**Introduction to ANOVA**



This section extends the work of comparing two population means to ≥ 2 population means!

Examples:

* Wheat varieties: Which variety of hard red winter wheat (Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof) provides higher yield on average?
* Grocery Stores: Is there a difference between average prices at these grocery stores: HyVee, Wal-Mart, or Super Saver?
* Allegry medication: Which drugs provide better relief on average to allergy sufferers: Seldane, Allegra, or Claritin?
* Fertilizer: Which type of fertilizer (A, B, or C) applied to a specific variety of wheat produces the highest yield on average?

Each example uses a hypothesis test of the form:

Ho: μ1 = μ2 = … = μt (No difference in population means)

Ha: At least one pair of means are not equal

The statistical procedure used for comparing the population means is called Analysis of Variance (ANOVA).

This name may seem strange since we are examining means; however, through examining the variances, differences between population means can be determined. Dot plots later in these notes will help illustrate this.

Terminology

Much of the terminology comes from agricultural applications, which is one of the first places where ANOVA methods were used.

* The process for collecting the sample data is called an experiment.
* The plan for collecting the data is the design of the experiment.
* Thus, “Experimental Design” is a general term used to denote how to properly design experiments.

Example: Grocery stores

Design: Randomly select 10 different items to be selected at each of three different grocery store chains. Preferably, these should be the same items at each store.

Experiment: Collecting the prices of the items.

Example: Fertilizer

Design: Randomly select field locations where each type of fertilizer is used separately.

Experiment: Planting the wheat and obtaining the yield for each field location.

* The variable measured is called the response variable.

Example: Grocery stores

Response variable: Prices

Example: Fertilizer

Response variable: Yield

The interest is in how the wheat “responds” to the fertilizer; hence, the name “response variable”.

* The variables that may be related to the response variables are called factors.

Example: Grocery stores

Factor: Grocery store chain

Example: Fertilizer

Factor: Fertilizer type

* The different levels of the factors are called factor levels. These levels could be equated to the different populations that are being compared. For example, we compared “two” factor levels when examining μ1 – μ2 earlier in the course.

Example: Grocery stores

Factor levels: HyVee, Wal-Mart, or Super Saver.

Example: Hard red winter wheat

Factor levels: Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof

* A treatment is a particular combination of levels of factor**s**

For an experiment with only one factor (all the examples so far), treatments are the same as factor levels. See the example below for when they are different.

Example: Which variety of hard red winter wheat (Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof) and fertilizer (A, B, and C) provides higher yield on average? One treatment would be variety = Newton and fertilizer = A.

The name “treatment” comes from items being “treated” with a combination of factor levels.

* The experimental unit is the object in which treatment is applied.

This is the “unit” of the “experiment” in which “treatment” is being applied to.

Example: Fertilizers

Suppose 12 field locations are available for the experiment. Four fields each receive one of the 3 (A, B, or C) fertilizer treatments (factor levels) in a random manner.

|  |  |  |
| --- | --- | --- |
| A | C | A |
| C | B | B |
| B | B | A |
| C | A | C |

Each field is representative of the population to which inferences will be made from examining the sample.

The fertilizer is applied to a field location, so the experimental unit is a field.

Example: Grocery Stores

The experimental unit is an item at the grocery stores for which a price is recorded.