7.11: Other Considerations in Hypothesis Testing

There are several other considerations we need to keep in mind when performing hypothesis testing.

Errors in Hypothesis Testing

In the Physicians' Reactions case study, the probability value associated with the significance test is 0.0057. Therefore, the null hypothesis was rejected, and it was concluded that physicians intend to spend less time with obese patients. Despite the low probability value, it is possible that the null hypothesis of no true difference between obese and average-weight patients is true and that the large difference between sample means occurred by chance. If this is the case, then the conclusion that physicians intend to spend less time with obese patients is in error. This type of error is called a Type I error. More generally, a Type I error occurs when a significance test results in the rejection of a true null hypothesis.

By one common convention, if the probability value is below 0.05 then the null hypothesis is rejected. Another convention, although slightly less common, is to reject the null hypothesis if the probability value is below 0.01. The threshold for rejecting the null hypothesis is called the α level or simply α. It is also called the significance level. As discussed in the introduction to hypothesis testing, it is better to interpret the probability value as an indication of the weight of evidence against the null hypothesis than as part of a decision rule for making a reject or do-not-reject decision. Therefore, keep in mind that rejecting the null hypothesis is not an all-or-nothing decision.

The Type I error rate is affected by the α level: the lower the α level the lower the Type I error rate. It might seem that α is the probability of a Type I error. However, this is not correct. Instead, α is the probability of a Type I error given that the null hypothesis is true. If the null hypothesis is false, then it is impossible to make a Type I error.

The second type of error that can be made in significance testing is failing to reject a false null hypothesis. This kind of error is called a Type II error. Unlike a Type I error, a Type II error is not really an error. When a statistical test is not significant, it means that the data do not provide strong evidence that the null hypothesis is false. Lack of significance does not support the conclusion that the null hypothesis is true. Therefore, a researcher should not make the mistake of incorrectly concluding that the null hypothesis is true when a statistical test was not significant. Instead, the researcher should consider the test inconclusive. Contrast this with a Type I error in which the researcher erroneously concludes that the null hypothesis is false when, in fact, it is true.

A Type II error can only occur if the null hypothesis is false. If the null hypothesis is false, then the probability of a Type II error is called β (beta). The probability of correctly rejecting a false null hypothesis equals 1- β and is called power. Power is simply our ability to correctly detect an effect that exists. It is influenced by the size of the effect (larger effects are easier to detect), the significance level we set (making it easier to reject the null makes it easier to detect an effect, but increases the likelihood of a Type I Error), and the sample size used (larger samples make it easier to reject the null).

Misconceptions in Hypothesis Testing

Misconceptions about significance testing are common. This section lists three important ones.

1. Misconception: The probability value is the probability that the null hypothesis is false.
    - Proper interpretation: The probability value is the probability of a result as extreme or more extreme given that the null hypothesis is true. It is the probability of the data given the null hypothesis. It is not the probability that the null hypothesis is false.
2. Misconception: A low probability value indicates a large effect.
    - Proper interpretation: A low probability value indicates that the sample outcome (or one more extreme) would be very unlikely if the null hypothesis were true. A low probability value can occur with small effect sizes, particularly if the sample size is large.
3. Misconception: A non-significant outcome means that the null hypothesis is probably true.
    - Proper interpretation: A non-significant outcome means that the data do not conclusively demonstrate that the null hypothesis is false.