2.5: Statistical Power

Learning Objectives

At the end of this chapter you should be able to answer the following:

- Explain the idea of Statistical Power.
- Define Type I and Type ll Decision Errors.

As course examiners, we know that students always want to get the 'right' answer. However, it is possible to make errors in the interpretation of statistical effects. When researchers misinterpret the significance of results, these mistakes are often more to do with errors in judgment rather than about mistakes in effect calculations.

You may recall that the essence of significance lies in an inference between a sample and the population to which it corresponds.

Therefore, there are two factors that must be considered.

Firstly, we need to consider the conclusion we reach about a sample result showing a significant or insignificant effect.

Secondly, we need to consider whether there is a true state of reality in the population from which the sample corresponds.

It is important to understand that the conclusion of the sample result and the true state of reality in the population must be aligned.

For example, if we conclude there is no difference between groups in a sample – and at the same time conclude there is no difference between groups in the true state of reality in the population – then our judgment about the sample result and the difference between the true state of the population are aligned.

Misalignment of the sample result and the true state of reality in the population can be thought of either as Type I and Type ll Decision Errors.

A Type l (α) Error is known as a false positive or thinking something is there when it is not. This is where, based on the sample, we say there is a true effect, but in the population, there is in fact no effect. For example, a Type I error occurs when a healthy person is diagnosed with a particular disease like cancer.

A Type ll (β) Error is known as a false negative or thinking something is not there when it is. This is where, based on the sample, we say there is a not a true effect, but in the population, there is in fact an actual true effect. For example, a Type II error occurs when a sick person is diagnosed as being disease-free.

The statistical power of the statistical test is related to the Type ll (β) Error as the probability of rejecting the null hypothesis when it is true.

Power = $1 - \beta$

Statistical power is, therefore, the probability of making a correct decision, or saying there is an effect based on the sample, and at the same time there is in fact an actual true effect in the population.

The smaller the probability of a type II error or a false negative, the bigger the power. However, with large amounts of data, it is possible to have samples that are overpowered and find statistically significant effects that are not practically significant.

There are tests that can be used to check if the sample size you have is large enough to detect a relationship or a difference. These are called apriori and posthoc power analyses. An apriori power analysis allows researchers to understand how large a sample should be so that it has adequate sensitivity to detect true effects. Posthoc power analyses allow researchers to determine if the sample size was inadequate for calculating a statistically significant finding.