

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά –
Χρηματοοικονομικά Μαθηματικά

2015

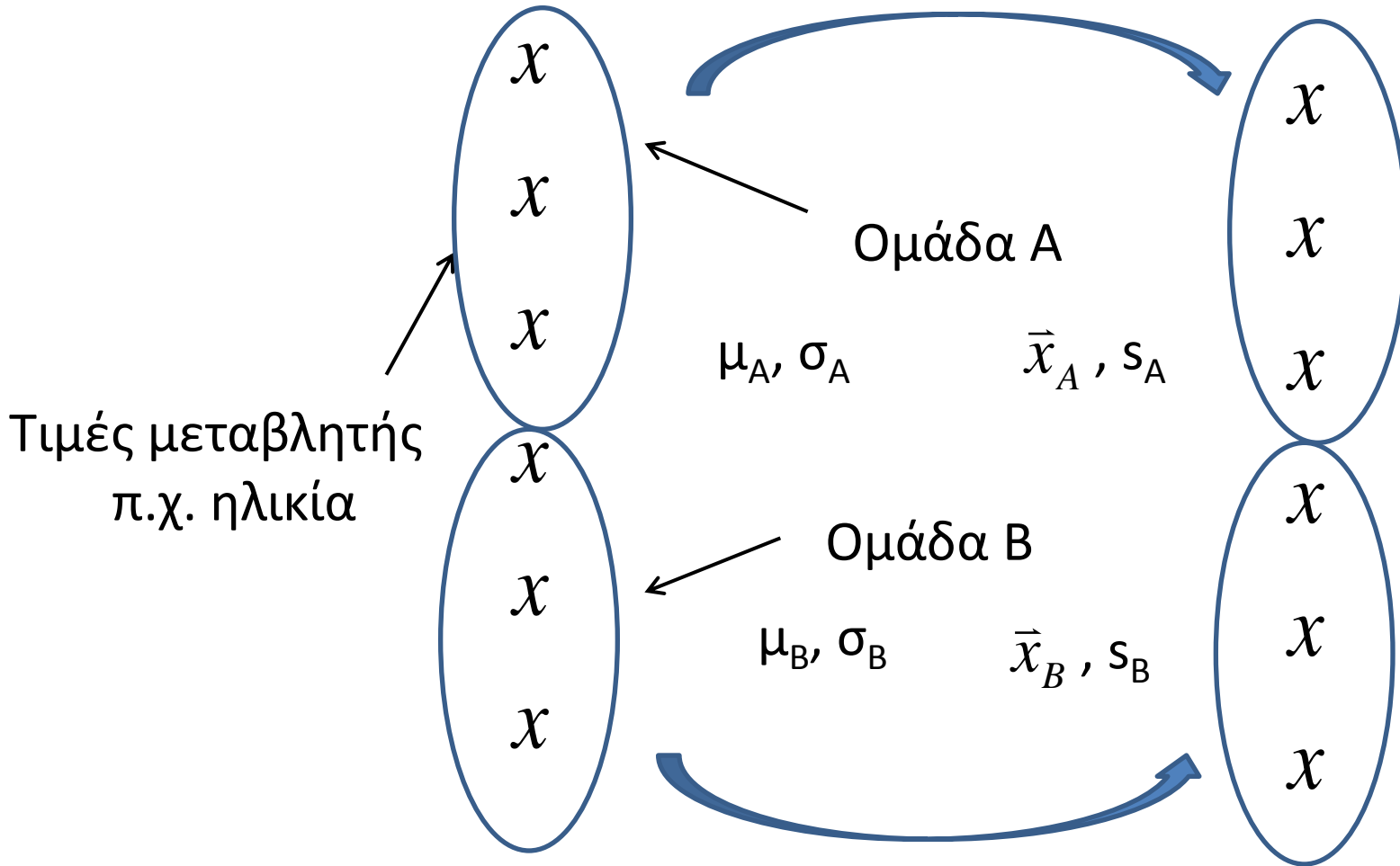
Ελεγχος μεσών τιμών μέσω διακύμανσης

- **Ανεξάρτητα δείγματα:**

Αφορά δύο κανονικούς πληθυσμούς με παραμέτρους (μ_A, σ^2_A) και (μ_B, σ^2_B) αντίστοιχα, με μέσους μ_A και μ_B και τυπικές αποκλίσεις σ_X και σ_Y αντίστοιχα.

Από τον πρώτο πληθυσμό (A) λαμβάνεται δείγμα n_A με αντίστοιχες μέσες τιμές \bar{x}_A και τυπικές αποκλίσεις s_A και από τον δεύτερο πληθυσμό (B) δείγμα n_B με αντίστοιχες μέσες τιμές \bar{x}_B και τυπικές αποκλίσεις s_B .

ΕΛΕΓΧΟΣ ΜΕΣΩΝ ΤΙΜΩΝ ΜΕΣΩ ΔΙΑΚΥΜΑΝΣΗΣ



Q: Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο ηλικιακών ομάδων?

Έλεγχος μεσών τιμών μέσω διακύμανσης

Ο έλεγχος γίνεται με βάση την μηδενική υπόθεση

$$H_0: \mu_A = \mu_B$$

έναντι της εναλλακτικής

$$H_1: \mu_A \neq \mu_B$$

Υπάρχουν δύο περιπτώσεις:

1. Πληθυσμιακές διακυμάνσεις ίσες
2. Πληθυσμιακές διακυμάνσεις άνισες

ΕΛΕΓΧΟΣ ΜΕΣΩΝ ΤΙΜΩΝ ΜΕΣΩ ΔΙΑΚΥΜΑΝΣΗΣ

- Πληθυσμιακές διακυμάνσεις ίσες

Ως εκτιμήτρια συνάρτηση της κοινής πληθυσμιακής διακύμανσης χρησιμοποιούμε την σταθμισμένη μέση τιμή των δειγματικών διακυμάνσεων. Υπολογίζεται με βάση τους βαθμούς ελευθερίας της κάθε δειγματικής διακύμανσης. Άρα δίνεται από τον τύπο:

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

ΕΛΕΓΧΟΣ ΜΕΣΩΝ ΤΙΜΩΝ ΜΕΣΩ ΔΙΑΚΥΜΑΝΣΗΣ

- Θεωρώντας ότι ισχύει η μηδενική υπόθεση έχουμε της ελεγκοσυνάρτηση t με

$$t^* = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}} \rightarrow t_{n_A+n_B-2}$$

- Αν η τιμή της $t^* > t$ -κατανομή τότε η ισότητα των δύο μέσων απορρίπτεται. Αντίθετα γίνεται αποδεκτή

Έλεγχος μέσων τιμών μέσω διακύμανσης

- Πληθυσμιακές διακυμάνσεις άνισες

Όταν οι πληθυσμιακές διακυμάνσεις είναι άνισες τότε η κάθε μία είναι δειγματική διακύμανση είναι εκτιμήτρια της αντίστοιχης πληθυσμιακής

Στην περίπτωση που ισχύει η μηδενική υπόθεση η ελεγκοσυνάρτηση γίνεται:

$$t^* = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \rightarrow t_n$$
$$n = \frac{\left[\left(\frac{s_A^2}{n_A} \right) + \left(\frac{s_B^2}{n_B} \right) \right]^2}{\frac{\left(\frac{s_A^2}{n_A} \right)^2}{n_A - 1} + \frac{\left(\frac{s_B^2}{n_B} \right)^2}{n_B - 1}}$$

όπου n δίνεται από την σχέση

Έλεγχος μέσων τιμών μέσω διακύμανσης

Βασικές προϋποθέσεις χρήσης της ελεγκοσυνάρτησης t

1. Το κάθε δείγμα μας είναι τυχαίο. Γενικότερα ο έλεγχος αυτός γίνεται από τον τρόπο λήψης του δείγματος και ο έλεγχος μπορεί να γίνει και με στατιστικό τεστ.
2. Τα δύο δείγματα είναι ανεξάρτητα μεταξύ τους. Η ανεξαρτησία των δύο δειγμάτων βασίζεται κυρίως στον τρόπο επιλογής των δειγμάτων. Ανεξαρτησία δειγμάτων έχουμε όταν μετράμε το ίδιο χαρακτηριστικό σε διαφορετικά πειράματα

Έλεγχος μέσων τιμών μέσω διακύμανσης

3. Η κατανομή του πληθυσμού, από τον οποίο προήλθε το κάθε δείγμα, είναι η κανονική κατανομή. Ο έλεγχος κανονικότητας, για κάθε δείγμα χωριστά, γίνεται γραφικά και με στατιστικά
4. Δεν υπάρχουν ακραίες παρατηρήσεις. Ο έλεγχος υπάρξης ακραίων τιμών, για κάθε δείγμα χωριστά, γίνεται κυρίως γραφικά με τη βοήθεια του θηκογράμματος (boxplot)

Έλεγχος μεσών τιμών μέσω διακύμανσης

- Έλεγχος ισότητας των διακυμάνσεων

Ο έλεγχος της ισότητας των διακυμάνσεων με βάση την μηδενική υπόθεση

$$H_0: \sigma^2_A = \sigma^2_B$$

έναντι της εναλλακτικής

$$H_1: \sigma^2_A \neq \sigma^2_B$$

πραγματοποιείται με το test του Levene

test του Levene

Ελέγχουμε την ισότητα των διακυμάνσεων μιας μεταβλητής με διαφορετικούς παράγοντες.

Το τεστ βασίζεται στην ανάλυση διακύμανσης προς ένα παράγοντα που πραγματοποιείται όχι με βάση τις αρχικές τιμές x_{ij} με $i=1,2,\dots,n_j$ και $j=1,2,\dots,k$ μιας μεταβλητής αλλά με βάση τις τροποποιημένες μεταβλητές y_{ij} ύστερα από μετασχηματισμό τους.

Ο μετασχηματισμός αυτός δίνεται από την παρακάτω σχέση

$$y_{ij} = |x_{ij} - \bar{x}_j|$$

όπου n_j ο αριθμός των παρατηρήσεων της ομάδας j και k ο αριθμός των ομάδων

test του Levene

Όταν δημιουργηθούν οι νέες τιμές y_{ij} πραγματοποιείται έλεγχος υποθέσεων ανάλυσης διακύμανσης για ένα παράγοντα με

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

και εναλλακτική

H_1 : τουλάχιστον δύο από τις $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ να διαφέρουν

test του Levene

Το αντίστοιχο τεστ για το έλεγχο διακυμάνσεων (ομοιογένειας) δίνεται από τον τύπο

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k N_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_{i.})^2}$$

Συνολικός μέσος

$$z_{ij} = |x_{ij} - \bar{x}_j|$$

Μέσοι των γραμμών

Απορρίπτουμε την υπόθεση αν ισχύει

$$W > F_{k-1, N-k, \alpha}$$

Έλεγχος μέσων τιμών μέσω διακύμανσης

Η πιθανότητα να ισχύει η μηδενική υπόθεση στο test του Levene, αποτελεί το κριτήριο σχετικά με τον τρόπο που θα γίνει ο έλεγχος σημαντικότητας των δύο μέσων. Αν είναι μεγαλύτερος του 0.05 ο έλεγχος γίνεται με βάση τον τύπο

$$t^* = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}} \quad \text{Ίσες πληθυσμιακές διακυμάνσεις}$$

Αν είναι μικρότερος του 0.05 ο έλεγχος γίνεται με βάση τον τύπο

$$t^* = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \quad \text{Άνισες πληθυσμιακές διακυμάνσεις}$$

Παράδειγμα

Στον παρακάτω πίνακα δίνονται οι τιμές της αμόλυβδης βενζίνης (σε χρηματικές μονάδες) 20 τυχαία επιλεγμένων πρατηρίων υγρών καυσίμων δύο διαφορετικών εταιρειών (Εταιρεία 1 και Εταιρεία 2).

Επιθυμούμε να ελέγξουμε, με επίπεδο σημαντικότητας 5%, αν η μέση τιμή μ_1 του πληθυσμού των πρατηρίων υγρών καυσίμων της εταιρείας 1 είναι ίση με τη μέση τιμή μ_2 του πληθυσμού των πρατηρίων υγρών καυσίμων της εταιρείας 2.

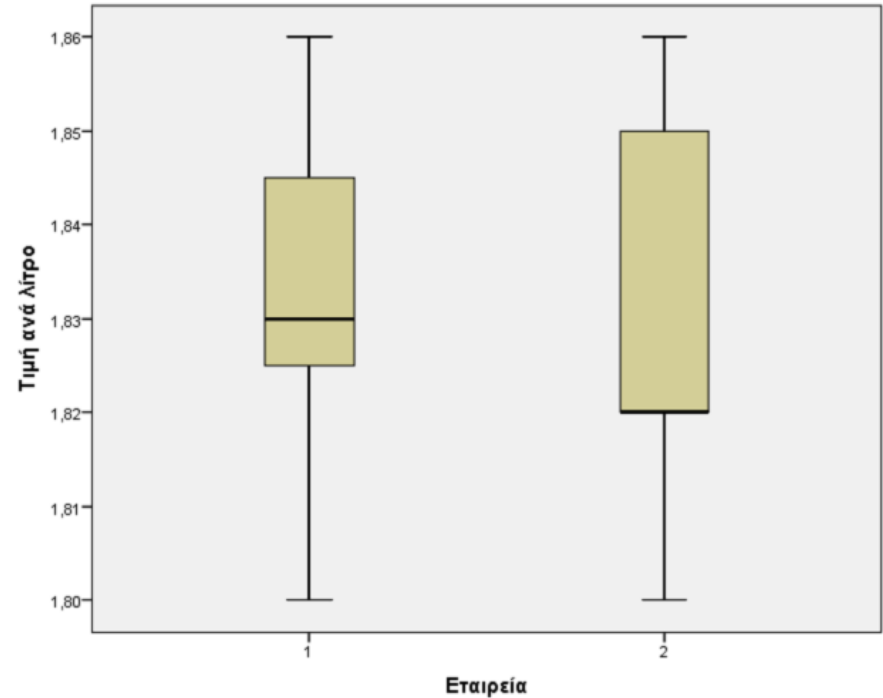
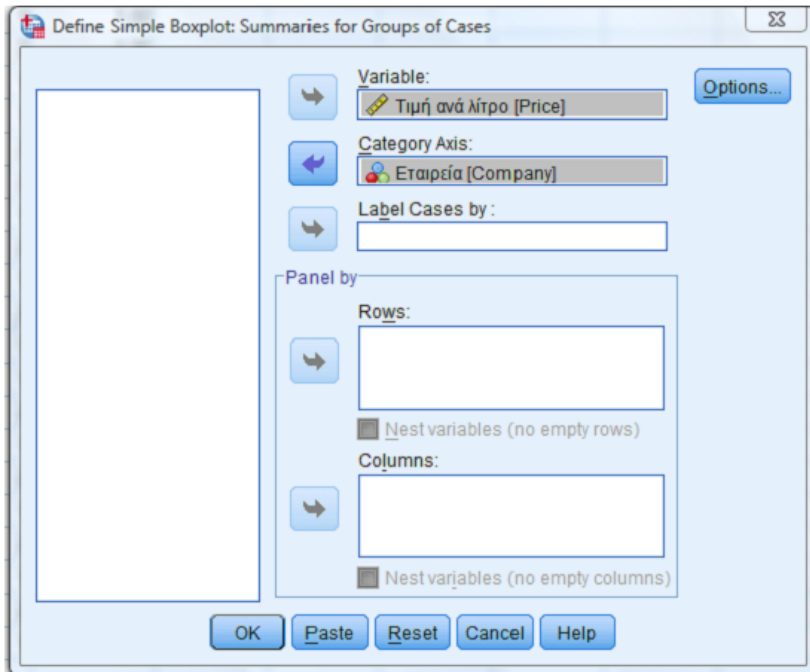
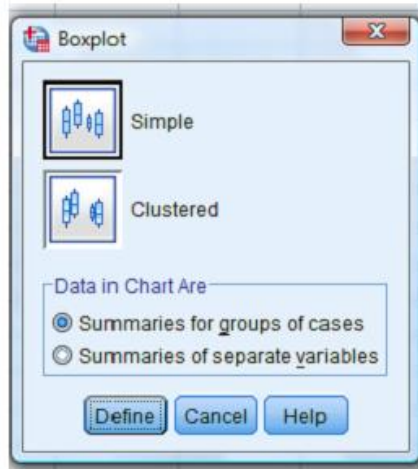
Εταιρεία	Τιμή/ανά λίτρο βενζίνης
1	1,82
2	1,85
1	1,81
1	1,83
2	1,86
2	1,82
1	1,84
2	1,80
1	1,83
2	1,82
1	1,85
1	1,83
2	1,82
1	1,86
2	1,82
1	1,85
2	1,83
1	1,80
2	1,85
1	1,83

Παράδειγμα

Αρχικά θα ελέγξουμε, αν ισχύουν οι προϋποθέσεις

1. Το κάθε δείγμα μας είναι τυχαίο.
 2. Τα δύο δείγματα είναι ανεξάρτητα μεταξύ τους.
 3. Δεν υπάρχουν ακραίες παρατηρήσεις. Για κάθε δείγμα ξεχωριστά, θα ελέγξουμε την ύπαρξη ακραίων παρατηρήσεων με τη βοήθεια θηκογράμματος (boxplot).
- Για να κατασκευάσουμε τα θηκογράμματα για κάθε δείγμα ξεχωριστά ακολουθούμε την παρακάτω διαδρομή:).

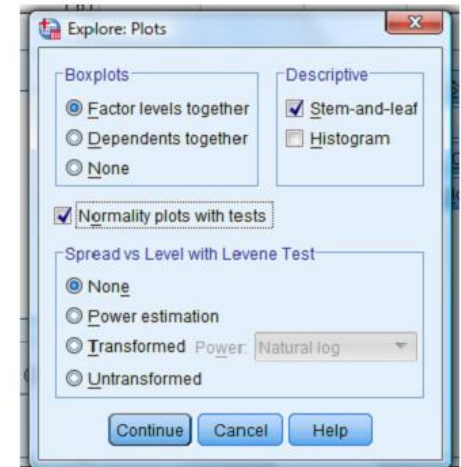
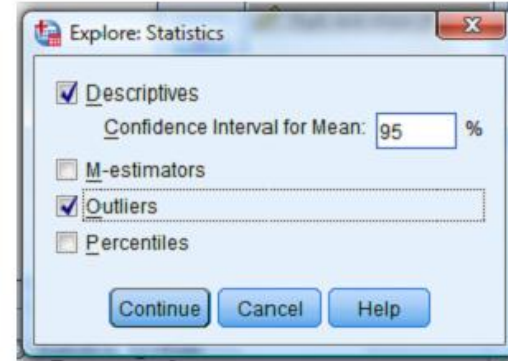
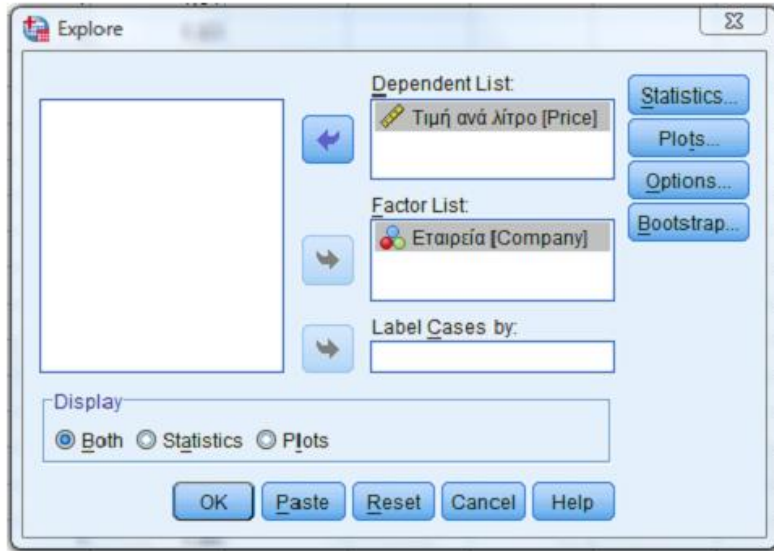
Παράδειγμα



Όχι ακραίες τιμές

Παράδειγμα

4. Η κατανομή του πληθυσμού, από τον οποίο προήλθε το κάθε δείγμα, είναι η κανονική κατανομή.



	Εταιρεία	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Τιμή ανά λίτρο	1	,187	11	,200*	,959	11	,756
	2	,253	9	,101	,899	9	,249

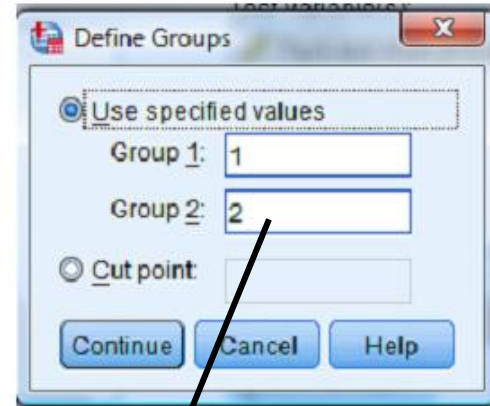
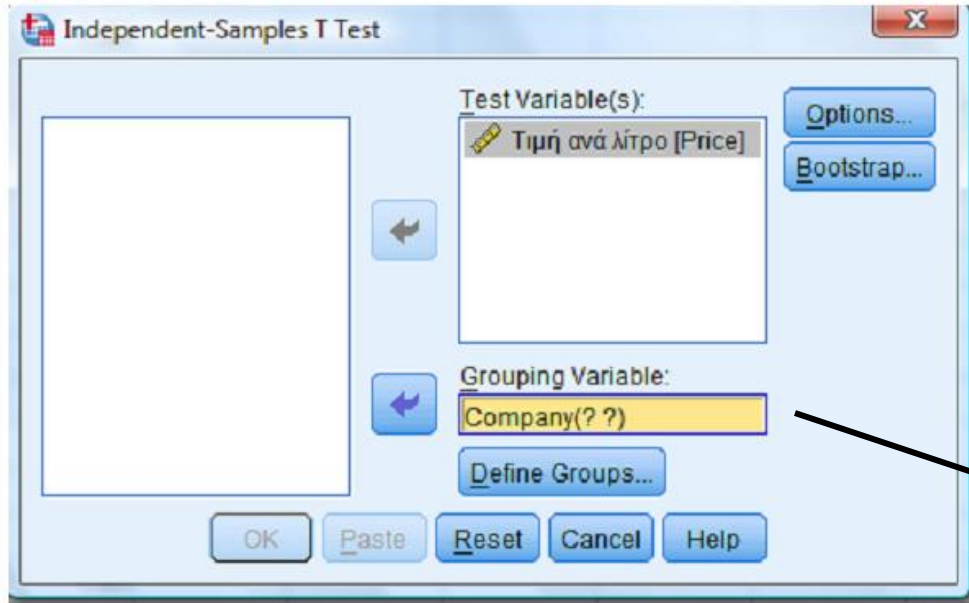
*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

$p > 0.05$ – κανονικότητα

Μικρό δείγμα

Παράδειγμα

$p > 0.05$ – κανονικότητα



κατηγορίες

Έλεγχος μέσων

Εταιρεία	N	Mean	Std. Deviation	Std. Error Mean
Τιμή ανά λίτρο 1	11	1,8318	,01779	,00536
2	9	1,8300	,01936	,00645

	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
Τιμή ανά λίτρο	,235	,634	,219	18	,829	,00182	,00832	-,01566	,01929
			,217	16,548	,831	,00182	,00839	-,01592	,01956

$p > 0.05$ – ίσες διακυμάνσεις

$p > 0.05$ – μέσοι ίσοι