

**Sampling distributions and the central limit theorem**

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Parameter and statistic

Statistic

A statistic is any function of the observations in a random sample.

We have encountered statistics before. For example, if X, X2, . . . , Xn is a random sample of size n, the sample mean  , the sample variance S2 , and the sample standard deviation S are statistics. Since a statistic is a random variable, it has a probability distribution.

Sampling distribution

The probability distribution of a statistic is called a sampling distribution.

For example, the probability distribution of is called the sampling distribution of the mean. The sampling distribution of a statistic depends on the distribution of the population, the size of the sample, and the method of sample selection. We now present perhaps the most important sampling distribution. Other sampling distributions and their applications will be illustrated extensively in the following two chapters.

Consider determining the sampling distribution of the sample mean . Suppose that a random sample of size n is taken from a normal population with mean and variance 2 . Now each observation in this sample, say, X1, X2, , Xn, is a normally and independently distributed random variable with mean and variance 2 . Then, because linear functions of independent, normally distributed random variables are also normally distributed with mean  and standard deviation 

Central limit theorem

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean and variance , if the sample size n is large. This is one of the most useful theorems in statistics, called the central limit theorem. The statement is as follows:

Sampling distribution of various statistics

Confidence interval

Hypothesis testing

Definition

Steps in hypothesis testing ( Explain each step)

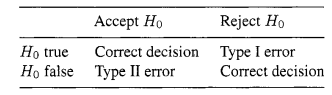
CONCEPTS USED IN STATISTICAL TESTING In this section we discuss the decision to accept or to reject a null hypotheses. The two types of error are defined and explained. 8.4.1 After testing a null hypothesis, one of two decisions is made; either HO is accepted or it is rejected. If we use a significance level Q = .05 and reject HO : p = 12.0, we feel reasonably sure that p is not 12.0 because in the long run, in repeated experimentation,

if it actually is 12.0, the mistake of rejecting p = 12.0 will occur only 5% of the time. On the other hand, if, with a = .05; HO is accepted, it should be emphasized that we should not say that p is 12.0; instead, we say that p may be 12.0. In a sense, the decision to reject HO is a more satisfactory decision than is the decision to accept Ho, since if we reject Ho, we are reasonably sure that p is not 12.0, whereas if we accept Ho, we simply conclude that p may be 12.0. The statistical test is better adapted to “disproving” than to “proving.” This is not surprising. If we find that the available facts do not fit a certain hypothesis, we discard the hypothesis. If, on the other hand, the available facts seem to fit the hypothesis, we do not know whether the hypothesis is correct or whether some other explanation might be even better. After we test and accept Ho, we have just a little more faith in HO than we had before making the test. If a hypothesis has been able to stand up under many attempts to disprove it, we begin to believe that it is a correct hypothesis, or at least nearly correct.

Two Kinds of Error

In a statistical test, two types of mistake can occur in making a decision. We may reject HO when it is actually true, or we may accept it when it is actually false. The two kinds of error are analogous to the two kinds of mistake that a jury can make. The jury may make the mistake of deciding that the accused is guilty when he or she is actually innocent, or the jury may make the mistake of deciding that the accused is innocent when he or she is actually guilty. The first type of error (rejecting Ho when it is really true) is called a type Z error and the chance of making a type I error is called a (alpha). If we reject the null hypothesis and if a is selected to be .01, then if HO is true, we have a 1% chance of making an error and deciding that HO is false. Earlier in this chapter we noted that often, though not always, a is set at .05. In certain types of biomedical studies, the consequences of making a type I error are worse than in others. For example, suppose that the medical investigator is testing a new vaccine for a disease for which no vaccine exists at present. Normally, either the new vaccine or a placebo vaccine is randomly assigned to volunteers. The null hypothesis is that both the new vaccine and the placebo vaccine are equally protective against the disease. If we reject the null hypothesis and conclude that the vaccine has been successful in preventing the disease when actually it was not successful, the consequences are serious. A worthless vaccine might be given to millions of people. Any vaccine has some risks attached to using it, so many people could be put at risk. Also, a great deal of time and money would have been wasted. In this case, an a of .05 does not seem small enough. It may then seem that a type I error should always be very small. But this is not true because of a second type of error. The second type of error, accepting HO when it is really false, is called a type ZZ error. The probability of making this type of error is called /3 (beta). The value of B depends on the numerical value of the unknown

Decisions and outcomes



true population mean, and the size of p is larger if the true population mean is close to the null hypothesis mean than if they are far apart. In some texts and statistical programs a type I error is called an a-error and type I1 error is called a p-error. Decisions and outcomes are illustrated in Table 8.4. Since we either accept or reject the null hypothesis, we know after making our decision which type of error we may have made. Ideally, we would like our chance of making an error to be very small. If our decision is to reject the Ho, we made either a correct decision or a type I error. Note that we set the size of a. The chances of our making a correct decision is 1 - a. If our decision is to accept the null hypothesis, we have made either a correct decision or a type I1 error. We do not set the size of p before we make a statistical test and do not know its size after we accept the null hypothesis. We can use a desired size of ,!? in estimating the needed sample size as seen in Section 8.5.

1. For a given size of a, 0, and n, the farther the hypothesized p is from the actual p, the smaller ,O will be if the null hypothesis is accepted. If we accept the null hypothesis, we are most apt to make a type I1 error if the hypothesized mean and the true mean are close together.
2. If we set the value of a very small, the acceptance region is larger than it would otherwise be. We are then more likely to accept the null hypothesis if it is not true; we will have increased our chance of making a type I1 error. Refer to Figure 8.4(a) and imagine the rejection region being made smaller in both tails, thus increasing the area of ,O in Figure 8.4(b).
3. If we increase the size of n, we will decrease the (standard error of the mean) and both normal curves will get narrower and higher (imagine them being squeezed together from both sides). Thus, the overlap between the two normal curves will be less. In this case, for a given preset value of Q and the same mean values, we will be less apt to make a type I1 error since the curves will not overlap so much. Since ,!? is the chance of making a type I1 error, 1 - ,O is the chance of not making a type I1 error and is called the power of the test. When making a test, we decide how small to make a and try to take a sufficient sample size so that the power is high. If we want a test to prove that the two population means are the same, a test called the test of equivalence would be used (see Wellek [2003]). 8.5 SAMPLE SIZE When planning a study, we want our sample size to be large enough so that we can reach correct conclusions but we do not want to take more observations than necessary. We illustrate the computation of the sample size for a test of the mean for two independent samples. Here we assume that the estimated sample sizes for the two groups are equal and call the sample size simply n. The formula for sample size is derived by solving n for a given value of a) ,!?, 0, and 1-11 - p2 (the difference between the hypothesized two means). Here, we assume that the estimated sample

REPORTING THE RESULTS In typical biomedical examples involving patients, numerous variables are measured. A few of these variables relate to the main purpose of the study and are called outcome variables. Here, tests of hypotheses are commonly reported. But other variables are collected that describe the individual patients, such as age, gender, seriousness of medical condition, results from a wide range of medical tests, attitudes, and previous treatment. These data are taken so that the patients being treated can be adequately described and to see if the patients in the various treatment groups were similar prior to treatment. In experiments using animals and in laboratory experiments, usually fewer variables are measured. In reporting the results, we have to decide what is of interest to the reader. When we report the results of tests of hypotheses, we should include not only the P value but also other information from the sample that will aid in interpreting the results. Simply reporting the P value is not sufficient. It is common when reporting the results of HO : p1 = p2 to also report the two sample means so that the reader can see their actual values. Usually, either the standard deviations or the standard errors of the means are also included. The standard deviation is given if we want the reader to have a measure of the variation of the observations and the standard error of the mean is given if it is more important to know the variation of the mean value. The standard deviation is sometimes easier to interpret than the standard error of the mean if the sample sizes in the two groups are unequal since its size does not depend on the sample size. Ranges are useful in warning the reader that outliers may be present. Often tables of means and standard deviations are given for variables that simply describe the patients in the study. This information is often given in a graphical form with bar graphs. If the shape of the distribution is of interest, then using histograms, box plots, or other available options are useful. These plots can be obtained directly from the statistical programs, and many readers find them easier to interpret than tables of statistics.