

---

---

# MODULE 15

---

## HYPOTHESIS TEST ERRORS

### Contents

---

15.1 Errors . . . . .	121
15.2 Statistical Power . . . . .	122

---

**D**ECISIONS ABOUT HYPOTHESES BASED ON STATISTICS may, at times, be incorrect. In this module, two types of errors that can be made are described and the probability of making these errors is described. The concepts in Module 14 should be understood before proceeding here.

### 15.1 Errors

The goal of hypothesis testing is to make a decision about  $H_0$ . Unfortunately, because of sampling variability, there is always a risk of making an incorrect decision. Two types of incorrect decisions can be made (Table 15.1). A Type I error occurs when a true  $H_0$  is falsely rejected. In other words, even if  $H_0$  is true, there is a chance that a rare sample will occur and  $H_0$  will be deemed incorrect. The probability of making a Type I error is set when  $\alpha$  is chosen. A Type II error occurs when a false  $H_0$  is not rejected. The probability of a Type II error is denoted by  $\beta$ .

Table 15.1. Types of decisions that can be made from a hypothesis test.

		Decision from Data	
		Reject	Not Reject
Truth About Population	$H_0$	Type I	Correct
	$H_A$	Correct	Type II

The decision in the Square Lake example of Module 14 produced a Type II error because  $H_0 : \mu = 105$  was not rejected even though we know that  $\mu = 98.06$  (Table 2.1). Unfortunately, in real life, it will never be known exactly when a Type I or a Type II error has been made because the true  $\mu$  is not known. However, it is known that a Type I error will be made  $100\alpha\%$  of the time. The probability of a type II error ( $\beta$ ), though, is never known because this probability depends on the true but unknown  $\mu$ . Decisions can be made, however, that affect the magnitude of  $\beta$  (discussed below with power).

## 15.2 Statistical Power

A concept that is very closely related to decision-making errors is the idea of statistical power, or just **power** for short. Power is the probability of correctly rejecting a false  $H_0$ . In other words, it is the probability of detecting a difference from the hypothesized value if a difference really exists. Power is used to demonstrate how sensitive a hypothesis test is for identifying a difference. High power related to a  $H_0$  that is not rejected implies that the  $H_0$  really should not have been rejected. Conversely, low power related to a  $H_0$  that was not rejected implies that the test was very unlikely to detect a difference, so not rejecting  $H_0$  is not surprising nor particularly conclusive.

Power is equal to  $1 - \beta$  and, thus, like  $\beta$  it cannot be computed directly because the actual mean ( $\mu_A$ ) is not known. However, in the Square Lake example,  $\mu_A$  is known and power can be calculated in four steps:

1. Draw the sampling distribution assuming the  $H_0$  is true (called the null distribution).
  - The null distribution is  $N(105, \frac{31.49}{\sqrt{50}})$  because  $H_0 : \mu = 105$ ,  $\sigma = 31.49$ , and  $n = 50$ .
2. Find the rejection region borders (based on  $\alpha$  and  $H_A$ ) in terms of the value of the statistic (a “reverse” calculation on the null distribution).
  - The rejection region is delineated by the  $\bar{x}$  that has  $\alpha = 0.10$  to the left (because  $H_A$  is a “less than”). This reverse calculation on the null distribution gives  $\bar{x}=99.2928$ .

```
> ( rejreg <- distrib(0.10,mean=105,sd=31.49/sqrt(50),type="q") )
[1] 99.29279
```

3. Draw the sampling distribution corresponding to the “actual” parameter value (SE is the same as that for the null distribution).
  - The actual  $\mu$  is 98.06. Thus, the actual sampling distribution is  $N(98.06, \frac{31.49}{\sqrt{50}})$ .
4. Compute the portion of the “actual” sampling distribution in the REJECTION region of the null distribution (i.e., a “forward” calculation on the actual distribution).
  - This computation is to find the area to the left of  $\bar{x}=99.2928$  on  $N(98.06, \frac{31.49}{\sqrt{50}})$ . The area to the left of this Z is 0.6090.

```
> ( distrib(rejreg,mean=98.06,sd=31.49/sqrt(50)) )
[1] 0.6090419
```

Thus, the power to detect a  $\mu_A = 98.06$  was 0.6090. This means that in only about 61% of the samples will the false  $H_0 : \mu = 105$  be correctly rejected. Thus, it is not too surprising that  $H_0$  was not rejected in this example.

Even though power can usually not be calculated, a researcher can make decisions that will positively affect power (Figure 15.1). For example, a researcher can increase power by increasing  $\alpha$  or  $n$ . Increasing  $n$  is more beneficial because it does not result in an increase in Type I errors as would occur with increasing  $\alpha$ .

In addition, power decreases as the difference between the hypothesized mean ( $\mu_0$ ) and the actual mean ( $\mu_A$ ) decreases (Figure 15.1). This means that the ability to detect increasingly smaller differences decreases. In addition, power decreases with an increasing amount of natural variability (i.e.,  $\sigma$ ; Figure 15.1). In other words, the ability to detect a difference decreases with increasing amounts of variability among individuals. A researcher cannot control the difference between  $\mu_0$  and  $\mu_A$  or the value of  $\sigma$ . However, it is important to know that if a situation with a “large” amount of variability is encountered or the difference to be detected is small, the researcher will need to increase  $n$  to gain power. For example, if  $n$  could be doubled in the Square Lake example to 100, then the power to correctly reject  $H_0 : \mu = 105$  would increase to approximately 0.82 (Figure 15.1).

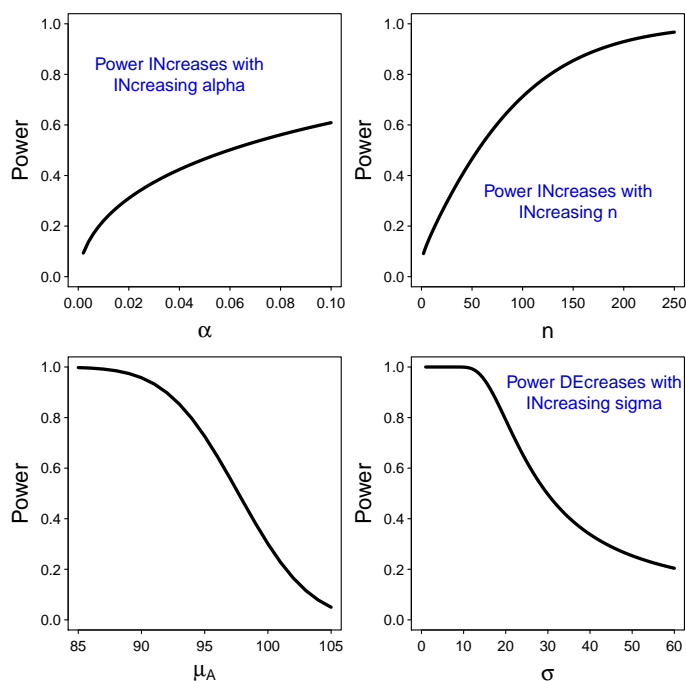


Figure 15.1. The relationship between one-tailed (lower) power and  $\alpha$ ,  $n$ , actual mean ( $\mu_A$ ), and  $\sigma$ . In all situations where the variable does not vary,  $\mu_0 = 105$ ,  $\mu_A = 98.06$ ,  $\sigma = 31.49$ ,  $n = 50$ , and  $\alpha = 0.05$ .