Unit 6 Lecture: ANOVA

ANOVA stands for **Analysis of Variance.** It is much easier to understand ANOVA if you have a good understanding of that which has been covered in the course up to this point. So let us begin to understand ANOVA by reviewing:

We have already covered a two-group, **between-subjects design.** In a between-subjects design, the researcher (1) selects a sample from the population, (2) randomly assigns the sample to **two** conditions (e.g., control group and experimental group), (3) manipulates the independent variable (IV), and (4) collects data on the dependent variable (DV). The researcher would then calculate a mean (M) for the control group and a mean (M) for the experimental group. The appropriate hypothesis test that would allow the researcher to analyze the data from ***a two-group between-subject design*** is the Independent-Measures t-test.

Do you recall the following design from our initial discussion of between-subjects design?

Population
↓
Random Sample
↓
(Random Assignment)
(Splits the Sample)
↓ ↓

Control Group (Placebo)	Experimental group (IV)
(Measure DV)	(Measure DV)
3 4 5 7	5 4 3 7
M	M

In this Unit, we **expand on the randomized Between-Subject Research Design** by randomly assigning the sample to MORE THAN TWO GROUPS, which is the major difference between the two-group, between-subject (independent measures) design, which we have already covered, and this new thing called “ANOVA.” Again, the main difference is that the sample is randomly assigned to MORE than just two groups, like this:

Population

↓

Random Sample

↓

(Random Assignment)

↓

Control group	Experimental	Experimental	Experimental	Experimental
Placebo (measure DV)	IV Level (measure DV)	IV Level (measure DV)	IV Level (measure DV)	IV Level (measure DV)
3 5 6 7	3 4 2 1	4 5 3 2 1	5 6 4 5	2 3 4 5
4 7 6 6	8 5 5 3	6 5 4 1 2	0 2 3 4	6 7 4 6
M	M	M	M	M

In the above diagram, notice that the sample is **randomly assigned** to FIVE groups. Therefore, we are no longer comparing two groups, but rather we must analyze differences among FIVE group means.

When there is random assignment to more than two groups, and we need to compare more than two means, then the Independent Measures t-test is no longer appropriate.

You can think of t-test as “t for two.” And, when there are more than two groups, the t-tests will not be able to analyze the differences among more than two means.

When there are more than two groups in a between-subjects design, such as is the case here, we need to use ANOVA in order to analyze the data.

ANOVA

The “letter” for the ANOVA hypothesis test is **F**, officially called the **F-ratio**, but plain old F is good enough.

While ANOVA can become quite complex, we will focus on the most straightforward ANOVA called the ONE-WAY ANOVA.

To make it easier to follow, I am replicating the above design here:

Population
↓
Random Sample
↓
(Random Assignment)
↓

Control group	Experimental	Experimental	Experimental	Experimental
Placebo (measure DV)	IV Level (measure DV)	IV Level (measure DV)	IV Level (measure DV)	IV Level (measure DV)
3 5 6 7	3 4 2 1	4 5 3 2	5 6 4 5	2 3 4 5
4 7 6 6	8 5 5 3	6 5 4 1	0 2 3 4	6 7 4 6
M	M	M	M	M

ONE-WAY INDEPENDENT MEASURES ANOVA.

"One-Way" means that there is only one independent variable. In ANOVA, independent variables are referred to as factors (keep that in mind). If we had two independent variables, then it would be called a TWO-WAY ANOVA (we do not cover TWO-WAY ANOVA). I am sure you are relieved!

The **One-Way Independent Measures ANOVA** has the same general idea as any other hypothesis test. Therefore, you should recognize this general formula for the logic of ALL hypotheses tests:

$$\textbf{ALL HYPOTHESIS TESTS} = \frac{\textbf{Differences among means because of the IV}}{\textbf{Differences among means due to ERROR}}$$

I hope you can recognize that the numerator has the differences we believe is due to the independent variable and the denominator has the differences due to anything else but the independent variable (ERROR).

Now, for the F-test or One-Way Independent Measures ANOVA, we would have the following:

$$F = \frac{MS_{between} \quad \text{(Variance between = differences among means due to IV)}}{MS_{within} \quad \text{(Variance within = differences among means NOT due to IV effects) ERROR}}$$

Important New Terminology Specific to ANOVA

A new term is MS (mean square). MS is simply another word for variance. You may remember that simplistically speaking, variance means “differences.” In other words, if we had a basket of different kinds of fruits, we would have ‘VARIETY’ or variance among our fruits. Similarly, variance among data means that we want to analyze the differences (variance) that exist among our data. Therefore, the “analysis of variance, ANOVA,” analyzes the differences in our data.

Specifically, ANOVA, analyzes two very specific variances. It analyzes the differences (variance) among the data that we believe is due to the IV, and the differences (variance) among the data that we believe is due to error, but the overall logic of hypothesis testing remains the same:

$$F = \frac{MS_{between} \text{ (Variance between treatments) are the differences due to IV}}{MS_{within} \text{ (Variance within treatments) are the differences NOT due to IV (Error)}}$$

As you can hopefully see, we are still going to analyze the differences among the means due to the IV and compare those differences to the differences among the means due to ERROR.

Again, I can’t stress this enough: The major difference between the t-tests with which you are already familiar and this new thing called ANOVA is that in the t-tests, the independent variable had ONLY **two** levels. By levels, I mean HOW MUCH of the independent variable each group receives. For example, in a simple drug test, the researcher may randomly assign the sample only to **two** levels of the independent variable:

Control	vs	Experimental
(placebo)		(drug)
zero (0) amount of drug		some amount of drug

IF this were the case, I hope you can see that it is only two groups, where one group gets ZERO level of the drug and the other groups gets SOME AMOUNT (or level) of the drug. However, suppose the researcher wanted to know whether there would be different effects caused by different dosages (more levels) of the drug (the independent variable, the factor)? The researcher could simply randomly assign the sample to more than two levels of the factor and he would need a group for each of the levels he wanted to test:

Population

↓

Random Sample

↓

(random assignment)

↓

Control group	Experimental	Experimental	Experimental	Experimental
Placebo (0mg drug)	IV Level (10mg drug)	IV Level (20mg drug)	IV Level (40mg drug)	IV Level (50mg drug)
(measure DV)	(measure DV)	(measure DV)	(measure DV)	(measure DV)
3 5 6 7 4 7 6 6	3 4 2 1 8 5 5 3	4 5 3 2 1 6 5 4 1 2	5 6 4 5 0 2 3 4	2 3 4 5 6 7 4 6
M	M	M	M	M

Hopefully, you can see that this experiment has 5 levels of the factor (the independent variable, the drug). And because there is **more than two groups**, the F-ratio (ANOVA) is necessary in order to compare all 5 means. You will learn how to conduct the One-Way Independent-Measures ANOVA by viewing the Video Lecture. But for now, let me mention a few other important things you need to know to better understand the video.

Interpreting ANOVA Results

Once you run an ANOVA on SPSS, you will make the same decisions as you did in the t-tests. You will either reject the null hypothesis or fail to reject the null hypothesis depending on the probability outcomes. When you rejected the null hypothesis in the t-tests, and concluded that there was an effect, you KNEW WHERE the effect was: That is, you knew that the difference was found between the control and the experimental group or between whatever two groups you were comparing. After all, if you are just comparing two groups, then the difference you find MUST be between those two groups because there simply aren't any other groups.

However, in ANOVA, you DO have more than two groups. Therefore, if you reject the null hypothesis, you KNOW that there is a significant difference SOMEWHERE among the groups, perhaps even among all the groups; but, you do NOT KNOW which groups. This is important: you DO NOT KNOW which groups show the significant differences.

In order to find out which groups are statistically different, you must run a POST-HOC test. This is quite easy on SPSS, and is explained and demonstrated in the Video Lecture for this unit.

In our world of statistical packages, such as SPSS, no one is calculating hypothesis tests by hand and neither will you ☺ Therefore, the important thing is to UNDERSTAND what I have written here, gain a sense of the LOGIC OF HYPOTHESIS TESTING, observe the SPSS Video lecture to learn how to run the ANOVA, run the POST-HOC test, and read and interpret what the SPSS results show.