



ΣΥΓΚΡΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΔΥΟ ΠΛΗΘΥΣΜΩΝ Κλωνάρης Στάθης

ΠΜΣ: Οργάνωση & Διοίκηση Επιχειρήσεων Τροφίμων και Γεωργίας

Σύγκριση δύο πληθυσμών

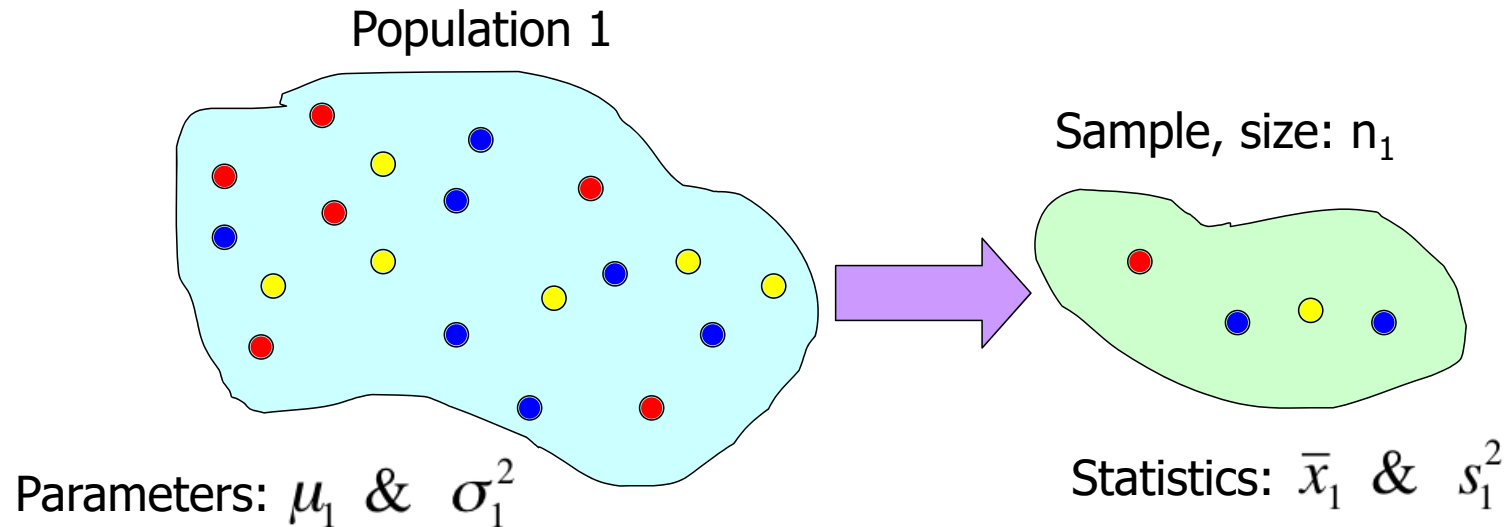
Μέχρι τώρα ασχοληθήκαμε με τις τεχνικές εκτίμησης παραμέτρων για ένα πληθυσμό όπως:
τον Μέσο μ
και το ποσοστό p

Θα συνεχίσουμε με τις μεθόδους σύγκρισης **δύο πληθυσμών**, και οι παράμετροι που θα μας απασχολήσουν θα είναι:

- Η διαφορά μεταξύ δύο μέσων.
- Η διαφορά των ποσοστών.

Διαφορά μεταξύ δύο μέσων

Για να ελέγξουμε ή να εκτιμήσουμε την διαφορά μεταξύ των μέσων δύο πληθυσμών, πρέπει να επιλέξουμε τυχαία δείγματα από τους δύο πληθυσμούς. Αρχικά θα εξετάσουμε την περίπτωση όπου τα δείγματα είναι ανεξάρτητα.



(με το ίδιο τρόπο θεωρούμε $\mu_2, \sigma_2^2, n_2, \bar{x}_2, \& s_2^2$ τον πληθυσμό 2)

Διαφορά μεταξύ δύο μέσων

Επειδή συγκρίνουμε τους μέσους δύο πληθυσμών, χρησιμοποιούμε την στατιστική,

$$\bar{x}_1 - \bar{x}_2$$

η οποία είναι αμερόληπτη και συνεπής εκτιμήτρια της διαφοράς $\mu_1 - \mu_2$.

Κατανομή δειγματοληψίας της διαφοράς δύο μέσων

$\bar{x}_1 - \bar{x}_2$ ακολουθεί κανονική κατανομή εφόσον οι δύο πληθυσμοί έχουν κανονική κατανομή, και κατά προσέγγιση κανονική κατανομή ή όταν οι δύο πληθυσμοί δεν έχουν κανονική κατανομή αλλά τα μεγέθη των δειγμάτων είναι μεγάλα. ($n_1, n_2 > 30$).

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Και τυπικό σφάλμα $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Κατανομή δειγματοληψίας της διαφοράς δύο μέσων

Έτσι η μεταβλητή

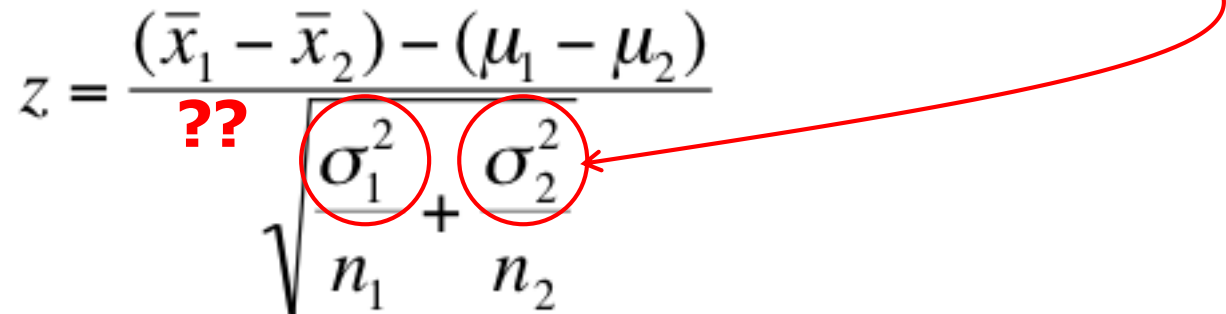
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Είναι μια κανονική (ή κατά προσέγγιση κανονική) τυχαία μεταβλητή.

Μπορούμε να χρησιμοποιήσουμε αυτή την μεταβλητή για τον έλεγχο και τον εκτιμητή του διαστήματος εμπιστοσύνης της διαφοράς δύο μέσων $\mu_1 - \mu_2$.

Κατανομή δειγματοληψίας της διαφοράς δύο μέσων

...στην πράξη η στατιστική z χρησιμοποιείται σπάνια επειδή σχεδόν πάντοτε οι διασπορές των πληθυσμών είναι άγνωστες.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$


Αντ' αυτού χρησιμοποιούμε την στατιστική t (t-statistic). Διακρίνουμε δυο περιπτώσεις για τις άγνωστες διακυμάνσεις του πληθυσμού:

- i) όταν πιστεύουμε ότι είναι μεταξύ τους **ίσες** και
- ii) όταν είναι **άνισες**

Έλεγχος υπόθεσης για ίσες διασπορές

Υπολογίζουμε τον σταθμισμένο εκτιμητή διασποράς (*pooled variance estimator*)

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

...και τον χρησιμοποιούμε :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad \nu = n_1 + n_2 - 2$$

Βαθμοί ελευθερίας

Εκτιμητής διαστήματος εμπιστοσύνης για ίσες διακυμάνσεις

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad \nu = n_1 + n_2 - 2$$

Σταθμισμένος εκτιμητής διασποράς

Βαθμοί ελευθερίας

Έλεγχος υποθέσεων για άνισες διακυμάνσεις

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}, \quad \nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Ανάλογα ο εκτιμητής διαστήματος: *Βαθμοί ελευθερίας*

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}, \quad \nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Έλεγχος για ίσες διασπορές

Η υπόθεση που πρέπει να ελέγξουμε είναι:

$$H_0: \sigma_1^2 / \sigma_2^2 = 1$$

$$H_1: \sigma_1^2 / \sigma_2^2 \neq 1$$

Το κριτήριο ελέγχου είναι το $F = s_1^2 / s_2^2$, η οποία ακολουθεί F-κατανομή με $\nu_1 = n_1 - 1$ και $\nu_2 = n_2 - 1$ βαθμούς ελευθερίας και προϋπόθεση την κανονική κατανομή των δύο πληθυσμών.

Επειδή έχουμε δικατάληκτο έλεγχο, υπάρχει περιοχή απόρριψης και στα δύο άκρα της κατανομής (two-tails), συνεπώς το κριτήριο απόρριψης της μηδενικής υπόθεσης είναι

$$F < F_{1-\alpha/2, \nu_1, \nu_2} \quad \text{ή} \quad F > F_{\alpha/2, \nu_1, \nu_2}$$

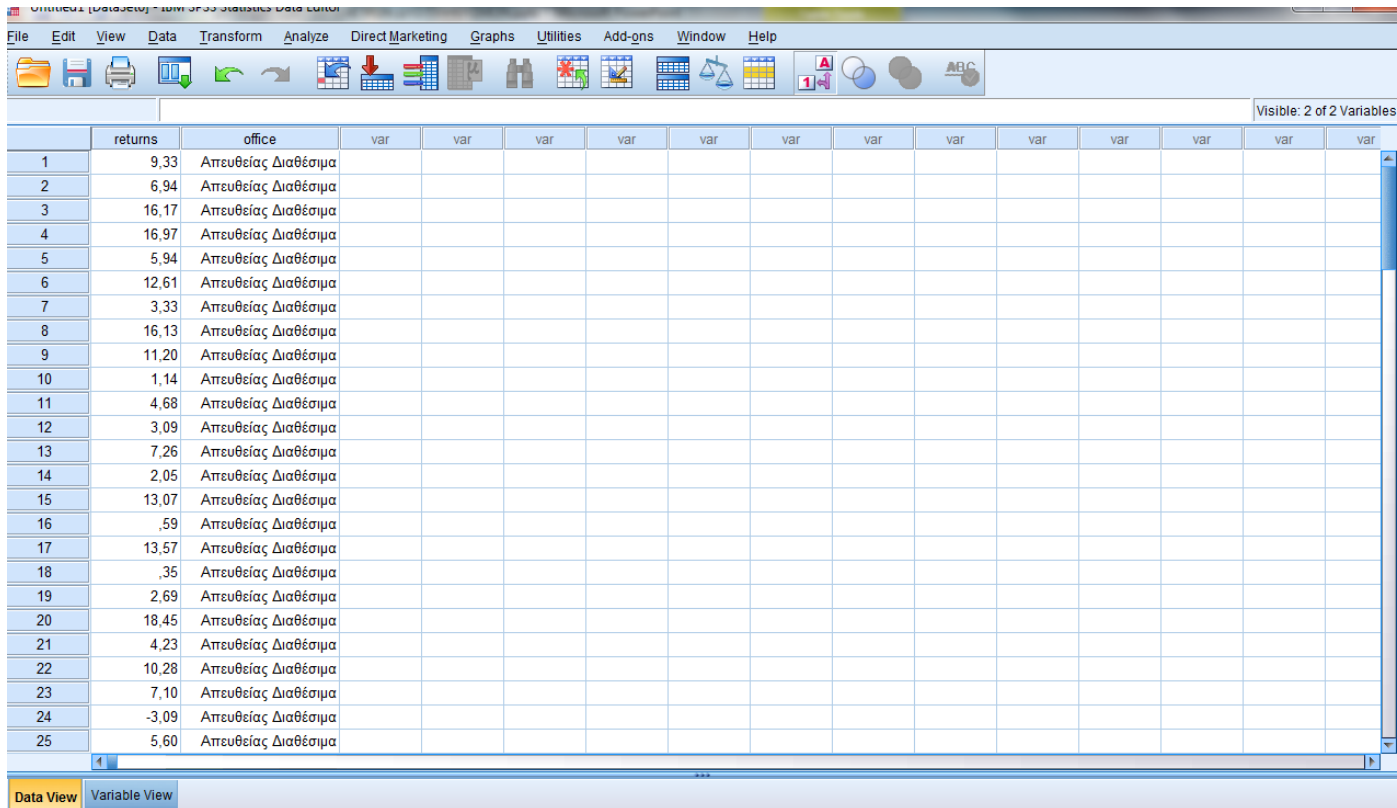
Παράδειγμα

Πολλοί επενδυτές αγοράζουν αμοιβαία κεφάλαια είτε απευθείας από τράπεζες ή μέσω χρηματιστών. Το ερώτημα που τίθεται είναι: κατά πόσο η απόδοση των αμοιβαίων κεφαλαίων που προτείνουν οι χρηματιστές είναι αρκετά υψηλότερη ώστε να δικαιολογούν την αμοιβή τους; Για το έλεγχο αυτού του ισχυρισμού επιλέχθηκε ένα τυχαίο δείγμα αμοιβαίων κεφαλαίων που είναι άμεσα διαθέσιμα στους επενδυτές από τις Τράπεζες και αμοιβαίων κεφαλαίων που διατίθενται μέσω χρηματιστών και καταγράφηκε η ετήσια απόδοσή τους.

Απευθείας διαθέσιμα									
9.33	6.94	16.17	16.97	5.94	12.61	3.33	16.13	11.2	1.14
4.68	3.09	7.26	2.05	13.07	0.59	13.57	0.35	2.69	18.45
4.23	10.28	7.1	-3.09	5.6	5.27	8.09	15.05	13.21	1.72
14.69	-2.97	10.37	-0.63	-0.15	0.27	4.59	6.38	-0.24	10.32
10.29	4.39	-2.06	7.66	10.83	14.48	4.8	13.12	-6.54	-1.06
Μέσω Χρηματιστών									
3.24	-6.76	12.8	11.1	2.73	-0.13	18.22	-0.8	-5.75	2.59
3.71	13.15	11.05	-3.12	8.94	2.74	4.07	5.6	-0.85	-0.28
16.4	6.39	-1.9	9.49	6.7	0.19	12.39	6.54	10.92	-2.15
4.36	-11.07	9.24	-2.67	8.97	1.87	-1.53	5.23	6.87	-1.69
9.43	8.31	-3.99	-4.44	8.63	7.06	1.57	-8.44	-5.72	6.95

Μπορούμε από τα δεδομένα να συμπεράνουμε με στάθμη σημαντικότητας 5% ότι τα αμοιβαία κεφάλαια που αγοράζονται απευθείας από τους επενδυτές έχουν υψηλότερη απόδοση;

Δύο ανεξάρτητα δείγματα

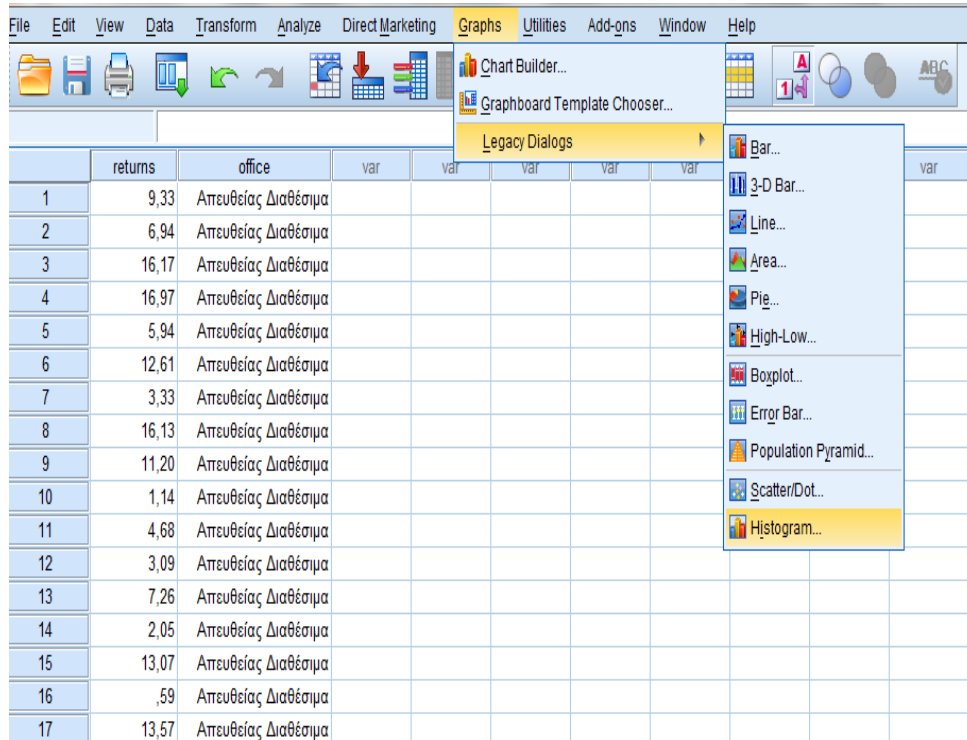


SPSS Statistics Data Editor - Visible: 2 of 2 Variables

	returns	office	var	var	var	var	var	var	var	var	var	var	var	var
1	9,33	Απευθείας Διαθέσιμα												
2	6,94	Απευθείας Διαθέσιμα												
3	16,17	Απευθείας Διαθέσιμα												
4	16,97	Απευθείας Διαθέσιμα												
5	5,94	Απευθείας Διαθέσιμα												
6	12,61	Απευθείας Διαθέσιμα												
7	3,33	Απευθείας Διαθέσιμα												
8	16,13	Απευθείας Διαθέσιμα												
9	11,20	Απευθείας Διαθέσιμα												
10	1,14	Απευθείας Διαθέσιμα												
11	4,68	Απευθείας Διαθέσιμα												
12	3,09	Απευθείας Διαθέσιμα												
13	7,26	Απευθείας Διαθέσιμα												
14	2,05	Απευθείας Διαθέσιμα												
15	13,07	Απευθείας Διαθέσιμα												
16	,59	Απευθείας Διαθέσιμα												
17	13,57	Απευθείας Διαθέσιμα												
18	,35	Απευθείας Διαθέσιμα												
19	2,69	Απευθείας Διαθέσιμα												
20	18,45	Απευθείας Διαθέσιμα												
21	4,23	Απευθείας Διαθέσιμα												
22	10,28	Απευθείας Διαθέσιμα												
23	7,10	Απευθείας Διαθέσιμα												
24	-3,09	Απευθείας Διαθέσιμα												
25	5,60	Απευθείας Διαθέσιμα												

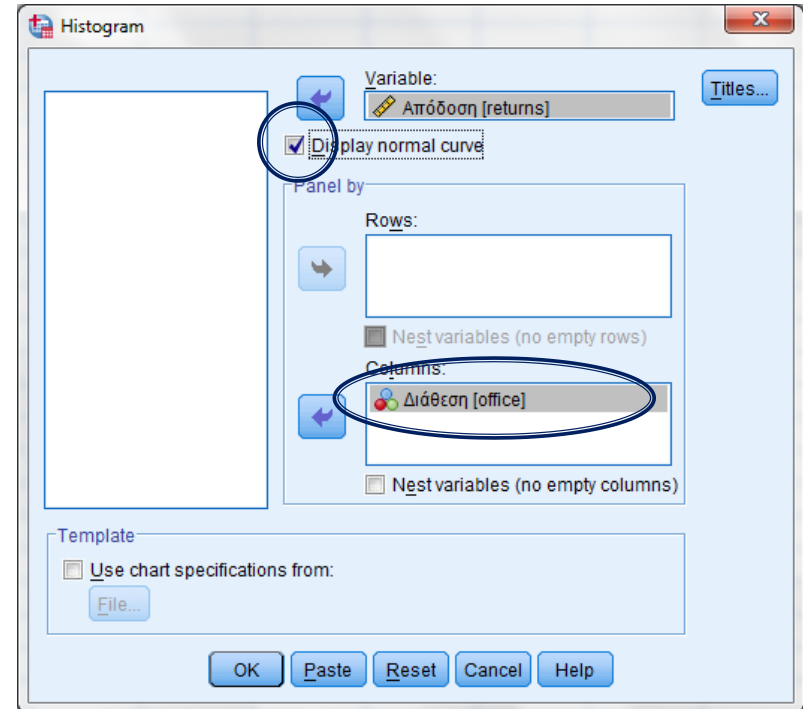
Παρόλο που τα μεγέθη του κάθε δείγματος (σε κάθε ομάδα) είναι αρκετά μεγάλο) (50 παρατηρήσεις για κάθε ομάδα) ελέγχουμε και γραφικά κατά πόσο ευσταθεί η υπόθεση της κανονικότητας)

Δύο ανεξάρτητα δείγματα

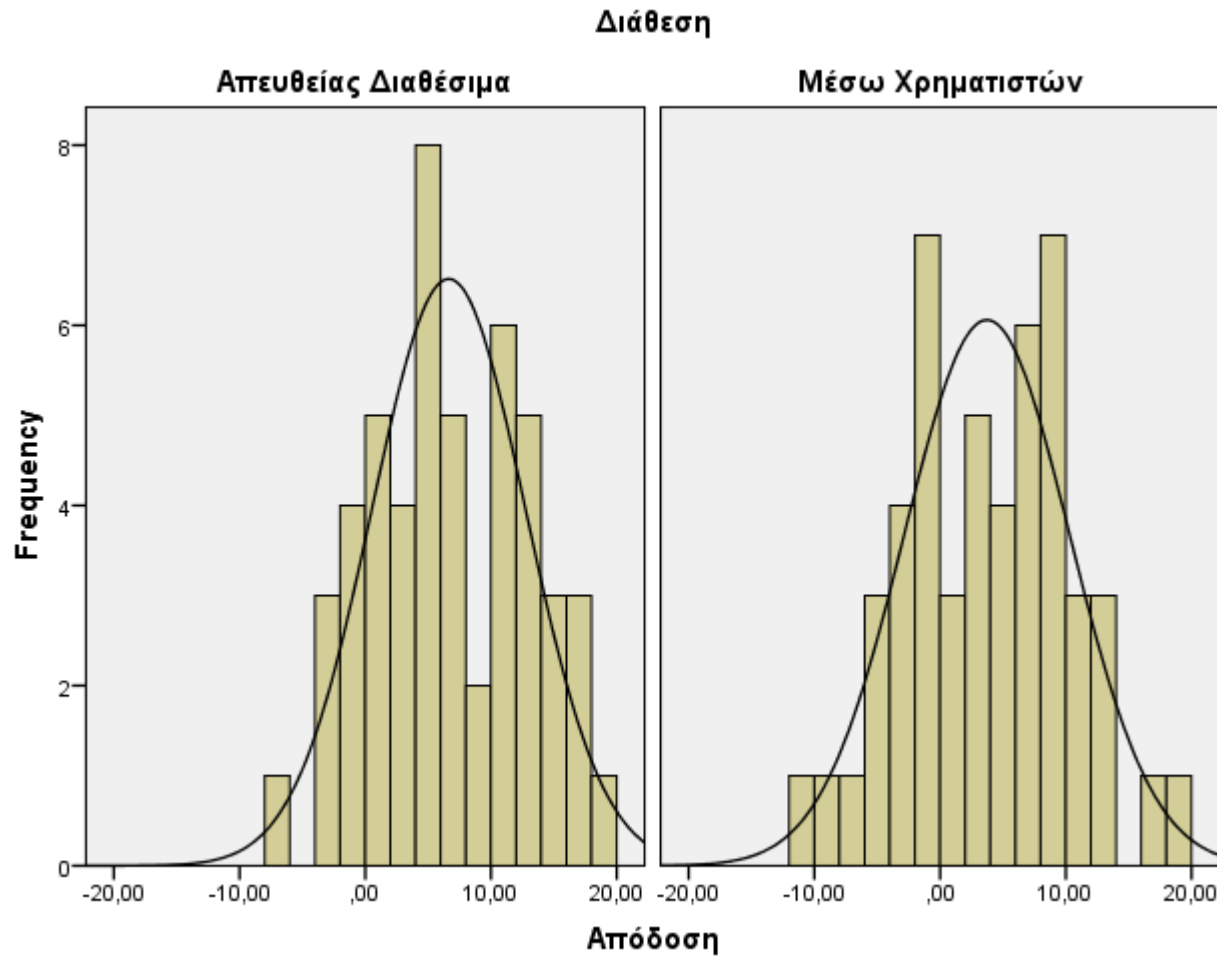


The screenshot shows the Minitab software interface. The main window displays a data table with two columns: 'returns' and 'office'. The 'returns' column contains numerical values, and the 'office' column contains the text 'Απευθείας Διαθέσιμα'. The 'Legacy Dialogs' menu is open, and the 'Histogram...' option is highlighted.

	returns	office	var	var	var	var	var
1	9,33	Απευθείας Διαθέσιμα					
2	6,94	Απευθείας Διαθέσιμα					
3	16,17	Απευθείας Διαθέσιμα					
4	16,97	Απευθείας Διαθέσιμα					
5	5,94	Απευθείας Διαθέσιμα					
6	12,61	Απευθείας Διαθέσιμα					
7	3,33	Απευθείας Διαθέσιμα					
8	16,13	Απευθείας Διαθέσιμα					
9	11,20	Απευθείας Διαθέσιμα					
10	1,14	Απευθείας Διαθέσιμα					
11	4,68	Απευθείας Διαθέσιμα					
12	3,09	Απευθείας Διαθέσιμα					
13	7,26	Απευθείας Διαθέσιμα					
14	2,05	Απευθείας Διαθέσιμα					
15	13,07	Απευθείας Διαθέσιμα					
16	,59	Απευθείας Διαθέσιμα					
17	13,57	Απευθείας Διαθέσιμα					



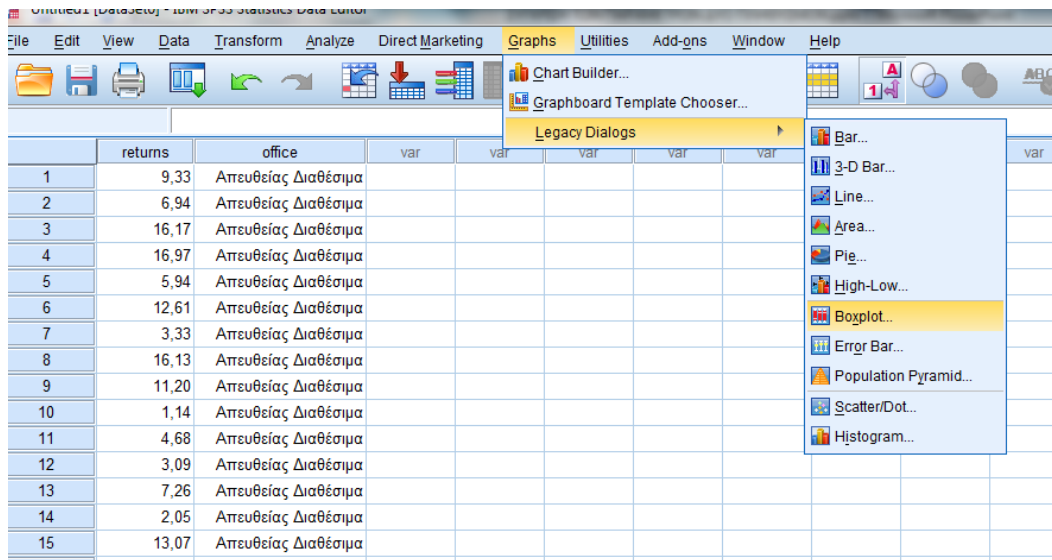
Δύο ανεξάρτητα δείγματα



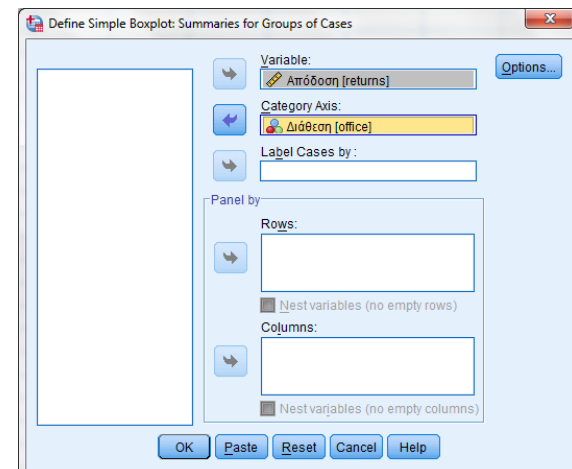
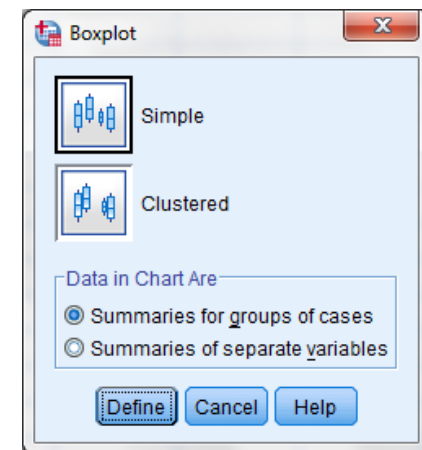
Η υπόθεση
της
κανονικότητας
φαίνεται
λογική και
στις 2
περιπτώσεις

Δύο ανεξάρτητα δείγματα

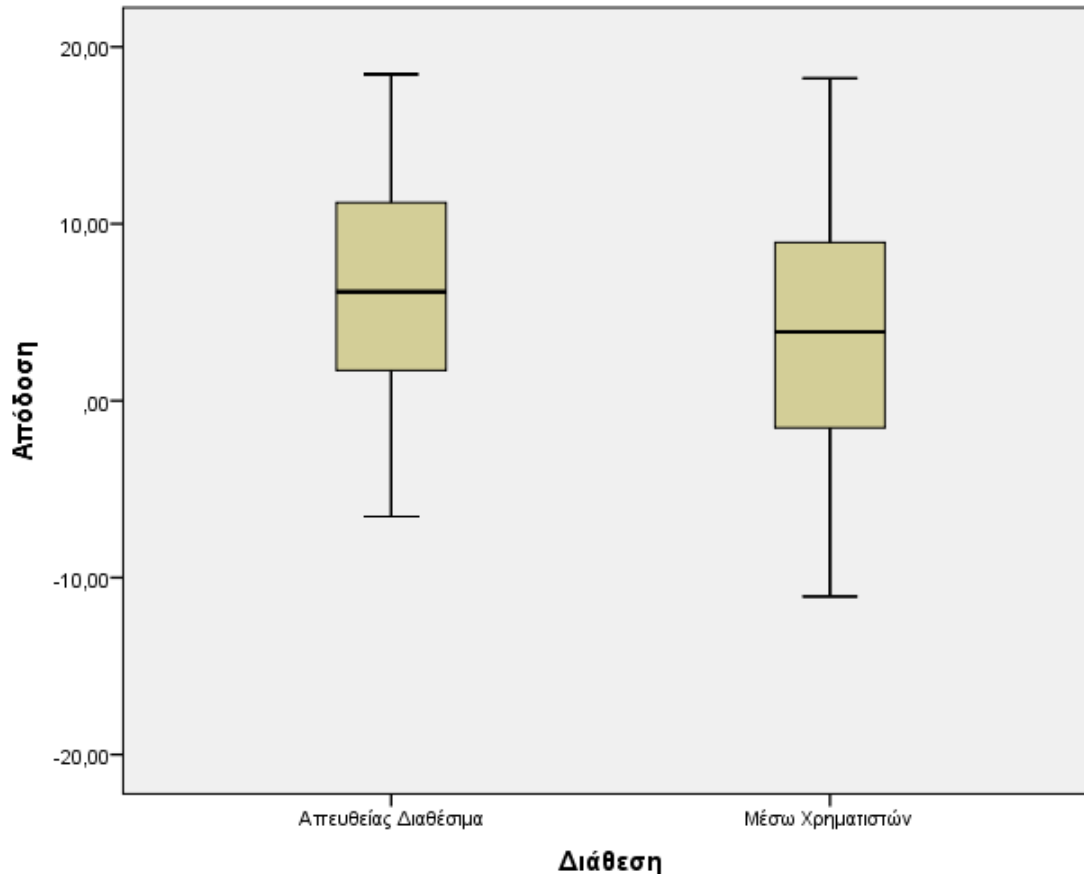
Στην συνέχεια με την χρήση των θηκογραμμάτων συγκρίνουμε στο δείγμα μας τα επίπεδα των αποδόσεων των μετοχών και από τις δύο ομάδες (Τράπεζες-Χρηματιστηριακές εταιρείες)



	returns	office	var	var	var	var	var
1	9,33	Απευθείας Διαθέσιμα					
2	6,94	Απευθείας Διαθέσιμα					
3	16,17	Απευθείας Διαθέσιμα					
4	16,97	Απευθείας Διαθέσιμα					
5	5,94	Απευθείας Διαθέσιμα					
6	12,61	Απευθείας Διαθέσιμα					
7	3,33	Απευθείας Διαθέσιμα					
8	16,13	Απευθείας Διαθέσιμα					
9	11,20	Απευθείας Διαθέσιμα					
10	1,14	Απευθείας Διαθέσιμα					
11	4,68	Απευθείας Διαθέσιμα					
12	3,09	Απευθείας Διαθέσιμα					
13	7,26	Απευθείας Διαθέσιμα					
14	2,05	Απευθείας Διαθέσιμα					
15	13,07	Απευθείας Διαθέσιμα					

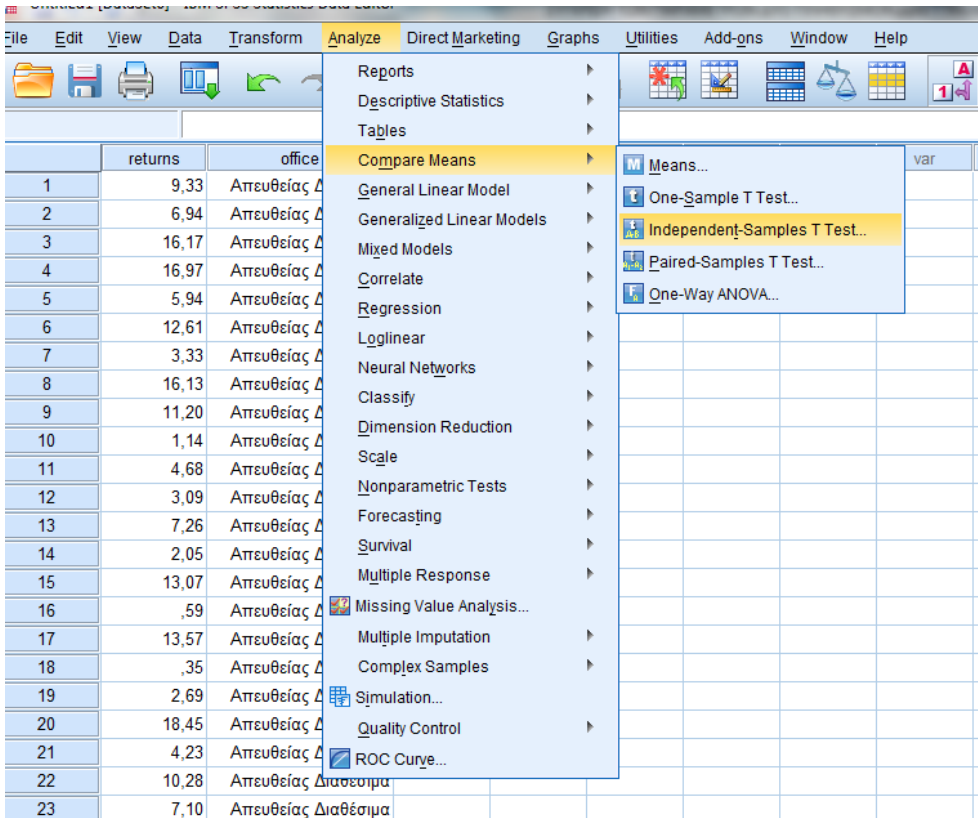


Δύο ανεξάρτητα δείγματα



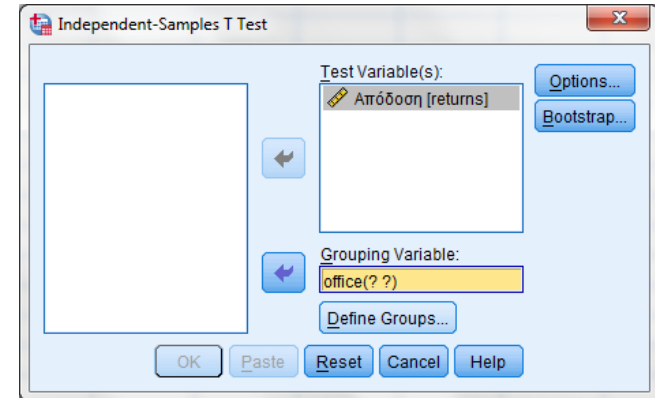
- Από το διπλανό σχήμα καταλαβαίνουμε ότι οι αποδόσεις των αμοιβαίων κεφαλαίων από τις τράπεζες ήταν μεγαλύτερες σε σχέση με τις αποδόσεις των αμοιβαίων από τις χρηματιστηριακές εταιρείες.
- Επίσης παρατηρήστε μία μεγαλύτερη μεταβλητότητα (διασπορά στις αποδόσεις των αμοιβαίων από τις χρηματιστηριακές από ότι στις αποδόσεις των αμοιβαίων από τις Τράπεζες. Είναι όμως και πάλι αυτή η διαφορά στατιστικά σημαντική ή μπορούμε να θεωρήσουμε ότι έχουμε ισότητα διασπορών;

Δύο ανεξάρτητα δείγματα

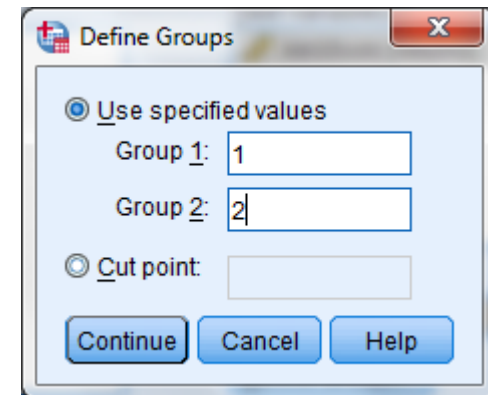


The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and 'Independent-Samples T Test...' is selected. The data view shows a table with columns 'returns' and 'office'.

	returns	office
1	9,33	Απευθείας Δ
2	6,94	Απευθείας Δ
3	16,17	Απευθείας Δ
4	16,97	Απευθείας Δ
5	5,94	Απευθείας Δ
6	12,61	Απευθείας Δ
7	3,33	Απευθείας Δ
8	16,13	Απευθείας Δ
9	11,20	Απευθείας Δ
10	1,14	Απευθείας Δ
11	4,68	Απευθείας Δ
12	3,09	Απευθείας Δ
13	7,26	Απευθείας Δ
14	2,05	Απευθείας Δ
15	13,07	Απευθείας Δ
16	,59	Απευθείας Δ
17	13,57	Απευθείας Δ
18	,35	Απευθείας Δ
19	2,69	Απευθείας Δ
20	18,45	Απευθείας Δ
21	4,23	Απευθείας Δ
22	10,28	Απευθείας Διασπορά
23	7,10	Απευθείας Διασπορά



The 'Independent-Samples T Test' dialog box is shown. The 'Test Variable(s):' field contains 'Απόδοση [returns]'. The 'Grouping Variable:' field contains 'office(? ?)'. Buttons for 'Options...', 'Bootstrap...', 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' are visible.



The 'Define Groups' dialog box is shown. The 'Use specified values' radio button is selected. The 'Group 1:' field contains '1' and the 'Group 2:' field contains '2'. Buttons for 'Continue', 'Cancel', and 'Help' are visible.

Δύο ανεξάρτητα δείγματα

Group Statistics

Διάθεση	N	Mean	Std. Deviation	Std. Error Mean
Απόδοση Απευθείας Διαθέσιμα	50	6,6312	6,12276	,86589
Μέσω Χρηματιστών	50	3,7232	6,58326	,93101

Ο παραπάνω πίνακας μας δίνει περιγραφικά στοιχεία του δείγματος ξεχωριστά. Παρατηρούμε ότι η μέση απόδοση είναι μεγαλύτερη όταν τα αμοιβαία κεφάλαια έχουν διατεθεί απευθείας από ότι ,έσω Χρηματιστηριακών εταιρειών. Επίσης η τυπική απόκλιση είναι μεγαλύτερη για τις αποδόσεις αμοιβαίων κεφαλαίων που διατέθηκαν από Χρηματιστηριακές εταιρείες.

Δύο ανεξάρτητα δείγματα

Δεν μπορούμε να απορρίψουμε την υπόθεση των ίσων διακυμάνσεων

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
Απόδοση	Equal variances assumed	,197	,658	2,287	98	,024	2,90800	1,27144	,38488	5,43112
	Equal variances not assumed			2,287	97,489	,024	2,90800	1,27144	,38471	5,43129

Τιμή της t στατιστικής

Βαθμοί ελευθερίας
100-2

P-Τιμή
αμφίπλευρου
ελέγχου

95% Δ.Ε. της
διαφοράς των
μέσων

Δύο ανεξάρτητα δείγματα

- ✓ Αρχικά δεχόμαστε την H_0 ισότητας διασπορών.
- ✓ Η P -τιμή του ελέγχου είναι αρκετά μικρή οπότε απορρίπτουμε την μηδενική υπόθεση ισότητας των μέσων. Με την βοήθεια των Δ.Ε., συμπεραίνουμε ότι η μέση απόδοση των αμοιβαίων κεφαλαίων που διατέθηκαν από Χρηματιστηριακές εταιρείες είναι κατά έσο όρο χαμηλότερη από ατή των αμοιβαίων κεφαλαίων που διατέθηκαν απευθείας.
- ✓ Όταν τα μεγέθη είναι μικρά και δεν ισχύσει η υπόθεση της κανονικότητας, εφαρμόζουμε τον αντίστοιχο Μη-παραμετρικό έλεγχο που καλείται **Wilcoxon rank sum test ή Mann-Whitney test**

Δύο ανεξάρτητα δείγματα – Μη Παραμετρικός έλεγχος

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Nonparametric Tests' > 'Independent Samples...' is highlighted. The data table below shows two columns: 'returns' and 'office'.

	returns	office
1	9.33	Απευθείας Δ
2	6.94	Απευθείας Δ
3	16.17	Απευθείας Δ
4	16.97	Απευθείας Δ
5	5.94	Απευθείας Δ
6	12.61	Απευθείας Δ
7	3.33	Απευθείας Δ
8	16.13	Απευθείας Δ
9	11.20	Απευθείας Δ
10	1.14	Απευθείας Δ
11	4.68	Απευθείας Δ
12	3.09	Απευθείας Δ
13	7.26	Απευθείας Δ
14	2.05	Απευθείας Δ
15	13.07	Απευθείας Δ
16	.59	Απευθείας Δ
17	13.57	Απευθείας Δ
18	.35	Απευθείας Δ
19	2.69	Απευθείας Δ
20	18.45	Απευθείας Δ
21	4.23	Απευθείας Δ
22	10.28	Απευθείας Διασπαρα
23	7.10	Απευθείας Διασπαρα
24	-3.09	Απευθείας Διασπαρα

The 'Fields' tab of the dialog box shows the 'Test Fields' list containing 'Απόδοση' and the 'Groups' list containing 'Διασπαρα'.

The 'Settings' tab of the dialog box shows the 'Choose Tests' section with 'Mann-Whitney U (2 samples)' selected. Other options include 'Kruskal-Wallis 1-way ANOVA (k samples)', 'Kolmogorov-Smirnov (2 samples)', 'Test for ordered alternatives (Jonckheere-Terpstra for k samples)', 'Test for randomness (Wald-Wolfowitz for 2 samples)', 'Moses extreme reaction (2 samples)', 'Compare Medians across Groups', and 'Hodges-Lehmann estimate (2 samples)'.

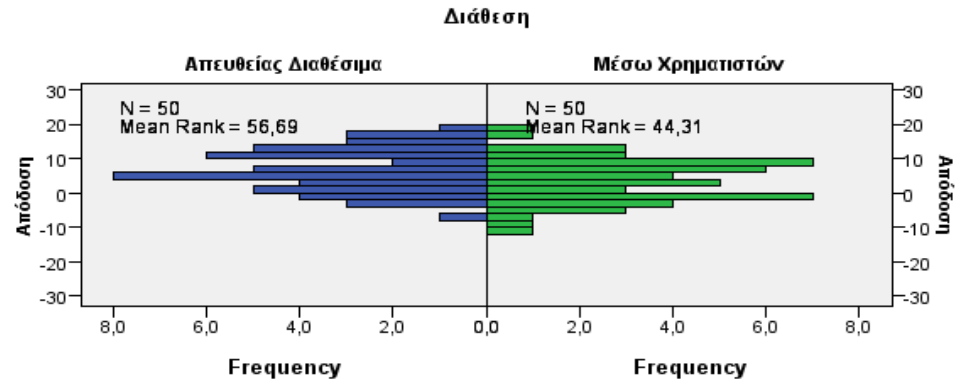
Δύο ανεξάρτητα δείγματα – Μη Παραμετρικός έλεγχος

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Απόδοση is the same across categories of Διάθεση.	Independent-Samples Mann-Whitney U Test	,033	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Independent-Samples Mann-Whitney U Test



Total N	100
Mann-Whitney U	940,500
Wilcoxon W	2.215,500
Test Statistic	940,500
Standard Error	145,057
Standardized Test Statistic	-2,134
Asymptotic Sig. (2-sided test)	,033

Παρατηρούμενα και Πειραματικά δεδομένα

- ▶ Ερευνητική υπόθεση: Η κατανάλωση δημητριακών με υψηλή περιεκτικότητα σε φυτικές ίνες για πρωινό, καταναλώνουν λιγότερες θερμίδες για μεσημεριανό.
- ▶ Δείγμα 150 ατόμων
- ▶ Έλεγχος διαφοράς μέσων των πληθυσμών.
- ▶ $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 < 0$
- ▶ Ο έλεγχος για ίσες διασπορές έδωσε $F=0.3845$ και τιμή $p=0.0008$

...αποτελέσματα Excel

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		<i>Consumers</i>	<i>Nonconsumers</i>
4	Mean	604.02	633.23
5	Variance	4103	10670
6	Observations	43	107
7	Hypothesized Mean Difference	0	
8	df	123	
9	t Stat	-2.09	
10	P(T<=t) one-tail	0.0193	
11	t Critical one-tail	1.6573	
12	P(T<=t) two-tail	0.0386	
13	t Critical two-tail	1.9794	

Βάση των αποτελεσμάτων συμπεραίνουμε ότι αυτοί που καταναλώνουν για πρωινό δημητριακά με φυτικές ίνες, καταναλώνουν λιγότερες θερμίδες το μεσημεριανό.

Είναι αυτή η μοναδική ερμηνεία;

Πειραματικά δεδομένα

- Πρέπει να εξασφαλίσουμε ότι οι δύο πληθυσμοί έχουν παρόμοια χαρακτηριστικά.
- Όσο μεγαλύτερο είναι το δείγμα τόσο πιθανότερη είναι η ομοιομορφία των πληθυσμών.
- Τα πειραματικά δεδομένα απαιτούν περισσότερη προετοιμασία και σημαντικό μεγαλύτερο κόστος.
- Πολλές φορές δεν είναι δυνατό να προκύψει ελεγχόμενο περιβάλλον.
- Η διαφορά μεταξύ παρατηρούμενων και πειραματικών δεδομένων επηρεάζει την ερμηνεία κάθε στατιστικού αποτελέσματος και όχι μόνο τη διαφορά των μέσων δύο πληθυσμών.

Διαφορά Δύο Μέσων: Σύγκριση κατά Ζεύγη

Προηγουμένως, όταν συγκρίναμε δύο πληθυσμούς τα δύο δείγματα που εξετάζαμε ήταν ανεξάρτητα.

Εάν παρόλο αυτά, κάποιο πείραμα σχεδιαστεί με τέτοιο τρόπο όπου κάθε παρατήρηση του ενός δείγματος να αντιστοιχεί σε μία παρατήρηση του άλλου δείγματος, αυτό λέγεται *πείραμα σύγκρισης κατά ζεύγη*

Παράδειγμα

Σε ένα πείραμα γευσιγνωσίας οίνου, 30 καταναλωτές κρασιού βαθμολόγησαν την ποιότητα ενός κρασιού (σε μία κλίμακα από το 1: Πολύ Κακό μέχρι το 7: Παρά πολύ καλό) δύο φορές. Την πρώτη φορά χωρίς να έχουν λάβει καμία πληροφορία όσον αφορά τις ποικιλίες των σταφυλιών, τον χρόνο ωρίμανσης καθώς και τον τόπο παραγωγής. Την δεύτερη φορά αφού λάβανε όλες τις πληροφορίες που αφορούσε το κρασί. Θέλουμε να ελέγξουμε την υπόθεση ότι η μέση ποιότητα του κρασιού είναι ίδια πριν και μετά την χορήγηση των πληροφοριών που αφορά το κρασί με εναλλακτική ότι η ποιότητα του κρασιού **αυξάνεται** μετά την χορήγηση των απαραίτητων πληροφοριών.

Παράδειγμα

Η παράμετρος είναι η διαφορά μεταξύ δύο μέσων (μ_1 = μέση ποιότητα κρασιού πριν την χορήγηση πληροφοριών και μ_2 = μέση ποιότητα κρασιού μετά την χορήγηση πληροφοριών) .

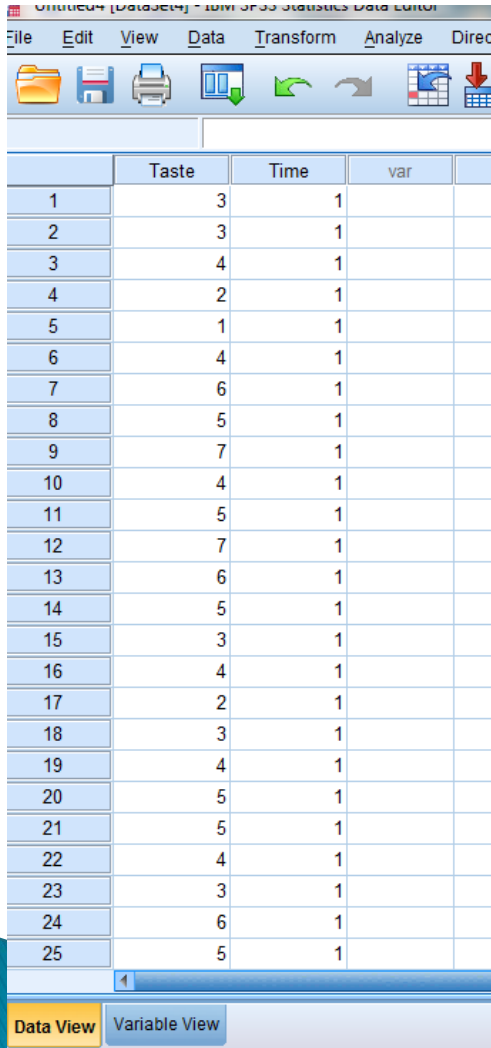
Επειδή ενδιαφερόμαστε να διαπιστώσουμε εάν η ποιότητα του κρασιού αυξήθηκε μετά την χορήγηση πληροφοριών ο έλεγχος θα είναι μονόπλευρος μικρότερη από...

$$H_0 : (\mu_1 - \mu_2) = 0$$

$$H_1 : (\mu_1 - \mu_2) < 0$$

Παράδειγμα Αποτελέσματα SPSS θεωρώντας ότι έχω 2 Ανεξάρτητα δείγματα Πριν και Μετά

Υπολογισμός

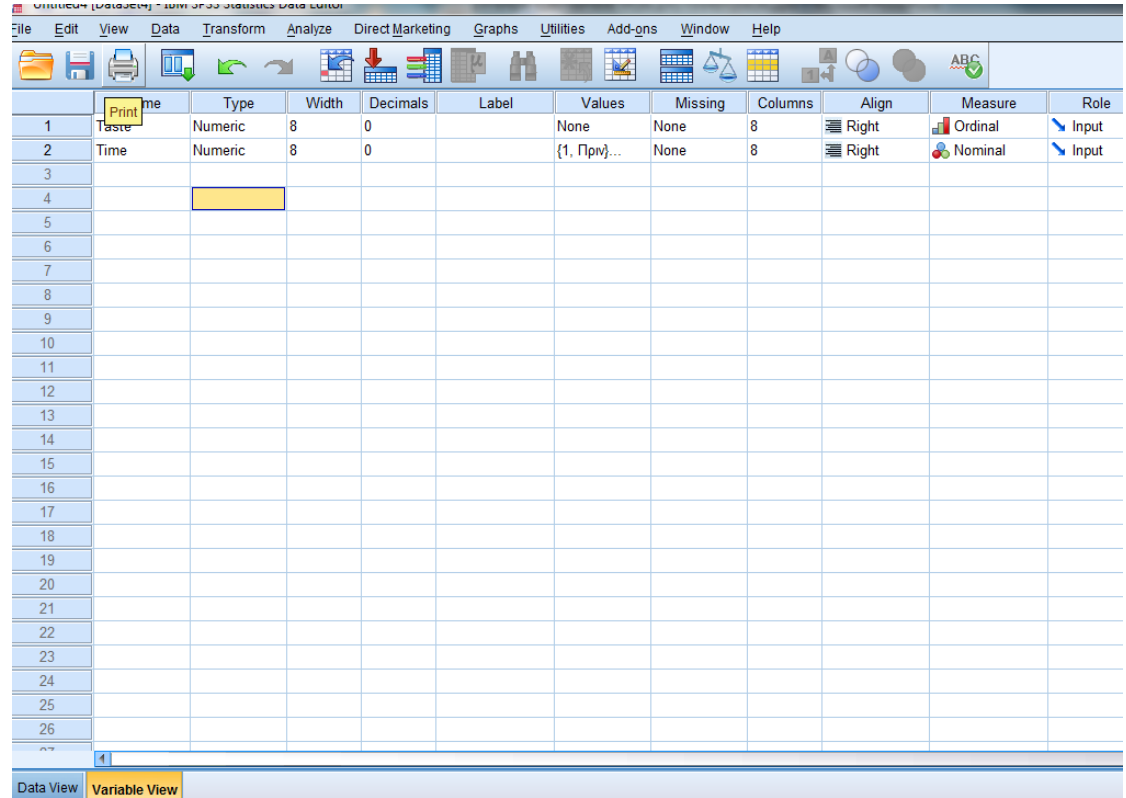


SPSS Statistics Data Editor - Untitled4 [Database] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct

	Taste	Time	var
1	3	1	
2	3	1	
3	4	1	
4	2	1	
5	1	1	
6	4	1	
7	6	1	
8	5	1	
9	7	1	
10	4	1	
11	5	1	
12	7	1	
13	6	1	
14	5	1	
15	3	1	
16	4	1	
17	2	1	
18	3	1	
19	4	1	
20	5	1	
21	5	1	
22	4	1	
23	3	1	
24	6	1	
25	5	1	

Data View Variable View



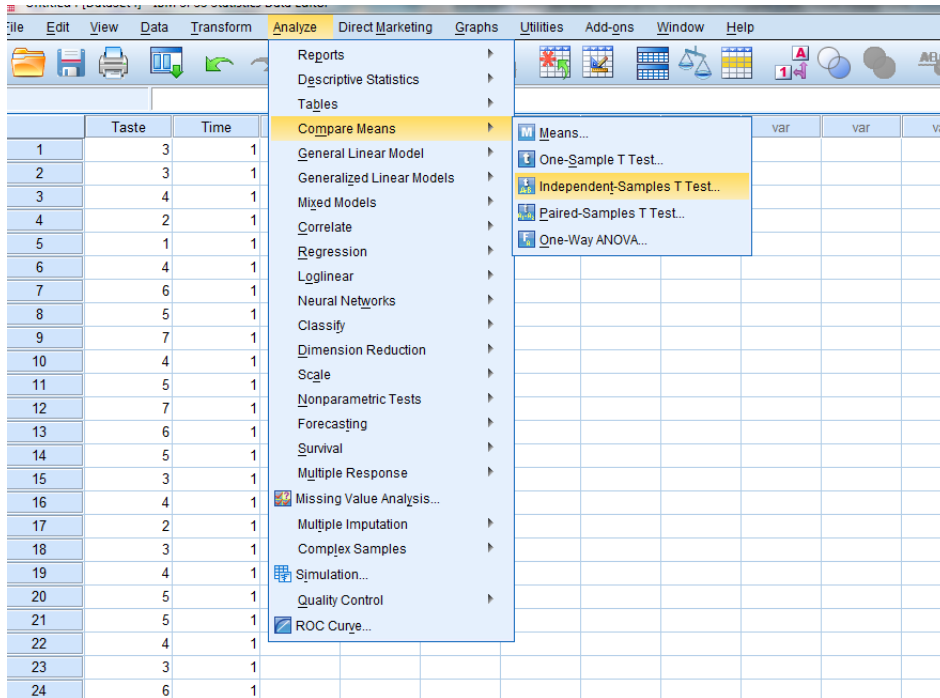
SPSS Statistics Data Editor - Untitled4 [Database] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

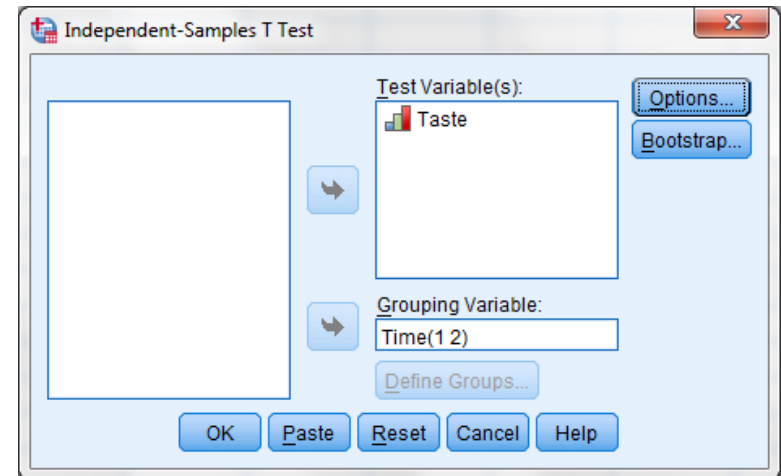
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Taste	Numeric	8	0		None	None	8	Right	Ordinal	Input
2	Time	Numeric	8	0		{1, Πριν}...	None	8	Right	Nominal	Input
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											

Data View Variable View

Παράδειγμα Αποτελέσματα SPSS θεωρώντας ότι έχω 2 Ανεξάρτητα δείγματα Πριν και Μετά



	Taste	Time
1	3	1
2	3	1
3	4	1
4	2	1
5	1	1
6	4	1
7	6	1
8	5	1
9	7	1
10	4	1
11	5	1
12	7	1
13	6	1
14	5	1
15	3	1
16	4	1
17	2	1
18	3	1
19	4	1
20	5	1
21	5	1
22	4	1
23	3	1
24	6	1



Παράδειγμα Αποτελέσματα SPSS θεωρώντας ότι έχω 2 Ανεξάρτητα δείγματα Πριν και Μετά

Υπολογισμός

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	
VAR00001	Equal variances assumed	,003	,956	1,759	58	,084	,73333	,41698	-,10134	1,56800
	Equal variances not assumed			1,759	57,888	,084	,73333	,41698	-,10137	1,56804

Γίνεται δεκτή η υπόθεση των ίσων διακυμάνσεων

P-τιμή για τον αμφίπλευρο έλεγχο. Για τον μονόπλευρο η P-τιμή είναι ίση με αυτή του αμφίπλευρου/2=0.042

Παράδειγμα Αποτελέσματα SPSS θεωρώντας ότι έχω 2 Ανεξάρτητα δείγματα Πριν και Μετά

Ερμηνεία

Η τιμή του ελέγχου είναι ($t = 1.759$) και η τιμή $p = 0.042$.

Συνεπώς τα δεδομένα μας δεν μπορούν να στηρίξουν την υπόθεση ότι η μέση ποιότητα όπως βαθμολογείται από τους καταναλωτές είναι ίση πριν και μετά την παροχή πληροφοριών. Συνεπώς οι καταναλωτές θεωρούν το κρασί ανώτερο ποιοτικά μετά την παροχή των κατάλληλων πληροφοριών.

Παράδειγμα

Επαναλαμβάνουμε το πείραμα με τον εξής τρόπο

Το πείραμα έχει σχεδιαστεί με τέτοιο τρόπο ώστε το ίδιο άτομο να αξιολογεί την ποιότητα του κρασιού σε δύο φάσεις: μία χωρίς να γνωρίζει τίποτε για το κρασί που δοκιμάζει και σε μία δεύτερη φάση να αξιολογεί το κρασί αφού του έχουν δοθεί οι σχετικές πληροφορίες. Άρα μιλάμε για μία μεταβλητή (ποιότητα) στην ίδια μονάδα μελέτης (καταναλωτής) αλλά σε διαφορετικές στιγμές (πριν και μετά).

Για το έλεγχο της υπόθεσης αρχικά υπολογίζουμε την διαφορά μεταξύ των παρατηρήσεων κάθε ζεύγους.

Παράδειγμα

Recall recently used dialogs

	Before	After	var	var	var	var	var
1	3	2					
2	3	4					
3	4	3					
4	2	2					
5	1	3					
6	4	2					
7	6	4					
8	5	5					
9	7	4					
10	4	5					
11	5	3					
12	7	6					
13	6	5					
14	5	3					
15	3	5					
16	4	2					
17	2	1					
18	3	1					
19	4	2					
20	5	4					
21	5	6					
22	4	2					
23	3	3					
24	6	2					
25	5	4					

Data View Variable View

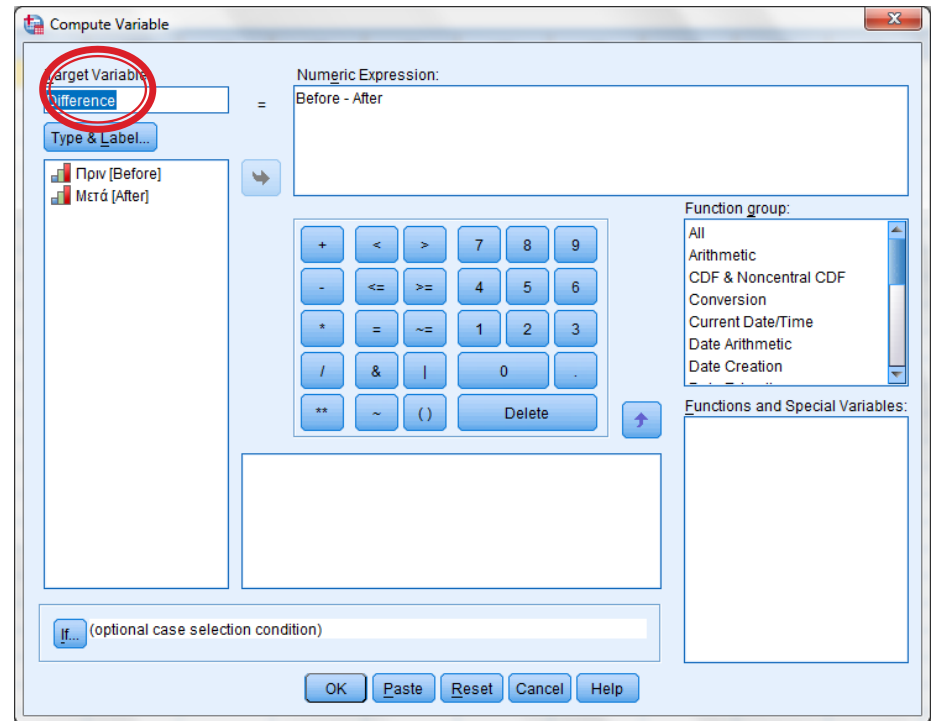
Αναγνωρίζω την παράμετρο

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Before	Numeric	8	0	Πριν	None	None	8	Right	Ordinal
2	After	Numeric	8	0	Μετά	None	None	8	Right	Ordinal
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

Data View Variable View

Δύο Εξαρτημένα δειγματα

	Before	After
1	3	
2	3	
3	4	
4	2	
5	1	
6	4	
7	6	
8	5	
9	7	
10	4	
11	5	
12	7	
13	6	
14	5	3
15	3	5
16	4	2
17	2	1
18	3	1
19	4	2
20	5	4
21	5	6
22	4	2

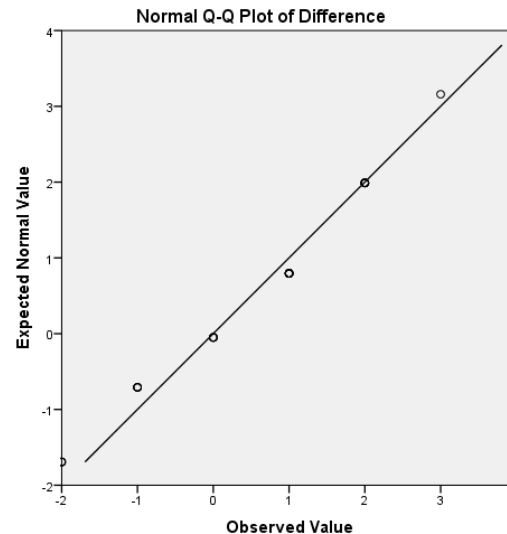
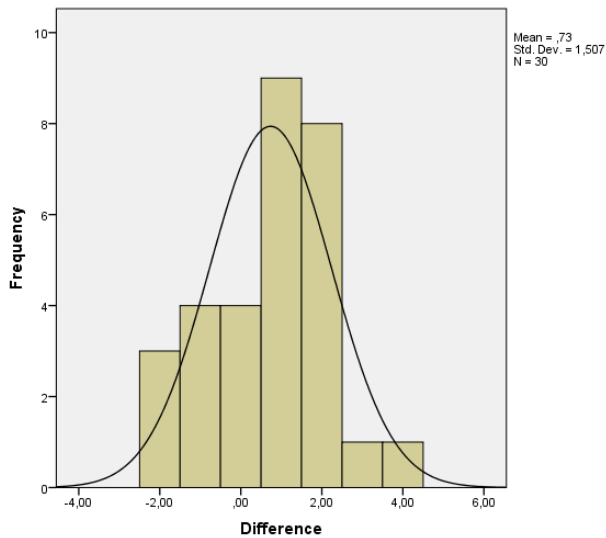


Η **διαφορά των μέσων** είναι ίση με τον **μέσο των διαφορών**, και ως εκ τούτου θα θεωρήσουμε τον «**μέσο του πληθυσμού των διαφορών**» ως την παράμετρο που πρέπει να ελέγξουμε :

$$\mu_1 - \mu_2 = \mu_D$$

Δύο Εξαρτημένα δείγματα

- ▶ Μιας και το μέγεθος του δείγματος είναι σχετικά μικρό ελέγχουμε γραφικά την υπόθεση της κανονικότητας για την διαφορά (μεταβλητή: Difference)



Η υπόθεση της κανονικότητας είναι λογική

A. Έλεγχος της διαφοράς

- ▶ Εφαρμόζουμε τον έλεγχο για την νέα μεταβλητή που δηλώνει την διαφορά (difference) ότι η μέση τιμή της είναι 0 (μηδενική υπόθεση) με εναλλακτική ότι είναι μικρότερη, διάφορη ή μεγαλύτερη του μηδενός)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Difference	30	,7333	1,50707	,27515

One-Sample Test

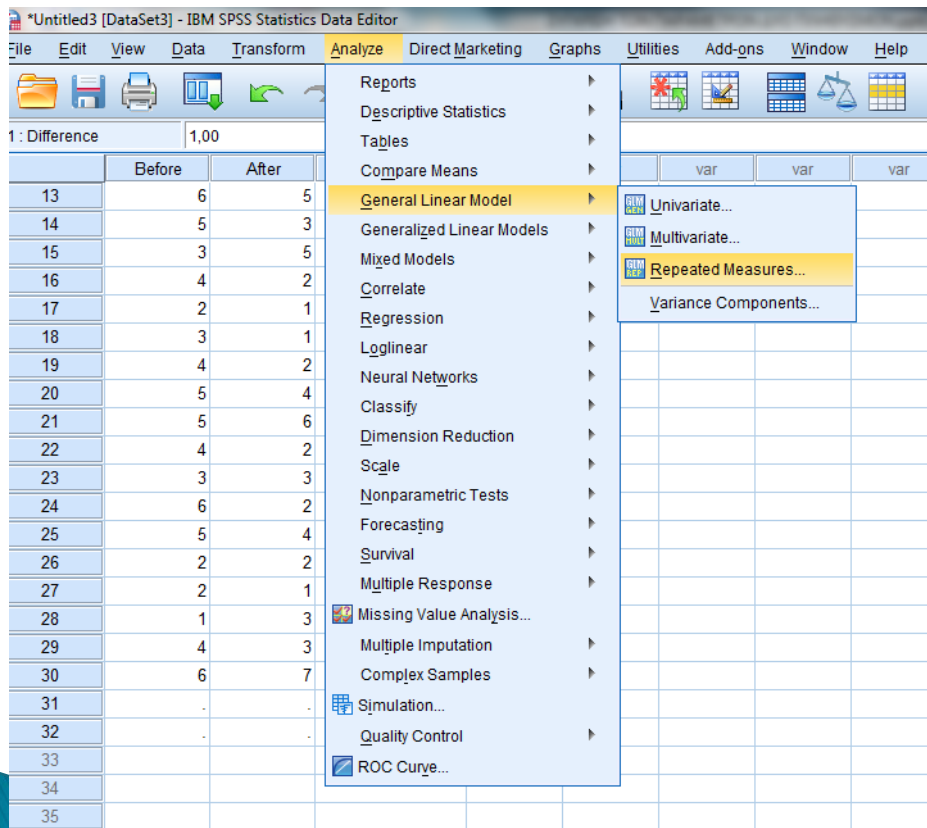
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Difference	2,665	29	,012	,73333	,1706	1,2961

Η θετική τιμή της διαφοράς μας πληροφορεί ότι η αξιολόγηση της ποιότητας ήταν καλύτερη πριν

P-τιμή αμφίπλευρου ελέγχου. Για τον μονόπλευρο έλεγχο η τιμή αυτή πρέπει να διαιρεθεί με το 2

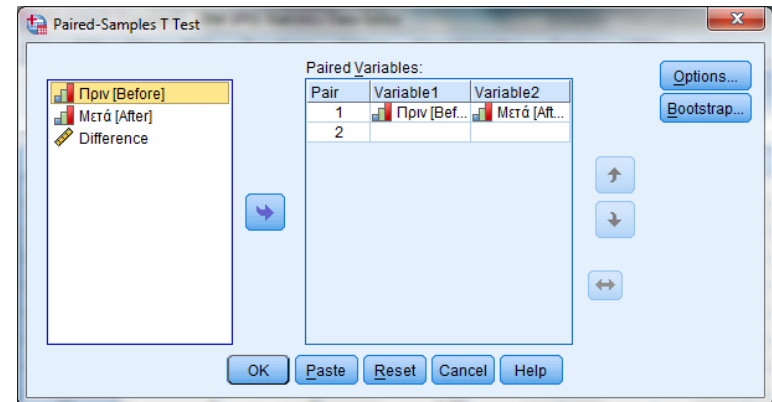
B. Έλεγχος κατά Ζεύγη

Στο ίδιο συμπέρασμα καταλήγουμε αν εφαρμόσουμε το **Paired Sample test**



The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and 'Repeated Measures...' is selected. The data table below shows the following values:

	Before	After
13	6	5
14	5	3
15	3	5
16	4	2
17	2	1
18	3	1
19	4	2
20	5	4
21	5	6
22	4	2
23	3	3
24	6	2
25	5	4
26	2	2
27	2	1
28	1	3
29	4	3
30	6	7
31	.	.
32	.	.
33		
34		
35		



Β. Έλεγχος κατά Ζεύγη

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Πριν	4,03	30	1,650	,301
	Μετά	3,30	30	1,579	,288



Περιγραφικά στατιστικά

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Πριν & Μετά	30	,565	,001

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Πριν - Μετά	,733	1,507	,275	,171	1,296	2,665	29	,012

t στατιστική

P-τιμή
αμφίπλευρου
ελέγχου

Β. Έλεγχος κατά Ζεύγη

Ερμηνεία

Με βάση τα παραπάνω αποτελέσματα έχουμε πολύ σοβαρές ενδείξεις εναντίον της μηδενικής υπόθεσης. Άρα πράγματι οι καταναλωτές αξιολογούν χειρότερα το κρασί μετά την παροχή των κατάλληλων πληροφοριών.

Έλεγχος των διαφορών

Αναγνωρίζω την παράμετρο

Ο έλεγχος για τον μέσο μ_D των διαφορών δύο πληθυσμών υπολογίζεται από τον τύπο:

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$$

Που ακολουθεί κατανομή t-Student με $n_D - 1$ βαθμούς ελευθερίας, με την προϋπόθεση ότι ο πληθυσμός των διαφορών έχει κανονική κατανομή..

Excel Αποτελέσματα

Data, Data Analysis, t-Test: Paired Two- Sample for Means

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK
Cancel
Help

Παράδειγμα

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		<i>Finance</i>	<i>Marketing</i>
4	Mean	65,438	60,374
5	Variance	444,981,810	469,441,785
6	Observations	25	25
7	Pearson Correlation	0.9520	
8	Hypothesized Mean Difference	0	
9	df	24	
10	t Stat	3.81	
11	P(T<=t) one-tail	0.0004	
12	t Critical one-tail	1.7109	
13	P(T<=t) two-tail	0.0009	
14	t Critical two-tail	2.0639	

Παράδειγμα

Η τιμή έλεγχου είναι ίση με $t=3.81$ ενώ η κριτική τιμή $t_{0.05, 24}=1.7109$ και η τιμή p ίση με 0.0004. Βρισκόμαστε σε περιοχή απόρριψης άρα έχουμε επαρκή στατιστικά στοιχεία για να απορρίψουμε την μηδενική υπόθεση.

Διαπιστώνουμε ότι με τη σύγκριση κατά ζεύγη μπορέσαμε να εξάγουμε από τα δεδομένα μας την πληροφορία ότι οι απόφοιτοι MBA που προέρχονται από οικονομικά Τμήματα δέχονται υψηλότερες προσφορές από αποφοίτους MBA που προέρχονται από το Πολυτεχνείο.

Διαφορά δύο ποσοστών

Τώρα θα μελετήσουμε μεθόδους που θα μπορούμε να καταλήγουμε σε συμπεράσματα σχετικά με την διαφορά μεταξύ δύο πληθυσμών όταν τα δεδομένα μας είναι **ονομαστικά**.

Όταν τα δεδομένα είναι ονομαστικά (nominal), το μόνο πράγμα που μπορεί να μετρηθεί είναι ο αριθμός των εμφανίσεων (συχνότητα) ενός ενδεχομένου και ο υπολογισμός του ποσοστού του ενδεχομένου στο σύνολο του πληθυσμού. Κατά συνέπεια, η μόνη παράμετρος που μπορούμε να ελέγξουμε ή να εκτιμήσουμε, είναι η διαφορά δύο ποσοστών σε δύο πληθυσμούς : **$p_1 - p_2$** .

Εκτιμητής και κατανομή δειγματοληψίας...

Για να εκτιμήσουμε ή να ελέγξουμε την διαφορά δύο ποσοστών $\mathbf{p_1-p_2}$, επιλέγουμε τυχαία δείγματα και υπολογίζουμε τα ποσοστά δειγμάτων.

$$\hat{p}_1 = \frac{x_1}{n_1} \text{ and } \hat{p}_2 = \frac{x_2}{n_2}$$

Η διαφορά $\hat{p}_1 - \hat{p}_2$ είναι ένας αμερόληπτος εκτιμητής της διαφοράς $\mathbf{p_1-p_2}$.

x_1 αριθμός των επιτυχιών σε ένα δείγμα μεγέθους n_1 \rightarrow $\hat{p}_1 = \frac{x_1}{n_1}$ και $\hat{p}_2 = \frac{x_2}{n_2}$

Κατανομή δειγματοληψίας

Ο έλεγχος $\hat{p}_1 - \hat{p}_2$ κατά προσέγγιση κανονική κατανομή, με την **προϋπόθεση** ότι τα μεγέθη των δειγμάτων είναι αρκετά μεγάλα ώστε τα τέσσερα γινόμενα $n_1\hat{p}_1, n_2\hat{p}_2, n_1(1-\hat{p}_1)$ και $n_2(1-\hat{p}_2)$, να είναι όλα μεγαλύτερα ή ίσα του 5.

Ο μέσος της διαφοράς

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

Η Διακύμανση και το τυπικό σφάλμα

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}, \text{ and}$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

...συνεπώς η στατιστική z είναι μία κατά προσέγγιση τυποποιημένη κανονική κατανομή:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Έλεγχος και Εκτίμηση της $p_1 - p_2$

Επειδή τα ποσοστά των πληθυσμών είναι άγνωστα, το τυπικό σφάλμα

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

είναι άγνωστο. Ως εκ τούτου, έχουμε δύο **διαφορετικούς εκτιμητές** για το τυπικό σφάλμα της διαφοράς $\hat{p}_1 - \hat{p}_2$, και η επιλογή του καταλληλότερου εξαρτάται από την μηδενική υπόθεση.

Έλεγχος της διαφοράς $p_1 - p_2$

Υπάρχουν δύο περιπτώσεις ...

Case 1:

If our null hypothesis is stated as:

$$H_0: (p_1 - p_2) = 0$$

our test statistic is...

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where \hat{p} is the "pooled proportion estimator" given by

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Case 2:

If our null hypothesis is stated as:

$$H_0: (p_1 - p_2) = D \quad (D \neq 0)$$

our test statistic is...

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

Έλεγχος της διαφοράς $p_1 - p_2$

- ▶ Ισοδύναμα θα μπορούσαμε τα δεδομένα μας να τα βλέπαμε σε μορφή ενός πίνακα συχνοτήτων 2×2

$Y \backslash X$	0	1
0	n11	n12
1	n21	n22

Παρατηρηθείσες συχνότητες στο δείγμα

Έλεγχος της διαφοράς $p_1 - p_2$

- ▶ Ισοδύναμα τότε με τον προηγούμενο αμφίπλευρο έλεγχο θα ήταν να ελέγξουμε αν οι δύο μεταβλητές μας Y και X είναι ανεξάρτητες (με εναλλακτική ότι δεν είναι).

$$\chi^2 = \sum \frac{(\text{παρατηρηθείσες συχνότητες} - \text{αναμενόμενες συχνότητες})^2}{(\text{αναμενόμενες συχνότητες})}$$

Όπου αναμενόμενες συχνότητες = $\frac{(\text{άθροισμα γραμμής}) \times (\text{άθροισμα στήλης})}{\text{Μέγεθος δείγματος}}$

Έλεγχος της διαφοράς $p_1 - p_2$

- ▶ Η στατιστική έλεγχου χ^2 , κάτω από την μηδενική υπόθεση της ανεξαρτησίας, ακολουθεί προσεγγιστικά την χ^2 κατανομή με 1 βαθμό ελευθερίας. Συχνά για να είναι πιο ικανοποιητική η προσέγγιση μας προχωράμε την διόρθωση συνέχειας του Yates και η στατιστική υπολογίζεται:

$$\chi^2 = \sum \frac{(|\text{παρατηρηθείσες συχνότητες} - \text{αναμενόμενες συχνότητες}| - 0.5)^2}{(\text{αναμενόμενες συχνότητες})}$$

Πολλοί Στατιστικολόγοι θεωρούν την διόρθωση Yates πολύ συντηρητική με αποτέλεσμα να μας οδηγεί συχνά σε Σφάλματα Τύπου II

Έλεγχος της διαφοράς $p_1 - p_2$

Παράδειγμα: Στο προηγούμενο παράδειγμα με την αγορά των αμοιβαίων κεφαλαίων, πέραν της δίτιμης τυχαίας μεταβλητής «Διάθεση» που μας ενημερώνει από πού έχουν αγοράσει τα αμοιβαία κεφάλαια, γνωρίζουμε και τις τιμές της δίτιμης τυχαίας μεταβλητής το φύλο του κάθε επενδυτή (1: Άνδρας ή 2: Γυναίκα).

Θέλουμε τώρα να ελέγξουμε αν υπάρχει διαφοροποίηση μεταξύ ανδρών και γυναικών στην αγορά αμοιβαίων κεφαλαίων.

Έλεγχος της διαφοράς $p_1 - p_2$

IBM SPSS Statistics Data Editor - ανεξάρτητα δειγματα.sav [DataSet1]

File Edit View Data Transform Analyze Direct Marketing

	returns	office	gender	var
1	9,3	1	1	
2	6,9	1	1	
3	16,2	1	1	
4	17,0	1	2	
5	5,9	1	1	
6	12,6	1	2	
7	3,3	1	2	
8	16,1	1	1	
9	11,2	1	1	
10	1,1	1	1	
11	4,7	1	2	
12	3,1	1	2	
13	7,3	1	1	
14	2,1	1	1	
15	13,1	1	1	
16	,6	1	2	
17	13,6	1	2	
18	,4	1	1	
19	2,7	1	2	
20	18,5	1	1	
21	4,2	1	2	
22	10,3	1	1	
23	7,1	1	2	
24	-3,1	1	2	
25	5,6	1	2	

Data View Variable View

IBM SPSS Statistics Data Editor - ανεξάρτητα δειγματα.sav [DataSet1]

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	R	
1	returns	Numeric	8	1	Απόδοση	None	None	8	Right	Scale	In
2	office	Numeric	8	0	Διάθεση	{1, Απρευθε...	None	7	Right	Nominal	In
3	gender	Numeric	8	0	Φύλο	None	None	8	Right	Nominal	In
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											

Data View Variable View

Έλεγχος της διαφοράς $p_1 - p_2$

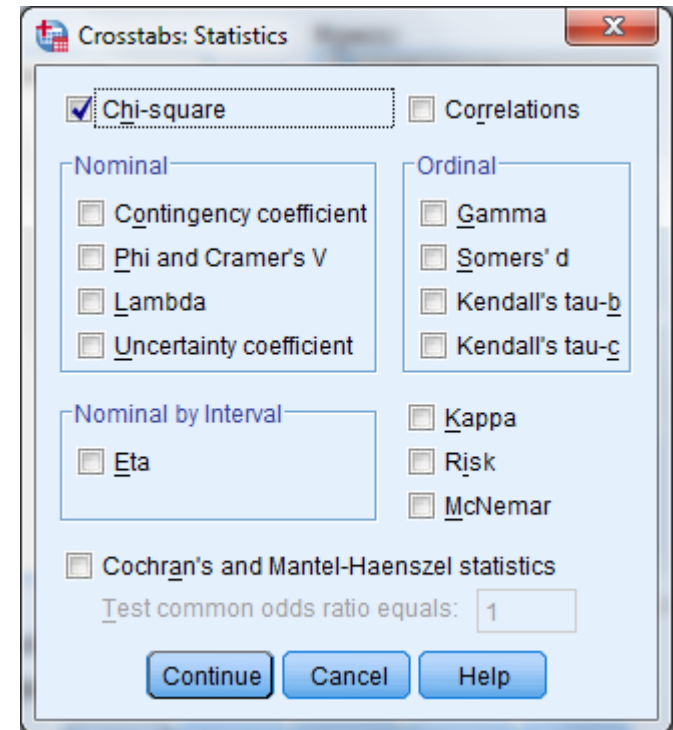
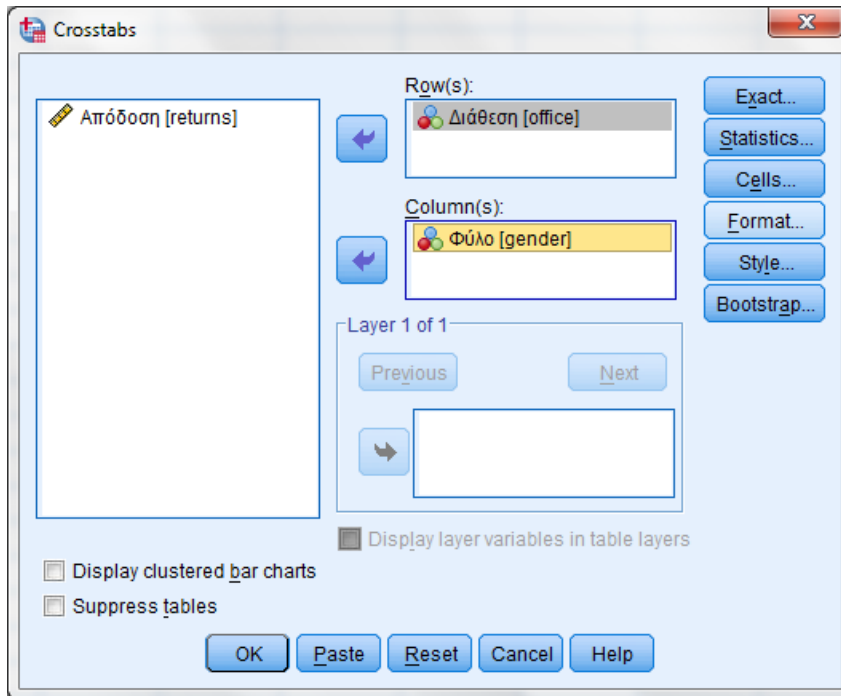
The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Descriptive Statistics' > 'Crosstabs...' is highlighted. The data table below shows the following values:

	returns	office
1	9,3	1
2	6,9	1
3	16,2	1
4	17,0	1
5	5,9	1
6	12,6	1
7	3,3	1
8	16,1	1
9	11,2	1
10	1,1	1
11	4,7	1
12	3,1	1
13	7,3	1
14	2,1	1
15	13,1	1
16	,6	1
17	13,6	1
18	,4	1
19	2,7	1
20	18,5	1
21	4,2	1
22	10,3	1
23	7,1	1

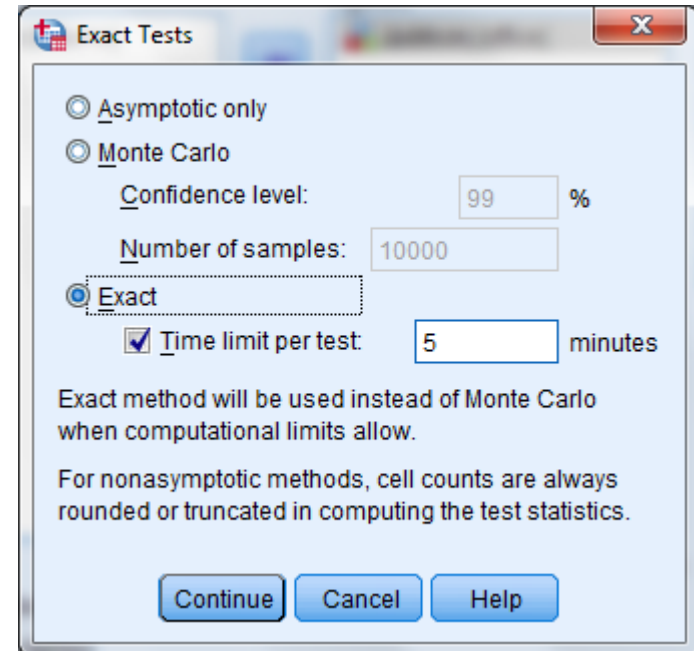
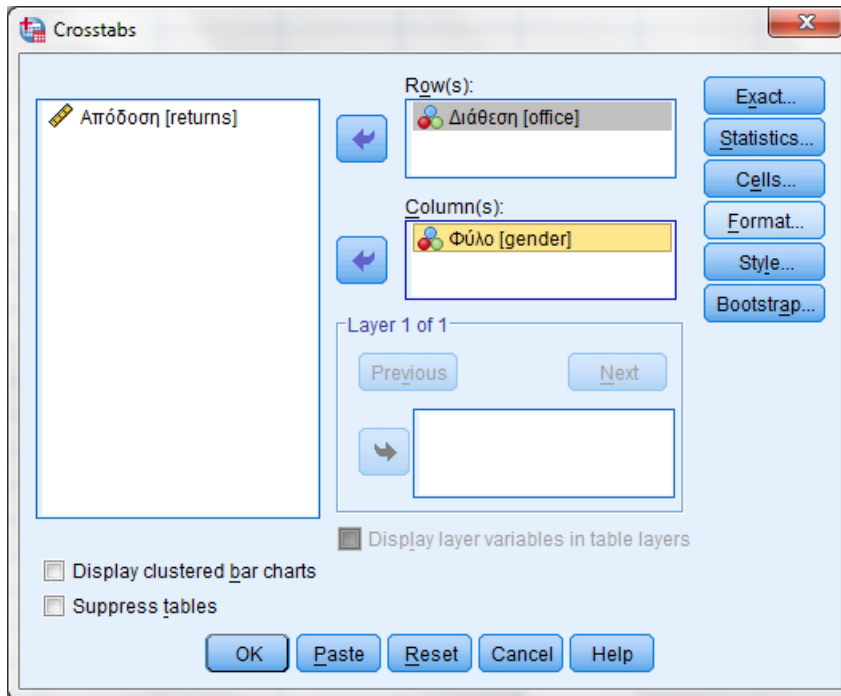
The 'Crosstabs' dialog box is shown. The 'Row(s)' field contains 'Διάρθρωση [office]' and the 'Column(s)' field contains 'Φύλο [gender]'. The 'Display clustered bar charts' checkbox is checked. The 'OK' button is highlighted.

The 'Crosstabs: Cell Display' dialog box is shown. The 'Counts' section has 'Observed', 'Expected', and 'Total' checked. The 'Percentages' section has 'Row', 'Column', and 'Total' checked. The 'Noninteger Weights' section has 'Round cell counts' selected. The 'Continue' button is highlighted.

Έλεγχος της διαφοράς $p_1 - p_2$



Έλεγχος της διαφοράς $p_1 - p_2$



Έλεγχος της διαφοράς $p_1 - p_2$

Διάθεση * Φύλο Crosstabulation

			Φύλο		Total
			Άνδρας	Γυναίκα	
Διάθεση	Απευθείας Διαθέσιμα	Count	26	24	50
		Expected Count	27,5	22,5	50,0
		% within Διάθεση	52,0%	48,0%	100,0%
		% within Φύλο	47,3%	53,3%	50,0%
		% of Total	26,0%	24,0%	50,0%
Μέσω Χρηματιστών		Count	29	21	50
		Expected Count	27,5	22,5	50,0
		% within Διάθεση	58,0%	42,0%	100,0%
		% within Φύλο	52,7%	46,7%	50,0%
		% of Total	29,0%	21,0%	50,0%
Total		Count	55	45	100
		Expected Count	55,0	45,0	100,0
		% within Διάθεση	55,0%	45,0%	100,0%
		% within Φύλο	100,0%	100,0%	100,0%
		% of Total	55,0%	45,0%	100,0%

Συχνότητες: 26 Άνδρες προμηθεύονται αμοιβαία κεφάλαια Απευθείας

Αναμενόμενες Συχνότητες: 27,5% των Ανδρών προμηθεύονται Αμοιβ. Κεφ. Απευθείας (50X55)/100

Το 52% που προμηθεύονται Απευθείας είναι Άνδρες

Το 47 των Ανδρών προμηθεύονται Αμ. Κεφ. Απευθείας

Έλεγχος της διαφοράς $p_1 - p_2$

Στατιστική Ελέγχου χ^2

Στατιστική Ελέγχου χ^2
με διόρθωση συνέχειας

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	,364 ^a	1	,546	,688	,344	
Continuity Correction ^b	,162	1	,688			
Likelihood Ratio	,364	1	,546	,688	,344	
Fisher's Exact Test				,688	,344	
Linear-by-Linear Association	,360 ^c	1	,549	,688	,344	,133
N of Valid Cases	100					

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 22,50.

b. Computed only for a 2x2 table

c. The standardized statistic is -,600.

Ισχύουν οι
προϋποθέσεις ελέγχου

Ασυμπτωτική P-
τιμή αμφίπλευρου
ελέγχου

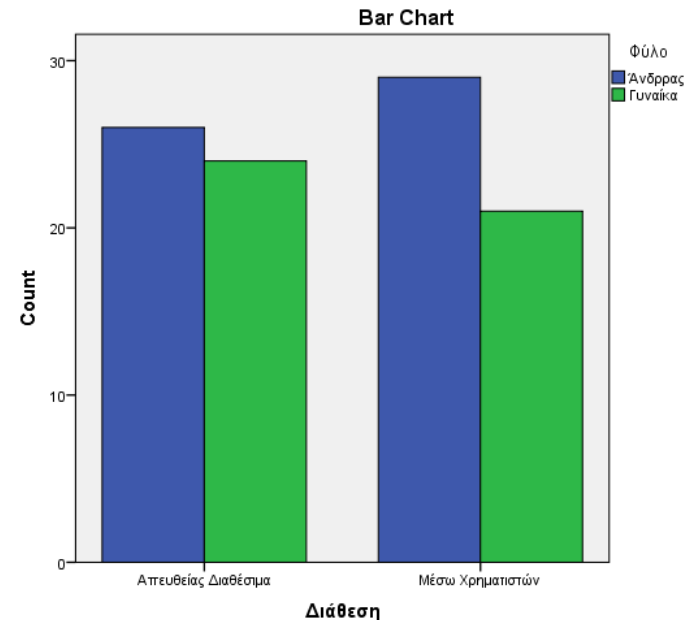
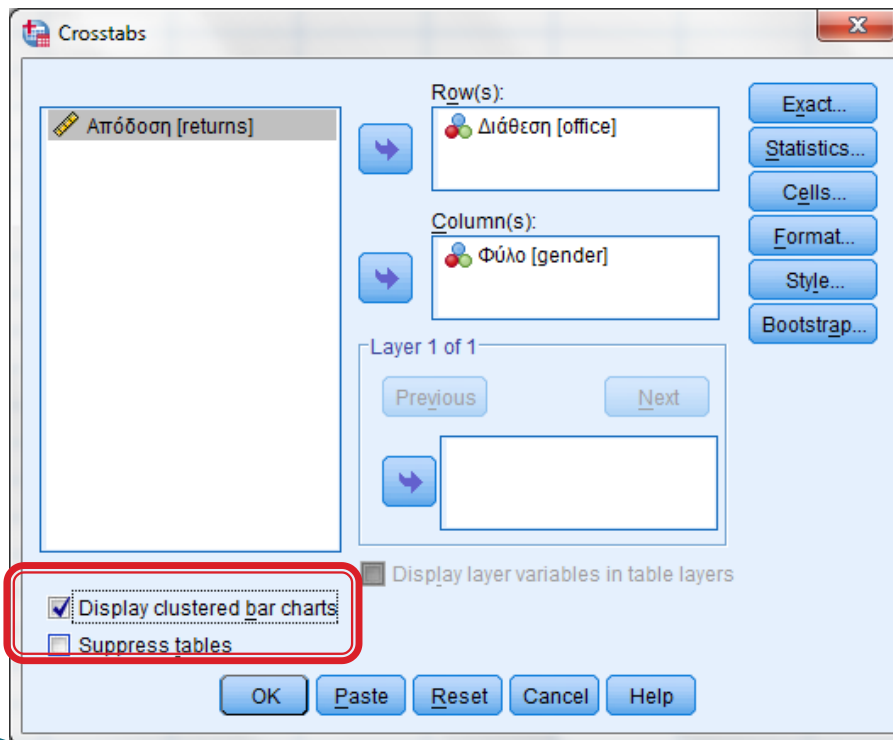
Ασυμπτωτική P-
τιμή αμφίπλευρου
ελέγχου με
διόρθωση

Έλεγχος της διαφοράς $p_1 - p_2$

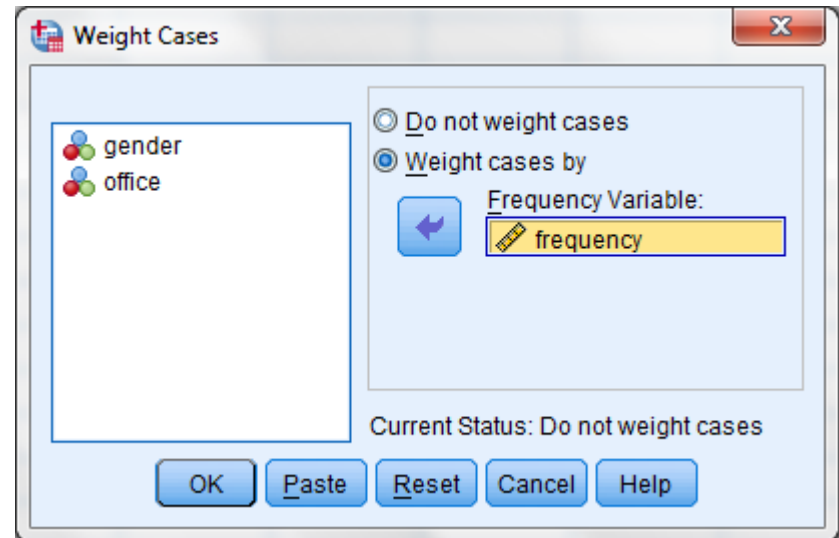
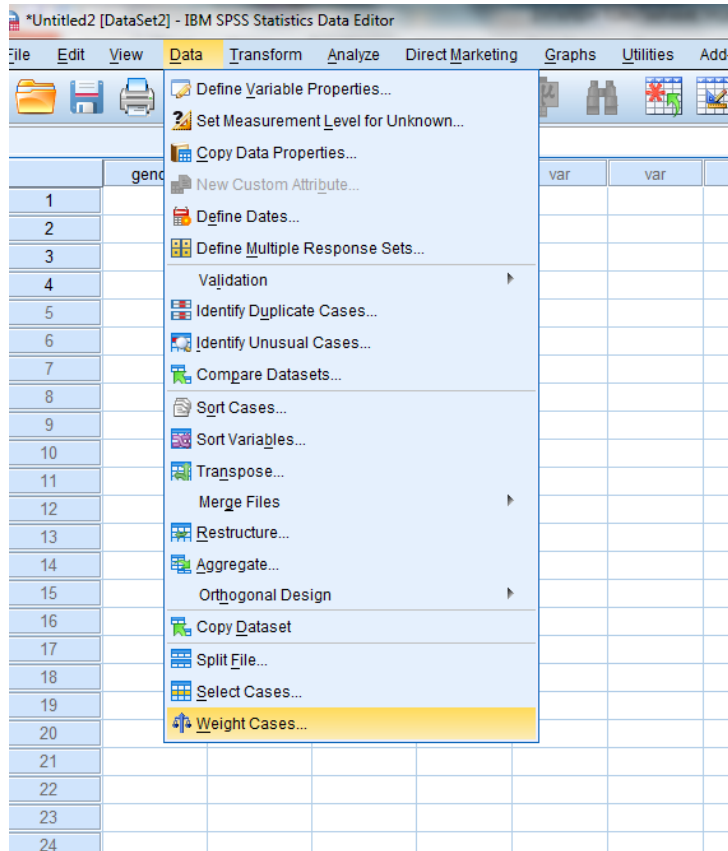
- Κατ' αρχήν παρατηρούμε ότι όλες οι αναμενόμενες συχνότητες είναι $>$ του 5. Άρα μπορούμε να βασίσουμε την συμπερασματολογία μας στα αποτελέσματα του ελέγχου χ^2 .
- Προχωρώντας σε διόρθωση συνέχειας παρατηρούμε ότι σε ε.σ. 5% δεν έχουμε ενδείξεις εναντίον της μηδενικής υπόθεσης, δηλαδή ότι δεν υπάρχει διαφοροποίηση μεταξύ ανδρών και γυναικών στον τρόπο προμήθειας Αμοιβαίων Κεφαλαίων

Έλεγχος της διαφοράς $p_1 - p_2$

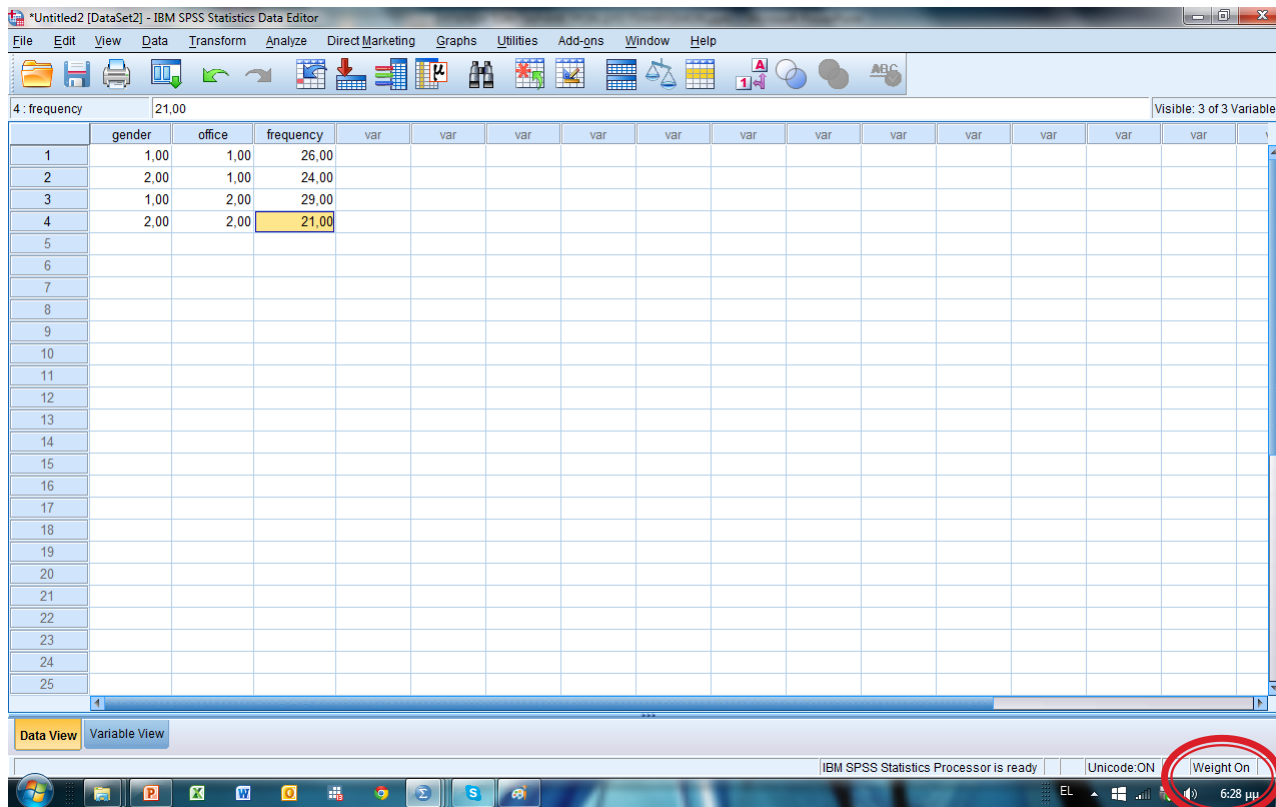
Στο ίδιο συμπέρασμα μπορούμε να καταλήξουμε και με την γραφική απεικόνιση



Έλεγχος της διαφοράς $p_1 - p_2$



Έλεγχος της διαφοράς $p_1 - p_2$



The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a data table with the following data:

	gender	office	frequency	var	var	var	var	var	var	var	var	var	var	var	var
1	1,00	1,00	26,00												
2	2,00	1,00	24,00												
3	1,00	2,00	29,00												
4	2,00	2,00	21,00												
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															

The status bar at the bottom of the window shows "IBM SPSS Statistics Processor is ready", "Unicode: ON", and "Weight On". The "Weight On" button is circled in red.

Τώρα πλέον είμαστε σε θέση να δημιουργήσουμε ένα 2x2 πίνακα συχνοτήτων καθώς και να υπολογίσουμε τα στατιστικά ελέγχου μας για τις μεταβλητές Φύλο και Διάθεση ακριβώς όπως πριν.

Ποια μέθοδο να χρησιμοποιήσω ?

t-test and estimator of μ

z-test and estimator of p



**t-test and estimator of $\mu_1 - \mu_2$
(unequal variances formulas)**

t-test and estimator of μ_D

χ^2 -test and estimator of σ^2

**t-test and estimator of $\mu_1 - \mu_2$
(equal variances formulas)**

F-test and estimator of σ_1^2 / σ_2^2

z-test (cases 1 and 2) and estimator of $p_1 - p_2$

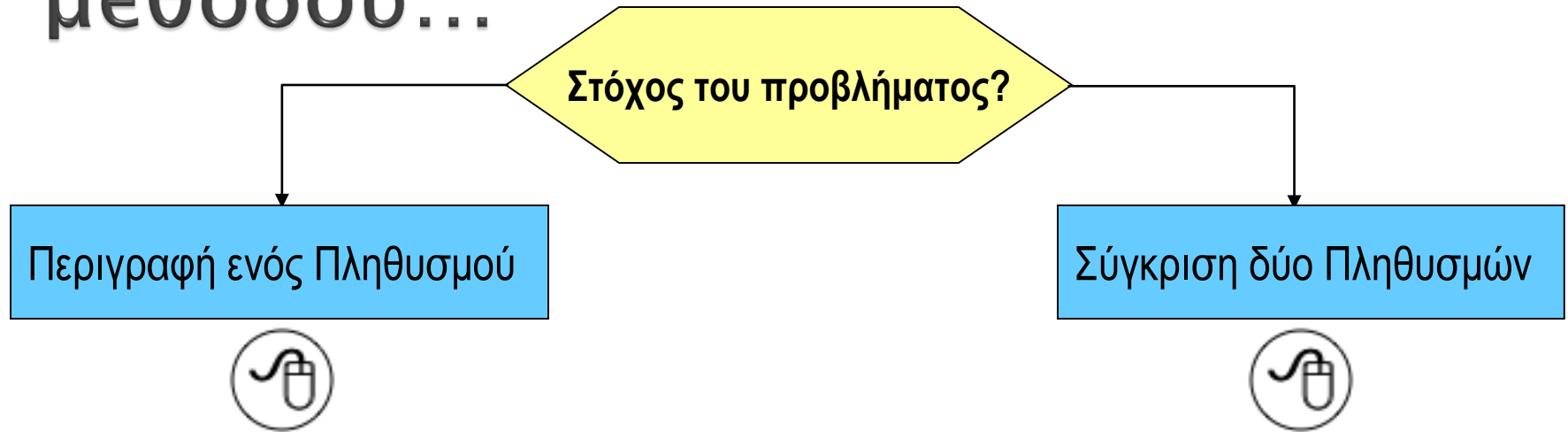


Εντοπισμός της ορθής μεθόδου...

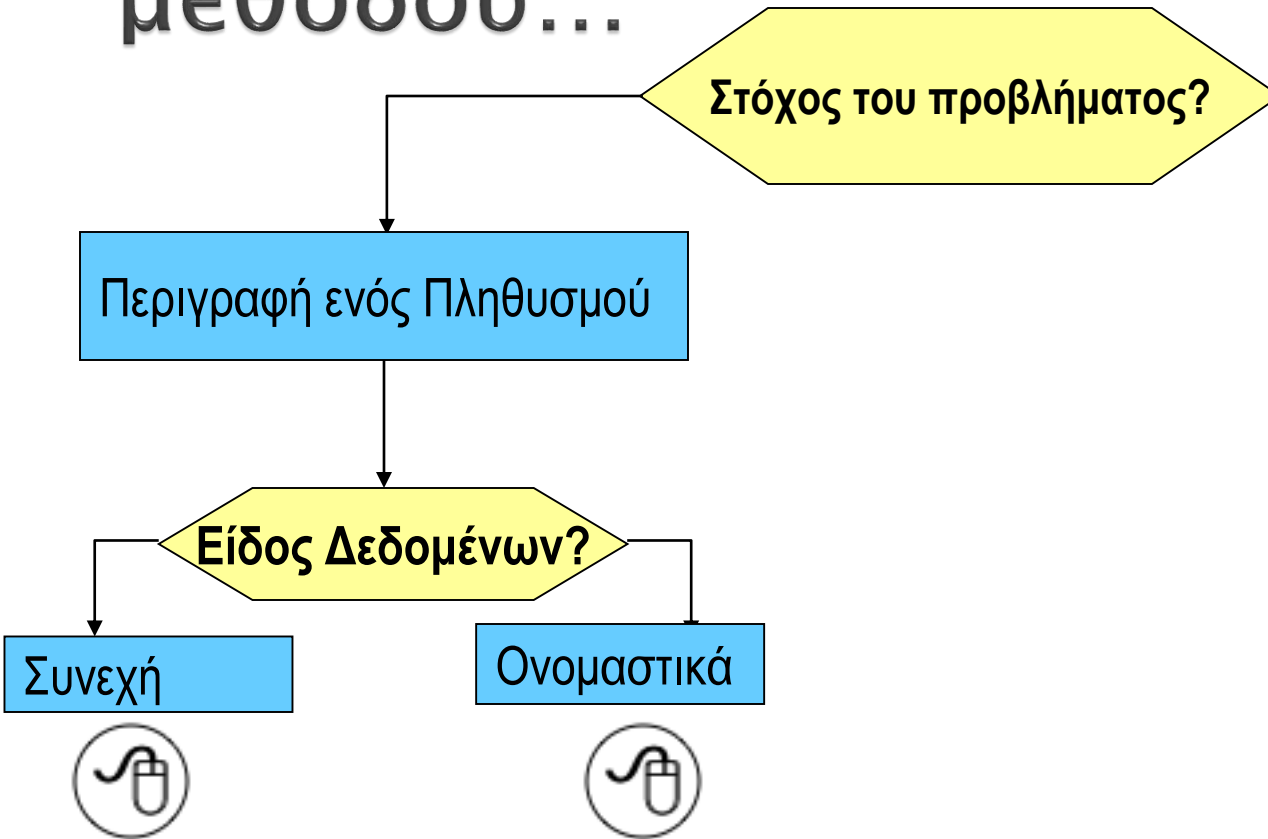
- ▶ Οι σημαντικότεροι παράγοντες για την επιλογή μεθόδου σε ένα στατιστικό πρόβλημα είναι:
 - ο *στόχος του προβλήματος*,
(δηλ. η περιγραφή ενός ή δύο πληθυσμών)
 - και *το είδος των δεδομένων*.
(δηλ. συνεχή ή ονομαστικά)

Μόλις καθοριστούν οι παραπάνω παράγοντες, η ανάλυσή μας εκτείνεται σε άλλους παράγοντες [π.χ. τύπος περιγραφικού μέτρου (Μέτρο κεντρικής τάσης ή μέτρο διασποράς)].

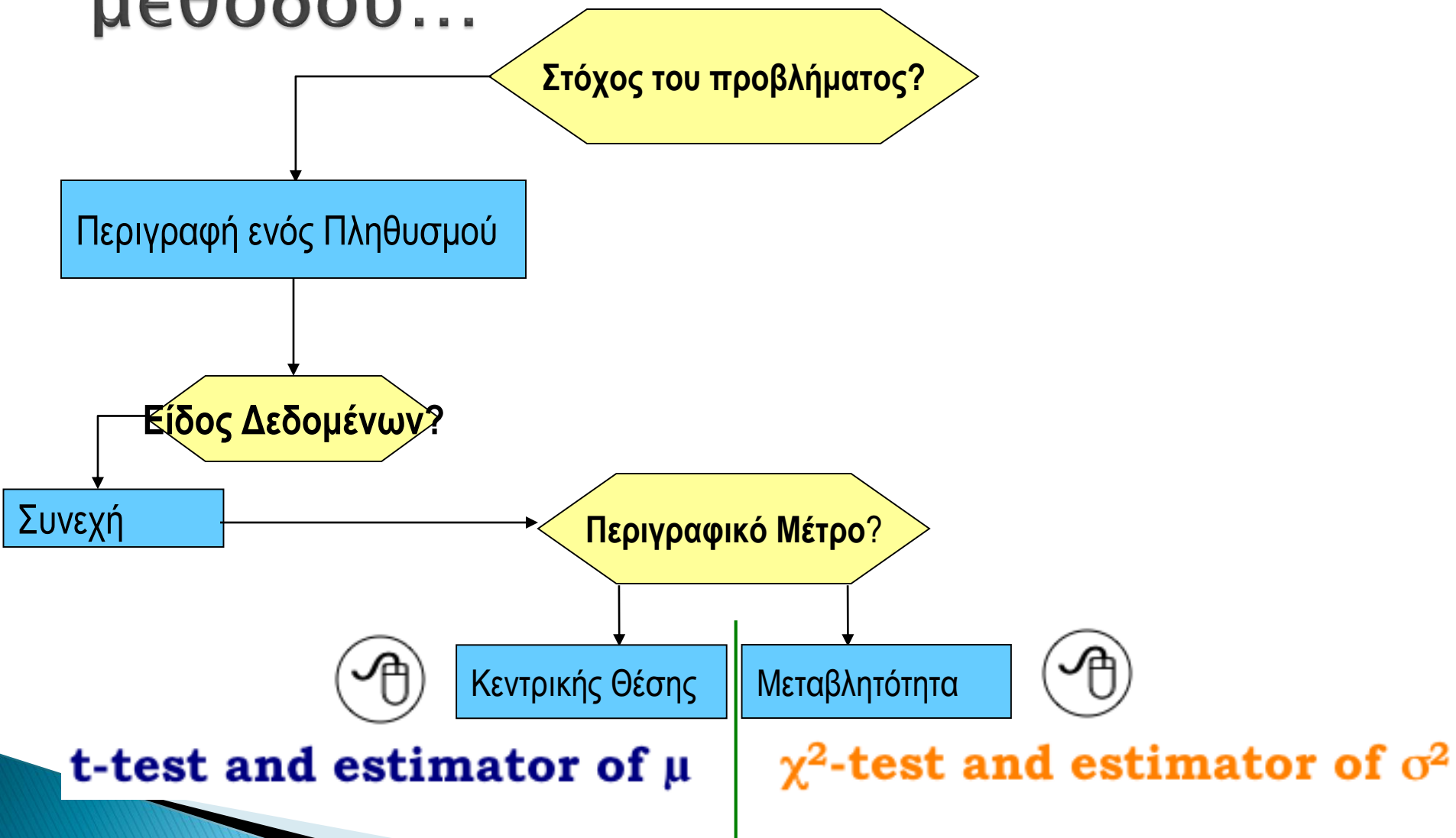
Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...



Μέθοδος...



Αρχή του
Διαγράμματος
Ροής

Describe a Population

Data Type?

Interval

Type of Descriptive Measurement?

Central Location

Parameter of interest: μ

Test Statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\nu = n - 1$ degrees of freedom

Interval Estimator:

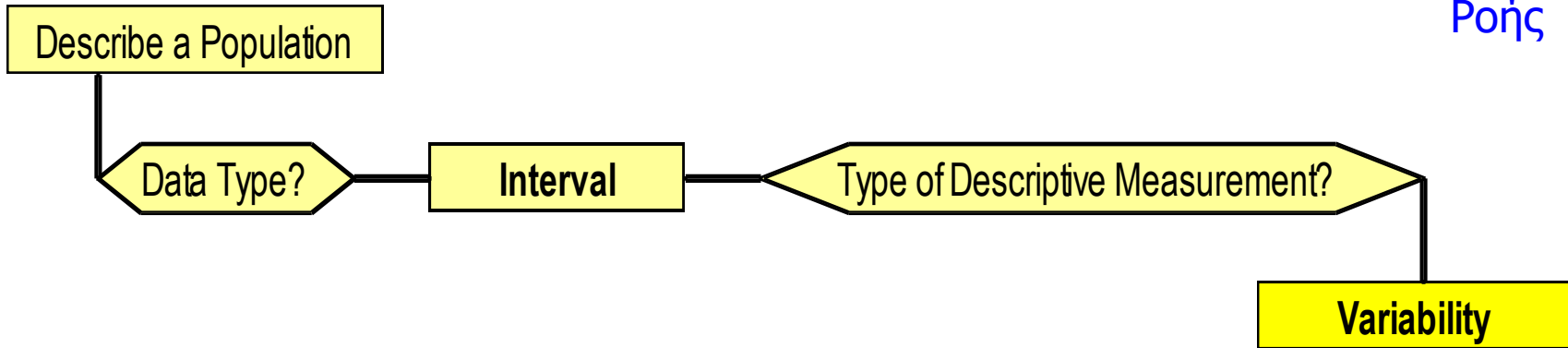
$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Required Condition: Population is normal (or at least "not extremely nonnormal")

Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Parameter of interest: σ^2

Test Statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

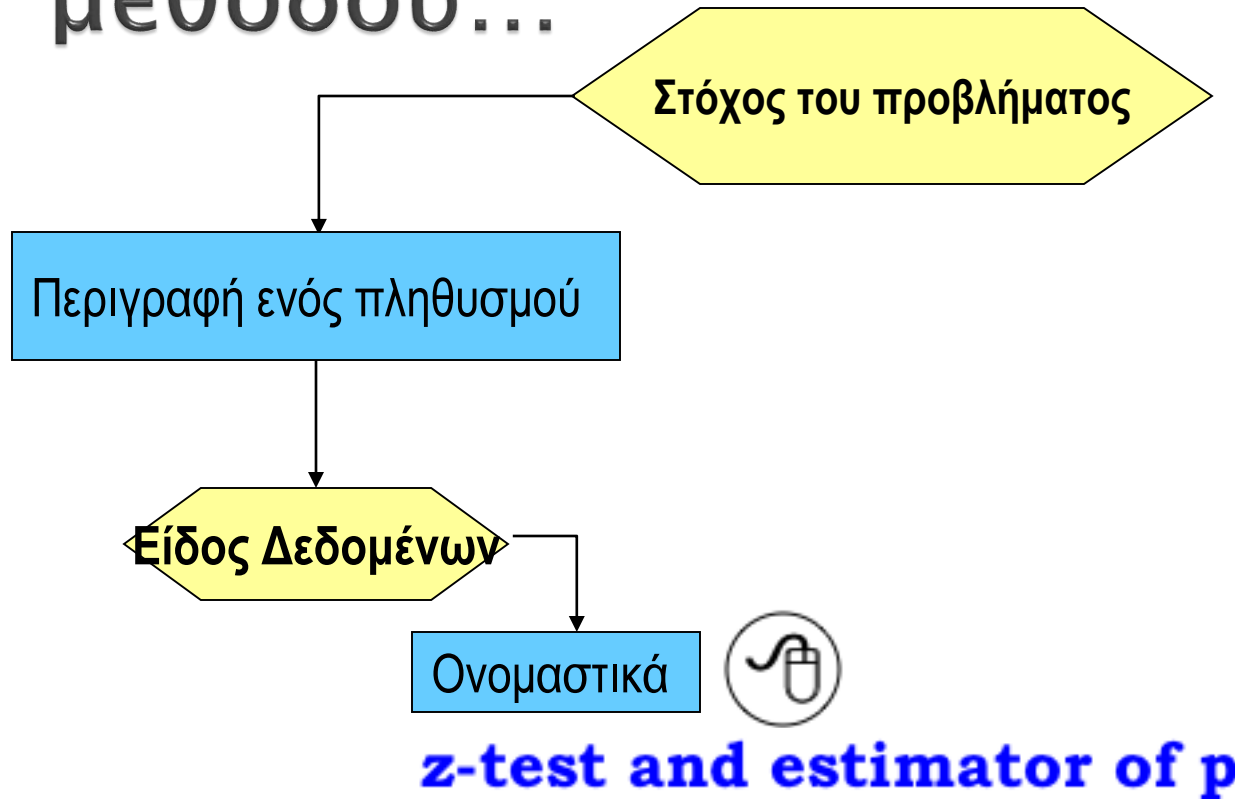
$\nu = n - 1$ degrees of freedom

Interval Estimator:

$$LCL = \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \quad UCL = \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Required Condition: Population is normal

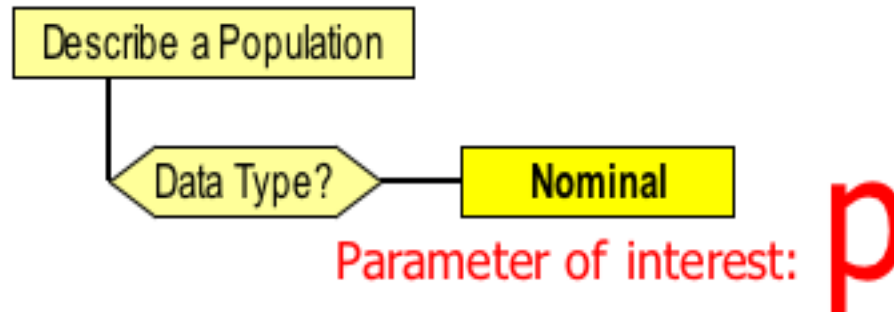
Διάγραμμα ροής επιλογής της μεθόδου...



Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Test Statistic:

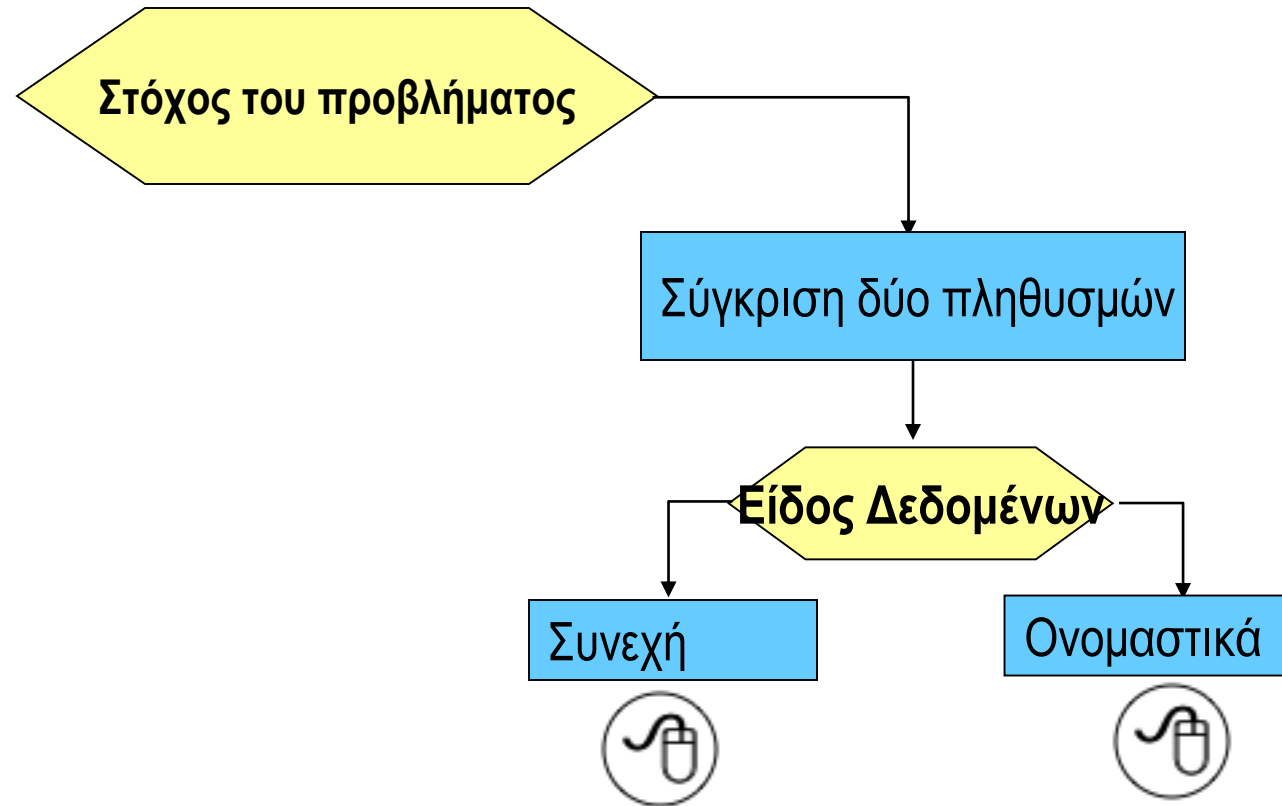
$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

Interval Estimator:

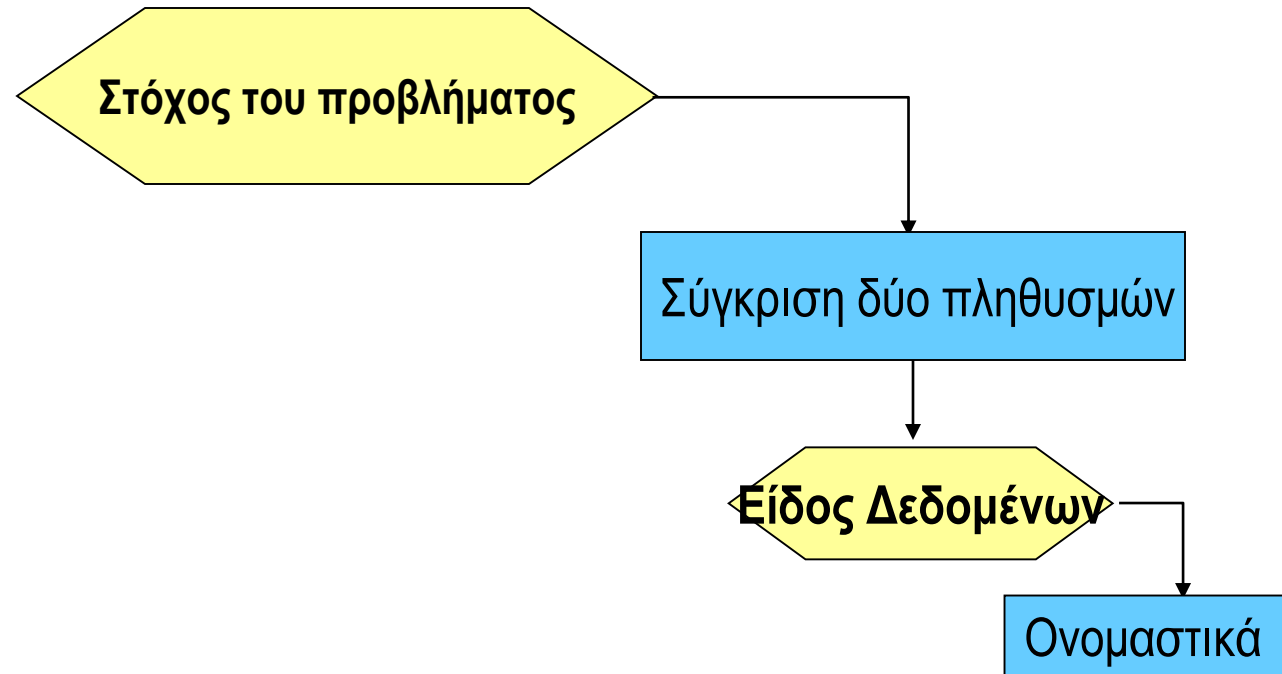
$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Required Conditions: $np \geq 5$ and $n(1-p) \geq 5$ (for test)
 $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$ (for estimate)

Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...



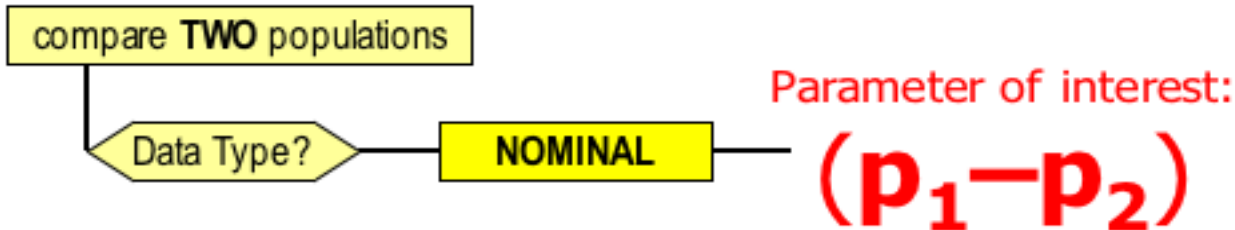
z-test (cases 1 and 2) and estimator of $p_1 - p_2$



Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Test Statistic:
depends on null hypothesis...

Case 1: $H_0: (p_1 - p_2) = 0$

Case 2: $H_0: (p_1 - p_2) = D$ ($D \neq 0$)

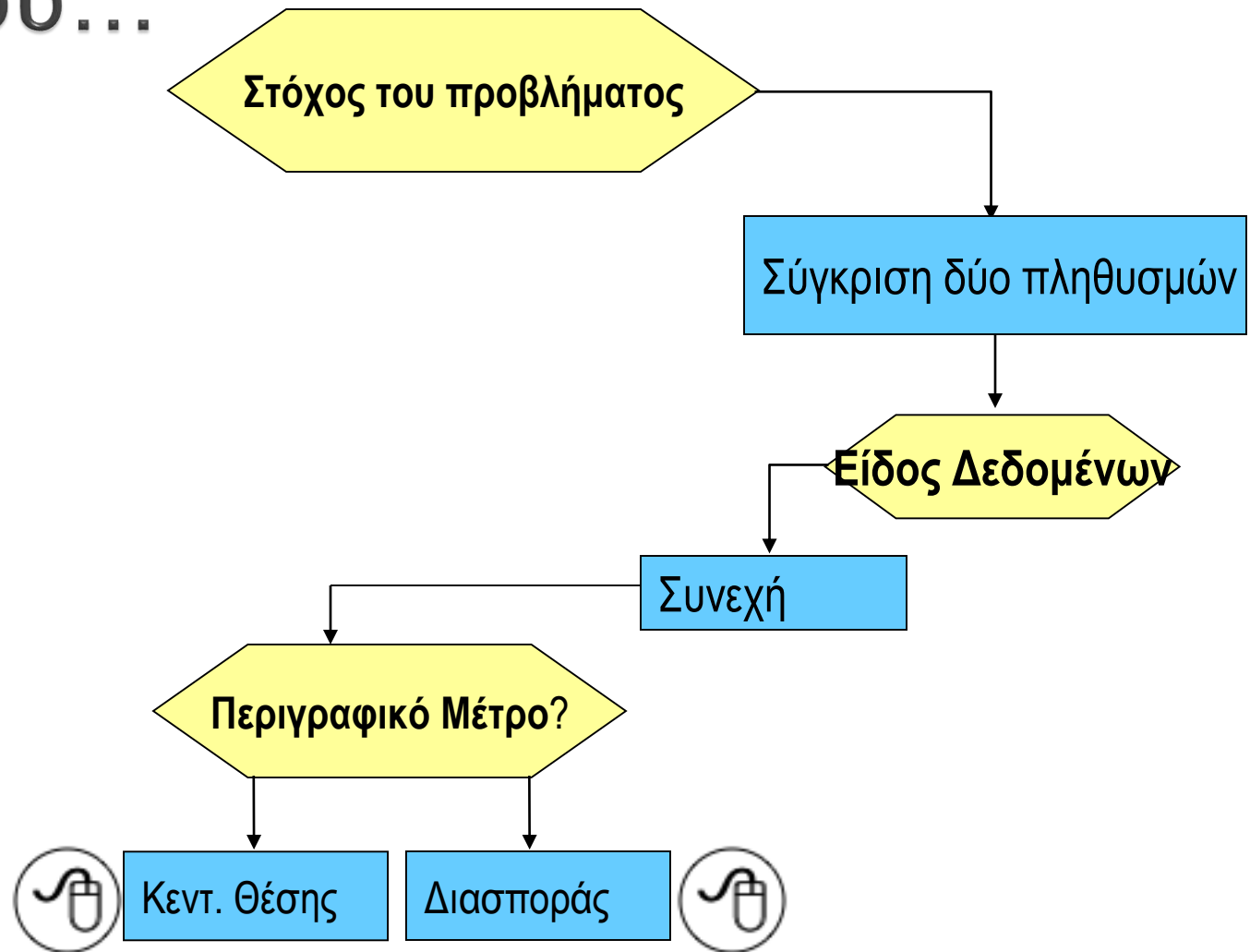
$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

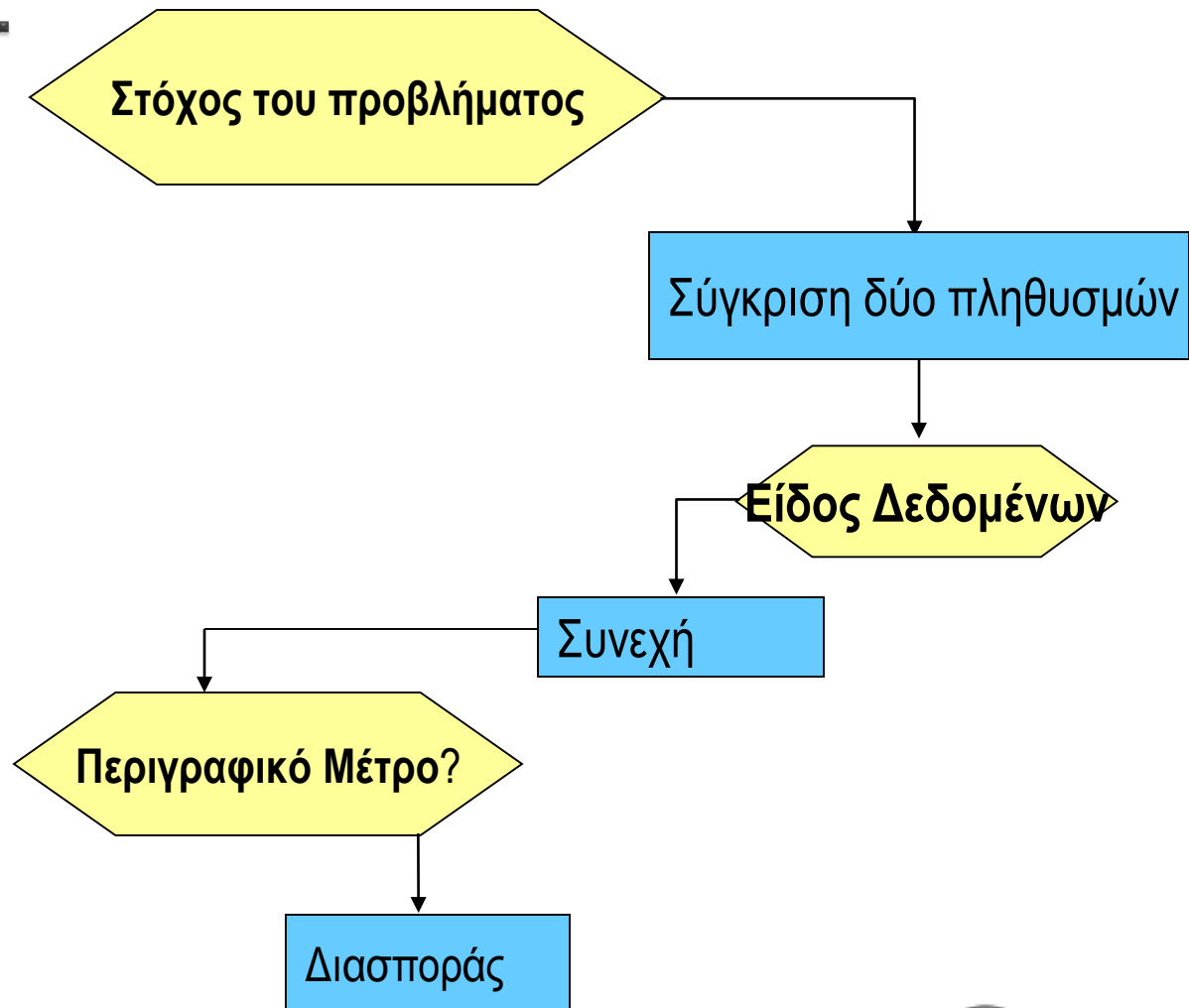
Interval Estimator: $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$


Required Condition: $n_1\hat{p}_1$, $n_2\hat{p}_2$, $n_1(1 - \hat{p}_1)$, and $n_2(1 - \hat{p}_2)$ are all ≥ 5

Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...



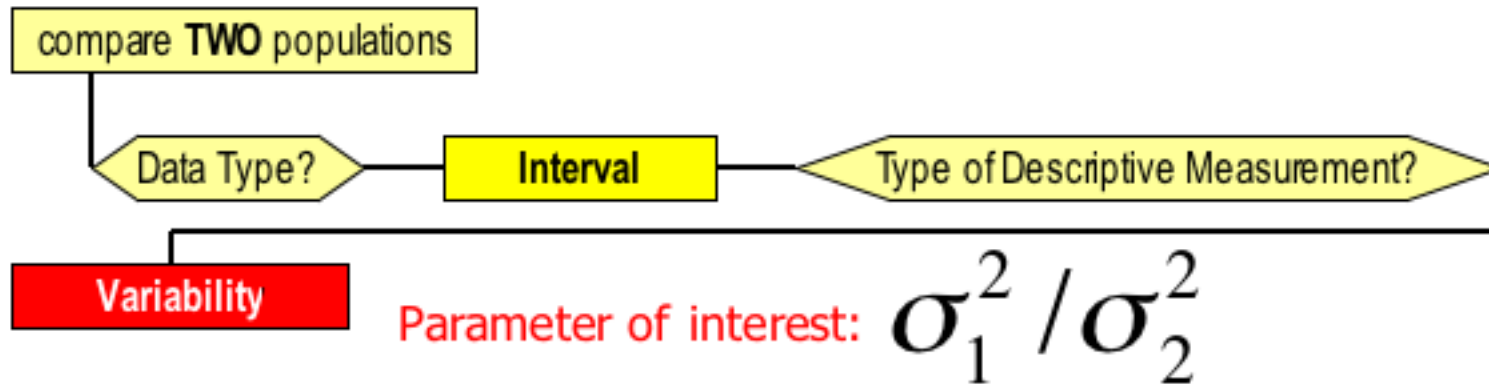
F-test and estimator of σ_1^2 / σ_2^2 

Μέθοδος...



Αρχή του
Διαγράμματος
Ροής

$$\sigma_1^2 / \sigma_2^2$$



Test Statistic:

$$F = s_1^2 / s_2^2$$

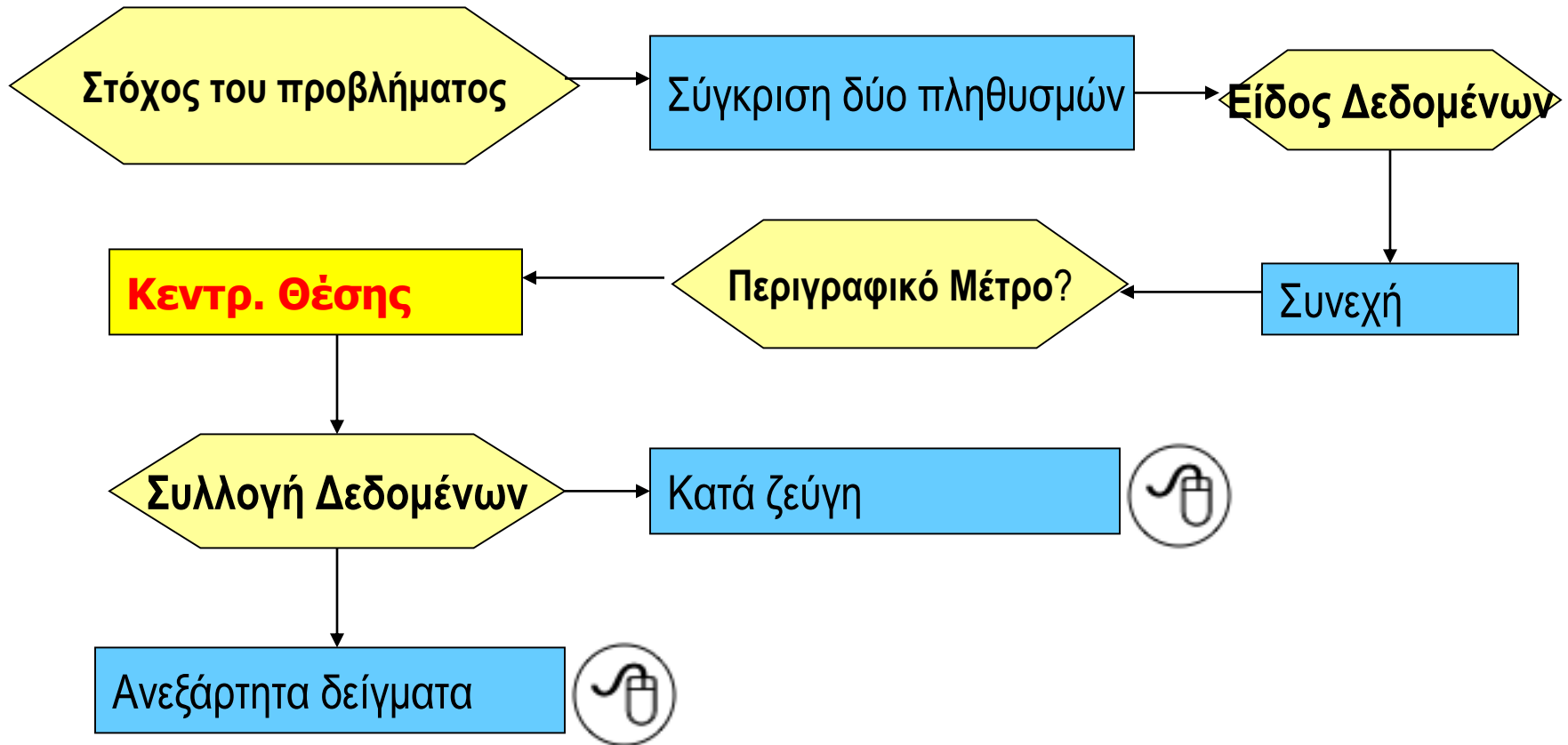
Interval Estimator:

$$LCL = \left(\frac{s_1^2}{s_2^2} \right) \frac{1}{F_{\alpha/2, \nu_1, \nu_2}}$$

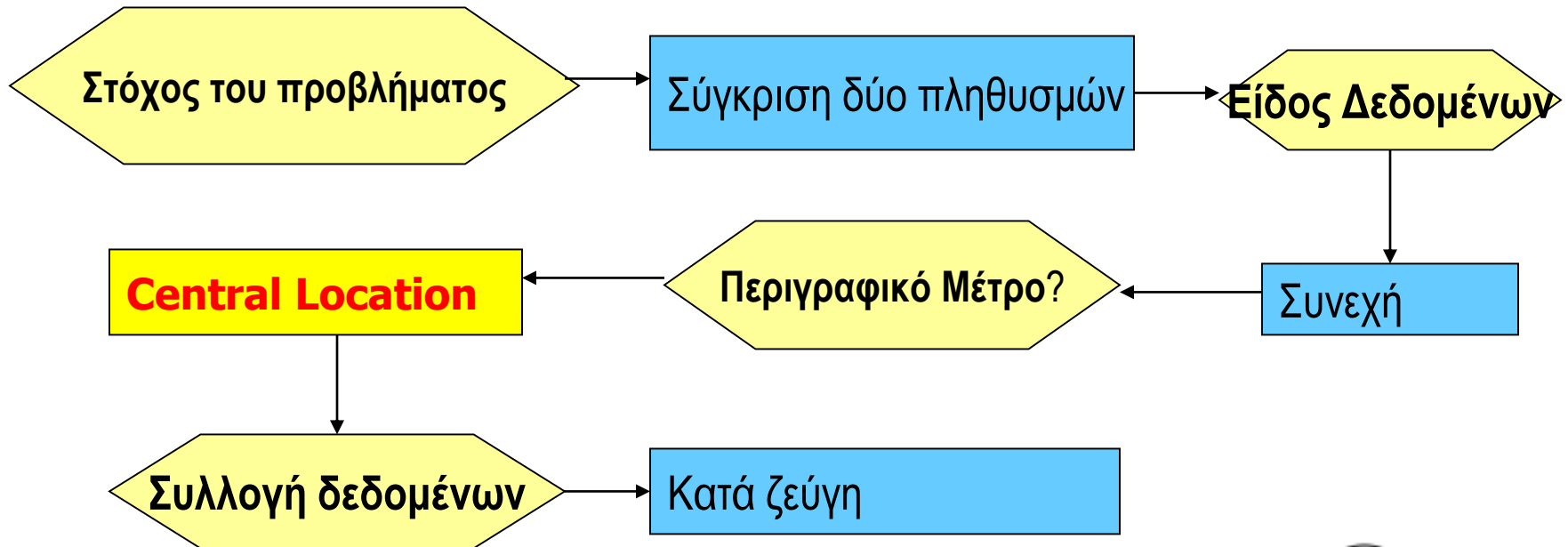
$$UCL = \left(\frac{s_1^2}{s_2^2} \right) F_{\alpha/2, \nu_2, \nu_1}$$

Required Condition: Populations are normal

Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...



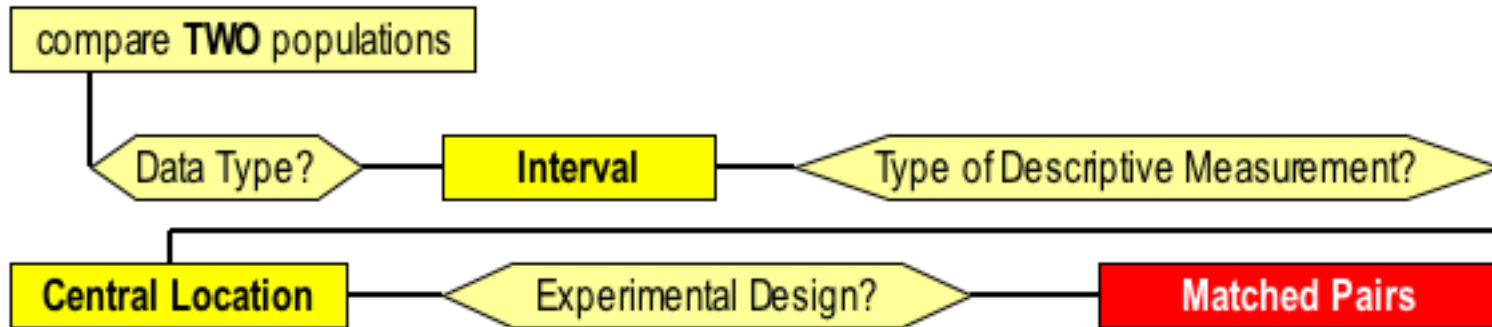
t-test and estimator of μ_D



Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Test Statistic:

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$$

Parameter of interest:

μ_D

Interval Estimator:

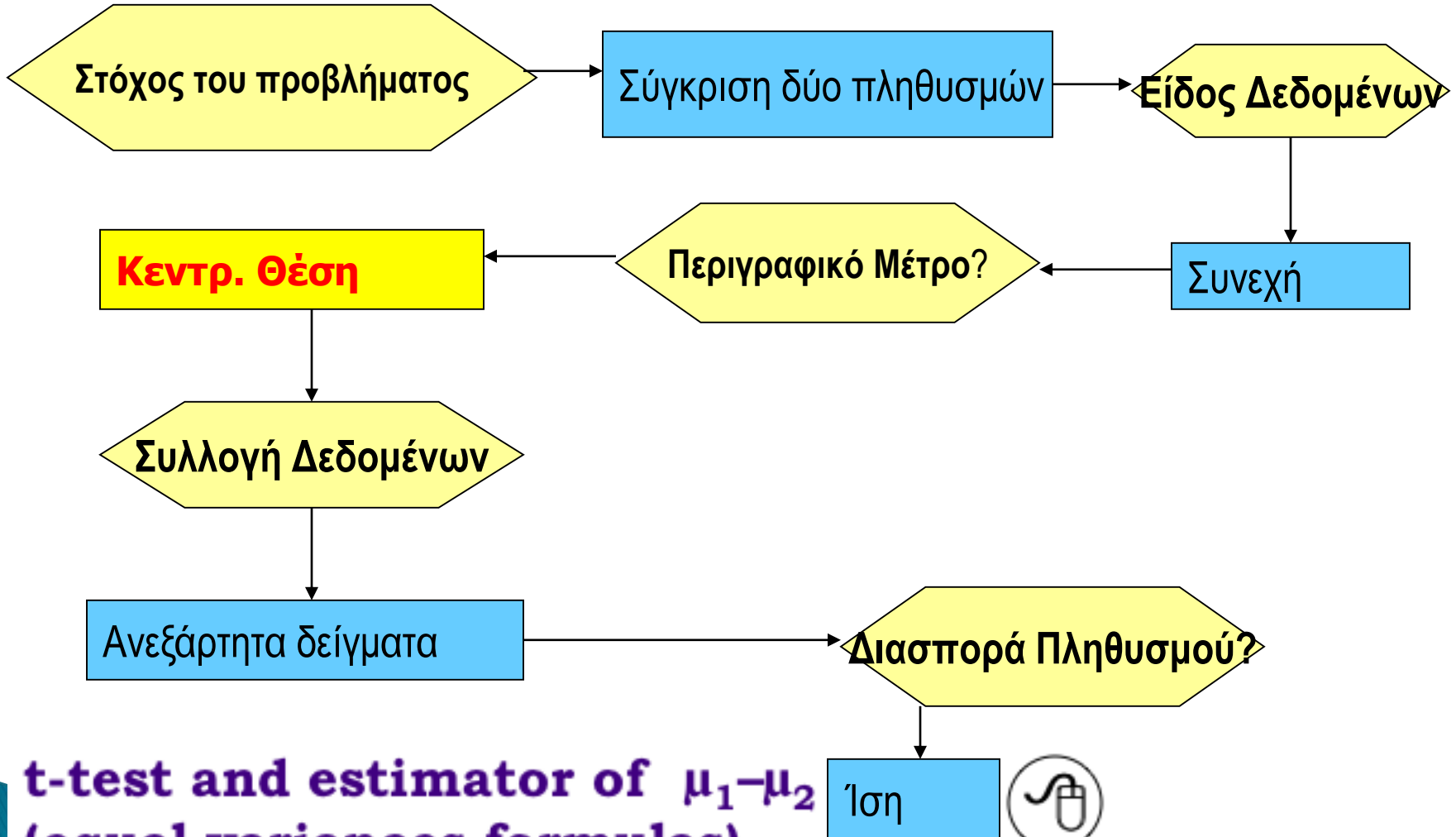
$$\bar{x}_D \pm t_{\alpha/2} \frac{s_D}{\sqrt{n_D}}$$

Required Condition: **Differences** are normal

Διάγραμμα ροής επιλογής της μεθόδου...



Διάγραμμα ροής επιλογής της μεθόδου...

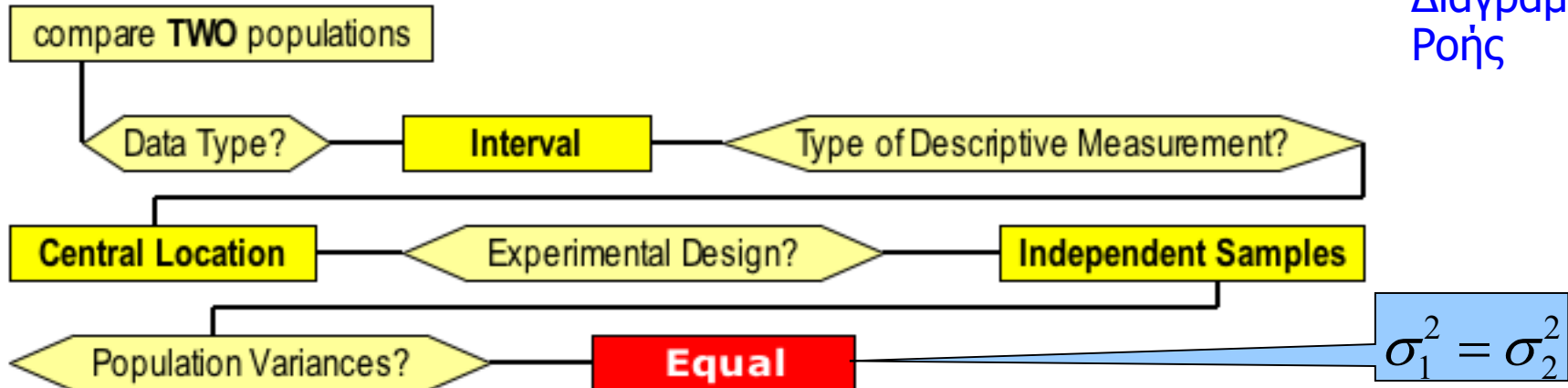


**t-test and estimator of $\mu_1 - \mu_2$
(equal variances formulas)**

Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Parameter of interest: $\mu_1 - \mu_2$

Test Statistic:

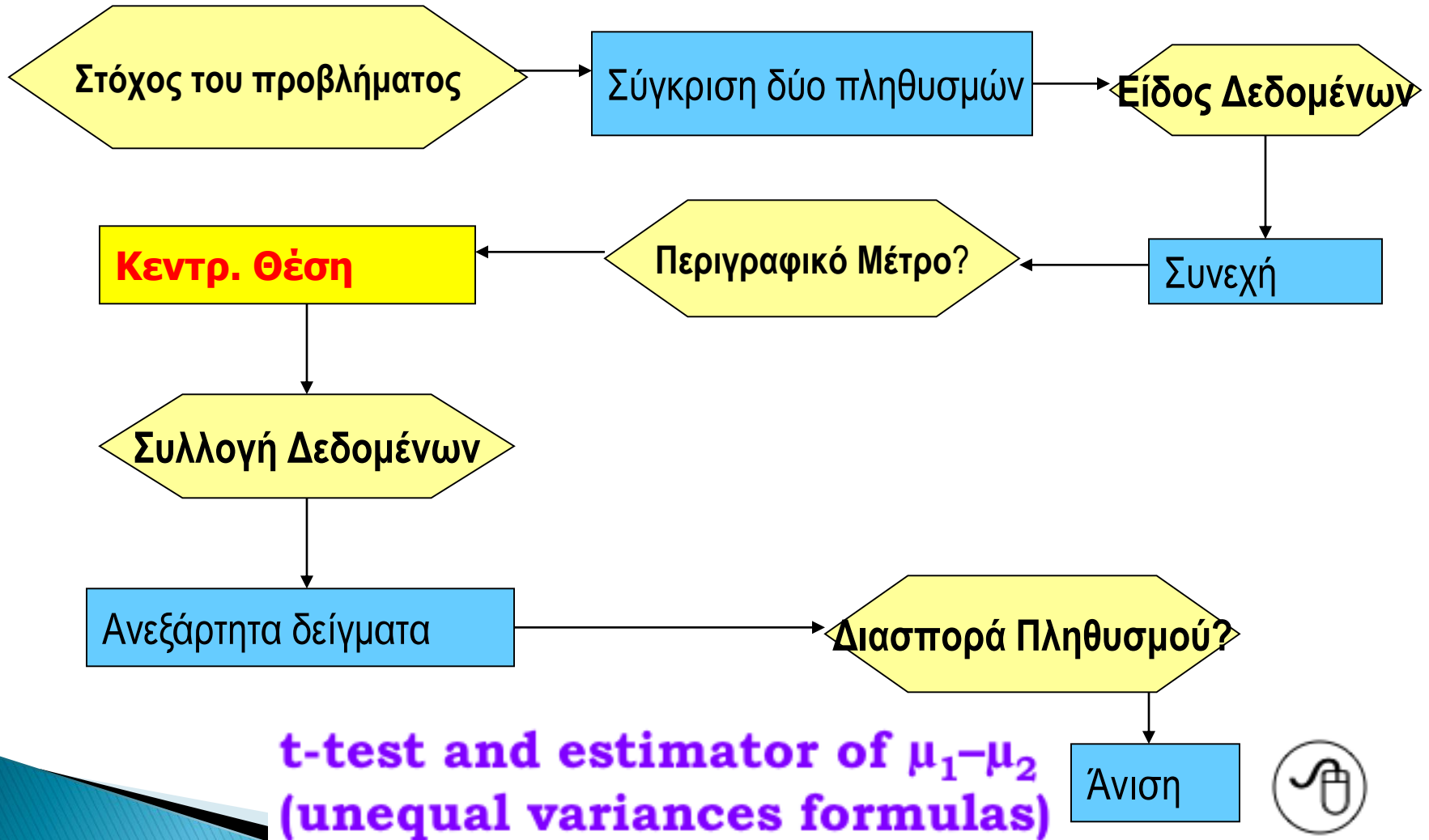
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Interval Estimator:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Required Condition: Populations are normal

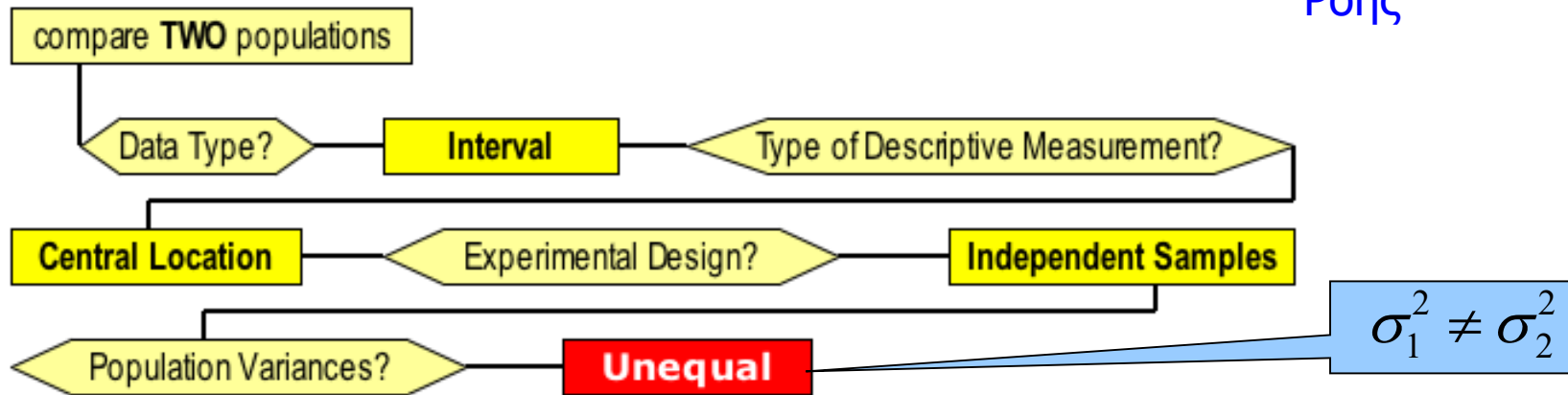
Διάγραμμα ροής επιλογής της μεθόδου...



Μέθοδος...



Αρχή του
Διαγράμματος
Ροής



Parameter of interest: $\mu_1 - \mu_2$

Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Interval Estimator:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

Required Condition: Populations are normal