Two-Way ANOVA without replication

Elias Zintzaras, M.Sc., Ph.D.

Professor in Biomathematics-Biometry
Department of Biomathematics
School of Medicine
University of Thessaly

Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA

Theodoros Mprotsis, MSc, PhD
Teacher & Research Fellow
(http://biomath.med.uth.gr)
University of Thessaly
Email: tmprotsis@uth.gr

When to use it

- When there are **two known factors** affecting the variability (or variation) in the data
- Then, the effect of each factor is analyzed using a **two-way analysis of variance (two-way ANOVA)**
- This analysis is typically performed using statistical software such as SPSS or Excel

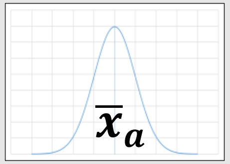
Example

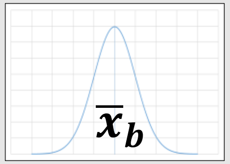
In an experiment to compare the effect of three **drugs** on lymphocyte counts in mice, a design with three mice from four different **cages** was used

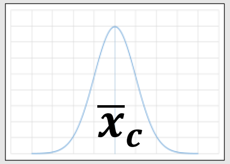
Factors of variation

There are two possible factors of variation: the cage and the drug

	Cages (columns/treatments/groups)				
Drugs	1	2	3	4	**Mean**
a	7.1	6.1	6.9	5.6	$\bar{x}_a = 6.43$
b	6.7	5.0	5.9	5.1	$\bar{x}_b = 5.68$
c	6.6	5.4	5.8	5.2	$\bar{x}_c = 5.75$







Overall mean:

The mean of all 12 lymphocyte counts together

$$\bar{\bar{x}} = 5.95$$

A quick look at the columns tells us what?

Quick note: two-way ANOVA with replication

	ph		
sa	6.0	5.5	5.0
0	8.2 8.4 8.3	5.9 6.0 6.1	4.3 4.3 4.2
100	7.6 7.8 7.6	5.0 5.3 5.8	4.1 4.4 4.2

multiple measurements per cell

	Cages				
Drugs	1	2	3	4	**Mean**
a	7.1	6.1	6.9	5.6	$\bar{x}_a = 6.43$
b	6.7	5.0	5.9	5.1	$\bar{x}_b = 5.68$
c	6.6	5.4	5.8	5.2	$\bar{x}_c = 5.75$
	$\bar{x}_1 = 6.80$	$\bar{x}_2 = 5.50$	$\bar{x}_3 = 6.20$	$\bar{x}_4 = 5.30$	$\bar{\bar{x}} = 5.95$

Drug variation

Cages variation

Splitting sum of squares

One Way ANOVA

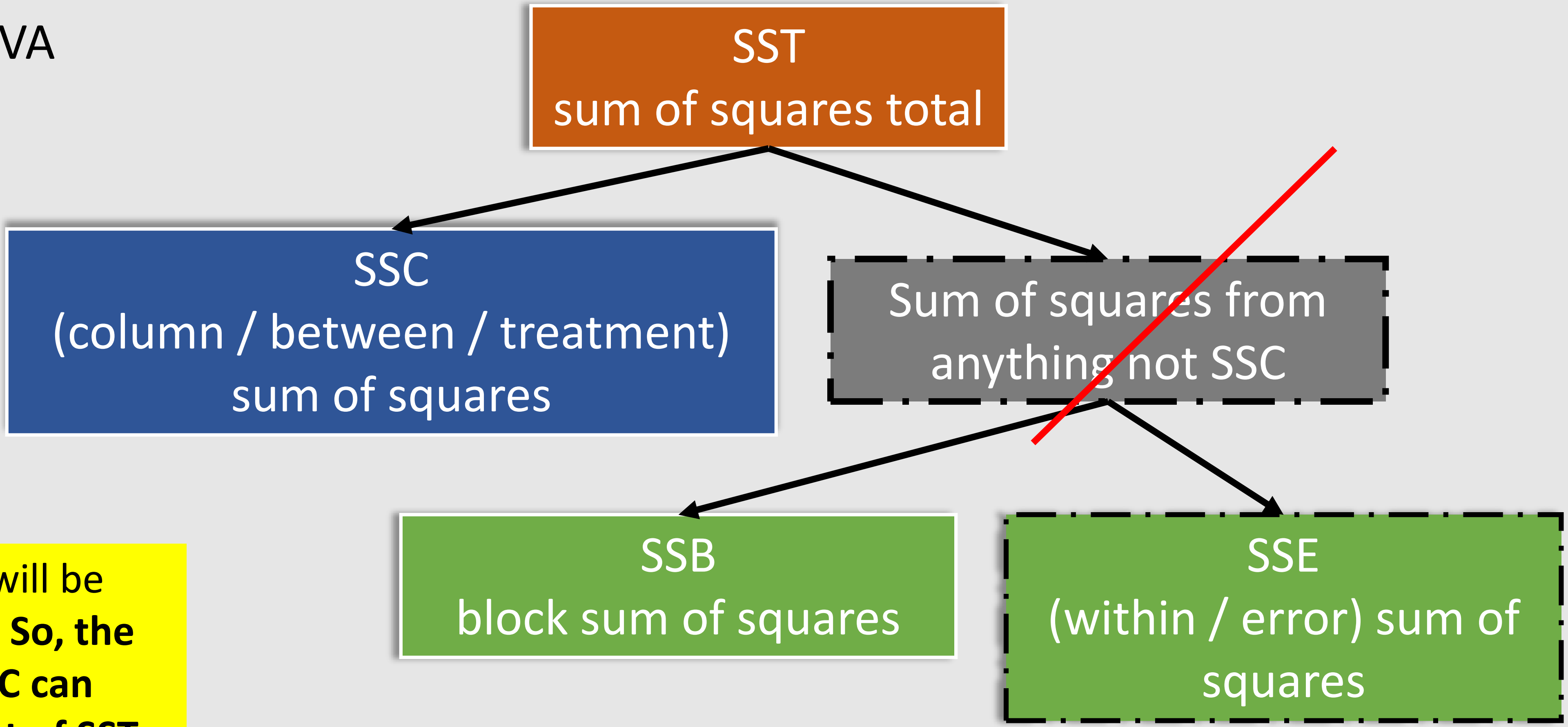
SST (sum of squares total)
(total / overall)

SSC
(column / between / treatment)
sum of squares

SSE
(within / error)
sum of squares

Splitting sum of squares

Two-Way ANOVA



In then end, SSC will be compared to SSE. **So, the smaller SSE is, SSC can claim a larger part of SST.**

SST calculation

	Cages (κλουβιά)				
Drugs	1	2	3	4	
a	7.1	6.1	6.9	5.6	6.425
b	6.7	5	5.9	5.1	5.675
c	6.6	5.4	5.8	5.2	5.75
	6.8	5.5	6.2	5.3	5.95

5.870

```
=VAR.S(B3:E5)*(COUNT(B3:E5)-1)
```

SST (total/overall) sum of squares

1. Find difference between each data point and the overall mean
2. Square the difference
3. Add them up

$$\bar{\bar{x}} = 5.95$$

SSC calculation

	A	B	C	D	E	F
1		Cages (κλουβιά)				
2	Drugs	1	2	3	4	
3	a	7.1	6.1	6.9	5.6	6.425
4	b	6.7	5	5.9	5.1	5.675
5	c	6.6	5.4	5.8	5.2	5.75
6		6.8	5.5	6.2	5.3	5.95

1.365

=VAR.S(F3:F5)*2*4

SSC (column/between) sum of squares

1. Find difference between each drug mean and the overall mean
2. Square the deviations
3. Add them up
4. Multiply by the number of cages

$$\bar{\bar{x}} = 5.95$$

SSB calculation

	A	B	C	D	E	F
1		Cages (κλουβιά)				
2	Drugs	1	2	3	4	
3	a	7.1	6.1	6.9	5.6	6.425
4	b	6.7	5	5.9	5.1	5.675
5	c	6.6	5.4	5.8	5.2	5.75
6		6.8	5.5	6.2	5.3	5.95
7		=AVERAGE(B3:B5)	=AVERAGE(C3:C5)	=AVERAGE(D3:D5)	=AVERAGE(E3:E5)	
8						
9		0.850	-0.450	0.250	-0.650	
10		=B6-$F6	=C6-$F6	=D6-$F6	=D6-$F6	
11						
12		0.722	0.203	0.063	0.423	
13		=POWER(B9,2)	=POWER(C9,2)	=POWER(D9,2)	=POWER(E9,2)	
14						
15		4.230				
16		=SUM(B12:E12)*3				

SSB
block sum of squares

1. Find difference between each cage mean and the overall mean
2. Square the deviations
3. Add them up
4. Multiply by the number of drugs

$$\bar{\bar{x}} = 5.95$$

Sum of squares

SS Total = SS Drugs + SS Cages + SSE

If we calculate 3 of the 4, we can find the 4^{th} by subtraction.

Rearrange algebraically

SSE = SS Total − SS Cages − SS Drugs

SSE calculation

SSE = SS Total − SS Cages − SS Drugs

SSE = 5.87 − 4.23 − 1.365

SSE = 0.275

two-way ANOVA without replication

Two-way ANOVA is presented in the form of the following table:

Source of variation	df (degrees of freedom)	SS (Sum of squares)
Between cages	4-1=3	4.23
Between drugs	3-1=2	1.365
Error	11-3-2=6	0.275
Total	12-1=11 (N-1)	5.87

N = number of observations

Tests of Between-Subjects Effects

Dependent Variable: cells

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.595[a]	5	1.119	24.415	.001
Intercept	424.830	1	424.830	9269.018	.000
drugs	1.365	2	.683	14.891	.005
cages	4.230	3	1.410	30.764	.000
Error	.275	6	.046		
Total	430.700	12			
Corrected Total	5.870	11			

a. R Squared = .953 (Adjusted R Squared = .914)

Two-Way ANOVA without replication

We are interested in the relationship between column variance (drug variance) and error variance

SS Drugs

sum of square drugs

÷

SSE

sum of square error

df

MS Drugs

mean square drugs

÷

MSE

mean square error

Since this a ratio of two variances, it will be an F-ratio and follow the F-distribution

Two-Way ANOVA without replication

We are interested in the relationship between column variance (drug variance) and error variance

$$\frac{\text{SS drugs}}{\text{df drugs}} \div \frac{\text{SSE}}{\text{df Error}} = \frac{1.365}{2} \div \frac{0.275}{6} = 0.6825 \div 0.04583 = \text{MS drugs} \div \text{MSE}$$

Two-Way ANOVA without replication

Two-way ANOVA is presented in the form of the following table:

Source of variation	df (degrees of freedom)	SS (Sum of squares)	MS=SS/df (Mean Squares)	F=MS/Error MS
Between cages	4-1=3	4.23	1.41	
Between drugs	3-1=2	1.365	0.683	14.89
Error	11-3-2=6	0.275	0.046	
Total	12-1=11 (N-1)	1.83		

N = number of observations

Tests of Between-Subjects Effects

Dependent Variable: cells

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.595[a]	5	1.119	24.415	.001
Intercept	424.830	1	424.830	9269.018	.000
drugs	1.365	2	.683	14.891	.005
cages	4.230	3	1.410	30.764	.000
Error	.275	6	.046		
Total	430.700	12			
Corrected Total	5.870	11			

a. R Squared = .953 (Adjusted R Squared = .914)

F-Test

- We test if the drugs differ from each other by comparing the F-value $F = \frac{MS_{Between\ drugs}}{MS_{error}} = 14.89$ with the 5% point of the F-distribution with 2 and 6 degrees of freedom (Between drugs df and Error df)
- Because the F-value of 14.89 is greater than the critical value of the F-distribution, which is 5.14 (see the F-distribution table at the end), there is a significant difference between the drugs with a probability of error P<0.05

Tests of Between-Subjects Effects

Dependent Variable: cells

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.595[a]	5	1.119	24.415	.001
Intercept	424.830	1	424.830	9269.018	.000
drugs	1.365	2	.683	14.891	.005
cages	4.230	3	1.410	30.764	.000
Error	.275	6	.046		
Total	430.700	12			
Corrected Total	5.870	11			

a. R Squared = .953 (Adjusted R Squared = .914)

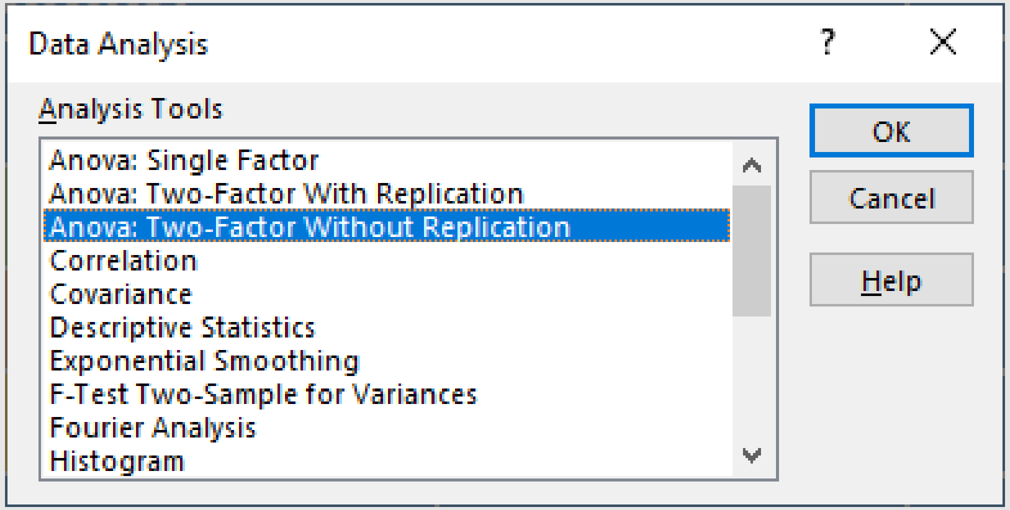
Finding the 5% point of the F-distribution with 2 and 6 degrees of freedom in Excel

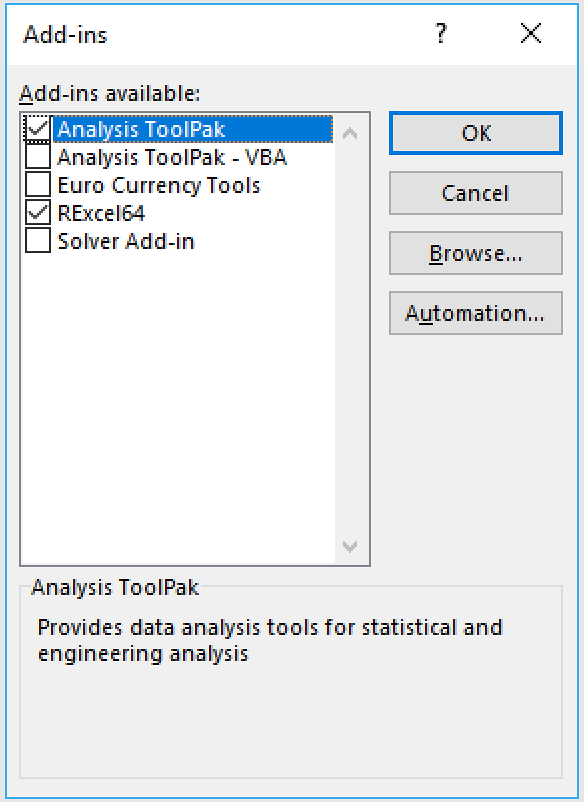
5.143
=F.INV.RT(0.05,2,6)

Excel Data Analysis

File -> Options -> Add Ins -> Go … -> Analysis ToolPak

Data -> Data Analysis

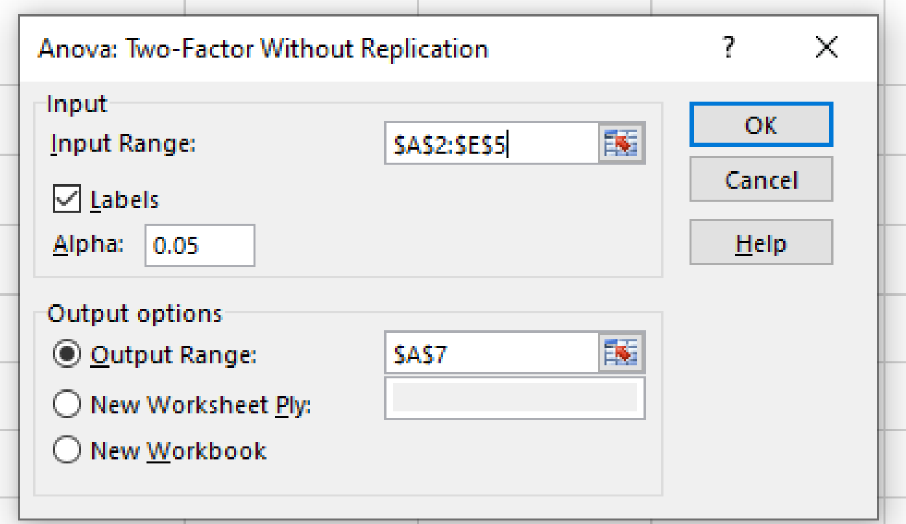




Excel Data Analysis

Data -> Data Analysis

	Cages (κλουβιά)			
Drugs	1	2	3	4
a	7.1	6.1	6.9	5.6
b	6.7	5	5.9	5.1
c	6.6	5.4	5.8	5.2



SPSS vs EXCEL

Source of variation	df (Degrees of freedom)	SS (Sum of squares)	MS=SS/df (Mean Squares)	F=MS/Error MS
Between cages	4-1=3	4.23	1.41	
Between drugs	3-1=2	1.365	0.683	14.89
Error	11-3-2=6	0.275	0.046	
Total	12-1=11	1.83		

Tests of Between-Subjects Effects

Dependent Variable: cells

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.595[a]	5	1.119	24.415	.001
Intercept	424.830	1	424.830	9269.018	.000
drugs	1.365	2	.683	14.891	.005
cages	4.230	3	1.410	30.764	.000
Error	.275	6	.046		
Total	430.700	12			
Corrected Total	5.870	11			

a. R Squared = .953 (Adjusted R Squared = .914)

20	ANOVA						
21	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
22	Rows	1.365	2	0.6825	14.89090909	0.004714835	5.14325285
23	Columns	4.23	3	1.41	30.76363636	0.000486046	4.757062663
24	Error	0.275	6	0.045833333			
25							
26	Total	5.87	11				

Two-way ANOVA with replication

Elias Zintzaras, M.Sc., Ph.D.

Professor in Biomathematics-Biometry
Department of Biomathematics
School of Medicine
University of Thessaly

Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA

Theodoros Mprotsis, MSc, PhD
Teacher & Research Fellow
(http://biomath.med.uth.gr)
University of Thessaly
Email: tmprotsis@uth.gr

When to use it

- When the data is categorized by two factors
- When there are multiple observations for each combination of the two factors (replication)
- Then the effects of each factor (i.e., the differences between the levels of the factor) or the interaction between the factors can be analyzed using a two-way analysis of variance (ANOVA) with replication
- **Note: It’s only possible to measure the interaction effect between the two predictor variables in an ANOVA with replication**

Effects of sorbic acid (sa) and water pH on salmonella survival

To investigate the effects of sorbic acid (sa) and water pH on **salmonella** survival, we used three pH levels (5.0, 5.5, 6.0) and two levels of sorbic acid (0, 100 p.p.m.)

For each sa and pH combination, there were three observations

One week later, the number of surviving **salmonella** was measured (log(density/ml))

	ph		
sa	6.0 (1)	5.5 (2)	5.0 (3)
0 (1)	8.2 8.4 8.3	5.9 6.0 6.1	4.3 4.3 4.2
100 (2)	7.6 7.8 7.6	5.0 5.3 5.8	4.1 4.4 4.2

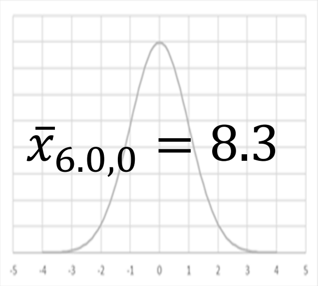
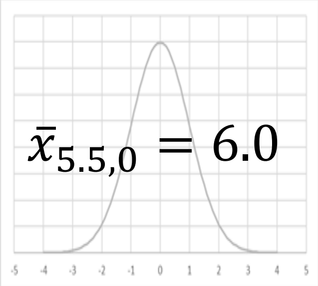
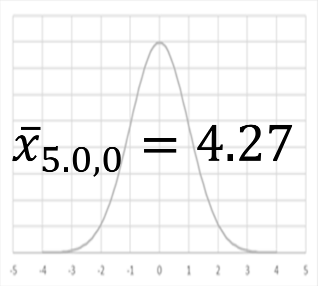
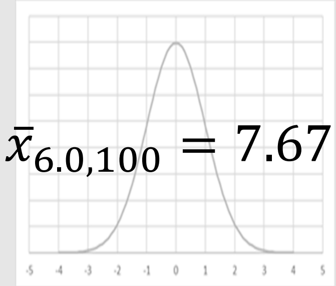
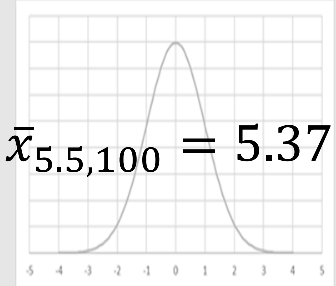
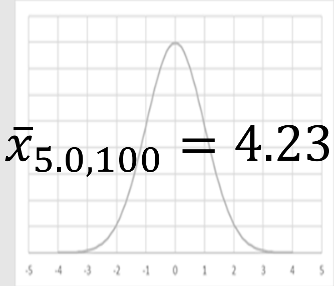
NOW we have a two factor or two-way ANOVA with replication

Effects of sorbic acid (sa) and water pH on salmonella survival

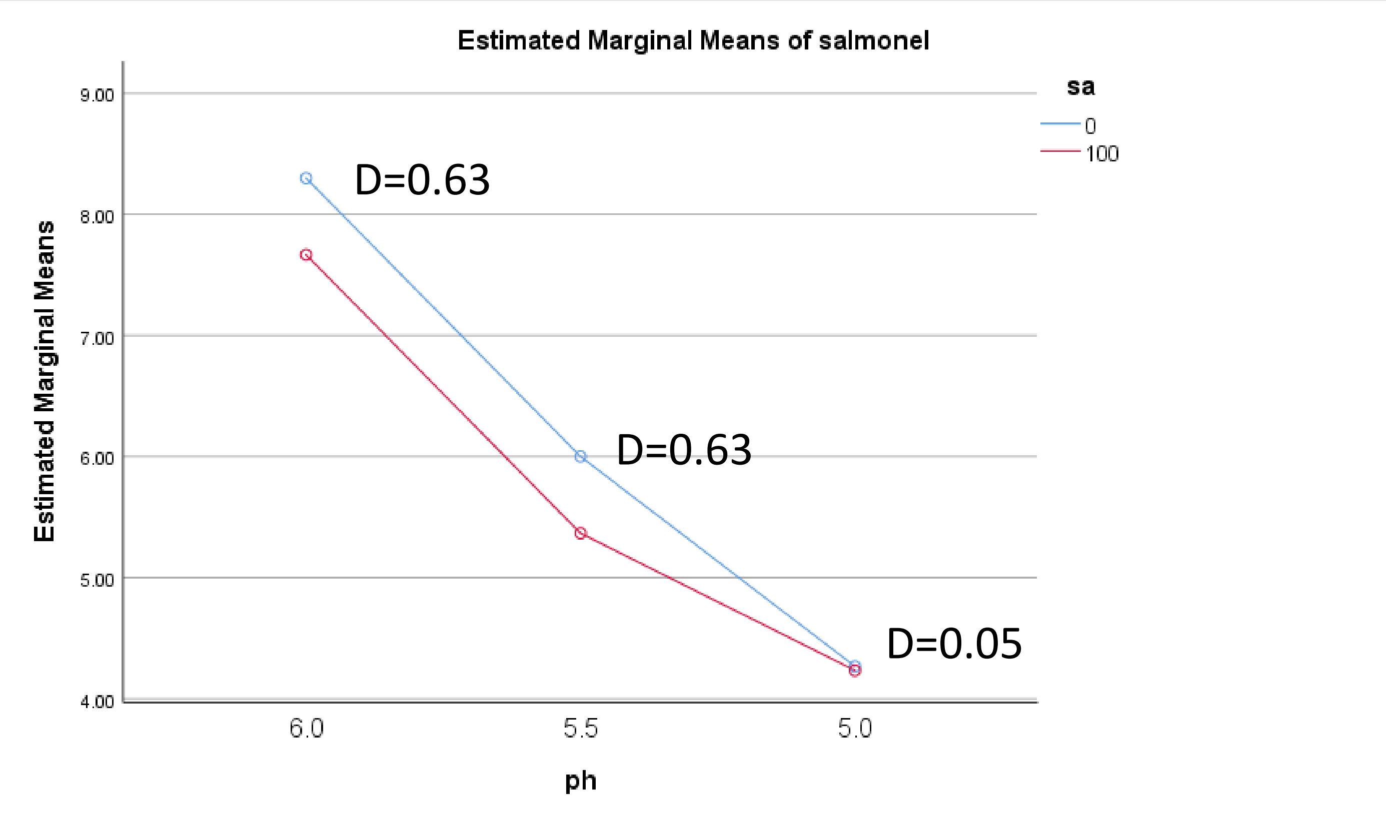
	ph			
sa	6.0	5.5	5.0	**Μέση τιμή**
0	$\bar{x}_{6.0,0} = 8.3$	$\bar{x}_{5.5,0} = 6.0$	$\bar{x}_{5.0,0} = 4.27$	$\bar{x}_{0} = 6.19$
100	$\bar{x}_{6.0,100} = 7.67$	$\bar{x}_{5.5,100} = 5.37$	$\bar{x}_{5.0,100} = 4.23$	$\bar{x}_{100} = 5.76$
	$\bar{x}_{6.0} = 7.99$	$\bar{x}_{5.5} = 5.69$	$\bar{x}_{6.0} = 4.25$	$\bar{\bar{x}} = 5.98$

Each cell has its own mean and its own distribution

Effects of sorbic acid (sa) and water pH on salmonella survival

	ph			
sa	6.0	5.5	5.0	**Μέση τιμή**
0	 $\bar{x}_{6.0,0} = 8.3$	 $\bar{x}_{5.5,0} = 6.0$	 $\bar{x}_{5.0,0} = 4.27$	$\bar{x}_{0} = 6.19$
100	 $\bar{x}_{6.0,100} = 7.67$	 $\bar{x}_{5.5,100} = 5.37$	 $\bar{x}_{5.0,100} = 4.23$	$\bar{x}_{100} = 5.76$
	$\bar{x}_{6.0} = 7.99$	$\bar{x}_{5.5} = 5.69$	$\bar{x}_{6.0} = 4.25$	$\bar{\bar{x}} = 5.98$

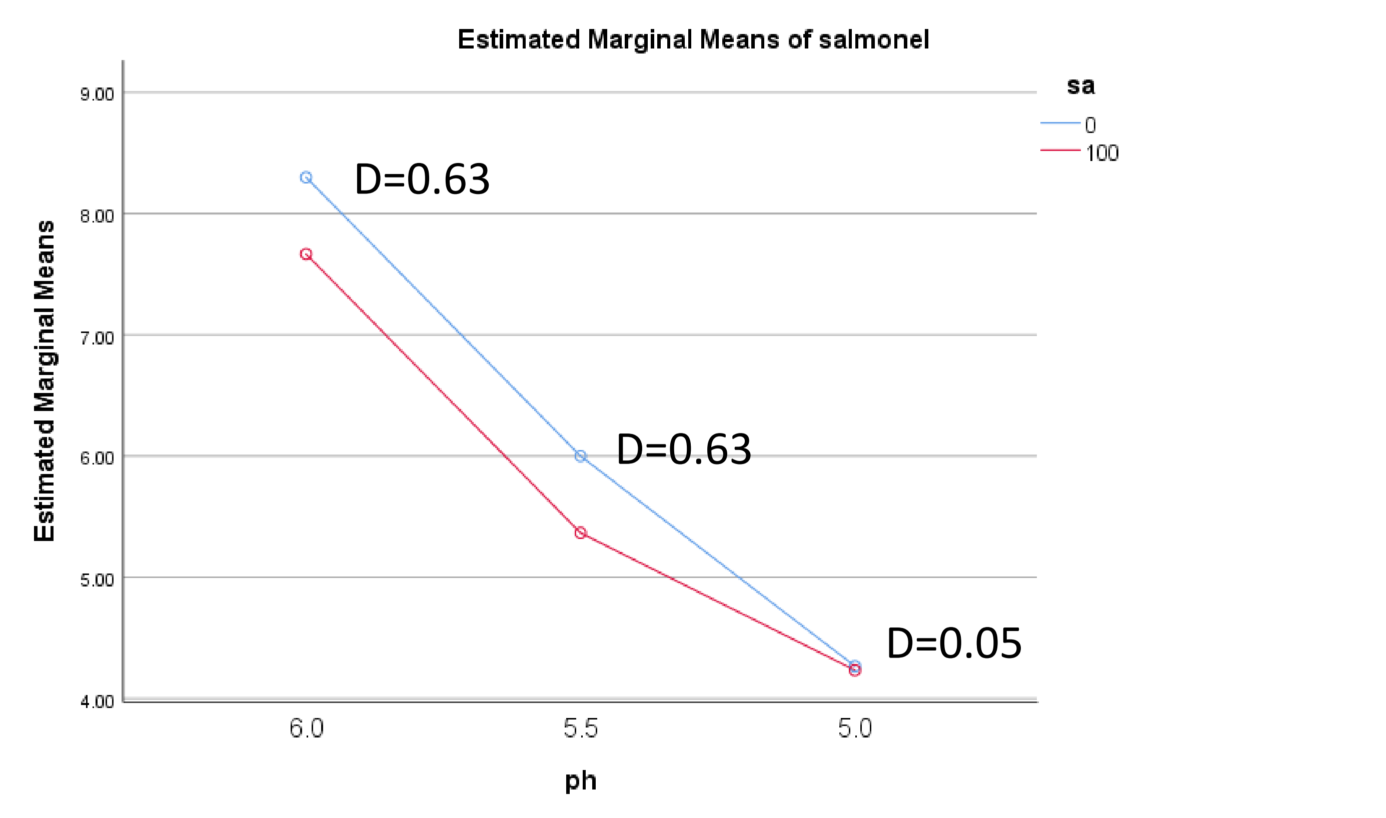
Estimated Marginal Means (SPSS)



- For the pH levels 5.5 and 6.0, we observe a higher number of surviving **salmonella** at the 0 level of sa
- However, at the pH 5.0 level, we observed no difference in the number of surviving **salmonella** between the two levels of sa

Do the two levels of sa consistently increase **salmonella** survival across all pH levels?

Estimated Marginal Means (SPSS)



The answer is: NO

This kind of situation is called an **interaction**

The presence of the **interaction** means that the difference (D) sa0–sa100 is not constant

In a **marginal means graph**, a general rule is to examine whether the lines converge or tend to converge, as this indicates a **statistically significant interaction**

Two-Way ANOVA with interaction

Two-way ANOVA is presented in the form of the following table:

Source of variation	df	SS	MS=SS/df	$F=MS/s^2$
pH	w-1=2	42.46	21.23	590
Sa	s-1=1	0.86	0.86	23.89
Interaction pH*sa	(w-1)*(s-1)=2	0.34	0.17	4.72
Error (Residual)	17-2-1-2=12	0.43	s^2=0.036	
Total n=18-1=17				

3.885293835

=F.INV.RT(0.05, 2, 12)

- We test if the pH levels differ by comparing the F-value $F = \frac{MS_{pH}}{MS_{Residual}} = 590$ with the 5% critical value of the F-distribution with 2 and 12 degrees of freedom (df for pH and residual, respectively), which is 3.89 (see F-distribution table at the end)
- Since the F-value of 590 is much greater than 3.89, there is a significant difference between the pH levels ($P < 0.05$)

Two-Way ANOVA with interaction

Source of variation	df	SS	MS=SS/df	F=MS/s^2
pH	w-1=2	42.46	21.23	590
Sa	s-1=1	0.86	0.86	23.89
Interaction pH*sa	(w-1)*(s-1)=2	0.34	0.17	4.72
Error (Residual)	17-2-1-2=12	0.43	s^2=0.036	
Total n=18-1=17				

4.747225347

=F.INV.RT(0.05, 1, 12)

- We test if the sa levels differ by comparing the F-value $F = \frac{MS_{sa}}{MS_{Residual}} = 23.89$ with the 5% critical value of the F-distribution with 1 and 12 degrees of freedom (df for sa and residual, respectively), which is 4.75 (see F-distribution table at the end)
- Since the F-value of 23.89 is much greater than 4.75, there is a significant difference between the pH levels ($P < 0.05$)

Two-Way ANOVA with interaction

Source of variation	df	SS	MS=SS/df	$F=MS/s^2$
pH	w-1=2	42.46	21.23	590
Sa	s-1=1	0.86	0.86	23.89
Interaction pH*sa	(w-1)*(s-1)=2	0.34	0.17	4.72
Error (Residual)	17-2-1-2=12	0.43	s^2=0.036	
Total n=18-1=17				

3.885293835
=F.INV.RT(0.05, 2, 12)

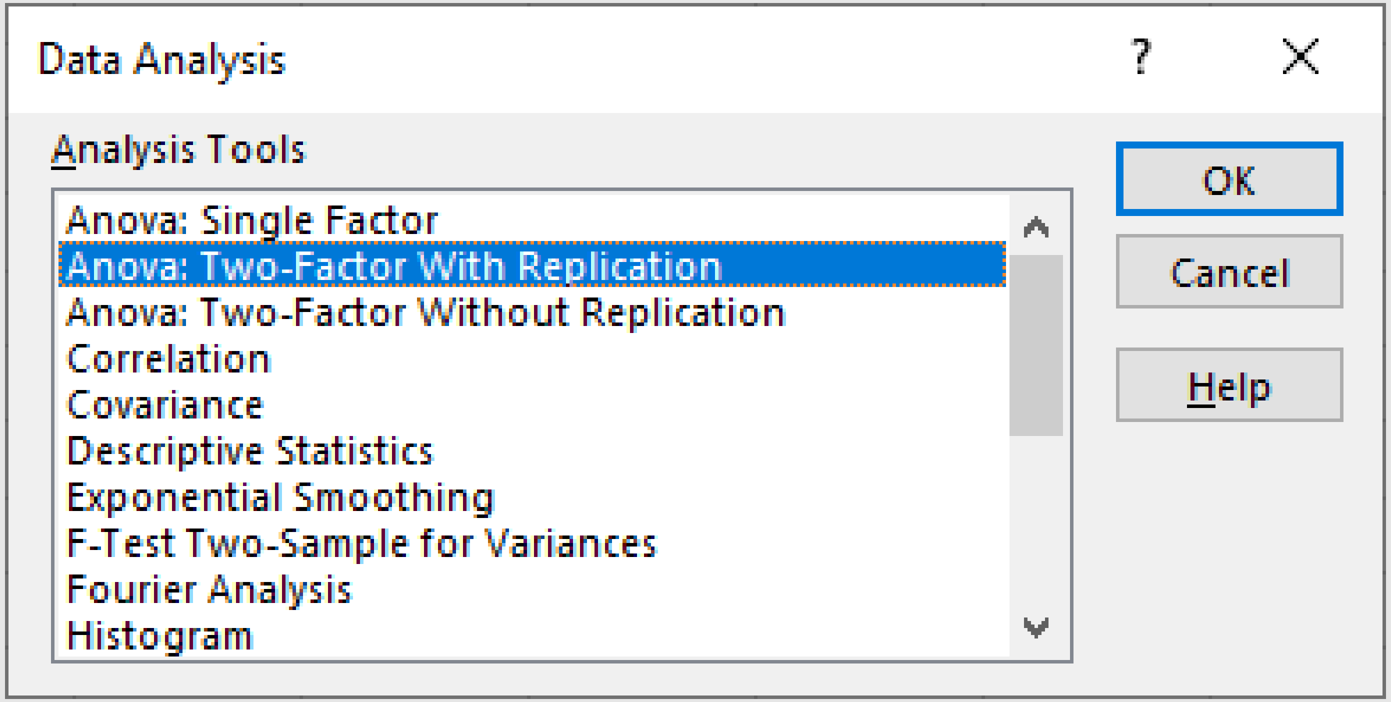
- We test if there is an interaction between pH and sa by comparing the F-value $F = \frac{MS_{Interaction\ pH*sa}}{MS_{Residual}} = 4.72$ with the 5% critical value of the F-distribution with 2 and 12 degrees of freedom (Interaction df and residual df, respectively), which is 3.89 (see F-distribution table at the end)
- Since the F-value of 4.72 is greater than 3.89, there is a significant interaction between pH and sa ($P < 0.05$)

Two-Way ANOVA with interaction

- We can compare the means of two pH levels or two sa levels using a t-test, similar to how it is done in one-way ANOVA
- Additionally, we can compare the mean values of two pH levels for one level of sa, again using a t-test, as in one-way ANOVA

Excel

Data -> Data Analysis



Excel

Tests of Between-Subjects Effects

Dependent Variable: salmonel

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	43.769[a]	5	8.754	235.179	.000
Intercept	642.014	1	642.014	17248.134	.000
sa	.845	1	.845	22.701	.000
ph	42.564	2	21.282	571.761	.000
sa * ph	.360	2	.180	4.836	.029
Error	.447	12	.037		
Total	686.230	18			
Corrected Total	44.216	17			

a. R Squared = .990 (Adjusted R Squared = .986)

32	ANOVA						
33	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
34	Sample	0.845	1	0.845	22.70149	0.000460488	4.747225
35	Columns	42.56444444	2	21.28222222	571.7612	1.25435E-12	3.885294
36	Interaction	0.36	2	0.18	4.835821	0.028822828	3.885294
37	Within	0.446666667	12	0.037222222			
38							
39	Total	44.21611111	17				

F-Distribution

Degrees of freedom in denominator	Degrees of freedom in numerator												
	1	2	3	4	5	6	7	8	9	10	20	40	∞
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90	248.00	251.10	254.30
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.47	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.59	8.50
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.72	5.60
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.46	4.40
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.77	3.70
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.34	3.20
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.04	2.90
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.83	2.70
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.66	2.50
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.53	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.43	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.34	2.20
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.27	2.10
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.20	2.10
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.15	2.00
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.10	2.00
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.06	1.90
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.03	1.90
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	1.99	1.80
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.79	1.60
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.69	1.50
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.66	1.50	1.30
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.57	1.39	1.00