

# Συσχέτιση

Ανάλυση συνεχών μεταβλητών

Γεωργία Σαλαντή

Λέκτορας

*Εργαστήριο υγιεινής και Επιδημιολογίας*

# Περιεχόμενα

A decorative graphic consisting of two groups of three circles. The first group on the left has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right. The second group on the right has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right.

- Συσχέτιση μεταξύ δύο συνεχών μεταβλητών
- Παλινδρόμηση μεταξύ
  - Μίας συνεχούς μεταβλητής
  - Μιας (ή και περισσότερων) μεταβλητών (συνεχών, διακριτών κτλ)

# Εισαγωγικά

A decorative graphic at the top of the slide consists of two groups of three circles. The first group on the left has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right. The second group on the right has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right.

- $y$  το αποτέλεσμα (ή δεσμευμένη μεταβλητή) που μας ενδιαφέρει
  - Π.χ. πίεση, χοληστερίνη
- $x$  η ανεξάρτητη μεταβλητή
  - Π.χ. ηλικία, φύλο

Ο **συντελεστής συσχέτισης** κοιτά το πως μεταβάλλεται το  $y$  σε σχέση με το  $x$

# Τυχαίες μεταβλητές και συσχέτιση

- Όσο πιο ψηλός τόσο πιο βαρύς
- Όσο πιο χαμηλό το βιοτικό επίπεδο, τόσο πιο υψηλή η παιδική θνησιμότητα
- Όσο πιο πολύ το  $x$   
τόσο πιο πολύ/λίγο το  $y$

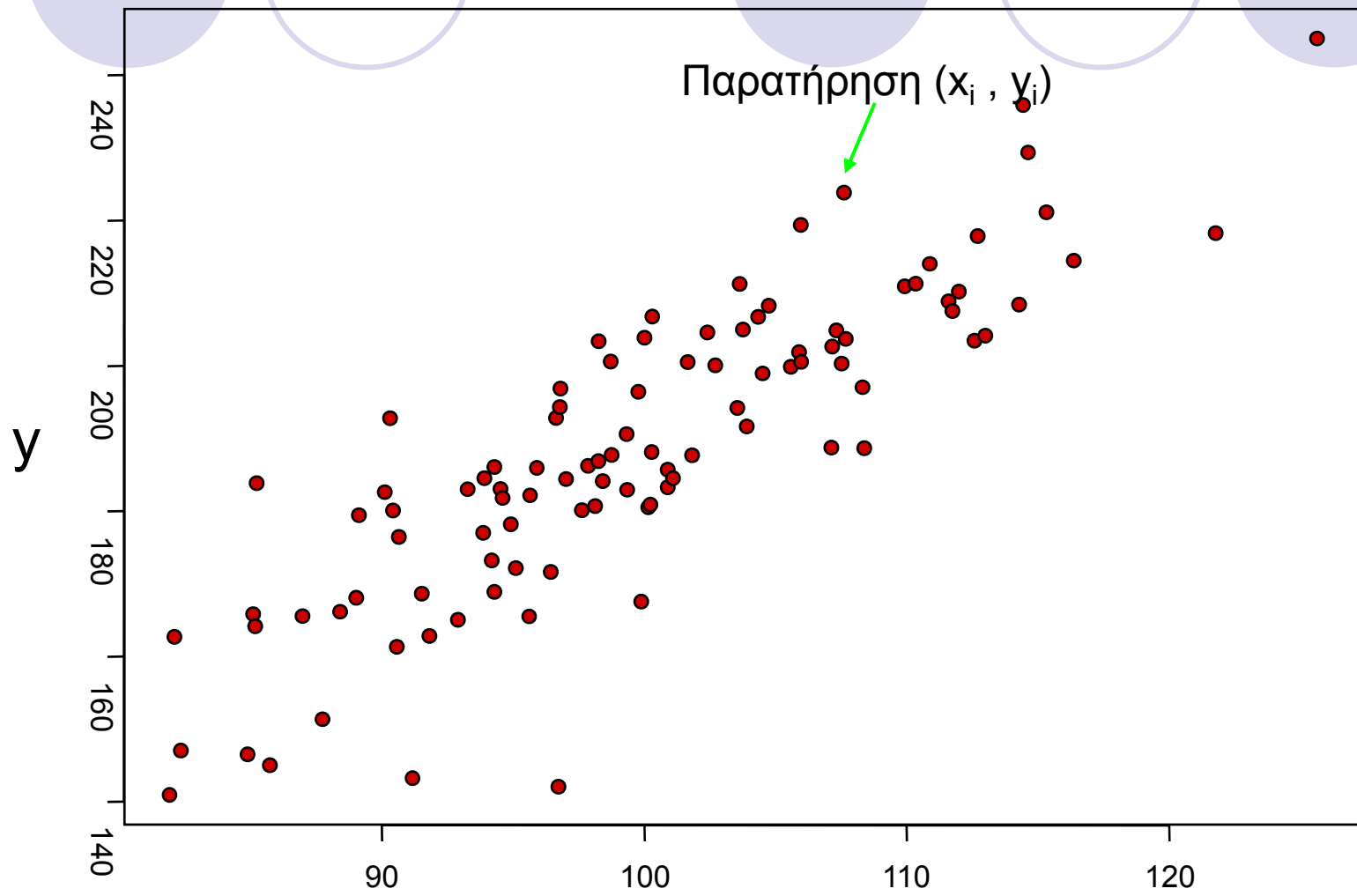
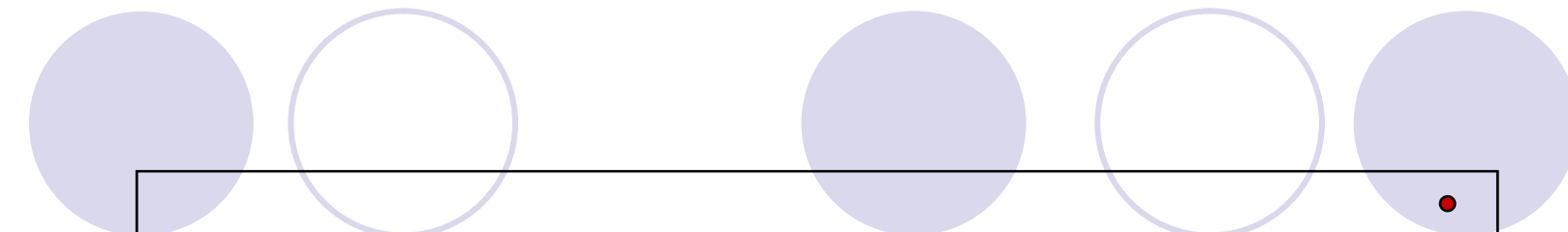
# Συσχέτιση και παλινδρόμηση

- **Συσχέτιση:** Πόσο έντονα μία αλλαγή στο ένα μέγεθος επηρεάζει το άλλο μέγεθος;
- **Παλινδρόμηση:** Αν ξέρουμε την τιμή του  $x$  μπορούμε να προβλέψουμε το  $y$ ?

Ξεκινάμε με ένα **διάγραμμα διασποράς** (scatterplot)

$n$  παρατηρήσεις από το  $x$  :  $x_1, x_2, \dots, x_n$

$n$  παρατηρήσεις από το  $y$  :  $y_1, y_2, \dots, y_n$



$m_x$  ο μέσος όρος των  $x$   
 $m_y$  ο μέσος όρος των  $y$

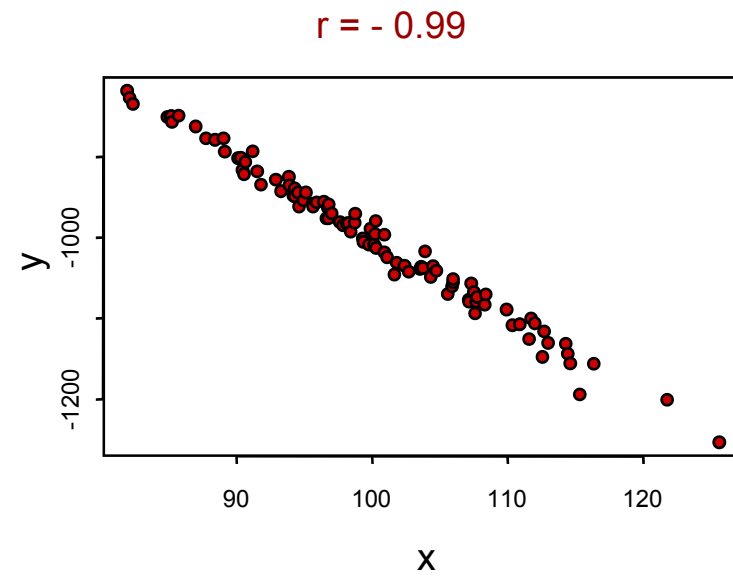
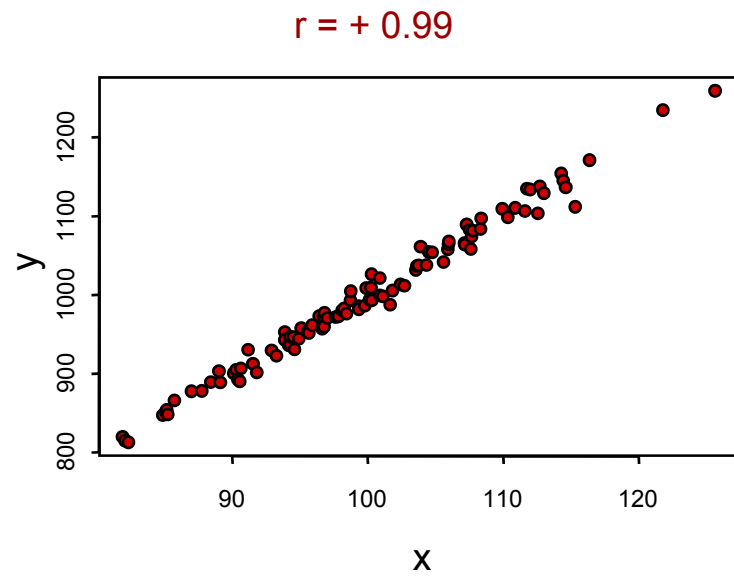
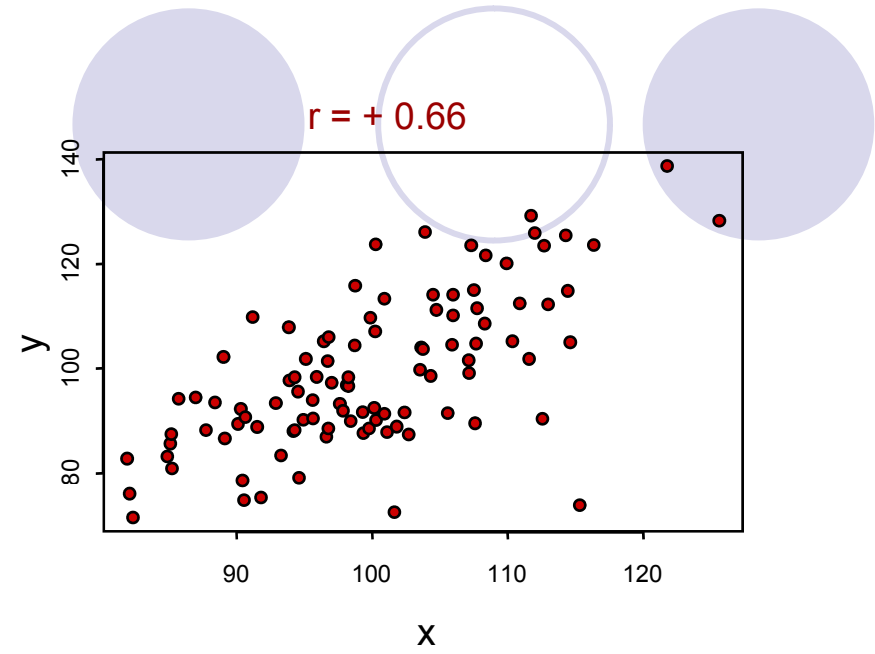
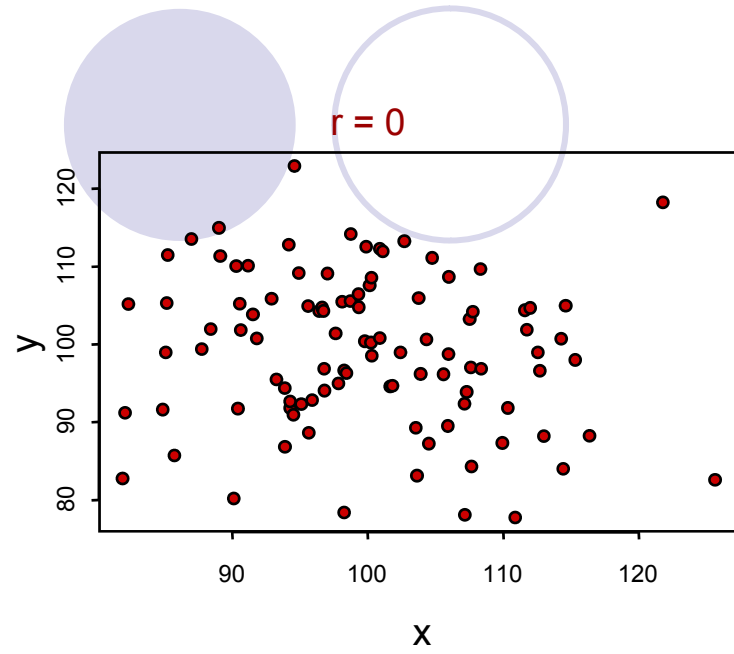
# Συντελεστής Συσχέτισης

Συνδιακύμανση

$$r = \frac{\sum (x_i - m_x)(y_i - m_y)}{\sqrt{\sum (x_i - m_x)^2 \sum (y_i - m_y)^2}} = \frac{\text{Cov}(x, y)}{sd_x sd_y}$$

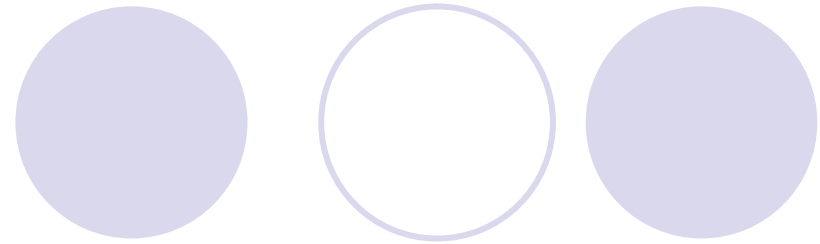
Τ.Α. ΤΟΥ X

- $-1 \leq r \leq 1$
- $r = 0$  καμία συσχέτιση
- $|r| = 1$  τέλεια συσχέτιση
- $r > 0$  ανάλογη συσχέτιση
- $r < 0$  αντιστρόφως ανάλογη συσχέτιση





# Ερμηνεία του $r$



- $0.75 \leq |r| \leq 1$  πολύ ισχυρή συσχέτιση
- $0.50 \leq |r| \leq 0.75$  ισχυρή συσχέτιση
- $0.30 \leq |r| \leq 0.50$  μέτρια συσχέτιση
- $0 \leq |r| \leq 0.30$  πολύ ασθενής συσχέτιση
  
- Έλεγχος
  - $H_0: r = 0$
  - Πίνακες
  - Βαθμοί ελευθερίας =  $n - 2$

# Υποθέσεις και προβλήματα

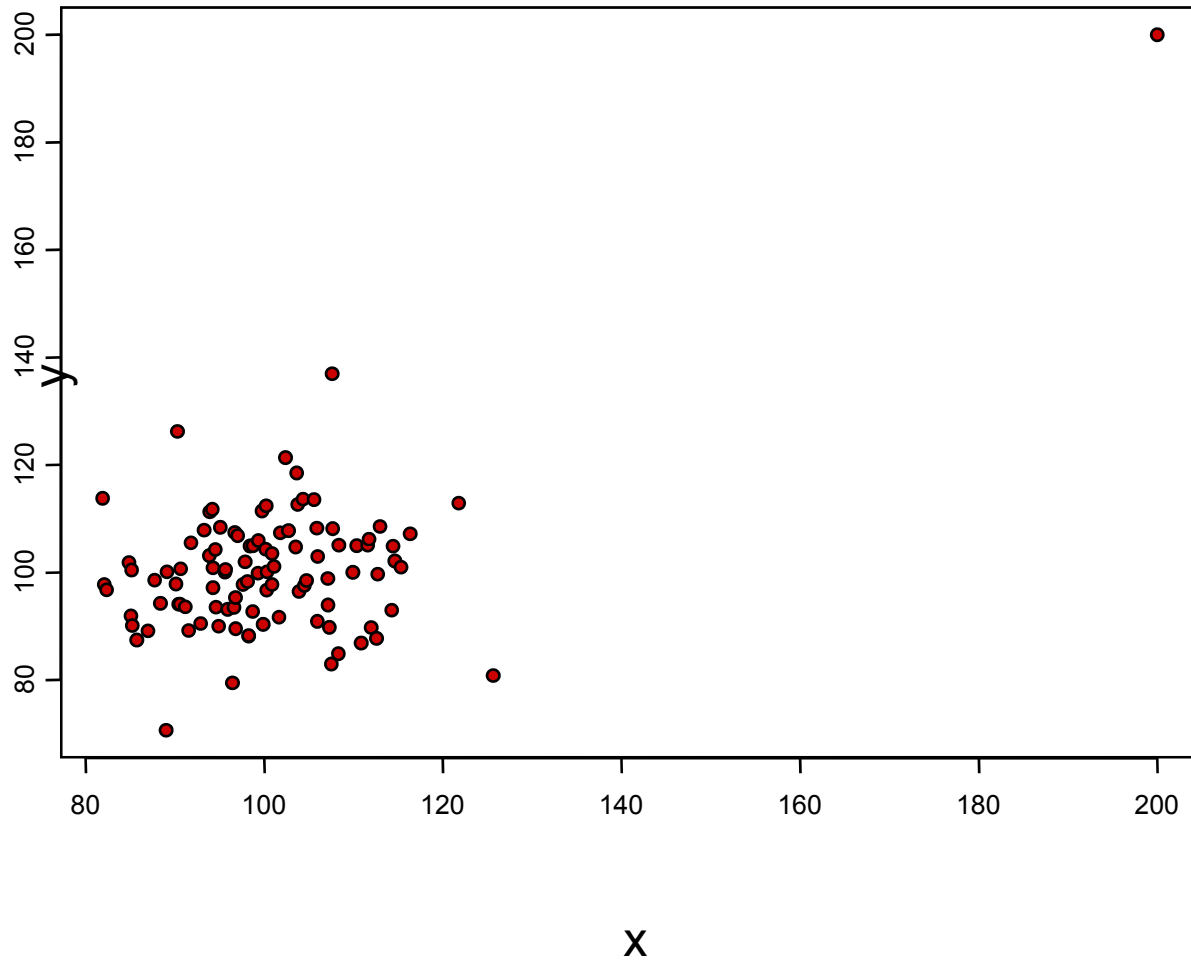
- Υπόθεση  
     $x, y \sim$  Κανονική κατανομή
- Αν δεν πληρείται η προϋπόθεση, τότε χρησιμοποιώ ‘μη παραμετρικούς συντελεστές’ – **spearman’s  $\rho$ , kendall’s  $\tau$**

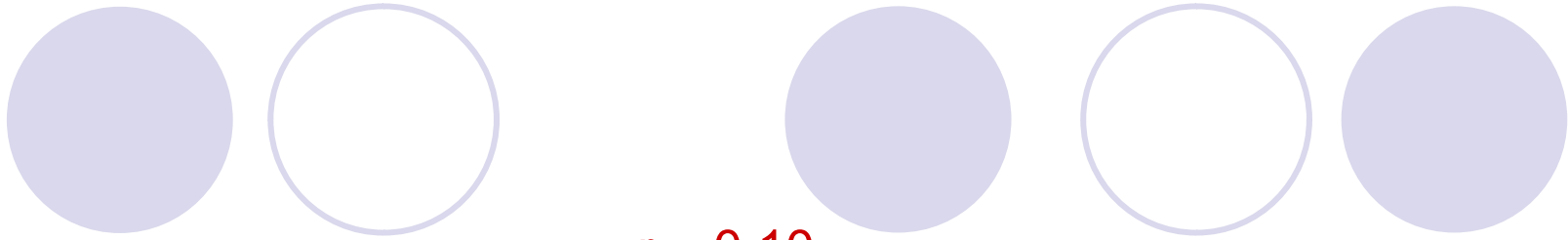
## Προσοχή!

- Μεμονωμένες παρατηρήσεις μπορεί να επηρεάσουν το  $r$ !
- Ο σ.σ.  $r$  δεν δίνει πληροφορίες για σχέση αιτίας αποτελέσματος
- Ο σ.σ.  $r$  δεν δίνει πληροφορίες για σχέση δόσης αποτελέσματος
- ΠΑΝΤΑ να κοιτάτε το γράφημα!

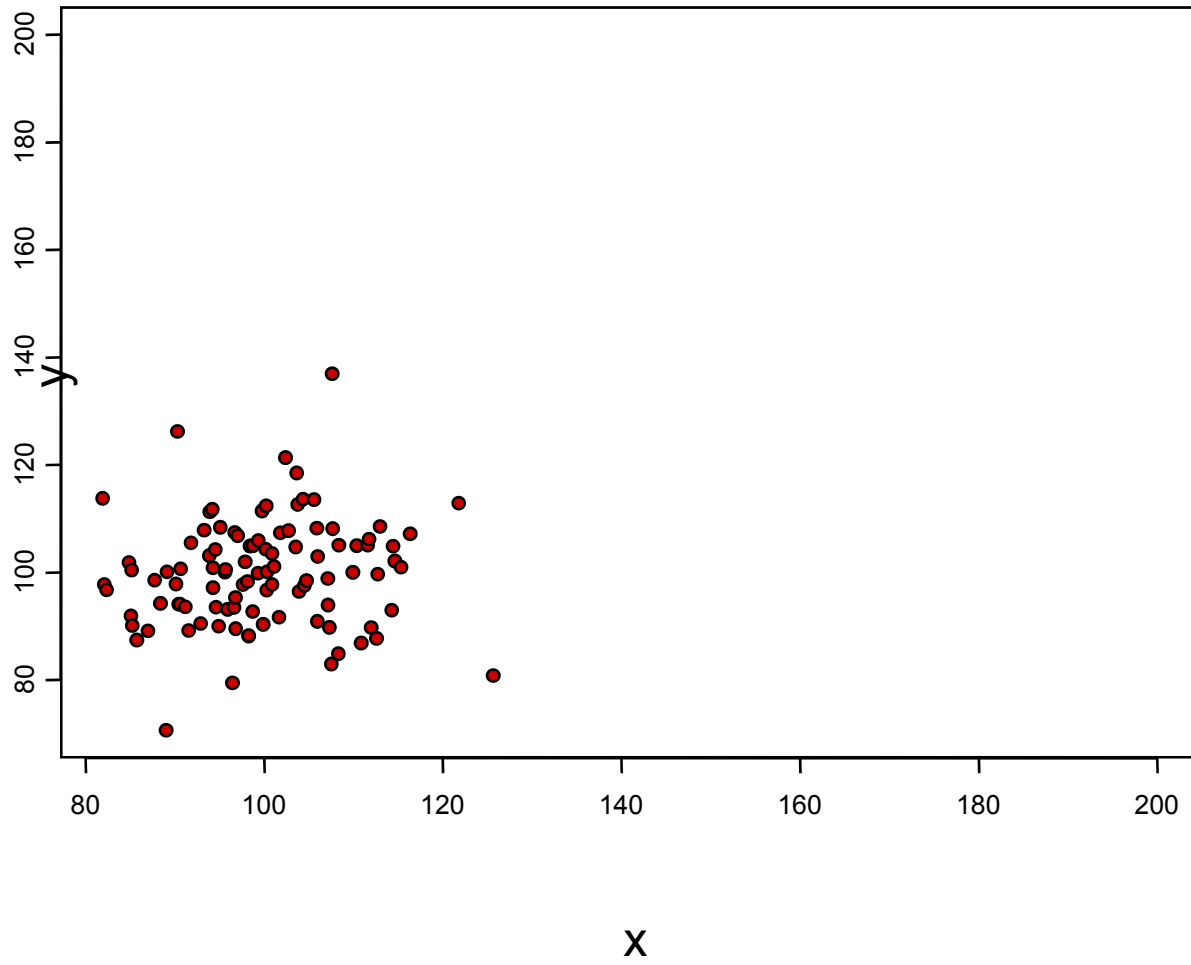


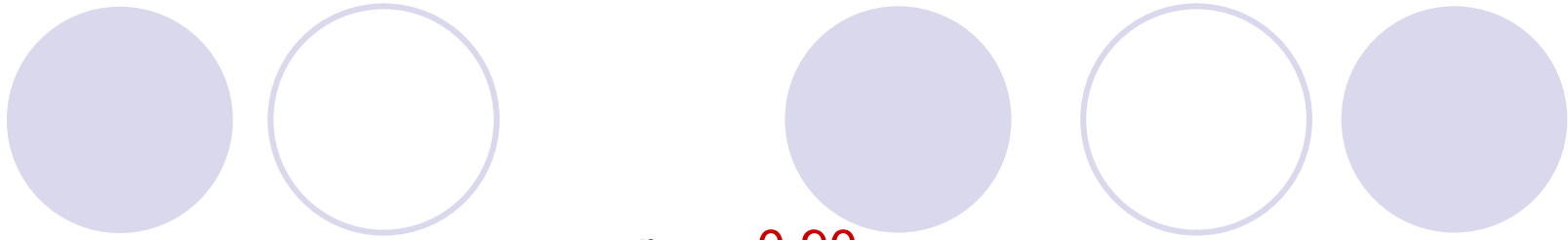
$r = 0.60$



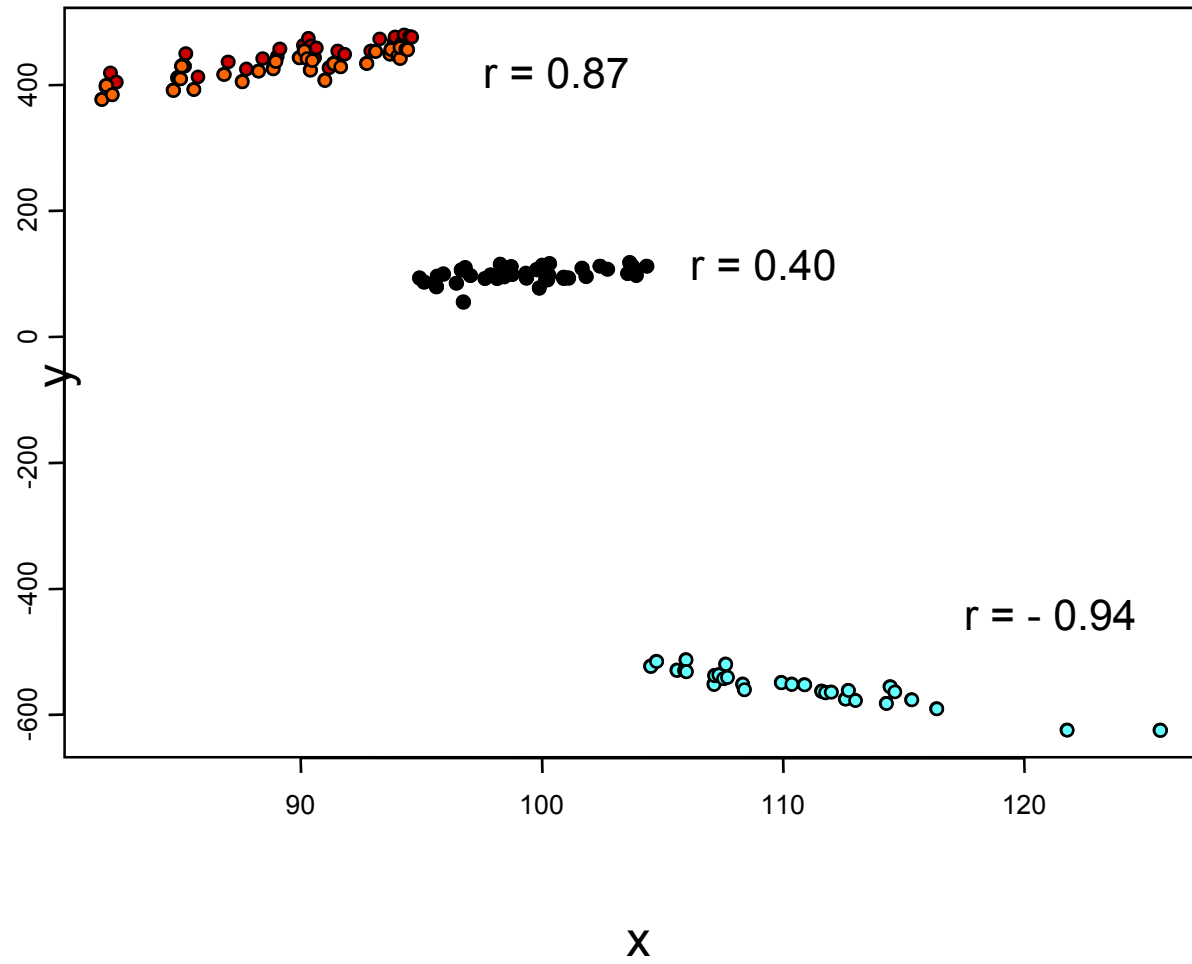


$r = 0.10$

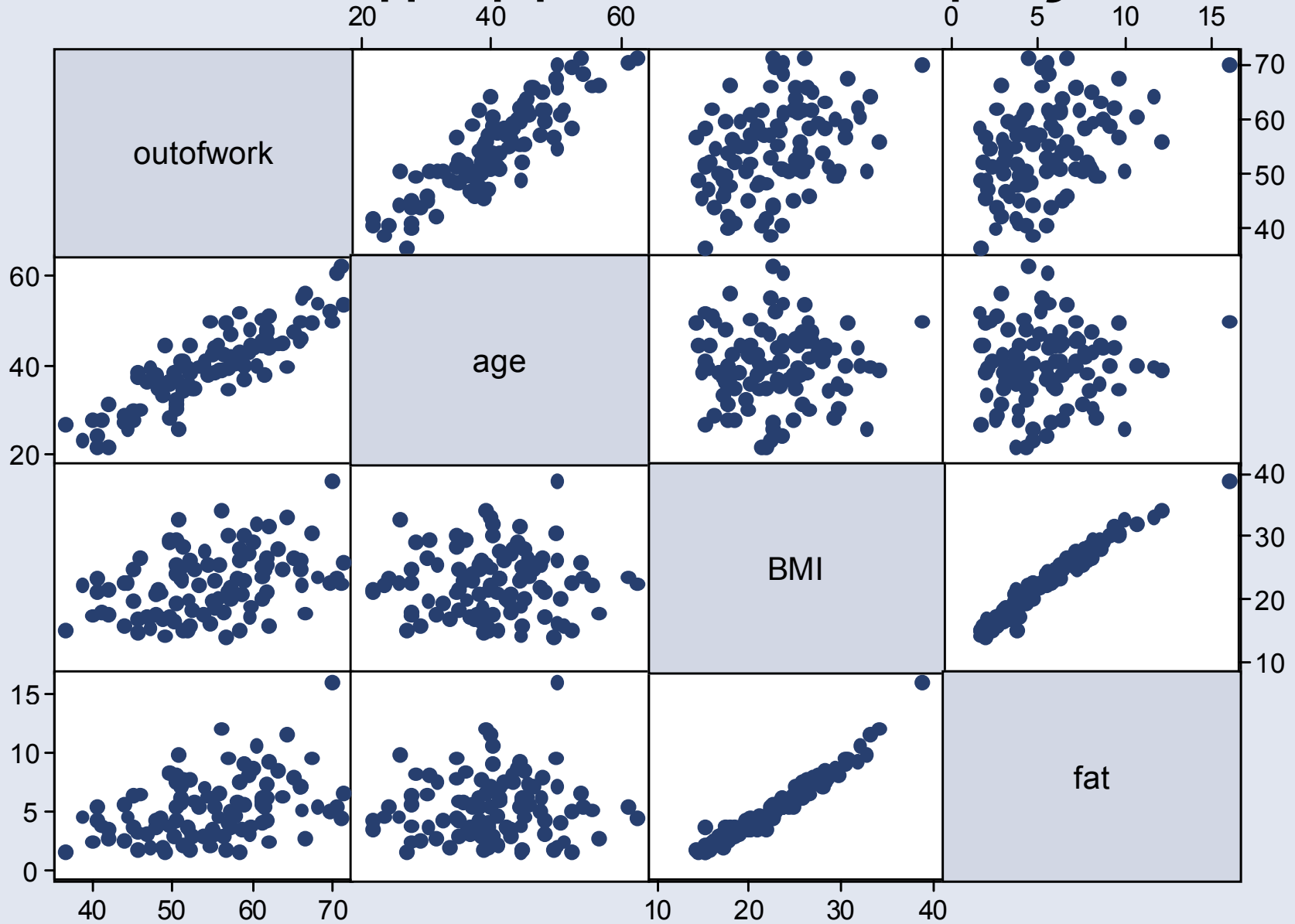




$r = -0.90$



# Διαγράμματα διασποράς



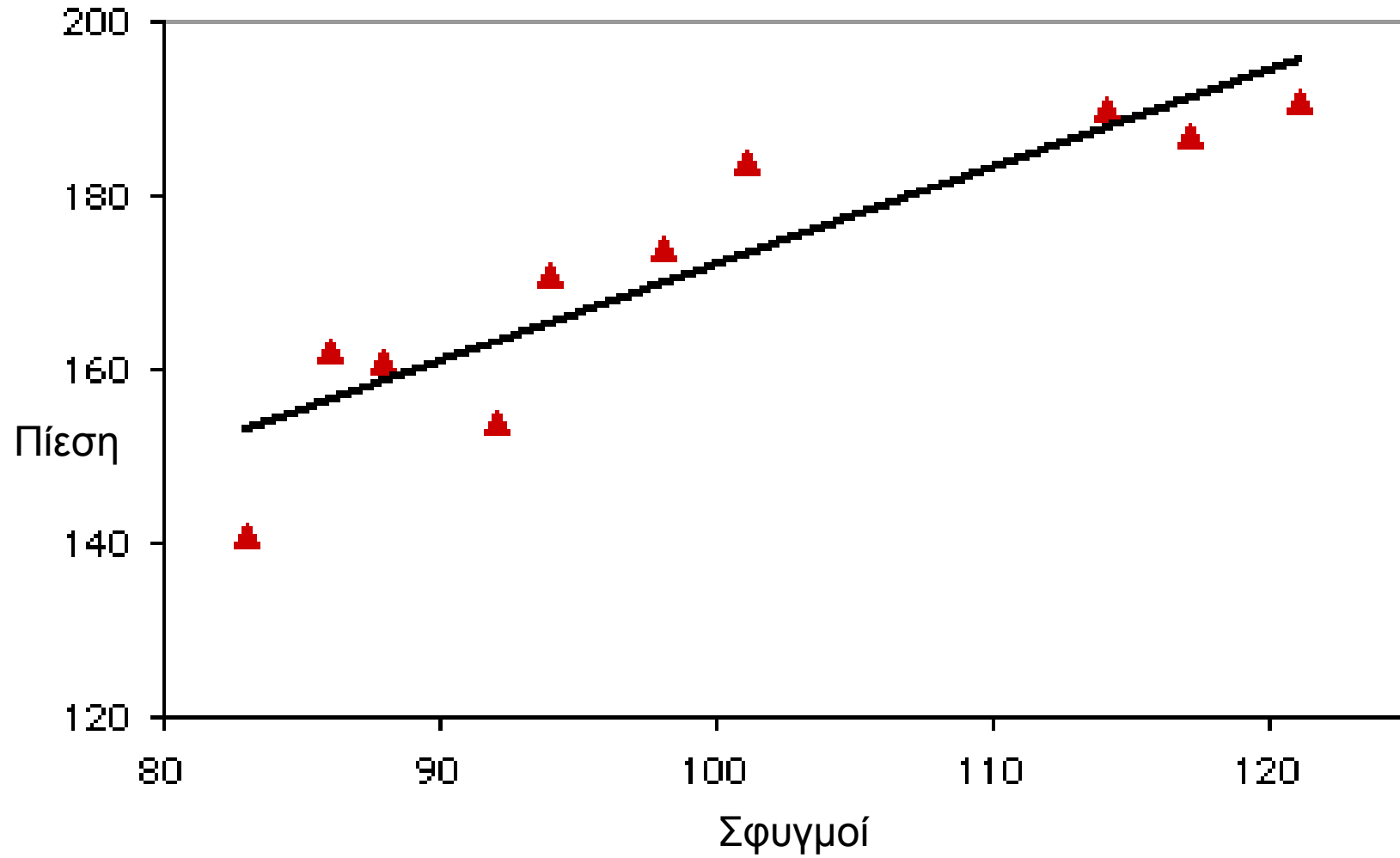
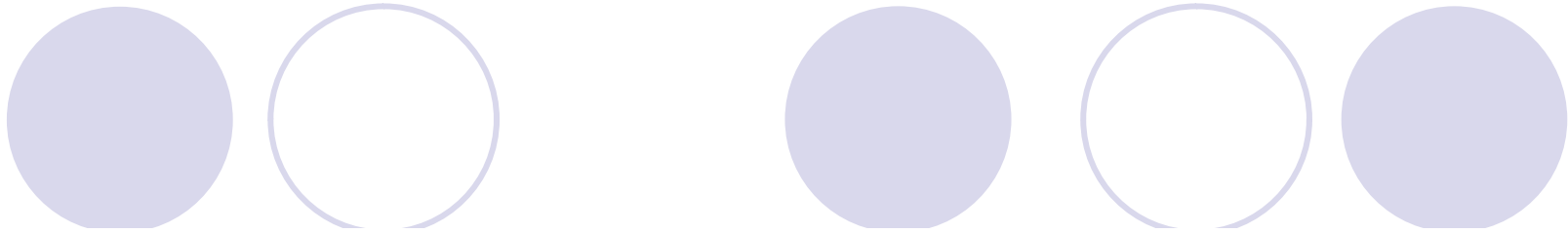
# Παράδειγμα

**Ασθενής**

**Σφυγμοί**

**Πίεση:**

1	83	141
2	86	162
3	88	161
4	92	154
5	94	171
6	98	174
7	101	184
8	114	190
9	117	187
10	121	191





## Ανάλυση

- $H_0$ : Δεν υπάρχει συσχέτιση μεταξύ σφυγμών και πίεσης

- Υπολογίζω  $r = 0.903$

- Βαθμοί ελευθερίας =  $n-2 = 10 - 2 = 8$

Κρίσιμη τιμή από τον πίνακα με 8 β.ε. = 0.632

- Η τιμή 0.903 είναι μεγαλύτερη από την κρίσιμη

→ απορρίπτουμε την  $H_0$

- Συμπέρασμα:

σφυγμοί και πίεση δεν είναι ανεξάρτητα μεγέθη

## Level of Significance (p) for a Two-Tailed Test

df (n-2):	0.10	0.05
1	0.988	0.997
2	0.900	0.950
3	0.805	0.878
4	0.729	0.811
5	0.669	0.754
6	0.622	0.707
7	0.582	0.666
8	0.549	0.632
9	0.521	0.602
10	0.497	0.576
11	0.476	0.553
12	0.458	0.532
13	0.441	0.514
14	0.426	0.497
15	0.412	0.482
16	0.400	0.468
17	0.389	0.456
18	0.378	0.444
19	0.369	0.433
20	0.360	0.423
21	0.352	0.413
22	0.344	0.404
23	0.337	0.396
24	0.330	0.388
25	0.323	0.381
26	0.317	0.374
27	0.311	0.367

Αυτές οι τιμές ορίζουν από πιο σημείο και πέρα έχουμε ακραίες' τιμές.

## Ανάλυση

$H_0$ : Δεν υπάρχει συσχέτιση μεταξύ σφυγμών και πίεσης

- Υπολογίζω  $r = 0.903$
- **Βαθμοί ελευθερίας** =  $n-2 = 10 - 2 = 8$
- Κρίσιμη τιμή από τον πίνακα με 8 β.ε. = 0.632
- Η τιμή 0.903 είναι μεγαλύτερη από την κρίσιμη  
→ απορρίπτουμε την  $H_0$
- **Συμπέρασμα:**  
σφυγμοί και πίεση δεν είναι ανεξάρτητα μεγέθη

Για επόμενη φορά...

Association between raised body temperature  
and acute mountain sickness: cross sectional  
study. *British Medical Journal*, 315, 403-4.