9.6 The Power of a Test

Section 9.1 defined Type I and Type II errors and their associated risks. Recall that \( \alpha \) represents the probability that you reject the null hypothesis when it is true and should not be rejected, and \( \beta \) represents the probability that you do not reject the null hypothesis when it is false and should be rejected. The power of the test, \( 1 - \beta \), is the probability that you correctly reject a false null hypothesis. This probability depends on how different the actual population parameter is from the value being hypothesized (under \( H_0 \)), the value of \( \alpha \) used, and the sample size. If there is a large difference between the population parameter and the hypothesized value, the power of the test will be much greater than if the difference between the population parameter and the hypothesized value is small. Selecting a larger value of \( \alpha \) makes it easier to reject \( H_0 \) and therefore increases the power of a test. Increasing the sample size increases the precision in the estimates and therefore increases the ability to detect differences in the parameters and increases the power of a test.

The power of a statistical test can be illustrated by using the Oxford Cereal Company scenario. The filling process is subject to periodic inspection from a representative of the consumer affairs office. The representative’s job is to detect the possible “short weighting” of boxes, which means that cereal boxes having less than the specified 368 grams are sold. Thus, the representative is interested in determining whether there is evidence that the cereal boxes have a mean weight that is less than 368 grams. The null and alternative hypotheses are as follows:

\[
H_0: \mu \geq 368 \quad \text{(filling process is working properly)} \\
H_1: \mu < 368 \quad \text{(filling process is not working properly)}
\]

The representative is willing to accept the company’s claim that the standard deviation, \( \sigma \), equals 15 grams. Therefore, you can use the \( Z \) test. Using Equation (9.1) on page 330, with \( \bar{X}_L \) (the lower critical \( \bar{X} \) value) substituted for \( \bar{X} \), you can find the value of \( \bar{X} \) that enables you to reject the null hypothesis:

\[
Z = \frac{\bar{X}_L - \mu}{\sigma / \sqrt{n}}
\]

\[
Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{X}_L - \mu
\]

\[
\bar{X}_L = \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

Because you have a one-tail test with a level of significance of 0.05, the value of \( Z_{\alpha/2} \) is equal to \(-1.645\) (see Figure 9.16). The sample size \( n = 25 \). Therefore,

\[
\bar{X}_L = 368 + (-1.645) \frac{15}{\sqrt{25}} = 368 - 4.935 = 363.065
\]

The decision rule for this one-tail test is

Reject \( H_0 \) if \( \bar{X} < 363.065 \);
otherwise, do not reject \( H_0 \).
The decision rule states that if in a random sample of 25 boxes, the sample mean is less than 363.065 grams, you reject the null hypothesis, and the representative concludes that the process is not working properly. The power of the test measures the probability of concluding that the process is not working properly for differing values of the true population mean.

What is the power of the test if the actual population mean is 360 grams? To determine the chance of rejecting the null hypothesis when the population mean is 360 grams, you need to determine the area under the normal curve below $X_L = 363.065$ grams. Using Equation (9.1), with the population mean $\mu = 360,$

$$Z_{STAT} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{363.065 - 360}{15 / \sqrt{25}} = 1.02$$

From Table E.2, there is an 84.61% chance that the $Z$ value is less than +1.02. This is the power of the test where $\mu$ is the actual population mean (see Figure 9.17). The probability ($\beta$) that you will not reject the null hypothesis ($\mu = 368$) is $1 - 0.8461 = 0.1539.$ Thus, the probability of committing a Type II error is 15.39%.

Now that you have determined the power of the test if the population mean were equal to 360, you can calculate the power for any other value of $\mu.$ For example, what is the power of the test if the population mean is 352 grams? Assuming the same standard deviation, sample size, and level of significance, the decision rule is

Reject $H_0$ if $\bar{X} < 363.065$
otherwise, do not reject $H_0.$

Once again, because you are testing a hypothesis for a mean, from Equation (9.1),

$$Z_{STAT} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

If the population mean shifts down to 352 grams (see Figure 9.18), then

$$Z_{STAT} = \frac{363.065 - 352}{15 / \sqrt{25}} = 3.69$$
9.6 The Power of a Test

From Table E.2, there is a 99.989% chance that the $Z$ value is less than $+3.69$. This is the power of the test when the population mean is 352 grams. The probability ($\beta$) that you will not reject the null hypothesis ($\mu = 368$) is $1 - 0.99989 = 0.00011$. Thus, the probability of committing a Type II error is only 0.011%.

In the preceding two examples, the power of the test is high, and the chance of committing a Type II error is low. In the next example, you compute the power of the test when the population mean is equal to 367 grams—a value that is very close to the hypothesized mean of 368 grams.

Once again, from Equation (9.1),

$$Z_{STAT} = \frac{X - \mu}{\sigma/\sqrt{n}}$$

If the population mean is equal to 367 grams (see Figure 9.19), then

$$Z_{STAT} = \frac{363.065 - 367}{15/\sqrt{25}} = -1.31$$

From Table E.2, the probability less than $Z = -1.31$ is 0.0951 (or 9.51%). Because the rejection region is in the lower tail of the distribution, the power of the test is 9.51%, and the chance of making a Type II error is 90.49%.

Figure 9.20 illustrates the power of the test for various possible values of $\mu$ (including the three values examined). This graph is called a **power curve**.
From Figure 9.20, you can see that the power of this one-tail test increases sharply (and approaches 100%) as the population mean takes on values farther below the hypothesized mean of 368 grams. Clearly, for this one-tail test, the smaller the actual mean the greater the power to detect this difference. For values of close to 368 grams, the power is small because the test cannot effectively detect small differences between the actual population mean and the hypothesized value of 368 grams. When the population mean approaches 368 grams, the power of the test approaches the level of significance (which is 0.05 in this example).

In the above discussion, a one-tail test with \( \alpha = 0.05 \) and \( n = 25 \) was used. The type of statistical test (one-tail vs. two-tail), the level of significance, and the sample size all affect the power. Three basic conclusions regarding the power of the test are summarized below:

1. A one-tail test is more powerful than a two-tail test.
2. An increase in the level of significance (\( \alpha \)) results in an increase in power. A decrease in \( \alpha \) results in a decrease in power.
3. An increase in the sample size, \( n \), results in an increase in power. A decrease in the sample size, \( n \), results in a decrease in power.
Figure 9.21
Determining statistical power for varying values of the population mean

Panel A
Given: $\alpha = .05$, $\sigma = 15$, $n = 25$
One-tail test
$\mu = 368$ (null hypothesis is true)

$X_0 = 368 - (1.645) \cdot \frac{15}{\sqrt{25}} = 363.065$

Decision rule: Reject $H_0$ if $X < 363.065$; otherwise, do not reject

Panel B
Given: $\alpha = .05$, $\sigma = 15$, $n = 25$
One-tail test
$H_0: \mu = 368$
$\mu = 367$ (true mean shifts to 367 grams)

$Z_{STAT} = \frac{X - \mu}{\sigma/\sqrt{n}} = \frac{363.065 - 367}{3} = -1.31$

Power = .0951

Panel C
Given: $\alpha = .05$, $\sigma = 15$, $n = 25$
One-tail test
$H_0: \mu = 368$
$\mu = 360$ (true mean shifts to 360 grams)

$Z_{STAT} = \frac{X - \mu}{\sigma/\sqrt{n}} = \frac{363.065 - 360}{3} = 1.02$

Power = .8461

Panel D
Given: $\alpha = .05$, $\sigma = 15$, $n = 25$
One-tail test
$H_0: \mu = 368$
$\mu = 352$ (true mean shifts to 352 grams)

$Z_{STAT} = \frac{X - \mu}{\sigma/\sqrt{n}} = \frac{363.065 - 352}{3} = 3.69$

Power = .99989

Region of Rejection
Region of Nonrejection
Problems for Section 9.6

APPLYING THE CONCEPTS

9.80 A coin-operated soft-drink machine is designed to discharge at least 7 ounces of beverage per cup, with a standard deviation of 0.2 ounce. If you select a random sample of 16 cups and you are willing to have an $\alpha = 0.05$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean amount dispensed is actually
a. 6.9 ounces per cup.
b. 6.8 ounces per cup.

9.81 Refer to Problem 9.78. If you are willing to have an $\alpha = 0.01$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean amount dispensed is actually
a. 6.9 ounces per cup.
b. 6.8 ounces per cup.
c. Compare the results in (a) and (b) of this problem and in Problem 9.78. What conclusion can you reach?

9.82 Refer to Problem 9.78. If you select a random sample of 25 cups and are willing to have an $\alpha = 0.05$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean amount dispensed is actually
a. 6.9 ounces per cup.
b. 6.8 ounces per cup.
c. Compare the results in (a) and (b) of this problem and in Problem 9.78. What conclusion can you reach?

9.83 A tire manufacturer produces tires that have a mean life of at least 25,000 miles when the production process is working properly. Based on past experience, the standard deviation of the tires is 3,500 miles. The operations manager stops the production process if there is evidence that the mean tire life is below 25,000 miles. If you select a random sample of 100 tires (to be subjected to destructive testing) and you are willing to have an $\alpha = 0.05$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean life is actually
a. 24,000 miles.
b. 24,900 miles.

9.84 Refer to Problem 9.81. If you are willing to have an $\alpha = 0.01$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean life is actually
a. 24,000 miles.
b. 24,900 miles.
c. Compare the results in (a) and (b) of this problem and (a) and (b) in Problem 9.81. What conclusion can you reach?

9.85 Refer to Problem 9.81. If you select a random sample of 25 tires and are willing to have an $\alpha = 0.05$ risk of committing a Type I error, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean life is actually
a. 24,000 miles.
b. 24,900 miles.
c. Compare the results in (a) and (b) of this problem and (a) and (b) in Problem 9.81. What conclusion can you reach?

9.86 Refer to Problem 9.81. If the operations manager stops the process when there is evidence that the mean life is different from 25,000 miles (either less than or greater than) and a random sample of 100 tires is selected, along with a level of significance of $\alpha = 0.05$, compute the power of the test and the probability of a Type II error ($\beta$) if the population mean life is actually
a. 24,000 miles.
b. 24,900 miles.
c. Compare the results in (a) and (b) of this problem and (a) and (b) in Problem 9.81. What conclusion can you reach?