Type I and Type II Error

We use inferences based on probability to make decisions in hypothesis testing. Basing decisions on probability, and not certainty, introduces the possibility of error – that is, of making incorrect inferences and drawing incorrect conclusions. These are NOT errors in calculation or process – these errors are inherent in hypothesis testing, even if you do everything right.

Two types of error in hypothesis testing:

1. **Type I Error:** Rejecting the null hypothesis (and accepting the alternative hypothesis), when the null hypothesis is actually true.
2. **Type II Error:** Not rejecting the null hypothesis, when the null hypothesis is actually false.

These errors have direct, real-world consequences. If you are testing whether a new program has decreased costs, and a Type I error occurs, then you would conclude that the program decreased costs when it actually had not done so. This could lead your company to prolong an ineffective program. On the other hand, consider the kettlebell hypothesis test we did on Handout #4. In that test we did not reject the null hypotheses, and so we concluded that there was no reason to shut down the manufacturing line. If a Type II error had occurred in that situation, then we would have let the line continue when actually the population mean weight of the bells was not 20lbs, thus leading us to ship product that did not conform to our 20lb specification.

This is why hypothesis testing should be done while considering other information and context if at all possible. If a single hypothesis test yields an unexpected result, then another sample should be taken and the test should be repeated. It is highly unlikely that two Type I or Type II errors would occur in successive samples from the same population.

The **α significance level** in a hypothesis test gives **the probability of making a Type I error**.

* If you do a hypothesis test at the significance level, you have a 0.01 probability (in other words, a 1% chance) of rejecting a null hypothesis that is actually true. To restate that, 1% of the time, a Type I error will happen.

*Example 1*:

A shampoo company has introduced a new formula of a particular shampoo. In the past, customers have been regularly surveyed about the shampoo, asked to rate the product on a scale of 1 to 10, and the shampoo has scored a mean of 6.2.

An analyst at the company would like to test whether the mean satisfaction score changed after the new formula was introduced.

Suppose the analyst did this hypothesis test and, based on the sample, rejected . However, in reality the population mean customer satisfaction score is 6.2 (i.e. a value that makes the Null Hypothesis true), this would be a Type I Error.

If the analyst did a hypothesis test and did not reject , but in reality the actual population mean customer satisfaction score was 5.1 (i.e. a value that means the Null Hypothesis is not true), this would be a Type II Error.

*Example 2*:

A human resources manager is interested in testing the effectiveness of a new training program. Prior to the training, employees could handle an average of 11.65 cases per day. The HR manager would like to determine whether the training has increased the average number of cases employees can handle per work day.

If the HR manager did a hypothesis test and did not reject , but the actual population mean of cases per day was 12, what type of error would this be?

What if the actual population mean was 11, and the HR manager rejected the null hypothesis?

What if the HR manager rejected the null hypothesis based on the sample, and the true population mean was 12.8?

*Example 3*:

A manufacturer makes test tubes for laboratory use. These test tubes must hold exactly 10 milliliters (ml) of liquid. If the production line goes out of adjustment, then it may produce test tubes that are on average too large or too small. A quality control analyst would like to check whether the production line is running properly.

What if the analyst rejected the null hypothesis based on the sample, and the true population mean was actually 10 ml?

What if the analyst rejected the null hypothesis based on the sample, and the true population mean was actually 10.06?

What if the analyst did not reject the null hypothesis, and the true population mean was actually 10ml?