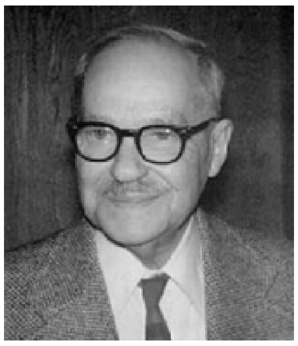
Chapter 7

Hypothesis Testing

Objective: In this chapter, various methods of testing hypotheses will be discussed.



Jerzy Neyman

(*Source: http://sciencematters.berkeley.edu/archives/volume2/issue12/legacy.php*)

Jerzy Neyman (1894–1981) made far-reaching contributions in hypothesis testing, confidence intervals, probability theory, and other areas of mathematical statistics. His work with Egon Pearson gave logical foundation and mathematical rigor to the theory of hypothesis testing. Their ideas made sure that samples were large enough to avoid false representation. Neyman made a broader impact in statistics throughout his lifetime.

7.1 INTRODUCTION

Statistics plays an important role in decision making. In statistics, one utilizes random samples to make inferences about the population from which the samples were obtained. Statistical inference regarding population parameters takes two forms: estimation and hypothesis testing, although both hypothesis testing and estimation may be viewed as different aspects of the same general problem of arriving at decisions on the basis of observed data. We already saw several estimation procedures in earlier chapters. Hypothesis testing is the subject of this chapter. Hypothesis testing has an important role in the application of statistics to real-life problems. Here we utilize the sampled data to make decisions concerning the unknown distribution of a population or its parameters. Pioneering work on the explicit formulation as well as the fundamental concepts of the theory of hypothesis testing are due to J. Neyman and E. S. Pearson.

A statistical hypothesis is a statement concerning the probability distribution of a random variable or population parameters that are inherent in a probability distribution. The following example illustrates the concept of hypothesis testing. An important industrial problem is that of accepting or rejecting lots of manufactured products. Before releasing each lot for the consumer, the manufacturer usually performs some tests to determine whether the lot conforms to acceptable standards. Let us say that both the manufacturer and the consumer agree that if the proportion of defectives in a lot is less than or equal to a certain number p, the lot will be released. Very often, instead of testing every item in the lot, we may test only a few items chosen at random from the lot and make decisions about the proportion of defectives in the lot; that is, we make the decisions about the population on the basis of sample information. Such decisions are called *statistical decisions*. In attempting to reach decisions, it is useful to make some initial conjectures about the population involved. Such conjectures are called *statistical hypotheses*. Sometimes the results from the sample may be markedly different from those expected under the hypothesis. Then we can say that the observed differences are significant and we would be inclined to reject the initial hypothesis. These procedures that enable us to decide whether to accept or reject hypotheses or to determine whether observed samples differ significantly from expected results are called *tests of hypotheses*, *tests of significance*, or *rules of decision*.

In any hypothesis testing problem, we formulate a *null hypothesis* and an *alternative hypothesis* such that if we reject the null, then we have to accept the alternative. The null hypothesis usually is a statement of either the "status quo" or "no effect." A guideline for selecting a null hypothesis is that when the objective of an experiment is to establish a claim, the nullification of the claim should be taken as the null hypothesis. The experiment is often performed to determine whether the null hypothesis is false. For example, suppose the prosecution wants to establish that a certain person is guilty. The null hypothesis would be that the person is innocent and the alternative would be that the person is guilty. Thus, the claim itself becomes the alternative hypothesis. Customarily, the alternative hypothesis is the statement that the experimenter believes to be true. For example, the alternative hypothesis is the reason a person is arrested (police suspect the person is not innocent). Once the hypotheses have been stated, appropriate statistical procedures are used to determine whether to reject the null hypothesis. For the testing procedure, one begins with the assumption that the null hypothesis is true. If the information furnished by the sampled data strongly contradicts (beyond a reasonable doubt) the null hypothesis, then we reject it in favor of the alternative hypothesis. If we do not reject the null, then we automatically reject the alternative. Note that we always make a decision with respect to the null hypothesis. Note that the failure to reject the null hypothesis does not necessarily mean that the null hypothesis is true. For example, a person being judged "not guilty" does not mean the person is innocent. This basically means that there is not enough evidence to reject the null hypothesis (presumption of innocence) beyond "a reasonable doubt."

We summarize the elements of a statistical hypothesis in the following.

THE ELEMENTS OF A STATISTICAL HYPOTHESIS

1. The *null hypothesis*, denoted by H_0, is usually the nullification of a claim. Unless evidence from the data indicates otherwise, the null hypothesis is assumed to be true.
2. The *alternate hypothesis*, denoted by H_a (or sometimes denoted by H_1), is customarily the claim itself.
3. The *test statistic*, denoted by *TS*, is a function of the sample measurements upon which the statistical decision, to reject or not reject the null hypothesis, will be based.
4. A *rejection region* (or a *critical region*) is the region (denoted by RR) that specifies the values of the observed test statistic for which the null hypothesis will be rejected. This is the range of values of the test statistic that corresponds to the rejection of H_0 at some fixed level of significance, α, which will be explained later.
5. **Conclusion:** If the value of the observed test statistic falls in the rejection region, the null hypothesis is rejected and we will conclude that there is enough evidence to decide that the alternative hypothesis is true. If the TS does not fall in the rejection region, we conclude that we cannot reject the null hypothesis.

In practice one may have hypotheses such as $H_0 : \mu = \mu_0$ against one of the following alternatives:

$$\begin{cases} & H_a : \mu \neq \mu_0, \quad \text{called a two-tailed alternative} \\ \text{or} & H_a : \mu < \mu_0, \quad \text{called a lower (or left) tailed alternative} \\ \text{or} & H_a : \mu > \mu_0, \quad \text{called an upper (or right) tailed alternative} \end{cases}$$

A test with a lower or upper tailed alternative is called a *one-tailed test*. In an applied hypothesis testing problem, we can use the following general steps.

GENERAL METHOD FOR HYPOTHESIS TESTING

1. From the (word) problem, determine the appropriate null hypothesis, H_0, and the alternative, H_a.
2. Identify the appropriate test statistics and calculate the observed test statistic from the data.
3. Find the rejection region by looking up the critical value in the appropriate table.
4. Draw the conclusion: Reject or fail to reject the null hypothesis, H_0.
5. Interpret the results: State in words what the conclusion means to the problem we started with.

It is always necessary to state a null and an alternate hypothesis for every statistical test performed. All possible outcomes should be accounted for by the two hypotheses.

Example 7.1.1

In a coin-tossing experiment, let p be the probability of heads. We start with the claim that the coin is fair, that is, $H_0: p = 1/2$. We test this against one of the following alternatives:

(a) H_a: The coin is not fair ($p \neq 1/2$). This is a two-tailed alternative.

(b) H_a: The coin is biased in favor of heads ($p > 1/2$). This is an upper tailed alternative.

(c) H_a: The coin is biased in favor of tails ($p < 1/2$). This is a lower tailed alternative.

It is important to observe that the test statistic is a function of a random sample. Thus, the test statistic itself is a random variable whose distribution is known under the null hypothesis. The value of a test statistic when specific sample values are substituted is called the observed test statistic or simply test statistic.

For example consider the hypothesis $H_0 : \mu = \mu_o$ versus $H_a : \mu \neq \mu_o$, where μ_o is known. Assume that the population is normal with a known variance σ^2. Consider $\overline{X}$, an unbiased estimator of μ based on the random sample $X_1, \ldots, X_n$. Then $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n})$ is a function of the random sample $X_1, \ldots, X_n$, and has a known distribution, a standard normal, under H_0. If $x_1, x_2, \ldots, x_n$ are specific sample values, then $z = (\overline{x} - \mu_0)/(\sigma/\sqrt{n})$ is called the *observed sample statistic* or simply *sample statistic*.

Definition 7.1.1 *A hypothesis is said to be a* **simple hypothesis** *if that hypothesis uniquely specifies the distribution from which the sample is taken. Any hypothesis that is not simple is called a* **composite hypothesis**.

Example 7.1.2

Refer to Example 7.1.1. The null hypothesis $p = 1/2$ is simple, because the hypothesis completely specifies the distribution, which in this case will be a binomial with $p = 1/2$ and with n being the number of tosses. The alternative hypothesis $p \neq 1/2$ is composite because the distribution now is not completely specified (we do not know the exact value of p).

Because the decision is based on the sample information, we are prone to commit errors. In a statistical test, it is impossible to establish the truth of a hypothesis with 100% certainty. There are two possible types of errors. On the one hand, one can make an error by rejecting H_0 when in fact it is true. On the other hand, one can also make an error by failing to reject the null hypothesis when in fact it is false. Because the errors arise as a result of wrong decisions, and the decisions themselves are based on random samples, it follows that the errors have probabilities associated with them. We now have the following definitions.

Table 7.1 Statistical Decision and Error Probabilities

Statistical decision	True state of null hypothesis	
	H_0 true	H_0 false
Do not reject H_0	Correct decision	Type II error (β)
Reject H_0	Type I error (α)	Correct decision

The decision and the errors are represented in Table 7.1.

Definition 7.1.2 **(a)** *A* **type I error** *is made if H_0 is rejected when in fact H_0 is true. The probability of type I error is denoted by α. That is,*

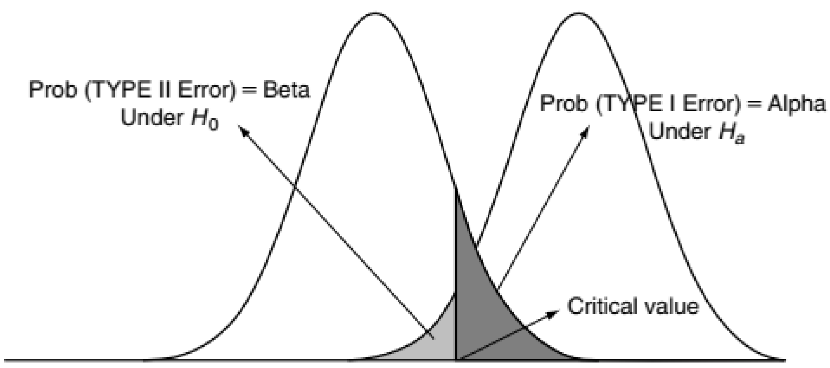
$$P(\text{rejecting } H_0 | H_0 \text{ is true}) = \alpha.$$

The probability of type I error, α, is called **the level of significance**.

(b) *A* **type II error** *is made if H_0 is accepted when in fact H_a is true. The probability of a type II error is denoted by β. That is,*

$$P(\text{not rejecting } H_0 | H_0 \text{ is false}) = \beta.$$

It is desirable that a test should have $a = \beta = 0$ (this can be achieved only in trivial cases), or at least we prefer to use a test that minimizes both types of errors. Unfortunately, it so happens that for a fixed sample size, as α decreases, β tends to increase and vice versa. There are no hard and fast rules that can be used to make the choice of α and β. This decision must be made for each problem based on quality and economic considerations. However, in many situations it is possible to determine which of the two errors is more serious. It should be noted that a type II error is only an error in the sense that a chance to correctly reject the null hypothesis was lost. It is not an error in the sense that an incorrect conclusion was drawn, because no conclusion is made when the null hypothesis is not rejected. In the case of type I error, a conclusion is drawn that the null hypothesis is false when, in fact, it is true. Therefore, type I errors are generally considered more serious than type II errors. For example, it is mostly agreed that finding an innocent person guilty is a more serious error than finding a guilty person innocent. Here, the null hypothesis is that the person is innocent, and the



alternate hypothesis is that the person is guilty. "Not rejecting the null hypothesis" is equivalent to acquitting a defendant. It does not prove that the null hypothesis is true, or that the defendant is innocent. In statistical testing, the significance level α is the probability of wrongly rejecting the null hypothesis when it is true (that is, the risk of finding an innocent person guilty). Here the type II risk is acquitting a guilty defendant. The usual approach to hypothesis testing is to find a test procedure that limits α, the probability of type I error, to an acceptable level while trying to lower β as much as possible.

The consequences of different types of errors are, in general, very different. For example, if a doctor tests for the presence of a certain illness, incorrectly diagnosing the presence of the disease (type I error) will cause a waste of resources, not to mention the mental agony to the patient. On the other hand, failure to determine the presence of the disease (type II error) can lead to a serious health risk.

To formulate a hypothesis testing problem, consider the following situation. Suppose a toy store chain claims that at least 80% of girls under 8 years old prefer dolls over other types of toys. We feel that this claim is inflated. In an attempt to dispose of this claim, we observe the buying pattern of 20 randomly selected girls under 8 years old, and we observe X, the number of girls under 8 years old who buy stuffed toys or dolls. Now the question is, how can we use X to confirm or reject the store's claim? Let p be the probability that a girl under 8 chosen at random prefers stuffed toys or dolls. The question now can be reformulated as a hypothesis testing problem. Is $p \geq 0.8$ or $p < 0.8$? Because we would like to reject the store's claim only if we are highly certain of our decision, we should choose the null hypothesis to be $H_0: p \geq 0.8$, the rejection of which is considered to be more serious. The null hypothesis should be $H_0: p \geq 0.8$, and the alternative $H_a: p < 0.8$. In order to make the null hypothesis simple, we will use $H_0: p = 0.8$, which is the boundary value with the understanding that it really represents $H_0: p \geq 0.8$. We note that X, the number of girls under 8 years old who prefer stuffed toys or dolls, is a binomial random variable. Clearly a large sample value of X would favor H_0. Suppose we arbitrarily choose to accept the null hypothesis if $X > 12$. Because our decision is based on only a sample of 20 girls under 8, there is always a possibility of making errors whether we accept or reject the store chain's claim. In the following example, we will now formally state this problem and calculate the error probabilities based on our decision rule.

Example 7.1.3

A toy store chain claims that at least 80% of girls under 8 years old prefer dolls over other types of toys. After observing the buying pattern of many girls under 8 years old, we feel that this claim is inflated. In an attempt to dispose of this claim, we observe the buying pattern of 20 randomly selected girls under 8 years old, and we observe X, the number of girls who buy stuffed toys or dolls. We wish to test the hypothesis $H_0 : p = 0.8$ against $H_a : p < 0.8$. Suppose we decide to accept the H_0 if $X > 12$ (that is $X \geq 13$). This means that if $\{X \leq 12\}$ (that is $X < 13$) we will reject H_0.

(a) Find α.
(b) Find β for $p = 0.6$.
(c) Find β for $p = 0.4$.
(d) Find the rejection region of the form $\{X \leq K\}$ so that **(i)** $\alpha = 0.01$; **(ii)** $\alpha = 0.05$.
(e) For the alternative $H_a: p = 0.6$, find β for the values of α in part (d).

Solution

The TS X is the number of girls under 8 years old who buy dolls. X follows the binomial distribution with $n = 20$ and p, the unknown population proportion of girls under 8 who prefer dolls. We now calculate α and β.

(a) *For $p = 0.8$, the probability of type I error is*

$$
\begin{aligned}
\alpha &= P\{\text{reject } H_0 | H_0 \text{ is true}\} \\
&= P\{X \le 12 | p = 0.8\} \\
&= \sum_{x=0}^{12} \binom{20}{x} (0.8)^x (0.2)^{20-x} \\
&= 0.0321.
\end{aligned}
$$

If we calculate α for any other value of $p > 0.8$, then we will find that it is smaller than 0.0321. Hence, there is at most a 3.21% chance of rejecting a true null hypothesis. That is, if the store's claim is in fact true, then the chance that our test will erroneously reject that claim is at most 3.21%.

(b) *Here $p = 0.6$. The probability of type II error is*

$$
\begin{aligned}
\beta &= P\{\text{accept } H_0 | H_0 \text{ false}\} \\
&= P\{X > 12 | p = 0.6\} \\
&= 1 - P\{X \le 12 | p = 0.6\} \\
&= 1 - 0.584 \\
&= 0.416
\end{aligned}
$$

so there is a 4.2% chance of accepting a false null hypothesis. Thus, in case the store's claim is not true, and the truth is that only 60% of girls under 8 years old prefer dolls over other types of toys, then there is a 4.2% chance that our test will erroneously conclude that the store's claim is true.

(c) *If $p = 0.4$, then*

$$
\begin{aligned}
\beta &= P\{\text{accept } H_0 | H_0 \text{ false}\} \\
&= P\{X > 12 | p = 0.4\} \\
&= 1 - P\{X \le 12 | p = 0.4\} \\
&= 1 - 0.979 \\
&= 0.021.
\end{aligned}
$$

That is, there is a 2.1% chance of accepting a false null hypothesis.

(d) **(i)** *To find K such that*

$$\alpha = P\{X \le K | p = 0.8\} = 0.01$$

from the binomial table, $K = 11$. Hence, the rejection region is: Reject H_0 if $\{X \le 11\}$.

(ii) *To find K such that*

$$\alpha = P\{X \le K | p = 0.8\} = 0.05$$

from the binomial table, $\alpha = 0.05$ falls between $K = 12$ and $K = 13$. However, for $K = 13$, the value for α is 0.087, exceeding 0.05. If we want to limit α to be no more than 0.05, we will have to take $K = 12$. That is, we reject the null hypothesis if $X \leq 12$, yielding an $\alpha = 0.0321$ as shown in (a).

(e) (i) *When $a = 0.01$, from* (d)*, the rejection region is of the form $\{X \leq 11\}$. For $p = 0.6$,*

$$\begin{aligned}\beta &= P\{\text{accept } H_0 | H_0 \text{ false}\} \\ &= P\{Y > 11 | p = 0.6\} \\ &= 1 - P\{Y \leq 11 | p = 0.6\} \\ &= 1 - 0.404 \\ &= 0.596.\end{aligned}$$

(ii) *From* (a) *and* (b) *for testing the hypothesis $H_0 : p = 0.8$ against $H_a : p < 0.8$ with $n = 20$. We see that when α is 0.0321, β is 0.416. From* (d)(i) *and* (e)(i) *for the same hypothesis, we see that when α is 0.01, β is 0.596. This holds in general. Thus, we observe that for fixed n as α decreases, β increases and vice versa.*

In the next example, we explore what happens to β as the sample size increases, with α fixed.

Example 7.1.4

Let X be a binomial random variable. We wish to test the hypothesis $H_0 : p = 0.8$ against $H_a : p = 0.6$. Let $\alpha = 0.03$ be fixed. Find β for $n = 10$ and $n = 20$.

Solution

For $n = 10$, using the binomial tables, we obtain $P\{X \leq 5 | p = 0.8\} \cong 0.03$. Hence the rejection region for the hypothesis $H_0 : p = 0.8$ vs. $H_a : p = 0.6$ is given by reject H_0 if $X \leq 5$. The probability of type II error is

$$\beta = P\{\textit{accept } H_0 | p = 0.6\}$$

$$\beta = P\{X > 5 | p = 0.6\} = 1 - P\{X \leq 5 | p = 0.6\} = 0.733.$$

For $n = 20$, as shown in Example 7.1.3, if we reject H_0 for $X \leq 12$, we obtain

$$P(X \leq 12 | p = 0.8) \cong 0.03$$

and

$$\beta = P(X > 12 | p = 0.6) = 1 - P\{X \leq 12 | p = 0.6\} = 0.416.$$

We see that for a fixed α, as n increases β decreases and vice versa. It can be shown that this result holds in general.

In order for us to compute the value of β, it is necessary that the alternate hypothesis is simple. Now we will discuss a three-step procedure to calculate β.

STEPS TO CALCULATE β

1. Decide an appropriate test statistic (usually this is a sufficient statistic or an estimator for the unknown parameter, whose distribution is known under H_0).
2. Determine the rejection region using a given α, and the distribution of the test statistic (TS).
3. Find the probability that the observed test statistic does not fall in the rejection region assuming H_a is true. This gives β. That is,

$$\beta = P(T.S. \text{ falls in the complement of the rejection region} | H_a \text{is true}).$$

Example 7.1.5

A random sample of size 36 from a population with known variance, $\sigma^2 = 9$, yields a sample mean of $\bar{x} = 17$. Find β, for testing the hypothesis $H_0 : \mu = 15$ versus $H_a : \mu = 16$. Assume $\alpha = 0.05$.

Solution

Here $n = 36, \bar{x} = 17$, and $\sigma^2 = 9$. In general, to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$, we proceed as follows. An unbiased estimator of μ is $\overline{X}$. Intuitively we would reject H_0 if $\overline{X}$ is large, say $\overline{X} > c$. Now using $\alpha = 0.05$, we will determine the rejection region. By the definition of α, we have

$$P(\overline{X} > c \, | \mu = \mu_0) = 0.05$$

or

$$P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}} \, | \mu = \mu_0\right) = 0.05$$

But if $\mu = \mu_0$, because the sample size $n \geq 30$, $[(\overline{X} - \mu_0)/(\sigma/\sqrt{n})] \sim N(0, 1)$. Therefore, $P\left(\frac{\overline{X}-\mu_0}{(\sigma/\sqrt{n})} > \frac{c-\mu_0}{(\sigma/\sqrt{n})}\right) = 0.05$ is equivalent to $P\left(Z > \frac{c-\mu_0}{(\sigma/\sqrt{n})}\right) = 0.05$. From standard normal tables, we obtain $P(Z > 1.645) = 0.05$. Hence $\frac{c-\mu_0}{(\sigma/\sqrt{n})} = 1.645$ or $c = \mu_0 + 1.645(\sigma/\sqrt{n})$.
Therefore, the rejection region is the set of all sample means $\bar{x}$ such that

$$\bar{x} > \mu_0 + 1.645\left(\frac{\sigma}{\sqrt{n}}\right).$$

Substituting $\mu_0 = 15$, and $\sigma = 3$, we obtain

$$\mu_0 + 1.645(\sigma/\sqrt{n}) = 15 + 1.645\left(\frac{3}{36}\right) = 15.8225.$$

The rejection region is the set of $\bar{x}$ such that $\bar{x} \geq 15.8225$.

Then by definition,

$$\beta = P(\overline{X} \leq 15.8225 \text{ when } \mu = 16).$$

Consequently, for $\mu = 16$,

$$\begin{aligned}\beta &= P\left(\frac{\overline{X} - 16}{\sigma/\sqrt{n}} \leq \frac{15.8225 - 16}{3/\sqrt{36}}\right)\\ &= P(Z \leq -0.36)\\ &= 0.3594.\end{aligned}$$

That is, under the given information, there is a 35.94% chance of not rejecting a false null hypothesis.

■

7.1.1 Sample Size

It is clear from the preceding example that once we are given the sample size n, an α, a simple alternative H_a, and a test statistic, we have no control over β and it is exactly determined. Hence, for a given sample size and test statistic, any effort to lower β will lead to an increase in α and vice versa. This means that for a test with fixed sample size it is not possible to simultaneously reduce both α and β. We also notice from Example 7.1.4 that by increasing the sample size n, we can decrease β (for the same α) to an acceptable level. The following discussion illustrates that it may be possible to determine the sample size for a given α and β.

Suppose we want to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$. Given α and β, we want to find n, the sample size, and K, the point at which the rejection begins. We know that

$$\begin{aligned}\alpha &= P(\overline{X} > K \text{ when } \mu = \mu_0)\\ &= P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{K - \mu_0}{\sigma/\sqrt{n}}, \text{ when } \mu = \mu_0\right) \qquad (7.1)\\ &= P(Z > z_a)\end{aligned}$$

and

$$\begin{aligned}\beta &= P(\overline{X} \leq K, \text{ when } \mu = \mu_a)\\ &= P\left(\frac{\overline{X} - \mu_a}{\sigma/\sqrt{n}} \leq \frac{K - \mu_a}{\sigma/\sqrt{n}}, \text{ when } \mu = \mu_a\right) \qquad (7.2)\\ &= P(z \leq -z_\beta).\end{aligned}$$

From Equations (7.1) and (7.2),

$$z_\alpha = \frac{K - \mu_0}{\sigma/\sqrt{n}}$$

and

$$-z_{\beta} = \frac{K - \mu_a}{\sigma/\sqrt{n}}.$$

This gives us two equations with two unknowns (K and n), and we can proceed to solve them. Eliminating K, we get

$$\mu_0 + z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - z_{\beta}\left(\frac{\sigma}{\sqrt{n}}\right).$$

From this we can derive

$$\sqrt{n} = \frac{(z_{\alpha} + z_{\beta})\sigma}{\mu_a - \mu_0}.$$

Thus, the sample size for an upper tail alternative hypothesis is

$$n = \frac{(z_{\alpha} + z_{\beta})^2\sigma^2}{(\mu_a - \mu_0)^2}.$$

The sample size increases with the square of the standard deviation and decreases with the square of the difference between mean value of the alternative hypothesis and the mean value under the null hypothesis. Note that in real-world problems, care should be taken in the choice of the value of μ_a for the alternative hypothesis. It may be tempting for a researcher to take a large value of μ_a in order to reduce the required sample size. This will seriously affect the accuracy (power) of the test. This alternative value must be realistic within the experiment under study. Care should also be taken in the choice of the standard deviation σ. Using an underestimated value of the standard deviation to reduce the sample size will result in inaccurate conclusions similar to overestimating the difference of means. Usually, the value of σ is estimated using a similar study conducted earlier. The problem could be that the previous study may be old and may not represent the new reality. When accuracy is important, it may be necessary to conduct a pilot study only to get some idea on the estimate of σ. Once we determine the necessary sample size, we must devise a procedure by which the appropriate data can be randomly obtained. This aspect of the design of experiments is discussed in Chapter 9.

Example 7.1.6

Let $\sigma = 3.1$ be the true standard deviation of the population from which a random sample is chosen. How large should the sample size be for testing $H_0 : \mu = 5$ versus $H_a : \mu = 5.5$, in order that $\alpha = 0.01$ and $\beta = 0.05$?

Solution

We are given $\mu_0 = 5$ *and* $\mu_a = 5.5$. *Also,* $z_{\alpha} = z_{0.01} = 2.33$ *and* $z_{\beta} = z_{0.05} = 1.645$. *Hence, the sample size*

$$n = \frac{(z_{\alpha} + z_{\beta})^2\sigma^2}{(\mu_a - \mu_0)^2} = \frac{(2.33 + 1.645)^2(3.1)^2}{(0.5)^2} = 607.37.$$

So, $n = 608$ will provide the desired levels. That is, in order for us to test the foregoing hypothesis, we must randomly select 608 observations from the given population.

From a practical standpoint, the researcher typically chooses α, and the sample size β is ignored. Because a trade-off exists between α and β, choosing a very small value of α will tend to increase β in a serious way. A general rule of thumb is to pick reasonable values of α, possibly in the 0.05 to 0.10 range so that β will remain reasonably small.

EXERCISES 7.1

7.1.1. An appliance manufacturer is considering the purchase of a new machine for stamping out sheet metal parts. If μ_0 (given) is the true average of the number of good parts stamped out per hour by their old machine and μ is the corresponding true unknown average for the new machine, the manufacturer wants to test the null hypothesis $\mu = \mu_0$ versus a suitable alternative. What should the alternative be if he does not want to buy the new machine unless it is **(a)** more productive than the old one? **(b)** At least 20% more productive than the old one?

7.1.2. Formulate an alternative hypothesis for each of the following null hypotheses.

(a) H_0: Support for a presidential candidate is unchanged after the start of the use of TV commercials.

(b) H_0: The proportion of viewers watching a particular local news channel is less than 30%.

(c) H_0: The median grade point average of undergraduate mathematics majors is 2.9.

7.1.3. It is suspected that a coin is not balanced (not fair). Let p be the probability of tossing a head. To test $H_0 : p = 0.5$ against the alternative hypothesis $H_a : p > 0.5$, a coin is tossed 15 times. Let Y equal the number of times a head is observed in the 15 tosses of this coin. Assume the rejection region to be $\{Y \geq 10\}$.

(a) Find α.

(b) Find β for $p = 0.7$.

(c) Find β for $p = 0.6$.

(d) Find the rejection region for $\{Y \geq K\}$ for $\alpha = 0.01$, and $\alpha = 0.03$.

(e) For the alternative $H_a : p = 0.7$, find β for the values of α given in (d).

7.1.4. In Exercise 7.1.3:

(a) Assume that the rejection region is $\{Y \geq 8\}$. Calculate α and β if $p = 0.6$. Compare the results with the corresponding values obtained in Exercise 7.1.3. (This gives the effect of enlarging the rejection region on α and β.)

(b) Assume that the rejection region is $\{Y \geq 8\}$. Calculate α and β if $p = 0.6$ and **(i)** the coin is tossed 20 times, or **(ii)** the coin is tossed 25 times. (This shows the effect of increasing the sample size on α and β for a fixed rejection region.)

7.1.5. Suppose we have a random sample of size 25 from a normal population with an unknown mean μ and a standard deviation of 4. We wish to test the hypothesis $H_0 : \mu = 10$ vs.

$H_a : \mu > 10$. Let the rejection region be defined by: reject H_0 if the sample mean $\overline{X} > 11.2$.

(a) Find α.

(b) Find β for $H_a : \mu = 11$.

(c) What should the sample size be if $\alpha = 0.01$ and $\beta = 0.8$?

7.1.6. A process for making steel pipe is under control if the diameter of the pipe has mean 3.0 in. with standard deviation of no more than 0.0250 in. To check whether the process is under control, a random sample of size $n = 30$ is taken each day and the null hypothesis $\mu = 3.0$ is rejected if $\overline{X}$ is less than 2.9960 or greater than 3.0040. Find **(a)** the probability of type I error; **(b)** the probability of type II error when $\mu = 3.0050$ in. Assume $\sigma = 0.0250$ in.

7.1.7. A bowl contains 20 balls, of which x are green and the remain- der red. To test $H_0 : x = 10$ versus $H_a : x = 15$, three balls are selected at random without replacement, and H_0 is rejected if all three balls are green. Calculate α and β for this test.

7.1.8. Suppose we have a sample of size 6 from a population with pdf $f(x) = (1/\theta)e^{-x/\theta}, x > 0, \theta > 0$. We wish to test $H_0 : \theta = 1$ vs. $H_a : \theta > 1$. Let the rejection region be defined by reject H_0 if $\sum_{i=1}^{6} X_i > 8$. **(a)** Find α. **(b)** Find β for $H_a : \theta = 2$.

7.1.9. Let $\sigma^2 = 16$ be the variance of a normal population from which a random sample is chosen. How large should the sample size be for testing $H_0 : \mu = 25$ versus $H_a : \mu = 24$, in order that $\alpha = 0.05$ and $\beta = 0.05$?

7.2 THE NEYMAN–PEARSON LEMMA

In practical hypothesis testing situations, there are typically many tests possible with significance level α for a null hypothesis versus alternative hypothesis (see Project 7A). This leads to some important questions, such as (1) how to decide on the test statistic and (2) how to know that we selected the best rejection region. In this section, we study the answer to these questions using the Neyman–Pearson approach.

Definition 7.2.1 *Suppose that W is the test statistic and RR is the rejection region for a test of hypothesis concerning the value of a parameter θ. Then the* **power** *of the test is the probability that the test rejects H_0 when the alternative is true. That is,*

$$\pi = Power(\theta)$$
$$= P(W \text{ in } RR \text{ when the parameter value is an alternative } \theta).$$

If $H_0 : \theta = \theta_0$ and $H_a : \theta \neq \theta_0$, then the power of the test at some $\theta = \theta_1 \neq \theta_0$ is

$$Power(\theta_1) = P(\text{reject } H_0 | \theta = \theta_1).$$

But, $\beta(\theta_1) = P(\text{accept } H_0 | \theta = \theta_1)$. Therefore,

$$Power(\theta_1) = 1 - \beta(\theta_1).$$

A good test will have high power.

Note that the power of a test H_0 cannot be found until some true situation H_a is specified. That is, the sampling distribution of the test statistic when H_a is true must be known or assumed. Because β depends on the alternative hypothesis, which being composite most of the time does not specify the distribution of the test statistic, it is important to observe that the experimenter cannot control β. For example, the alternative $H_a : \mu < \mu_0$ does not specify the value of μ, as in the case of the null hypothesis, $H_0 : \mu = \mu_0$.

Example 7.2.1

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, that is, the pdf is given by $f(x) = e^{-\lambda}\lambda^x/(x!)$. Then the hypothesis $H_0 : \lambda = 1$ uniquely specifies the distribution, because $f(x) = e^{-1}/(x!)$ and hence is a simple hypothesis. The hypothesis $H_a : \lambda > 1$ is composite, because $f(x)$ is not uniquely determined.

Definition 7.2.2 *A test at a given α of a simple hypothesis H_0 versus the simple alternative H_a that has the largest power among tests with the probability of type I error no larger than the given α is called a* **most powerful** *test.*

Consider the test of hypothesis $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_1$. If α is fixed, then our interest is to make β as small as possible. Because $\beta = 1 - Power(\theta_1)$, by minimizing β we would obtain a most powerful test. The following result says that among all tests with given probability of type I error, the likelihood ratio test given later minimizes the probability of a type II error, in other words, it is most powerful.

Theorem 7.2.1 (Neyman–Pearson Lemma) *Suppose that one wants to test a simple hypothesis H_0 : $\theta = \theta_0$ versus the simple alternative hypothesis $H_a : \theta = \theta_1$ based on a random sample $X_1, \ldots, X_n$ from a distribution with parameter θ. Let $L(\theta) \equiv L(\theta; X_1, \ldots, X_n) > 0$ denote the likelihood of the sample when the value of the parameter is θ. If there exist a positive constant K and a subset C of the sample space $\mathbb{R}^n$ (the Euclidean n-space) such that*

1. $\dfrac{L(\theta_0)}{L(\theta_1)} \le K$ *for* $(x_1, x_2, \ldots, x_n) \in C$
2. $\dfrac{L(\theta_0)}{L(\theta_1)} \ge K$ *for* $(x_1, x_2, \ldots, x_n) \in C'$, *where C' is the complement of C, and*
3. $P[(X_1, \ldots, X_n) \in C; \theta_0] = \alpha$.

Then the test with critical region C will be the most powerful test for H_0 versus H_a. We call α the size of the test and C the best critical region of size α.

Proof. We prove this theorem for continuous random variables. For discrete random variables, the proof is identical with sums replacing the integral. Let S be some region in $\mathbb{R}^n$, an n-dimensional Euclidean space. For simplicity we will use the following notation:

$$\int_S L(\theta) = \int_S \ldots \int_S L(\theta; x_1, x_2, \ldots, x_n)dx_1dx_2, \ldots, dx_n$$

Note that

$$P((X_1, \ldots, X_n) \in C; \theta_0) = \int\limits_C f(x_1, \ldots, x_n; \theta_0) dx_1, \ldots, dx_n$$

$$= \int\limits_C L(\theta_0; x_1, \ldots, x_n) dx_1, \ldots, dx_n.$$

Suppose that there is another critical region, say B, of size less than or equal to α, that is $\int_B L(\theta_0) \leq \alpha$. Then

$$0 \leq \int\limits_C L(\theta_0) - \int\limits_B L(\theta_0), \text{ because } \int\limits_C L(\theta_0) = \alpha \text{ by assumption 3.}$$

Therefore,

$$0 \leq \int\limits_C L(\theta_0) - \int\limits_B L(\theta_0)$$

$$= \int\limits_{C \cap B} L(\theta_0) + \int\limits_{C \cap B'} L(\theta_0) - \int\limits_{C \cap B} L(\theta_0) - \int\limits_{C' \cap B} L(\theta_0)$$

$$= \int\limits_{C \cap B'} L(\theta_0) - \int\limits_{C' \cap B} L(\theta_0).$$

Using assumption 1 of Theorem 7.2.1, $KL(\theta_1) \geq L(\theta_0)$ at each point in the region C and hence in $C \cap B'$. Thus

$$\int\limits_{C \cap B'} L(\theta_0) \leq K \int\limits_{C \cap B'} L(\theta_1).$$

By assumption 2 of the theorem, $KL(\theta_1) \leq L(\theta_0)$ at each point in C', and hence in $C' \cap B$. Thus,

$$\int\limits_{C' \cap B} L(\theta_0) \geq K \int\limits_{C' \cap B} L(\theta_1).$$

Therefore,

$$0 \leq \int\limits_{C \cap B'} L(\theta_0) - \int\limits_{C' \cap B} L(\theta_0)$$

$$\leq K \left\{ \int\limits_{C \cap B'} L(\theta_1) - \int\limits_{C' \cap B} L(\theta_1) \right\}.$$

That is,

$$0 \leq K \left\{ \int_{C \cap B} L(\theta_1) + \int_{C \cap B'} L(\theta_1) - \int_{C \cap B} L(\theta_1) - \int_{C' \cap B} L(\theta_1) \right\}$$

$$= K \left\{ \int_{C} L(\theta_1) - \int_{B} L(\theta_1) \right\}.$$

As a result,

$$\int_{C} L(\theta_1) \geq \int_{B} L(\theta_1).$$

Because this is true for every critical region B of size $\leq \alpha$, C is the best critical region of size α, and the test with critical region C is the most powerful test of size α. □

When testing two simple hypotheses, the existence of a best critical region is guaranteed by the Neyman–Pearson lemma. In addition, the foregoing theorem provides a means for determining what the best critical region is. However, it is important to note that Theorem 7.2.1 gives only the form of the rejection region; the actual rejection region depends on the specific value of α.

In real-world situations, we are seldom presented with the problem of testing two simple hypotheses. There is no general result in the form of Theorem 7.4.1 for composite hypotheses. However, for hypotheses of the form $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$, we can take a particular value $\theta_1 > \theta_0$ and then find a most powerful test for $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_1$. If this test (that is, the rejection region of the test) does not depend on the particular value θ_1, then this test is said to be a *uniformly most powerful test* for $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$.

The following example illustrates the use of the Neyman–Pearson lemma.

Example 7.2.2

Let $X_1, \ldots, X_n$ denote an independent random sample from a population with a Poisson distribution with mean λ. Derive the most powerful test for testing $H_0 : \lambda = 2$ versus $H_a : \lambda = 1/2$.

Solution

Recall that the pdf of Poisson variable is

$$p(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \lambda > 0, x = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the likelihood function is

$$L = \frac{\left[\lambda^{\left(\sum\limits_{i=1}^{n} x_i\right)} e^{-\lambda n} \right]}{\prod\limits_{i=1}^{n} (x_i!)}.$$

For $\lambda = 2$,

$$L(\theta_0) = L(\lambda = 2) = \frac{\left[2^{\left(\sum_{i=1}^{n} x_i\right)} e^{-2n}\right]}{\prod_{i=1}^{n} (x_i!)}$$

and for $\lambda = 1/2$,

$$L(\theta_1) = L(\lambda = 1/2) = \frac{\left[(1/2)^{\left(\sum_{i=1}^{n} x_i\right)} e^{-(1/2)n}\right]}{\prod_{i=1}^{n} (x_i!)}$$

Thus,

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{\left(2^{(\sum x_i)} e^{-n2}\right)}{\left(\frac{1}{2}\right)^{\sum x_i} e^{-\frac{n}{2}}} < K$$

which implies

$$(4)^{\sum x_i}\left(e^{-\frac{3n}{2}}\right) < K$$

or, taking natural logarithm,

$$\left(\sum x_i\right) \ln 4 - \frac{3n}{2} < \ln K.$$

Solving for $(\sum x_i)$ *and letting* $\{[\ln K + (3n/2)]/\ln 4\} = K'$, *we will reject* H_0 *whenever* $(\sum x_i) < K'$.

A step-by-step procedure in applying the Neyman–Pearson lemma is now given.

PROCEDURE FOR APPLYING THE NEYMAN–PEARSON LEMMA

1. Determine the likelihood functions under both null and alternative hypotheses.
2. Take the ratio of the two likelihood functions to be less than a constant K.
3. Simplify the inequality in step 2 to obtain a rejection region.

Example 7.2.3

Suppose $X_1, \ldots, X_n$ is a random sample from a normal distribution with a known mean of μ and an unknown variance of σ^2. Find the most powerful α-level test for testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_a : \sigma^2 = \sigma_1^2 (\sigma_1^2 > \sigma_0^2)$. Show that this test is equivalent to the χ^2-test. Is the test uniformly most powerful for $H_a : \sigma^2 > \sigma_0^2$?

Solution

To test $H_0 : \sigma^2 = \sigma_0^2$ *versus* $H_a : \sigma^2 > \sigma_1^2$. *We have*

$$L(\sigma_0^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0^n} e^{-\frac{(x_i - \mu)^2}{2\sigma_0^2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n \sigma_0^n} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma_0^n}}.$$

Similarly,

$$L(\sigma_1^2) = \frac{1}{(\sqrt{2\pi})^n \sigma_1^n} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma_1^2}}.$$

Therefore, the most powerful test is, reject H_0 *if,*

$$\frac{L(\sigma_0^2)}{L(\sigma_1^2)} = \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^n e^{\left[-\frac{(\sigma_1^2 - \sigma_0^2)^2}{2\sigma_1^2\sigma_0^2} \sum (x_i - \mu)^2\right]} \leq K$$

for some K.

Taking the natural logarithms, we have

$$n \ln\left(\frac{\sigma_1}{\sigma_0}\right) - \frac{(\sigma_1^2 - \sigma_0^2)}{2\sigma_1^2\sigma_0^2} \sum (x_i - \mu)^2 \leq \ln K$$

or

$$\sum (x_i - \mu)^2 \geq \left[n \ln\left(\frac{\sigma_1}{\sigma_0}\right) - \ln K\right] \left(\frac{2\sigma_1^2\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\right) = C.$$

To find the rejection region for a fixed value of α, *write the region as*

$$\frac{\sum (x_i - \mu)^2}{\sigma_0^2} \geq \frac{C}{\sigma_0^2} = C'.$$

Note that $\sum (x_i - \mu)^2 / \sigma_0^2$ *has a* χ^2*-distribution with* n *degrees of freedom. Under the* H_0 *because the same rejection region (does not depend upon the specific value of* σ_1^2 *in the alternative) would be used for any* $\sigma_1^2 > \sigma_0^2$, *the test is uniformly most powerful.*

■

The foregoing example shows that, in order to test for variance using a sample from a normal distribution, we could use the chi-square table to obtain the critical value for the rejection region given α.

EXERCISES 7.2

7.2.1. Suppose $X_1, \ldots, X_n$ is a random sample from a normal distribution with a known variance of σ^2 and an unknown mean of μ. Find the most powerful α-level test of $H_0 : \mu = \mu_0$ versus $H_a : \mu = \mu_a$ if **(a)** $\mu_0 > \mu_a$, and **(b)** $\mu_a > \mu_0$.

7.2.2. Show that the most powerful test obtained in Example 7.2.1 is uniformly most powerful for testing $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_a$, but not uniformly most powerful for testing $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$.

7.2.3. Suppose $X_1, \ldots, X_n$ is a random sample from a $U(0, \theta)$ distribution. Find the most powerful α-level test for testing $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_1$, where $\theta_0 < \theta_1$.

7.2.4. Let $X_1, \ldots, X_n$ be a random sample from a geometric distribution with parameter p. Find the most powerful test of $H_0 : p = p_0$ versus $H_a : p = p_a (> p_0)$. Is this uniformly most powerful test for $H_0 : p = p_0$ versus $H_a : p > p_0$?

7.2.5. Let $X_1, \ldots, X_n$ be a random sample from a distribution having a pdf of

$$f(y) = \begin{cases} \frac{2y}{\eta^2} e^{-\frac{y^2}{\eta^2}}, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find a uniformly most powerful test for testing $H_0 : \eta = \eta_0$ versus $H_a : \eta < \eta_0$.

7.2.6. Let X be a single observation from the pdf

$$f(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the most powerful test with a level of significance $\alpha = 0.01$ to test $H_0 : \theta = 3$ versus $H_a : \theta = 4$.

7.2.7. Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p. Find the most powerful test of $H_0 : p = p_0$ versus $H_a : p = p_a$, where $p_a > p_0$.

7.2.8. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with mean λ. Find a best critical region for testing $H_0 : \lambda = 3$ against $H_a : \lambda = 6$.

7.3 LIKELIHOOD RATIO TESTS

The Neyman–Pearson lemma provides a method for constructing most powerful tests for simple hypotheses. We also have seen that in some instances when a hypothesis is not simple, it is possible to find uniformly most powerful tests. In general, uniformly most powerful (UMP) tests do not exist for composite hypotheses. As an example, consider the two-sided hypothesis, at level α, given by

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_a : \mu \neq \mu_0$$

where μ is the mean of a normal population with known variance σ^2. If $\overline{X}$ is the sample mean of a random sample of size n, then as shown earlier, we can use the test statistic

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}.$$

For $H_a : \mu = \mu_1 > \mu_0$, the rejection region for the most powerful test would be

$$\text{Reject } H_0 \text{ if } z > z_\alpha.$$

On the other hand for $H_a : \mu = \mu_2 < \mu_0$, the rejection region for the most powerful test would be

$$\text{Reject } H_0 \text{ if } z < -z_\alpha.$$

Thus, the rejection region depends on the specific alternative. Consequently, the two-sided hypothesis just given has no UMP test.

In this section, we shall study a general procedure that is applicable when one or both H_0 and H_a are composite. In fact, this procedure works for simple hypotheses as well. This method is based on the maximum likelihood estimation and the ratio of likelihood functions used in the Neyman–Pearson lemma. We assume that the pdf or pmf of the random variable X is $f(x, \theta)$, where θ can be one or more unknown parameters. Let Θ represent the total parameter space that is the set of all possible values of the parameter θ given by either H_0 or H_1.

Consider the hypotheses

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_a : \theta \in \Theta_a = \Theta - \Theta_0.$$

where θ is the unknown population parameter (or parameters) with values in Θ, and Θ_0 is a subset of Θ.

Let $L(\theta)$ be the likelihood function based on the sample $X_1, \ldots, X_n$. Now we define the likelihood ratio corresponding to the hypotheses H_0 and H_a. This ratio will be used as a test statistic for the testing procedure that we develop in this section. This is a natural generalization of the ratio test used in the Neyman–Pearson lemma when both hypotheses were simple.

Definition 7.3.1 *The* **likelihood ratio** λ *is the ratio*

$$\lambda = \frac{\max\limits_{\theta \in \Theta_0} L(\theta; x_1, \ldots, x_n)}{\max\limits_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n)} = \frac{L_0^*}{L^*}.$$

We note that $0 \le \lambda \le 1$. Because λ is the ratio of nonnegative functions, $\lambda \ge 0$. Because Θ_0 is a subset of Θ, we know that $\max\limits_{\theta \in \Theta_0} L(\theta) \le \max\limits_{\theta \in \Theta} L(\theta)$. Hence, $\lambda \le 1$.

If the maximum of L in Θ_0 is much smaller as compared with the maximum of L in Θ, that is, if λ is small, it would appear that the data $X_1, \ldots, X_n$ do not support the null hypothesis $\theta \in \Theta_0$. On the other hand, if λ is close to 1, one could conclude that the data support the null hypothesis, H_0. Therefore, small values of λ would result in rejection of the null hypothesis, and large values nearer to 1 will result a decision in support of the null hypothesis.

For the evaluation of λ, it is important to note that $\max_{\theta \in \Theta} L(\theta) = L(\hat{\theta}_{ml.})$, where $\hat{\theta}_{ml.}$ is the maximum likelihood estimator of $\theta \in \Theta$, and $\max_{\theta \in \Theta_0} L(\theta)$ is the likelihood function with unknown parameters replaced by their maximum likelihood estimators subject to the condition that $\theta \in \Theta_0$. We can summarize the likelihood ratio test as follows.

LIKELIHOOD RATIO TESTS (LRTs)

To test

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_a : \theta \in \Theta_a$$

$$\lambda = \frac{\max\limits_{\theta \in \Theta_0} L(\theta; x_1, \ldots, x_n)}{\max\limits_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n)} = \frac{L_0^*}{L*}$$

will be used as the test statistic.

The rejection region for the likelihood ratio test is given by

$$\text{Reject } H_0 \text{ if } \lambda \leq K.$$

K is selected such that the test has the given significance level α.

Example 7.3.1

Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$. Assume that σ^2 is known. We wish to test, at level α, $H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$. Find an appropriate likelihood ratio test.

Solution

We have seen that to test

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_a : \mu \neq \mu_0$$

there is no uniformly most powerful test for this case. The likelihood function is

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}}.$$

Here, $\Theta_0 = \{\mu_0\}$ *and* $\Theta_a = \mathbb{R} - \{\mu_0\}$.

Hence,

$$L_0^* = \max_{\mu = \mu_0} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{2\sigma^2}}.$$

Similarly,

$$L^* = \max_{-\infty<\mu<\infty} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}.$$

Because the only unknown parameter in the parameter space Θ *is* μ, $-\infty < \mu < \infty$, *the maximum of the likelihood function is achieved when* μ *equals its maximum likelihood estimator, that is,*

$$\hat{\mu}_{ml.} = \overline{X}.$$

Therefore, with a simple calculation we have

$$\lambda = \frac{e^{-\left(\sum_{i=1}^{n}(x_i-\mu_0)^2\right)/2\sigma^2}}{e^{-\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)/2\sigma^2}} = e^{-n(\overline{x}-\mu_0)^2/2\sigma^2}.$$

Thus, the likelihood ratio test has the rejection region

$$\text{Reject } H_0 \text{ if } \lambda \le K$$

which is equivalent to

$$-\frac{n}{2\sigma^2}(\overline{X}-\mu_0)^2 \le \ln K \Leftrightarrow$$

$$\frac{(\overline{X}-\mu_0)^2}{\sigma^2/n} \ge 2\ln K \Leftrightarrow$$

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \ge 2\ln K = c_1, \text{ say.}$$

Note that we use the symbol $\Leftrightarrow$ *to mean "if and only if." We now compute* c_1. *Under* H_0, $[(\overline{X}-\mu_0)/(\sigma/\sqrt{n})] \sim N(0,1)$.
Observe that

$$\alpha = P\left\{\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \ge c_1\right\}$$

gives a possible value of c_1 *as* $c_1 = z_{\alpha/2}$. *Hence, LRT for the given hypothesis is*

$$\text{Reject } H_0 \text{ if } \left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \ge z_{a/2}.$$

Thus, in this case, the likelihood ratio test is equivalent to the z*-test for large random samples.*

In fact, when both the hypotheses are simple, the likelihood ratio test is identical to the Neyman–Pearson test. We can now summarize the procedure for the likelihood ratio test, LRT.

PROCEDURE FOR THE LIKELIHOOD RATIO TEST (LRT)

1. Find the largest value of the likelihood $L(\theta)$ for any $\theta_0 \in \Theta_0$ by finding the maximum likelihood estimate within Θ_0 and substituting back into the likelihood function.
2. Find the largest value of the likelihood $L(\theta)$ for any $\theta \in \Theta$ by finding the maximum likelihood estimate within Θ and substituting back into the likelihood function.
3. Form the ratio

$$\lambda = \lambda(x_1, x_2, \ldots, x_n) = \frac{L(\theta) \text{ in } \Theta_0}{L(\theta) \text{ in } \Theta}.$$

4. Determine a K so that the test has the desired probability of type I error, α.
5. Reject H_0 if $\lambda \leq K$.

In the next example, we find a LRT for a testing problem when both H_0 and H_a are simple.

Example 7.3.2

Machine I produces 5% defectives. Machine 2 produces 10% defectives. Ten items produced by each of the machines are sampled randomly; X = number of defectives. Let θ be the true proportion of defectives. Test $H_0 : \theta = 0.05$ versus $H_a : \theta = 0.1$. Use $\alpha = 0.05$.

Solution

We need to test $H_0 : \theta = 0.05$ *vs.* $H_a : \theta = 0.1$. *Let*

$$L(\theta) = \begin{cases} \binom{10}{x}(0.05)^x(0.95)^{10-x}, \text{ if } \theta = 0.05 \\ \binom{10}{x}(0.1)^x(0.90)^{10-x}, \text{ if } \theta = 0.10. \end{cases}$$

And

$$L_1 = L(0.05) = \binom{10}{x}(0.05)^x(0.95)^{10-x}$$

and

$$L_2 = L(0.1) = \binom{10}{x}(0.1)^x(0.90)^{10-x}.$$

Thus, we have

$$\frac{L_1}{L_2} = \frac{0.05^x}{0.1^x}\frac{(0.95)^{10-x}}{(0.9)^{10-x}} = \left(\frac{1}{2}\right)^x\left(\frac{19}{18}\right)^{10-x}.$$

The ratio

$$\lambda = \frac{L_1}{\max(L_1, L_2)}.$$

Note that if $\max(L_1, L_2) = L_1$*, then* $\lambda = 1$*. Because we want to reject for small values of* λ*,* $\max(L_1, L_2) = L_2$*, and we reject* H_0 *if* $(L_1/L_2) \le K$ *or* $(L_2/L_1) > K$ *(note that* $\frac{L_2}{L_1} = 2^x(\frac{18}{19})^{10-x}$*).*
That is, reject H_0 *if*

$$2^x \left(\frac{18}{19}\right)^{10-x} > K$$
$$\Leftrightarrow \left(\frac{2}{\frac{18}{19}}\right)^x > K_1$$
$$\Leftrightarrow \left(\frac{19}{9}\right)^x > K_1.$$

■

Hence, reject H_0 if $X > C$; $P(X > C|H_0 : \theta = 0.05) \le 0.05$.

Using the binomial tables, we have

$$P(X > 2|\theta = 0.05) = 0.0116$$

and

$$P(X \ge 2|\theta = 0.05) = 0.0862.$$

Reject H_0 if $X > 2$. If we want α to be exactly 0.05, we have to use randomized test. Reject with probability $\frac{0.0384}{0.0762} = 0.5039$ if $X = 2$.

The likelihood ratio tests do not always produce a test statistic with a known probability distribution such as the z-statistic of Example 7.3.1. If we have a large sample size, then we can obtain an approximation to the distribution of the statistic λ, which is beyond the level of this book.

EXERCISES 7.3

7.3.1. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$. Assume that σ^2 is unknown. We wish to test, at level α, $H_0 : \mu = \mu_0$ vs $H_a : \mu < \mu_0$. Find an appropriate likelihood ratio test.

7.3.2. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$. Assume that both μ and σ^2 are unknown. We wish to test, at level α, $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_a : \sigma^2 > \sigma_0^2$. Find an appropriate likelihood ratio test.

7.3.3. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu_1, \sigma^2)$ and let $Y_1, Y_2, \ldots, Y_n$ be an independent sample from an $N(\mu_2, \sigma^2)$, where σ^2 is unknown. We wish to test, at level α, $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \ne \mu_2$. Find an appropriate likelihood ratio test.

7.3.4. Let $X_1, \ldots, X_n$ be a sample from a Poisson distribution with parameter λ. Show that a likelihood ratio test of $H_0 : \lambda = \lambda_0$ vs. $H_a : \lambda \ne \lambda_0$ rejects the null hypothesis if $\overline{X} \ge m_1$ or $\overline{X} \le m_2$.

7.3.5. Let $X_1, \ldots, X_n$ be a sample from an exponential distribution with parameter θ. Show that a likelihood ratio test of $H_0 : \theta = \theta_0$ vs. $H_a : \theta \neq \theta_0$ rejects the null hypothesis if $\sum_{i=1}^{n} X_i \geq m_1$ or $\sum_{i=1}^{n} X_i \leq m_2$.

7.3.6. A clinical oncology program developed a set of guidelines for their cancer patients to follow. It is believed that the proportion of patients who are still living after 24 months is greater for those who follow the guidelines. Of the 40 patients who followed the guidelines, 30 are still living after 24 months, whereas of 32 patients who did not follow the guidelines, 21 are living after 24 months. Find a likelihood ratio test at $\alpha = 0.01$ to decide whether the program is effective.

7.4 HYPOTHESES FOR A SINGLE PARAMETER

In this section, we first introduce the concept of p-value. After that, we study hypothesis testing concerning a single parameter.

7.4.1 The p-Value

In hypothesis testing, the choice of the value of α is somewhat arbitrary. For the same data, if the test is based on two different values of α, the conclusions could be different. Many statisticians prefer to compute the so-called p-value, which is calculated based on the observed test statistic. For computing the p-value, it is not necessary to specify a value of α. We can use the given data to obtain the p-value.

Definition 7.4.1 *Corresponding to an observed value of a test statistic, the* **p-value** *(or* **attained significance level***) is the lowest level of significance at which the null hypothesis would have been rejected.*

For example, if we are testing a given hypothesis with $\alpha = 0.05$ and we make a decision to reject H_0 and we proceeded to calculate the p-value equal to 0.03, this means that we could have used an α as low as 0.03 and still maintain the same decision, rejecting H_0.

Based on the alternative hypothesis, one can use the following steps to compute the p-value.

STEPS TO FIND THE p-VALUE

1. Let TS be the test statistic.
2. Compute the value of TS using the sample $X_1, \ldots, X_n$. Say it is a.
3. The p-value is given by

$$p\text{-value} = \begin{cases} P(TS < a|H_0), & \text{if lower tail test} \\ P(TS > a|H_0), & \text{if upper tail test} \\ P(|TS| > |a||H_0), & \text{if two tail test.} \end{cases}$$

Example 7.4.1

To test $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$, suppose that the test statistic Z results in a computed value of 1.58. Then, the p-value $= P(|Z| > 1.58) = 2(0.0571) = 0.1142$. That is, we must have a type I error of 0.1142 in order to reject H_0. Also, if $H_a : \mu > 0$, then the p-value would be $P(Z > 1.58) = 0.0582$. In this case we must have an α of 0.0582 in order to reject H_0.

The p-value can be thought of as a measure of support for the null hypothesis: The lower its value, the lower the support. Typically one decides that the support for H_0 is insufficient when the p-value drops below a particular threshold, which is the significance level of the test.

REPORTING TEST RESULT AS *p*-VALUES

1. Choose the maximum value of α that you are willing to tolerate.
2. If the p-value of the test is less than the maximum value of α, reject H_0.

If the exact p-value cannot be found, one can give an interval in which the p-value can lie. For example, if the test is significant at $\alpha = 0.05$ but not significant for $\alpha = 0.025$, report that $0.025 \leq p\text{-value} \leq 0.05$. So for $\alpha > 0.05$, reject H_0, and for $\alpha < 0.025$, do not reject H_0.

In another interpretation, $1-(p\text{-}value)$ is considered as an index of the strength of the evidence against the null hypothesis provided by the data. It is clear that the value of this index lies in the interval $[0, 1]$. If the p-value is 0.02, the value of index is 0.98, supporting the rejection of the null hypothesis. Not only do p-values provide us with a yes or no answer, they provide a sense of the strength of the evidence against the null hypothesis. The lower the p-value, the stronger the evidence. Thus, in any test, reporting the p-value of the test is a good practice.

Because most of the outputs from statistical software used for hypothesis testing include the p-value, the p-value approach to hypothesis testing is becoming more and more popular. In this approach, the decision of the test is made in the following way. If the value of α is given, and if the p-value of the test is less than the value of α, we will reject H_0. If the value of α is not given and the p-value associated with the test is small (usually set at p-value < 0.05), there is evidence to reject the null hypothesis in favor of the alternative. In other words, there is evidence that the value of the true parameter (such as the population mean) is significantly different (greater, or lesser) than the hypothesized value. If the p-value associated with the test is not small ($p > 0.05$), we conclude that there is not enough evidence to reject the null hypothesis. In most of the examples in this chapter, we give both the rejection region and p-value approaches.

Example 7.4.2

The management of a local health club claims that its members lose on the average 15 pounds or more within the first 3 months after joining the club. To check this claim, a consumer agency took a random sample of 45 members of this health club and found that they lost an average of 13.8 pounds within the first 3 months of membership, with a sample standard deviation of 4.2 pounds.

(a) Find the p-value for this test.

(b) Based on the p-value in (a), would you reject the null hypothesis at $\alpha = 0.01$?

Solution

(a) *Let μ be the true mean weight loss in pounds within the first 3 months of membership in this club. Then we have to test the hypothesis*

$$H_0 : \mu = 15 \text{ versus } H_a : \mu < 15$$

Here $n = 45$, $\bar{x} = 13.8$, and $s = 4.2$. Because $n = 45 > 30$, we can use normal approximation. Hence, the test statistic is

$$z = \frac{13.8 - 15}{4.2/\sqrt{45}} = -1.9166$$

and

$$p\text{-value} = P(Z < -1.9166) \simeq P(Z < -1.92) = 0.0274.$$

Thus, we can use an α as small as 0.0274 and still reject H_0.

(b) *No. Because the p-value $= 0.0274$ is greater than $\alpha = 0.01$, one cannot reject H_0.*

In any hypothesis testing, after an experimenter determines the objective of an experiment and decides on the type of data to be collected, we recommend the following step-by-step procedure for hypothesis testing.

STEPS IN ANY HYPOTHESIS TESTING PROBLEM

1. State the alternative hypothesis, H_a (what is believed to be true).
2. State the null hypothesis, H_0 (what is doubted to be true).
3. Decide on a level of significance α.
4. Choose an appropriate TS and compute the observed test statistic.
5. Using the distribution of TS and α, determine the rejection region(s) (RR).
6. Conclusion: If the observed test statistic falls in the RR, reject H_0 and conclude that based on the sample information, we are $(1 - \alpha)100\%$ confident that H_a is true. Otherwise, conclude that there is not sufficient evidence to reject H_0. In all the applied problems, interpret the meaning of your decision.
7. State any assumptions you made in testing the given hypothesis.
8. Compute the *p-value* from the null distribution of the test statistic and interpret it.

7.4.2 Hypothesis Testing for a Single Parameter

Now we study the testing of a hypothesis concerning a single parameter, θ, based on a random sample $X_1, \ldots, X_n$. Let $\hat{\theta}$ be the sample statistic. First, we deal with tests for the population mean μ for large and small samples. Next, we study procedures for testing the population variance σ^2. We conclude the section by studying a test procedure for the true proportion p.

To test the hypothesis $H : \mu = \mu_0$ concerning the true population mean μ, when we have a large sample ($n \geq 30$) we use the test statistic Z given by

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

where S is the sample standard deviation and μ_0 is the claimed mean under H_0 (if the population variance is known, we replace S with σ.

For a small random sample ($n < 30$), the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

where μ_0 is the claimed value of the true mean, and $\overline{X}$ and S are the sample mean and standard deviation, respectively. Note that we are using the lowercase letters, such as z and t, to represent the observed values of the test statistics Z and T, respectively.

In practice, with raw data, it is important to verify the assumptions. For example, in the small sample case, it is important to check for normality by using normal plots. If this assumption is not satisfied, the nonparametric methods described in Chapter 12 may be more appropriate. In addition, because the sample statistic such as $\overline{X}$ and S will be greatly affected by the presence of outliers, drawing a box plot to check for outliers is a basic practice we should incorporate in our analysis.

We now summarize the typical test of hypothesis for tests concerning population (true) mean.

In order to compute the observed test statistic, z in the large sample case and t in the small sample case, calculate the values of $z = (\bar{x} - \mu_0)/(s/\sqrt{n})$ and $t = [(\bar{x} - \mu_0)/(s/\sqrt{n})]$, respectively.

SUMMARY OF HYPOTHESIS TESTS FOR μ

Large Sample ($n \geq 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{cases} \mu > \mu_0, & \text{upper tail test} \\ \mu < \mu_0, & \text{lower tail test} \\ \mu \neq \mu_0, & \text{two-tailed test} \end{cases}$$

Test statistic: $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

Replace σ by S, if σ is unknown.

$$\text{Rejection region}: \begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR} \end{cases}$$

Small Sample ($n < 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{cases} \mu > \mu_0, & \text{upper tail test} \\ \mu < \mu_0, & \text{lower tail test} \\ \mu \neq \mu_0, & \text{two-tailed test} \end{cases}$$

Test statistic: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

$$\text{RR}: \begin{cases} t > t_{\alpha,n-1}, & \text{upper tail RR} \\ t < -t_{\alpha,n-1}, & \text{lower tail RR} \\ |t| > t_{\alpha/2,n-1}, & \text{two tail RR} \end{cases}$$

Assumption: $n \geq 30$	**Assumption:** Random sample comes from a normal population

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, keep H_0 so that there is not enough evidence to conclude that H_a is true for the given α and more experiments may be needed.

Example 7.4.3

It is claimed that sports-car owners drive on the average 18,000 miles per year. A consumer firm believes that the average mileage is probably lower. To check, the consumer firm obtained information from 40 randomly selected sports-car owners that resulted in a sample mean of 17,463 miles with a sample standard deviation of 1348 miles. What can we conclude about this claim? Use $\alpha = 0.01$.

Solution

Let μ be the true population mean. We can formulate the hypotheses as $H_0 : \mu = 18{,}000$ versus $H_a : \mu < 18{,}000$.

The observed test statistic (for $n \geq 30$) is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \cong \frac{17{,}463 - 18{,}000}{1348/\sqrt{40}}$$

$$= -2.52.$$

Rejection region is $\{z < -z_{0.01}\} = \{z < -2.33\}$.

Decision: *Because $z = -2.52$ is less than -2.33, the null hypothesis is rejected at $\alpha = 0.01$. There is sufficient evidence to conclude that the mean mileage on sport cars is less than 18,000 miles per year.*

Example 7.4.4

In a frequently traveled stretch of the I-75 highway, where the posted speed is 70 mph, it is thought that people travel on the average of at least 75 mph. To check this claim, the following radar measurements of the speeds (in mph) is obtained for 10 vehicles traveling on this stretch of the interstate highway.

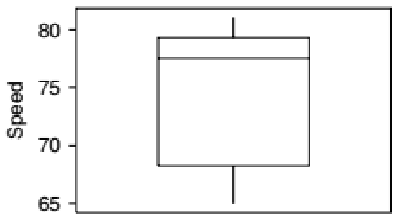
66 74 79 80 69 77 78 65 79 81

Do the data provide sufficient evidence to indicate that the mean speed at which people travel on this stretch of highway is at most 75 mph? Test the appropriate hypothesis using $\alpha = 0.01$. Draw a box plot and normal plot for this data, and comment.

Solution

We need to test

$$H_0 : \mu = 75 \text{ vs. } H_a : \mu > 75$$



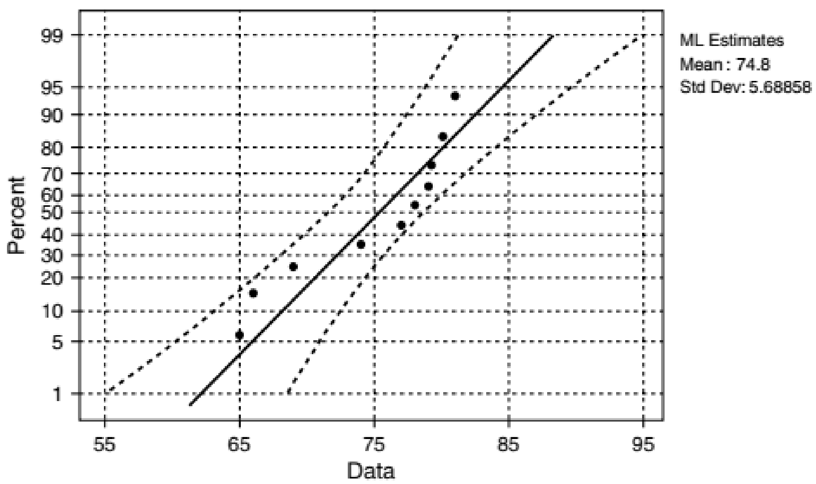
■ **FIGURE 7.1** Box plot of speed data.

For this sample, the sample mean is $\bar{x} = 74.8$ *mph and the standard deviation is* $\sigma = 5.9963$ *mph. Hence, the observed test statistic is*

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{74.8 - 75}{5.9963/\sqrt{10}}$$

$$= -0.10547.$$

From the t*-table,* $t_{0.019} = 2.821$*. Hence, the rejection region is* $\{t > 2.821\}$*.*
Because, $t = -0.10547$ *does not fall in the rejection region, we do not reject the null hypothesis at* $\alpha = 0.01$*.*
Note that we assumed that the vehicles were randomly selected and that collected data follow the normal distribution, because of the small sample size, $n < 30$*, we use the* t*-test.*
Figures 7.1 and 7.2 are the box plot and the normal plot of the data, respectively.



■ **FIGURE 7.2** Normal probability plot for speed.

The box plot suggests that there are no outliers present. However, the normal plot indicates that the normality assumption for this data set is not justified. Hence, it may be more appropriate to do a nonparametric test.

■

Example 7.4.5

In attempting to control the strength of the wastes discharged into a nearby river, an industrial firm has taken a number of restorative measures. The firm believes that they have lowered the oxygen consuming power of their wastes from a previous mean of 450 manganate in parts per million. To test this belief, readings are taken on $n = 20$ successive days. A sample mean of 312.5 and the sample standard deviation 106.23 are obtained. Assume that these 20 values can be treated as a random sample from a normal population. Test the appropriate hypothesis. Use $\alpha = 0.05$.

Solution

Here we need to test the following hypothesis:

$$H_0 : \mu = 450 \text{ vs. } H_a : \mu < 450$$

Given $n = 20$, $\overline{x} = 312.5$, *and* $s = 106.23$. *The observed test statistic is*

$$t = \frac{312.5 - 450}{106.23/\sqrt{20}} = -5.79.$$

The rejection region for $\alpha = 0.05$ *and with 19 degrees of freedom is the set of* t*-values such that*

$$\{t < -t_{0.05,19}\} = \{t < -1.729\}.$$

Decision: *Because* $t = -5.79$ *is less than* -1.729, *reject* H_0. *There is sufficient evidence to confirm the firm's belief.*

For large random samples, the following procedure is used to perform tests of hypotheses about the population proportion, p.

Example 7.4.6

A machine is considered to be unsatisfactory if it produces more than 8% defectives. It is suspected that the machine is unsatisfactory. A random sample of 120 items produced by the machine contains 14 defectives. Does the sample evidence support the claim that the machine is unsatisfactory? Use $\alpha = 0.01$.

Solution

Let Y *be the number of observed defectives. This follows a binomial distribution. However, because* np_0 *and* nq_0 *are greater than 5, we can use a normal approximation to the binomial to test the hypothesis. So we need to test* $H_0 : p = 0.08$ *versus* $H_a : p > 0.08$. *Let the point estimate of* p *be* $\hat{p} = (Y/n) = 0.117$, *the sample proportion. Then the value of the TS is*

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.117 - 0.08}{\sqrt{\frac{(0.08)(0.92)}{120}}} = 0.137.$$

For $\alpha = 0.01$, $z_{0.01} = 2.33$. *Hence, the rejection region is* $\{z > 2.33\}$.

Decision: *Because 0.137 is not greater than 2.33, we do not reject H_0. We conclude that the evidence does not support the claim that the machine is unsatisfactory.*

SUMMARY OF LARGE SAMPLE HYPOTHESIS TEST FOR *p*

To test

$$H_0 : p = p_0$$

versus

$$H_a : \begin{aligned} p &> p_0, && \text{upper tail test} \\ p &< p_0, && \text{lower tail test.} \end{aligned}$$

Test statistic:

$$Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}, \quad \text{where} \quad \sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}}, \quad \text{where} \quad q_0 = 1 - p_0.$$

$$\text{Rejection region} : \begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR,} \end{cases}$$

where z is the observed test statistic.

Assumption: n is large. A good rule of thumb is to use the normal approximation to the binomial distribution only when np_0 and $n(1 - p_0)$ are both greater than 5.

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for given α and more data are needed.

Note that this an approximate test, and the test can be improved by increasing the sample size.

Now we give the procedure for testing the population variance when the samples come from a normal population.

SUMMARY OF HYPOTHESIS TEST FOR THE VARIANCE σ^2

To test

$$H_0 : \sigma^2 = \sigma_0^2$$

versus

$$H_a : \begin{aligned} \sigma^2 &> \sigma_0^2, && \text{upper tail test} \\ \sigma^2 &< \sigma_0^2, && \text{lower tail test} \\ \sigma^2 &\neq \sigma_0^2, && \text{two-tailed test.} \end{aligned}$$

Test statistic:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

where S^2 is the sample variance.
Observed value of test statistic:

$$\frac{(n-1)s^2}{\sigma_0^2}$$

$$\text{Rejection region}: \begin{cases} \chi^2 > \chi^2_{\alpha,n-1}, & \text{upper tail RR} \\ \chi^2 < \chi^2_{1-\alpha,n-1}, & \text{lower tail RR} \\ \chi^2 > \chi^2_{\alpha/2,n-1} \text{ or } \chi^2 < \chi^2_{1-\alpha/2,n-1}, & \text{two tail RR} \end{cases}$$

where $\chi^2_{\alpha,n-1}$ is such that the area under the chi-square distribution with $(n-1)$ degrees of freedom to its right is equal to α.

Assumption: Sample comes from a normal population.

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for given α and more data are needed.

Because the chi-square distribution is not symmetric, the "equal tails" used for the two-sided alternative may not be the best procedure. However, in real-world problems we seldom use a two tail test for the population variance.

Example 7.4.7

A physician claims that the variance in cholesterol levels of adult men in a certain laboratory is at least 100. A random sample of 25 adult males from this laboratory produced a sample standard deviation of cholesterol levels as 12. Test the physician's claim at 5% level of significance.

Solution

To test

$$H_0 : \sigma^2 = 100 \text{ versus } H_a : \sigma^2 < 100$$

for $\alpha = 0.05$, *and 24 degrees of freedom, the rejection region is*

$$RR = \{\chi^2 < \chi^2_{1-\alpha,n-1}\} = \{\chi^2 < 13.484\}.$$

The observed value of the TS is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(24)(144)}{100} = 34.56.$$

Because the value of the test statistic does not fall in the rejection region, we cannot reject H_0 at 5% level of significance. Here, we assumed that the 25 cholesterol measurements follow the normal distribution.

EXERCISES 7.4

7.4.1. A random sample of 50 measurements resulted in a sample mean of 62 with a sample standard deviation 8. It is claimed that the true population mean is at least 64.

(a) Is there sufficient evidence to refute the claim at the 2% level of significance?
(b) What is the p-value?
(c) What is the smallest value of α for which the claim will be rejected?

7.4.2. A machine in a certain factory must be repaired if it produces more than 12% defectives among the large lot of items it produces in a week. A random sample of 175 items from a week's production contains 45 defectives, and it is decided that the machine must be repaired.

(a) Does the sample evidence support this decision? Use $\alpha = 0.02$.
(b) Compute the p-value.

7.4.3. A random sample of 78 observations produced the following sums:

$$\sum_{i=1}^{78} x_i = 22.8, \sum_{i=1}^{78} (x_i - \bar{x})^2 = 2.05.$$

(a) Test the null hypothesis that $\mu = 0.45$ against the alternative hypothesis that $\mu < 0.45$ using $\alpha = 0.01$. Also find the p-value.
(b) Test the null hypothesis that $\mu = 0.45$ against the alternative hypothesis that $\mu \neq 0.45$ using $\alpha = 0.01$. Also find the p-value.
(c) What assumptions did you make for solving (a) and (b)?

7.4.4. Consider the test $H_0 : \mu = 35$ vs. $H_a : \mu > 35$ for a population that is normally distributed.

(a) A random sample of 18 observations taken from this population produced a sample mean of 40 and a sample standard deviation of 5. Using $\alpha = 0.025$, would you reject the null hypothesis?
(b) Another random sample of 18 observations produced a sample mean of 36.8 and a sample standard deviation of 6.9. Using $\alpha = 0.025$, would you reject the null hypothesis?
(c) Compare and discuss the decisions of parts (a) and (b).

7.4.5. According to the information obtained from a large university, professors there earned an average annual salary of \$55,648 in 1998. A recent random sample of 15 professors from this university showed that they earn an average annual salary of \$58,800 with a sample standard deviation of \$8300. Assume that the annual salaries of all the professors in this university are normally distributed.

(a) Suppose the probability of making a type I error is chosen to be zero. Without performing all the steps of test of hypothesis, would you accept or reject the null hypothesis that the current mean annual salary of all professors at this university is \$55,648?

(b) Using the 1% significance level, can you conclude that the current mean annual salary of professors at this university is more than \$55,648?

7.4.6. A check-cashing service company found that approximately 7% of all checks submitted to the service were without sufficient funds. After instituting a random check verification system to reduce its losses, the service company found that only 70 were rejected in a random sample of 1125 that were cashed. Is there sufficient evidence that the check verification system reduced the proportion of bad checks at $\alpha = 0.01$? What is the p-value associated with the test? What would you conclude at the $\alpha = 0.05$ level?

7.4.7. A manufacturer of washers provides a particular model in one of three colors, white, black, or ivory. Of the first 1500 washers sold, it is noticed that 550 were of ivory color. Would you conclude that customers have a preference for the ivory color? Justify your answer. Use $\alpha = 0.01$.

7.4.8. A test of the breaking strength of six ropes manufactured by a company showed a mean breaking strength of 6425 lb and a standard deviation of 120 lb. However, the manufacturer claimed a mean breaking strength of 7500 lb.

(a) Can we support the manufacturer's claim at a level of significance of 0.10?

(b) Compute the p-value. What assumptions did you make for this problem?

7.4.9. A sample of 10 observations taken from a normally distributed population produced the following data:

44	31	52	48	46	39	43	36	41	49

(a) Test the hypothesis that $H_0 : \mu = 44$ vs. $H_a : \mu \neq 44$ using $\alpha = 0.10$. Draw a box plot and normal plot for this data, and comment.

(b) Find a 90% confidence interval for the population mean μ.

(c) Discuss the meanings of (a) and (b). What can we conclude?

7.4.10. The principal of a charter school in Tampa believes that the IQs of its students are above the national average of 100. From the past experience, IQ is normally distributed with a standard deviation of 10. A random sample of 20 students is selected from this school and their IQs are observed. The following are the observed values.

95	91	110	93	133	119	113	107	110	89
113	100	100	124	116	113	110	106	115	113

(a) Test for the normality of the data

(b) Do the IQs of students at the school run above the national average at $\alpha = 0.01$?

7.4.11. In order to find out whether children with chronic diarrhea have the same average hemoglobin level (Hb) that is normally seen in healthy children in the same area, a random sample of 10 children with chronic diarrhea are selected and their Hb levels (g/dL) are obtained as follows.

12.3 11.4 14.2 15.3 14.8 13.8 11.1 15.1 15.8 13.2

Do the data provide sufficient evidence to indicate that the mean Hb level for children with chronic diarrhea is less than that of the normal value of 14.6 g/dL? Test the appropriate hypothesis using $\alpha = 0.01$. Draw a box plot and normal plot for this data, and comment.

7.4.12. A company that manufactures precision special-alloy steel shafts claims that the variance in the diameters of shafts is no more than 0.0003. A random sample of 10 shafts gave a sample variance of 0.00027. At the 5% level of significance, test whether the company's claim can be substantiated.

7.4.13. It was claimed that the average annual expenditures per consumer unit had continued to rise, as measured by the Consumer Price Index annual averages (Bureau of Labor Statistics report, 1995). To test this claim, 100 consumer units were randomly selected in 1995 and found to have an average annual expenditure of \$32,277 with a standard deviation of \$1200. Assuming that the average annual expenditure of all consumer units was \$30,692 in 1994, test at the 5% significance level whether the annual expenditure per consumer unit had really increased from 1994 to 1995.

7.4.14. It is claimed that two of three Americans say that the chances of world peace are seriously threatened by the nuclear capabilities of other countries. If in a random sample of 400 Americans, it is found that only 252 hold this view, do you think the claim is correct? Use $\alpha = 0.05$. State any assumptions you make in solving this problem.

7.4.15. According to the Bureau of Labor Statistics (1996), the average price of a gallon of gasoline in all U.S. cities in the United States in January 1996 was \$1.129. A later random sample in 24 cities found the mean price to be \$1.24 with a standard deviation of 0.01. Test at $\alpha = 0.05$ to see whether the average price of a gallon of gas in the cities had recently changed.

7.4.16. A manufacturer claims that the mean life of batteries manufactured by his company is at least 44 months. A random sample of 40 of these batteries was tested, resulting in a sample mean life of 41 months with a sample standard deviation of 16 months. Test at $\alpha = 0.01$ whether the manufacturer's claim is correct.

7.5 TESTING OF HYPOTHESES FOR TWO SAMPLES

In this section we study the hypothesis testing procedures for comparing the means and variances of two populations. For example, suppose that we want to determine whether a particular drug is effective for a certain illness. The sample subjects will be randomly selected from a large pool of people with that particular illness and will be assigned randomly to the two groups. To one group we will administer a placebo; to the other we will administer the drug of interest. After a period of time, we measure a physical characteristic, say the blood pressure, of each subject that is an indicator of the severity of the illness. The question is whether the drug can be considered effective on the population from which our samples have been selected. We will consider the cases of independent and dependent samples.

7.5.1 Independent Samples

Two random samples are drawn independently of each other from two populations, and the sample information is obtained. We are interested in testing a hypothesis about the difference of the true means. Let $X_{11}, \ldots, X_{1n}$ be a random sample from population 1 with mean μ_1 and variance σ_1^2, and $X_{21}, \ldots, X_{2n}$ be a random sample from population 2 with mean μ_2 and variance σ_2^2. Let $\overline{X}_i, i = 1, 2,$ represent the respective sample means and $S_i^2, i = 1, 2,$ represent the sample variances. In this case, we shall consider following three cases in testing hypotheses about μ_1 and μ_2: **(i)** when σ_1^2 and σ_2^2 are known, **(ii)** when σ_1^2 and σ_2^2 are unknown and $n_1 \geq 30$ and $n_2 \geq 30$, and **(iii)** when σ_1^2 and σ_2^2 are unknown and $n_1 < 30$ and $n_2 < 30$. In case **(iii)** we have the following two possibilities, **(a)** $\sigma_1^2 = \sigma_2^2$, and **(b)** $\sigma_1^2 \neq \sigma_2^2$.

In the large sample case, knowledge of population variances σ_1^2 and σ_2^2 does not make much difference. If the population variances are unknown, we could replace them with sample variances as an approximation. If both $n_1 \geq 30$ and $n_2 \geq 30$ (large sample case), we can use normal approximation. The following box sums up a large sample hypothesis testing procedure for the difference of means for the large sample case.

SUMMARY OF HYPOTHESIS TEST FOR $\mu_1 - \mu_2$ FOR LARGE SAMPLES (n_1 & $n_2 \geq 30$)

To test

$$H_0 : \mu_1 - \mu_2 = D_0$$

versus

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0, & \text{upper tailed test} \\ \mu_1 - \mu_2 < D_0, & \text{lower tailed test} \\ \mu_1 - \mu_2 \neq D_0, & \text{two-tailed test.} \end{cases}$$

The test statistic is

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Replace σ_i by S_i, if $\sigma_i, i = 1,2$ are not known.
Rejection region is

$$RR : \begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR,} \end{cases}$$

where z is the observed test statistic given by

$$z = \frac{\overline{x}_1 - \overline{x}_2 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Assumption: The samples are independent and n_1 and $n_2 \geq 30$.

Decision: Reject H_0, if test statistic falls in the RR and conclude that H_a is true with $(1 - a)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for given α and more experiments are needed.

Example 7.5.1

In a salary equity study of faculty at a certain university, sample salaries of 50 male assistant professors and 50 female assistant professors yielded the following basic statistics.

	Sample mean salary	Sample standard deviation
Male assistant professor	$36,400	360
Female assistant professor	$34,200	220

Test the hypothesis that the mean salary of male assistant professors is more than the mean salary of female assistant professors at this university. Use $\alpha = 0.05$.

Solution

Let μ_1 be the true mean salary for male assistant professors and μ_2 be the true mean salary for female assistant professors at this university. To test

$$H_0 : \mu_1 - \mu_2 = 0 \text{ vs. } H_a : \mu_1 - \mu_2 > 0$$

the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{36,400 - 34,200}{\sqrt{\frac{(360)^2}{50} + \frac{(220)^2}{50}}} = 36.872.$$

The rejection region for $\alpha = 0.05$ is $\{z > 1.645\}$.

Because $z = 36.872 > 1.645$, we reject the null hypothesis at $\alpha = 0.05$. We conclude that the salary of male assistant professors at this university is higher than that of female assistant professors for $\alpha = 0.05$. Note that even though σ_1^2 and σ_2^2 are unknown, because $n_1 \geq 30$ and $n_2 \geq 30$, we could replace σ_1^2 and σ_2^2 by the respective sample variances. We are assuming that the salaries of male and female are sampled independently of each other.

Given next is the procedure we follow to compare the true means from two independent normal populations when n_1 and n_2 are small ($n_1 < 30$ or $n_2 < 30$) and we can assume homogeneity in the population variances, that is, $\sigma_1^2 = \sigma_2^2$. In this case, we pool the sample variances to obtain a point estimate of the common variance.

COMPARISON OF TWO POPULATION MEANS, SMALL SAMPLE CASE (POOLED t-TEST)

To test

$$H_0 : \mu_1 - \mu_2 = D_0$$

versus

$$H_a : \begin{array}{ll} \mu_1 - \mu_2 > D_0, & \text{upper tailed test} \\ \mu_1 - \mu_2 < D_0, & \text{lower tailed test} \\ \mu_1 - \mu_2 \neq D_0, & \text{two-tailed test.} \end{array}$$

The test statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Here the pooled sample variance is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Then the rejection region is

$$RR : \begin{cases} t > t_\alpha, & \text{upper tailed test} \\ t < -t_\alpha, & \text{lower tail test} \\ |t| > t_{\alpha/2}, & \text{two-tailed test} \end{cases}$$

where t is the observed test statistic and t_α is based on $(n_1 + n_2 - 2)$ degrees of freedom, and such that $P(T > t_\alpha) = \alpha$.

Decision: Reject H_0, if test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for given α.

Assumptions: The samples are independent and come from normal populations with means μ_1 and μ_2, and with the (unknown) but equal variances, that is, $\sigma_1^2 = \sigma_2^2$.

Now we shall consider the case where σ_1^2 and σ_2^2 are unknown and cannot be assumed to be equal. In such a case the following test is often used. For the hypothesis

$$H_0 : \mu_1 - \mu_2 = D_0 \text{ vs. } H_0 : \begin{cases} \mu_1 - \mu_2 > D_0 \\ \mu_1 - \mu_2 < D_0 \\ \mu_1 - \mu_2 = D_0 \end{cases}$$

define the test statistic T_ν as

$$T_\nu = \frac{\overline{X}_1 - \overline{X}_2 - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

where T_ν has a t-distribution with ν degrees of freedom, and

$$\nu = \frac{\left[(s_1^2/n_1) + (s_2^2/n_2)\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}.$$

The value of ν will not necessarily be an integer. In that case, we will round it down to the nearest integer. This method of hypothesis testing with unequal variances is called the ***Smith–Satterthwaite*** procedure. Even though this procedure is not widely used, some simulation studies have shown that the Smith–Satterthwaite procedure perform well when variances are unequal and it gives results that are more or less equivalent to those obtained with the pooled t-test when the variances are equal. However, when the sample sizes are approximately equal, the pooled t-test may still be used. Note that in addressing the question which of the cases (iii)(a) or (iii)(b) to use in a given problem, we suggest that if the point estimates S_1^2 of σ_1^2, and S_2^2 of σ_2^2 are approximately the same, then it is logical to assume homogeneity, $\sigma_1^2 = \sigma_2^2$ and use (iii)(a), whereas if S_1^2 and S_2^2 are significantly different we use (iii)(b). More appropriately, we have tests that can be used to test hypotheses concerning $\sigma_1^2 = \sigma_2^2$ or $\sigma_1^2 \neq \sigma_2^2$, known as the F-test, which we discuss at the end of this subsection.

Example 7.5.2

The intelligence quotients (IQs) of 17 students from one area of a city showed a sample mean of 106 with a sample standard deviation of 10, whereas the IQs of 14 students from another area chosen independently showed a sample mean of 109 with a sample standard deviation of 7. Is there a significant difference between the IQs of the two groups at $\alpha = 0.02$? Assume that the population variances are equal.

Solution

We test

$$H_0 : \mu_1 - \mu_2 = 0 \text{ vs. } H_a : \mu_1 - \mu_2 \neq 0$$

Here $n_1 = 17, \overline{x_1} = 106$, *and* $s_1 = 10$. *Also,* $n_2 = 14, \overline{x_2} = 109$, *and* $s_2 = 7$.
We have

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(16)(10)^2 + (13)(7)^2}{29} = 77.138.$$

The test statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2 - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{106 - 109}{\left(\sqrt{77.138}\right)\sqrt{\frac{1}{17} + \frac{1}{14}}} = -0.94644.$$

For $\alpha = 0.02$, $t_{0.01,29} = 2.462$. *Hence, the rejection region is* $t < -2.462$ *or* $t > 2.462$.

Because the observed value of the test statistic, $T = -0.94644$, *does not fall in the rejection region, there is not enough evidence to conclude that the mean IQs are different for the two groups. Here we assume that the two samples are independent and taken from normal populations.*

Example 7.5.3

Assume that two populations are normally distributed with unknown and unequal variances. Two independent samples were drawn from these populations and the data obtained resulted in the following basic statistics:

$$n_1 = 18 \quad \bar{x}_1 = 20.17 \quad s_1 = 4.3$$

$$n_2 = 12 \quad \bar{x}_2 = 19.23 \quad s_2 = 3.8$$

Test at the 5% significance level whether the two population means are different.

Solution

We need to test the hypothesis

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_a : \mu_1 - \mu_2 \neq 0.$$

Here $n_1 = 18, \bar{x}_1 = 20.17$, *and* $s_1 = 4.3$. *Also,* $n_2 = 12$, $\bar{x}_2 = 19.23$, *and* $s_2 = 3.8$.
The degrees of freedom for the **t**-*distribution are given by*

$$\begin{aligned} \nu &= \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \\ &= \frac{\left(\frac{(4.3)^2}{18} + \frac{(3.8)^2}{12}\right)^2}{\frac{\left(\frac{(4.3)^2}{18}\right)^2}{17} + \frac{\left(\frac{(3.8)^2}{12}\right)^2}{11}} = 25.685. \end{aligned}$$

Hence, we have $\nu = 25$ *degrees of freedom. For* $\alpha = 0.05$, $t_{0.025,25} = 2.060$. *Thus, the rejection region is* $t < -2.060$ *or* $t > 2.060$.
The test statistic is given by

$$T_\nu = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{20.17 - 19.23}{\sqrt{\frac{(4.3)^2}{18} + \frac{(3.8)^2}{12}}} = 0.62939.$$

Because the observed value of the test statistic, $T_\nu = 0.62939$, does not fall in the rejection region, we do not reject the null hypothesis. At $\alpha = 0.05$ there is not enough evidence to conclude that the population means are different. Note that the assumptions we made are that the samples are independent and came from two normal populations. No homogeneity assumption is made.

Example 7.5.4

Infrequent or suspended menstruation can be a symptom of serious metabolic disorders in women. In a study to compare the effect of jogging and running on the number of menses, two independent subgroups were chosen from a large group of women, who were similar in physical activity (aside from running), heights, occupations, distribution of ages, and type of birth control methods being used. The first group consisted of a random sample of 26 women joggers who jogged "slow and easy" 5 to 30 miles per week, and the second group consisted of a random sample of 26 women runners who ran more than 30 miles per week and combined long, slow distance with speed work. The following summary statistics were obtained (E. Dale, D. H. Gerlach, and A. L. Wilhite, "Menstrual Dysfunction in Distance Runners," *Obstet. Gynecol.* **54**, 47–53, 1979).

Joggers	$\overline{x}_1 = 10.1,$	$s_1 = 2.1$
Runners	$\overline{x}_2 = 9.1,$	$s_2 = 2.4$

Using $\alpha = 0.05$, **(a)** test for differences in mean number of menses for each group assuming equality of population variances, and **(b)** test for differences in mean number of menses for each group assuming inequality of population variances.

Solution

Here we need to test

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_a : \mu_1 - \mu_2 \neq 0.$$

Here, $n_1 = 26, \overline{x}_1 = 10.1$, and $s_1 = 2.1$. Also, $n_2 = 26, \overline{x}_2 = 9.1$, and $s_2 = 2.4$.

(a) *Under the assumption $\sigma_1^2 = \sigma_2^2$, we have*

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(25)(2.1)^2 + (25)(2.4)^2}{50} = 5.085.$$

The test statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2 - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{10.1 - 9.1}{\left(\sqrt{5.085}\right)\sqrt{\frac{1}{26} + \frac{1}{26}}} = 1.5989.$$

For $\alpha = 0.05$, $t_{0.025,50} \approx 1.96$. *Hence, the rejection region is* $t < -1.96$ *and* $t > 1.96$. *Because* $T = 1.589$ *does not fall in the rejection region, we do not reject the null hypothesis. At* $\alpha = 0.05$ *there is not enough evidence to conclude that the population mean number of menses for joggers and runners are different.*

(b) *Under the assumption* $\sigma_1^2 \neq \sigma_2^2$, *we have*

$$\nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

$$= \frac{\left(\frac{(2.1)^2}{26} + \frac{(2.4)^2}{26}\right)^2}{\frac{\left(\frac{(2.1)^2}{26}\right)^2}{25} + \frac{\left(\frac{(2.4)^2}{26}\right)^2}{25}} = 49.134.$$

Hence, we have $\nu = 49$ *degrees of freedom. Because this value is large, the rejection region is still approximately* $t < -1.96$ *and* $t > 1.96$. *Hence, the conclusion is the same as that of part (a). In both parts (a) and (b), we assumed that the samples are independent and came from two normal populations.*

■

Now we present the summary of the test procedure for testing the difference of two proportions, inherent in two binomial populations. Here, again we assume that the binomial distribution is approximated by the normal distribution and thus it is an approximate test.

SUMMARY OF HYPOTHESIS TEST FOR $(p_1 - p_2)$ FOR LARGE SAMPLES $(n_i p_i > 5$ AND $n_i q_i > 5$, FOR $i = 1, 2)$

To test

$$H_0 : p_1 - p_2 = D_0$$

versus

$$H_a : \begin{cases} p_1 - p_2 < D_0, & \text{upper tailed test} \\ p_1 - p_2 > D_0, & \text{lower tailed test} \\ p_1 - p_2 \neq D_0, & \text{two-tailed test} \end{cases}$$

at significance level α, the test statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

where z is the observed value of Z.

The rejection region is

$$RR : \begin{cases} z > z_{\alpha}, & \text{upper tailed RR} \\ z < -z_{\alpha}, & \text{lower tailed RR} \\ |z| > z_{\alpha/2}, & \text{two-tailed RR} \end{cases}$$

Assumption: The samples are independent and

$$n_i p_i > 5 \text{ and } n_i q_i > 5, \text{ for } i = 1,2.$$

Decision: Reject H_0 if the test statistic falls in the RR and conclude that H_a is true with $(1 - a)100\%$ confidence. Otherwise, do not reject H_0, because there is not enough evidence to conclude that H_a is true for given α and more experiments are needed.

Example 7.5.5

Because of the impact of the global economy on a high-wage country such as the United States, it is claimed that the domestic content in manufacturing industries fell between 1977 and 1997. A survey of 36 randomly picked U.S. companies gave the proportion of domestic content total manufacturing in 1977 as 0.37 and in 1997 as 0.36. At the 1% level of significance, test the claim that the domestic content really fell during the period 1977–1997.

Solution

Let p_1 be the domestic content in 1977 and p_2 be the domestic content in 1997.
Given $n_1 = n_2 = 36$, $\hat{p}_1 = 0.37$ and $\hat{p}_2 = 0.36$. We need to test

$$H_0 : p_1 - p_2 = 0 \text{ vs. } H_a : p_1 - p_2 > 0.$$

The test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_2}{n_1} + \frac{\hat{p}_1\hat{q}_2}{n_2}}}$$

$$= \frac{0.37 - 0.36}{\sqrt{\frac{(0.37)(0.63)}{36} + \frac{(0.36)(0.64)}{36}}} = 0.08813.$$

For $\alpha = 0.01$, $z_{0.01} = 2.325$. Hence, the rejection region is $z > 2.325$.
Because the observed value of the test statistic does not fall in the rejection region, at $\alpha = 0.01$, there is not enough evidence to conclude that the domestic content in manufacturing industries fell between 1977 and 1997.

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be two independent random samples from two normal populations with sample variances s_1^2 and s_2^2, respectively. The problem here is of testing for the equality of the variances, $H_0 : \sigma_1^2 = \sigma_2^2$. We have already seen in Chapter 4 that

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

follows the F-distribution with $\nu_1 = n_1 - 1$ numerator and $\nu_2 = n_2 - 1$ degrees of freedom. Under the assumption $H_0 : \sigma_1^2 = \sigma_2^2$, we have

$$F = \frac{S_1^2}{S_2^2}$$

which has an F-distribution with (ν_1, ν_2) degrees of freedom. We summarize the test procedure for the equality of variances.

TESTING FOR THE EQUALITY OF VARIANCES

To test

$$H_0 : \sigma_1^2 = \sigma_2^2$$

versus

$$H_a : \begin{array}{ll} \sigma_1^2 > \sigma_2^2, & \text{lower tailed test} \\ \sigma_1^2 < \sigma_2^2, & \text{upper tailed test} \\ \sigma_1^2 \neq \sigma_2^2, & \text{two-tailed test} \end{array}$$

at significance level α, the test statistic is

$$F = \frac{S_1^2}{S_2^2}.$$

The rejection region is

$$RR : \begin{cases} f > F_\alpha(\nu_1,\nu_2), & \text{upper tailed RR} \\ f < F_{1-\alpha}(\nu_1,\nu_2), & \text{lower tailed RR} \\ f > F_{\alpha/2}(\nu_1,\nu_2) \text{ or } f < F_{1-\alpha/2}(\nu_1,\nu_2), & \text{two-tailed RR} \end{cases}$$

where f is the observed test statistic given by $f = \frac{s_1^2}{s_2^2}$.

Decision: Reject H_0 if the test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, keep H_0, because there is not enough evidence to conclude that H_a is true for a given α and more experiments are needed.

Assumption:

(i) The two random samples are independent.
(ii) Both populations are normal.

Recall from Section 4.2 that in order to find $F_{1-\alpha}(\nu_1, \nu_2)$, we use the identity $F_{1-\alpha}(\nu_1, \nu_2) = (1/F_\alpha(\nu_2, \nu_1))$.

Example 7.5.6

Consider two independent random samples $X_1, \ldots, X_n$ from an $N(\mu_1, \sigma_1^2)$ distribution and $Y_1, \ldots, Y_n$ from an $N(\mu_2, \sigma_2^2)$ distribution. Test $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_a : \sigma_1^2 \neq \sigma_2^2$ for the following basic statistics:

$$n_1 = 25, \overline{x_1} = 410, s_1^2 = 95, \text{ and } n_2 = 16, \overline{x_2} = 390, s_2^2 = 300$$

Use $\alpha = 0.20$.

Solution

Test $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_a : \sigma_1^2 \neq \sigma_2^2$. This is a two-tailed test.
Here the degrees of freedom are $\nu_1 = 24$ and $\nu_2 = 15$. The test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{95}{300} = 0.317.$$

From the F-table, $F_{0.10}(24, 15) = 1.90$ and $F_{0.90}(24, 15) = (1/F_{0.10}(15, 24)) = 0.50$.

Hence, the rejection region is $F > 1.90$ or $F < 0.56$. Because the observed value of the test statistic, 0.317, is less than 0.56, we reject the null hypothesis. There is evidence that the population variances are not equal.

7.5.2 Dependent Samples

We now consider the case where the two random samples are not independent. When two samples are dependent (the samples are dependent if one sample is related to the other), then each data point in one sample can be coupled in some natural, nonrandom fashion with each data point in the second sample. This situation occurs when each individual data point within a sample is paired (matched) to an individual data point in the second sample. The pairing may be the result of the individual observations in the two samples: (1) representing before and after a program (such as weight before and after following a certain diet program), (2) sharing the same characteristic, (3) being matched by location, (4) being matched by time, (5) control and experimental, and so forth. Let (X_{1i}, X_{2i}), for $i = 1, 2, \ldots, n$, be a random sample. X_{1i}, and X_{2j} $(i \neq j)$ are independent. To test the significance of the difference between two population means when the samples are dependent, we first calculate for each pair of scores the difference, $D_i = X_{1i} - X_{2i}$, $i = 1, 2, \ldots, n$, between the two scores. Let $\mu_D = E(D_i)$. Because pairs of observations form a random sample $D_1, \ldots, D_n$ are independent and identically distributed random variables, if $d_1, \ldots, d_n$ are the observed values of $D_1, \ldots, D_n$, then we define

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2 = \frac{\sum_{i=1}^{n} d_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2}{n-1}.$$

Now the testing for these n observed differences will proceed as in the case of a single sample. If the number of differences is large $(n \geq 30)$, large sample inferential methods for one sample case can be used for the paired differences. We now summarize the hypothesis testing procedure for small samples.

SUMMARY OF TESTING FOR MATCHED PAIRS EXPERIMENT

To test

$$H_0 : \mu_D = d_0 \text{ versus } H_a : \begin{array}{ll} \mu_D > d_0, & \text{upper tail test} \\ \mu_D < d_0, & \text{lower tail test} \\ \mu_D \neq d_0, & \text{two-tailed test} \end{array}$$

the test statistic: $T = \frac{\overline{D}-D_0}{S_D/\sqrt{n}}$ (this approximately follows a Student t-distribution with $(n-1)$ degrees of freedom).

The rejection region is

$$\begin{cases} t > t_{\alpha,n-1}, & \text{upper tail RR} \\ t < -t_{\alpha,n-1}, & \text{lower tail RR} \\ |t| > t_{\alpha/2,n-1}, & \text{two-tailed RR} \end{cases}$$

where t is the observed test statistic.

Assumptions: The differences are approximately normally distributed.

Decision: Reject H_0 if the test statistic falls in the RR and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0, because there is not enough evidence to conclude that H_a is true for a given α and more data are needed.

Example 7.5.7

A new diet and exercise program has been advertised as remarkable way to reduce blood glucose levels in diabetic patients. Ten randomly selected diabetic patients are put on the program, and the results after 1 month are given by the following table:

Before	268	225	252	192	307	228	246	298	231	185
After	106	186	223	110	203	101	211	176	194	203

Do the data provide sufficient evidence to support the claim that the new program reduces blood glucose level in diabetic patients? Use $\alpha = 0.05$.

Solution

We need to test the hypothesis

$$H_0 : \mu_D = 0 \quad \text{vs.} \quad H_a : \mu_D < 0.$$

First we calculate the difference of each pair given in the following table.

Before	268	225	252	192	307	228	246	298	231	185
After	106	186	223	110	203	101	211	176	194	203
Difference (after−before)	−162	−39	−29	−82	−104	−127	−35	−122	−37	18

From the table, the mean of the differences is $\overline{d} = -71.9$ *and the standard deviation* $s_d = 56.2$. *The test statistic is*

$$t = \frac{\overline{d} - d_0}{s_d/\sqrt{n}} = \frac{-71.9}{56.2/\sqrt{10}} = -4.0457 \approx -4.05.$$

From the t-table, $t_{0.05,9} = 1.833$. *Because the observed value of* $t = -4.05 < -t_{0.05,9} = -1.833$, *we reject the null hypothesis and conclude that the sample evidence suggests that the new diet and exercise program is effective.*

We can also obtain a $(1-\alpha)100\%$ confidence interval for μ_D using the formula

$$\left(\overline{D} - t_{\alpha/2}\frac{S_d}{\sqrt{n}}, \overline{D} + t_{\alpha/2}\frac{S_d}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ is obtained from the t-table with $(n-1)$ degrees of freedom. The interpretation of the confidence interval is identical to the earlier interpretation.

Example 7.5.8

For the data in Example 7.5.7, obtain a 95% confidence interval for μ_D and interpret its meaning.

Solution

We have already calculated $\overline{d} = -71.9$ *and* $s_d = 56.2$. *From the t-table,* $t_{0.025,9} = 2.262$. *Hence, a 95% confidence interval for* μ_D *is* $(-112.1, -31.7)$. *That is,* $P(-112.1 \le \mu_D \le -31.7) = 0.95$. *Note that* $\mu_D = \mu_1 - \mu_2$, *and from the confidence limits we can conclude with 95% confidence that* μ_2 *is always greater than* μ_1, *that is,* $\mu_2 > \mu_1$.

It is interesting to compare the matched pairs test with the corresponding two independent sample test. One of the natural questions is, why must we take paired differences and then calculate the mean and standard deviation for the differences—why can't we just take the difference of means of each sample, as we did for independent samples? The answer lies in the fact that $\sigma^2_{\overline{D}}$ need not be equal to $\sigma^2_{(\overline{X}_1 - \overline{X}_2)}$. Assume that

$$E(X_{ji}) = \mu_j, \; Var(X_{ji}) = \sigma_j^2, \; \text{for } j = 1, 2,$$

and

$$Cov(X_{1i}, X_{2i}) = \rho\sigma_1\sigma_2$$

where ρ denotes the assumed common correlation coefficient of the pair (X_{1i}, X_{2i}) for $i = 1, 2, \ldots, n$. Because the values of D_i, $i = 1, 2, \ldots, n$, are independent and identically distributed,

$$\mu_D = E(D_i) = E(X_{1i}) - E(X_{2i}) = \mu_1 - \mu_2$$

and

$$\sigma_D^2 = Var(D_i) = Var(X_{1i}) + Var(X_{2i}) - 2Cov(X_{1i}, X_{2i})$$
$$= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2.$$

From these calculations,

$$E(\overline{D}) = \mu_D = \mu_1 - \mu_2$$

and

$$\sigma_{\overline{D}}^2 = Var(\overline{D}) = \frac{\sigma_D^2}{n} = \frac{1}{n}(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2).$$

Now, if the samples were independent with $n_1 = n_2 = n$,

$$E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$$

and

$$\sigma^2_{(\overline{X}_1 - \overline{X}_2)} = \frac{1}{n}(\sigma_1^2 + \sigma_2^2).$$

Hence, if $\rho > 0$, then $\sigma_{\overline{D}}^2 < \sigma^2_{(\overline{X}_1 - \overline{X}_2)}$. As a result, we can see that the matched pairs test reduces any variability introduced by differences in physical factors in comparison to the independent samples test when $\rho > 0$. It is also important to observe that normality assumption for the difference does not imply that the individual samples themselves are normal. Also, in a matched pairs experiment, there is no need to assume the equality of variances for the two populations. Matching also reduces degrees of freedom, because in case of two independent samples, the degrees of freedom is $(n_1 + n_2 - 2)$, whereas for the case of two dependent samples it is only $(n - 1)$.

EXERCISES 7.5

7.5.1. Two sets of elementary school children were taught to read by different methods, 50 by each method. At the conclusion of the instructional period, a reading test gave results $\overline{y_1} = 74$, $\overline{y_2} = 71$, $s_1 = 9$, and $s_2 = 10$. What is the attained significance level if you wish to see if there is evidence of a real difference between the two population means? What would you conclude if you desired an α-value of 0.05?

7.5.2. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal variances.

Sample 1	14	15	11	14	10	8	13	10	12	16	15		
Sample 2	17	16	21	12	20	18	16	14	21	20	13	20	13

Test at the 2% significance level whether μ_1 is lower than μ_2.

7.5.3. In the academic year 1997–1998, two random samples of 25 male professors and 23 female professors from a large university produced a mean salary for male professors of \$58,550 with a standard deviation of \$4000 and an average for female professors of \$53,700 with a standard deviation of \$3200. At the 5% significance level, can you conclude that the mean salary of all male professors for 1997–1998 was higher than that of all female professors? Assume that the salaries of male and female professors are both normally distributed with equal standard deviations.

7.5.4. It is believed that the effects of smoking differ depending on race. The following table gives the results of a statistical study for this question.

	Number in the study	Average number of cigarettes per day	Number of lung cancer cases
Whites	400	15	78
African Americans	280	15	70

Do the data indicate that African Americans are more likely to develop lung cancer due to smoking? Use $\alpha = 0.05$.

7.5.5. A supermarket chain is considering two sources A and B for the purchase of 50-pound bags of onions. The following table gives the results of a study.

	Source A	Source B
Number of bags weighed	80	100
Mean weight	105.9	100.5
Sample variance	0.21	0.19

Test at $\alpha = 0.05$ whether there is a difference in the mean weights.

7.5.6. In order to compare the mean Hemoglobin (Hb) levels of well-nourished and undernourished groups of children, random samples from each of these groups yielded the following summary.

	Number of children	Sample mean	Sample standard deviation
Well nourished	95	11.2	0.9
Undernourished	75	9.8	1.2

Test at $\alpha = 0.01$ whether the mean Hb levels of well-nourished children were higher than those of undernourished children.

7.5.7. An aquaculture farm takes water from a stream and returns it after it has circulated through the fish tanks. In order to find out how much organic matter is left in the waste water after the circulation, some samples of the water are taken at the intake and other samples are taken at the downstream outlet and tested for biochemical oxygen demand (BOD). BOD is a common environmental measure of the quantity of oxygen consumed by microorganisms during the decomposition of organic matter. If BOD increases, it can be said that the waste matter contains more organic matter than the stream can handle. The following table gives data for this problem.

Upstream	9.0	6.8	6.5	8.0	7.7	8.6	6.8	8.9	7.2	7.0
Downstream	10.2	10.2	9.9	11.1	9.6	8.7	9.6	9.7	10.4	8.1

Assuming that the samples come from a normal distribution,

(a) Test that the mean BOD for the downstream samples is less than for the samples upstream at $\alpha = 0.05$. Assume that the variances are equal.

(b) Test for the equality of the variances at $\alpha = 0.05$.

(c) In parts (a) and (b), we assumed samples are independent. Now, we feel this assumption is not reasonable. Assuming that the difference of each pair is approximately normal, test that the mean BOD for the downstream samples is less than for the upstream samples at $\alpha = 0.05$.

7.5.8. Suppose we want to know the effect on driving of a drug for cold and allergy, in a study in which the same people were tested twice, once after 1 hour of taking the drug and once when no drug is taken. Suppose we obtain the following data, which represent the number of cones (placed in a certain pattern) knocked down by each of the nine individuals before taking the drug and after an hour of taking the drug.

No drug	0	0	3	2	0	0	3	3	1
After drug	1	5	6	5	5	5	6	1	6

Assuming that the difference of each pair is coming from an approximately normal distribution, test if there is any difference in the individuals' driving ability under the two conditions. Use $\alpha = 0.05$.

7.5.9. Suppose that we want to evaluate the role of intravenous pulse cyclophosphamide (IVCP) infusion in the management of nephrotic syndrome in children with steroid resistance. Children were given a monthly infusion of IVCP in a dose of 500 to 750 mg/m^2. The following data (source: S. Gulati and V. Kher, "Intravenous pulse cyclophosphamide—A new regime for steroid resistant focal segmental glomerulosclerosis," *Indian Pediatr*. **37**, 2000) represent levels of serum albumin (g/dL) before and after IVCP in 14 randomly selected children with nephrotic syndrome.

Pre-IVCP	2.0	2.5	1.5	2.0	2.3	2.1	2.3	1.0	2.2	1.8	2.0	2.0	1.5	3.4
Post-IVCP	3.5	4.3	4.0	4.0	3.8	2.4	3.5	1.7	3.8	3.6	3.8	3.8	4.1	3.4

Assuming that the samples come from a normal distribution:

(a) Test whether the mean Pre-IVCP is less than the mean Post-IVCP at $\alpha = 0.05$. Assume that the variances are equal.

(b) Test for the equality of the variances at $\alpha = 0.05$.

(c) In parts (a) and (b), we assumed that the samples are independent. Now, we feel this assumption is not reasonable. Assuming that the difference of each pair is approximately normal, test that the mean Pre-IVCP is less than the Post-IVCP at $\alpha = 0.05$.

7.5.10. Show that S_D^2 is an unbiased estimator of σ_D^2.

7.5.11. Test $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_a : \sigma_1^2 \neq \sigma_2^2$ for the following data.

$$n_1 = 10, \overline{x_1} = 71, s_1^2 = 64 \quad \text{and} \quad n_2 = 25, \overline{x_2} = 131, s_2^2 = 96.$$

Use $\alpha = 0.10$.

7.5.12. The IQs of 17 students from one area of a city showed a mean of 106 with a standard deviation of 10, whereas the IQs of 14 students from another area showed a mean of 109 with a standard deviation of 7. Test for equality of variances between the IQs of the two groups at $\alpha = 0.02$.

7.5.13. The following data give SAT mean scores for math by state for 1989 and 1999 for 20 randomly selected states (source: *The World Almanac and Book of Facts 2000*).

State	1989	1999
Arizona	523	525
Connecticut	498	509
Alabama	539	555
Indiana	487	498
Kansas	561	576
Oregon	509	525
Nebraska	560	571
New York	496	502
Virginia	507	499
Washington	515	526
Illinois	539	585
North Carolina	469	493
Georgia	475	482
Nevada	512	517
Ohio	520	568
New Hampshire	510	518

Assuming that the samples come from a normal distribution:

(a) Test that the mean SAT score for math in 1999 is greater than that in 1989 at $\alpha = 0.05$. Assume the variances are equal.

(b) Test for the equality of the variances at $\alpha = 0.05$.

7.6 CHI-SQUARE TESTS FOR COUNT DATA

In this section, we study several commonly used tests for count data. These are basically large sample tests based on a χ^2-approximation. Suppose that we have outcomes of a multinomial experiment that consists of K mutually exclusive and exhaustive events $A_1, \ldots, A_k$. Let $P(A_i) = p_i$, $i = 1, 2, \ldots, k$. Then $\sum_{i=1}^{n} p_i = 1$. Let the experiment be repeated n times, and let $X_i (i = 1, 2, \ldots, k)$ represent the number of times the event A_i occurs. Then $(X_1, \ldots, X_k)$ have a multinomial distribution with parameters $n, p_1, \ldots, p_k$.

Let

$$Q^2 = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{(X_i - np_i)^2}.$$

It can be shown that for large n, the random variable Q^2 is approximately χ^2-distributed with $(k-1)$ degrees of freedom. It is usual to demand $np_i \geq 5$ $(i = 1, 2, \ldots, k)$ for the approximation to be valid, although the approximation generally works well if for only a few values of i (about 20%), $np_i \geq 1$ and the rest (about 80%) satisfy the condition $np_i \geq 5$. This statistic was proposed by Karl Pearson in 1900.

It should be noted that the χ^2-tests that we discuss in this section are approximate tests valid for large samples. Often X_i is called the observed frequency and is denoted by O_i (this is the observed value in class i), and np_i is called the expected frequency and is denoted by E_i (this is the theoretical distribution frequency under the null hypothesis). Thus, with these notations, we get

$$Q^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

Example 7.6.1

A plant geneticist grows 200 progeny from a cross that is hypothesized to result in a 3 : 1 phenotypic ratio of red-flowered to white-flowered plants. Suppose the cross produces 170 red- to 30 white-flowered plants. Calculate the value of Q^2 for this experiment.

Solution

There are two categories of data totaling $n = 200$. *Hence,* $k = 2$. *Let* $i = 1$ *represent red-flowered and* $i = 2$ *represent white-flowered plants. Then* $O_1 = 170$, *and* $O_2 = 30$.
Here, H_0 : *The flower color population ratio is not different from 3 : 1, and the alternate is* H_a : *The flower color population sampled has a flower color ratio that is not 3 red : 1 white.*
Under the null hypothesis, the expected frequencies are $E_1 = (200)(3/4) = 150$, *and* $E_2 = (200)(1/4) = 50$. *Hence,*

$$\begin{aligned} Q^2 &= \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(170-150)^2}{150} + \frac{(30-50)^2}{50} = 10.667. \end{aligned}$$

The type of calculation in Example 7.6.1 gives a measure of how close our observed frequencies come to the expected frequencies and is referred to as a measure of *goodness of fit*. Smaller values of Q^2 values indicate better fit.

One of the most frequent uses of the χ^2-test is in comparison of observed frequencies. Unless the sample size is exactly 100, percentages cannot be used. These are approximate tests. Let the random variables $(X_1, \ldots, X_k)$ have a multinomial distribution with parameters $n, p_1, \ldots, p_k$. Let n be known. We will now present some important tests based on the chi-square statistic.

7.6.1 Testing the Parameters of Multinomial Distribution: Goodness-of-Fit Test

Let an experiment have k mutually exclusive and exhaustive outcomes $A_1, A_2, \ldots, A_k$. We would like to test the null hypothesis that all the $p_i = p(A_i), i = 1, 2, \ldots, k$ are equal to known numbers $p_{i0}, i = 1, 2, \ldots, k$. We now summarize the test procedure.

TESTING THE PARAMETERS OF A MULTINOMIAL DISTRIBUTION (SUMMARY)

To test

$$H_0 : p_1 = p_{10}, \ldots, p_k = p_{k0}$$

versus

$$H_a : \text{At least one of the probabilities is different from the hypothesized value.}$$

The test is always a one-sided upper tail test.

Let O_i be the observed frequency, $E_i = np_{i0}$ be the expected frequency (frequency under the null hypothesis), and k be the number of classes. The test statistic is

$$Q^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

The test statistic Q^2 has an approximate chi-square distribution with $k - 1$ degrees of freedom.

The rejection region is

$$Q^2 \geq \chi^2_{\alpha,k-1}.$$

Assumption: $E_i \geq 5$: Exact methods are available. Computing the power of this test is difficult.

This test is known as the goodness-of-fit test. It implies that if the observed data are very close to the expected data, we have a very good fit and we accept the null hypothesis. That is, for small Q^2 values, we accept H_0.

Example 7.6.2

A TV station broadcasts a series of programs on the ill effects of smoking marijuana. After the series, the station wants to know whether people have changed their opinion about legalizing marijuana. Given in the following tables are the data based on a survey of 500 randomly chosen people:

Before the Series Was Shown

For legalization	Decriminalization	Existing law (fine or imprisonment)	No opinion
7%	18%	65%	10%

After the Series Was Shown

For legalization	Decriminalization	Existing law (fine or imprisonment)	No opinion
39%	9%	36%	16%

Here, $n = 4$, and we wish to test

$$H_0 : p_1 = 0.07; \;\; p_2 = 0.18; \;\; p_3 = 0.65; \;\; p_4 = 0.1$$

versus

H_a : At least one of the probabilities is different from the hypothesized value.

The test is always an upper tail test. Test this hypothesis using $\alpha = 0.01$.

Solution

We have

$$E_1 = (500)(0.07) = 35; E_2 = 90; E_3 = 325; E_4 = 50.$$

The observed frequencies are

$$O_1 = (500)(0.39) = 195; O_2 = 45; O_3 = 180; O_4 = 80.$$

The test statistic is

$$\begin{aligned} Q^2 &= \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \\ &= \left[\frac{(195-35)^2}{35} + \frac{(45-90)^2}{90} + \frac{(180-325)^2}{325} + \frac{(80-50)^2}{50} \right] \\ &= 836.62. \end{aligned}$$

From the χ^2*-table,* $\chi^2_{0.01,3} = 11.3449$*. Because the test statistic* $Q^2 = 836.62 > 11.3449$*, we reject* H_0 *at* $\alpha = 0.01$*. Hence, the data suggest that people have changed their opinion after the series on the ill effects of smoking marijuana was shown.*

■

Example 7.6.3

A die is rolled 60 times and the face values are recorded. The results are as follows.

Up face	1	2	3	4	5	6
Frequency	8	11	5	12	15	9

Is the die balanced? Test using $\alpha = 0.05$.

Solution

If the die is balanced, we must have

$$p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$$

where $p_i =$ *P(face value on the die is i),* $i = 1, 2, \ldots, 6$. *This has the discrete uniform distribution. Hence,*

$$H_0 : p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$$

versus

H_a *: At least one of the probabilities is different from the hypothesized value of 1/6*

$$E_1 = n_1 p_1 = (60)(1/6) = 10, \ldots, E_6 = 10.$$

We summarize the calculations in the following table:

Face value	1	2	3	4	5	6
Frequency, O_i	8	11	5	12	15	9
Expected value, E_i	10	10	10	10	10	10

The test statistic value is given by

$$Q^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 6.$$

From the chi-square table with 5 d.f., $\chi^2_{0.05,5} = 11.070$.

Because the value of the test statistic does not fall in the rejection region, we do not reject H_0. *Therefore, we conclude that the die is balanced.*

7.6.2 Contingency Table: Test for Independence

One of the uses of the χ^2-statistic is in contingency (dependence) testing where n randomly selected items are classified according to two different criteria, such as when data are classified on the basis of two factors (row factor and column factor) where the row factor has r levels and the column factor has c levels. The obtained data are displayed as shown in the following table, where n_{ij} represents the number of data values under row i and column j. Our interest here is to test for independence of two methods of classification of observed events. For example, we might classify a sample of students by sex and by their grade on a statistics course in order to test the hypothesis that the grades are dependent on sex. More generally the problem is to investigate a *dependency* (or *contingency*) between two classification criteria.

		Levels of column factor				
		1	2	…	c	Row total
Row	1	n_{11}	n_{12}		n_{1c}	n_1
levels	2	n_{21}	n_{21}		n_{2c}	n_2
	.					
	.					
	r	n_{r1}	n_{r2}		an_{rc}	n_r
Column total		$n_{.1}$	$n_{.2}$		$n_{.c}$	N

where $N = \sum_{j=1}^{c} n_{.j} = \sum_{i=1}^{r} n_{i.} = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}$ is the grand total.

We wish to test the hypothesis that the two factors are independent. We summarize the procedure in the following table for testing that the factors represented by the rows are independent with that represented by the columns.

TESTING FOR THE INDEPENDENCE OF TWO FACTORS

To test

$$H_0 : \text{The factors are independent}$$

versus

$$H_a : \text{The factors are dependent}$$

the test statistic is,

$$Q^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where

$$O_{ij} = n_{ij}$$

and

$$E_{ij} = \frac{n_i n_j}{N}.$$

Then under the null hypothesis the test statistic Q^2 has an approximate chi-square distribution with $(r-1)(c-1)$ degrees of freedom.
Hence, the rejection region is $Q^2 > \chi^2_{\alpha,(r-1)(c-1)}$.

Assumption: $E_{ij} \geq 5$.

Example 7.6.4

The following table gives a classification according to religious affiliation and marital status for 500 randomly selected individuals.

		Religious affiliation					
		A	B	C	D	None	**Total**
Marital status	Single	39	19	12	28	18	116
	With spouse	172	61	44	70	37	384
	Total	211	80	56	98	55	**500**

For $\alpha = 0.01$, test the null hypothesis that marital status and religious affiliation are independent.

Solution

We need to test the hypothesis

$$H_0 : \textit{Marital status and religious affiliation are independent}$$

versus

$$H_a : \textit{Marital status and religious affiliation are dependent.}$$

Here, $c = 5$, *and* $r = 2$. *For* $\alpha = 0.01$, *and for* $(c-1)(r-1) = 4$ *degrees of freedom, we have*

$$\chi^2_{0.01,4} = 13.2767$$

Hence, the rejection region is $Q^2 > 13.2767$.
We have $E_{ij} = \frac{n_i n_j}{N}$. *Thus,*

$$E_{11} = \frac{(116)(211)}{500} = 48.952; E_{12} = \frac{(116)(80)}{500} = 18.5;$$

$$E_{13} = \frac{(116)(56)}{500} = 12.992, E_{14} = \frac{(116)(98)}{500} = 22.736;$$

$$E_{15} = \frac{(116)(55)}{500} = 12.76, E_{21} = \frac{(384)(211)}{500} = 162.05;$$

$$E_{22} = \frac{(384)(80)}{500} = 61.44; E_{23} = \frac{(384)(56)}{500} = 43.008;$$

and

$$E_{24} = \frac{(384)(98)}{500} = 75.264; \quad E_{25} = \frac{(384)(55)}{500} = 42.24.$$

The value of the test statistic is

$$Q^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \left[\frac{(39-48.952)^2}{48.952} + \frac{(19-18.5)^2}{18.5} + \frac{(12-12.992)^2}{12.992} + \frac{(28-22.736)^2}{22.736} \right]$$
$$+ \frac{(18-12.76)^2}{12.76} + \frac{(172-162.05)^2}{162.05} + \frac{(61-61.44)^2}{61.44} + \frac{(44-43.08)^2}{43.08}$$
$$+ \frac{(70-75.264)^2}{75.264} + \frac{(37-42.24)^2}{42.24}$$
$$= 7.1351.$$

Because the observed value of Q^2 does not fall in the rejection region, we do not reject the null hypothesis at $\alpha = 0.01$. Therefore, based on the observed data, the marital status and religious affiliation are independent.

■

7.6.3 Testing to Identify the Probability Distribution: Goodness-of-Fit Chi-Square Test

Another application of the chi-square statistic is using it for goodness-of-fit tests in a different context. In hypothesis testing problems we often assume that the form of the population distribution is known. For example, in a χ^2-test for variance, we assume that the population is normal. The goodness-of-fit tests examine the validity of such an assumption if we have a large enough sample. We now describe the goodness-of-fit test procedure for such applications.

GOODNESS-OF-FIT TEST PROCEDURES FOR PROBABILITY DISTRIBUTIONS

Let $X_1, \ldots, X_n$ be a sample from a population with cdf $F(x)$, which may depend on the set of unknown parameters θ. We wish to test $H_0 : F(x) = F_0(x)$, where $F_0(x)$ is completely specified.

1. Divide the range of values of the random variables X_1 into K nonoverlapping intervals $I_1, I_2, \ldots, I_K$. Let O_j be the number of sample values that fall in the interval $I_j (j = 1, 2, \ldots, K)$.
2. Assuming the distribution of X to be $F_0(x)$, find $P(X \in I_j)$. Let $P(X \in I_j) = \pi_j$. Let $e_j = n\pi_j$ be the expected frequency.
3. Compute the test statistic Q^2 given by
$$Q^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}.$$
The test statistic Q^2 has an approximate χ^2-distribution with $(K-1)$ degrees of freedom.
4. Reject the H_0 if $Q^2 \geq \chi^2_{\alpha,(K-1)}$.
5. **Assumptions:** $e_j \geq 5, j = 1, 2, \ldots, K$.

If the null hypothesis does not specify $F_0(x)$ completely, that is, if $F_0(x)$ contains some unknown parameters $\theta_1, \theta_2, \ldots, \theta_p$, we estimate these parameters by the method of maximum likelihood. Using these estimated values we specify $F_0(x)$ completely. Denote the estimated $F_0(x)$ by $\hat{F}_0(x)$. Let

$$\hat{\pi}_i = P\left\{X \in I_i | \hat{F}_0(x)\right\} \quad \text{and} \quad \hat{E}_i = n\hat{\pi}_i.$$

The test statistic is

$$Q^2 = \sum_{i=1}^{K} \frac{(O_i - \hat{e}_i)^2}{\hat{e}_i}.$$

The statistic Q^2 has an approximate chi-square distribution with $(K - 1 - p)$ degrees of freedom. We reject H_0 if $Q^2 \geq \chi^2_{a,(K-1-p)}$.

We now illustrate the method of goodness-of-fit with an example.

Example 7.6.5

The grades of students in a class of 200 are given in the following table. Test the hypothesis that the grades are normally distributed with a mean of 75 and a standard deviation of 8. Use $\alpha = 0.05$.

Range	0–59	60–69	70–79	80–89	90–100
Number of students	12	36	90	44	18

Solution

We have $O_1 = 12, O_2 = 36, O_3 = 90, O_4 = 44, O_5 = 18.$
We now compute $\pi_i (i = 1, 2, \ldots, 5)$, *using the continuity correction factor,*

$$\pi_1 = P\{X \leq 59.5 | H_0\} = P\{z \leq \tfrac{59.5-75}{8}\} = 0.0262,$$

$$\pi_2 = 0.2189, \pi_3 = 0.4722, \pi_4 = 0.2476, \pi_5 = 0.0351,$$

and

$$E_1 = 5.24, E_2 = 43.78, E_3 = 94.44, E_4 = 49.52, E_5 = 7.02.$$

The test statistic results in

$$\begin{aligned} Q^2 &= \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i} \\ &= \frac{(12-5.74)^2}{5.74} + \frac{(36-43.78)^2}{43.78} + \frac{(90-94.44)^2}{94.44} + \frac{(44-49.52)^2}{49.52} + \frac{(18-7.02)^2}{7.02} \\ &= 26.22. \end{aligned}$$

Q^2 has a chi-square distribution with $(5-1)=4$ degrees of freedom. The critical value is $\chi^2_{0.05,4}=7.11$. Hence, the rejection region is $Q^2>7.11$. Because the observed value of $Q^2=26.22>7.11$, we reject H_0 at $\alpha=0.05$. Thus, we conclude that the population is not normal.

EXERCISES 7.6

7.6.1. The following table gives the opinion on collective bargaining by a random sample of 200 employees of a school system, belonging to a teachers' union.

Opinion on Collective Bargaining by Teachers' Union

	For	Against	Undecided	Total
Staff	30	15	15	60
Faculty	50	10	40	100
Administration	10	25	5	40
Column totals	90	50	60	**200**

Test the hypotheses

H_0 : Opinion on collective bargaining is independent of employee classification

versus

H_a : Opinion on collective bargaining is dependent on employee classification

using $\alpha=0.05$.

7.6.2. A random sample was taken of 300 undergraduate students from a university. The students in the sample were classified according to their gender and according to the choice of their major. The result is given in the following table.

College

Gender	Arts and sciences	Engineering	Business	Other	Total
Male	75	40	24	66	205
Female	45	12	15	23	95
Total	120	52	39	89	300

Test the hypothesis that the choice of the major by undergraduate students in this university is independent of their gender. Use $\alpha=0.01$.

7.6.3. The speeds of vehicles (in mph) passing through a section of Highway 75 are recorded for a random sample of 150 vehicles and are given below. Test the hypothesis that the speeds are normally distributed with a mean of 70 and a standard deviation of 4. Use $a=0.01$.

Range	40–55	56–65	66–75	76–85	>85
Number	12	14	78	40	6

7.6.4. Based on the sample data of 50 days contained in the following table, test the hypothesis that the daily mean temperatures in the city are normally distributed with mean 77 and variance 6. Use $\alpha=0.05$.

Temperature	46–55	56–65	66–75	76–85	86–95
Number of days	4	6	13	23	4

7.6.5. A presidential candidate advertises on TV by comparing his positions on some important issues with those of his opponent. After a series of advertisements, a pollster wants to know whether people have changed their opinion about the candidate. The following are the data based on a survey of 950 randomly chosen people:

Before the Advertisement Was Shown

Support the candidate	Oppose the candidate	Need to know more about the candidate	Undecided
40%	20%	5%	35%

After the Advertisement Was Shown

Support the candidate	Oppose the candidate	Need to know more about the candidate	Undecided
45%	25%	2%	28%

Let p_i, $i = 1, 2, 3, 4$, represent the respective true proportions.
Test

$$H_0 : p_1 = 0.35;\ p_2 = 0.20;\ p_3 = 0.15;\ p_4 = 0.3$$

versus

$$H_a : \text{At least one of the probabilities is different from the hypothesized value.}$$

Test this hypothesis using $\alpha = 0.05$.

7.6.6. A survey of footwear preferences of a random sample of 100 undergraduate students (50 females and 50 males) from a large university resulted in the following data.

	Boots	Leather shoes	Sneakers	Sandals	Other
Female	12	9	12	10	7
Male	10	12	17	7	4

(a) Let $p_i, i = 1, 2, 3, 4, 5$, represent the respective true proportions of students with a particular footwear preference, and let

$$H_0 : p_1 = 0.20;\ p_2 = 0.20;\ p_3 = 0.30;\ p_4 = 0.20;\ p_5 = 0.10$$

versus

$$H_a : \text{At least one of the probabilities is different from the hypothesized value.}$$

Test this hypothesis using $\alpha = 0.05$.

(b) Test the hypothesis that the choice of footwear by undergraduate students in this university is independent of their gender, using $\alpha = 0.05$.

7.7 CHAPTER SUMMARY

In this chapter, we have learned various aspects of hypothesis testing. First, we dealt with hypothesis testing for one sample where we used test procedures for testing hypotheses about true mean, true variance, and true proportion. Then we discussed the comparison of two populations through their true means, true variances, and true proportions. We also introduced the Neyman–Pearson lemma and discussed likelihood ratio tests and chi-square tests for categorical data.

We now list some of the key definitions in this chapter.

- Statistical hypotheses
- Tests of hypotheses, tests of significance, or rules of decision
- Simple hypothesis
- Composite hypothesis
- Type I error
- Type II error
- The level of significance
- The p-value or attained significance level
- The Smith–Satterthwaite procedure
- Power of the test
- Most powerful test
- Likelihood ratio

In this chapter, we also learned the following important concepts and procedures:

- General method for hypothesis testing
- Steps to calculate β
- Steps to find the p-value
- Steps in any hypothesis testing problem
- Summary of hypothesis tests for μ
- Summary of large sample hypothesis tests for p
- Summary of hypothesis tests for the variance σ^2
- Summary of hypothesis tests for $\mu_1 - \mu_2$ for large samples (n_1 & $n_2 \geq 30$)
- Summary of hypothesis tests for $p_1 - p_2$ for large samples
- Testing for the equality of variances
- Summary of testing for a matched pairs experiment
- Procedure for applying the Neyman–Pearson lemma
- Procedure for the likelihood ratio test
- Testing the parameters of a multinomial distribution (summary)
- Testing the independence of two factors
- Goodness-of-fit test procedures for probability distributions

7.8 COMPUTER EXAMPLES

In the following examples, if the value of α is not specified, we will always take it as 0.05.

7.8.1 Minitab Examples

Example 7.8.1
(***t*-Test**): Consider the data

66 74 79 80 69 77 78 65 79 81

Using Minitab, test $H_0 : \mu = 75$ vs. $H_1 : \mu > 75$.

Solution
*Enter the data in **C1**. Then*

Stat > Basic Statistics > 1-sample t. . . > In ***Variables:*** enter ***C1*** > choose ***Test Mean*** > enter ***75*** > in ***Alternative:*** choose ***greater than*** and click ***OK***

We obtain the following output.

```
T-Test of the Mean
Test of mu = 75.00 vs mu > 75.00

            Variable  N   Mean   StDev  SE Mean  T      P
            C1        10  74.80  6.00   1.90     −0.11  0.54
```

Example 7.8.2
For the following data:

Sample 1: 16 18 21 13 19 16 18 15 20 19 14 21 14
Sample 2: 14 15 10 13 11 7 12 11 12 15 14

Test $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$. Use $\alpha = 0.02$.

Solution
*Enter sample 1 data in **C1** and sample 2 data in **C2**. Then*

Stat > Basic Statistics > 2-sample t. . . > Choose ***Samples in different columns*** > in ***Alternative:*** choose ***less than*** > in ***Confidence level:*** enter ***98*** > click ***Assumed equal variances*** and click ***OK***

We obtain the following output.

```
Two Sample T-test and Confidence Interval
Two sample T for C1 vs C2
```

```
                         N   Mean  StDev  SE Mean
                     C1  13  17.23  2.74    0.76
                     C2  11  12.18  2.40    0.72

98% CI for mu C1 – mu C2: (2.38, 7.71)
T-Test mu C1 = mu C2 (vs <): T = 4.75 P = 1.0 DF = 22
Both use Pooled StDev = 2.59
```

If we did not select ***Assumed equal variances***, we will obtain the following output.

```
Two Sample T-Test and Confidence Interval
Two sample T for C1 vs C2
                         N   Mean  StDev  SE Mean
                     C1  13  17.23  2.74    0.76
                     C2  11  12.18  2.40    0.72

98% CI for mu C1 - mu C2: (2.40, 7.69)
T-Test mu C1 = mu C2 (vs <): T = 4.81 P = 1.0 DF = 21
```

Example 7.8.3

For the following data:

6.8	5.6	8.5	8.5	8.4	7.5	9.3	9.4	7.8	7.1
9.9	9.6	9.0	9.4	13.7	16.6	9.1	10.1	10.6	11.1
8.9	11.7	12.8	11.5	12.0	10.6	11.1	6.4	12.3	12.3
11.4	9.9	14.3	11.5	11.8	13.3	12.8	13.7	13.9	12.9
14.2	14.0	15.5	16.9	18.0	17.9	21.8	18.4	34.3	

Test $H_0 : \mu = 12$ versus $H_1 : \mu \neq 12$. Use $\alpha = 0.05$.

Solution

*Enter the data in **C1**. Then*

Stat > Basic Statistics > 1-sample z. . . > in ***Variables:*** Type ***C1 >*** choose ***Test Mean*** and enter ***12 >*** choose ***not equal*** in ***Alternative***, and Type ***4.7*** for ***sigma >*** Click ***OK***

We obtain the following output.

```
Z-Test
Test of mu = 12.000 vs mu not = 12.000
The assumed sigma = 4.70

          Variable  N   Mean   StDev  SE Mean   Z     P
             C1     49  12.124 4.700   0.671   0.19  0.85
```

Here the test statistic is 0.19 and the p-value is 0.85, which is larger than 0.05. Hence, we cannot reject the null hypothesis.

Example 7.8.4

(Contingency Table): Consider the following data with five levels and two factors. Test for dependence of the factors.

Factors	Levels				
	1	2	3	4	5
1	39	19	12	28	18
2	172	61	44	70	37

Solution

*In **C1** enter the data in column 1 (39 and 172), and continue to **C5**. Then*

Stat** > **Tables** > **Chi-Square-Test. . .** > in **Columns containing the table:** Type **C1 C2 C3 C4 C5** > click **OK

We will obtain the following output.

```
Chi-Square Test
Expected counts are printed below observed counts
       C1      C2     C3     C4     C5     Total
   1   39      19     12     28     18     116
       48.95   18.56  12.99  22.74  12.76

   2   172     61     44     70     37     384
       162.05  61.44  43.01  75.26  42.24
 Total 211     80     56     98     55     500

Chi-Sq = 2.023 + 0.010 + 0.076 + 1.219 + 2.152 +
0.611 + 0.003 + 0.023 + 0.368 + 0.650 = 7.135

DF = 4, p-value = 0.129
```

Example 7.8.5

(Paired *t*-Test): Consider the data of Example 7.5.7. Using Minitab, perform a paired t-test.

Solution

*Enter sample 1 in column **C1** and sample 2 in column **C2**. Then:*

Stat** > **Basic Statistics** > **Paired t. . .** >* in ***First Sample: Type ***C2***, and in the ***Second sample:*** Type ***C1* >** click ***options* >** and click ***less than*** (if α is other than 0.05, enter appropriate percentage in ***Confidence level:*** and enter appropriate number if it is not zero in ***Test mean:***) **>** click ***OK* > *OK***

We obtain the following output.

```
Paired T-test and Confidence Interval

Paired  T for C2 - C1
              N        Mean       StDev    SE Mean
C2           10       171.3       47.1      14.9
C1           10       243.2       40.1      12.7
Difference   10       -71.9       56.2      17.8
95% CI for mean difference: (-112.1, -31.7)
T-Test of mean difference = 0 (vs < 0): T-Value = -4.05
p-value = 0.001
```

because the p-value $0.001 < 0.05 = \alpha$.

7.8.2 SPSS Examples

Example 7.8.6
Consider the data

66 74 79 80 69 77 78 65 79 81

Using SPSS, test $H_0 : \mu = 75$ vs. $H_1 : \mu > 75$.

Solution
Use the following procedure:

1. Enter the data in column 1.
2. Click ***Analyze > Compare Means > One-sample t Test. . .*** , Move ***var00001*** to ***Test Variable(s)***, and change ***Test Value: 0*** to ***75***. Click ***OK***

We obtain the following output.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	10	74.8000	5.99630	1.89620

One-Sample Test

	Test Value = 75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00001	−.105	9	.918	−.2000	−4.4895	4.0895

For the one sample t-test $H_0 : \mu = 75$ vs. $H_1 : \mu > 75$, the t-statistic is −0.105 with 9 degrees of freedom. The p-value is $0.46 > 0.02$. Hence, we will not reject the null hypothesis.

If we want the computer to calculate the p-value in the previous example, use the following procedure.

1. Enter the test statistic ***(−0.105)*** in the data editor using '***teststat***'.
2. Click ***Transform > compute. . .***
3. Type '***p-value***' in the box called ***Tarobtain value***. In the box called ***Functions:*** scroll and click on ***CDF.T(q,df)*** and move to Numeric Expressions.
4. The CDF(q,df) will appear as ***CDF(?,?)*** in the Numeric Expressions box. Replace teststat for ***q*** and ***9*** for ***df*** (the degree of freedom in this example is 9). Click ***OK***

We obtain the p-value as 0.46.

Example 7.8.7

For the following data

Sample 1:	16	18	21	13	19	16	18	15	20	19	14	21	14
Sample 2:	14	15	10	13	11	7	12	11	12	15	14		

Test $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$. Use $\alpha = 0.02$.

Solution

*In column 1, under the title "group" enter 1s to identify the sample 1 data and 2s to identify sample 2 data. In column **C2**, under the title "data" enter the data corresponding to samples 1 and 2. Then:*

Analyze > Compare Means > Independent Samples t-test. . . > bring ***Data*** to ***Test Variable(s):*** and ***group*** to ***Grouping Variable:,*** click ***Define Groups. . . ,*** and enter ***1*** for ***sample 1, 2*** for ***sample 2 >*** click ***continue >*** click ***Options. . . .*** Enter ***98*** in ***Confidence interval: >*** click ***continue > OK***

We obtain the following output.

Group Statistics					
	GROUP	N	Mean	Std. Deviation	Std. Error Mean
DATA	1.00	13	17.2308	2.74329	.76085
	2.00	11	12.1818	2.40076	.72386

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	98% Confidence Interval of the Difference	
									Lower	Upper
DATA	Equal variances assumed	.975	.334	4.753	22	.000	5.0490	1.06237	2.38419	7.71372
	Equal variances not assumed			4.808	21.963	.000	5.0490	1.05017	2.41443	7.68347

Looking at the statistical significance values, which are greater than 0.05, we do not reject the null hypothesis.

Example 7.8.8

(Paired *t*-Test) For the data of Example 7.5.7, use SPSS to test whether the data provide sufficient evidence for the claim that the new program reduces blood glucose level in diabetic patients. Use $\alpha = 0.05$.

Solution

*Enter **after** data in column **C1** and **before** data in column **C2**. Then:*

Analyze **>** ***Compare Means*** **>** ***Paired-Sample T-Test*** **>** bring after and before to ***Paired Variables:*** so that it will look ***after-before*** **>** click ***OK***

We obtain the following output.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	AFTER	171.3000	10	47.11228	14.89821
	BEFORE	243.2000	10	40.12979	12.69015

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	AFTER & BEFORE	10	.179	.621

Paired Samples Test

		Paired Differences					*t*	*df*	*Sig. (2-tailed)*
		Mean	*Std. Deviation*	*Std. Error Mean*	*95% Confidence Interval of the Difference*				
					Lower	*Upper*			
Pair 1	*AFTER -- BEFORE*	*−71.9000*	*56.15544*	*17.75791*	*−112.0712*	*−31.7288*	*−4.049*	*9*	*.003*

Because the significance level for the test is 0.003, which is less than $\alpha = 0.05$*, we reject the null hypothesis.*

7.8.3 SAS Examples

To conduct a hypothesis test using SAS, we could use proc ttest, or proc means with option of computing the t-value and corresponding probability. However, to use this, we need a hypothesis of the form $H_0 : \mu = 0$. For testing nonzero values, $H_0 : \mu = \mu_0$, we must create a new variable by subtracting μ_0 from each observation, and then use the test procedure for this new variable. The following example illustrates this concept.

Example 7.8.9

(t-Test): The following radar measurements of speed (in miles per hour) are obtained for 10 vehicles traveling on a stretch of interstate highway.

66 74 79 80 69 77 78 65 79 81

Do the data provide sufficient evidence to indicate that the mean speed at which people travel on this stretch of highway is at least 75 mph? Test using $\alpha = 0.01$. Use an SAS procedure to do the analysis.

Solution

In the SAS editor, type in the following commands.

```
data speed;
   title 'Test on highway speed';
   input X @@;
   Y=X-75;
   datalines;
66 74 79 80 69 77 78 65 79 81
;
PROC TTEST data=speed;
run;
```

We obtain the following output.

```
Test on highway speed
The TTEST Procedure
Statistics
```

		Lower CL		Upper CL		Lower CL	Upper CL	
Variable	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
X	10	70.511	74.8	79.089	4.1245	5.9963	10.947	1.8962
Y	10	−4.489	−0.2	4.0895	4.1245	5.9963	10.947	

T-Tests

Variable	DF	t Value	Pr > |t|
X	9	39.45	<.0001
Y	9	−0.11	0.9183

To test $H_0 : \mu = 75$, we need to look at the Y-values. The corresponding t-value is -0.11, and because this is a one-sided test, we need to divide 0.9183 by 2 to obtain the p-value as $p = 0.45915$. Because the p-value is larger than $0.01 = \alpha$, we cannot reject the null hypothesis.

One of the easier ways to conduct large sample hypothesis testing using SAS procedures is through the computation of the p-value. The following example illustrates the procedure.

Example 7.8.10

(z-Test): It is claimed that the average miles driven per year for sports cars is at least 18,000 miles. To check the claim, a consumer firm tests 40 of these cars randomly and obtains a mean of 17,463 miles with standard deviation of 1348 miles. What can it conclude if $\alpha = 0.01$?

Solution

Here we will find the p-value and compare that with α to test the hypothesis. We use the following SAS procedure:

```
Data ex888;
z=(17463-18000)/(1348/(SQRT(40)));
           pval=probnorm(z);
run;
proc print data=ex888;
title 'Test of mean, large sample';
run;
```

We obtain the following output.

Test of mean,		large sample
Obs	z	pval
1	2.51950	.005876079

Because the p-value of 0.005876079 is less than $\alpha = 0.01$, we reject the null hypothesis. There is sufficient evidence to conclude that the mean miles driven per year for sport cars is less than 18,000.

Note that in the previous example, the value of z was negative. If the value of z is positive, use `pval=probnorm(-z);`, also, if it is a two-sided hypothesis, we need to multiply by 2, so use `pval=probnorm(z)*2;` to obtain the p-value.

Example 7.8.11

(Paired t-Test): For the data of Example 7.5.7, use SAS to test whether the data provide sufficient evidence for the claim that the new program reduces blood glucose level in diabetic patients. Use $\alpha = 0.05$.

Solution

We can use the following commands.

```
data dietexr;
  input before after;
  diff = after - before;
  datalines;
      268 106
      225 186
      252 223
      192 110
      307 203
      228 101
      246 211
      298 176
      231 194
      185 203
;
run;
proc means data=dietexr t prt;
var diff;
title 'Test of mean, Paired difference';
run;
```

We obtain the following output.

Test of mean, Paired difference
The MEANS Procedure
Analysis Variable : diff

t Value	Pr > ltl
−4.05	0.0029

Because the p*-value 0.0029 is less than* $\alpha = 0.05$*, we reject the null hypothesis.*

PROJECTS FOR CHAPTER 7

7A. Testing on Computer-Generated Samples

(a) Small sample test:
Generate a sample of size 20 from a normal population with $\mu = 10$, and $\sigma^2 = 4$.
- **(i)** Perform a t-test for the test $H_0 : \mu = 10$ versus $H_a : \mu \neq 10$ at level $\alpha = 0.05$.
- **(ii)** Perform the test $H_0 : \sigma^2 = 4$ versus $H_a : \sigma^2 \neq 4$ at level $\alpha = 0.05$.

Repeat the procedure 10 times, and comment on the results.

(b) Large sample test:

Generate a sample of size 50 from a normal population with $\mu = 10$, and $\sigma^2 = 4$. Perform a z-test for the test $H_0 : \mu = 10$ versus $H_a : \mu \neq 10$ at level $\alpha = 0.05$. Repeat the procedure 10 times and comment on the results.

7B. Conducting a Statistical Test with Confidence Interval

Let θ be any population parameter. Consider the three tests of hypotheses

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta > \theta_0 \tag{1}$$

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta < \theta_0 \tag{2}$$

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta \neq \theta_0 \tag{3}$$

The following procedure can be exploited to test a statistical hypothesis utilizing the confidence intervals.

Procedure to Use Confidence Interval for Hypothesis Testing

Let θ be any population parameter.

(a) For test (1), that is,

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta > \theta_0$$

choose a value for α. From a random sample, compute a confidence interval for θ using a confidence coefficient equal to $1 - 2\alpha$. Let L be the lower end point of this confidence interval.

$$\text{Reject } H_0 \text{ if } \theta_0 < L.$$

That is, we will reject the null hypothesis if the confidence interval is completely to the right of θ_0.

(b) For test (2), that is,

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta < \theta_0$$

choose a value for α. From a random sample, compute a confidence interval for θ using a confidence coefficient equal to $1 - 2\alpha$. Let U be the upper end point of this confidence interval.

$$\text{Reject } H_0 \text{ if } U < \theta_0.$$

That is, we will reject the null hypothesis if the confidence interval is completely to the left of θ_0.

(c) For test (3), that is,

$$H_0 : \theta = \theta_0 \text{ vs. } H_a : \theta \neq \theta_0$$

choose a value for α. From a random sample, compute a confidence interval for θ using a confidence coefficient equal to $1 - \alpha$. Let L be the lower end point and U be the upper end point of this confidence interval.

$$\text{Reject } H_0 \text{ if } \theta_0 < L \quad \text{or } U < \theta_0.$$

That is, we will reject the null hypothesis if the confidence interval does not contain θ_0.

(i) For any large data set, conduct all three of these hypothesis tests using a confidence interval for the population mean.

(ii) For any small data set, conduct all three of these hypothesis tests using a confidence interval for the population mean.