

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Παναγιώτα Λάλου, Αλέξανδρος Γρυπάρης

Έλεγχος χ^2

- Ο έλεγχος αυτός πραγματοποιείται όταν έχουμε 2 ποιοτικές μεταβλητές
- Η υποθέσεις υπό έλεγχο είναι οι:
 - H_0 : Δεν υπάρχει σχέση μεταξύ των 2 μεταβλητών μας, δηλαδή οι μεταβλητές μας είναι ανεξάρτητες
 - H_1 : Υπάρχει σχέση μεταξύ των 2 μεταβλητών μας, δηλαδή οι μεταβλητές μας δεν είναι ανεξάρτητες

Έλεγχος Ανεξαρτησίας δυο κατηγορικών μεταβλητών: Έλεγχος χ^2

- Με τον έλεγχο αυτό, ελέγχουμε αν δύο κατηγορικές μεταβλητές είναι ανεξάρτητες ή όχι, δηλαδή αν επηρεάζει η μία την άλλη.
 - Αν βρεθεί εξάρτηση δεν είμαστε σε θέση να πούμε ποια είναι η αιτία και ποιο το αποτέλεσμα.
- Εστω ότι ένας πληθυσμός εξετάζεται ως προς 2 χαρακτηριστικά (μεταβλητές) A και B.

A \ B	B1	B2	...	Bs	Σύνολο
A1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
A2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
⋮					
A _r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
Σύνολο	$n_{.1}$	$n_{.2}$...	$n_{.s}$	n

H_0 : Οι μεταβλητές A και B είναι ανεξάρτητες

H_1 : Οι μεταβλητές A και B δεν είναι ανεξάρτητες

Κρίσιμη περιοχή: $R = \{X^2 > X^2_{(r-1)*(s-1);a}\}$

Στατιστικό κριτήριο: $X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

όπου: $O_{ij} = n_{ij}$ οι παρατηρούμενες τιμές (*Observed*)

$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ οι αναμενόμενες τιμές (*Expected*)

Ή αλλιώς:

$$E_{ij} = \frac{\text{Αθροισμα } i \text{ γραμμης} \times \text{Αθροισμα } j \text{ στηλης}}{n}$$

Στην κρίσιμη τιμή $X^2_{(r-1)*(s-1);a}$ **r**: αριθμός γραμμών

s: αριθμός στηλών

Προϋποθέσεις εφαρμογής του ελέγχου:

1. Τα δείγματα είναι τυχαία.
2. Οι παρατηρήσεις είναι ανεξάρτητες.
3. Όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 1.
4. Το πολύ 20% από τις αναμενόμενες συχνότητες είναι μικρότερες από 5.
 - Δηλαδή, πάνω από το 80% των αναμενόμενων συχνοτήτων είναι μεγαλύτερες του 5.

Παράδειγμα 2

Ο παρακάτω πίνακας παρουσιάζει την κατανομή 80 καρκινοπαθών (με καρκίνο του μαστού) και 160 “υγιών” κατά οικογενειακή κατάσταση

	Οικογενειακή κατάσταση		
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα
Ναι	30	44	6
Όχι	30	106	24

Σχετίζεται ο καρκίνος του μαστού με την οικογενειακή κατάσταση; ($\alpha=0,05$)

- Στο παράδειγμα αυτό θα ελέγξουμε αν ο καρκίνος του μαστού(A) σχετίζεται με την οικογενειακή κατάσταση(B).
- Επειδή και οι 2 μεταβλητές είναι ποιοτικές, θα χρησιμοποιήσουμε τον έλεγχο χ^2 .
- Για τον έλεγχο αυτό έχουμε ότι:

H_0 : Οι μεταβλητές A και B είναι ανεξάρτητες

H_1 : Οι μεταβλητές A και B δεν είναι ανεξάρτητες

Στην επόμενη διαφάνεια υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης.

Υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης:

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα	ΣΥΝΟΛΟ
Ναι	30	44	6	80
Όχι	30	106	24	160
ΣΥΝΟΛΟ	60	150	30	240

• Οι αναμενόμενες συχνότητες είναι:

$$\bullet E_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{80 \cdot 60}{240} = 20$$

Προϋποθέσεις;;;

$$\bullet E_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{80 \cdot 150}{240} = 50$$

$$\bullet E_{13} = \frac{n_{1.} \times n_{.3}}{n} = \frac{80 \cdot 30}{240} = 10$$

$$\bullet E_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{160 \cdot 60}{240} = 40$$

$$\bullet E_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{160 \cdot 150}{240} = 100$$

$$\bullet E_{23} = \frac{n_{2.} \times n_{.3}}{n} = \frac{160 \cdot 30}{240} = 20$$

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα	ΣΥΝΟΛΟ
Ναι	30	44	6	80
Όχι	30	106	24	160
ΣΥΝΟΛΟ	60	150	30	240

Παρατηρούμενες
συχνότητες

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα	ΣΥΝΟΛΟ
Ναι	20	50	10	80
Όχι	40	100	20	160
ΣΥΝΟΛΟ	60	150	30	240

Αναμενόμενες
συχνότητες

Το στατιστικό κριτήριο είναι:
$$X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$$
$$\frac{(30 - 20)^2}{20} + \frac{(44 - 50)^2}{50} + \frac{(6 - 10)^2}{10} + \frac{(30 - 40)^2}{40} + \frac{(106 - 100)^2}{100} + \frac{(24 - 20)^2}{20} = 10,98$$

Η κρίσιμη τιμή, απο πίνακες είναι:

$$X^2_{(r-1)*(s-1);a} = X^2_{(2-1)*(3-1);0.05} = X^2_{2;0.05} = 5,99$$

Η κρίσιμη περιοχή: $R = \{X^2 > X^2_{(r-1)(s-1);a}\}$

Άρα, είμαστε στην κρίσιμη περιοχή, αφού $10,98 > 5,99$

Επομένως απορρίπτουμε την μηδενική υπόθεση και δεχόμαστε ότι ο καρκίνος του μαστού σχετίζεται με την οικογενειακή κατάσταση.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Παράδειγμα

- Κατανομή 3042 ανδρών και γυναικών βάση των διατροφικών τους συνηθειών (πόσο κοντά είναι στη μεσογειακή διαίτα)

Διατροφικές συνήθειες				
Φύλο	Μακριά	Κοντά	Πολύ κοντά	Σύνολο
Γυναίκες	246	428	854	1528
Άνδρες	767	587	160	1514
Σύνολο	1013	1015	1014	3042

- Να ελεγχθεί αν το φύλο σχετίζεται με τις διατροφικές συνήθειες.

Μηδενική και εναλλακτική υπόθεση

- H_0 : Το φύλο δεν σχετίζεται με τις διατροφικές συνήθειες
- H_1 : Το φύλο **σχετίζεται** με τις διατροφικές συνήθειες

Ο έλεγχος χ^2 έδειξε ...

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	767,803 ^a	2	,000
Likelihood Ratio	827,697	2	,000
Linear-by-Linear Association	728,051	1	,000
N of Valid Cases	3042		

p-value<0,001

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 504,17.

Με βάση τα στοιχεία του πίνακα ...

- Το εκτιμώμενο p-value $< 0,001$, το οποίο είναι $< 0,05$
- Ποια η τιμή του χ^2 κριτηρίου;
- Πόσοι είναι οι βαθμοί ελευθερίας;

Με βάση τα στοιχεία του πίνακα ...

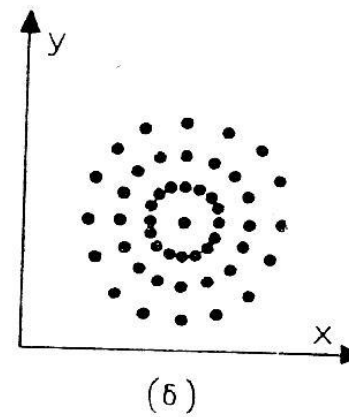
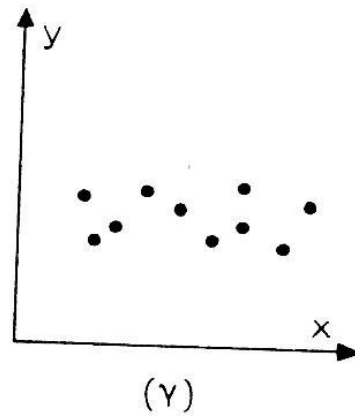
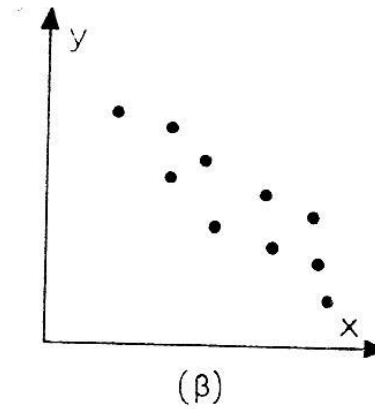
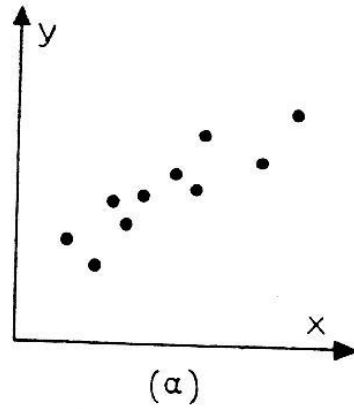
- Ποιο το συμπέρασμά σας;
 - Επειδή $p\text{-value} < 0,05$, απορρίπτουμε τη μηδενική υπόθεση και αποδεχόμαστε την H_1
 - Άρα, οι διατροφικές συνήθειες εξαρτώνται από το φύλο

Συντελεστής συσχέτισης του Pearson: r

- Ο επόμενος έλεγχος που θα μελετήσουμε διερευνά αν υπάρχει γραμμική σχέση μεταξύ 2 ποσοτικών μεταβλητών.

Διαγράμματα συσχέτισης:

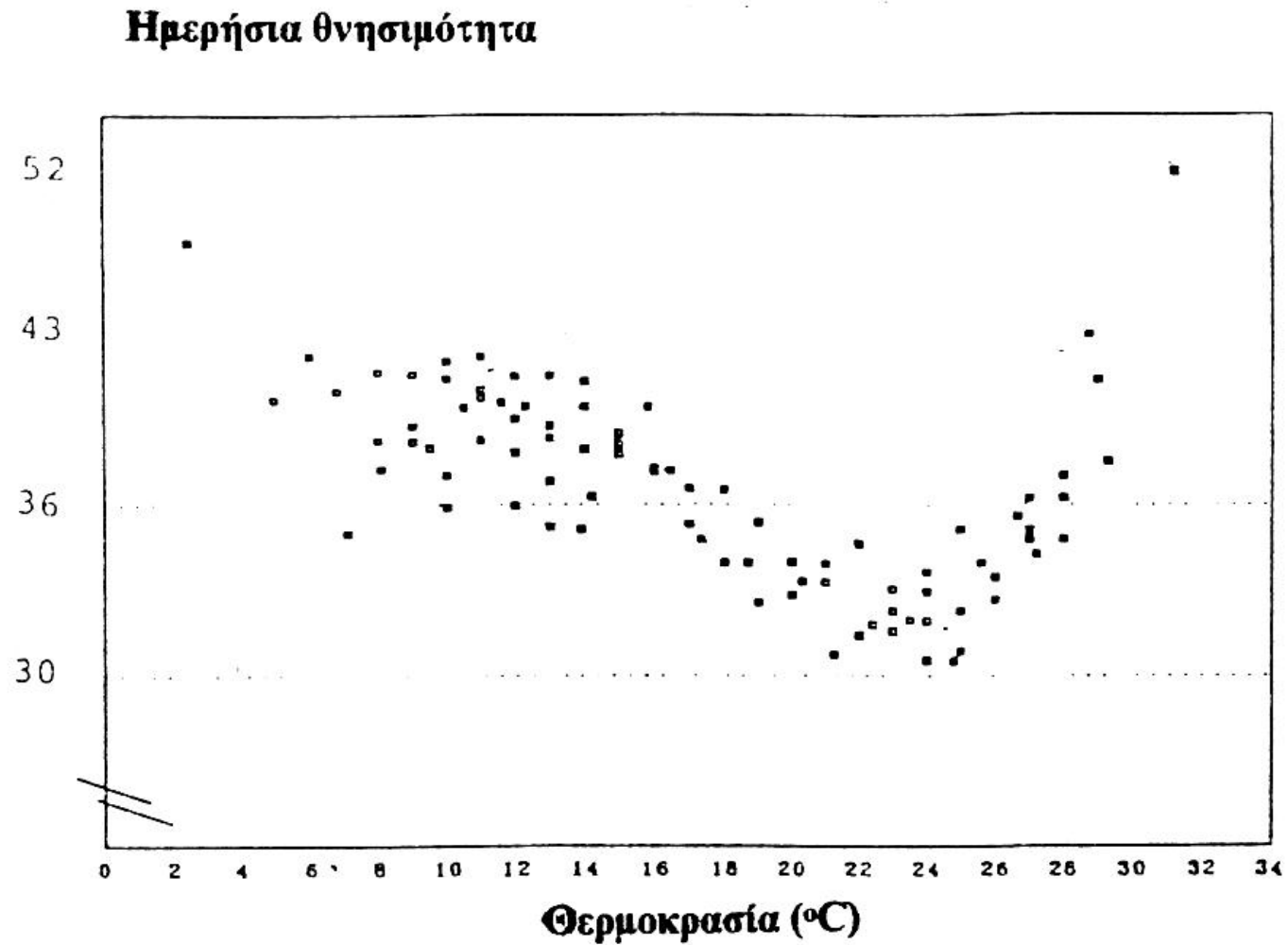
(α) θετική συσχέτιση, (β) αρνητική συσχέτιση,
(γ) και (δ) απουσία γραμμικής συσχέτισης



Έλλειψη γραμμικής σχέσης;

- Έλλειψη γραμμικής σχέσης δεν συνεπάγεται και απουσία σχέσης.
 - Δηλαδή, μπορεί να υπάρχει σχέση μεταξύ 2 μεταβλητών, απλά να μην είναι γραμμική!
- Στο επόμενο σχήμα παρουσιάζεται το στικτόγραμμα της μέσης ημερήσιας θερμοκρασίας με τον μέσο ημερήσιο αριθμό θανάτων από όλες τις αιτίες.

Σχέση μεταξύ μέσης ημερήσιας θνησιμότητας και μέσης ημερήσιας θερμοκρασίας



(συν.)

- Παρατηρούμε μια μη-γραμμική σχέση μεταξύ θνησιμότητας και θερμοκρασίας:
 - Μεγαλύτερη θνησιμότητα παρατηρείται κατά τις πολύ κρύες ή πολύ ζεστές ημέρες.
 - Στην περίπτωση αυτή, ο συντελεστής συσχέτισης του Pearson δεν θα μας ήταν χρήσιμος.

Ιδιότητες του συντελεστή συσχέτισης

1. Είναι καθαρός αριθμός

2. Παίρνει τιμές από -1 ως $+1$ (αρνητική ή θετική συσχέτιση).

Θετική συσχέτιση: σε υψηλές τιμές του ενός μεγέθους αντιστοιχούν υψηλές τιμές του άλλου.

Αρνητική: σε υψηλές τιμές του ενός μεγέθους αντιστοιχούν χαμηλές τιμές του άλλου.

- Όσο ο r πλησιάζει την τιμή $+1$ (ή -1) τόσο ισχυρότερη είναι η συσχέτιση, ενώ όσο πιο κοντά είναι στο 0 τόσο πιο αδύναμη η γραμμική συσχέτιση.

3. Μετρά μόνο την γραμμική συσχέτιση

Οι τιμές του συντελεστή ...

Εμπειρική κατηγοριοποίηση του συντελεστή συσχέτισης.



Ισχυρή αρνητική συσχέτιση	Μέτρια αρνητικά συσχετισμένες	Ελαφρά αρνητικά συσχετισμένες	Ασυσχέτιστες ή πολύ ελαφρά συσχετισμένες	Ελαφρά θετικά συσχετισμένες	Μέτρια θετικά συσχετισμένες	Ισχυρή θετική συσχέτιση
<-0,80	-0,80...-0,50	-0,50...-0,20	-0,20...+0,20	+0,20...+0,50	+0,50...+0,80	> +0,80
Τιμές του συντελεστή συσχέτισης						

Ο συντελεστής συσχέτισης r του Pearson

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) * (\sum_{i=1}^n y_i^2 - n \bar{y}^2)}}$$

Έλεγχος υπόθεσης για τον συντελεστή συσχέτισης του Pearson

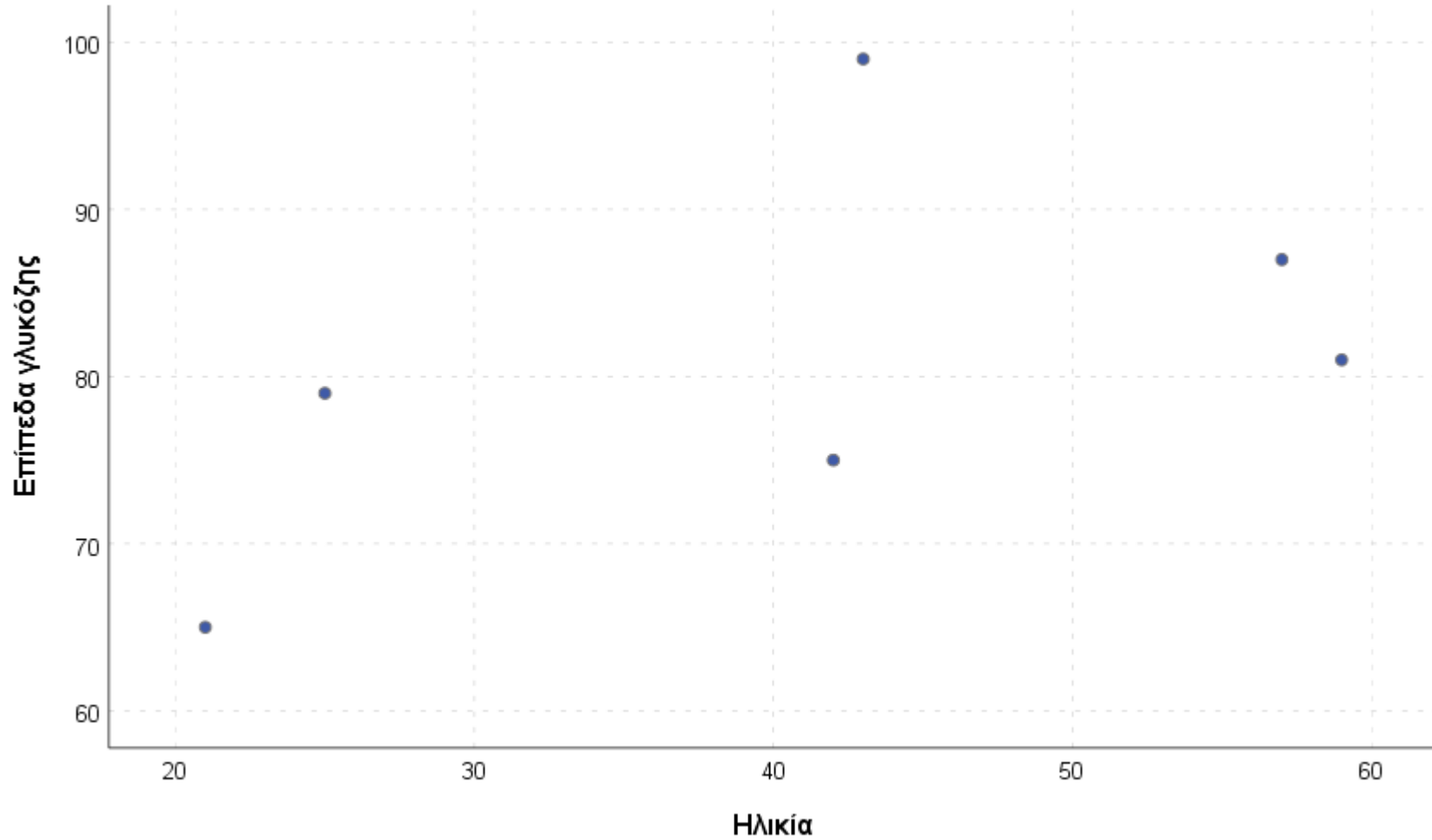
- Όταν έχουμε 2 ποσοτικά μεγέθη, μας ενδιαφέρει να ελέγξουμε αν υπάρχει ή όχι γραμμική συσχέτιση μεταξύ τους.
- Αυτό είναι ισοδύναμο με το να ελέγξουμε αν ο r είναι ίσος με 0 (οπότε δεν υπάρχει γραμμική συσχέτιση), ή όχι.
- Ο έλεγχος υπόθεσης που χρησιμοποιούμε είναι ο εξής:
 - $H_0: r=0$ (οι μεταβλητές μας δεν συσχετίζονται γραμμικά)
 - $H_1: r \neq 0$ (οι μεταβλητές μας συσχετίζονται γραμμικά)
 - Η κρίσιμη περιοχή είναι η: $R = \{t > t_{n-2; \alpha}\}$
 - Το στατιστικό κριτήριο: $t = r * \sqrt{\frac{n-2}{1-r^2}}$

Παράδειγμα

Σε 6 άτομα που έπασχαν από διαβήτη τύπου 2, μετρήθηκαν οι τιμές γλυκόζης στο αίμα. Το ερώτημα που μας απασχολεί είναι αν τα επίπεδα αυτά σχετίζονται με την ηλικία τους. Ο παρακάτω πίνακας παρουσιάζει τα στοιχεία που έχουμε:

Ηλικία (X)	Επίπεδα Γλυκόζης (Y)
43	99
21	65
25	79
42	75
57	87
59	81

Γράφημα (σικτόγραμμα)



Θέλουμε να ελέγξουμε αν:

- H_0 : Τα επίπεδα γλυκόζης δεν συσχετίζονται γραμμικά με την ηλικία
- H_1 : Τα επίπεδα γλυκόζης συσχετίζονται γραμμικά με την ηλικία

	Ηλικία (X)	Επίπεδα γλυκόζης (Y)	XY	X ²	Y ²
	43	99	4257	1849	9801
	21	65	1365	441	4225
	25	79	1975	625	6241
	42	75	3150	1764	5625
	57	87	4959	3249	7569
	59	81	4779	3481	6561
ΣΥΝΟΛΟ	247	486	20485	11409	40022

$$\bar{x} = \frac{247}{6} = 41,17$$

$$\bar{y} = \frac{486}{6} = 81$$

Ο συντελεστής συσχέτισης r του Pearson δίνεται από τη σχέση:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) * (\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = \frac{20.485 - 6 * 41,17 * 81}{\sqrt{(11.409 - 6 * 41,17^2)(40.022 - 6 * 81^2)}} =$$
$$= 0,53$$

Το στατιστικό κριτήριο δίνεται από τη σχέση:

$$t = r * \sqrt{\frac{n - 2}{1 - r^2}} = 0,53 * \sqrt{\frac{6 - 2}{1 - 0,53^2}} = 0,53 * \sqrt{5,56} = 0,53 * 2,35 = 1,25$$

Η κρίσιμη τιμή είναι η $t_{n-2;\alpha} = t_{4;0,05} = 2,132$

Η κρίσιμη περιοχή είναι η: $R = \{t > t_{n-2;\alpha}\}$

Έχουμε $1,25 = t < t_{n-2;\alpha} = 2,132$. Άρα δεν βρισκόμαστε στην κρίσιμη περιοχή.

Οπότε δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση. Έτσι, συμπεραίνουμε ότι η ηλικία και τα επίπεδα της γλυκόζης δεν σχετίζονται γραμμικά.

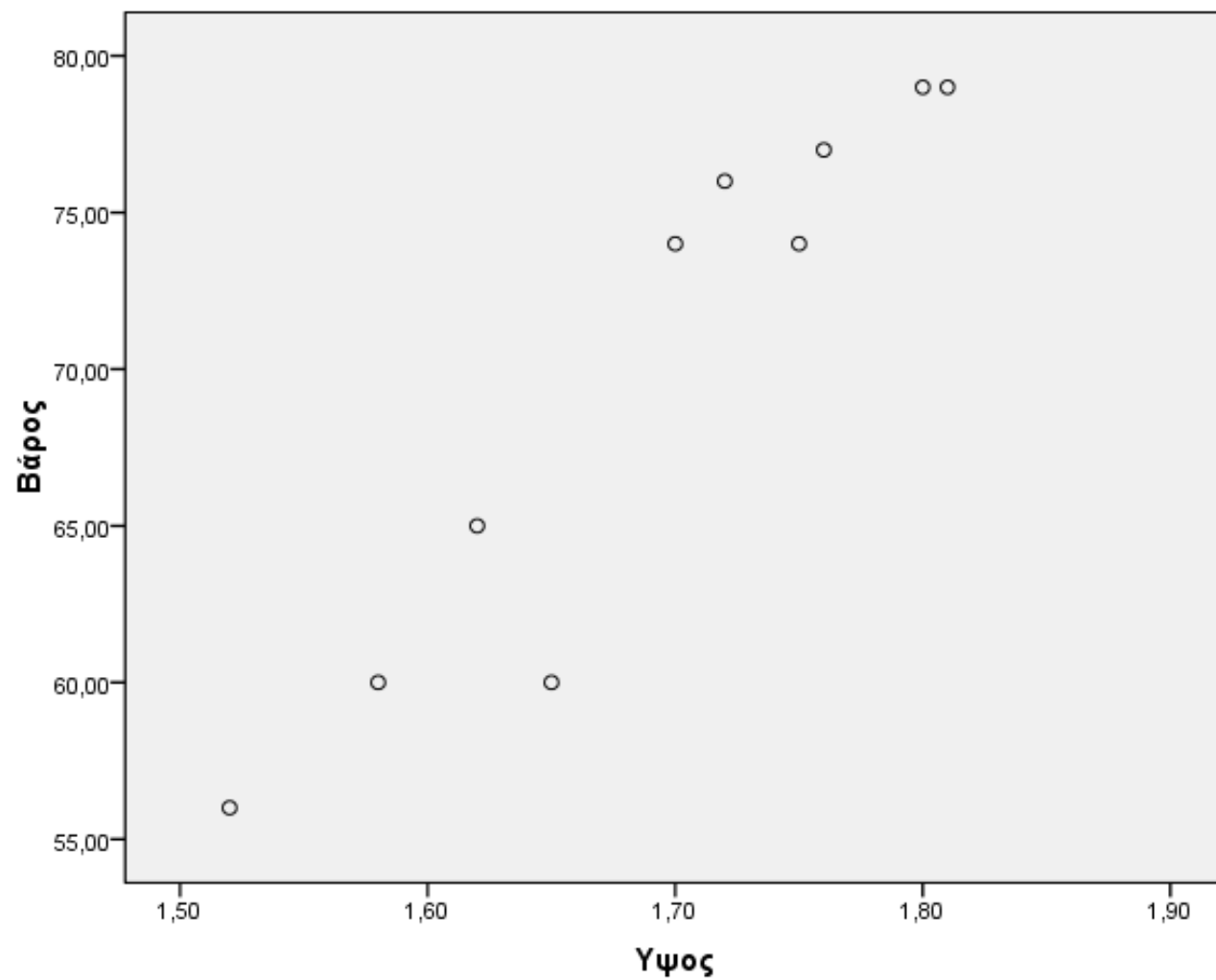
DF	A = 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
∞	ta = 1.282	1.645	1.96	2.326	2.576	3.091	3.291
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3

Παράδειγμα 2

- Ο παρακάτω πίνακας δίνει τα ύψη και τα αντίστοιχα βάρη 10 μαθητών

Υψος σε m (X)	Βάρος σε κιλά (Y)
1,52	56
1,58	60
1,62	65
1,65	60
1,70	74
1,72	76
1,75	74
1,76	77
1,8	79
1,81	79

- Υπάρχει γραμμική συσχέτιση του ύψους με το βάρος;



Στικτόγραμμα του βάρους σε σχέση με το ύψος

- H_0 : Το ύψος δεν συσχετίζεται γραμμικά με το βάρος
- H_1 : Το ύψος συσχετίζεται γραμμικά με το βάρος

	Υψος σε m (X)	Βάρος σε κιλά (Y)	XY	X ²	Y ²
	1,52	56	85,12	2,3104	3136
	1,58	60	94,8	2,4964	3600
	1,62	65	105,3	2,6244	4225
	1,65	60	99	2,7225	3600
	1,70	74	125,8	2,89	5476
	1,72	76	130,72	2,9584	5776
	1,75	74	129,5	3,0625	5476
	1,76	77	135,52	3,0976	5929
	1,8	79	142,2	3,24	6241
	1,81	79	142,99	3,2761	6241
ΣΥΝΟΛΟ	16.91	700	1190,95	28,6783	49700

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{16,91}{10} = 1,691$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{700}{10} = 70$$

- Συντελεστής συσχέτισης Pearson:

$$\begin{aligned} r &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}} = \frac{1190,95 - 10 \cdot 1,691 \cdot 70}{\sqrt{28,6783 - 10 \cdot 1,691^2} \cdot \sqrt{49700 - 10 \cdot 70^2}} \\ &= \frac{7,25}{0,289 \cdot 26,458} = 0,948 \end{aligned}$$

Το στατιστικό κριτήριο : $t = r \cdot \sqrt{\frac{n-2}{1-r^2}} = 0,948 \cdot \sqrt{\frac{10-2}{1-0,948^2}} = 0,948 \cdot 8,887 = 8,425$

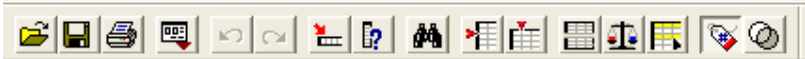
Η κρίσιμη τιμή: $t_{n-2;\alpha} = t_{8;0,05} = 1,86$

Η κρίσιμη περιοχή: $R = \{t > t_{n-2;\alpha}\}$

Άρα βρισκόμαστε στην κρίσιμη περιοχή, κι επομένως απορρίπτουμε την μηδενική υπόθεση. Δηλαδή το ύψος σχετίζεται γραμμικά με το βάρος.

Παράδειγμα

- Έστω ότι στο επόμενο αρχείο θέλουμε να συσχετίσουμε την ηλικία με τον μεταβολικό ρυθμό.
 - Και οι 2 μεταβλητές είναι συνεχείς
- Αρχικά ξεκινάμε από ένα στικτόγραμμα μεταξύ των 2 μεταβλητών
 - Σε αυτό ελέγχουμε αν υπάρχει γραμμική σχέση μεταξύ των 2 μεταβλητών



1 : country Afghanistan

	country	populatn	density	urban	religion	lifeexpf	lifeexpm	literacy	pop_incr	babymort	gdp_cap	region	calorie
1	Afghanistan	20500	25,0	18	Muslim	44	45	29	2,8	168,0	205	Pacific/Asi	
2	Argentina	33900	12,0	86	Catholic	75	68	95	1,3	25,6	3408	Latn Ameri	31
3	Armenia	3700	126,0	68	Orthodox	75	68	98	1,4	27,0	5000	Middle Eas	
4	Australia	17800	2,3	85	Protstnt	80	74	100	1,4	7,3	16848	OECD	32
5	Austria	8000	94,0	58	Catholic	79	73	99	,2	6,7	18396	OECD	34
6	Azerbaijan	7400							1,4	35,0	3000	Middle Eas	
7	Bahrain	600							2,4	25,0	7875	Middle Eas	
8	Bangladesh	125000							2,4	106,0	202	Pacific/Asi	20
9	Barbados	256							,2	20,3	6950	Latn Ameri	
10	Belarus	10300							,3	19,0	6500	East Europ	
11	Belgium	10100							,2	7,2	17912	OECD	
12	Bolivia	7900							2,7	75,0	730	Latn Ameri	19
13	Bosnia	4600							,7	12,7	3098	East Europ	
14	Botswana	1359							2,7	39,3	2677	Africa	25
15	Brazil	156600							1,3	66,0	2354	Latn Ameri	27
16	Bulgaria	8900							-,2	12,0	3831	East Europ	
17	Burkina Faso	10000							2,8	118,0	357	Africa	22
18	Burundi	6000							2,3	105,0	208	Africa	19
19	Cambodia	10000							2,9	112,0	260	Pacific/Asi	21
20	Cameroon	13100							2,9	77,0	993	Africa	22
21	Canada	29100							,7	6,8	19904	OECD	34
22	Cent. Afri.R	3300	5,0	47	Protstnt	44	41	27	2,4	137,0	457	Africa	20
23	Chile	14000	18,0	85	Catholic	78	71	93	1,7	14,6	2591	Latn Ameri	25
24	China	1205200	124,0	26	Taoist	69	67	78	1,1	52,0	377	Pacific/Asi	25
25	Colombia	35600	31,0	70	Catholic	75	69	87	2,0	28,0	1538	Latn Ameri	25
26	Costa Rica	3300	64,0	47	Catholic	79	76	93	2,3	11,0	2031	Latn Ameri	25
27	Croatia	4900	85,0	51	Catholic	77	70	97	-,1	8,7	5487	East Europ	
28	Cuba	11100	99,0	74	Catholic	78	74	94	1,0	10,2	1382	Latn Ameri	
29	Czech Rep.	10400	132,0	.	Catholic	77	69	.	,2	9,3	7311	East Europ	35
30	Denmark	5200	120,0	85	Protstnt	79	73	99	,1	6,6	18277	OECD	35
31	Dominican R	7800	159,0	60	Catholic	70	66	83	1,8	51,5	1034	Latn Ameri	25

Bivariate Correlations

Variables:

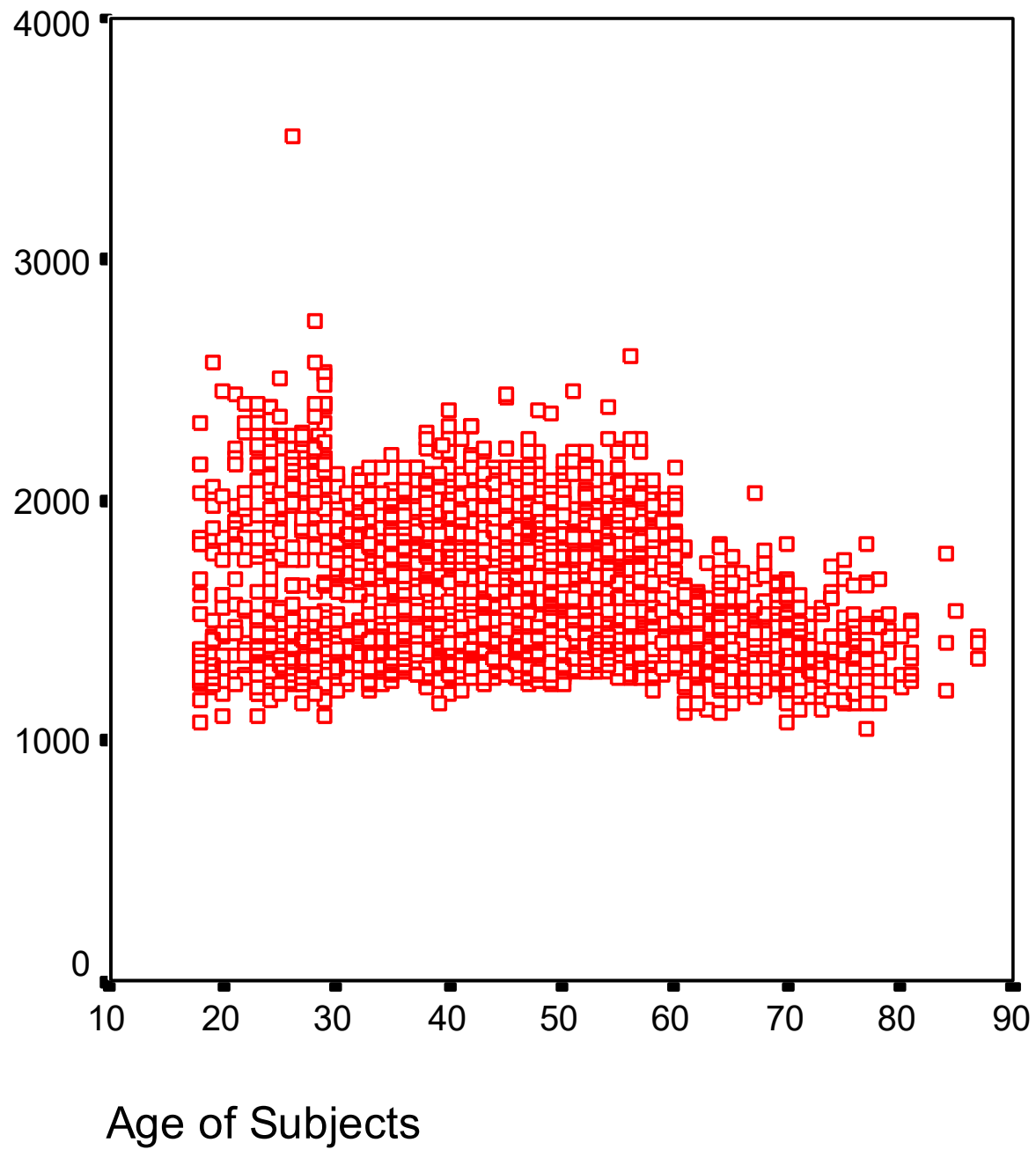
Population in thous.
 Number of people /
 People living in citie
 Average female life
 Average male life ex
 People who read (%
 Population increase
 Infant mortality idea

Correlation Coefficients
 Pearson Kendall's tau-b Spearman

Test of Significance
 Two-tailed One-tailed

Flag significant correlations

Options...



Το στικτόγραμμα φαίνεται
στο διπλανό σχήμα.

Τα αποτελέσματα από το SPSS δίνονται στον παρακάτω πίνακα:

Correlations

		Age of Subjects	Basic Metabolic Rythm (Harris & Benedict)
Age of Subjects	Pearson Correlation	1	-,194**
	Sig. (2-tailed)	,	,000
	N	3035	3017
Basic Metabolic Rythm (Harris & Benedict)	Pearson Correlation	-,194**	1
	Sig. (2-tailed)	,000	,
	N	3017	3017

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

Ο συντελεστής συσχέτισης του **Pearson** είναι $r = -0,194$ (πολύ ελαφρά αρνητική συσχέτιση), η οποία είναι όμως στατιστικά σημαντική ($p\text{-value} < 0,001$).

(συν.)

- Η μηδενική υπόθεση που αντιστοιχεί στο p-value που μας δίνει ο πίνακας είναι η εξής:
- $H_0: r=0$ (δηλαδή οι δύο μεταβλητές μας δεν συσχετίζονται γραμμικά)
- $H_1: r \neq 0$ (δηλαδή οι δύο μεταβλητές μας συσχετίζονται γραμμικά)
- Το $p\text{-value} < 0,001$, άρα απορρίπτουμε την μηδενική υπόθεση. Έτσι, οι μεταβλητές μας σχετίζονται γραμμικά.
- Επιπλέον, ο συντελεστής συσχέτισης του Pearson είναι $r = -0,194$ (πολύ ελαφρά αρνητική συσχέτιση).