic understanding of the underlying principles of these statistical tests rather than implying that what has been presented is appropriate for every situation.

Further information about these statistical tests and other applications can be found in the following references:

Statistical First Aid: Interpretation of Health Research Data by Robert P Hirsch and Richard K. Riegelman. Blackwell Scientific Publications, Cambridge, MA 1992.

Categorical Data Analysis, Using the SAS System by ME Stokes, CS Davis, and GG Koch. SAS Institute Inc., Cary, NC, 2001.

References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

The University of North Carolina at Chapel Hill, Department of Epidemiology Courses: Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, and Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-2013.

Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH and Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP) to promote the strategic growth of the epidemiologic capacity of the DVA.