20. Analysis of Variance

In our previous problems we had one sample and tested whether our sample differed from a known population mean. In the analysis of variance we will test multiple samples – and hence multiple treatment groups – against each other. The basic idea used by the *AN*alysis *O*f *VA*riance (ANOVA) is that the total variability within the pooled sample has three components:

- the variability within each treatment group;
- the variability between the groups; and
- the residual variability-everything else.

ANOVA tests to see of the "between groups" differences are large enough to conclude that the groups are really different. For motivation, consider a couple of examples.

20.1. Example.

Suppose that there are three different diet plans each enrolling five clients. At the end of one month the following losses are recorded:

Plan A	Plan B	Plan C
5	15	10
5	15	10
5	15	10
5	15	10
5	15	10

The average loss in each plan is

$$\bar{\boldsymbol{x}}_{\boldsymbol{A}} = 5 \quad \bar{\boldsymbol{x}}_{\boldsymbol{B}} = 15 \quad \text{and} \ \bar{\boldsymbol{x}}_{\boldsymbol{C}} = 10$$

Remember that the population variance involved calculating how much each observation differed from the mean, $x_i - \bar{x}$, squaring it $(x_i - \bar{x})^2$, and then averaging the squared differences

$$rac{1}{n}\sum_i (x_i-ar{x})^{f 2}$$

ANOVA involves comparing sums of squared differences like the above. The differences

$$(\boldsymbol{x_i} - \bar{\boldsymbol{x}})$$

represent the error between the observed value and the average value. With the diet plan data, the sum of the squared differences over the entire sample

$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = 250$$

May 30, 2017

is the numerator in the variance

$$\sigma_{Total}^{2} = \left(\frac{\sum_{i=1}^{15} (x_{i} - \bar{x})^{2}}{15}\right) = 16.667$$

But there's another way to think about the variability. We could instead calculate the variability between the treatments, i.e., between the plans. Since the average loss in the three plans is

$$\frac{1}{3}(5+15+10) = 10,$$

the variability in the column averages is

$$\sigma^{2}_{\ plans} = \frac{(5-10)^{2} + (15-10)^{2} + (10-10)^{2})}{3} = \frac{50}{3} = 16.667.$$

In the above example, the two numbers, σ^2_{Total} and σ^2_{plans} are the same.

20. Analysis of Variance

This isn't surprising since

In this example, *all* of the variability in the sample is due to the differences in treatments (column effects) and *none* of the variability is due to individual differences in the clients

In principle, there are three sources of variability in our diet plan example:

- the variability within each treatment group;
- the variability between the groups; and
- the residual variability–all other sources of variability having nothing to do with the groups.

In our idealized example, the only variability is between the groups.

In the real world you will never get such perfect data. There will always be some variability due to influences other than the treatments (column effects). In the diet plan, initial weight, gender, age, exercise and other uncontrolled variables (including measurement error) will result in other sources of variability. More realistic data might similar to that given in the next example.

20.2. Example.

Suppose that the data from the three diet plans had been:

Plan A	Plan B	Plan C
5	10	5
5	15	5
5	15	10
10	15	10
15	15	10

Using the two methods above, find the overall variance and the variance due

to the treatment (column) effects.

Solution. We still have the same fifteen numbers as in the first example, except now those numbers are distributed differently between the plans. However, this won't change the total variance, so

$$\sigma^2_{Total} = 16.667$$

as before. The sample means are now

$$\bar{\boldsymbol{x}}_{\boldsymbol{A}} = 8$$
 $\bar{\boldsymbol{x}}_{\boldsymbol{B}} = 14$ and $\bar{\boldsymbol{x}}_{\boldsymbol{C}} = 8$

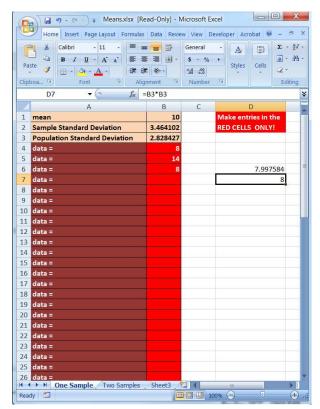
and so we can find the variance due to the column effects using the MEANS spreadsheet:

• enter the column means 8, 14 and 8 into the spreadsheet;

• find the population variance by squaring the population standard deviation. Since the spreadhsheet calculates the value of the population standard deviation in cell B3, you can do this by entering

=b3*b3

into a cell on the spreadsheet and pressing enter. This gives you a value of $\sigma^2_{PLans} = 8.$



Thus about 50% $(\frac{8}{16.667})$ of the total variability in the population is due to the differences in treatments. The remaining variability is due to differences within the treatment groups and to residual effects not attributable to the diet plans.

ANOVA lets you decide whether or not the kinds of differences observed in the previous example are statistically significant. In particular, ANOVA will test

 $H_0: \mu_A = \mu_B = \mu_C$ against

 H_A : the means are not all the same

Another way of saying the alternative hypothesis is:

 $H_0: \mu_A = \mu_B = \mu_C$ against $H_A:$ at least 2 differ one from the other

since if the means are not all the same, at least two must be different, one from the other.

The test will be accomplished by comparing the *variance* of the overall sample against the *variance* of the sample means. In practice, however, programs like the AnalyzeThis spreadsheet calculate sums of squared differences, as we did above. As we shall see, the report generated by

the ANOVA tab of AnalyzeThis includes sums of squared differences and averages—mean square differences—for each of the sources of variability mentioned above.

The basic assumptions of ANOVA are

- Each group is drawn from a normally distributed population;
- All populations have a common variance;
- all samples are independent;

• within each group sample, all observations are random and independent.

The test is fairly robust against violations of the normality assumption, provided that the sample sizes are equal, sufficiently large, and symmetrical. The test is less robust against violations the second assumption, that all the group populations have the same variance. It's also possible to test for violations of the above assumptions, although we won't cover that in this course.

Remark. Usually ANOVA is only done for three or more treatment groups. When you have just two groups, hypothesis tests for means

are theoretically equivalent to ANOVA and computationally simpler. If you are given the actual data, the AnalyzeThis spreadsheet will do all the necessary calculations and perform the ANOVA test for you. If you are just given summary information—for example on the final exam—you will use the Formulas spreadsheet and will be given the following information:

- Sample means for each treatment group.
- Overall sample standard deviation *s Total*.
- Sample size *n* for each treatment group.
- Significance level.

20.3. Example.

An attorney is interested in whether knowledge of prior bad acts-criminal convictions-will influence potential juror's view of the guilt or innocence of a person accused of a crime. She randomly samples potential jurors and then divides her sample into three groups. Group I is told that the accused has a criminal record, Group II is told that the accused has no criminal record, and Group III is given no information on prior bad acts. Each subject then fills out a questionnaire measuring how likely it is, in the view of the subject, that the accused is guilty.

Group			
mean	8	3.8	5
sample size	5	5	5

The pooled standard deviation was 2.8. At the 1% significance level do the data show that information on prior bad acts influences juror perceptions?

Solution.

Step 1. The above table, together with the pooled standard deviation of 2.8 provides the dictionary needed for the spreadsheet.

Step 2. Now enter the summary data and the pooled standard deviation into the spreadsheet FORMULAS.XLSX, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled ANOVA.

Step 3. Read the significance level from the spreadsheet. If it's less than the pre-set target, then we reject the null hypothesis, otherwise we accept the null hypothesis.

	B				l Willian					
F	File Home Insert D	raw Pag	e Layout	Formulas	Data	Review	View	Q Tel	l me	۶
1	Calibri + 11	* 3	9	6 Con	ditional Forr	natting *		2		
	ste	A Alian	ment Num	ber Form	nat as Table	÷	Cells	Editing		
Р <i>а</i>	ste 💉 🗄 - 🙆 - 🗛 -		v v	Cell	Styles *		v v	*		
Clip	pboard 5 Font	6			Styles					
F1	2 * 1 × ~	fx								
Å	A	В		с	D	E		F	G	
1	Analysis of Variance									
2	For up to EIGHT treatment g	roups								
3	Data									
4	Pooled Standard Deviation	2.8								
5		Mean	Sample S	Size						
6	Treatment Group 1	8		5						
7	Treatment Group 2	3.8		5						
8	Treatment Group 3	5		5						
9	Treatment Group 4			1						
0	Treatment Group 5									
1	Treatment Group 6			1						
2	Treatment Group 7									
3	Treatment Group 8			1						
4	Leave sample data BLANK i	f you have	no data fe	or that samp	le!					
5	Calculations									
6			Degrees	of Freedom						
7	Total Variance	7.84		14						
8	Column-Effect Variance	3.12		2						
9	Residual Variance	4.72		12						
20										
21	Test Statistic	3.966102								
2	p-value	4.7616%								
3							_			
4		SS	dF		MSSq	F	р			
25	Total	117.6		14	8.4	3.966102	2 4	.76%		
26	Within Groups	70.8		12	5.9					
27	Between Groups	46.8		2	23.4					
8										
9										

In this case, the pre-set significance level was 1% and the p-value is 4.76%, so we do NOT reject the null hypothesis. If we believe that the

means are not all the same, then the chance of a Type I Error is 4.76% Another way of stating this concluding is that the results are not highly significant but are significant since 1% .

Remark. Note that an ANOVA is *always* a right-tailed test.

Solution Template

Step 1. Build your dictionary if it's not already in the problem:

pooled standard deviation	σ
Mean for treatment group 1	$ar{x}_1$
Mean for treatment group 2	$ar{x}_{2}$
Mean for treatment group 3	$ar{x}_{3}$

The spreadsheet permits up to eight treatment groups, although ANOVA in general works for any finite number of treatment groups.

Step 2. Now enter the summary data and the pooled standard deviation into the spreadsheet FORMULAS.XLSX, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled ANOVA.

Step 3. Read the significance level from the spreadsheet. If it's less than the pre-set target, then we reject the null hypothesis, otherwise we accept the null hypothesis. ANOVA is always a right-tailed test

	⊟ ち・♂・६ + ÷	raw Page	Formulas.xlsx - e Layout Formul					Tell me	6
Pa	Calibri • 11		ment Number	Form	ditional Forr nat as Table Styles * Styles			P liting	
F1	2 * : × ~	fx							
Å	A	В	с		D	E	F	G	
1	Analysis of Variance				-				
2	For up to EIGHT treatment g	roups							
3	Data								
4	Pooled Standard Deviation	2.8							
5		Mean	Sample Size	1					
6	Treatment Group 1	8		5					
7	Treatment Group 2	3.8		5					
8	Treatment Group 3	5		5					
9	Treatment Group 4			1					
10	Treatment Group 5			Î					
11	Treatment Group 6			1					
12	Treatment Group 7			Ĩ					
13	Treatment Group 8			Ĩ					
14	Leave sample data BLANK in	fyou have	no data for that s	amp	le!				
15	Calculations								
16			Degrees of Freed	lom					
17	Total Variance	7.84		14					
18	Column-Effect Variance	3.12		2					
19	Residual Variance	4.72		12					
20									
21	Test Statistic	3.966102		1					
22	p-value	4.7616%							
23									
24		SS	dF		MSSq	F	р		
25	Total	117.6		14	8.4	3.966102	4.76	%	
26	Within Groups	70.8		12	5.9				
27	Between Groups	46.8		2	23.4				
28									
29									

End of Solution Template

The spreadsheet reports some other calculations in addition to the p-value.

First, it calculates the **total variance**, which is the square of the pooled standard deviation.

Second, it calculates the **column effect variance**, which is the amount of variability due to the treatments. The spreadsheet uses the following formula (in the case of four treatment groups):

$$V_C = rac{n_A ar{x}_A^2 + n_B ar{x}_B^2 + n_C ar{x}_C^2 + n_D ar{x}_D^2}{n_A + n_B + n_C + n_D} - ar{x}^2$$

where \bar{x} is the pooled average.

In the example with the lawyer and the potential jurors, all three groups had the same sample size—a balanced sample. This is not necessary for ANOVA—the groups can have different sizes.

Third, the spreadsheet calculates the **residual variance**, which is the variability due to all other sources than the treatments. This is the column variance subtracted from the pooled variance.

In addition to the above items, it calculates the **degrees of freedom** for each quantity. For the total variance, this is the size of the pooled

sample minus one. For the column effects, it is the number of treatment groups minus one. For the residual variance, it's the difference between these two numbers.

The above all needed to calculate the test statistic for ANOVA:

 $F = rac{v_C/(ext{degrees of freedom for } v_C)}{v_R/(ext{degrees of freedom for } v_R)}$

Similar to our initial motivating example, the F statistic compares two sources of variation in the data:

- Variation due to differences in the columns; and
- Variation due to residual effects.

Since the spreadsheet does all of these calculations for you, they are included here just for completeness.

Note that the spreadsheet also has a second table using "Sum of Squares" and "Mean Sum of Squares" statistic. The values for SS_{Total} and SS_{Between} are calculated as before, while

$$SS_{\text{Within}} = SS_{\text{Total}} - SS_{\text{Between}}$$
.

The degrees of freedom are

SS Total	total sample size - 1
$SS_{Between}$	number of groups - 1
$old SS$ $_{Within}$	difference of the above

while the Mean Sum of Squares column MSSq is the Sum of Squares SS divided by the corresponding degrees of Freedom dF. The F Statistics F is

 $rac{MSSq_{Between\ Groups}}{MSSq_{Within\ Groups}}.$

It turns out that the two approaches to calculating the F statistic are equivalent. The "sum of squares" approach takes somewhat better advantage of the precision of the computer and is slightly faster on large data sets, which is why it's the preferred method.

Thus, the F statistic, the p-value, and the conclusions are the same in this both tables.

The good news is that the spreadsheet does all of these calculations for you. You need to understand conceptually that the test is comparing sources of variability in the data and be able to interpret the p-value.

Finally, the AnalyzeThis spreadsheet produces the same results, but starts with the raw scores. The final numbers may be slightly different since AnalyzeThis uses the full precision of the computer while we rounded the test averages and pooled standard deviation in creating the Formulas Version.

	File Home Insert [)raw Page L	avout 6	ormulas	Data F	leview \	/iew ♀ Tell	me what yo	unumet to	do 0	. Sha
r.							100				. 300
1	Calibri 👻	11 · A A					litional Formattin			$\sum \cdot \frac{1}{2} \cdot \cdot$	
as	ste <u>∛</u> B <u>I U</u> → ⊞			≣ ⊞ -	\$ • % *		at as Table *		elete 🔹	¥ - 2 -	
	• • • • • • • •	· 🎽 · 🏯 ·	<u>€</u> <u>∃</u>	87 -	00, 00 →.0	Cell :	Styles *	F F	ormat *	🧶 v	
Clip	pboard 🕞 Font	ŗ	a Aligne	ient 🕠	Number	1 5	Styles	0	Cells	Editing	1.3
D7	7 * 1 × V	fx									
		J.4									
4	A	В	С	D	E	F	G	н	1	J	
-	Pooled Mean	5.6					Variables a	are quantit	tative and	d generally	C
3	Pooled StdDev	2.8					continunous a	as opposed	d to discr	ete.	
	Alpha	0.05									
0											Pu
1		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	3	in Do
	Mean	8		5							Do
3	Sample Size	5	5	5							Th
-											
4					1					-	
	Table I	Value	Degrees o	f Freedom			, this spreadsh			ne	
5	Table I Total Variance	Value 7.84	_		analys	is in Table	and in Table II	. Table II o	contains		
5			14		analys inform	is in Table I nation on "s	and in Table II ums of squares	. Table II o	contains an square		
7 8	Total Variance Column Effect Var Residual Variance	7.84	14 2		analys inform	is in Table I nation on "s	and in Table II	. Table II o	contains an square		
5 6 7 8 9	Total Variance Column Effect Var Residual Variance Test Statistic F	7.84 3.12 4.72 3.966101695	14 2 12		analys inform	is in Table I nation on "s	and in Table II ums of squares	. Table II o	contains an square		
5 6 7 8 9	Total Variance Column Effect Var Residual Variance	7.84 3.12 4.72	14 2 12		analys inform	is in Table I nation on "s	and in Table II ums of squares	. Table II o	contains an square		
5 6 7 8 9	Total Variance Column Effect Var Residual Variance Test Statistic F	7.84 3.12 4.72 3.966101695	14 2 12		analys inform	is in Table I nation on "s	and in Table II ums of squares as the same con	. Table II o	contains an square		
5 5 7 8 9 0 1 2	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II	7.84 3.12 4.72 3.966101695 4.76% SS	14 2 12 dF	MSSq	analys inform statist	is in Table nation on "s ics, but give	and in Table II ums of squares the same con Critical Value	. Table II o	contains an square		
5 6 7 8 9 0 1 2 3	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total	7.84 3.12 4.72 3.966101695 4.76% SS 117.6	14 2 12 dF 14	MSSq 8.4	analys inform statist F 3.966102	is in Table nation on "s ics, but give	and in Table II ums of squares the same con Critical Value	. Table II o	contains an square		
5 6 7 8 9 0 1 2 3 4	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups	7.84 3.12 4.72 3.966101695 4.76% SS 117.6 70.8	14 2 12 dF 14 12	MSSq 8.4 5.9	analys inform statist F 3.966102	is in Table nation on "s ics, but give	and in Table II ums of squares the same con Critical Value	. Table II o	contains an square		
5 5 7 8 9 0 1 2 8 4 5	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total	7.84 3.12 4.72 3.966101695 4.76% SS 117.6	14 2 12 dF 14	MSSq 8.4	analys inform statist F 3.966102	is in Table nation on "s ics, but give	and in Table II ums of squares the same con Critical Value	. Table II o	contains an square		
5 6 7 8 9 0 1 2 3 4 5 6	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups	7.84 3.12 4.72 3.966101695 4.76% 5S 117.6 70.8 46.8	14 2 12 dF 14 12 2	MSSq 8.4 5.9 23.4	F 3.966107	p 4.76%	and in Table II ums of squares es the same con Critical Value 3.8853	. Table II of " and "me inclusions a	contains an square		
5 6 7 8 9 0 1 2 3 4 5 6 7	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups	7.84 3.12 4.72 3.966101695 4.76% 5S 117.6 70.8 46.8	14 2 12 dF 14 12 2	MSSq 8.4 5.9 23.4	F 3.966107	p 4.76%	and in Table II ums of squares es the same con Critical Value 3.8853	. Table II of " and "me inclusions a	contains an square		
5 6 7 8 9 0 1 2 3 4 5 6 7 8	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups	7.84 3.12 4.72 3.966101695 4.76% SS 117.6 70.8 46.8 at all the column	14 2 12 dF 14 12 2 mn means	MSSq 8.4 5.9 23.4 are the san	F 3.966102	p 4.76%	and in Table II ums of squares is the same con Critical Value 3.8853 tive that they a	. Table II d " and "me inclusions a	contains an squard s Table I.	e"	
5 5 7 B 9 0 1 2 B 4 5 5 7 B 9	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hypothesis th	7.84 3.12 4.72 3.966101695 4.76% SS 117.6 70.8 46.8 at all the colum Group 1	dF 14 12 dF 14 12 2 mn means Group 2	MSSq 8.44 5.9 23.4 are the sam Group 3	F analys inform statist 3,966102 Group 4	p 4.76%	and in Table II ums of squares es the same con Critical Value 3.8853	. Table II of " and "me inclusions a	contains an square	e"	
5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hypothesis th Subject1	7.84 3.12 4.72 3.966101695 4.76% 55 117.6 70.8 46.8 46.8 at all the coluu Group 1	14 2 12 dF 14 12 2 mn means Group 2 5	MSSq 8,4 5,9 23,4 are the sam Group 3 4	F 3.966102 Group 4	p 4.76%	and in Table II ums of squares is the same con Critical Value 3.8853 tive that they a	. Table II d " and "me inclusions a	contains an squard s Table I.	e"	
5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hypothesis th Subject1 Subject2	7.84 3.12 4.72 3.966101695 4.76% 55 55 117.6 70.8 46.8 at all the colum Group 1 10 7	14 2 12 dF 14 12 2 mn means Group 2 5 0	MSSq 8.4 5.9 23.4 are the sam Group 3 4 6	F 3.966102 Group 4	p 4.76%	and in Table II ums of squares is the same con Critical Value 3.8853 tive that they a	. Table II d " and "me inclusions a	contains an squard s Table I.	e"	
5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 1 2	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hypothesis th Subject1 Subject2 Subject3	7,84 3.12 4,72 3,96601069 4,76% SS 117.6 70.8 46.8 at all the colum Group 1 10 7 5	14 2 12 dF 14 12 2 mn means Group 2 5 0 3	MSSq 8.4 5.9 23.4 are the sam Group 3 4 6 9	F 3.966102 Group 4	p 4.76%	and in Table II ums of squares is the same con Critical Value 3.8853 tive that they a	. Table II d " and "me inclusions a	contains an squard s Table I.	e"	
5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3	Total Variance Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hypothesis th Subject1 Subject2	7.84 3.12 4.72 3.966101695 4.76% 55 55 117.6 70.8 46.8 at all the colum Group 1 10 7	14 2 12 dF 14 12 2 mn means Group 2 5 0 3 3 7	MSSq 8.4 5.9 23.4 are the sam Group 3 4 6	F 3.966107 Group 4	p 4.76%	and in Table II ums of squares is the same con Critical Value 3.8853 tive that they a	. Table II d " and "me inclusions a	contains an squard s Table I.	e"	

20.4. Example.

A researcher is interested whether the presence of pets will influence the socialization of tenants in long-term care facilities. She selects four group homes and chooses a random sample of tenants from each home as follows:

Group 1	10 tenants	group areas include dogs
Group 2	7 tenants	group areas in include cats
Group 3	15 tenants	both cats and dogs
Group 4	13 tenants	no pets

After four weeks, the researcher scores each subject on a standardized socialization scale and obtains the following results.

Group 1	Group 2	Group 3	Group 4
67	69	80	64
78	65	78	81
82	83	62	57
79	66	67	65
84	64	74	84
100	69	71	80
83	71	87	78
75		78	66
73		70	71
75		86	74
		72	72
		76	69
		59	58
		81	
		71	

Can the researcher conclude that there is a difference in the socialization between the three groups?

Solution.

Notice that in this problem we have the raw data instead of summary data. Using the AnalyzeThis spreadsheet makes this easy—all we have to do is enter the data. See also the spreadsheet ANOVAExample.XLSX, which has the data pre-loaded.

Since the p-value is 4.39%, we have significant but not highly significant evidence that the treatment groups have different socialization levels.

	∃ <u>∱ਾ ੇ -</u> %			Analyze	This.xlsx - Ex	cel	William Ray	- 🖬		×
	ile Home	Insert	Draw Page	Layout f	ormulas	Data R	eview V	liew ♀ Tel	Ime 🖇	≩ Sha
-	📉 🔏 🛛 Calibri	* 11	• = =	=	General 🔻	Cond	itional Form	atting *	P	
	B - B I	<u>U</u> - A	• A' ≡ ≡	≣ ⊡ •	\$ • % •	Forma	at as Table *	Cells		
'a:	ste 😽 🔟 - 📘	- A	• •	87-	€.0 .00 .00 →.0	🕎 Cell S	tyles 🕶	v (Celi:	* Lutting	
lip	board 🕞 I	Font	🕞 Aligni	nent 🕞	Number 5		Styles			
32	9 + 1	× .	f= Gr	oup 1						
1	A		В	с	D	E	F	G	н	
	Pooled Mean		73,6444444					10%	are quanti	itativ
	Pooled StdDev		8.57944716					continunous		
	Alpha		0.0	5						
0										
1			Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Gr
2	Mean		79.	6 69.57143	74.133333	70.69231				
3	Sample Size		1	0 7	15	13				
4										
5	Table I		Value	Degrees o	of Freedom			, this spreads		
-	Total Variance		73.6069135	8 44				and in Table		
D										
2	Column Effect Var	r	13.0598724	7 3			ation on "s			
7	Column Effect Var Residual Variance		13.0598724 60.5470411					is the same co		
7 8 9	Column Effect Var Residual Variance Test Statistic F		60.5470411 2.94787194	1 41 4						
7 8 9	Column Effect Var Residual Variance		60.5470411	1 41 4						
7 8 9 0	Column Effect Var Residual Variance Test Statistic F p-value		60.5470411 2.94787194 4.399	1 41 4 6		statisti	cs, but give	s the same co	onclusions a	
6 7 8 9 0 1 2	Column Effect Var Residual Variance Test Statistic F p-value Table II		60.5470411 2.94787194 4.399 SS	dF	MSSq	statisti F	cs, but give	the same co Critical Value	e	
7 8 9 0 1 2 3	Column Effect Var Residual Variance Test Statistic F p-value Table II Total		60.5470411 2.94787194 4.399 55 3312.31111	1 41 4 6 dF 1 44	75.279798	statisti F	cs, but give	s the same co	e	
7 8 9 0 1 2 3 4	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups		60.5470411 2.94787194 4.399 55 3312.31111 2724.6168	1 41 4 6 dF 1 44 5 41	75.279798 66.45407	statisti F	cs, but give	the same co Critical Value	e	
7 8 9 0 1 2 3 4 5	Column Effect Var Residual Variance Test Statistic F p-value Table II Total		60.5470411 2.94787194 4.399 55 3312.31111	1 41 4 6 dF 1 44 5 41	75.279798	statisti F	cs, but give	the same co Critical Value	e	
7 8 9 0 1 2 3 4 5 6	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261	1 41 4 6 1 44 5 41 3 3	75.279798 66.45407 195.89809	F 2.947872	p 4.39%	Critical Value	e P	
7 B 9 0 1 2 3 4 5 5 7	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261	1 41 4 6 1 44 5 41 3 3	75.279798 66.45407 195.89809	F 2.947872	p 4.39%	Critical Value	e P	
7 B 9 0 1 2 3 4 5 6 7 B	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261	1 41 4 6 1 44 5 41 3 3	75.279798 66.45407 195.89809	F 2.947872	p 4.39%	Critical Value	e P	
7 B 9 0 1 2 3 4 5 6 7 B 9	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261	1 41 4 6 7 44 5 41 3 3 3 3 9 10 10 10 10 10 10 10 10 10 10 10 10 10	75.279798 66.45407 195.89809 are the sam	F 2.947872 e against t	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 B 9 0 1 2 3 4 5 6 7 B 9 0 1 2 3 4 5 6 7 B	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp		60.5470411 2.94787194 4.399 55 3312.31111 2724.6168 587.694261 hat all the colu Group 1	dF dF 1 44 5 41 3 3 3 6 6 7 69	75.279798 66.45407 195.89809 are the sam Group 3	F 2.947872 Group 4	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 B 9 0 1 2 3 4 5 6 7 B 9 0 1	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 hat all the column Group 1 6 7 8	1 41 4 6 1 444 5 411 3 3 3 4 6 6 6 7 69 8 65 2 83	75.279798 66.45407 195.89809 are the sam Group 3 80	F 2.947872 Group 4 64	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 B 9 0 1 2 3 4 5 5 7 B 9 0 1 2 3 4 5 5 7 B 9 0 1 2 3 4 5 5 7 B 9 0 1 2 3 4 5 5 5 7 8 9 0 1 2 3 4 5 5 5 7 8 9 0 1 2 3 4 5 5 7 8 9 0 1 2 3 4 5 5 7 8 9 0 1 2 3 4 5 5 7 8 9 0 1 2 3 4 5 5 7 8 8 9 0 1 2 3 4 5 5 7 8 8 8 8 8 8 8 8 8 8 8 8 8	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 hat all the colu Group 1 6 7 8 8 7	1 41 4 6 1 44 5 41 3 3 3 3 6 7 7 69 8 65 2 83 9 666	75.279798 66.45407 195.89809 are the sam Group 3 80 78 62 67	F 2.947872 Group 4 64 811 577 655	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 B 9 0 1 2 3 4 5 6 7 B 9 0 1 2 3 4	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4 Subject5		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 hat all the colu Group 1 6 7 8 8 7 8	1 41 4 6 5 41 3 3 6 7 6 7 6 7 7 69 8 65 2 83 9 66 4 64	75.279798 66.45407 195.89809 are the sam Group 3 80 78 62 67 74	F 2.947872 Group 4 64 81 577 65 84	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 8 9 9 0 1 1 2 3 3 4 4 5 5 6 7 7 8 9 9 0 0 1 1 1 2 3 3 4 4 5 5 5 5 5 5 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4 Subject5 Subject6		60.5470411 2.94787194 4.389 SS 3312.31111 2724.6168 587.694261 hat all the colu Group 1 6 7 8 7 8 10	1 41 4 6 6 1 7 69 8 65 2 83 9 66 4 64 0 69	75.279798 66.45407 195.89809 are the sam Group 3 80 78 62 67 74 71	F 2.947872 Group 4 64 81 577 65 844 80	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 8 9 9 0 1 2 3 3 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4 Subject5 Subject6 Subject7		60.5470411 2.94787194 4.389 SS 3312.31111 2724.6168 587.694261 hat all the colu Group 1 6 7 8 7 8 7 8 10 8 10 8	1 41 4 6 dF 1 4 5 4 5 3 3 3 3 3 3 4 6 5 411 3 3 3 3 3 3 4 6 0 699 3 71	75.279798 66.45407 195.89809 are the sam Group 3 800 78 62 67 74 74 71 87	F 2.947872 Group 4 644 811 577 655 844 800 78	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 8 9 9 0 1 1 2 3 3 4 5 5 6 7 7 8 8 9 9 0 1 1 2 2 3 3 4 4 5 5 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4 Subject5 Subject6 Subject7 Subject8		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 Mat all the column Group 1 6 6 77 8 8 77 8 8 77 8 8 77 8 8 77 8 8 77 8 8 77	1 41 4 6 1 44 5 41 3 3 3 3 4 6 6 6 7 69 8 655 2 83 9 66 4 640 0 69 3 71 5 5	75.279798 66.45407 195.89809 are the sam Group 3 800 78 62 67 74 71 877 877	F 2.947872 Group 4 64 81 57 65 84 80 78 8 66	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 8 9 9 0 1 2 2 3 4 4 5 5 6 7 7 8 9 9 0 0 1 1 2 2 3 3 4 4 5 5 6 7 7 7 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject3 Subject4 Subject5 Subject6 Subject7 Subject8 Subject9		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 hat all the colu 6 6 7 8 8 7 8 8 7 8 8 7 7 8 8 7 7 8 8 7 7 8 8 7 7 8 7 7	1 41 4 6 1 441 5 41 3 3 Group 2 7 7 69 8 655 2 833 9 666 4 640 0 693 3 71 5 3	75.279798 66.45407 195.89809 are the sam Group 3 80 78 62 67 74 71 87 78 778 778 778	F 2.947872 Group 4 64 81 577 655 844 800 78 666 71	p 4.39% he alterna	Critical Value 2.832	e r are not.	
7 8 9 9 9 0 1 1 2 2 3 4 4 5 5 7 7 8 9 9 0 0 1 1 1 2 2 3 3 4 4 5 5 7 7 7 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9	Column Effect Var Residual Variance Test Statistic F p-value Table II Total Within Groups Between Groups Tests the null hyp Subject1 Subject2 Subject3 Subject4 Subject5 Subject6 Subject7 Subject8		60.5470411 2.94787194 4.399 SS 3312.31111 2724.6168 587.694261 At all the column Group 1 6 6 77 8 8 77 8 8 77 8 8 77 8 8 77 8 8 77 8 8 77	1 41 4 6 1 441 5 41 3 3 Group 2 7 7 69 8 655 2 833 9 666 4 640 0 693 3 71 5 3	75.279798 66.45407 195.89809 are the sam Group 3 800 78 62 67 74 71 877 877	F 2.947872 Group 4 64 811 577 655 844 800 788 666	p 4.39% he alterna	Critical Value 2.832	e r are not.	

20.5. Example.

Researchers are interested in the relationship, if any, between alcohol consumption and blood pressure. A group of 32 non-smoking females, aged 20-65, were divided into four groups as follows

- Group A: no alcohol consumption for two weeks;
- Group B: 13 ml of red wine daily for two weeks;
- Group C: 13 ml of non-alcoholic red wine daily for two weeks;
- Group D: 1125 ml of beer daily for two weeks.

	Group A	Group B	Group C	Group D
change in blood pressure	-0.1	1.9 mm	1.9 mm	2.9 mm
# of subjects	8	8	9	9

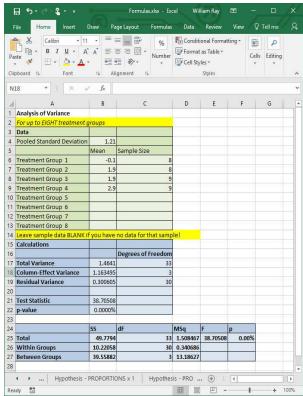
For this sample, the pooled standard deviation was 1.21 mm. At the 5% significance level, do the above data show that there is a difference in the blood pressure response times between groups?

Solution.

Step 1. The above table provides the summary data for four treatment groups and for a pooled standard deviation of 1.21.

Step 2. Now enter the summary data and the pooled standard deviation into the spreadsheet FORMULAS.XLSX, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled ANOVA.

Step 3. Read the significance level from the spreadsheet. If it's less than the pre-set target, then we reject the null hypothesis, otherwise we accept the null hypothesis.



In this case, the pre-set significance level was 5% and the $p\mbox{-value}$ is .0000%, so we reject the null hypothesis. The chance of a Type I Error

is less than 0.0000%.