

Experimental design and biometric research. **Toward innovations**

Sylwester Białowąs Editor





elSBN 978-83-8211-079-1 https://doi.org/10.18559/978-83-8211-079-1

PUEB PRESS



POZNAŃ UNIVERSITY **OF ECONOMICS** AND BUSINESS

© Copyright by Poznań University of Economics and Business Poznań 2021



This textbook is available under the Creative Commons 4.0 license – Attribution-Noncommercial-No Derivative Works



INDEPENDENT SAMPLES— MORE HYPOTHESES TESTING



Blaženka Knežević

Faculty of Economics and Business, University of Zagreb



Berislav Žmuk

Faculty of Economics and Business, University of Zagreb

Abstract: Two-way analysis of variance (ANOVA) without replication is called a factorial ANOVA with two factors. It is used to test if there is a significant difference between means of several sets of data (groups) dependable on two independent factors. It is applied when we have one measurement variable and two nominal variables (usually called 'factors' or 'main effects'). In this chapter hypotheses and assumptions of the method are given. Then the example of the procedure of two-way analysis of variance (ANOVA) without replication is described in details. The two-way analysis of variance (ANOVA) with replication is utilized to simultaneously test the effects of varying two variables for a sample which consists of more than one respondent per a certain combination of variables. The example of the procedure of two-way analysis of variance (ANOVA) with replication is described in details in this chapter. For both procedures the easy to follow examples shows the procedure step-by-step. The practical part includes the guidance for SPSS and for Excel.

Keywords: analysis of variance, ANOVA, two-way analysis of variance without replication, two-way analysis of variance with replication.

2.1. Two-way analysis of variance (ANOVA) without replication

General information

Two-way analysis of variance (ANOVA) without replication is used to determine if there is a significant difference between means of several subpopulations (groups) dependable on two independent factors (Balakirshnan, Render, & Stair, 2007; Fraser, 2016; Randolph & Myers, 2013). In other words, a two-way ANOVA (also called factorial ANOVA, with two factors) is applied when we have one measurement variable and two nominal variables (usually called 'factors' or 'main effects'). For instance, we could apply ANOVA with two factors without replication when explaining differences of revenue generation in different stores for different seasons where store would be one factor and particular season other factor by which we test differences in revenue generation. Or we can apply ANOVA without replication when we want to test in-field results of promotional activities of various sales representatives and various location where activities are applied (in this case representatives are one factor, locations are second factor and results or effect of promotion is variable which is tested for differences according to those two factors).

Hypothesis

In two-way ANOVA without replication, there is a single observation for each combination of the nominal variables, therefore we have only two null hypotheses: H0(1): There is no difference between means of observations grouped by one fac-

- tor.
- H0(2): There is no difference between means of observations grouped by other factor.

In two-way ANOVA without replication we assume that there is no interaction between factors.

Assumptions

There are the following assumptions associated with the two-way ANOVA without replication (Dean & Ilowsky, 2013; Field, 2013; Winston, 2016):

- dependent variable—should be continuous;
- two independent variables (factors)—should be in two or more categorical, independent groups;
- each sample has to be drawn independently of the other samples (the samples are disjoint);
- the variance of data in the different groups should homogenous;

- each sample should be taken from a normally distributed population;
- there is no dependence or interaction between factors;
- there are no significant outliers.

Example

Dataset: In Retail Company Tradex we collected data on value of expired or spoiled merchandize in EUR per week. Now we want to analyze if this value varies according to sales region and according to product categories.

Data info:

- variable 1: product category—nominal (1—Fruits and vegetables, 2—Diaries, 3—Meat);
- variable 2: sales region nominal (1—East, 2—West, 3—North, 4—South);
- variable 3: value of expired or spoiled merchandize in EUR per week—numeric.

Hypotheses:

- H0(1): There is no difference between means of value of expired or spoiled merchandize grouped by product category.
- H1(1): There is a difference between means of value of expired or spoiled merchandize grouped by product category.
- H0(2): There is no difference between means of value of expired or spoiled merchandize grouped by sales region.
- H1(2): There is a difference between means of value of expired or spoiled merchandize grouped by sales region.

Testing the hypotheses in SPSS

In our example, we observed value of expired or spoiled merchandize in EUR per week in a particular sales region and in particular product categories. In Table 1, dataset is shown.

	1—Region East	2—Region West	3—Region North	4—Region South
1—Fruits and vegetables	130	150	280	140
2—Diaries	250	320	330	230
3—Meat	120	130	250	180

Table 1. Dataset for two-way	ANOVA without replication
------------------------------	---------------------------

Source: The authors' own elaboration.

For analysis in SPSS we will form three variables "Type", "Region" and "FW", then we will proceed to enter data. In the first row (see Figure 1, Row 1) we will enter 1, 1, 130 (meaning: 1—Fruit and vegetables, 1—Region East, 130 EUR of

expired and spoiled merchandize per week). Then we will proceed to enter 1, 2, 150 (see Figure 1, row 2) for Fruits and vegetables in Region West—value 150 EUR.

jile Edit	Yew D	ata Iransi	form	Analyze	Graphs 1	tilities E	gtensions	Window E	jelp	_							0	
	1 🖨 [1 K	1	1	ii =	N.	<i>n</i>		1	• Q								
																	Visible: 3 of 3	Variable
	💰 Type	💰 Reg	ion	🖉 FW	var	Yar	var	var	var	Var	var	var	var	var	var	var	var	13
1	1,	00	1,00	130,00														
2	1.	00	2,00	150,00														
3	1,	00	3,00	280,00														
4	1,	00	4,00	140,00														
5	2.	00	1.00	250.00														
6	2.	00	2,00	320,00														
7	2.	00	3,00	330,00														
8	2,	00	4,00	230,00														
9	3,	00	1,00	120,00														
10	3.	00	2.00	130,00														
11	3.	00	3,00	250,00														
12	3.	00	4,00	180,00														
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		
23																		
	<																	>
Data View	Variable	liaw																
						_	_					IBM SPSS S	Ratistics Pro	cessor is rea	dy 🛃	Unicod	e:ON	
- C) 🖽	2					5) 8								- A - A	10.00	the 1745	

Figure 1. Dataset prepared for two-way ANOVA without replication analysis in SPSS

Source: The authors' own elaboration.

When all data is entered, we will choose type of analysis (see Figure 2). For twoway ANOVA without replication, we will select General Linear Model, Univariate option and then we will specify further options.

🚰 Untitled2 [[DataSet1] - IBM S	PSI Statistics D	ata Editor	- 0 X
Ele Edit	View Data	Iransform	Analyze Graphs Utilities Extens Power Analysis	sians Window Help
			Description Photostics	Visible: 3 of 3 Variable
	S Type	Region	Bayesian Statistics	s var
2	1,00	2.00	Tables	2
3	1,00	3,00	Compare Means	,
4	1,00	4,00	General Linear Model	" 📴 Unvariate
6	2.00	2.00	Mixed Models	Multivariate.
7	2.00	3.00	Correlate	Bepeated Measures
8	2,00	4,00	Remassion	Variance Components
9	3,00	1,00	Loginear	
10	3,00	2,00	Neural Networks	,
11	3,00	3,00	Classify	
13	3,00	1,00	Dimension Reduction	3
14			Scale	3
15			Nonparametric Tests	s
16			Forecasting	A state of the
17			Sunival	s
18			Multiple Response	,
20			Missing Value Analysis	
21			Multiple Imputation	,
22			Complex Samples	• 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
23	<		Simulation	
Data View	Variable View		Quality Control Spatial and Temporal Modeling	· · · · · · · · · · · · · · · · · · ·
Univariate			Direct Margeting	> IBM SPSS Statistics Processor is ready if Unicode ON
م ∎	Π	9 🛤	📅 🎦 🏮 🚖 🗵	и на

Figure 2. Choosing type of analysis—General Linear Model—Univariate

Firstly, we will have to set variables as shown in Figure 3. Our dependent variable is "FW" and our Fixed Factors are "Type" and "Region".

🔒 Untitled2 (DutaSet1] - IBM (SPSS Statistics D	ata Editor															- 6	3 ×
Ele Edit	Yew Data	Iransform	Analyze	Graphs U	tilities [4]	Extensions)	Mindow He		@ [<u> </u>									***
	& Tune	& Region	✓ FW	var	var	uar	var	1	var	var	1		-	Var	var	var	var	Visible: 3 of 3	3 vanable
1	1.00	1.00	130.00										-						
2	1.00	2.00	150.00			ta Univariate							×						
3	1.00	3.00	280.00																
4	1.00	4.00	140.00				_	4	Debe	ndent Vanable	e:	Mo	del						
5	2.00	1.00	250.00					1	1	·W		Cont	asts						
6	2.00	2.00	320,00						Exec	d Factor(s):		- Di	-						
7	2.00	3.00	330.00					4		ype			4						
8	2.00	4.00	230,00						i ø5₽	Region		Post	Hoc						
9	3,00	1,00	120,00						Rand	om Factor(s):		EM N	leans						
10	3,00	2,00	130,00				b					Sa	N9						
11	3,00	3,00	250,00					*				Ort							
12	3,00	4,00	180,00									-							
13									Çova	riate(s):		Boot	strap						
14							- 1	4											
15																			
16									WLS	Weight									
17								\$					- 1						
18									-			-							
19							OK B	aste	Rest	Cancel	melp	<u> </u>	_						
20																			
21																			

Figure 3. Setting options for analysis—variable specification

Source: The authors' own elaboration.

Secondly, we have to customize type of analysis. Therefore, we will have to specify our model. Therefore we select option "Model". We will use option "Build terms" and we will observe "Main effects" (see the central part of the screen at Figure 4). Moreover, by using arrow at the middle of the screen we have to transfer our factors from column "Factors & Covariates" to column "Model" (see right part of Figure 4). Then we will click to "Continue" and by clicking on "OK" at the previous screen, we will perform our analysis.

Ele Edit	Xiew Data	Iransform	Analyze Gr	aphs	Utilities Extensions	Window Help	ا 🗨 🔍						
					ta Univeriete: Model			×				Visible: 3 of 3	Variables
	💑 Туре	뤚 Region	🖋 FW	var	and the second second				var	var	var	var	10
1	1,00	1,00	130,00		Specify Model								_
2	1,00	2,00	160,00		O Full factorial	Build terms	Build custom terms						
3	1,00	3,00	280,00		Eactors & Covariates:		Model:						
4	1,00	4,00	140,00		M Type		Type						
6	2,00	1.00	250,00		H Region		Region						
6	2,00	2,00	320,00										
7	2,00	3,00	330,00			Build Term(s)							
8	2,00	4,00	230,00			Type:							
9	3,00	1,00	120,00			Main effects 🛩							
10	3,00	2,00	130,00			[m]							
11	3,00	3.00	250,00			•							-
12	3,00	4,00	180,00										
13													
14													
16													
16					+		Clear Term Add Bernow						
17					Build Term:								
18													
19					Concernance of the American	Name of Street Str		_					
20					Sum of squares: Type	• ·	Include intercept in model						
21						Continue	Cancel Help						
22						Zentinee							
23	<									-			>

Figure 4. Setting options—Specify Model—customization of analysis

Source: The authors' own elaboration.

133



In Figure 5 results of analysis are shown. For quick interpretation, we will pay attention to the column "Sig." and we will search for value lower than .05 in order to not reject or reject our hypotheses. At the factor Type the significance value is .006 whereas at the factor Region the significance value is .022. Because both significance values are lower than .05, we can reject H0 hypotheses for both factors (Type and Region).

🚰 ^Output7 [Document8] - IBM SPSS Sta	tistics Viewer					
<u>File Edit View Data Transf</u>	orm <u>I</u> nsert F <u>o</u> rn	nat <u>A</u> nalyze (Graphs J	Utilities E <u>x</u> t	ensions <u>W</u> in	dow <u>H</u> elp
			±_ =	🖩 🕞		
					•	
	4,00	3				
- 🗟 Univariate Analysis of ۱						
→ (iii) Title		Tests of Bet	ween-Sul	bjects Effec	ts	
Between-Subjects	Dependent Variab	le: FW				
Tests of Between-		Type III Sum				01-
Title	Source	of squares	ar	Mean Square	F	SIg.
🖨 🕒 Турс	Corrected Model	58008,333°	5	11601,667	9,557	,008
(a) Title	Intercept	525008,333	1	525008,333	432,501	,000,
	Туре	32316,667	2	16158,333	13,311	.006
E Title	Region	25691,667	3	8563,889	7,055	.022
- 🖓 FW	Error	7283,333	6	1213,889		
E Region	Total	590300,000	12		·	
Hultiple (Corrected Total	65291,667	11		· · · · · · · · · · · · · · · · · · ·	
Profile Plots	Post Hoc Tests Type	Ми	Itiple Co	mparisons		
	Dependent Variab	10: 510/				
	Tukey HSD	16. 199				
		Mean Difference //-			95% Confid	ence interval
	(I) Type (J) Type	- J)	Std. Error	Sig.	Lower Bound	Upper Bound
	1,00 2,00	-107,5000	24,63624	.011	-183,0908	-31,9092
< >	3,00	5,0000	24,63624	,978	-70,5908	80,5908
				W I		

Figure 5. Result of two-way ANOVA without replication analysis in SPSS

Source: The authors' own elaboration.

Based on this we conclude that we can conclude that there is the difference between means of value of expired or spoiled merchandize grouped by product category and that there is difference between means of value of expired or spoiled merchandize grouped by sales region.

Testing the hypotheses in Excel

Same dataset is entered to Excel. In Figure 6 collected data is shown in format suitable for analysis in Excel.



2.

Independent samples—more hypotheses testing

Figure 6. Dataset for two-way ANOVA (without replication) analysis in Excel

Source: The authors' own elaboration.

Then we have to select Data tab and we have to click Data Analysis (within Analysis group of commands). Within the list of methods, we choose ANOVA: two factor without replication (see Figure 7).

٩)	8	5.0.					khei	evic ANOVA twowing without replication, Excel. spis - Word		7 30 - 6 ×
043	01	H				anovaluasy		ordata - Excel 7 (8)	- 0 x	Prijaw
	DATO	TEKA POLAZNO	UMETANIE 12G	ED STRANKE	FORMULE PODA	PREGLED	PREC	Z DODACI	Prijau	- Minaterje -
Lije L	Dohra	adanje vanjskih odstaka *	Veze Svojstva Uredivanje vez Veze	刻 <u>武</u> 章 弘 Sortikanje	Filtar Dodate Anter Filtar Dodate Anteriorije i filologije	o primijeni o	An	egiled predidensish undas 👘 🗰 📆 Grupisanje - 1 💆 🐾 Alat za rježavanje dom doglatane 👘 🖏 Kang-sporanje - 1 🖏 Analiza poduralas odena valjanda Grupisatana - n ili Mang-sporanje - 1 kaj Analiza poduralas Alati na podutile Alati na podutile	~	B_BBC_ ABDCCC ADDCCC - Fodnacion Neupadiji Izziknuto u ↓ Dasbie - ↓ Destinge A
	110	•] • [)	× √ Jx						^	
	4	A	8	c	D	t	r		N .	
	2	Value of expired or s	polled merchandu	te in EUR per wee	*			Assina Toola		
	3		Region East	Region West	Region North	Region South		Anovs: Two if some Without Replication		
	4.8	Fruits and vegetables	130	150	280	140		Consideration Cancel		
	5 6	Diaries	250	320	330	230		Descriptive Statistics Forned		
1	7 8 9 10	weat	120	1.0	250	150		Linter Tarri Sanger Rol Vationes Pourier Anglin Hologram Morroy Jureger Bandos Number Canaston		
-	11 12 13									
~	14									
-	16									
	-	DATA1	List? DATAT	orilacodbaspss	DATA2 List	anova2	æ			
1	SPREM	AN T					0		+ 100%	
~					1			- A large the Annual Annua		
						0.000				
						· · · ·	1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
~					1	100 100 100 100 100 100 100 100 100 100	1001 (100 (100)) 100	and the second s		
						1 1	-			
							-			
						5 5	45 25 36	3		
5704	NICA 3	0 CO 16 BROJ RUCO 3	264 QB ENGLESK	(SAD) 10				·	_	III III II;
		, O Upišite ovdje	za pretraživanje	e	H 🤇		i i	C 🖬 📧	(a) ~ 6	🖸 🛒 🛢 🗭 😋 🖦 👼 🐜 di qi 2023

Figure 7. Data Analysis tab in Excel—selection of the method ANOVA: two-factor without replication

Source: The authors' own elaboration.

135



In the dialog box of ANOVA: two-factor without replication we have to configure as follows (see Figure 8):

- input range of the dataset including labels, in our example it is A3:E6;
- check existence of data labels (see Labels);
- output range, we can choose to show data at some position at the active Worksheet, then we have to specify exact cell from which our results are going to be presented (such as F3); but in our case we rather specified New Worksheet as the location of our results, also we can specify a name for our output (in our example "ANOVA");
- the last thing is the level of significance, i.e. alpha value. There we can use default value as it is already set to .05.

	B				anovažinay	toreplicate	undata - Excel		7 GE	a - a	×	P
Jac	račanje konjakih Opjorati podstala * 1 2	METANEE IZGLED	STRANEC IN State Softwarpe So	XIMULE PODA Total Filter ToPonovne Total Toponovne Total Toponovne Total Toponovne Total Toponovne Total Toponovne Total Toponovne Toponovn	5 PROCED	And a second sec	Z DOCACI unjikel predictivels knows line (Expense)e - die diplikate line - m ² / ₂ (Effectivels) opens unjinnets postatika - m ² / ₂ (Effectivels) Mei aspodelle (Expense) Strukter	9 % N 9 % N	let za rješavanje saliza podataka Analiza	PU	AgBbC AsbCcc AsbCcc .	2 Introdye - 2 Zamjena 2 Odabit * Londvanje
Q.	A	в	c	D	E	F	Anova: Two-Factor Without Replication		? X	N		
1	Value of expired or sp	oiled merchandize i	in EUR per week	1			nput	-	OK			
2		Resion East B	inelon West	Berlon North	Reelos South		Hput Renge: 14533130	181	Cancel			
4	Fruits and vegetables	130	150	280	140		- Mulati		Report			
5	Diaries	250	320	330	230	4	Block Cop		Enner			
6	Meat	120	130	250	180	/	Output options	-				
2							O Quiput Range					
9							These monksharet ply: ANOVA analysis					
10							O Hew Berkbeck					
11	2						L		_			
12												
13												
15												
16												
•	DATA1	List7 DATAtori	lacodbasos	DATA2 List	ancua2		1.61					
gree .	IMAN 55			_							-	
-												
				The corr (se Re	en we have mmands). W se Figure 2). pure 2: Data plication	to select Jithin the Analysis	Data tab and we have to click Data Analysis (list of methods, we choose Anova: Two Fact tab in Excel—selection of the method Ano	rithin Anal or Without Two-Fa	ysis group of t Replication ctor <u>Without</u>			

Figure 8. Dialog box ANOVA: two-factor without replication

Source: The authors' own elaboration.

In Figure 9 results of data analysis is shown and we can interpret our results. First of all, we have basic descriptive statistical data grouped by both factors. From this part we can read how many observations we had in which group (see column Number), then we can see what the average value of expired or spoiled merchandize in EUR per week in each group is, and what the variance within each group is. For instance, when we observe product categories, the lowest average of 170 EUR is expired or spoiled in Meat category and if we observe regions, the lowest average value of spoiled or expired merchandize is in region East and it is 166.6667 EUR.

In addition, ANOVA results are shown. In this table, the most important reading is *p*-value because by it we can decide not to reject or to reject the null hypothesis. In



our case the *p*-value for rows .006222 (remember, in our dataset product categories are entered to Excel, see Figure 1). In this case *p*-value is lower than *alpha* value of .05 and it means that we can reject the null hypothesis H0(1) and we can conclude that there is difference between means of value of expired or spoiled merchandize grouped by product category. So, this difference is statistically significant at the level of .05.

In addition, we can observe that in our case the *p*-value for columns (in our case, Regions) is .021524 which is lower than .05. Therefore, we can reject the null hypothesis H0(2), and we can conclude that there is difference between means of value of expired or spoiled merchandize grouped by sales region. So, the test results have shown that this difference is statistically significant at the level of .05.

A B C D E F G H J 1 Anova: Two-Factor Without Replication - <th>Ukloni Pro duplikate Ale</th> <th>fed t ih unosa du</th> <th>Pregl predloženi</th> <th>Tekst u stupce</th> <th>ti who primijeni atho</th> <th>tar Doc arje i filtrirarj</th> <th>rtiranje Fi</th> <th>ž↓ ž↓ So</th> <th>tze rojstva ređivanje w ete</th> <th>Osvježi vre • 🕞 U</th> <th>rograma S Iz Iz drugih Postojeće veze Dotvačanje varjstih podataka</th>	Ukloni Pro duplikate Ale	fed t ih unosa du	Pregl predloženi	Tekst u stupce	ti who primijeni atho	tar Doc arje i filtrirarj	rtiranje Fi	ž↓ ž↓ So	tze rojstva ređivanje w ete	Osvježi vre • 🕞 U	rograma S Iz Iz drugih Postojeće veze Dotvačanje varjstih podataka
Anova: Two-Factor Without Replication B C D t P G H I J Anova: Two-Factor Without Replication -		_									· · · ·
ANOVA Sum Average Variance Fuits and vargetables 4 700 175 9666.667 Diares 4 1130 282.5 2491.667 Region Rest 4 660 170 553.333 Region Rest 3 500 166.6667 523.333 Region Rest 3 500 166.6667 523.333 Region South 3 500 166.6667 523.333 Region South 3 550 183.3332 203.333 ANOVA 3 550 183.3332 1033.33 4 ANOVA 2 1055.33 153.1121 0.00222 5.14223 7 Columns 2591.67 11 7 Columns 55291.67 11 7 Tetal 65291.67 <t< th=""><th>K</th><th>,</th><th>1</th><th>н</th><th>G</th><th>P</th><th>Ł</th><th>D</th><th>C</th><th>в</th><th>A</th></t<>	K	,	1	н	G	P	Ł	D	C	в	A
SUMMARY Court Sum Average Variance 4 Futura and vegetables 4 700 175 9466.667 5 Diartes 4 1130 9466.667 - 6 Meat 130 252.5 2491.667 - 7											Anova: Two-Factor Without Replication
Joint Source of Variation							Mariana		Com.	Count	CILLARADY
Anton organizations 4 100 12.5 seess and mathematical sets and mathmatical sets and mathematical sets and mathmatemathmatical set							Variance	Average	300	Count	SUMMARY
Source of Variation SS df MS F Produc F crit 10 ANCVA 550 183.333 203.333 1 <							2491.667	282.5	1120	4	Distor
New Color Display 8 Region Mexit 3 500 166.6667 523.333 9 Region Wexit 3 660 200 10000 10 Region North 3 860 286.6667 163.333 12 Region South 3 550 188.3333 203.333 14 MOVA Source of Variation SS df MS F P-value F crit 15 Source of Variation SS df MS F P-value F crit 16 Rows 32316.67 2 10158.38 13.31121 0.006222 5.143253 17 Cohernes 25901.67 11 10 Tetal 65291.67 11 21 Cohernes Cohernes Cohernes Cohernes 22 Cohernes Cohernes Cohernes Cohern							3533 333	170	690	4	Ment
8 Region East 3 500 166.6667 523.333 9 Region West 3 600 200 10900 0 Region North 3 860 266.6667 163.333 10 Region South 3 550 183.333 203.333 11 Region South 3 550 183.333 203.333 12 14 NOVA 5 6 Purolue F crit 15 Source of Variation S5 df M5.33 13.31121 0.00222 16 Rows 32316.67 2 16138.33 13.31121 0.00222 16 Rows 32216.67 3 858.889 7.05492 0.021524 4.757063 16 Foro 7283.333 6 1213.889 1 1 1 17 Calar 65291.67 31 1 1 1 18 Eror 7283.133 1 1 1 1 19 1 1 1 1 1 1 10 1 1 1 1 1 1 12 1 1 1 1 1 1							3333.333	110	000	-	and of
9 Region West 3 600 200 10900 10 Region North 3 860 286.6667 1633.333 10 Region North 3 550 183.333 203.333 11 Region North 3 550 183.333 203.333 11 ANCVA							5233.333	166,6667	500	3	Region Fast
Nome 3 860 286.6667 1633.3333 1633.333.333 1633.3							10900	200	600	3	Region West
Region South 3 550 183.3333 2033.333 12 ANCVA Annovember 2014 Annovember 2014 Annovember 2014 4 ANCVA Source of Variation SS df MS F P-value F-crit 5 Source of Variation SS df MS F P-value F-crit 7 Colomers 2591.67 2 1055.33 13.31121 0.00222 5.143253 8 Fror 7283.333 6 1213.889 0.021524 4.757063 9 Total 65291.67 11							1633.333	286,6667	860	3	Region North
12 ANCVA Source of Variation SS df MS F P-value F.crt 15 Source of Variation SS df MS F P-value F.crt 16 Rows 32316.67 2 10458.33 13.31121 0.006222 5.143233 17 Cohemas 25691.67 3 8563.889 7.05492 0.021524 4.757063 18 Error 7283.333 6 121.3889 - - - 20 Total 65291.67 11 - - - 21 - - - - - - 22 - - - - - - 23 - - - - - - 24 - - - - - - 25 - - - - - - 26 - - - - - - 28 - - - - - -							2033.333	183,3333	550	3	Region South
13 ANCVA 6 7 P-value F crit 5 Source of Variation SS df MS F P-value F crit 6 Rows 32516.67 2 16158.38 13.31121 0.006222 5.143253 7 Columns 25616.67 3 8658.89 7.05462 0.025124 4.757063 8 Fror 7283.333 6 1213.889 - - - 9							20001000	20010000		-	The point of the second s
ANOVA Source of Variation So df MS F P-value F crt 5 Source of Variation 3216.67 2 16158.33 13.1121 0.00222 5.143253 7 Columns 25691.67 3 8563.889 7.05492 0.021524 4.757063 8 Error 7283.133 6 1213.889 9 Total 65291.67 1 22 24 25											
Source of Variation SS df M5 F P-value F-rit 6 Rows 32316.67 2 16138.33 13.31121 0.006222 5.143253 7 Columns 25691.67 3 858.889 7.05462 0.021524 4.757063 8 Eror 7283.333 6 1213.889 - - - 9											ANOVA
16 Rows 32316.67 2 16158.33 13.3122 0.006222 5.141253 17 Colorms 25691.67 3 8563.889 7.05492 0.021524 4.757063 18 Error 7283.33 6 1213.889 7.05492 0.021524 4.757063 19 7 7 7 1 <td></td> <td></td> <td></td> <td></td> <td>Fcrit</td> <td>P-value</td> <td>F</td> <td>MS</td> <td>df</td> <td>\$\$</td> <td>Source of Variation</td>					Fcrit	P-value	F	MS	df	\$\$	Source of Variation
7 Colomons 25691.67 3 8563.889 7.05492 0.021524 4.757063 8 Error 7283.333 6 1213.889 7.05492 0.021524 4.757063 0 Total 65291.67 1 20 Total 65291.67 1 21 <					5.143253	0.006222	13.31121	16158.33	2	32316.67	Rows
Bitror 7283.333 6 1213.889 10 total 65291.67 11 11 11 11 11 12 11 11 11 12 11 11 11 12 11 11 11 12 11 11 11 12 11 11 11 12 11 11 11 13 11 11 11 14 11 11 11 15 11 11 11 16 11 11 11 17 11 11 11 18 11 11 11					4.757063	0.021524	7.05492	8563.889	3	25691.67	Columns
19 tetal 65291.67 11 20 tetal 65291.67 11 21 1 11 11 22 1 11 11 23 1 11 11 24 1 11 11 25 1 11 11 26 1 11 11 27 1 11 11 28 1 11 11								1213.889	6	7283.333	Error
20 Tetal 65291.67 11 22 23 24 23 24 25 26 27 28 28 28											
21									11	65291.67	Total
22											Sec. 1
23		_									
24											
25 26 27 28											
26 27 28											
27 28 28 29 29 29 20 20 20 20 20 20 20 20 20 20 20 20 20											
28											
ANOVA analysis DATA1 List7 DATA1prilagodbaspss DATA2 List2 anova2 (+)				+	anova2	List2	DATA	godbaspss	ATA1prila	List7 D	ANOVA analysis DATA1

Figure 9. Two-way ANOVA without replication results

Source: The authors' own elaboration.

If ANOVA shows us that there is a statistically significant difference between observed groups, we have to do post hoc analysis by comparing pair of groups in order to explain which group differs in comparison to other group. For this purpose in Excel we can perform several *t*-tests. Procedure of *t*-tests is already explained in details in chapter about one-way ANOVA.

Summary of the example

Dataset: The value of expired or spoiled merchandize in EUR per week in different sales regions is observed. In addition, the value of expired or spoiled merchandize in EUR per week is observed according to the product category as well. For the purpose of the analysis four sales regions and three product categories are defined. For each combination of a sales region and a product category the value of expired or spoiled merchandize in EUR per week is collected. On that way, 12 data values of expired or spoiled merchandize in EUR per week were on disposal for our analysis needs.

Data info:

- variable 1: product category—nominal (1—Fruits and vegetables, 2—Diaries, 3—Meat);
- variable 2: sales region—nominal (1—East, 2—West, 3—North, 4—South);
- variable 3: value of expired or spoiled merchandize in EUR per week—numeric.

The two-way ANOVA approach was used to inspect whether the average value of expired or spoiled merchandize grouped by product category can be considered the same across all three product categories. Also, in the same time, the two-way ANOVA approach was used to inspect whether the average value of expired or spoiled merchandize grouped by sales categories can be considered the same across all four sales categories. The results of the two-way ANOVA have shown that there was a statistically significant difference between product categories groups (F(2,6) = 13.311, p = .0062). On the other side, the results of the two-way ANOVA have slown that there was a statistically significant difference between sales region groups (F(3,6) = 7.055, p = .0215).

More info about two-way ANOVA without replication

The two-way ANOVA without replication can be observed as an extension of the one-way ANOVA. Whereas at the one-way analysis just one factor is observed, here at two-way ANOVA without replication two factors are inspected in the same time. Despite the fact that the two-way ANOVA without replication analysis is more complex than the one-way ANOVA they are sharing the same assumptions with additional assumption that there is no dependence or interaction between factors (Barrow, 2017; Randolph & Myers, 2013).

In the analysis, at both observed factors statistically significant differences are found. However, we do not know whether all means between all three product categories or all means between all four sales regions are different or the difference is statistically significant just between some groups. In order to find out that, in the following step in the analysis Tukey post hoc tests can be conducted to observe differences between pairs of categories at given factors. The Tukey post hoc tests procedure and interpretations are analogous to those explained at the one-way ANOVA.

References

Balakirshnan, N., Render, B., & Stair, R. M. (2007). *Managerial decision modeling with spreadsheets*. Pearson Prentice Hall.

Barrow, M. (2017). Statistics for economics, accounting and business studies. Pearson.

Dean, S., & Illowsky, B. (2013). Introductory statistics. OpenStax College.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (5th ed.). Sage edge.

Fraser, C. (2016). Business statistics for competitive advantage with Excel 2016. Springer.

Randolph, K. A., & Myers, L. L. (2013). Basic statistics in multivariate analysis. Oxford University Press.

Winston, W. (2016). Microsoft Excel 2016: Data analysis and business modelling. Microsoft Press.

2.2. Two-way analysis of variance (ANOVA) with replication

General information

The two-way analysis of variance (ANOVA) with replication simultaneously tests the effects of varying two variables (such as gender and age or wealth and geographic area, and their interaction) for a sample which consists of more than one respondent per a certain combination of variables. While in two-factor ANOVA without replication there was only one sample item (observation) for each combination of factors.

Replication refers to the number of cases observed within the same combination of factors (Field, 2013; Fraser, 2016; Winston, 2016). Usually, we use this method in a case of a balanced research design when the size of each subgroup according to two factor is equal. Because then, we can calculate the mean square for each of the two factors, for their interaction, and for each combination of factors.

Hypotheses

In two-way ANOVA with replication, there are more than one observation for each combination of the nominal variables, therefore it is possible to examine interaction between factors as well. So, we will have three null-hypotheses:

- H0(1): There is no difference between means of observations grouped by one factor.
- H0(2): There is no difference between means of observations grouped by other factor.
- H0(3): There is no interaction between factors.

The alternative hypothesis to every above stated is its negation.

The interaction test shows us if the effects of one factor depend on the other factor (Balakirshnan et al., 2007; Dean & Illowsky, 2013).

Assumptions

There are the following assumptions associated with the two-way ANOVA with replication (Barrow, 2017; Field, 2013; Randolph & Myers, 2016):

- dependent variable is continuous;
- two independent variables should consist of at least two or more categorical, independent groups;
- dependent variable should be approximately normally distributed for each combination of independent variables;
- independence of observations, i.e. no relationship between the observations in each group or between the groups;
- no significant outliers because they can have a negative effect on the two-way ANOVA results;
- homogeneity of variances for each combination of the groups of the two independent variables.

Example

Situation: Recently we finished our series of seminars in area of Sustainable consumption and Social responsibility in our local community center. So, we want to test effects of this education in real-life environment. We had 3 educations on various topics which included topics on Eco-friendly products and Fair trade. Therefore, together with the owner of the local supermarket we are collecting data on consumer will to purchase sustainable products in everyday life. We want to test is our education effective or not.

Dataset: observed number of products in a shopping cart at the checkout in a local supermarket according to product labels, number of educations taken (no education, one education, two educations or three educations). For each out of 4 levels of education we collect same number of observations (5), i.e. we have same number of respondents (5). Sample consisted of 20 respondents. For each respondent we count number of products in a shopping cart according to 3 types of product labels.

Data info:

- variable 1: number of educations taken—nominal (1—No education, 2—One education, 3—Two educations, 4—Three educations);
- variable 2: label of product—nominal (1—Eco-friendly, 2—Fair trade, 3—No label);
- variable 3: number of items (products) in the shopping cart—numerical.

Hypotheses:

- H0(1): There is no difference between means of number of items in shopping cart grouped by number of educations taken.
- H1(1): There is a difference between means of number of items in shopping cart grouped by number of educations taken.
- H0(2): There is no difference between means of number of items in shopping cart grouped by label of products.

- H1(2): There is a difference between means of means of number of items in shopping cart grouped by label of product.
- H0(3): There is no interaction between factors.
- H1(3): There is interaction between factors.

Testing the hypotheses in SPSS

In our example we observed how many items (product) labeled with labels in field of sustainable business some customer had and we observed how many educations in our local community this person attended in the field of sustainability and social responsibility, then we entered data into SPSS (see Figure 10). For instance, in the first row we have recorded data for respondent who did not attended seminars (see column "Education" where 1 means No education), and this person had only 2 items (see "Numberofitems" column where we recorded quantity 2) labeled as Eco-friendly products (see column "Label" where 1 meaning Eco-friendly is entered). For the same respondent we recorded 1, 2, 3—see row 6 at Figure 10 (as he/she did not attend education (1), and had 3 items labeled as 2—Fair trade) and we concluded entering data on contents of his/her shopping cart by entering 1, 3, 7—see row 11 at Figure 10 (as he/she did not attend education (1), and had 7 items without any label in field of sustainability, 3 is entered for No label in column "Label"). Then we entered data for all other respondents and items in their shopping carts.

As there were five respondents from each group according to 4 education levels, and for each respondent we counted number of items in their shopping carts for 3 labels, our dataset in SPSS at the end had 60 rows.

🖬 🖽	🔒 🗔		× ₩±		i A			149		Q									
																		Visible: 3 of	3 Variable
	Education	💰 Label	Numberofi tems	var	var	var	19	r	var	var	v	ar	var	var	var	var	var	var	va
1	1,00	1,00	2,00		🛛 🙆 U	nivariate							×						1
2	1,00	1,00	4,00																
3	1,00	1,00	8,00		_			-	Depen	Sent Variabi	e:	Mo	del						
4	1,00	1,00	9,00						1 PR	mberohtem	5	Contr	asts						
5	1,00	1,00	5,00						Exced	Factor(s):		0							
6	1,00	2,00	3,00					4	- € E	ducation			49						
7	1,00	2,00	5,00						a 🕹 🕹	bel		Post	Hoc						
8	1,00	2,00	6,00						Dando	n Easteria)		EMN	leans						
9	1,00	2,00	8,00						1 cgrioo	arr accorday		Sa	-						
10	1,00	2,00	3,00					*											
11	1,00	3,00	7,00									<u>Obe</u>	ons						
12	1,00	3,00	10,00						Covaria	rte(s):		Hoot	strap						
13	1,00	3,00	12,00																
14	1,00	3,00	15,00					-											
15	1,00	3,00	9,00						140.01	Unioht .									
16	2.00	1,00	6.00					-	Hrs .	veignt.									
17	2,00	1,00	7,00		_						_								
18	2,00	1,00	9,00				OK	Easte	Beset	Cancel	Help								
19	2,00	1,00	8,00		-		-	-	_	-	-	-							
20	2,00	1,00	4,00																
21	2,00	2,00	5,00																
22	2,00	2,00	7,00																

Figure 10. Excerpt of the dataset suitable for two-way ANOVA with replication



When data is entered, we proceed to Method selection (see Figure 11). For two-way ANOVA with replication as a method we will choose Analyze—General Linear Model and select Univariate.

Sile Edit	View D	ata Iransform	Analyze Graphs Utilities Extensi	ons Window E	jelp	1.88								
			Poger Analysis 3		Co 🗖	Q								
			Reports										Maible: 3 of 3	Mariable
	B. Counti	a R. Label	Descriptive Statistics										answer o or o	
	. Concan	a a cave	 Bayesian Statistics 	• var	var	var	var	var	var	var	var	var	var	Vit
1	1,	1,00	Tables											
2	1,	1,00	Compare Means 3											
3	1,	1,00	General Linear Model	Univariate										
4	1,	1.00	Generalized Linear Models	- Multivariate										
5	1.	1.00	Mixed Models 3											
6	1,	2,00	Correlate	Fill Rebeated M	neasures									
1	1.	2.00	Regression	Ed Variance Co	omponents									
8	1,	2,00	Lasterer											
9	1.	2.00	Cogenear											
10	1.	2,00	Neural Networks 3											
11	1,	3,00	Classify											
12	1	3,00	Dimension Reduction 3	·										
13		3,00	Scale											
14	1.	3.00	Nonparametric Tests											
15		3.00	Forecasting											
10	2,	1,00	Survival											
10		1.00	Multiple Descourse											
19	2	1.00	The same links frankrain											
20	2	1 00	De meening Zante manyere											
21	2	2.00	Multiple Imputation											
22	2	2,00	Complex Samples											
	<[]]		W Simulation											>
Data View	Variable V	law	Quality Control											
	-		Spatial and Temporal Modeling						and the second					
nvanate			Direct Margeting 3					IBM SPSS SI	atistics Proc	essor is read	y m	Unic	ode.ON	

Figure 11. Choosing method—General Linear Model—Univariate

Source: The authors' own elaboration.

After that, we will define variables as shown at Figure 12. Dependent variable is "Numberofitems", while fixed factors are "Education" and "Label". Now we can specify several options which will enable us to interpret results of the analysis. In this chapter we will set options for Plot diagram and Post hoc analysis.



Figure 12. Setting basics options for the analysis—variables

Source: The authors' own elaboration.

142



Plot diagram settings are shown at Figure 13. Plot diagram will enable us to visualize results of ANOVA in quick and intuitive manner. It will show us interrelation between factors in a graphic mode. We will set Education to be shown at horizontal axis and Label as a separate lines, then we will click Add and we will just check if in the text box under word "Plots" "Education*Label" is shown and if the checkbox is selected for "Line Chart". If everything is correct, we can confirm everything by clicking the button "Continue".

			¥ 🔚		1	n	Univariate: Profile	Plots		×						
							Eactors:		Horizontal Axis:						Visible: 3 of 3	3 Variables
	Education	🗞 Label	Numberofi tems	var	var		Education		Education	var	var	var	var	var	var	VB
1	1,00	1,00	2,00			ta ur			Separate Lines:	×						<u>^</u>
2	1,00	1,00	4,00					*	Label							
3	1,00	1,00	8,00							8						
4	1,00	1,00	9,00						Segarate Plots:	sts						
5	1,00	1,00	5,00													
6	1,00	2,00	3,00				_			A						
7	1,00	2,00	5,00				Plots:	Add Ch	ange Bernove	oc						
8	1,00	2.00	6,00							ans						
9	1,00	2,00	8,00													
10	1,00	2,00	3,00													
11	1,00	3,00	7,00				Chart Type:			15						
12	1,00	3,00	10,00				@ Line Chart			rap						
13	1,00	3,00	12,00				Bar Chart									
14	1,00	3,00	15,00				C. C. C. M.			- 1						
15	1,00	3,00	9,00				Error Bars			- 1						
16	2,00	1,00	6,00				Include Error b	ars								
17	2,00	1,00	7,00				Confidence I	nterval (95.0%								
18	2,00	1,00	9,00				O Standard En	or Multip	lier 2							
19	2,00	1,00	8,00					and an order		_						
20	2,00	1,00	4,00				Include reference	line for grant	d mean							
21	2,00	2,00	5,00				Y axis starts at)								
22	2,00	2,00	7,00				Cart		(Hele							
	<[]						Zones	Cance	rieip							>
Onto Minut	Variable View							-								



Figure 13. Settings for Plot diagram



Finally, we will define options for the post hoc analysis. We will perform post hoc tests for Education and Label and we will mark Tukey post hoc analysis (see Figure 14). Again, we will click button "Continue" and then at previous screen confirm analysis by clicking to "OK" button.

Ele Ed	Yew .	Data	Iransform	Analyze Graphs	Utilities Extensions	Window Help								
													Visible: 3 of 3	3 Variable:
	💰 Educi	ation 🤞	b Label 🥖	Numberofi		Port Mor Multiple Con	nations for Observed Means	×	var	var	var	var	var	va
1		1.00	1.00	2.00										
2		1.00	1.00	4.00	Eactor(s):		Post Hoc Tests for:							
3		1.00	1.00	8.00	Education		Education							
4		1.00	1.00	9.00	Label		Label							
5		1.00	1.00	5.00			*							
6		1.00	2.00	3.00										
7		1.00	2.00	5.00										
8		1.00	2.00	6.00	Equal Varia	inces Assumed								
9		1.00	2.00	8.00	LSD	S-N-K	Waller-Duncan							
10		1.00	2.00	3.00	Bonfern	ani 🗹 Iukey	Type UType II Error Ratio: 100							
11		1.00	3.00	7.00	Sidak	Tukey's-b	Dunngtt							
12		1.00	3.00	10.00	Scheffe	Duncan	Control Category:							
13		1.00	3.00	12.00	DREGI	N.E. Hochberris C	T2 Tast	_						
14		1.00	3.00	15.00	DREAM		@2.sided @ < Control @ > Co	Inter						
15		1.00	3.00	9.00	LINE W	1.24 C Macriel	• Farrier • + officer • + of							
16		2.00	1.00	6.00	Equal Varia	mose Net Assumed								
17		2.00	1.00	7.00	Tamba	ele T2 Descettie 1	3 Gamas Housel Depositie C							
18		2.00	1.00	9.00		estra 🖬 contretts i	2 Contestionen Coginitas c							
19		2.00	1.00	8.00		Continue	Cancel Help							
20		2.00	1.00	4.00										
21		2.00	2.00	5.00										
22		2.00	2.00	7.00										- ·
	<													>
Data Via	- Variable	View					0.00							
Data vie														
							IBM	SPSS St	latistics Pro	cessor is rea	dy 🛃	Unic	ode:ON	

Figure 14. Setting post hoc analysis options

Source: The authors' own elaboration.

When analysis is finished, we will get results separated in several segments. Firstly, we will have two-way ANOVA general results (see Figure 15). On the basis of this table we will be able not to reject or reject our starting null hypotheses.

Dependent Variable	Numberofitems				
Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	611.783 ^a	11	55.617	9.715	.000
Intercept	4950.417	1	4950.417	864.702	.000
Education	144.850	3	48.283	8.434	.000
Label	39.033	2	19.517	3.409	.041
Education * Label	427.900	6	71.317	12.457	.000
Error	274.800	48	5.725		
Total	5837.000	60			
Corrected Total	886.583	59			

Tests of Between-Subjects Effects

a. R Squared = .690 (Adjusted R Squared = .619)

Figure 15. Two-way ANOVA with replication results

Source: The authors' own elaboration.

144

We will pay attention to column "Sig." where *p*-values are shown. In this column we will observe which values are less than .05. In our case all significance values are lower than .05 (Education p < .001, Label p = .041, Education * Label p < .001). Therefore, we can reject all three H0 hypotheses. Therefore, at significance level .05 we can conclude that:

- there is a difference between means of number of items in shopping cart grouped by number of educations taken;
- there is a difference between means of means of number of items in shopping cart grouped by label of product;
- there is interaction between factors.

In Figure 16 the plot diagram is shown. We can observe that Education and Label curve intersect which means that those variables are in interaction and that we have to take this fact into account when interpret our post hoc data.



Figure 16. Plot diagram

Source: The authors' own elaboration.

As the two-way ANOVA results have shown that there are differences between groups (factors), it is recommended to do post hoc analysis to get better insight into the results. Results of the post hoc analysis help us to observe between which groups are strongest differences and are there some groups which do not differ from each other.

In Figure 17 Post hoc analysis results are shown based on Label. We can observe that between those who did not attend any education (1) and those who attended

two educations (3), then between those who did not attend any education (1) and those who attended three educations (4), and between those who attended one education (2) and those who attended three educations (4) at level of .05, there is statistically significant difference between means of number of items in shopping cart. While other pairs of groups did not show statistically significant difference (p-values in column "Sig." are higher than .05).

Dependent Va Tukey HSD	riable: Number	ofitems				
(I) Education	(J) Education	Mean Difference (I- J)	Std. Error	Sig.	95% Confid Lower Bound	ence Interval Upper Bound
1.00	2.00	-1.4000	.87369	.387	-3.7252	.9252
	3.00	-2.4000	.87369	.041	-4.7252	0748
	4.00	-4.2667	.87369	.000	-6.5919	-1.9415
2.00	1.00	1.4000	.87369	.387	9252	3.7252
	3.00	-1.0000	.87369	.664	-3.3252	1.3252
	4.00	-2.8667	.87369	.010	-5.1919	5415
3.00	1.00	2.4000	.87369	.041	.0748	4.7252
	2.00	1.0000	.87369	.664	-1.3252	3.3252
	4.00	-1.8667	.87369	.156	-4.1919	.4585
4.00	1.00	4.2667	.87369	.000	1.9415	6.5919
	2.00	2.8667	.87369	.010	.5415	5.1919
	3.00	1.8667	.87369	.156	4585	4.1919

Figure 17. Post hoc analysis results-Education

Source: The authors' own elaboration.

Multiple Comparisons

Tukey HS	D					
(I) Label	(J) Label	Mean Difference (I- J)	Std. Error	Sig.	95% Confid Lower Bound	ence Interval Upper Bound
1.00	2.00	-1.2500	.75664	.234	-3.0799	.5799
	3.00	.7000	.75664	.627	1.1299	2.5299
2.00	1.00	1.2500	.75664	.234	5799	3.0799
	3.00	1.9500	.75664	.034	.1201	3.7799
3.00	1.00	7000	.75664	.627	-2.5299	1.1299
	2.00	-1.9500	.75664	.034	-3.7799	1201

Based on observed means.

Dependent Variable: Numberofitems

The error term is Mean Square(Error) = 5.725.

*. The mean difference is significant at the 0.05 level.

Figure 18. Post hoc analysis results-Label

In Figure 18 Post hoc analysis results are shown based on Education. We can observe that between products with label Fair trade (2) and products with no label (3) at level of .05, there is a statistically significant difference between means of number of items in shopping cart. While other pairs of groups did not show statistically significant difference (*p*-values in column "Sig." are higher than .05).

Testing the hypotheses in Excel

To perform two-way ANOVA with replication in Excel, it is extremely important to have the same number of observations for one factor, in our case—education level. Therefore, we made and entered 5 observations for each level of education (see rows in Figure 19).



Figure 19. Dataset for two-way ANOVA (with replication) analysis in Excel

Source: The authors' own elaboration.

Data is entered into Excel in format suitable for data analysis grouped by number of educations as we have same number of respondents for each level of education, then in columns we enter another grouping variable, i.e. product labels (see Figure 19). In each row we enter data on one survey participant. For instance, we enter data for participants which did not attend any seminar (no education) in five rows, each row for one participant. First participant without education put 2 eco-friendly items, 3 fair-trade items and 7 no label items to his/her shopping cart, while second participant without education put 4 eco-friendly items, 5 fair trade items and 10 items without labels into his/her shopping cart.

To perform two-way ANOVA with replication, we have to click Data tab and we have to choose Data Analysis tab (within Analysis group of commands). Within the list of methods, we choose ANOVA: two factor with replication (see Figure 20).



ata Analysis		? ×
Analysis Tools		OK
Anova: Single Factor	~	UK
Anova: Two-Factor With Replication		Cancel
Anova: Two-Factor Without Replication		
Correlation		Demet
Covariance		Pomoc
Descriptive Statistics		
Exponential Smoothing		
F-Test Two-Sample for Variances		
Fourier Analysis		
Histogram	\sim	

Figure 20. Data Analysis tab in Excel—selection of the method ANOVA: two-factor with replication

Source: The authors' own elaboration.

In the dialog box of ANOVA: two-factor with replication we have to configure as follows (see Figure 21):

- input range of the dataset including labels, in our example it is B3:E23;
- input number of rows per sample (i.e. number of observations), in our case 5;
- output range, we can choose to show data at some position at the active Worksheet, then we have to specify exact cell from which our results are going to be presented (such as F3); but in our case we rather specified New Worksheet as the location of our results, also we can specify a name for our output (in our example ANOVA);
- the last thing is the level of significance, i.e. alpha value. There we can use default value as it is already set to .05.

nova: Two-Factor With Replic	cation	? >
Input		OK
Input Range:	\$B\$3:\$E\$23	
Rows per sample:	5	Cancel
<u>A</u> lpha:	0.05	<u>P</u> omoć
Output options		
O Output Range:		1
New Worksheet <u>P</u> ly:	ANOVA twoway with	

Figure 21. Dialog box ANOVA: two-factor with replication

	А	В	С	D	E	F
1	Anova: Two-Factor	With Replic	ation			
2						
3	SUMMARY	Eco-friend	Fair trade	No label	Total	
4	No education					
5	Count	5	5	5	15	
6	Sum	28	25	53	106	
7	Average	5.6	5	10.6	7.066667	
8	Variance	8.3	4.5	9.3	13.06667	
9						
10	One education					
11	Count	5	5	5	15	
12	Sum	34	42	51	127	
13	Average	6.8	8.4	10.2	8.466667	
14	Variance	3.7	5.8	5.7	6.409524	
15						
16	Two educations					
17	Count	5	5	5	15	
18	Sum	46	66	30	142	
19	Average	9.2	13.2	6	9.466667	
20	Variance	3.7	5.7	2.5	12.69524	
21						
22	Three educations					
23	Count	5	5	5	15	
24	Sum	70	70	30	170	
25	Average	14	14	6	11.33333	
26	Variance	14.5	2.5	2.5	20.80952	
27						
28	Total					

Independent samples—more hypotheses testing

14	A	В	С	D	E	F	G
34							
35	ANOVA						
36	Source of Variation	SS	df	MS	F	P-value	F crit
37	Sample	144.85	3	48.28333	8.43377	0.000132	2.798061
38	Columns	39.03333	2	19.51667	3.409025	0.041269	3.190727
39	Interaction	427.9	6	71.31667	12.45706	2.08E-08	2.294601
40	Within	274.8	48	5.725			
41							
42	Total	886.5833	59				
43				S		5	

Figure 22. Two-way ANOVA with replication results

Source: The authors' own elaboration.

In Figure 22 there are results of two-way ANOVA with replication analysis. First of all, in Excel we have basic descriptive statistical data grouped by factor in rows (number of educations) systemized by factor in columns (in our case product labels). From this part we can read how many observations we had in which combination of factors (see rows Count in each sub table), then we can see what the average number of items for each combination of factors, together with data

on variance. For instance, when we observe sub table "No education", we will see that there are on average 5.5 items labeled as Eco-friendly in the shopping cart, 5 items labeled as Free trade, and 10.6 items with no label. On the other hand, in sub table "Three educations" there are on average 14 items labeled as Eco-friendly, 14 labeled as Fair-trade and 6 with no label.

In addition, ANOVA results are shown. In this table, the most important reading is p-value because by it we can decide not to reject or to reject the null hypothesis. In our case the p-value for Sample (in our case this data refers to number of education) is .000132 and it is less than significance level of .05 which means that we can reject the null hypothesis H0(1), and we can conclude that there is a difference between means of number of items in shopping cart grouped by number of educations taken. So, this difference is statistically significant at the level of .05.

In addition, we can observe that in our case the *p*-value for columns (in this case, product labels) is .041269 which is lower than .05. Therefore, we can reject the null hypothesis H0(2), and we can conclude that there is difference between means of means of number of items in shopping cart grouped by label of product. In other words, this difference is statistically significant at the level of .05.

Moreover, the *p*-value for testing interaction between our two factors (number of education and label of products) is 2.08E-08 which is lower than .05 and means that we can reject H0(3) and consequently conclude that there is an interaction between factors and we have to take it into account when we interpret our data. Because if an interaction effect is present, the impact of one factor depends on the level of the other factor.

Summary of the example

Dataset: The number of products in a shopping cart at the checkout in a local supermarket according to product labels is observed. Three product labels have been defined. Overall, 20 respondents have been selected to participate in the study according to their number of taken educations. Four levels of educations taken are recognized. In the study participated equal number of respondents according to the number of educations taken. Consequently, at each level of educations taken we had 5 respondents for which we measured the number of products with different product labels. On that way, three measurement for each respondent have been conducted.

Data info:

- variable 1: number of educations taken—nominal (1—No education, 2—One education, 3—Two educations, 4—Three educations);
- variable 2: label of product—nominal (1—Eco-friendly, 2—Fair trade, 3—No label);
- variable 3: number of items (products) in the shopping cart—numerical.

The two-way ANOVA with replication approach was used to inspect whether our education is effective in changing consumer habits to purchase sustainable products in everyday life. The results have shown that there was a statistically significant interaction between the number of educations taken and label of product (F (6, 48) = 12.457, p < .001).

More info about two-way ANOVA with replication

All previously mentioned additional information about two-way ANOVA without replication apply to two-way ANOVA with replication.

The main difference between two-way ANOVA without replication and two-way ANOVA with replication is the sample structure. In the ANOVA without replication we have only a single observation for each combination of nominal variables while in two-way ANOVA with replication we have more than one observation. In other words, in two-way ANOVA without replication design there is only 1 experimental unit for each combination of the factors. While in two-way ANOVA with replication "there are more than one experimental unit per combination of the factors. In such design we have enough degrees of freedom and the interaction between factors can be estimated" (Field, 2013; Fraser, 2016). It is recommended that in two-way ANOVA with replication we should use balanced data or sample with uniform size to get results in efficient manner".

References

Balakirshnan, N., Render, B., & Stair, R. M. (2007). *Managerial decision modeling with spreadsheets*. Pearson Prentice Hall.

Barrow, M. (2017). Statistics for economics, accounting and business studies. Pearson

Dean, S., & Illowsky, B. (2013). Introductory statistics. OpenStax College.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (5th ed.). Sage edge.

Fraser, C. (2016). Business statistics for competitive advantage with Excel 2016. Springer.

Randolph, K. A., & Myers, L. L. (2013). Basic statistics in multivariate analysis. Oxford University Press.

Winston, W. L. (2016). Microsoft Excel 2016: Data analysis and business modelling. Microsoft Press.