

## 11.3: Two-Way ANOVA (Factorial Design)

Two-way analysis of variance (two-way ANOVA) is an extension of one-way ANOVA. It can be used to compare the means of two independent variables or **factors** from two or more populations. It can also be used to test for interaction between the two independent variables.

We will not be doing the sum of squares calculations by hand. These numbers will be given to you in a partially filled out ANOVA table or an Excel output will be given in the problem.

There are three sets of hypotheses for testing the equality of  $k$  population means from two independent variables, and to test for interaction between the two variables (two-way ANOVA):

Row Effect (Factor A):	$H_0$ : The <u>row variable</u> has no effect on the average _____. $H_1$ : The <u>row variable</u> has an effect on the average _____. 
Column Effect (Factor B):	$H_0$ : The <u>column variable</u> has no effect on the average _____. $H_1$ : The <u>column variable</u> has an effect on the average _____. 
Interaction Effect (A×B):	$H_0$ : There is no interaction effect between <u>row variable</u> and <u>column variable</u> on the average _____. $H_1$ : There is an interaction effect between <u>row variable</u> and <u>column variable</u> on the average _____. 

These ANOVA tests are always right-tailed F-tests.

The F-test (for two-way ANOVA) is a statistical test for testing the equality of  $k$  independent quantitative population means from two nominal variables, called factors. The two-way ANOVA also tests for interaction between the two factors.

Assumptions:

1. The populations are normal.
2. The observations are independent.
3. The variances from each population are equal.
4. The groups must have equal sample sizes.

The formulas for the F-test statistics are:	
Factor 1:	$F_A = \frac{MS_A}{MSE}$ with $df_A = a - 1$ and $df_E = ab(n - 1)$
Factor 2:	$F_B = \frac{MS_B}{MSE}$ with $df_B = b - 1$ and $df_E = ab(n - 1)$
Interaction:	$F_{A \times B} = \frac{MS_{A \times B}}{MSE}$ with $df_{A \times B} = (a - 1)(b - 1)$ and $df_E = ab(n - 1)$

Where:

$SS_A$  = sum of squares for factor A, the row variable

$SS_B$  = sum of squares for factor B, the column variable

$SS_{A \times B}$  = sum of squares for interaction between factor A and B

$SSE$  = sum of squares of error, also called sum of squares within groups

$a$  = number of levels of factor A

$b$  = number of levels of factor B

$n$  = number of subjects in each group

It will be helpful to make a table. Figure 11-5 is called a two-way ANOVA table.

Source	SS = Sum of Squares	$df$	MS = Mean Square	F
A (row factor)	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{df_A}$	$F_A = \frac{MS_A}{MSE}$
B (column factor)	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{df_B}$	$F_B = \frac{MS_B}{MSE}$
A×B (interaction)	$SS_{A \times B}$	$(a - 1)(b - 1)$	$MS_{A \times B} = \frac{SS_{A \times B}}{df_{A \times B}}$	$F_{A \times B} = \frac{MS_{A \times B}}{MSE}$
Error (within)	$SSE$	$ab(n - 1)$	$MSE = \frac{SSE}{df_E}$	
Total	$SST$	$N - 1$		

Figure 11-5: Two-way ANOVA table template.

Since the computations for the two-way ANOVA are tedious, this text will not cover performing the calculations by hand. Instead, we will concentrate on completing and interpreting the two-way ANOVA tables.

A farmer wants to see if there is a difference in the average height for two new strains of hemp plants. They believe there also may be some interaction with different soil types so they plant 5 hemp plants of each strain in 4 types of soil: sandy, clay, loam and silt. At  $\alpha = 0.01$ , analyze the data shown, using a two-way ANOVA as started below in Figure 11-6. See below for raw data.

	Sandy	Clay	Loam	Silt
Strain A	60	54	80	62
	53	63	82	76
	58	62	62	55
	62	71	88	48
	57	76	71	61
Strain B	36	62	68	63
	41	61	72	65
	54	77	71	72
	65	53	82	71
	53	64	86	63

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	F
Strain	12.1			
Soil Type	2501			
Interaction	268.1			
Error				
Total	5108			

Figure 11-6: Data and initial ANOVA table.

### Solution

First, find the missing values for the ANOVA table in Figure 11-6. The sum of squares should add up to the total in the *SS* column of 5108.

Subtract the 3 sum of squares that were already given in the table from the total sum of squares to find the missing error sum of squares:  $5108 - 12.1 - 2501 - 268.1 = 2326.8$ .

Find the degrees of freedom for each category.

There are 2 types of hemp strains so  $a = 2$  and the  $df_A = 2 - 1 = 1$ .

There are 4 soil types so  $b = 4$  and the  $df_B = 4 - 1 = 3$ .

The interaction  $df_{A \times B} = (2 - 1)(4 - 1) = 3$ .

There are 5 plants within each group so  $n = 5$ , and  $df_E = 2 \cdot 4(5 - 1) = 32$ .

The total  $df$  should add up to  $N - 1$ . There are 8 groups of 5 so there are 40 data points all together;  $df_{Total} = 39$ . Fill in your *SS* and  $df$  to the ANOVA table and then divide the sum of squares by  $df$  to get the mean squares. See below.

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	F
Strain	12.1	1	12.1	
Soil Type	2501	3	833.6667	
Interaction	268.1	3	89.3667	
Error	2326.8	32	72.7125	
Total	5108	39		

Next, find the test statistic for by dividing the *MS* for each category by the *MSE*. For the row effect  $F_A = 12.1/72.7125 = 0.1664$  for the column effect  $F_B = 833.6667/72.7125 = 11.4652$  and for the interaction effect  $F_{A \times B} = 89.3667/72.7125 = 1.2290$  Add these to complete the ANOVA table.

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Strain	12.1	1	12.1	0.1664
Soil Type	2501	3	833.6667	11.4652
Interaction	268.1	3	89.3667	1.229
Error	2326.8	32	72.7125	
Total	5108	39		

There are three hypothesis tests that we are performing: the strain of hemp plant has an effect on the mean plant height (row effect), the type of soil has an effect on the mean plant height (column effect), and the interaction effect between strain of hemp plant and soil type on the mean plant height.

#### Row Effect-Factor A:

$H_0$  : The strain of hemp plant has no effect on the mean plant height.

$H_1$  : The strain of hemp plant has an effect on the mean plant height.

Test Statistic:  $F_A = 0.1664$ .

Critical Value: Use an F-distribution with  $\alpha = 0.01$  right-tail area,  $df_A = 1$  and  $df_E = 32$ .

The critical value is 7.4993. See Figure 11-7.

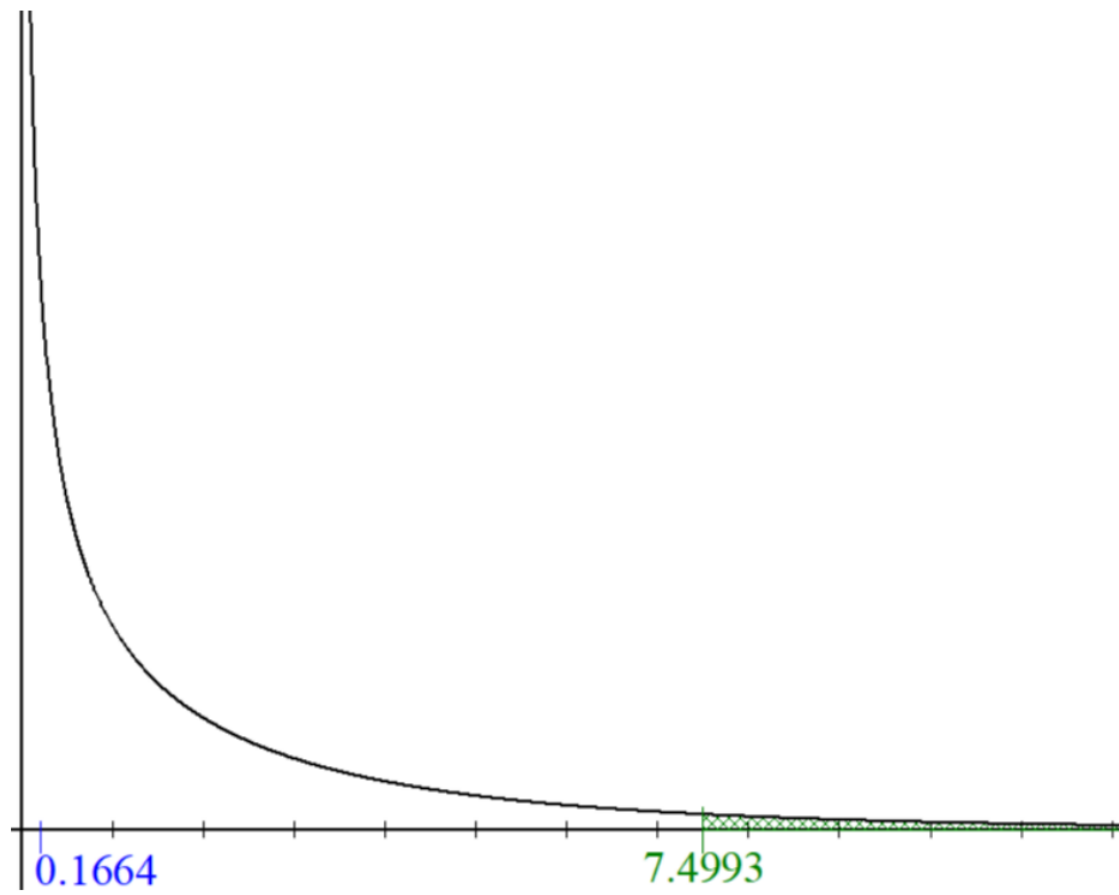


Figure 11-7: Graph of F-distribution for the row effect (plant strain).

The test statistic of 0.1664 is less than the critical value and not in the shaded tail; therefore, do not reject  $H_0$ . There is no significant difference in the mean plant height between the two hemp strains.

#### Column Effect-Factor B:

$H_0$  : The type of soil has no effect on the mean plant height.

$H_1$  : The type of soil has an effect on the mean plant height.

Test Statistic:  $F_B = 11.4652$

Critical Value: Use an F-distribution with  $\alpha = 0.01$  right-tail area,  $df_B = 3$  and  $df_E = 32$ . The critical value is 4.4594.

### Interaction Effect-Factor A $\times$ B

$H_0$  : There is no interaction effect between the strain of hemp and the soil type on the mean plant height.

$H_1$  : There is an interaction effect between the strain of hemp and the soil type on the mean plant height.

Test Statistic:  $F_{A \times B} = 1.229$  Figure 11-8: Graph of F-distribution for the column effect (soil type).

The test statistic is greater than the critical value and is in the shaded tail; therefore, reject  $H_0$ . The type of soil has a significant effect on the mean plant height.

Critical Value: Use an F-distribution with  $\alpha = 0.01$  right-tail area,  $df_{A \times B} = 3$  and  $df_E = 32$ . The critical value is 4.4594.

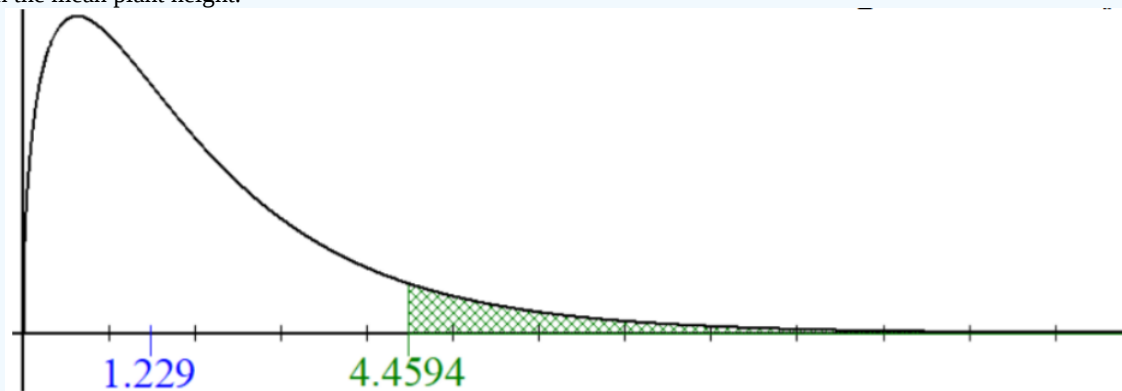


Figure 11-9: Graph of F-distribution for the interaction effect.

The test statistic is not in the shaded tail; therefore, do not reject  $H_0$ . There is no significant interaction effect between the hemp strain and the soil type on the mean plant height.

*Rough drawings from memory were futile. He didn't even know how long it had been, beyond Ford Prefect's rough guess at the time that it was "a couple of million years" and he simply didn't have the maths. Still, in the end he worked out a method which would at least produce a result. He decided not to mind the fact that with the extraordinary jumble of rules of thumb, wild approximations and arcane guesswork he was using he would be lucky to hit the right galaxy, he just went ahead and got a result. He would call it the right result. Who would know?*

*As it happened, through the myriad and unfathomable chances of fate, he got it exactly right, though he of course would never know that. He just went up to London and knocked on the appropriate door. "Oh. I thought you were going to phone me first."*

(Adams, 2002)

This page titled [11.3: Two-Way ANOVA \(Factorial Design\)](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Rachel Webb](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.