



One-Way ANOVA

Ανάλυση διακύμανσης ως προς έναν παράγοντα (One-Way ANOVA)

Ζιντζαράς Ηλίας, M.Sc., Ph.D.

Καθηγητής Βιομαθηματικών-Βιομετρίας

Εργαστήριο Βιομαθηματικών

Τμήμα Ιατρικής

Πανεπιστήμιο Θεσσαλίας

Institute for Clinical Research and Health Policy Studies

Tufts University School of Medicine

Boston, MA, USA

Θεόδωρος Μπρότσης, MSc, PhD Candidate

Ακαδημαϊκός Υπότροφος

(<http://biomath.med.uth.gr>)

Πανεπιστήμιο Θεσσαλίας

Email: tmprotsis@uth.gr



Πότε;

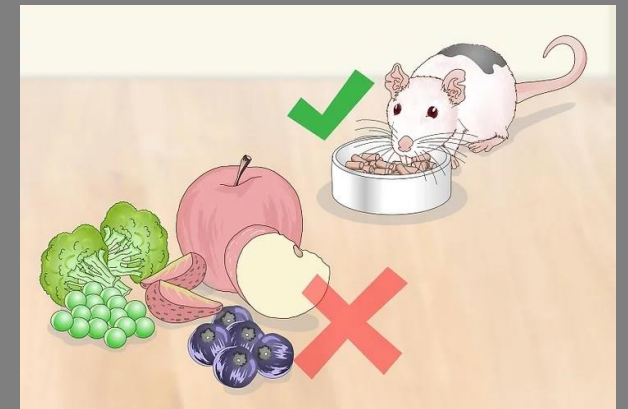
- Αν θέλουμε να ελέγξουμε αν η μέση τιμή μίας ποσοτικής μεταβλητής διαφέρει ανάμεσα σε δύο ανεξάρτητα δείγματα, εφαρμόζουμε το **t-test**.
- Στην περίπτωση που η ποιοτική μεταβλητή έχει περισσότερες από δύο κατηγορίες **δεν** εφαρμόζουμε το t-test.
- Έτσι για να ελέγξουμε αν διαφέρουν οι μέσες τιμές μίας ποσοτικής μεταβλητής, ανάμεσα στις κατηγορίες μίας ποιοτικής, όταν αυτή έχει **περισσότερες από δύο** κατηγορίες χρησιμοποιούμε την Ανάλυση διασποράς μίας κατεύθυνσης (One Way Anova).
- Εναλλακτικά θα μπορούσαμε να χρησιμοποιήσουμε το t-test σε όλα τα ζεύγη δειγμάτων. Η διαδικασία όμως αυτή είναι αρκετά επίμονη όταν οι ομάδες είναι πολλές, π. χ. για 5 ομάδες θα πρέπει να γίνουν 10 t – έλεγχοι. Επίσης, αυξάνεται η πιθανότητα λάθους.
- Για να δούμε ποιες ομάδες είναι διαφορετικές κάνουμε χρήση **post-hoc test**.



Προϋποθέσεις

- Η ποσοτική μεταβλητή κατανέμεται κανονικά σε κάθε κατηγορία της ποιοτικής (Shapiro-Wilk test για $n < 50$ και Kolmogorov Smirnov για $n > 50$)
- Οι διασπορές της ποσοτικής μεταβλητής σε κάθε κατηγορία της ποιοτικής να είναι ίσες ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2 = \sigma^2$) (Levene's test)
- Οι ομάδες είναι ανεξάρτητες

Διαφορές μεταξύ διατροφών σε ποντίκια





Διαφορές μεταξύ διαιτών



Σε μία μελέτη, έχει καταγραφεί το βάρος του ήπατος (εκφρασμένο ως ποσοστό του βάρους του σώματος) ποντικών που ανήκουν σε $k=4$ ομάδες που τράφηκαν με 4 δίαιτες. Θέλουμε να ερευνήσουμε αν υπάρχουν συστηματικές διαφορές μεταξύ των 4 ομάδων.

	a	b	c	d
	3.42	3.17	3.34	3.64
	3.96	3.63	3.72	3.93
	3.87	3.38	3.81	3.77
	4.19	3.47	3.66	4.18
	3.58	3.39	3.55	4.21
	3.76	3.41	3.51	3.88
Μέση τιμή	3.80	3.41	3.60	3.94



Εισαγωγή δεδομένων

Στη μεταβλητή **group** τα **1, 2, 3** και **4** αντιπροσωπεύουν τις ομάδες **a, b, c** και **d**, αντίστοιχα.

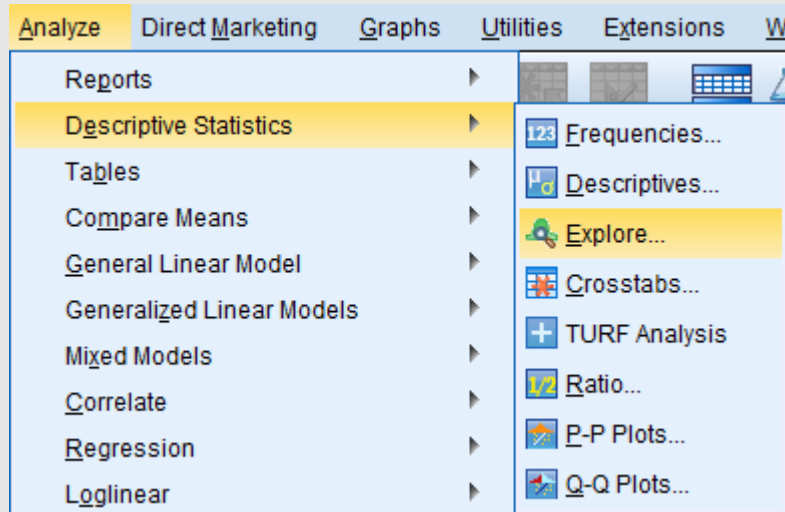
	 group	 weight
1	1	3.42
2	1	3.96
3	1	3.87
4	1	4.19
5	1	3.58
6	1	3.76
7	2	3.17
8	2	3.63
9	2	3.38
10	2	3.47
11	2	3.39
12	2	3.41

13	3	3,34
14	3	3,72
15	3	3,81
16	3	3,66
17	3	3,55
18	3	3,51
19	4	3,64
20	4	3,93
21	4	3,77
22	4	4,18
23	4	4,21
24	4	3,88

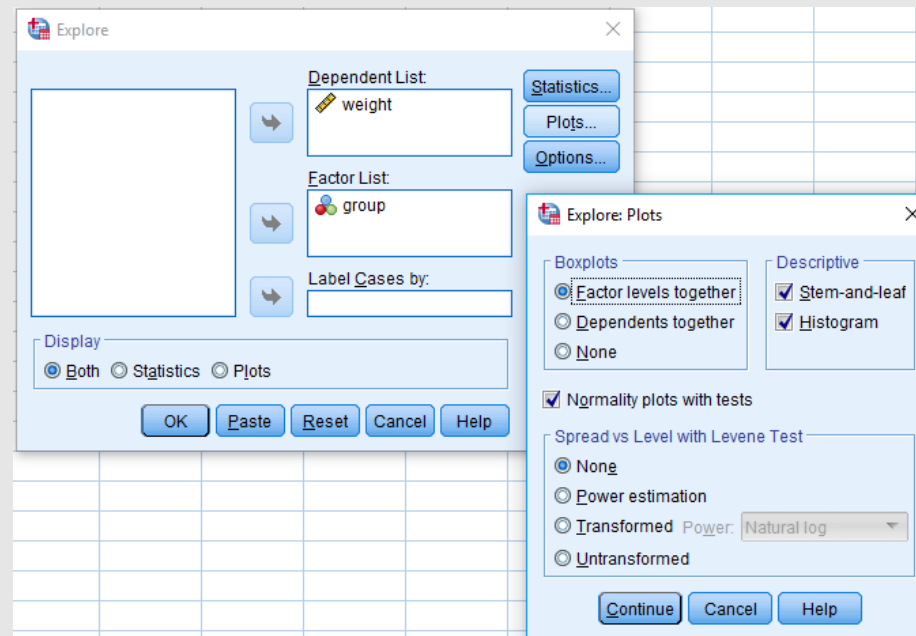


Έλεγχος κανονικότητας

Για τον έλεγχο κανονικότητας επιλέγουμε **Analyze -> Descriptive Statistics -> Explore**



Κάνουμε τις ρυθμίσεις όπως παρακάτω, πατάμε **Continue** και στην συνέχεια **OK**





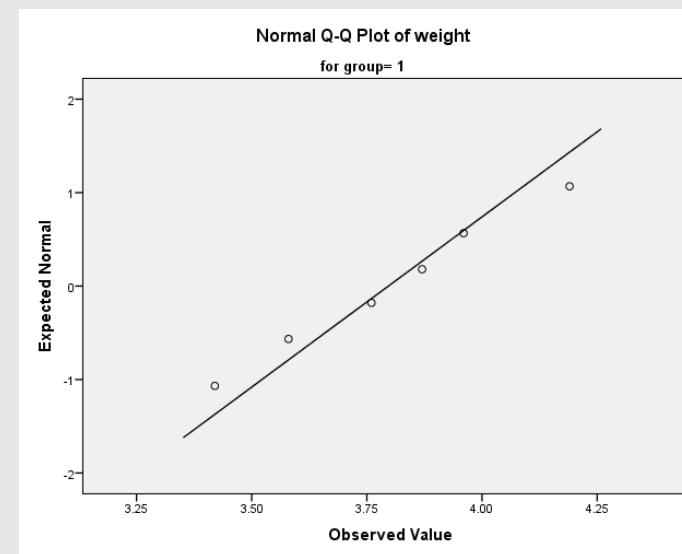
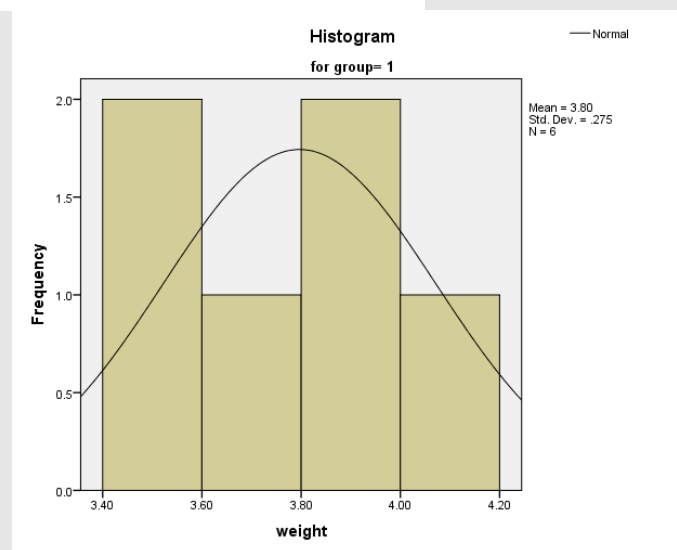
Έλεγχος κανονικότητας

group	Kolmogorov-Smirnov ^a			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
weight	1	.118	6	.200 [*]	.992	6	.993
	2	.258	6	.200 [*]	.940	6	.657
	3	.144	6	.200 [*]	.980	6	.953
	4	.195	6	.200 [*]	.934	6	.609

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

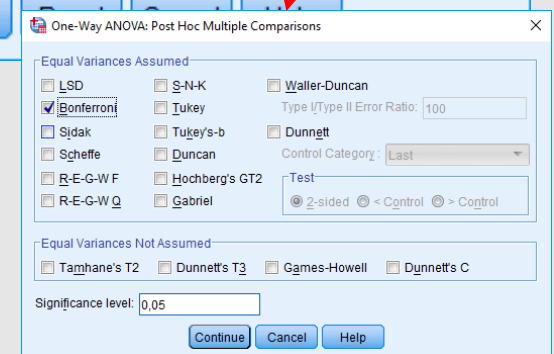
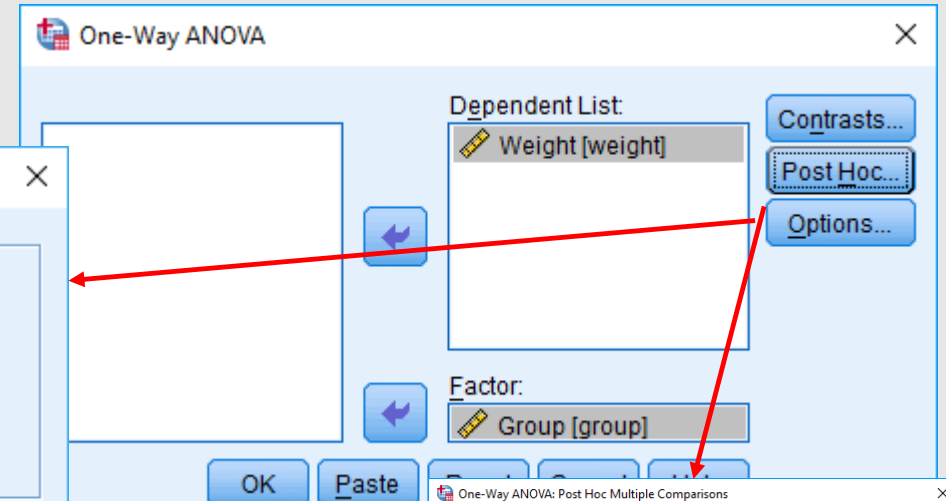
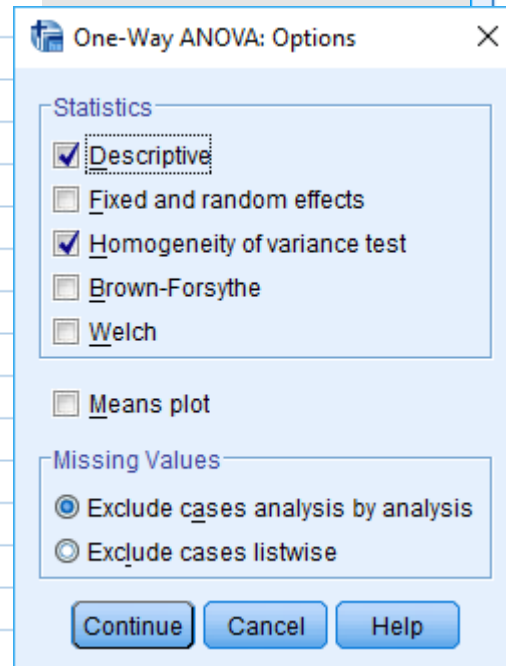
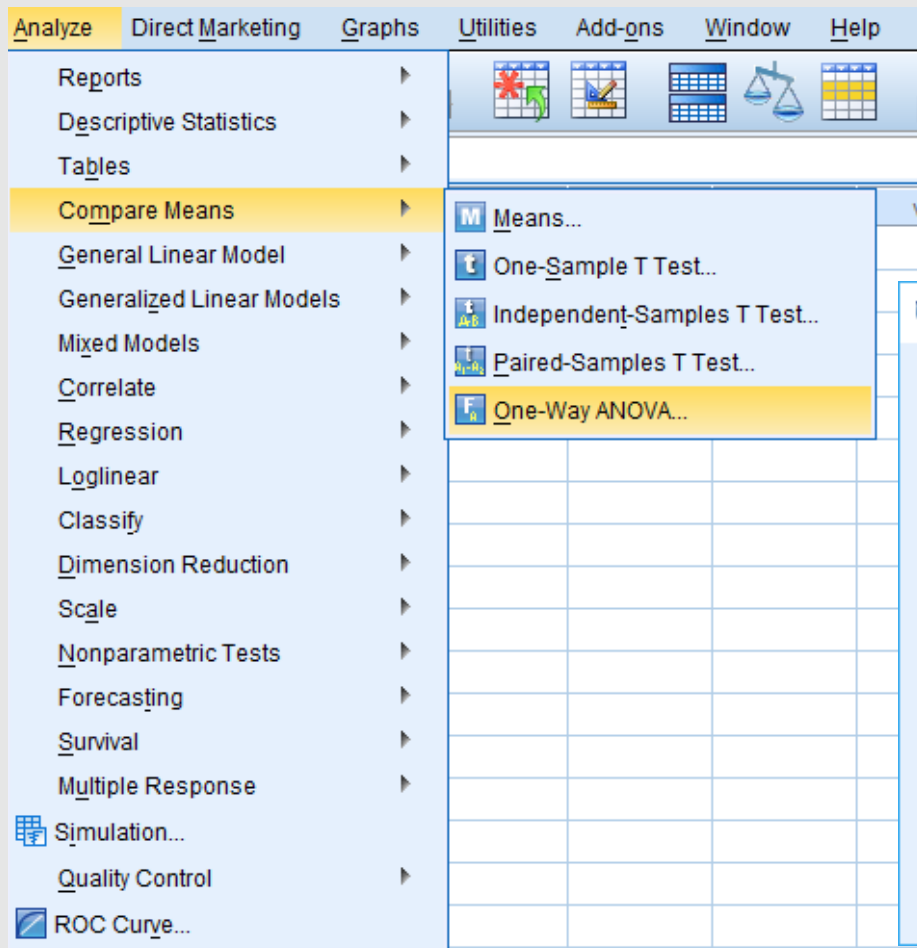
Στον πίνακα **Test of Normality -> Shapiro-Wilk (n<50)** βλέπουμε πως η τιμή **Sig. (p-value)** για κάθε μία κατηγορία είναι > 0.05 (δεν απορρίπτουμε την H_0), οπότε η ποσοτική μεταβλητή κατανέμεται κανονικά σε κάθε κατηγορία της ποιοτικής





Εκτέλεση One-Way ANOVA

Για να αναλύσετε τα δεδομένα επιλέξτε **Analyze -> Compare Means -> One Way ANOVA**





Συμπεράσματα

Descriptives								
weight	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	6	3.7967	.27457	.11209	3.5085	4.0848	3.42	4.19
2	6	3.4083	.14892	.06080	3.2521	3.5646	3.17	3.63
3	6	3.5983	.16750	.06838	3.4226	3.7741	3.34	3.81
4	6	3.9350	.22492	.09182	3.6990	4.1710	3.64	4.21
Total	24	3.6846	.28247	.05766	3.5653	3.8039	3.17	4.21

Test of Homogeneity of Variances				
weight	Levene Statistic	df1	df2	Sig.
	1.076	3	20	.382

Από το test του **Levene** συμπεραίνουμε πως οι διασπορές (διακυμάνσεις) της ποσοτικής μεταβλητής σε κάθε κατηγορία της ποιοτικής είναι ίσες καθώς το p value είναι > 0.05 (3.82)



Συμπεράσματα

- Το Output παράθυρο εμφανίζει την ANOVA και τα αποτελέσματα των πιθανών συγκρίσεων μεταξύ των ομάδων
- Στον πίνακα ANOVA, το **P-value** μεταξύ των ομάδων είναι ("Sig") $P=0.002$, οπότε υπάρχει σημαντική διαφορά μεταξύ των ομάδων ($P<0.05$)
- Ο πίνακας **Post Hoc Tests** δείχνει τις επιμέρους συγκρίσεις μεταξύ των ομάδων: Η σύγκριση/διαφορά μεταξύ των μέσων τιμών των ομάδων 1 και 2 είναι σημαντική, $P=0.027$ (Sig.) και το 95% CI της διαφοράς είναι (0.0336, 0.7430), το οποίο δεν περιέχει το μηδέν. Διαφορά υπάρχει επίσης και μεταξύ των ομάδων 2 και 4.

ANOVA

Weight

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,954	3	,318	7,220	,002
Within Groups	,881	20	,044		
Total	1,835	23			

Post Hoc Tests

Multiple Comparisons

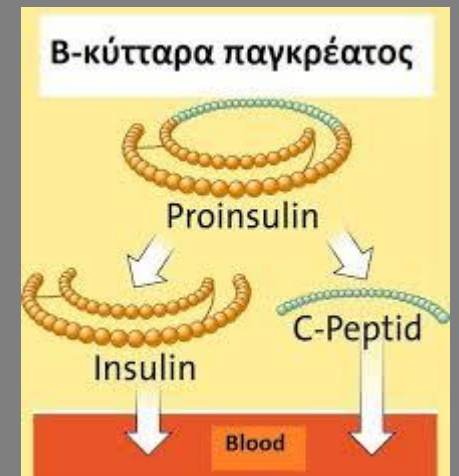
Dependent Variable: Weight

Bonferroni

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	,38833 [*]	,12118	,027	,0336	,7430
	3	,19833	,12118	,704	-,1564	,5530
	4	-,13833	,12118	1,000	-,4930	,2164
2	1	-,38833 [*]	,12118	,027	-,7430	-,0336
	3	-,19000	,12118	,795	-,5447	,1647
	4	-,52667 [*]	,12118	,002	-,8814	-,1720
3	1	-,19833	,12118	,704	-,5530	,1564
	2	,19000	,12118	,795	-,1647	,5447
	4	-,33667	,12118	,070	-,6914	,0180
4	1	,13833	,12118	1,000	-,2164	,4930
	2	,52667 [*]	,12118	,002	,1720	,8814
	3	,33667	,12118	,070	-,0180	,6914

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

Έκκριση Ινσουλίνης σε πειραματόζωα





Έκκριση Ινσουλίνης

Σε ένα πείραμα μετρήθηκε η έκκριση ινσουλίνης σε δείγματα παγκρεατικού ιστού πειραματόζων. Τα δείγματα χωρίστηκαν σε 5 ομάδες με βάση τα επίπεδα γλυκόζης. Κάθε μία ομάδα αντιστοιχεί σε διαφορετικά επίπεδα γλυκόζης. Θέλουμε να δούμε αν υπάρχει στατιστικά σημαντική διαφορά στις μέσες τιμές των κατηγοριών αυτών.

Ομάδα 1	1.53	1.69	3.75	2.89	3.26	2.83	2.86	2.59
Ομάδα 2	3.15	3.96	3.59	1.89	1.45	3.49	1.56	2.44
Ομάδα 3	3.89	4.80	3.68	5.70	5.62	5.79	4.75	5.33
Ομάδα 4	8.18	5.64	7.36	5.33	8.82	5.26	8.75	7.10
Ομάδα 5	5.86	5.46	5.69	6.49	7.81	9.03	7.49	8.98



Εισαγωγή δεδομένων

	group	level
1	1	1,53
2	1	1,69
3	1	3,75
4	1	2,89
5	1	3,26
6	1	2,83
7	1	2,86
8	1	2,59
9	2	3,15
10	2	3,96
11	2	3,59
12	2	1,89
13	2	1,45
14	2	3,49
15	2	1,56
16	2	2,44

1

17	3	3,89
18	3	4,80
19	3	3,68
20	3	5,70
21	3	5,62
22	3	5,79
23	3	4,75
24	3	5,33
25	4	8,18
26	4	5,64
27	4	7,36
28	4	5,33
29	4	8,82
30	4	5,26
31	4	8,75
32	4	7,10

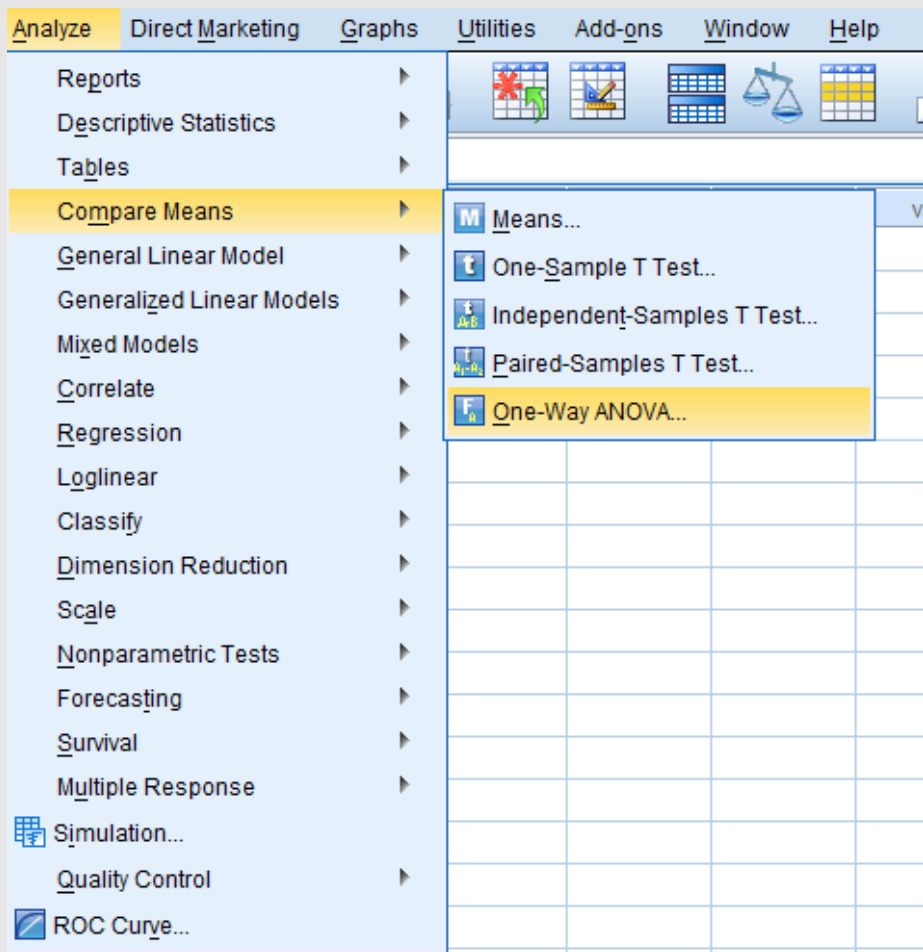
2

33	5	5,86
34	5	5,46
35	5	5,69
36	5	6,49
37	5	7,81
38	5	9,03
39	5	7,49
40	5	8,98

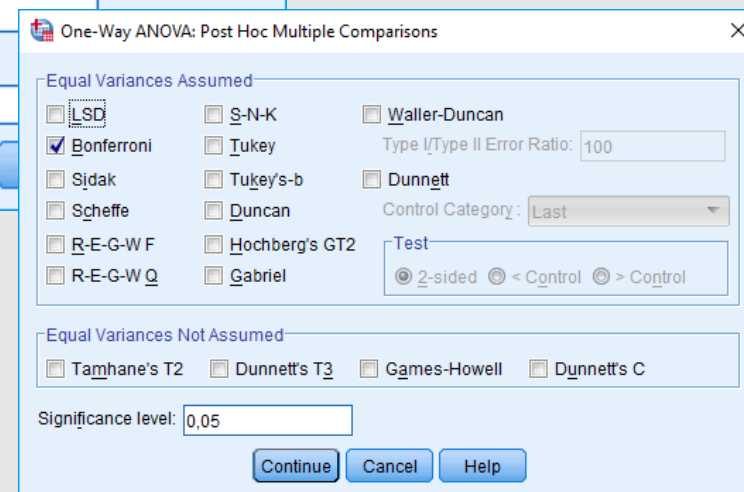
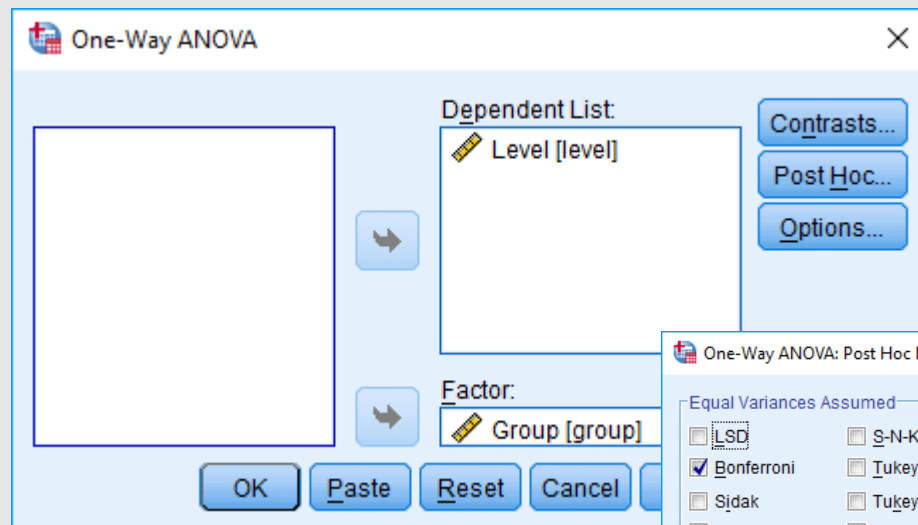
3



Εκτέλεση One-Way ANOVA



Για να αναλύσετε τα δεδομένα επιλέξτε **Analyze -> Compare Means -> One Way ANOVA**





Συμπεράσματα

Level	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	154,564	4	38,641	29,793	,000
Within Groups	45,394	35	1,297		
Total	199,959	39			

Level	Levene Statistic	df1	df2	Sig.
	2,960	4	35	,033

- Το **Output** παράθυρο εμφανίζει την ANOVA και τα αποτελέσματα των πιθανών συγκρίσεων μεταξύ των ομάδων.
 - Στον πίνακα ANOVA, το **P-value** μεταξύ των ομάδων είναι ("Sig") $P=0.000$, οπότε τουλάχιστον μία ομάδα έχει διαφορετική μέση τιμή από τις υπόλοιπες ($P<0.05$).
-
- **Μηδενική υπόθεση:** Το Levene's test ελέγχει την μηδενική υπόθεση ότι οι διακυμάνσεις των ομάδων είναι ίδιες
 - Στο παράδειγμά μας η τιμή Sig. είναι $0.033 < 0.05$, οπότε οι διακυμάνσεις διαφέρουν μεταξύ τους (μικρό δείγμα)



Post Hoc Test

Multiple Comparisons

Dependent Variable: Level
Bonferroni

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-,01625	,56942	1,000	-1,7223	1,6898
	3	-2,27000*	,56942	,003	-3,9760	-,5640
	4	-4,38000*	,56942	,000	-6,0860	-2,6740
	5	-4,42625*	,56942	,000	-6,1323	-2,7202
2	1	,01625	,56942	1,000	-1,6898	1,7223
	3	-2,25375*	,56942	,004	-3,9598	-,5477
	4	-4,36375*	,56942	,000	-6,0698	-2,6577
	5	-4,41000*	,56942	,000	-6,1160	-2,7040
3	1	2,27000*	,56942	,003	,5640	3,9760
	2	2,25375*	,56942	,004	,5477	3,9598
	4	-2,11000*	,56942	,007	-3,8160	-,4040
	5	-2,15625*	,56942	,006	-3,8623	-,4502
4	1	4,38000*	,56942	,000	2,6740	6,0860
	2	4,36375*	,56942	,000	2,6577	6,0698
	3	2,11000*	,56942	,007	,4040	3,8160
	5	-,04625	,56942	1,000	-1,7523	1,6598
5	1	4,42625*	,56942	,000	2,7202	6,1323
	2	4,41000*	,56942	,000	2,7040	6,1160
	3	2,15625*	,56942	,006	,4502	3,8623
	4	,04625	,56942	1,000	-1,6598	1,7523

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

Ο πίνακας **Post Hoc Tests** δείχνει τις επιμέρους συγκρίσεις μεταξύ των ομάδων: Η σύγκριση/διαφορά μεταξύ των μέσων τιμών των ομάδων 1 και 3, 1 και 4, 1 και 5 είναι σημαντική, $P=0.003$ (Sig.), $P=0,000$ (Sig.), $P=0.000$ (Sig.) αντίστοιχα, και το 95% CI της διαφοράς είναι (-3.976, -0,564), (-6.086, -2.674), (-6.1323, -2.7202) αντίστοιχα, το οποίο δεν περιέχει το μηδέν.



Πρακτική άσκηση

Να ελέγξετε εάν ένα νέο αναλγητικό φάρμακο (A) διαφέρει από ένα παλιό φάρμακο (B) και από το placebo (C), ως προς τους χρόνους ανακούφισης. Για τη σύγκριση των αναλγητικών φαρμάκων, 21 ασθενείς διαιρέθηκαν τυχαία στις τρεις ομάδες. Οι χρόνοι ανακούφισης που παρατηρήθηκαν ήταν:

A/A	A	B	C
1	4.56	4.32	2.30
2	5.61	4.89	3.42
3	4.67	5.01	4.21
4	5.09	5.90	1.50
5	6.21	4.88	2.03
6	4.28	5.29	1.87
7	5.23	4.89	3.12