ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΓΕΩΠΟΝΙΚΗ ΣΧΟΛΗ
Α.Π.Θ.

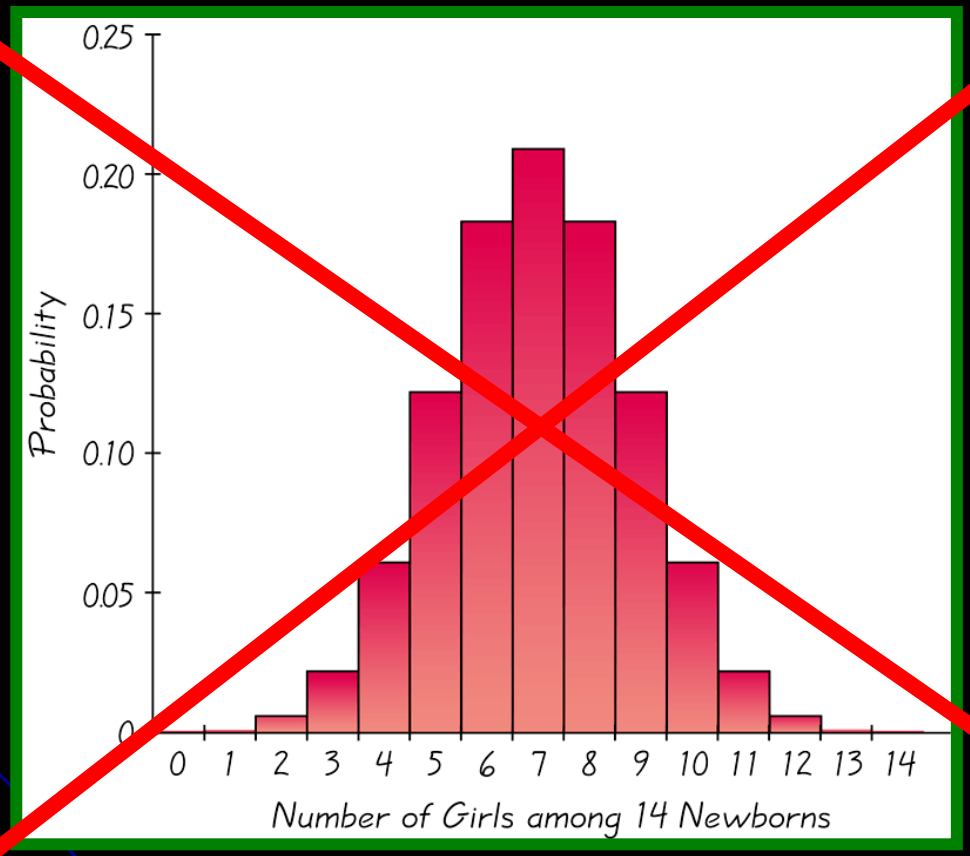
Μη Παραμετρικοί Έλεγχοι

Επιστημονική Επιμέλεια:

Δρ. Γεώργιος Μενεξές

Τομέας Φυτών Μεγάλης Καλλιέργειας
και Οικολογίας
Εργαστήριο Γεωργίας

Καταρχήν…


0.25
0.20
0.15
0.10
0.05
0
Probability
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Number of Girls among 14 Newborns

Μη Παραμετρικοί Έλεγχοι

- Δεν απαιτούν κανονικότητα στις κατανομές των μεταβλητών.
- Η κλίμακα μέτρησης των μεταβλητών πρέπει να είναι τουλάχιστον διάταξης (ordinal).
- Μπορούν να εφαρμοστούν όταν τα δείγματα είναι μικρά.
- Δεν απαιτούνται εκτιμήσεις των παραμέτρων των κατανομών.
- Εφαρμόζονται στις περιπτώσεις που δεν μπορούν να εφαρμοστούν έλεγχοι της Παραμετρικής Στατιστικής.

Παραδείγματα Ελέγχων

Έλεγχος Kruskal-Wallis (1)

Ο έλεγχος *Kruskal-Wallis* είναι Μη Παραμετρική διαδικασία που μπορεί να εφαρμοστεί για τη σύγκριση τριών ή περισσότερων πληθυσμών.

Πιο συγκεκριμένα:

H_o: Τα *k* δείγματα προέρχονται από τον ίδιο πληθυσμό.

H_1: Τουλάχιστον ένα δείγμα προέρχεται από διαφορετικό πληθυσμό.

σε ε.σ. α

Έλεγχος Kruskal-Wallis (2)

- Αν η μηδενική υπόθεση απορριφθεί τότε τουλάχιστον δύο πληθυσμοί διαφέρουν στατιστικά σημαντικά ως προς τη διάμεση τιμή τους (και με χαλαρή ερμηνεία ως προς τους αντίστοιχους μέσους όρους).

Προεργασία (1)

- Τα k δείγματα-ομάδες ενοποιούνται σε ένα μεγάλο δείγμα-ομάδα μεγέθους n, $n=n_1+n_2+\ldots+n_k$
- Για κάθε μία από τις n μετρήσεις–τιμές-παρατηρήσεις υπολογίζεται η βαθμίδα της (τάξη ή σειρά κατάταξης-rank). Δηλαδή, στις αρχικές μετρήσεις αποδίδονται βαθμοί από το 1 μέχρι το n. Το 1 αποδίδεται στη μέτρηση-τιμή-παρατήρηση με τη μικρότερη αριθμητική τιμή. Σε περίπτωση ισοπαλίας-δεσμού (tie) αποδίδεται ο μέσος όρος των αρχικών βαθμίδων. Υπάρχει και διορθωμένη εκδοχή του ελέγχου σε περίπτωση ισοπαλιών.
- Για κάθε δείγμα-ομάδα υπολογίζεται το άθροισμα T_i των βαθμίδων των αντίστοιχων μετρήσεων-τιμών-παρατηρήσεων.

Ισοπαλίες-Δεσμοί (1)

- Όταν η μεταβλητή που εξετάζεται έχει **πολλαπλές τιμές**. Δηλαδή, ορισμένες τιμές της επαναλαμβάνονται δύο ή περισσότερες φορές.
- Καταγράφουμε:
  - Το **πλήθος των πολλαπλών τιμών** (π.χ. 10 από τις 100 τιμές επαναλαμβάνονται δύο ή περισσότερες φορές).
  - Την **πολλαπλότητα** ρ (πόσες φορές εμφανίζεται) της κάθε πολλαπλής τιμής (π.χ. αν η τιμή 38 έχει πολλαπλότητα 4, εμφανίζεται 4 φορές ή 4 τιμές της εξεταζόμενης μεταβλητής είναι ίσες με 38).

Ισοπαλίες-Δεσμοί (2)

3	5	5	10	12	Πραγματικές τιμές
1	2,5	2,5	4	5	Βαθμίδες

2 και 3 έχουν ισοβαθμία (ισοπαλία, δεσμός)

Παράδειγμα ανάθεσης βαθμίδας (χωρίς ισοπαλίες-δεσμούς)

*ranks.sav [DataSet2] - SPSS Data Editor

21 : weight

	weight	var	var
1	3504		
2	3693		
3	3436		
4	3433		
5	3449		
6	4341		
7	4354		
8	4312		
9	4425		
10	3850		
11	3090		
12	4142		
13	4034		
14	4166		
15	3850		
16	3563		
17	3609		
18	3353		
19	3761		
20	3086		
21			

*ranks.sav [DataSet2] - SPSS Data Editor

21 : weight

	weight	Rank
1	3086	1
2	3090	2
3	3353	3
4	3433	4
5	3436	5
6	3449	6
7	3504	7
8	3563	8
9	3609	9
10	3693	10
11	3761	11
12	3850	13
13	3850	13
14	4034	14
15	4142	15
16	4166	16
17	4312	17
18	4341	18
19	4354	19
20	4425	20
21		

*ranks.sav [DataSet3] - SPSS Data Editor

1 : Rank 7 Visible

	weight	Rank	var
1	3504	7	
2	3693	10	
3	3436	5	
4	3433	4	
5	3449	6	
6	4341	18	
7	4354	19	
8	4312	17	
9	4425	20	
10	3850	13	
11	3090	2	
12	4142	15	
13	4034	14	
14	4166	16	
15	3850	13	
16	3563	8	
17	3609	9	
18	3353	3	
19	3761	11	
20	3086	1	
21			

Παράδειγμα ανάθεσης βαθμίδας (με ισοπαλίες-δεσμούς)

*ranks.sav [DataSet3] - SPSS Data Editor

	weight	var	var
1	3504		
2	3693		
3	3090		
4	3433		
5	3449		
6	4341		
7	4354		
8	4312		
9	4425		
10	3850		
11	3090		
12	4312		
13	4034		
14	4166		
15	3850		
16	3563		
17	4312		
18	3353		
19	3761		
20	3086		

Υπάρχουν 3 πολλαπλές τιμές: η 3090 με πολλαπλότητα 2, η 4312 με πολλαπλότητα 3 και η 3850 με πολλαπλότητα 2

*ranks.sav [DataSet3] - SPSS Data Editor

	weight	Rweight	var
1	3086	1,0	
2	3090	2,5	
3	3090	2,5	
4	3353	4,0	
5	3433	5,0	
6	3449	6,0	
7	3504	7,0	
8	3563	8,0	
9	3693	9,0	
10	3761	10,0	
11	3850	11,5	
12	3850	11,5	
13	4034	13,0	
14	4166	14,0	
15	4312	16,0	
16	4312	16,0	
17	4312	16,0	
18	4341	18,0	
19	4354	19,0	
20	4425	20,0	

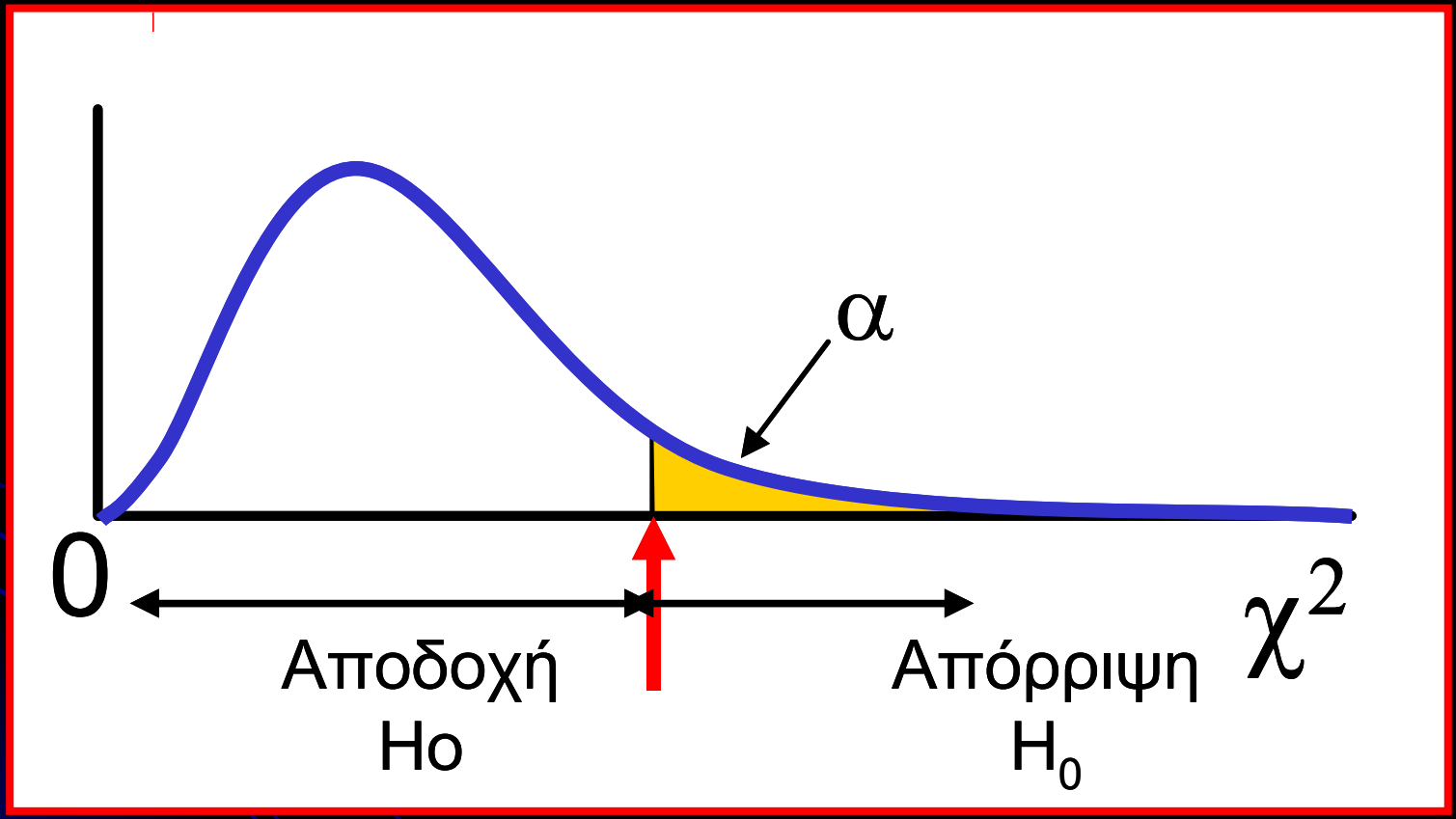
Διαδικασία

- Υπολογίζεται το στατιστικό:

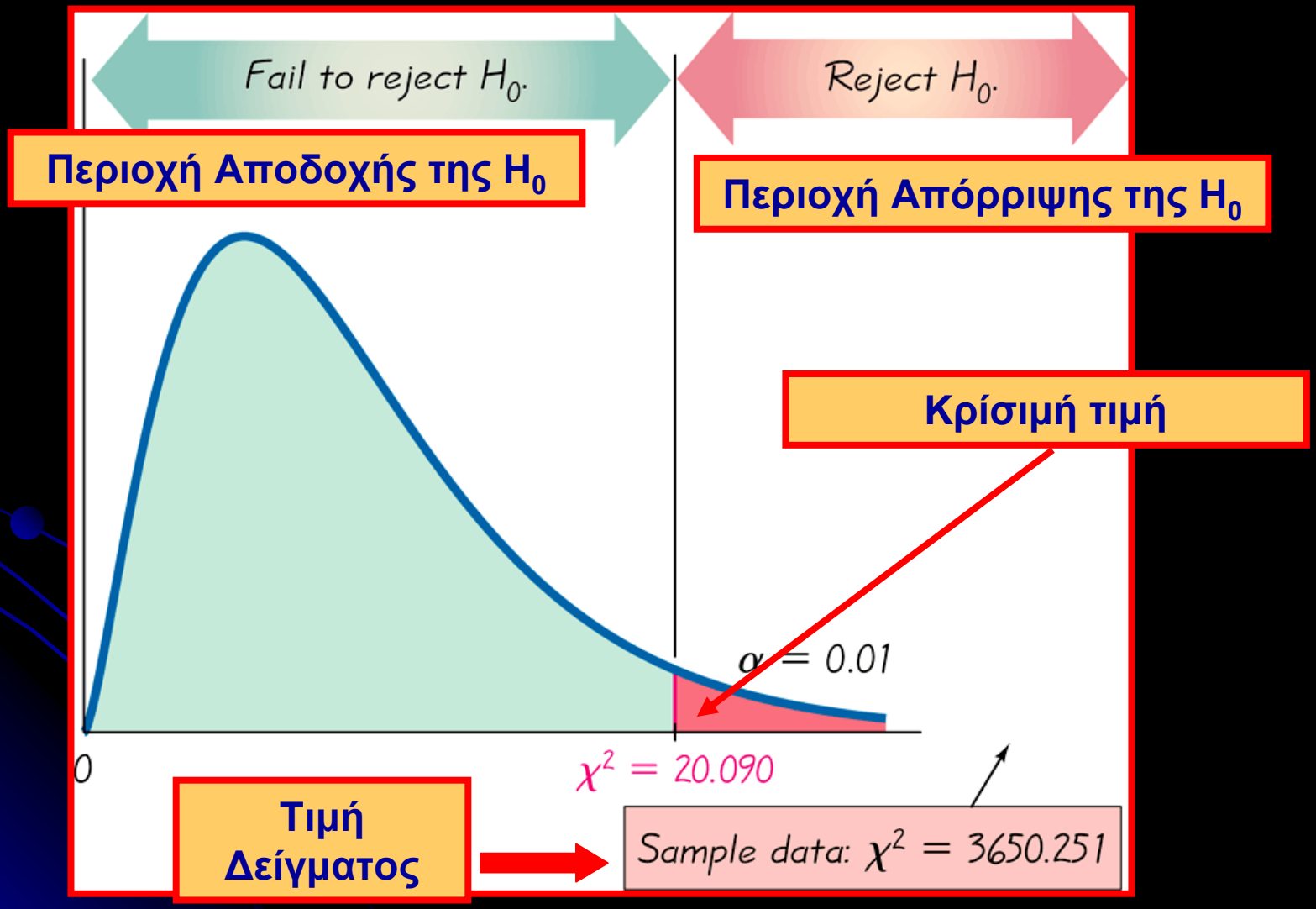
$$H = \frac{12}{n(n+1)} \sum \frac{T_i^2}{n_i} - 3(n+1)$$

- Κάτω από την ορθότητα της μηδενικής υπόθεσης το στατιστικό H ακολουθεί προσεγγιστικά X^2 κατανομή με k-1 β.ε.
- Η μηδενική υπόθεση απορρίπτεται σε ε.σ. α=0,05 αν $H > X^2_{0,05}$ για k-1 β.ε. Η προσέγγιση είναι ικανοποιητική αν κάθε ομάδα-δείγμα έχει τουλάχιστον 5 μετρήσεις-τιμές-παρατηρήσεις.

Περιοχή Απόρριψης της Ho



Παράδειγμα


Fail to reject H_0.
Reject H_0.
Περιοχή Αποδοχής της H_0
Περιοχή Απόρριψης της H_0
Κρίσιμη τιμή
$\alpha = 0.01$
0
$\chi^2 = 20.090$
Τιμή Δείγματος
Sample data: $\chi^2 = 3650.251$

Προϋποθέσεις Εφαρμογής

- Δείγματα τυχαία
- Παρατηρήσεις ανεξάρτητες
- Κλίμακα μέτρησης της μεταβλητής τουλάχιστον διάταξης (ordinal)
- Συμμετρικές κατανομές μέσα σε κάθε ομάδα-δείγμα

Παρατηρήσεις (1)

- Αποτελεί τη Μη Παραμετρική εκδοχή της one-way ANOVA.
- Εφαρμόζεται σε Πλήρως Τυχαιοποιημένα Πειραματικά Σχέδια.
- Εν γένει, έχει μικρότερη ισχύ (power) από την Παραμετρική ANOVA.
- Δεν απαιτεί κανονικότητα στην κατανομή των μετρήσεων μέσα σε κάθε ομάδα.

Παρατηρήσεις (2)

- Δεν έχουμε εκτιμήσεις των παραμέτρων των αντίστοιχων πληθυσμών.
- Χρησιμοποιεί λιγότερη πληροφορία από ότι η Παραμετρική ANOVA.
- Μπορεί να εφαρμοστεί σε περιπτώσεις που τα δείγματα είναι μικρά σε μέγεθος.
- Η παρατηρούμενη στάθμη σημαντικότητας του ελέγχου (*p*-value) μπορεί να υπολογιστεί είτε με τη μέθοδο Exact είτε με τη μέθοδο προσομοίωσης Monte-Carlo (SPSS, Exact tests)

Παρατηρήσεις (3)

Ο στατιστικός έλεγχος συχνά εμφανίζεται και ως:

H_o: $\Delta T_1 = \Delta T_2 = \ldots = \Delta T_k$

H_1: Τουλάχιστον δύο Διάμεσες Τιμές (ΔT) διαφέρουν

σε ε.σ. α

Παράδειγμα

- Σε 4 ποικιλίες ενός φυτού μετρήθηκαν οι τιμές ενός αγρονομικού χαρακτηριστικού. Τα στοιχεία δίνονται στον παρακάτω πίνακα. Διαφέρουν στατιστικά σημαντικά οι 4 ποικιλίες σε ε.σ. α=0,05 ως προς το αγρονομικό χαρακτηριστικό;

Παράδειγμα (*συνέχεια*)

Πίνακας Δεδομένων

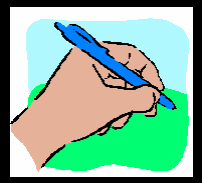
Π1	Π2	Π3	Π4
65	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88

Παράδειγμα (*συνέχεια*)

Βοηθητικοί Υπολογισμοί

	Π1	**Π2**	**Π3**	**Π4**
	65 (3)	75 (7)	59 (1)	94 (16)
	87 (13)	69 (5)	78 (8)	89 (15)
	73 (6)	83 (12)	67 (4)	80 (10)
	79 (9)	81 (11)	62 (2)	88 (14)
T_i	31	35	15	55

Παράδειγμα (*συνέχεια*)



$$H = \frac{12}{n(n+1)} \sum \frac{T_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{16(17)} \left(\frac{31^2 + 35^2 + 15^2 + 55^2}{4} \right) - 3(17) = 8,96$$

- Η κρίσιμη της X^2 κατανομής σε ε.σ. α=0,05 για 4-1=3 β.ε. είναι 7,81.
- Απόφαση:

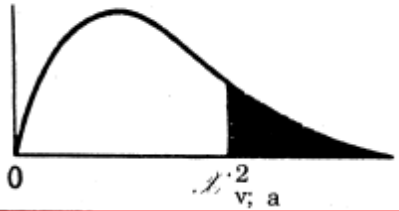
8,96>7,81→ Απορρίπτεται η H_o.

Συμπέρασμα: Υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των 4 ποικιλιών (τουλάχιστον δύο διαφέρουν).

Παράδειγμα (*συνέχεια*)

Πίνακας

Τιμών $\chi^2_{\nu;\,a}$ *της* χ^2 *κατανομής για τις οποίες* $P(X^2 > \chi^2_{\nu;\,a}) = a$



a = .10	a = .05	a = .025	a = .010	a = .005	β.ε.
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9

Παράδειγμα με Ισοπαλίες-Δεσμούς

(2+3)/2

X	Δείγμα	Βαθμίδα	X	Δείγμα	Βαθμίδα	X	Δείγμα	Βαθμίδα
33.45	IN	1	41.70	CA	11	47.78	CA	21
35.97	IN	2.5	42.05	IN	12.5	48.79	MD	22
35.97	IN	2.5	42.05	CA	12.5	48.80	MD	23
35.98	IN	4	43.20	CA	14	49.20	MD	24
37.10	CA	5	43.57	MD	15	49.68	MD	25
38.25	IN	6	43,85	CA	16	49.88	CA	26
38.65	IN	7	44.94	MD	17	52.94	CA	27
40.12	IN	8	44.98	CA	18	53.25	MD	28
40.37	MD	9	45.35	MD	19	53.82	MD	29
40.50	CA	10	45.70	IN	20			

3 ομάδες-δείγματα (IN, CA και MD)

Κατατάσσουμε τις μετρήσεις X σε αύξουσα σειρά

Αναθέτουμε βαθμίδες

Σύνοψη

Κριτήριο Kruskal - Wallis

H_0: τα k δείγματα προέρχονται από τον ίδιο πληθυσμό

H_1: όχι η H_0

Αντί του T_i

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) > \chi^2_{k-1;\,a}$$

όπου: n_i = το μέγεθος του i δείγματος, $i = 1, 2, \ldots, k$

k = το πλήθος των δειγμάτων

R_i = το άθροισμα των βαθμών του i δείγματος στο ενιαίο δείγμα

$n = n_1+n_2+\ldots+n_k$

Αν υπάρχουν δεσμοί: $H' = \frac{H}{c} > \chi^2_{k-1;\,a}$ όπου:

$$c = 1 - \frac{1}{n(n^2-1)} \sum_{i=1}^{\varrho} \mu_i\,(\mu_i^2-1)$$ όπου:

ϱ = το πλήθος των πολλαπλών τιμών

μ_i = η πολλαπλότητα της τιμής μ_i

Διόρθωση για δεσμούς

Παράδειγμα με το SPSS

1

Report

Miles per Gallon

Country of Origin	Minimum	Maximum	Mean	Median	Std. Deviation	N
American	10	39	20,13	18,55	6,377	248
European	16	44	27,89	26,50	6,724	70
Japanese	18	47	30,45	31,60	6,090	79
Total	10	47	23,55	23,00	7,792	397

2

Ranks

	Country of Origin	N	Mean Rank
Miles per Gallon	American	248	148,10
	European	70	264,56
	Japanese	79	300,70
	Total	397	

3

Test Statistics[b,c]

			Miles per Gallon
Chi-Square			133,793
df			2
Asymp. Sig.			,000
Monte Carlo Sig.	Sig.		,000[a]
	99% Confidence Interval	Lower Bound	,000
		Upper Bound	,000

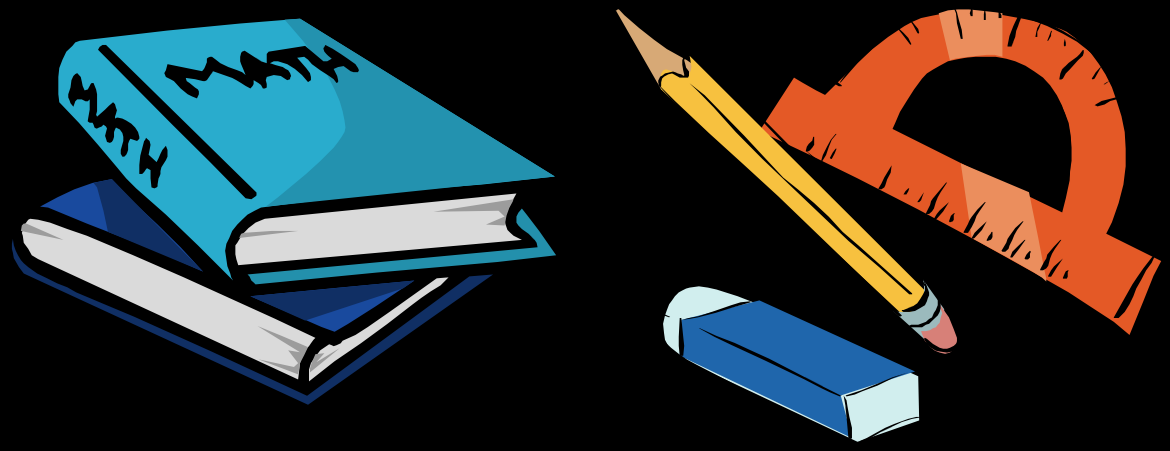
a. Based on 10000 sampled tables with starting seed 2000000.

b. Kruskal Wallis Test

c. Grouping Variable: Country of Origin

Αφού Sig. (*p*-value) =0,000 <*α*=0,05 η Μηδενική Υπόθεση Απορρίπτεται. Συνεπώς τουλάχιστον δύο χώρες διαφέρουν στατιστικά σημαντικά ως προς την κατανάλωση καυσίμου

Απορίες-Ερωτήσεις



Ο Συντελεστής Συσχέτισης των Τάξεων του Spearman (1)

Αν έχουμε n ζευγάρια τιμών (X_i, Y_i) των μεταβλητών X και Y αντίστοιχα τότε ο συντελεστής συσχέτισης των τάξεων rho του Spearman υπολογίζεται ως εξής:

① Υπολογίζουμε τις διαμοίρες των τιμών X_i
Υπολογίζουμε τις διαμοίρες των τιμών Y_i

Ο Συντελεστής Συσχέτισης των Τάξεων του Spearman (2)

(2) Υπολογίζουμε τις διαφορές $d_i = X_i' - Y_i'$, όπου X_i' οι τάξεις των X_i κ Y_i' οι τάξεις των Y_i

(3) Υπολογίζουμε τις d_i^2

(4) Υπολογίζουμε το άθροισμα $\sum d_i^2$.

(5) Αν δεν υπάρχουν ισοκατάταξεις - δεσμοί ή αν υπάρχουν είναι πολύ λίγες σε σχέση με το μέγεθος του δείγματος, τότε

rho ή $r_s = 1 - \frac{6 \sum d_i^2}{n.(n^2-1)}$

$$\text{rho ή } r_s = 1 - \frac{6.\sum d_i^2}{n(n^2-1)}$$

Ο Συντελεστής Συσχέτισης των Τάξεων του Spearman (3)

Αν υπάρχουν ισοψηφίες στα X_i' ή Y_i' τότε ο συντελεστής του Spearman δίνεται:

$$r_s \text{ ή } rho = \frac{\sum x^2 + \sum y^2 - \sum d_i^2}{2\sqrt{\sum x^2 \cdot \sum y^2}}$$

όπου:

$$\sum x^2 = \frac{n^3 - n}{12} - \frac{1}{12}\sum_j t_j(t_j^2 - 1), \text{ για την } X$$

$$\sum y^2 = \frac{n^3 - n}{12} - \frac{1}{12}\sum_h t_h(t_h^2 - 1), \text{ για την } Y.$$

με j ή h το πλήθος των ομάδων ισοψηφιών

ή t_j, t_h το μέγεθος κάθε ομάδας ισοψηφιών.

Ο Συντελεστής Συσχέτισης των Τάξεων του Spearman (4)

Ο Συντελεστής Συσχέτισης rho του Spearman παίρνει τιμές από $\boxed{-1 \text{ ως } 1.}$

Έλεγχος Σημαντικότητας

$H_0: \rho = 0$

$H_1: \rho \neq 0$, ή $\rho > 0$ ή $\rho < 0$ σε $\alpha = 0{,}05$

* Αν $4 \leq n \leq 30$ τότε η κρίσιμη τιμή του ελέγχου βρίσκεται από πίνακες.

Ο Συντελεστής Συσχέτισης των Τάξεων του Spearman (5)

** Γενικά, αν $10 \le n \le 30$ τότε:

$$r_S \cdot \sqrt{\frac{n-2}{1-r_S^2}} \sim t(\nu = n-2 \text{ β.ε.})$$

*** Αν $n > 30$ τότε

$$r_S \sim N\left(0, \frac{1}{n-1}\right)$$

Συμπεράσματα λογικά

* Αν r_S μεγαλύτερο ή ίσο από την κρίσιμη τιμή $|r_S(\alpha)|$ ή $|r_S(\alpha/2)|$ ανάλογα με τον έλεγχο, απορρίπτουμε την H_0.

** ή *** χρησιμοποιούμε κρίσιμες τιμές για τους ελέγχους t ή Z αντίστοιχα.

Παράδειγμα

Στον παρακάτω πίνακα καταγράφεται ο αριθμός των τσιγάρων/ημέρα που κάπνιζαν 15 γυναίκες κατά τη διάρκεια της εγκυμοσύνης τους, οι οποίες είχαν την ίδια σωματική διάπλαση, και το βάρος αντίστοιχα των παιδιών τους την ημέρα του τοκετού. Να υπολογιστεί ο συντελεστής συσχέτισης του τάξεως του Spearman. Να ερμηνευτεί η να δοκιμαστεί η στατιστική του σημαντικότητα σε ε.σ. $\alpha = 0,05$. Υποθέτουμε ότι θα υπάρχει αρνητική συσχέτιση.

Παράδειγμα (συνέχεια)

Επομένως ο πληθυσμός θα είναι

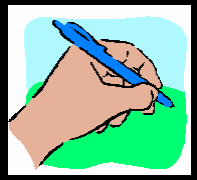
$\boxed{\text{μηδενικός}}$

$H_0: \rho = 0$

$H_1: \rho < 0$

σε $\alpha = 0{,}05$.

Παράδειγμα (*συνέχεια*)

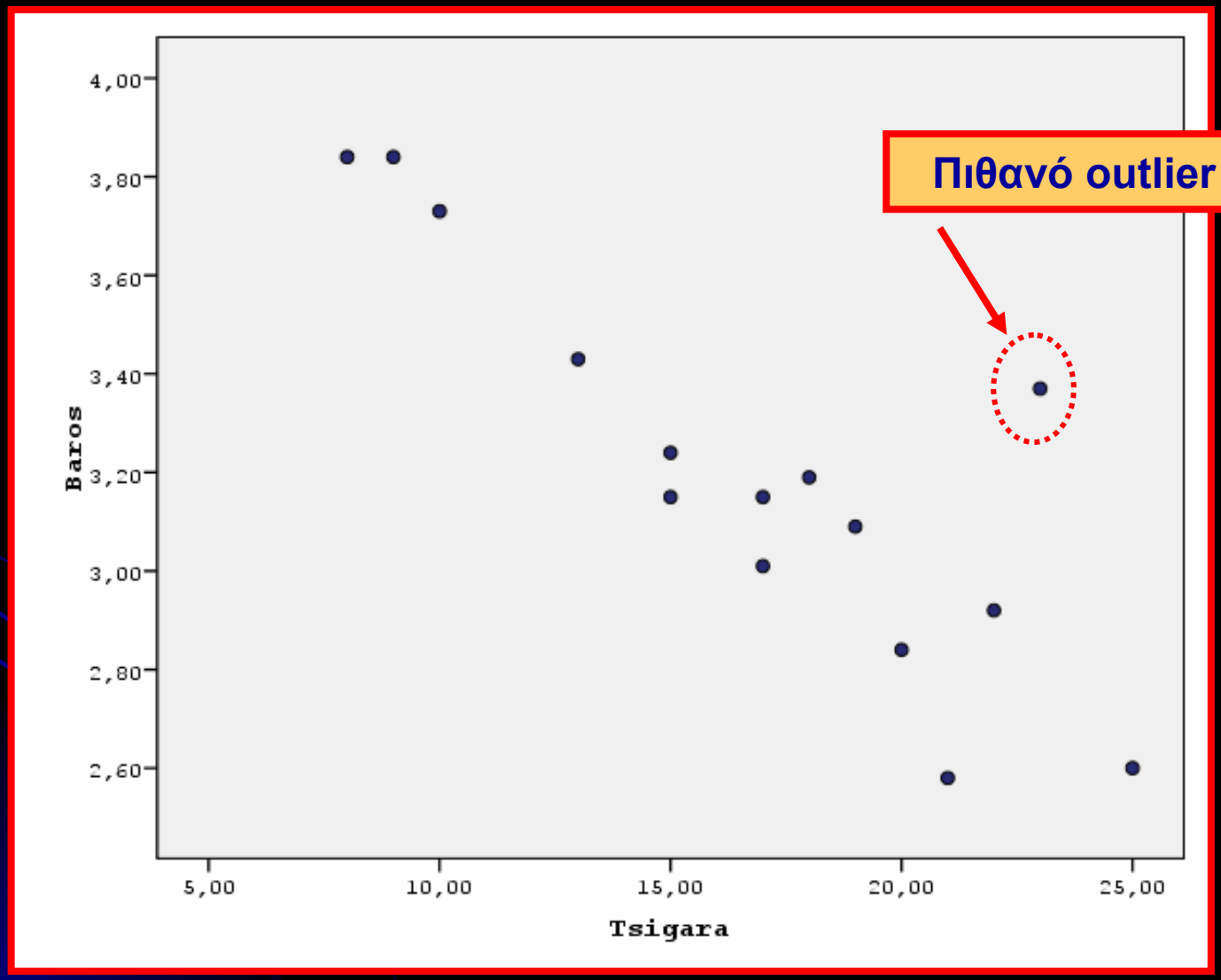


Α/Α	X_i	X_i' R_i	Y_i	Y_i' R_i	$d_i = X_i' - Y_i'$	d_i^2
1	10	3	3,73	13	-10	100
2	15	5,5	3,24	10	-4,5	20,25
3	17	7,5	3,15	7,5	0	0
4	25	15	2,60	1	16	256
5	13	4	3,43	12	-8	64
6	17	7,5	3,01	5	2,5	6,25
7	20	11	2,84	4	7	49
8	19	10	3,09	6	4	16
9	18	9	3,19	9	0	0
10	9	2	3,84	14,5	-12,5	156,25
11	15	5,5	3,15	7,5	-2	4
12	23	14	3,37	11	3	9
13	8	1	3,84	14,5	-13,5	182,25
14	22	13	2,92	3	10	100
15	21	12	2,58	2	10	100
						1063

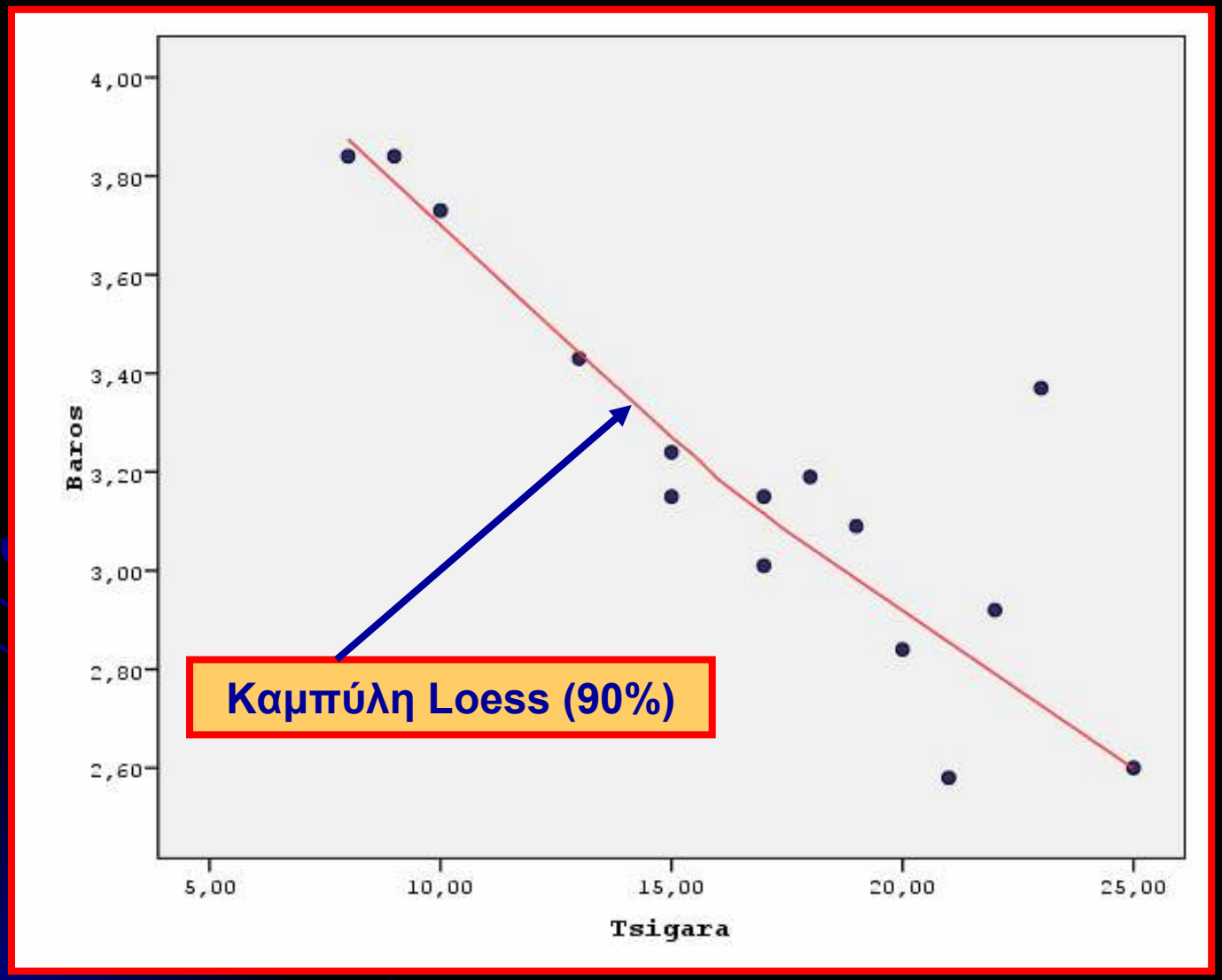
X_i = Αριθμός Τσιγάρων

Y_i = Βάρος παιδιών (Kg).

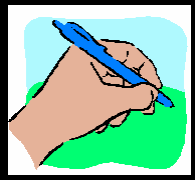
Οπτικοποίηση της σχέσης (1)



Οπτικοποίηση της σχέσης (2)



Παράδειγμα (συνέχεια)



(2)

Αγνοώντας τις ισοψυχίες:

$$r_s = 1 - \frac{6 \cdot 1063}{15 \cdot (15^2 - 1)} = -0{,}898$$

Λαμβάνοντας υπόψη τις ισοψυχίες:

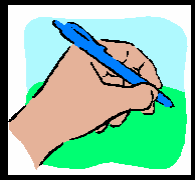
$$\sum x^2 = \frac{15^3 - 15}{12} - \frac{1}{12}\left(2(2^2-1) + 2(2^2-1)\right) = 279$$

$$\sum y^2 = \frac{15^3 - 15}{12} - \frac{1}{12}\left(2(2^2-1) + 2(2^2-1)\right) = 279$$

η ο συντελεστής διορθώνεται σε:

$$r_s = \frac{279 + 279 - 1063}{2\sqrt{279 \cdot 279}} = -0{,}905$$

Παράδειγμα (συνέχεια)



Επειδή $10 < n = 15$ έχουμε ότι

$$z = (-0,905)\sqrt{\frac{15-2}{(1-(0,905)^2}} = -7,67$$

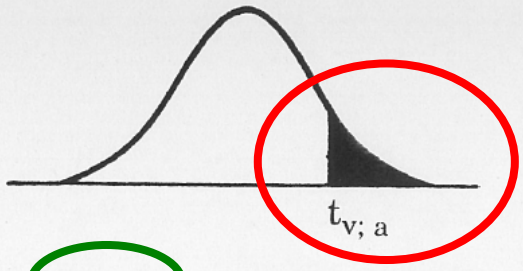
Η κρίσιμη τιμή $z_{0,05;13} = -1,771$.

Επειδή $z = -7,67 < -1,771$ ή

$|t| > 1,771$ απορρίπτεται η

H_0.

Παράδειγμα (*συνέχεια*)



β.ε.	a = .10	a = .05	a = .025	a = .010	a = .005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947

Παράδειγμα (συνέχεια)

(3)

Συμπεράσματα:

Βρήκαμε πολύ ισχυρή και απόλυτα αρνητική και στατιστικά σημαντική συσχέτιση των τάξεων μεταξύ αρ. τεταρτημορίων ή λάρος μωρού ($r_S = -0,905$).

Παρατήρηση

Η διόρθωση για ισοπαλίες – ισοψηφίες δεν μετέβαλε σημαντικά τον αρχικό υπολογισμό του r_S.

Παράδειγμα με το SPSS (1)

Data Example MR.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : x2 897,3

	var	y	x1	x2	x3	x4	var
1	Georgie	50331,00	52,70	897,30	23,85	37189,00	
2	Carina	44208,00	59,22	880,00	22,50	39044,00	
3	Igri	45779,00	58,52	866,60	21,40	47828,00	
4	Plaisant	42894,00	57,42	440,00	36,17	38133,00	
5	Sonjia	34688,00	56,49	702,60	18,35	45128,00	
6	Arma	45575,00	55,11	549,30	32,70	36378,00	
7	Rihane	47283,00	50,37	580,00	33,15	39172,00	
8	Beecher	41975,00	50,73	578,60	28,69	45756,00	
9	Assala	43472,00	51,40	610,60	29,98	43600,00	
10	CP/Apm	43735,00	54,56	733,30	19,79	44900,00	
11	WI 2291	45156,00	54,84	902,60	19,06	41494,00	
12	Matico s	32662,00	50,35	545,30	29,42	35517,00	
13	Trombill	42618,00	52,84	814,60	19,72	37033,00	
14	Γ-016252	44286,00	59,20	721,30	21,82	44217,00	
15	Oglow	49617,00	57,71	846,60	24,81	43728,00	
16	Zenit	46227,00	53,75	681,30	30,40	42378,00	
17	Γ-02020	41947,00	54,45	654,60	21,68	41228,00	
18							
19							
20							
21							

Data View / Variable View

SPSS

Δείκτης Συγκομιδής	y
Εκατολιτρικό Βάρος	x1
Στάχυα m^2	x2
Αρ. κόκκων / στάχυ	x3
Βάρος 1000 κόκκων	x4

Παράδειγμα με το SPSS (2)

Συντελεστές Συσχέτισης των Τάξεων *Spearman*

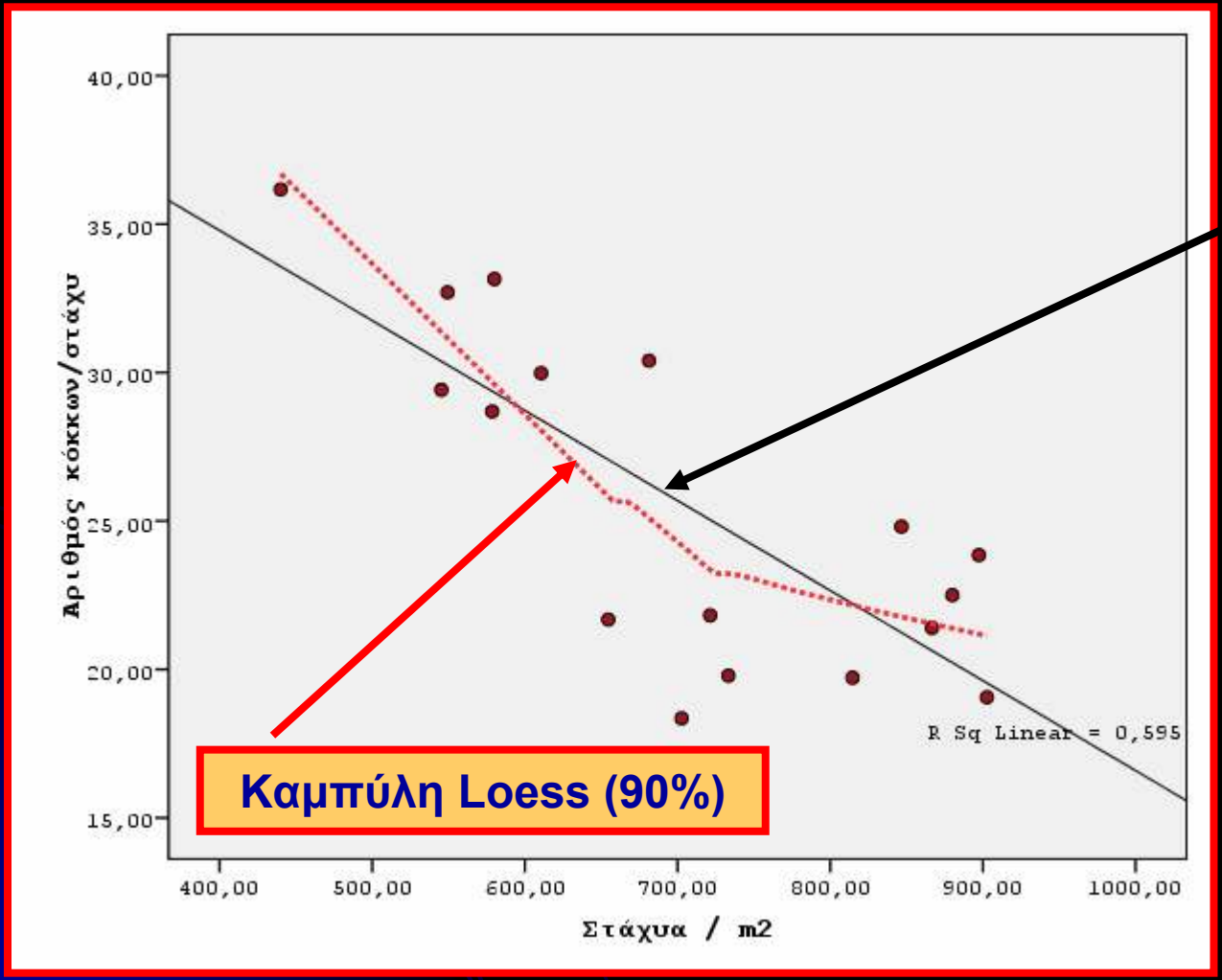
Correlations

			Δείκτης Συγκομιδής	Εκατολιτρικό βάρος	Στάχυα / m2	Αριθμός κόκκων/στάχυ	Βάρος κόκκων
Spearman's rho	Δείκτης Συγκομιδής	Correlation Coefficient	1,000	,194	,449	,238	,022
		Sig. (2-tailed)	.	,456	,071	,358	,933
		N	17	17	17	17	17
	Εκατολιτρικό βάρος	Correlation Coefficient	,194	1,000	,404	-,299	,289
		Sig. (2-tailed)	,456	.	,107	,244	,260
		N	17	17	17	17	17
	Στάχυα / m2	Correlation Coefficient	,449	,404	1,000	-,681**	,228
		Sig. (2-tailed)	,071	,107	.	,003	,379
		N	17	17	17	17	17
	Αριθμός κόκκων/στάχυ	Correlation Coefficient	,238	-,299	-,681**	1,000	-,375
		Sig. (2-tailed)	,358	,244	,003	.	,138
		N	17	17	17	17	17
	Βάρος κόκκων	Correlation Coefficient	,022	,289	,228	-,375	1,000
		Sig. (2-tailed)	,933	,260	,379	,138	.
		N	17	17	17	17	17

**. Correlation is significant at the 0.01 level (2-tailed).

Στατιστικά Σημαντική Αρνητική Συσχέτιση
p=0,003<α=0,05

Παράδειγμα με το SPSS (3)

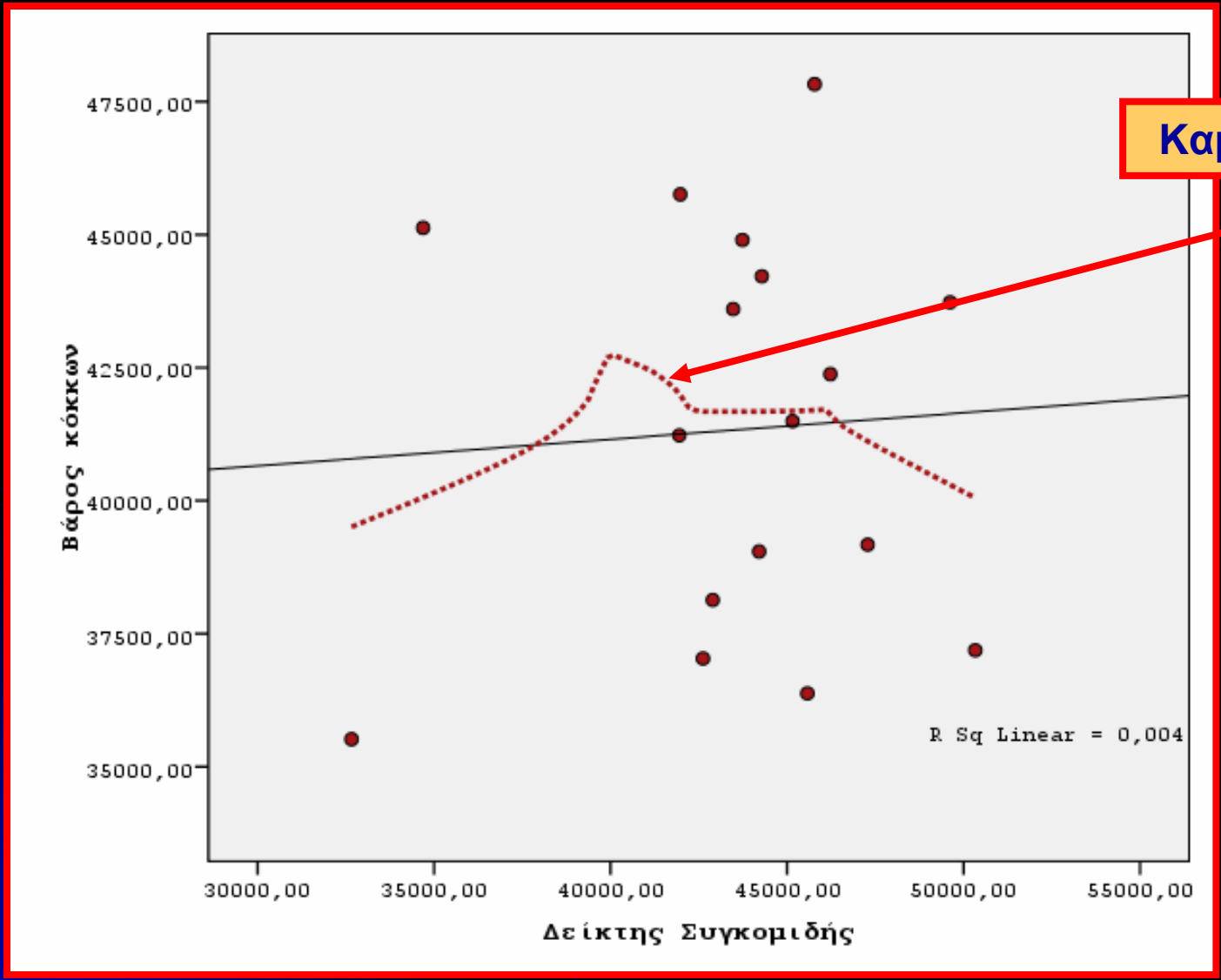


Καμπύλη Loess (90%)

Ευθεία Γραμμικής Συμμεταβολής (Ελαχίστων Τετραγώνων)

Ισχυρή Συσχέτιση

Παράδειγμα με το SPSS (4)



Καμπύλη Loess (90%)

Δεν φαίνεται να υπάρχει συσχέτιση (γραμμική ή των τάξεων)

Παράδειγμα με το SPSS (5)

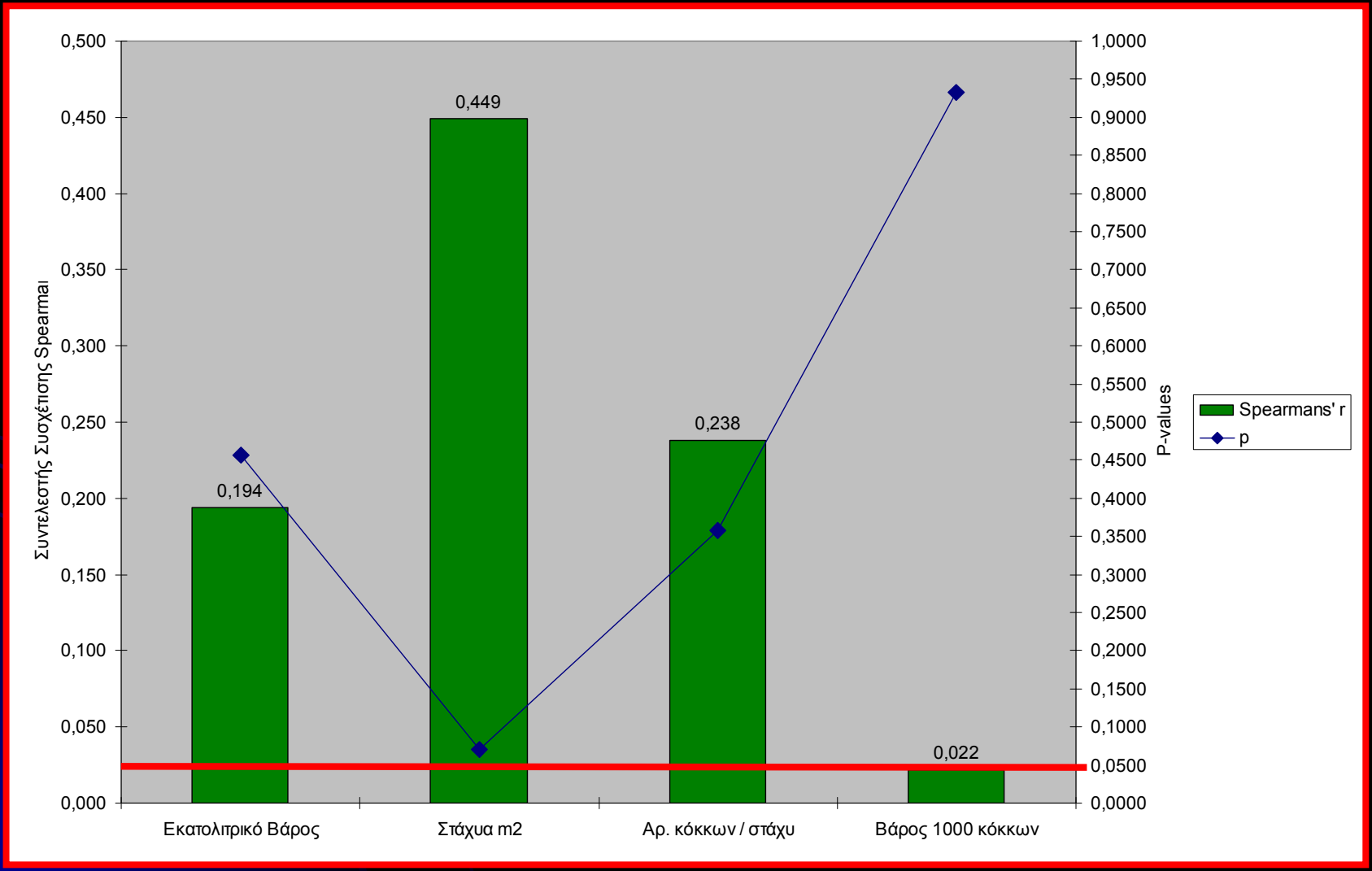
Συντελεστές Γραμμικής Συμμεταβολής *Pearson*

Correlations

		Δείκτης Συγκομιδής	Εκατολιτρικό βάρος	Στάχυα / m2	Αριθμός κόκκων/στάχυ	Βάρος κόκκων
Δείκτης Συγκομιδής	Pearson Correlation	1	,185	,414	,088	,060
	Sig. (2-tailed)		,477	,099	,737	,820
	N	17	17	17	17	17
Εκατολιτρικό βάρος	Pearson Correlation	,185	1	,378	-,350	,304
	Sig. (2-tailed)	,477		,135	,169	,236
	N	17	17	17	17	17
Στάχυα / m2	Pearson Correlation	,414	,378	1	-,771**	,223
	Sig. (2-tailed)	,099	,135		,000	,389
	N	17	17	17	17	17
Αριθμός κόκκων/στάχυ	Pearson Correlation	,088	-,350	-,771**	1	-,380
	Sig. (2-tailed)	,737	,169	,000		,132
	N	17	17	17	17	17
Βάρος κόκκων	Pearson Correlation	,060	,304	,223	-,380	1
	Sig. (2-tailed)	,820	,236	,389	,132	
	N	17	17	17	17	17

**. Correlation is significant at the 0.01 level (2-tailed).

Συσχέτιση Δείκτη Συγκομιδής με τα άλλα χαρακτηριστικά (*n*=17)



Επισημοποίηση

Συντελεστής συσχέτισης του Spearman

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad \text{όπου} \quad d_i = r_{x_i} - r_{y_i}$$

H_0: οι τ.μ. x και y είναι ανεξάρτητες

H_1: οι τ.μ. x και y δεν είναι ανεξάρτητες

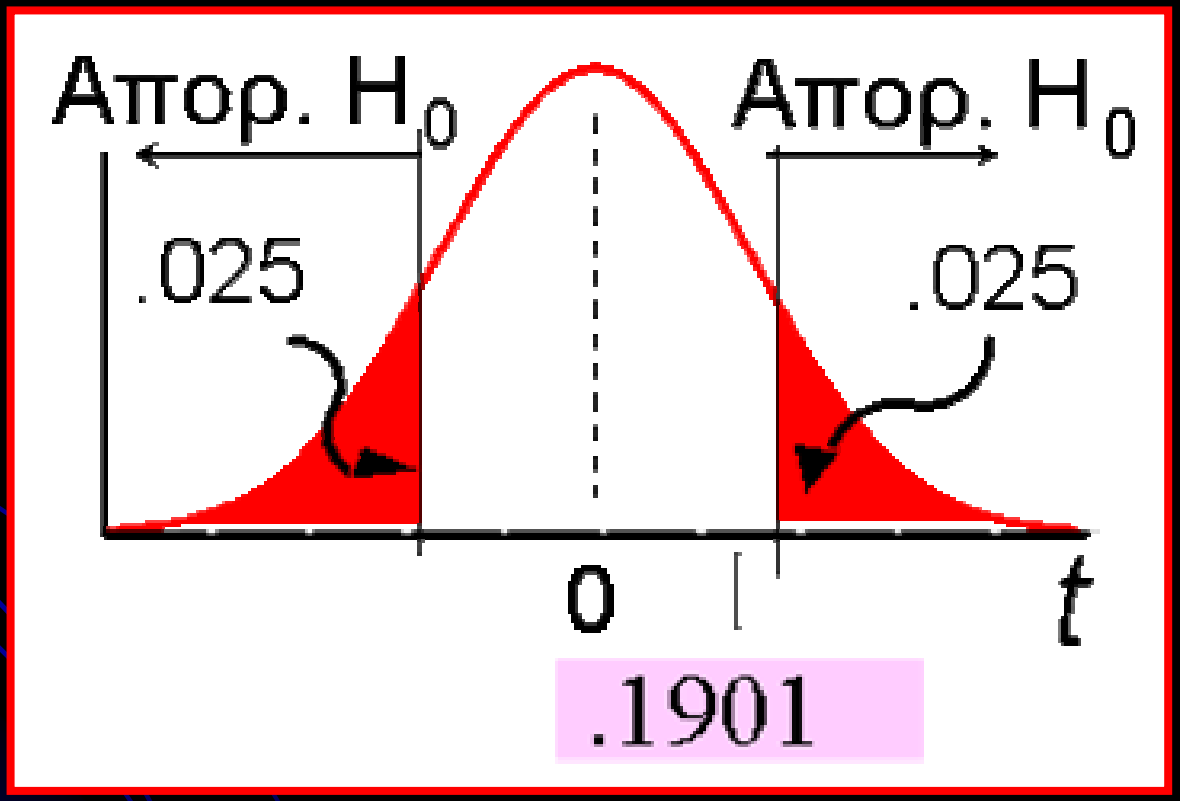
$n \leq 10$ $\quad |r_s| \geq k$ από πίνακες

$n > 10$ $\quad |t_s| \geq t_{n-2;\, a/2}$ όπου $t_s = r_s \sqrt{\frac{n-2}{1-r_s^2}}$

Προϋποθέσεις Εφαρμογής του Ελέγχου Σημαντικότητας

- Δείγματα τυχαία
- Παρατηρήσεις ανεξάρτητες
- Κλίμακα μέτρησης της μεταβλητής τουλάχιστον διάταξης (ordinal)
- Συμμετρικές κατανομές μεταβλητών

Περιοχή Απόρριψης (δίπλευρος έλεγχος)



Παράδειγμα (*n* μικρό)

Beef	Rank	Lamb	Rank	d	d^2
63.4	1	74.5	3	-2	4
68.6	3	79.8	4	-1	1
71.3	4	66.9	1	3	9
66.5	2	74.1	2	0	0
79.7	5	94.4	5	0	0
85.8	6	101.0	6	0	0
89.7	7	110.0	7	0	0

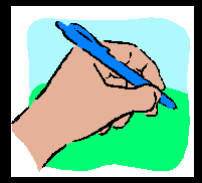
$$\Sigma d^2 = 14$$

Παράδειγμα (*συνέχεια*)

- H_0: $\rho_s = 0$
- H_a: $\rho_s \neq 0$

$\alpha = 0.05$

Παράδειγμα (*συνέχεια*)



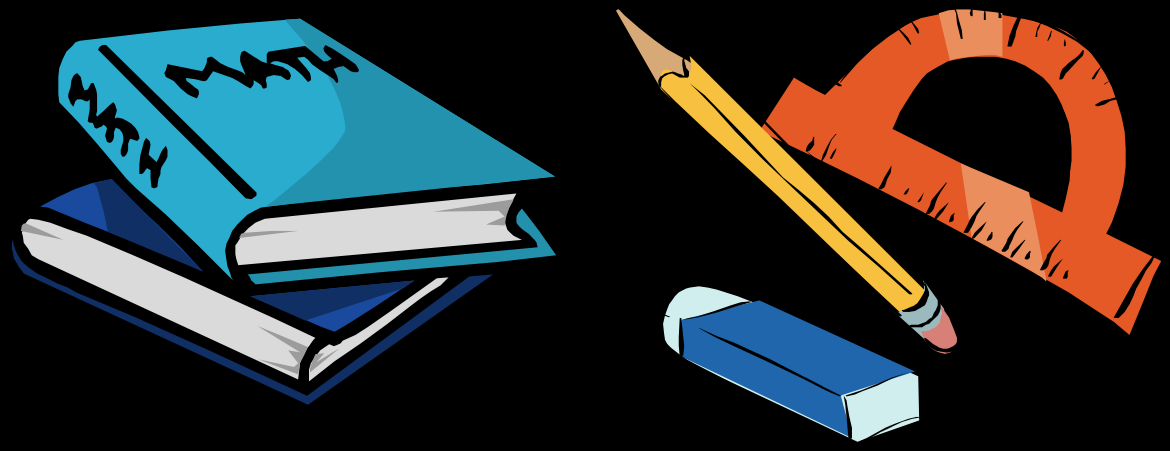
$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(14)}{7(7^2-1)} = 0.75$$

Παράδειγμα (*συνέχεια*)

- Από πίνακες (n=7) βρίσκουμε ότι η κρίσιμη τιμή σε ε.σ. α=0,05 είναι 0,786.
- Η μηδενική υπόθεση δεν μπορεί να απορριφθεί αφού 0,750<0,786.
- Σε επίπεδο σημαντικότητας α=0,05 δεν μπορούμε να ισχυριστούμε ότι υπάρχει στατιστικά σημαντική συσχέτιση των τάξεων μεταξύ των δύο μεταβλητών.

Απορίες-Ερωτήσεις

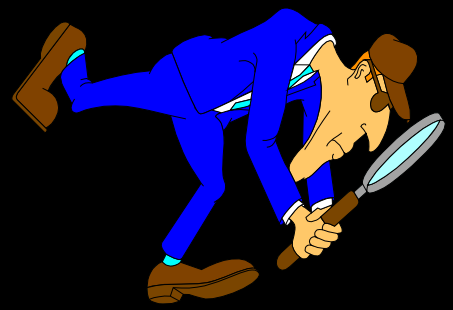


Αντιστοιχίες Ελέγχων

Παραμετρικοί Έλεγχοι	Μη Παραμετρικοί Έλεγχοι
t-test δείγματα ανεξάρτητα	Mann-Whitney test
t-test ζευγαρωτές παρατηρήσεις	Wilcoxon test
ANOVA (one-way)	Kruskal-Wallis test
Συντελεστής Γραμμικής Συμμεταβολής (Pearson)	Συντελεστής Συσχέτισης των Τάξεων (Spearman)

Μη Παραμετρικοί Έλεγχοι

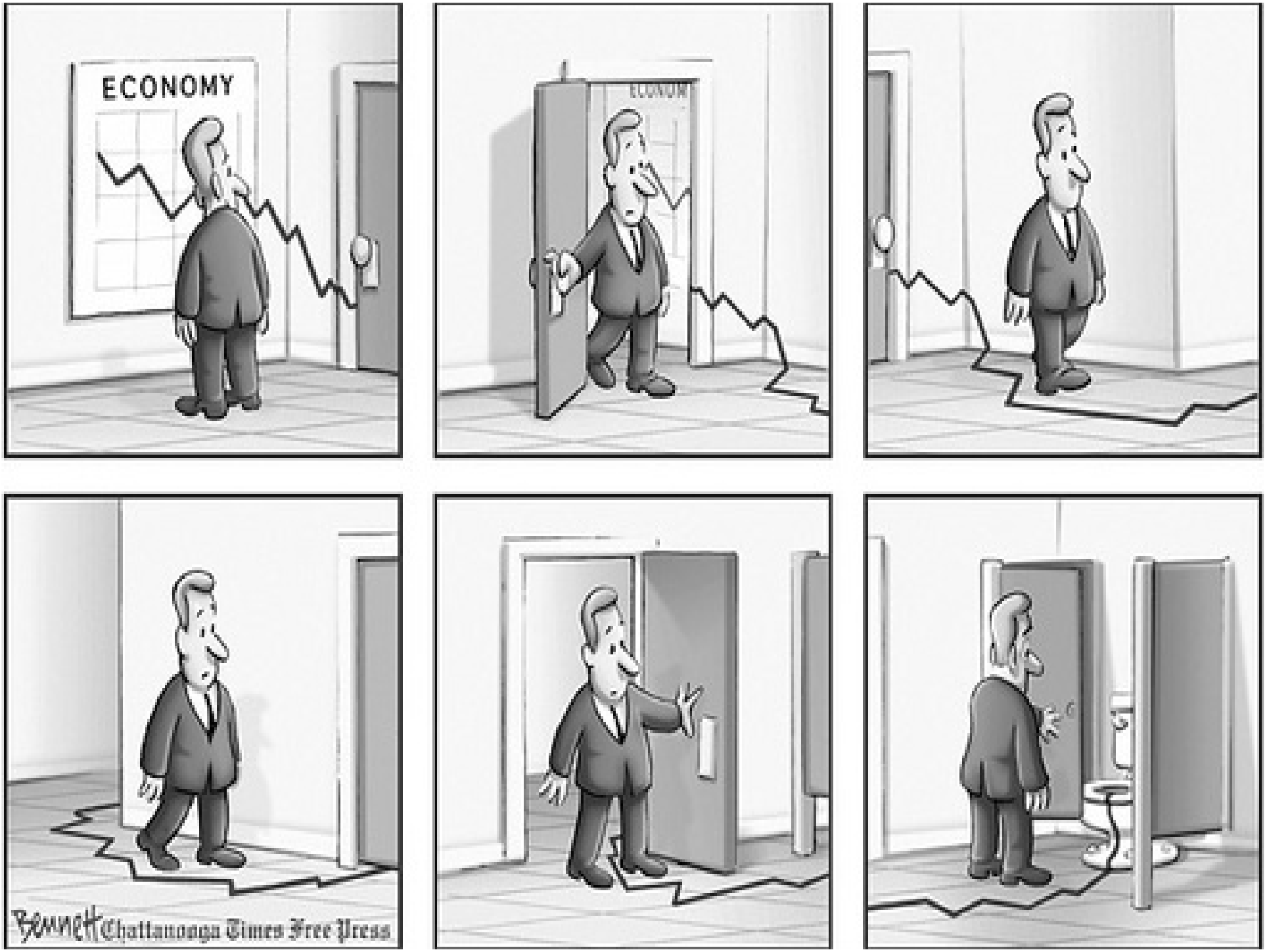
	Τύπος μεταβλητής	Κριτήρια
1 δείγμα	ποιοτική ποσοτική	X^2, ροών Kolmogorov - Smirnov
2 δείγματα ανεξάρτητα	ποιοτική ποσοτική	X^2 διαμέσου, Mann-Whitney, Kolmogorov - Smirnov, Wald - Wolfowitz,
2 δείγματα εξαρτημένα	ποιοτική ποσοτική	Mc Nemar προσημικό, Wilcoxon
k δείγματα ανεξάρτητα	ποιοτική ποσοτική	X^2 διαμέσου, Kruskal - Wallis
k δείγματα εξαρτημένα	ποιοτική ποσοτική	Cochran Friedman



Ποια είναι τα μειονεκτήματα των Μη Παραμετρικών ελέγχων;

Βιβλιογραφία

- Φωτιάδης, Ν. (1995). *Εισαγωγή στη Στατιστική για Βιολογικές Επιστήμες*. Θεσσαλονίκη: University Studio Press.
- Κολυβά, Φ. και Μπόρα-Σέντα, Ε. (1995). *Στατιστική: Θεωρία-Εφαρμογές*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.


ECONOMY
Bennett Chattanooga Times Free Press