UNIT 3

One-Way Analysis of Variance

- ***ANOVA is used to abbreviate analysis of variance .***

VARIANCE In ANOVA

- Total variance is separated into two kinds: *within-treatment variance and between- treatments variance.*
- Within-treatment variance refers to the variability within a particular sample.
- Between-treatments variance refers to the variability between the treatment groups***.***
- *Between-treatments variance = Individual differences + Experimental error + Treatment*
- *Within-treatment variance = Individual differences + Experimental error*

THE *F-STATISTIC*

$$F = \frac{\text{Between} - \text{treatments variance}\left(\text{individual differences, experimental error, treatment effects}\right)}{\text{Within} - \text{treatment variance}\left(\text{individual differences, experimental error}\right)}$$

HYPOTHESIS TESTING WITH THE *F-STATISTIC*

The *null hypothesis asserts no difference between means*

H0: *$\mu 1 = \mu 2 = \mu 3 = \mu k$ (the k subscript indicates that there will be as many μs as there are groups)*

The *alternative hypothesis asserts that there will be significant differences between some of the means.*

H1: Some *μs are not equal*

THE *F* -DISTRIBUTION TABLE

- The *F -distribution, like the t -distribution, is a family of curves with varying degrees of freedom.*

There are several features of each table of which you should take note:

- There will be two *df:* one associated with between-groups variance (df_{bet}), which is the numerator in the *F*-statistic, and one associated with within-group variance (df_{wi}), the denominator. Df_{bet} runs across the top of the table and df_{wi} runs along the left column.
- Once again, the chart is abbreviated and not all *df* values are shown for the denominator. If the *df* for your particular research problem are not shown, use the next-lowest *df* value.
- If your obtained *F*-value is greater than the critical value listed in the chart, you can reject the null hypothesis because differences between means that large are unlikely to be due to chance. It is more likely that the differences occurred because of the effects of treatment.
- If your obtained *F*-value is less than the listed critical value of *F*, you will fail to reject H_0 because differences that small could have occurred by chance alone.

NOTATIONS FOR ANOVA

Analysis of variance involves working through a series of steps and learning some new notations. After reviewing these notations, we will work through the calculations of ANOVA within the context of an example.

- *Sum of squares (SS)* refers to the sum of the squared deviations of scores from the mean. We will be calculating three *SS* values: SS_{total}, SS_{wi}, and SS_{bet}.
- ΣX_{tot} refers to the total of all the scores in the study, which is determined by adding the sum of the X scores for each group.
- ΣX^2_{tot} refers to the total of all the squared scores in the study as determined by adding the sum of the squared scores for each group.
- N refers to the total number of scores in all groups.
- n refers to the number of scores in a sample group; particular groups are designated by subscripts.
- t as a subscript refers to individual treatment groups.
- k refers to the number of groups in the study.

Sample Research Question

A researcher is interested in determining if interactive activities influence the subjective experience of life satisfaction in older individuals. She obtains a random sample of 15 residents from an assisted living residential center, all of whom report only moderate degrees of life satisfaction. The subjects are randomly assigned to one of three interactive conditions: (1) an online chat group, (2) caring for pets, and (3) talking with volunteer students about their life experiences. After three months, the residents are given the Life Satisfaction Questionnaire. Higher scores designate greater life satisfaction. Do the scores indicate significant differences between the three groups? Use $\alpha = .05$.

Online chat		*Pets*		*Students*	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
3	9	8	64	3	9
3	9	12	144	7	49
4	16	9	81	10	100
6	36	7	49	6	36
4	16	9	81	9	81

$\Sigma X_1 = 20$	$\Sigma X_2 = 45$	$\Sigma X_3 = 35$
$\Sigma X_1^2 = 86$	$\Sigma X_2^2 = 419$	$\Sigma X_3^2 = 275$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$M_1 = 4$	$M_2 = 9$	$M_3 = 7$

$$\Sigma X_{tot} = 100$$
$$\Sigma X^2_{tot} = 780$$
$$N = 15$$
$$k = 3$$

THE CALCULATIONS

Sum of Squares (SS)

The first calculations in ANOVA will be three sum of squares values: SS_{total}, SS_{wi}, and SS_{bet}.

- SS_{total} refers to the sum of squares for all of the N scores. We will change the notation in the SS formula that we are used to:

$$SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

to make it relevant to our current statistical procedure for ANOVA:

$$SS_{total} = \Sigma X^2{}_{tot} - \frac{(\Sigma X_{tot})^2}{N}$$

For our research problem,

$$SS_{total} = 780 - \frac{(100)^2}{15}$$
$$= 113.33$$

- SS_{wi} refers to the variability within each group as measured by

$$SS_{wi} = \Sigma\left[\Sigma X_t^2 - \frac{(\Sigma X_t)^2}{n_t}\right]$$

This formula instructs you to calculate the sum of squares (*SS*) for each treatment group and then to sum each of the *SS* values. For our research problem,

$$\begin{aligned} SS_{wi} &= \left(86 - \frac{(20)^2}{5}\right) + \left(419 - \frac{(45)^2}{5}\right) + \left(275 - \frac{(35)^2}{5}\right) \\ &= 6 + 14 + 30 \\ &= 50 \end{aligned}$$

- SS_{bet} refers to the variability between each group as measured by:

$$SS_{bet} = \Sigma\left[\frac{(\Sigma X_t)^2}{n_t}\right] - \frac{(\Sigma X_{tot})^2}{N}$$

This formula instructs you to perform the operations in the brackets for each group before subtracting the expression at the end. For our research problem,

$$\begin{aligned} SS_{\text{bet}} &= \frac{20^2}{5} + \frac{45^2}{5} + \frac{35^2}{5} - \frac{100^2}{15} \\ &= 80 + 405 + 245 - 666.67 \\ &= 63.33 \end{aligned}$$

Degrees of Freedom (*df*)

We will be determining *df* for each of the three *SS* values above according to the following formulas:

- $df_{wi} = N - k$ For our data: $df_{wi} = 15 - 3 = 12$
- $df_{bet} = k - 1$ For our data: $df_{bet} = 3 - 1 = 2$
- $df_{total} = N - 1$ For our data: $df_{total} = 15 - 1 = 14$

df_{total} is not needed in our calculations for ANOVA. However, it also gives us a check because df_{wi} and df_{bet} should equal df_{total}. Thus,

$$df_{wi} + df_{bet} = df_{total}$$
$$12 + 2 = 14$$

Mean Square (*MS*)

You have previously learned that the variance is obtained by dividing *SS* by *df*. In ANOVA, ***mean square*** is a measure of variance that is also calculated by dividing *SS* by its corresponding *df*. We will calculate the between and within mean squares as follows:

$$MS_{\text{bet}} = \frac{SS_{\text{bet}}}{df_{\text{bet}}}$$

For our data:

$$MS_{\text{bet}} = \frac{63.33}{2} = 31.67$$

$$MS_{\text{wi}} = \frac{SS_{\text{wi}}}{df_{\text{wi}}}$$

For our data:

$$MS_{\text{wi}} = \frac{50}{12} = 4.17$$

F-Statistic

The final calculation in ANOVA is the *F*-statistic. You have been presented with a couple of different formulations for the *F*-statistic, the latest being:

$$F = \frac{\text{Treatment variance}}{\text{Error variance}}$$

Let us reformulate it one more time to make it usable for our ANOVA calculations. MS_{bet} is an expression for treatment variance, and MS_{wi} is an expression for error variance. Thus, the working formula we will use for the *F*-statistic is

$$F = \frac{MS_{\text{bet}}}{MS_{\text{wi}}} \quad \text{For our data:} \quad F_{\text{obt}} = \frac{31.67}{4.17} = 7.59$$

Within groups degrees of freedom	Between groups degrees of freedom (numerator) (df_{bet})												
(df_{wi})	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*14*
1	161.00	200.00	216.00	225.00	230.00	234.00	237.00	239.00	241.00	242.00	243.00	244.00	161.00
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.72	2.77	2.63	2.63	2.60	2.55

- We now have to compare our calculated *F-value to a critical value of F in order to determine if H0 should be rejected.*
- *We obtained an F-value of 7.59 with dfwi = 12 and dfbet = 2. We are using an α = .05.*
- *Consulting the F-distribution table, we find that the critical value for reject-ing H0 is 3.88*
- *. Because our obtained value exceeds the critical value for F, we will reject the H0 that interactive activities have no effect on life satisfaction.*

Summary Data

Finally, we will arrange our *obtained* values into a summary table that includes the following components:

Source	*SS*	*df*	*MS*	*F*	*p*
Between-treatments	63.33	2	31.67	7.59	$< .05$
Within-treatment	50.00	12	4.17		
Total	113.33	14			

Statement of Conclusion

Reject H_0. Interactive activities have a significant effect on life satisfaction.

Step 1: Formulate Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: some μs are not equal

Step 2: Indicate the Alpha Level and Determine Critical Values

$$\alpha = .05$$

$$df_{wi} = N - k = 15 - 3 = 12 \qquad df_{bet} = k - 1 = 3 - 1 = 2$$

$$F_{crit} = 3.88$$

Step 3: Calculate Relevant Statistics

1. Sum of Squares (SS)

$$SS_{total} = \Sigma X_{tot}^2 - \frac{(\Sigma X_{tot})^2}{N} = 780 - \frac{(100)^2}{15} = 113.33$$

$$SS_{wi} = \Sigma\left[\Sigma X_t^2 - \frac{(\Sigma X_t)^2}{n_t}\right] = \left(86 - \frac{(20)^2}{5}\right) + \left(419 - \frac{(45)^2}{5}\right) + \left(275 - \frac{(35)^2}{5}\right) = 50$$

$$SS_{bet} = \Sigma\left[\frac{(\Sigma X_t)^2}{n_t}\right] - \frac{(\Sigma X_{tot})^2}{N} = \frac{20^2}{5} + \frac{45^2}{5} + \frac{35^2}{5} - \frac{100^2}{15} = 63.33$$

2. Mean Square (MS)

$$MS_{\text{bet}} = \frac{SS_{\text{bet}}}{df_{\text{bet}}} = \frac{63.33}{2} = 31.67$$

$$MS_{\text{wi}} = \frac{SS_{\text{wi}}}{df_{\text{wi}}} = \frac{50}{12} = 4.17$$

3. The F-ratio

$$F_{\text{obt}} = \frac{MS_{\text{bet}}}{MS_{\text{wi}}} = \frac{31.67}{4.17} = 7.59$$

Research Question. A counseling psychologist conducts a study to determine the effectiveness of different approaches to treating generalized anxiety disorder. Subjects with this disorder are randomly assigned to three groups. One group undergoes cognitive therapy, a second group is trained in biofeedback, and a third untreated group serves as a control. Anxiety scores for each group are listed in the table below. Lower scores indicate lower levels of anxiety.

Cognitive		*Biofeedback*		*Control*	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
2		3		8	
4		4		7	
5		2		10	
3		4		6	
4		3		8	
6		2		9	

II. Analysis of Variance

A. Step 1:

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: some μs are not equal

Step 2:

$\alpha = .05$

$df_{wi} = 15$

$df_{bet} = 2$

$F_{crit} = 3.68$

Step 3:

1. $SS_{tot} = 108$
   $SS_{wi} = 24$
   $SS_{bet} = 84$

2. $MS_{bet} = 42$
   $MS_{wi} = 1.6$

3. $F_{obt} = 26.25$

Step 4:

There were significant differences among the treatment groups for generalized anxiety disorder. Reject H_0.

Source	*SS*	*df*	*MS*	*F*	*p*
Between-treatments	84	2	42	26.25	<.05
Within-treatment	24	15	1.6		
	108	17			