Hypothesis Testing

Overview

Chapter 6 dealt with estimating the values for parameters of various distributions (point estimates and interval estimates). A more common situation is testing whether or not the value of a parameter estimated using a sample is equal to some preconceived value. Through previous experiments, from anecdotal evidence, or through observation, we may have some idea of what the value of some parameter ought to be. We can use hypothesis testing as a way of seeing whether or not sample data support that belief. We may also have to sets of sample data and want o know if we can say that our two groups differ in some way. In chapter 6, the approach taken to answering such questions was creation of confidence intervals. Chapters 7 and 8 address these questions using hypothesis testing.

Hypothesis testing provides us with a framework for making decisions using probabilistic methods instead of relying on subjective impressions or opinions. Chapter 7 of Rosner covers one-sample problems, testing hypotheses involving a single distribution. Chapter 8 compares two different distributions in two-sample problems. Much of the material is also covered in Chapters 8 and 9 of the *Cartoon Guide* and you can take a look at that material to assist you in understanding the methods of hypothesis testing.

Learning Objectives:

Chapter 7

1. Understand what a null hypothesis and an alternative hypothesis are.
2. Know what Type I and Type II errors are.
3. Know how to perform one sample tests for the mean of a normal distribution.
4. Understand what a p-value represents.
5 Understand the concept of the power of a test.
6. Know how to calculate the required sample size when testing for the mean of a normal distribution.
7. Understand the relationship between hypothesis testing and confidence intervals.
8. Know how to perform the Chi-Square test for the variance of a normal distribution.
9. Know how to perform one sample tests for binomial proportions.
10. Know how to calculate the required sample size when testing for binomial proportions.

Chapter 8

11. Know how to perform a paired t test.
12. Know how to perform a two sample t test for independent samples.
13. Know how to perform the F test for the equality of two variances.
14. Understand what the impact of outliers are and how they can be treated.
15. Know how to estimate the required sample size for comparing two means.

Problems

7.1-7.9, 7.21- 7.25

8.1- 8.21, 8.153- 8.156

What You Should Know

Chapter 7

Hypothesis testing is very closely related to the material in the previous chapter, confidence intervals. In many ways it can be considered a different way to look at the same problem. It involves a process of testing whether or not the value of a particular parameter is equal to a specified value. If you read Examples 7.1 and 7.2 you should get a pretty good idea of the general procedure one would follow to test a hypothesis in a one-sample problem. It is very important that you fully understand what is meant by the null and alternative hypotheses and how to formulate them. Definitions 7.2 and 7.3 detail what is meant by Type I and Type II errors and Definition 7.6 gives the calculation for the power of a test. These are all important concepts. Once you understand these concepts it should become obvious to you that the general aim is to use statistical tests that make probabilities of committing Type I and Type II errors as small as possible. That often involves compromise because we are moving in two different directions. If we want to make the probability of committing a Type I error small, we will need to reject the null hypothesis less often. If we want to make the probability of committing a Type II error small we will need to accept the null hypothesis less often. In general, our approach will be to set the value of committing a Type I error at some specific level and then to select a test that minimizes the probability of committing a Type II error (or equivalently maximizing the power of the test).

The bulk of Chapter 7 in Rosner formulates various one sample tests. Equation 7.2 gives the formula for testing whether the mean of a normal distribution with unknown variance equals a specified value versus the alternative hypothesis that the mean is less than that specified value. This situation calls for the use of the t-test. The discussion in the text following Equation 7.2 focuses on the critical value of the test and the difference between using the critical value method and obtaining the p-value of a test. Make sure you fully understand the topics covered in this discussion. Equation 7.6 gives the details of the same testing process, only with the alternative hypothesis that the mean is greater than the specified value. Section 7.4 covers the same material with the exception that the alternative hypothesis is now two sided (a "not equal" situation as opposed to the "greater than" or "less than" situation previously covered). One needs to be able to differentiate between the one sided and two sided cases and be able to calculate p-values in both instances. Figure 7.1, 7.2 and 7.3 illustrates these differences.

If we have an idea of what the variance is from some previous studies, then the z-test can be substituted for the t-test. The one sample tests for the mean of a normal distribution with known variance are given in Equations 7.13, 7.14 and 7.15. These tests correspond to the three t-tests previously discussed with the only difference being that the underlying variances are now assumed to be known.

The next two sections of Chapter 7 focus on a rather detailed discussion of power and sample size determination. You need to understand the concept of the power of a test, the factors affecting power (Equation 7.20) and how to calculate power for the one sample z-test. Likewise, you need to understand why sample size determination is important, what the factors affecting sample size are (Equation 7.27) and how to calculate the required sample size for testing the mean of a normal distribution. As we stated at the beginning of this discussion, the topics in Chapter 6 and Chapter 7 are really two different ways of approaching the same problem. In particular, a two sided hypothesis test and a two sided confidence interval are analogous. Equations 7.30 and 7.31 illustrate this quite well.

You can skip the section on Bayesian Inference.

The remainder of the chapter covers some additional test procedures for the variance, the binomial proportion, and the expected number of events for a Poisson distribution. You learn how to use the chi-square test for testing whether or not the variance of a normal distribution equals a specified value (Equation 7.40) and how to compute the p-value for this chi-square test (Equation 7.41). You also learn how to use the z-test for testing whether or not a binomial proportion equals a specified value (Equation 7.42) and how to compute the p-value for this z-test (Equation 7.43). These latter two equations assume that the normal approximation to the binomial is valid. This assumption is true when npq is greater than or equal to 5. Pages 273 and 274 outline how to use an exact method for computing the p-value when the value of npq is less than 5. Finally, we learn how to conduct the one sample Poisson test for testing hypotheses concerning the parameter μ of a Poisson distribution and the concept of the standardized mortality ratio is introduced (Definition 7.16).

Chapter 8

Chapter 8 extends the discussion of hypothesis testing to cover the case of two samples. While the previous chapter dealt with testing in one sample whether the value of some parameter estimated with sample data equaled a specific value that we assumed was the known value of the population parameter, we now cover the case where we have two different populations and want to compare the values of some parameter in these populations when we do not know what the true value of the parameter is in either population. The comparison of two samples differs depending on whether the two samples are independent or paired. Please make sure you understand the material presented in the Section 8.1 on how paired and independent samples differ and what is meant by cross sectional studies and longitudinal studies.

When we have a paired design and wish to test whether the mean of two populations are equal, the paired t-test is used to test whether or not the underlying mean difference between the paired values is equal to zero. Equation 8.4 outlines how to conduct the paired t-test and Equation 8.5 illustrates how to compute the p-value for a paired t-test. The construction of a confidence interval for the difference between paired samples is covered in Section 8.3. Again, this is a different way of looking at the same problem.

When dealing with non-paired or independent samples, testing whether two means are equal depends upon whether or not we assume the underlying populations have equal variances. If they are equal, Equation 8.10 is used to calculate the pooled estimate of the variance from two independent samples and Equation 8.11 is used to compute the test statistic to test whether or not the means from the two populations are equal. This is the two-sample t-test for independent samples with equal variances. The computation of the p-value for this case is given in Equation 8.12. The construction of confidence intervals for this same case is covered in Section 8.5.

Section 8.6 covers the method used to validate the assumption that the variances are equal. You learn how to test whether or not the variances from two normal populations are equal, based on analyzing two independent random samples. The testing of the differences between variances is accomplished using the F-test. Equations 8.15 and 8.16 illustrate how to conduct the F-test for the equality of two variances and compute the associated p-value. If you want to test the difference in means between two independent samples, you can first use the F-test to see if the variances are equal. If they are, then you can use the method outlined in Equation 8.11 above, if not, then you can use the following method.

With unequal variances, we can test for the differences between two means for independent samples using the methodology outlined in Equation 8.21 and we can compute the p-value for this test using Equation 8.22. The related two sided confidence interval is constructed according to Equation 8.23. A flow chart which outlines the decision making process for deciding which test to use (depending on the result of the F-test) is given in Figure 8.10.

The next section in the text deals with outliers. After defining the ESD statistic (Definition 8.7), two methods are described for determining whether or not outliers are present in a data set. One of the procedures is for testing whether a single outlier is present (Equation 8.24) and the second is for testing if many outliers are present (Equation 8.25). The many outlier procedure is the suggested way of determining whether outliers exist unless we are very sure that only one outlier is present. Next follows a general discussion of what you can consider doing if outliers are present in a data set which you are analyzing. As you will note, there really is no universally accepted or recommended way to treat outliers but you should be familiar with the alternatives available.

Chapter 8 concludes with discussion about estimating sample size and power for some of the tests covered in the chapter. Section 8.10 covers the case where we have 2 independent samples. Equation 8.26 gives us the formula for determining the sample size needed for comparing means in two normally distributed samples of equal size and Equation 8.27 gives us the formula for the case of unequal sample size. Section 8.11 outlines the procedure for determining the necessary sample size for longitudinal studies.

Problems

7.1-7.9, 7.21- 7.25
These problems are a mix hypothesis testing using continuous and discrete data. They also include a few problems that require you to compute power and sample size. There are a number of equations in chapter 7 (for example 7.21, 7.28, 7.45, 7.46) used for power and sample size and you must choose the appropriate one(s) based on the type of data and the type of alternative hypothesis.

8.1- 8.21, 8.153- 8.156
Once again, these problems are a mix of hypothesis testing using continuous and discrete data. However, these are two-sample rather than one-sample problems. They again include a few problems that require you to compute power and sample size and it is your task to choose the appropriate equation(s) in chapter 8 based on the type of data and the type of alternative hypothesis.