



T-test for independent samples

t-test για ανεξάρτητα δείγματα

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*

Επίδραση του φύλου στην επιθετικότητα





Επίδραση του φύλου στην επιθετικότητα

Άνδρες (N=12)	Γυναίκες (N=13)
12	12
18	9
12	12
10	8
10	10
14	8
14	7
18	13
12	16
8	11
14	15
14	13
	9

Μηδενική υπόθεση (H_0): Οι άνδρες δεν θα παρουσιάσουν υψηλότερα επίπεδα επιθετικότητας σε σχέση με τις γυναίκες



Εισαγωγή δεδομένων και ρύθμιση μεταβλητών

*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct

1 : sex 1

	sex	aggressiveness	var
1	1	12	
2	1	18	
3	1	12	
4	1	10	
5	1	10	
6	1	14	
7	1	14	
8	1	18	
9	1	12	
10	1	8	
11	1	14	
12	1	14	
13	2	12	
14	2	9	
15	2	12	
16	2	8	

Value Labels

Value Labels

Value: 2

Label: Γυναίκες

1,00 = "Ανδρες"

Add Change Remove

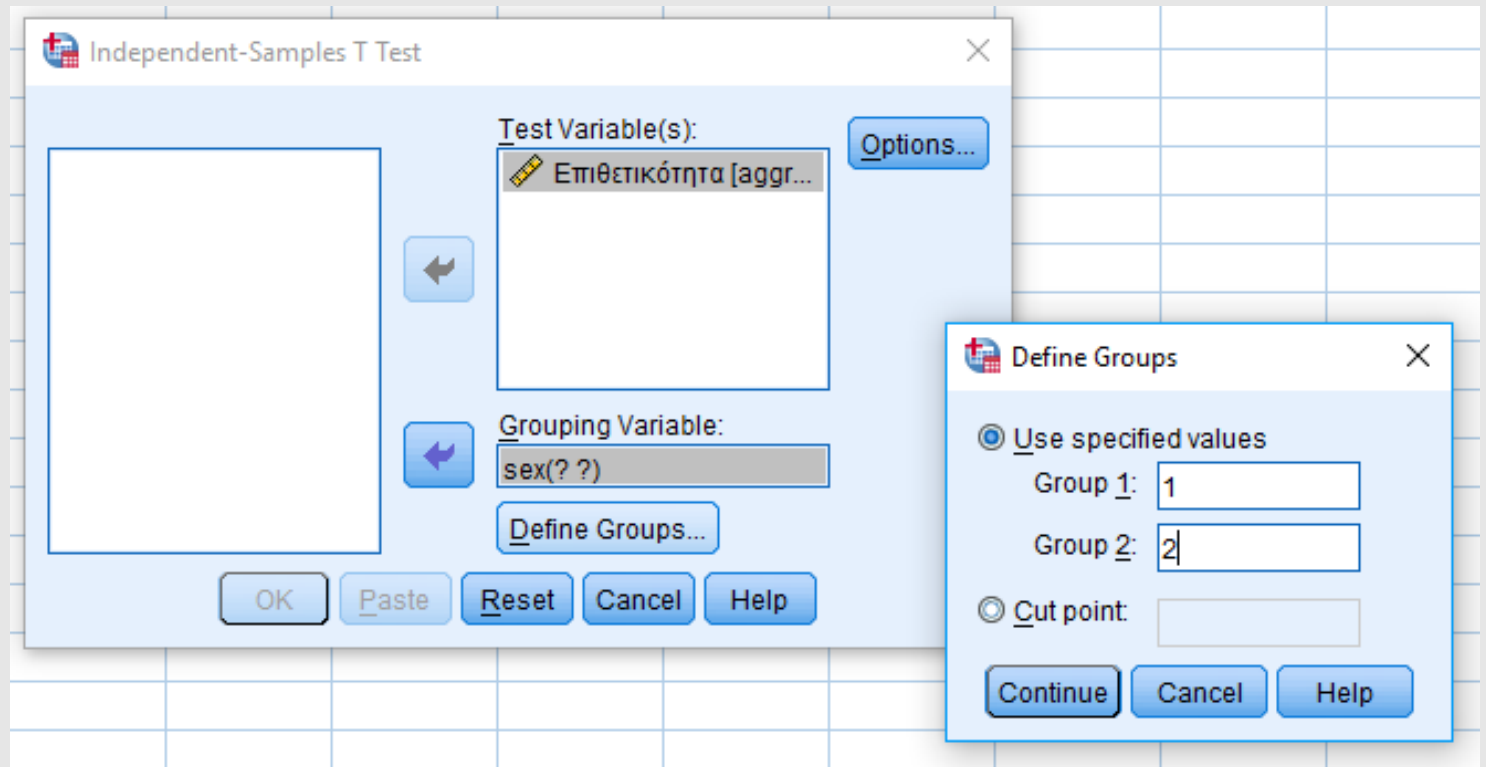
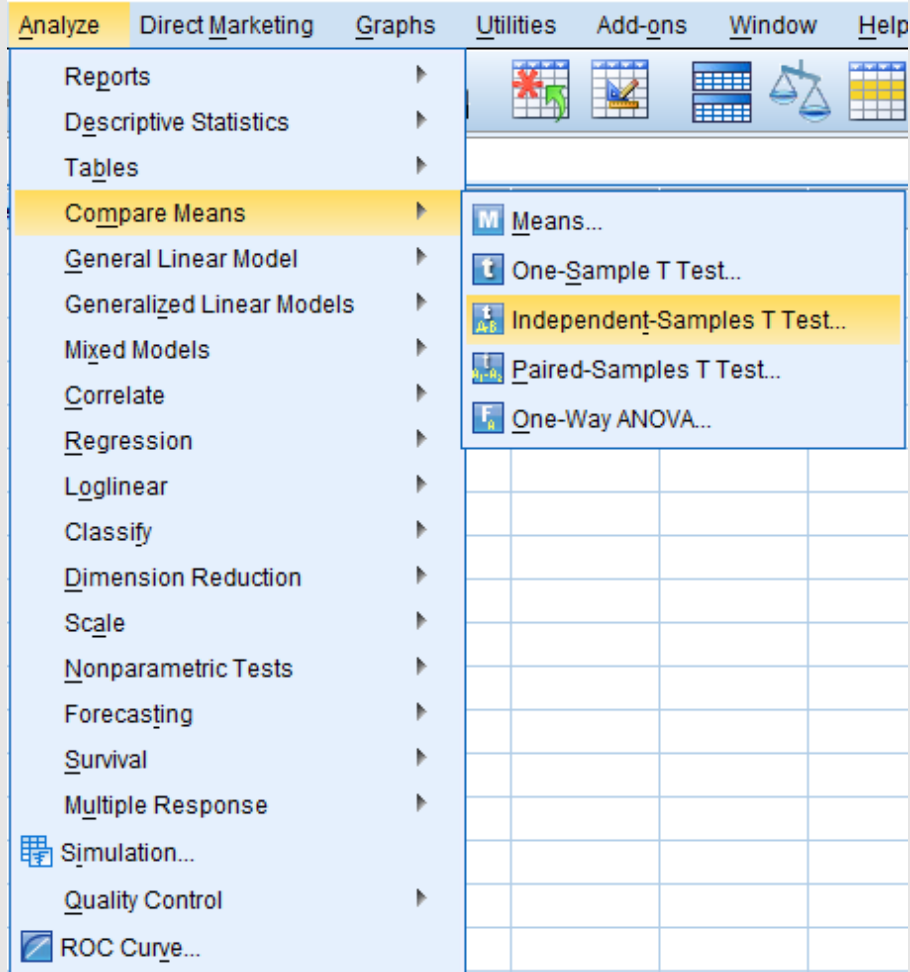
Spelling...

OK Cancel Help

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
sex	Numeric	8	2	Φύλο	{1,00, Ανδρ...	None	8	Right	Nominal	Input
aggressiveness	Numeric	8	2	Επιθετικότητα	None	None	8	Right	Scale	Input



Εκτέλεση του Independent Samples T Test





Ερμηνεία αποτελέσματος

→ T-Test

[DataSet0]

Τα δύο φύλα έχουν ίδια διακύμανση, καθώς το p-value (Sig.) > 0.05 στο Levene's Test.

Group Statistics

φύλο	N	Mean	Std. Deviation	Std. Error Mean
Επιθετικότητα Άντρες	12	13,00	3,015	,870
Επιθετικότητα Γυναίκες	13	11,00	2,799	,776

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Επιθετικότητα	Equal variances assumed	,002	,969	1,720	23	,099	2,000	1,163	-,405	4,405
	Equal variances not assumed			1,715	22,443	,100	2,000	1,166	-,416	4,416

Επειδή το p-value είναι > 0.05 (p-value = 0,099) δεν απορρίπτουμε την μηδενική υπόθεση, δηλ. οι άνδρες δεν παρουσιάζουν υψηλότερα επίπεδα επιθετικότητας σε σχέση με τις γυναίκες (Στατιστικώς μη σημαντικό αποτέλεσμα)

Χρόνος ανακούφισης δύο αναλγητικών φαρμάκων





Χρόνος ανακούφισης δύο αναλγητικών φαρμάκων

Drug (N=16)	time (N=16)
1	7.50
1	3.60
1	8.90
1	6.40
1	9.50
1	7.40
1	10.10
1	4.70
2	3.80
2	2.40
2	4.70
2	5.90
2	2.10
2	3.20
2	4.50
2	3.80

- Έστω ότι θέλουμε να ελέγξουμε εάν δύο αναλγητικά φάρμακα (“1” και “2”) διαφέρουν ως προς τους χρόνους ανακούφισης.
- Για τη σύγκριση των δύο αναλγητικών φαρμάκων “1” και “2”, 16 ασθενείς διαιρέθηκαν τυχαία στις δύο ομάδες.
- Καταρχήν, από το “Variable View” ονομάζουμε τις δύο μεταβλητές (“drug” και “time”) και καταχωρούμε τα φάρμακα και τους χρόνους ανακούφισης που παρατηρήθηκαν για τον κάθε ασθενή.

Μηδενική υπόθεση (H_0): Δεν υπάρχει διαφορά στους χρόνους ανακούφισης των δύο αναλγητικών φαρμάκων



Εισαγωγή δεδομένων και εκτέλεση του Independent Samples T Test

- Για να ελέγξουμε εάν υπάρχει διαφορά μεταξύ των δύο φαρμάκων, θα εφαρμόσουμε το **t-test** καθορίζοντας την στάθμη σημαντικότητας (significance level ή P-value).
- Έστω ότι θέλουμε να υπολογίσουμε το 95% δ. ε. της διαφοράς των μέσων τιμών του "1" και "2".
- Από το μενού επιλογών, διαλέγουμε

Analyze-> Compare Means-> Independent Samples T-test.

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a data table with columns 'drug' and 'time'. The 'Independent-Samples T Test' dialog box is open, with 'time' selected as the Test Variable(s) and 'drug(1 2)' as the Grouping Variable. The 'Define Groups...' dialog box is also open, showing 'Use specified values' selected, with Group 1 set to 1 and Group 2 set to 2. The 'Independent-Samples T Test: Options' dialog box is open, showing 'Confidence Interval Percentage' set to 95% and 'Exclude cases analysis by analysis' selected.

	drug	time	var	var	var	var	var	var
1	1	7.50						
2	1	3.60						
3	1	8.90						
4	1	6.40						
5	1	9.50						
6	1	7.40						
7	1	10.10						
8	1	4.70						
9	2	3.80						
10	2	2.40						
11	2	4.70						
12	2	5.90						
13	2	2.10						
14	2	3.20						
15	2	4.50						
16	2	3.80						

→ T-Test

[DataSet0]

Τα δύο φάρμακα έχουν ίδια διακύμανση, καθώς το p-value (Sig.) > 0.05 στο Levene's Test. Τότε $P=0.002$ δηλ. $P<0.05$ οπότε διαφέρουν.

Group Statistics

	drug	N	Mean	Std. Deviation	Std. Error Mean
time	1	8	7.2625	2.28469	.80776
	2	8	3.8000	1.24900	.44159

Independent Samples Test

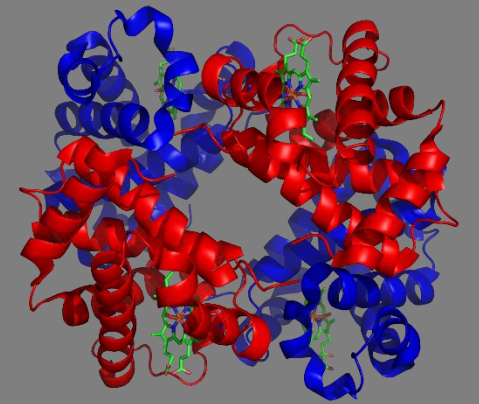
Levene's Test for Equality of Variances

t-test for Equality of Means

95% Confidence Interval of the Difference

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
time	Equal variances assumed	2.592	.130	3.761	14	.002	3.46250	.92059	1.48804	5.43696
	Equal variances not assumed			3.761	10.841	.003	3.46250	.92059	1.43267	5.49233

Μεταβολή της αιμοσφαιρίνης μετά την χορήγηση
δύο ΕΡΟ φαρμάκων





Μεταβολή της αιμοσφαιρίνης μετά την χορήγηση δύο EPO φαρμάκων

Test	Reference
0.38	0.80
0.37	0.70
0.24	0.39
0.08	0.06
0.04	0.49
0.03	0.07
0.07	0.63
0.19	0.83
0.35	0.62
0.38	0.95
0.27	0.92
0.20	0.81

Έστω ότι θέλουμε να συγκρίνουμε την μεταβολή της αιμοσφαιρίνης δύο EPO (ερυθροποιητίνη) φαρμάκων, Test (T) και Reference (R). 12 ασθενείς λαμβάνουν το Test και 12 το Reference.

Μηδενική υπόθεση (H_0): Δεν υπάρχει διαφορά στη μέση τιμή των τιμών της αιμοσφαιρίνης μεταξύ του Test και Reference φαρμάκου

Σημειώστε πως τα δεδομένα δεν παρουσιάζουν ατομική αντιστοιχία.



Εισαγωγή δεδομένων και ρύθμιση μεταβλητών

Εισάγουμε στο **Data View** τα δεδομένα και κάνουμε τις ρυθμίσεις των μεταβλητών μας στο **Variable View** όπως παρακάτω (data1.sav)

	drug	hgb
1	1	.38
2	1	.37
3	1	.24
4	1	.08
5	1	.04
6	1	.03
7	1	.07
8	1	.19
9	1	.35
10	1	.38
11	1	.27
12	1	.20
13	2	.80
14	2	.70
15	2	.39
16	2	.06

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	drug	Numeric	8	0		{1, Test}...	None	8	Right	Nominal	Input
2	hgb	Numeric	8	2		None	None	8	Right	Scale	Input

Value Labels

Value Labels

Value:

Label:

1 = "Test"
2 = "Reference"

Add Change Remove

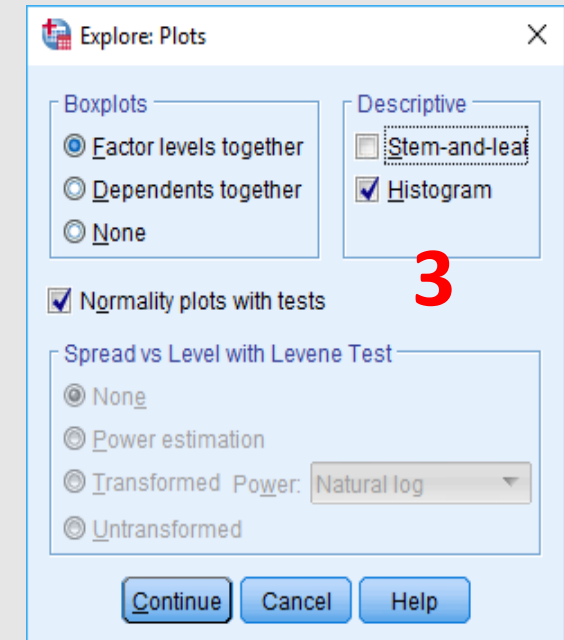
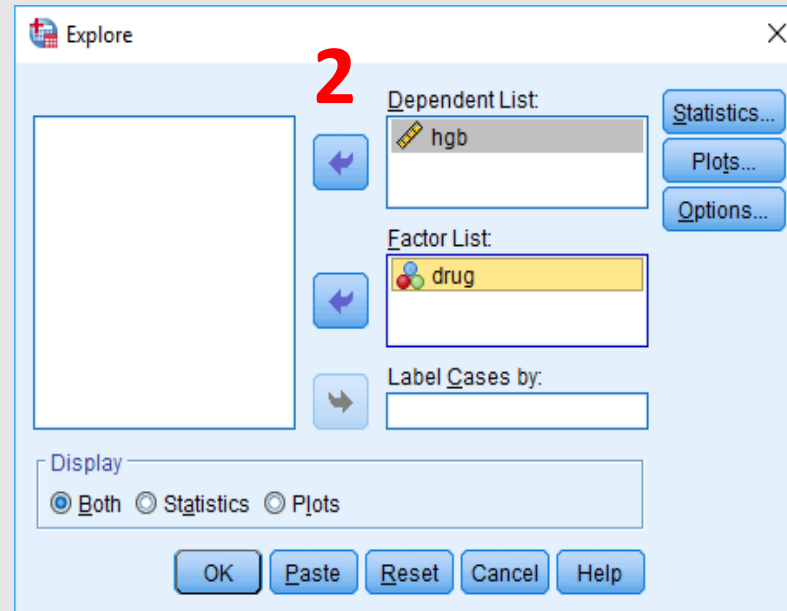
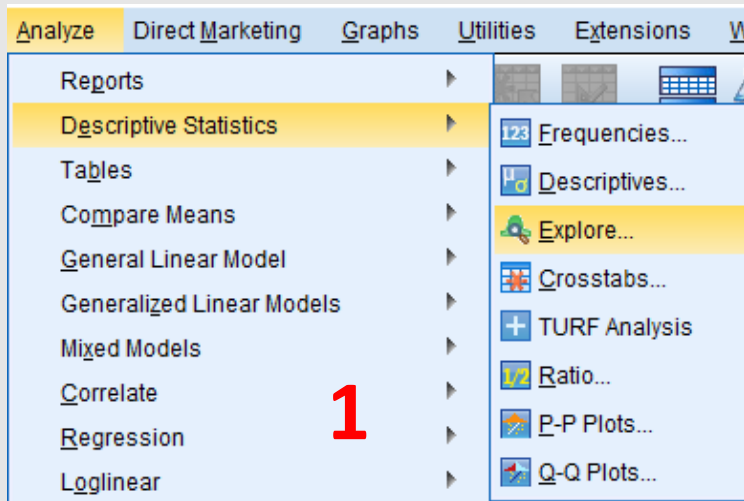
Spelling...

OK Cancel Help



Περιγραφική Στατιστική

Αρχικά εμφανίζουμε κάποια περιγραφικά στατιστικά μέσω του μενού **Analyze -> Descriptive -> Statistics -> Explore**. Στο παράθυρο (2) σέρνομαι τη μεταβλητή **hgb** από το αριστερό πλαίσιο στο πλαίσιο **Dependent List** και την μεταβλητή **drug** στο πλαίσιο **Factor List**.



Στο παράθυρο (3), επιλέγουμε **Histogram** και **Normality plots with tests**. Πατάμε Continue και στην συνέχεια OK.

Case Processing Summary

		Valid		Missing		Total	
drug		N	Percent	N	Percent	N	Percent
hgb	Test	12	100.0%	0	0.0%	12	100.0%
	Reference	12	100.0%	0	0.0%	12	100.0%

Tests of Normality

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
drug		Statistic	df	Sig.	Statistic	df	Sig.
hgb	Test	.175	12	.200 [*]	.886	12	.106
	Reference	.185	12	.200 [*]	.887	12	.109

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

a. Lilliefors Significance Correction

Από το test **Shapiro – Wilk** ($n < 50$) συμπεραίνουμε πως οι τιμές της αιμοσφαιρίνης του test φαρμάκου ακολουθούν κανονική κατανομή καθώς το $p > 0.05$ ($p = 0.106$). Το ίδιο ισχύει και το reference φάρμακο ($p > 0.05$).

Σημείωση: Το test κανονικότητας **Kolmogorov-Smirnov** ελέγχεται όταν ο αριθμός των ασθενών είναι μεγαλύτερος από 50.

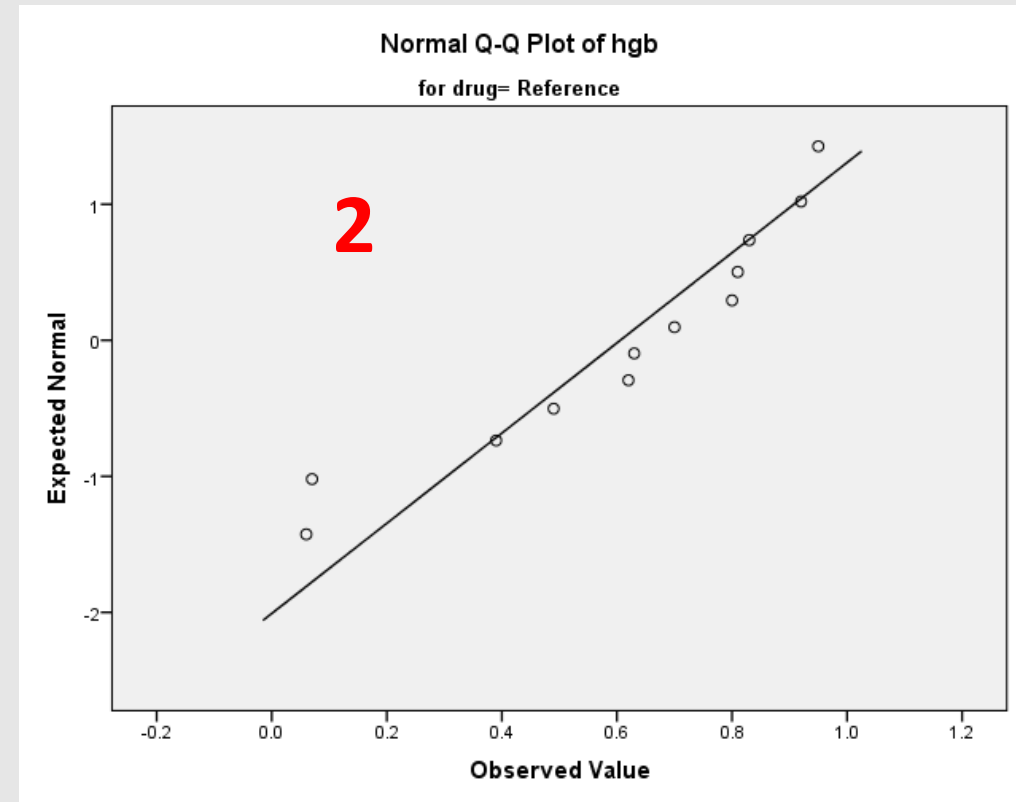
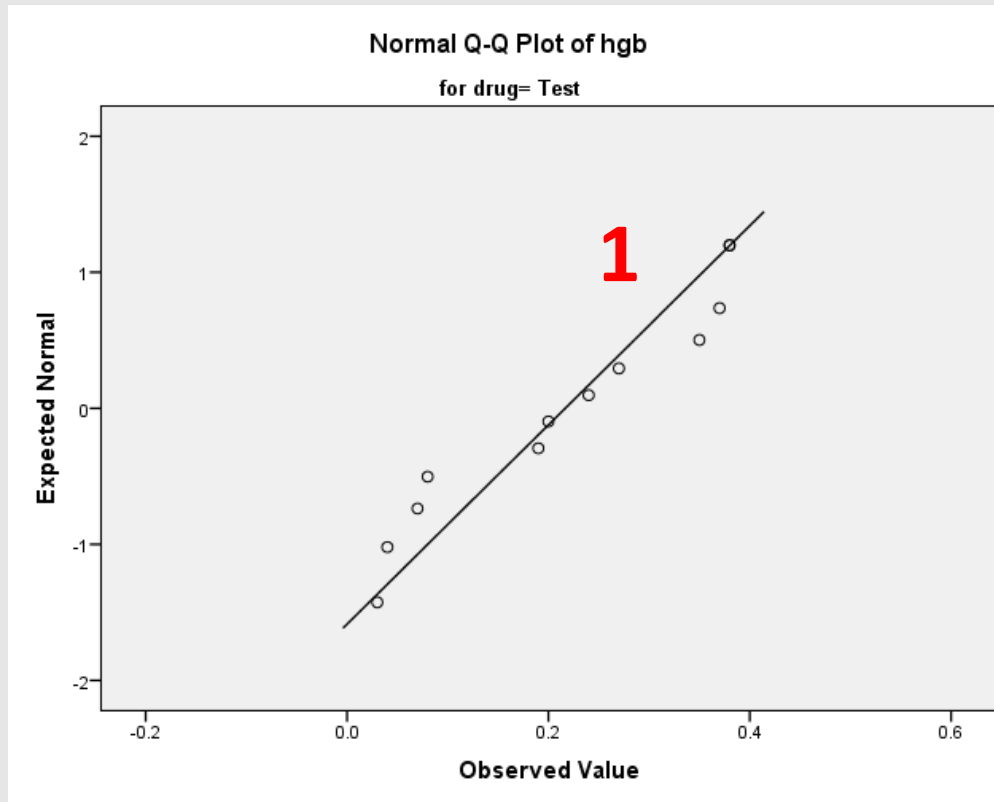
Descriptives

drug				Statistic	Std. Error		
hgb	Test	Mean		.2167	.03943		
		95% Confidence Interval for Mean		Lower Bound	.1299		
				Upper Bound	.3035		
		5% Trimmed Mean		.2180			
		Median		.2200			
		Variance		.019			
		Std. Deviation		.13660			
		Minimum		.03			
		Maximum		.38			
		Range		.35			
		Interquartile Range		.29			
		Skewness		-.124	.637		
		Kurtosis		-1.634	1.232		
		Reference		Mean		.6058	.08710
				95% Confidence Interval for Mean		Lower Bound	.4141
Upper Bound	.7975						
5% Trimmed Mean				.6170			
Median				.6650			
Variance				.091			
Std. Deviation				.30174			
Minimum				.06			
Maximum				.95			
Range				.89			
Interquartile Range				.41			
Skewness				-.907	.637		
Kurtosis				-.173	1.232		



Έλεγχος κανονικότητας με Q-Q Plots

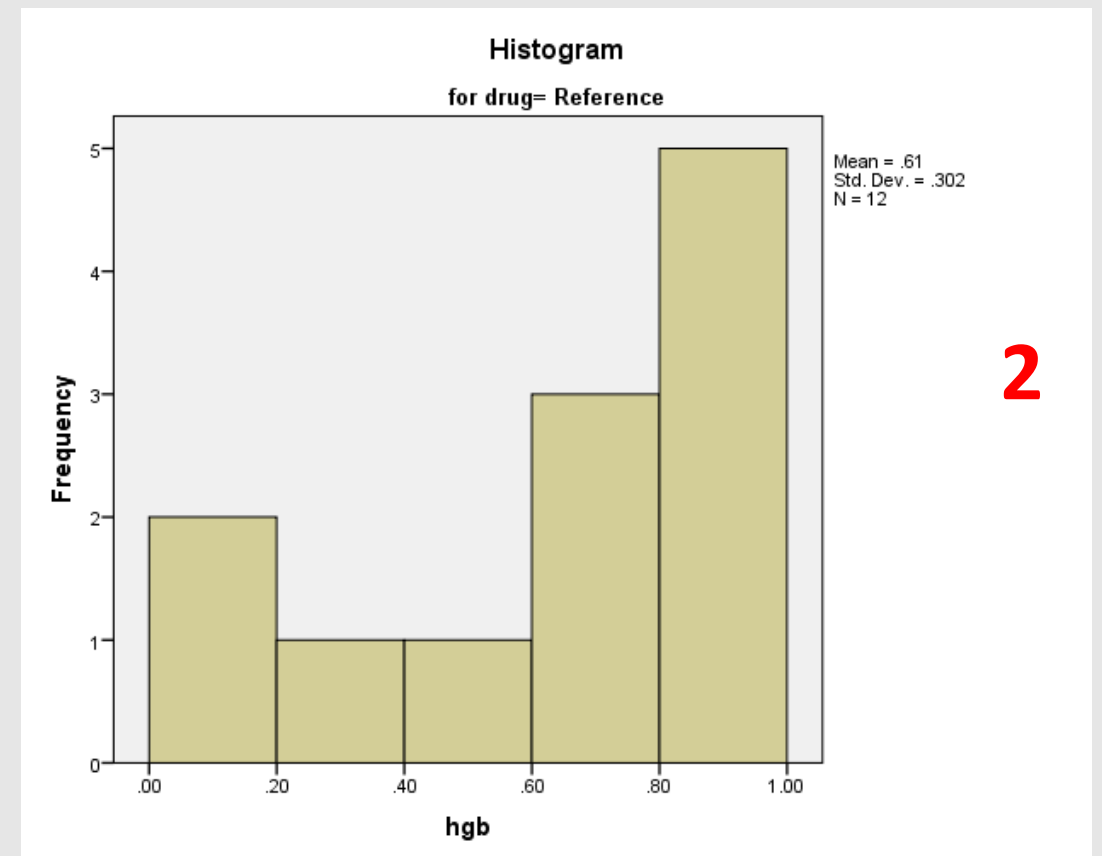
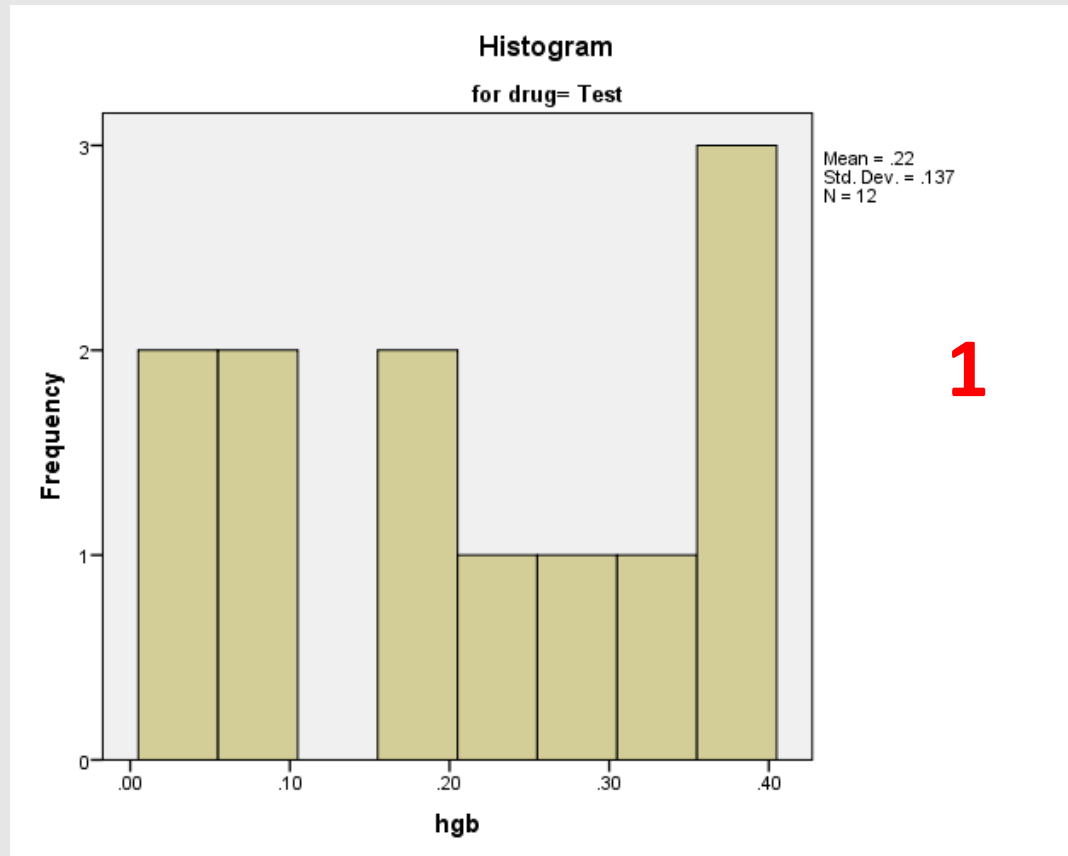
Στις εικόνες (1) και (2) εμφανίζονται τα διαγράμματα Q-Q των τιμών της αιμοσφαιρίνης των ασθενών για το Test και Reference φάρμακο, από όπου φαίνεται πως οι τιμές της αιμοσφαιρίνης ακολουθούν σε καλή προσέγγιση κανονική κατανομή





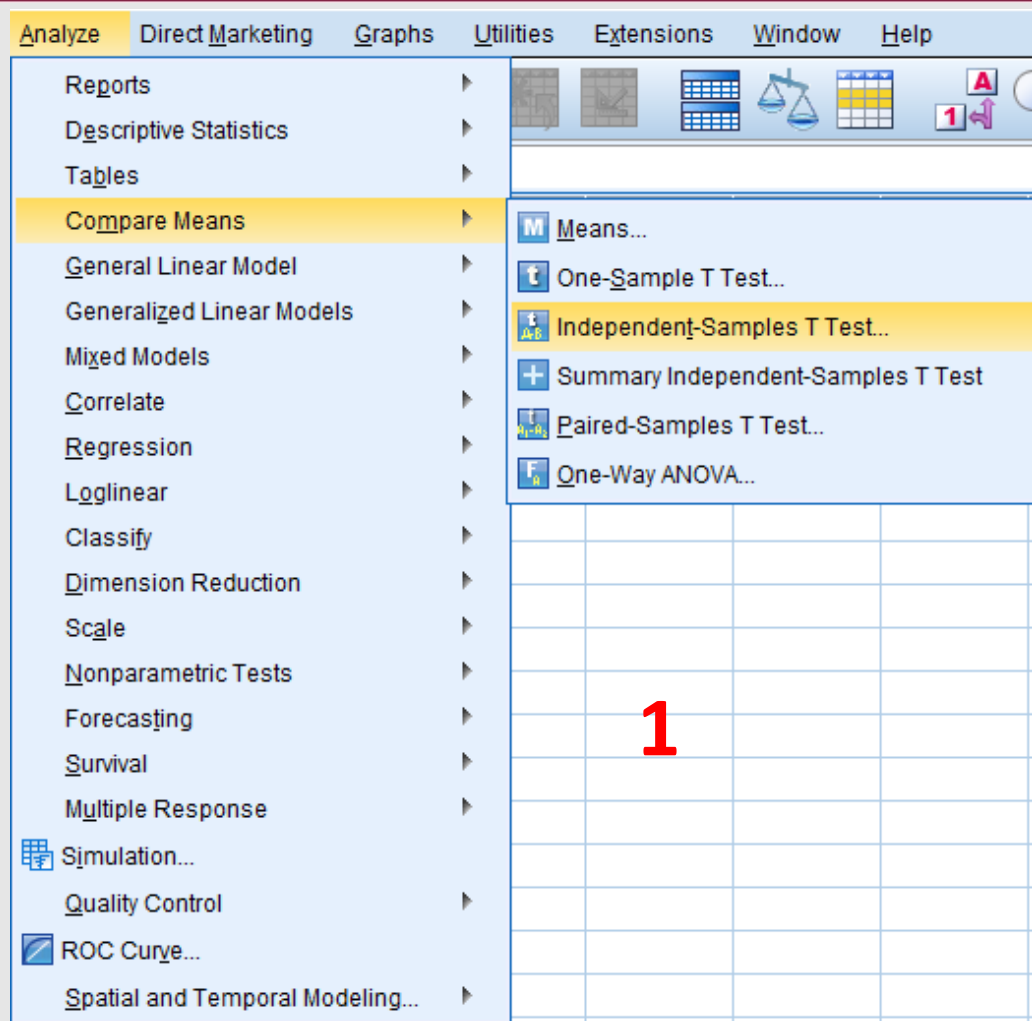
Έλεγχος κανονικότητας με Ιστογράμματα

Στις εικόνες (1) και (2) εμφανίζονται τα ιστογράμματα των τιμών της αιμοσφαιρίνης των δύο ΕΡΟ φαρμάκων (Test και Reference).

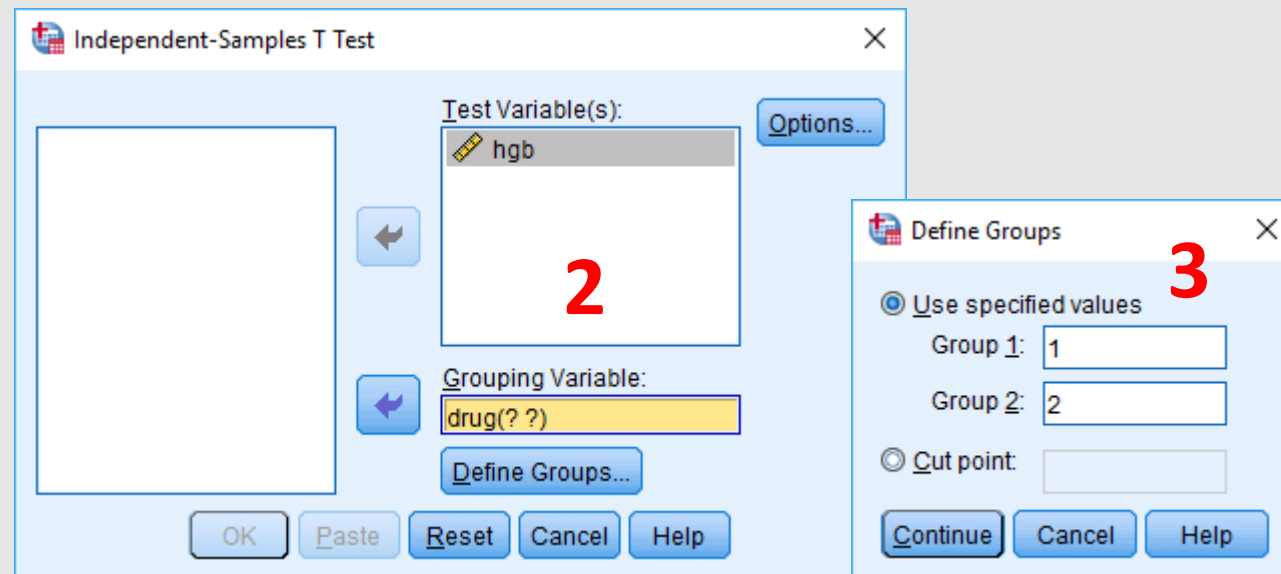




Εκτέλεση



1. Για την ανάλυση των δεδομένων, επιλέγουμε από το μενού **Analyze -> Compare Means -> Independent-Samples T Test...** (1)



2. Στο παράθυρο (2) σέρνουμε την μεταβλητή **hgb** στο πλαίσιο **Test Variable(s)** και την μεταβλητή **drug** στο πεδίο **Grouping Variable**; ορίζουμε τα δύο group στο παράθυρο (3) και πατάμε **OK**

Group Statistics

drug		N	Mean	Std. Deviation	Std. Error Mean
hgb	Test	12	.2167	.13660	.03943
	Reference	12	.6058	.30174	.08710

Τα δύο φάρμακα δεν έχουν κατά προσέγγιση την ίδια διακύμανση, καθώς το p - value (Sig.) < 0.05 στο Levene's Test. Τότε $P=0.001$ δηλ. $P<0.05$ οπότε η διαφορά στις μέσες τιμές της αιμοσφαιρίνης είναι στατιστικά σημαντική.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
hgb	Equal variances assumed	5.019	.035	-4.070	22	.001	-.38917	.09561	-.58746	-.19087
	Equal variances not assumed			-4.070	15.327	.001	-.38917	.09561	-.59259	-.18575

- Υπάρχει διαφορά στη μέση τιμή (-0.389) της αιμοσφαιρίνης μεταξύ του Test και Reference φαρμάκου. Η διαφορά αυτή είναι στατιστικά σημαντική, με επίπεδο στατιστικής σημαντικότητας $P < 0.05$
- Η μέση τιμή της διαφοράς αυτής εκτιμήθηκε δειγματοληπτικά σε -0.389 (Mean = -0.389)
- Με εμπιστοσύνη 95%, η πραγματική μέση διαφορά βρίσκεται μεταξύ των ορίων -0.5926 και -0.1858.



Πρακτική άσκηση

Το βάρος των νεογέννητων παιδιών που γεννήθηκαν από 15 μη-καπνίστριες και από 14 καπνίστριες είναι:

Δείγμα 1 Μη-καπνίστριες	Δείγμα 2 Καπνίστριες
3.99	3.18
3.79	2.84
3.60	2.90
3.73	3.27
3.21	3.85
3.60	3.52
4.08	3.23
3.61	2.76
3.83	3.60
3.31	3.75
4.13	3.59
3.26	3.63
3.54	2.38
3.51	2.34
2.71	

Υπάρχει διαφορά στο βάρος των παιδιών μεταξύ των δύο ομάδων;