



# Περιγραφική Στατιστική

Κωνσταντίνος Μπουγιούκας, MSc, PhD  
mpougioukas@auth.gr

2025-2026



## Στόχοι του σημερινού μαθήματος

- ☒ Σύνοψη κατηγορικών δεδομένων και γραφήματα
- ☒ Μέτρα κεντρικής θέσης (μέσος όρος, επικρατούσα τιμή, διάμεσος)
- ☒ Μέτρα διασποράς (εύρος, διακύμανση, τυπική απόκλιση)
- ☒ Γραφήματα για ποσοτικά δεδομένα





# Συνοψίζοντας τις κατηγορικές μεταβλητές

## Summarizing categorical data

### (Frequency Statistics)



## Πίνακας απόλυτων, σχετικών και αθροιστικών συχνοτήτων

- Μεταβλητή **φύλο** (sex)

Sex	Absolute frequency, n	Relative frequency (%)	Cumulative relative frequency (%)
male	109	42.2	42.2
female	149	57.8	100.0
Total	258	100.0	

### Ορολογία:

- Absolute frequency: Απόλυτη συχνότητα
- Relative frequency: Σχετική συχνότητα (εκφρασμένη ως ποσοστό %)
- Cumulative relative frequency: Αθροιστική σχετική συχνότητα (εκφρασμένη ως ποσοστό %)



## Πίνακας απόλυτων, σχετικών και αθροιστικών συχνοτήτων

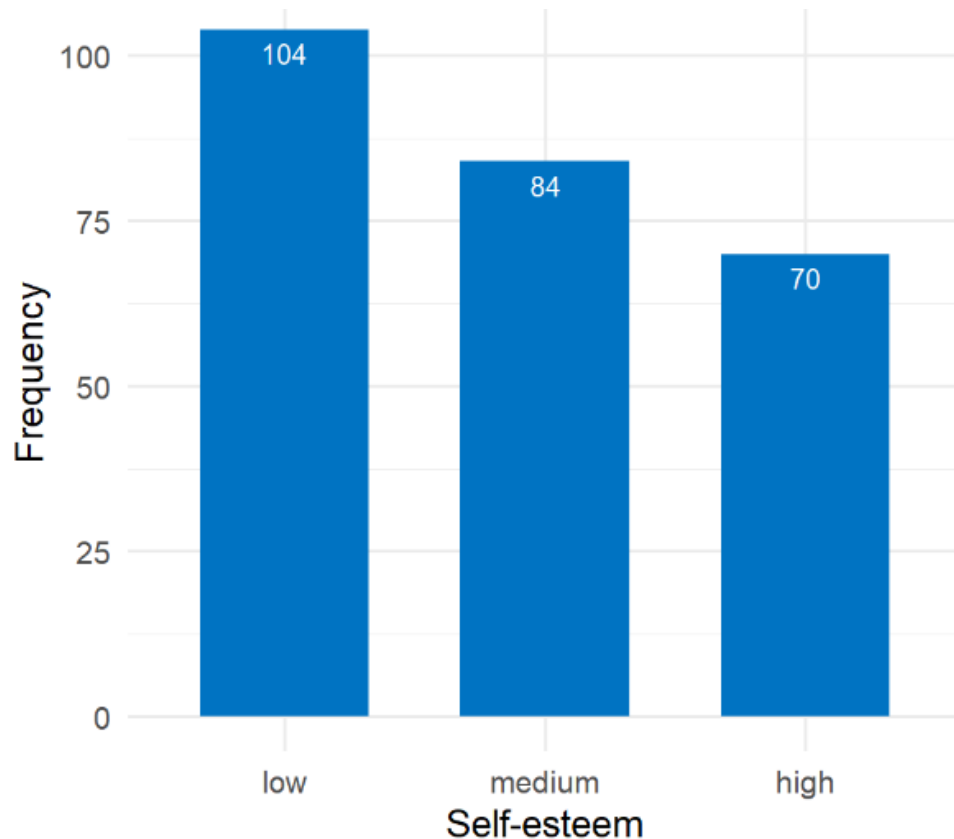
- Μεταβλητή αυτοεκτίμησης (self-esteem)

Self-esteem	Absolute frequency n	Relative frequency (%)	Cumulative percentage (%)
low	104	40.3	40.3
medium	84	32.6	72.9
high	70	27.1	100.0
Total	258	100.0	

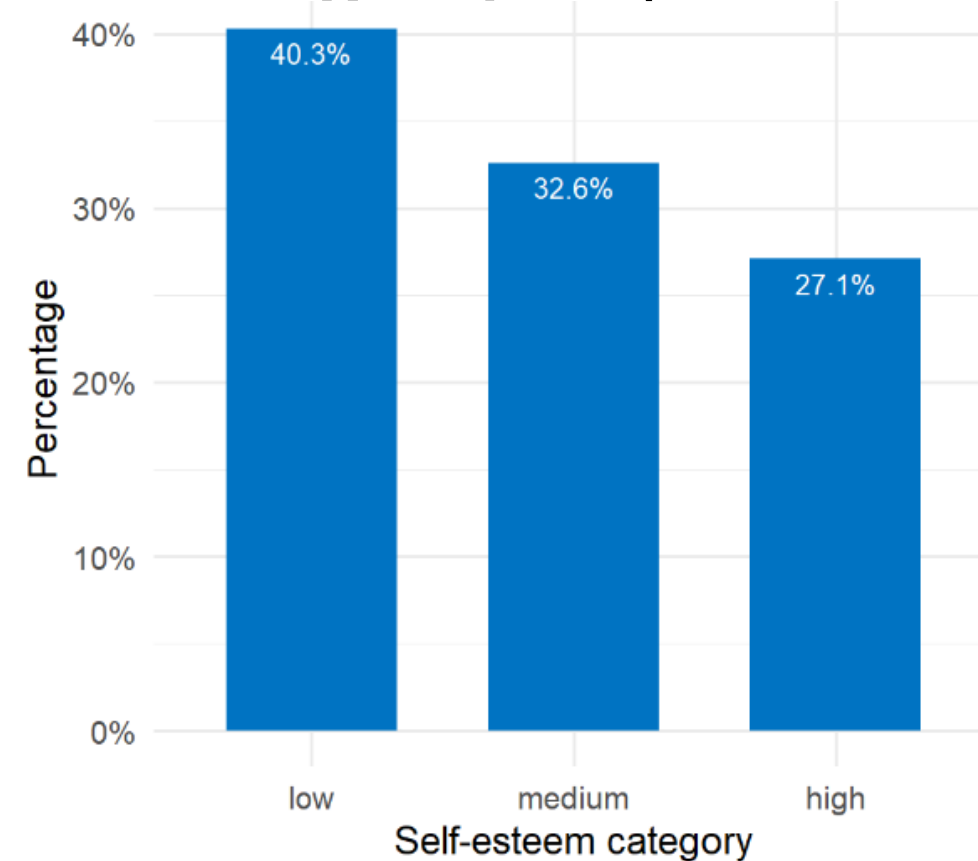
Πώς μπορούμε να αναπαραστήσουμε γραφικά της απόλυτες και σχετικές συχνότητες;

## Bar plots (self-esteem)

### Ραβδόγραμμα συχνοτήτων



### Ραβδόγραμμα σχετικών συχνοτήτων (ποσοστών)



- Όλες οι ράβδοι πρέπει να έχουν ίσο πλάτος και ίση απόσταση μεταξύ τους.
- Το ύψος κάθε ράβδου πρέπει να αντιστοιχεί στα δεδομένα που αντιπροσωπεύει.
- Οι ράβδοι πρέπει να σχεδιάζονται με αρχή το μηδέν.



# Σχέση δυο κατηγορικών μεταβλητών

## Πίνακας συνάφειας (contingency table) με απόλυτες συχνότητες

- Φύλο και αυτοεκτίμηση

		Self-esteem			Marginal Total
		low	medium	high	
Sex	male	26	47	36	109
	female	78	37	34	149
Marginal Total		104	84	70	Total 258



## Πίνακας συνάφειας (contingency table) με σχετικές συχνότητες

- Φύλο και αυτοεκτίμηση

		Self-esteem			Marginal Total
		low	medium	high	
Sex	male	26/258 x100 (10.1%)	47/258 x100 (18.2%)	36/258 x100 (14.0%)	109
	female	78/258 x100 (30.2%)	37/258 x100 (14.3%)	34/258 x100 (13.2%)	149
Marginal Total		104	84	70	Total 258





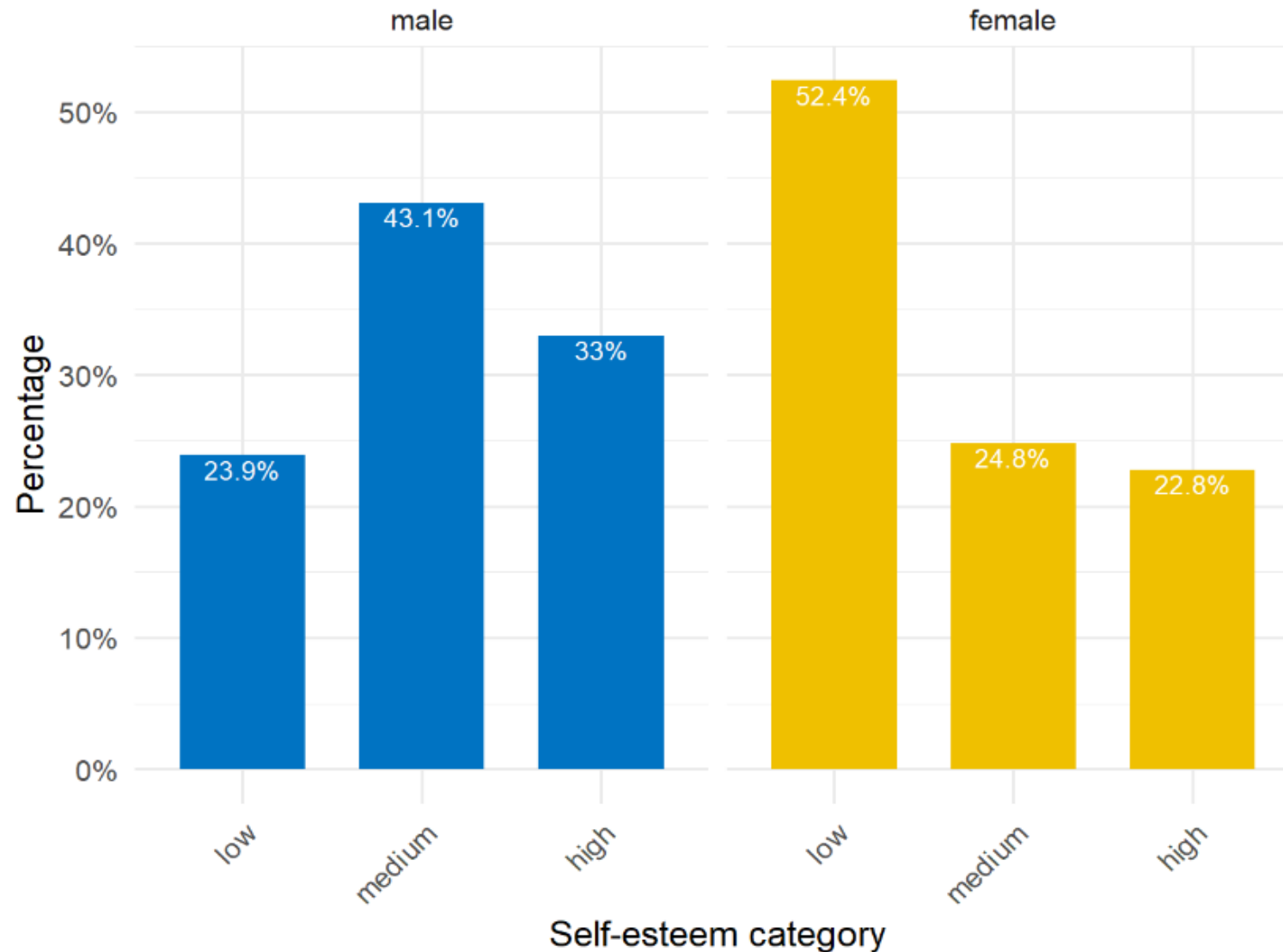
## Πίνακας συνάφειας (contingency table) με σχετικές συχνότητες σειρών

- Φύλο και αυτοεκτίμηση

		Self-esteem			Marginal Total
		low	medium	high	
Sex	male	26/109 x100 (23.9%)	47/109 x100 (43.1%)	36/109 x100 (33.0%)	109
	female	78/149 x100 (52.4%)	37/149 x100 (24.8%)	34/149 x100 (22.8%)	149
Marginal Total		104	84	70	Total 258



## Side-by-side bar plot





# Συνοψίζοντας τις ποσοτικές μεταβλητές

## Summarizing continuous data



## Μέτρα κεντρικής θέσης (Measures of central location)

- ☒ Αριθμητική μέση τιμή (mean)
- ☒ Η διάμεσος (median)
- ☒ Η επικρατούσα τιμή (mode)



## Αριθμητική μέση τιμή (Arithmetic mean) του δείγματος

Το άθροισμα όλων των τιμών ενός συνόλου αριθμητικών δεδομένων, διαιρούμενο με τον συνολικό αριθμό τους.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_{n-1} + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Π.χ. 5      4      12      8      4

$$\bar{x} = \frac{5 + 4 + 12 + 8 + 4}{5} = \frac{33}{5} = 6.6$$

- + Χρησιμοποιεί όλα τα δεδομένα και είναι το σημείο ισορροπίας τους
- + Έχει καλές μαθηματικές ιδιότητες.
- Επηρεάζεται από ακραίες τιμές. Ποια θα ήταν η μέση τιμή αν είχαμε μια επιπλέον τιμή π.χ. 24 στα δεδομένα μας.



## Διάμεσος (median) του δείγματος

Η τιμή που χωρίζει τη διατεταγμένη σειρά των παρατηρήσεων σε δύο ίσα μέρη (50% αριστερά, 50% δεξιά).

$$md = \begin{cases} x_{(\frac{n+1}{2})}, & \text{for } n \text{ odd} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{for } n \text{ even} \end{cases}$$

Δεδομένα:                      5                      4                      12                      8                      4

Διατεταγμένες τιμές:              4                      4                      **5**                      8                      12

Διάμεσος = 5

- + Ανθεκτικό μέτρο στις ακραίες τιμές. Ποιο θα είναι το αποτέλεσμα αν προστεθεί η τιμή 24;
- Λαμβάνει υπόψη μόνο την διάταξη των τιμών και όχι τις πραγματικές τιμές.



## Επικρατούσα τιμή (mode) του δείγματος

Η πιο συχνά εμφανιζόμενη τιμή στο δείγμα.

Π.χ. 5      4      12      8      4

Επικρατούσα τιμή = 4



## Μέτρα διασποράς (Measures of Dispersion)

- ☒ Εύρος των τιμών του δείγματος (range)
- ☒ Ενδοτεταρτημοριακό εύρος (interquartile range) IQR
- ☒ Διακύμανση (variance)
- ☒ Τυπική απόκλιση (standard deviation)





## Εύρος των τιμών του δείγματος (range)

Η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής σε ένα σύνολο δεδομένων.

$$\text{Range} = \text{max} - \text{min} = x_{(n)} - x_{(1)}$$

Π.χ. 5      4      12      8      4

**4**      4      5      8      **12**

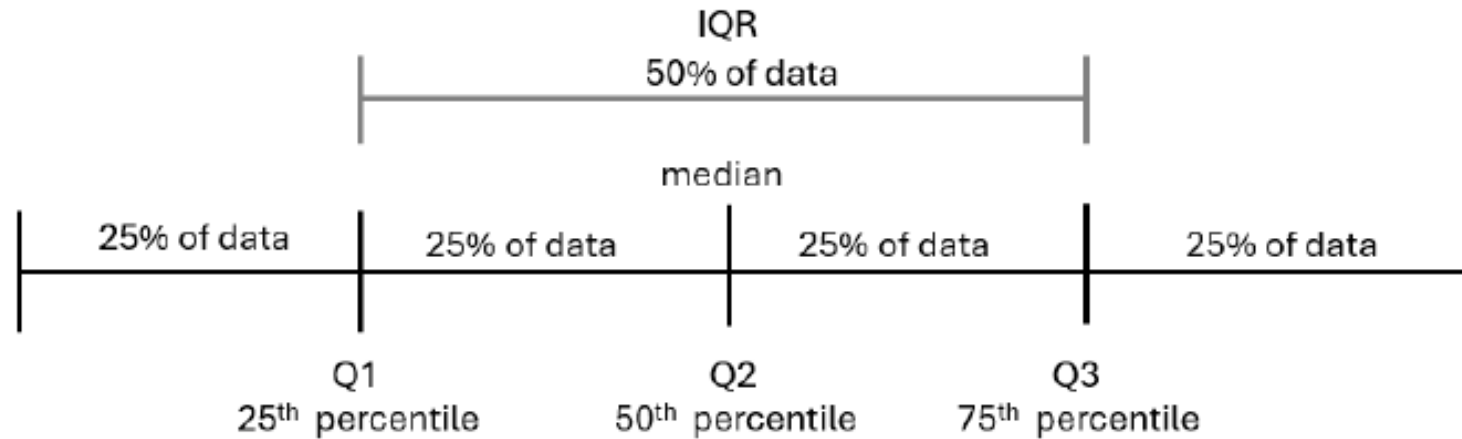
$$\text{Εύρος} = \text{Μέγιστη τιμή} - \text{Ελάχιστη τιμή} = 12 - 4 = 8$$

+ Απλό στον υπολογισμό.

- Επηρεάζεται από ακραίες τιμές.



## Τεταρτημόρια (quartiles) και εκατοστημόρια (percentiles)



**Interquartile range (IQR):** Η διαφορά μεταξύ του 3<sup>ου</sup> και 1<sup>ου</sup> τεταρτημορίου σε ένα διατεταγμένο σύνολο τιμών.

Π.χ. 5      4      12      8      4  
4      4      5      8      12  
Q1                      Q3

$$IQR = Q3 - Q1 = 8 - 4 = 4$$

+ Δεν επηρεάζεται ισχυρά από τις ακραίες τιμές.



## Διακύμανση (variance) του δείγματος

Η μέση τιμή των τετραγωνικών αποκλίσεων.

Π.χ. 5      4      12      8      4

$$\text{variance} = s^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n - 1}$$

Έχουμε υπολογίσει ότι:  $\bar{x} = 6.6$

$$\text{variance} = \frac{(5 - 6.6)^2 + (4 - 6.6)^2 + (12 - 6.6)^2 + (8 - 6.6)^2 + (4 - 6.6)^2}{5 - 1} = 11.8$$

- + Χρησιμοποιεί όλα τα δεδομένα.
- Επηρεάζεται από ακραίες τιμές.



## Τυπική απόκλιση (standard deviation) του δείγματος

Η τετραγωνική ρίζα της διακύμανσης. Μετρά την μεταβλητότητα των τιμών του δείγματος.

Π.χ. 5      4      12      8      4

$$s^2 = \text{variance} = 11.8$$

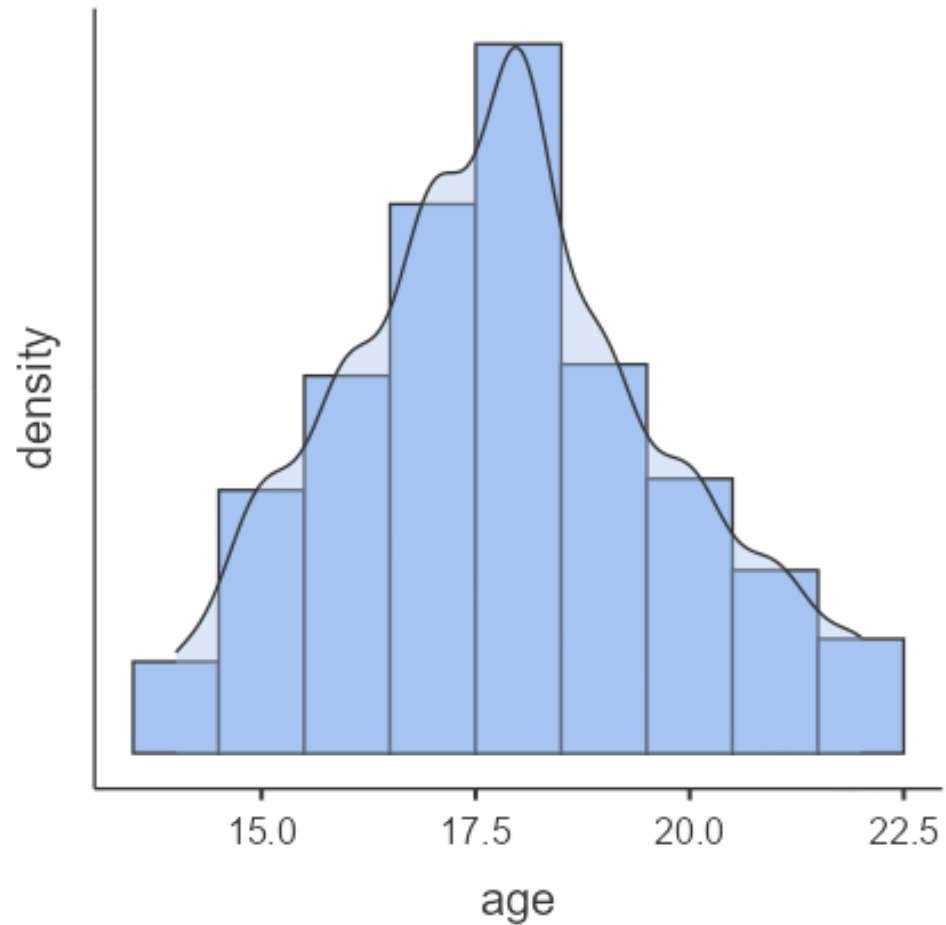
$$s = \sqrt{\text{variance}} = \sqrt{11.8} \approx 3.4$$

- + Χρησιμοποιεί όλα τα δεδομένα.
- + Είναι στην ίδια μονάδα μέτρησης με την αρχική μεταβλητή.
- Επηρεάζεται από ακραίες τιμές.

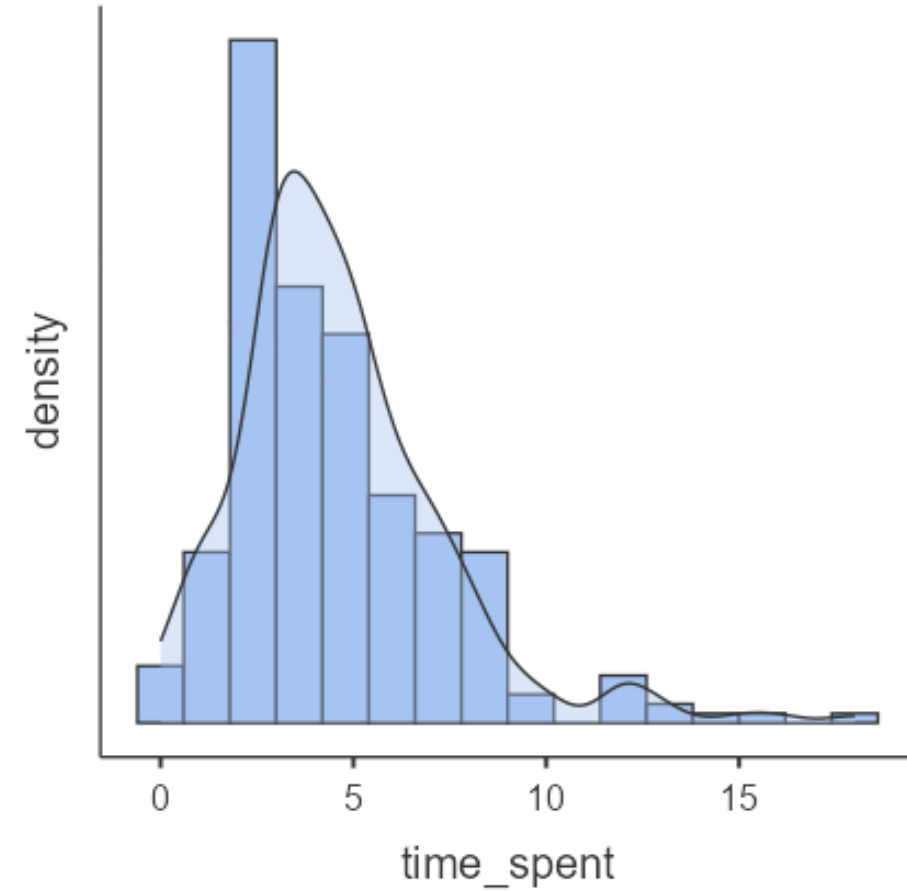


# Ιστογράμματα (histograms)

Συμμετρική κατανομή.

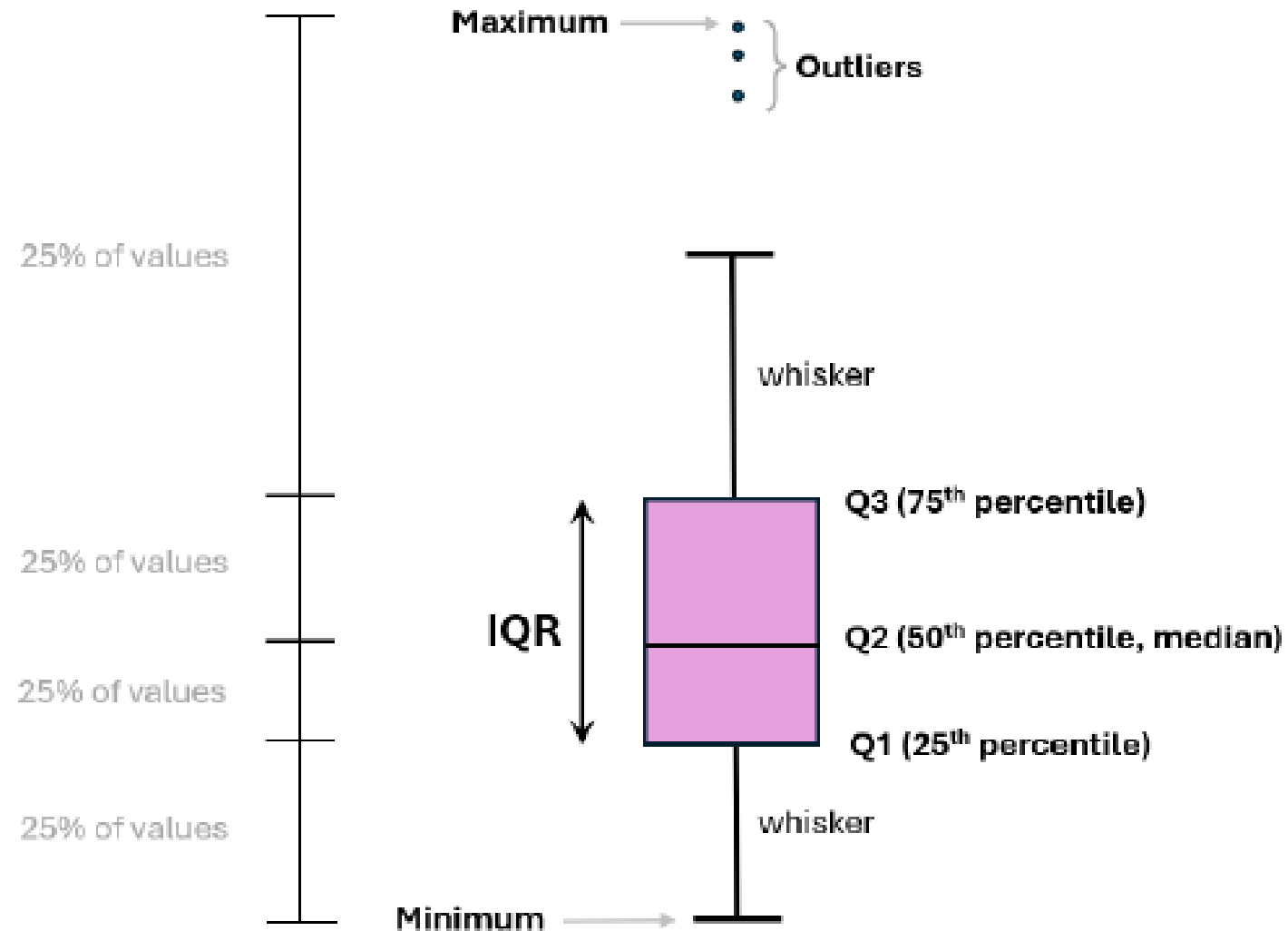


Δεξιά λοξότητα.



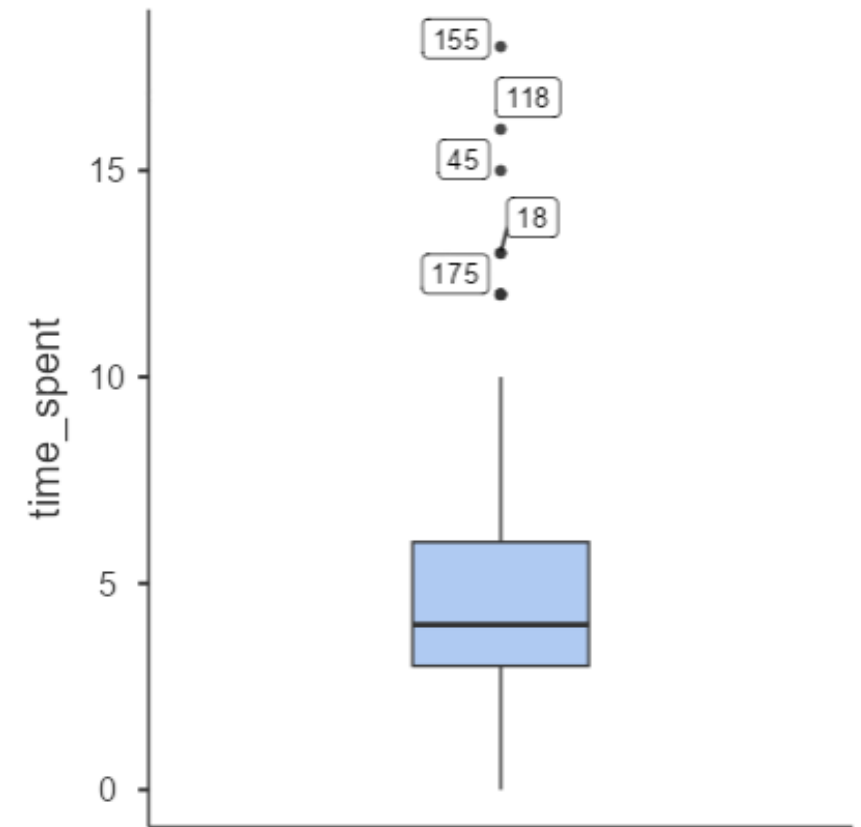
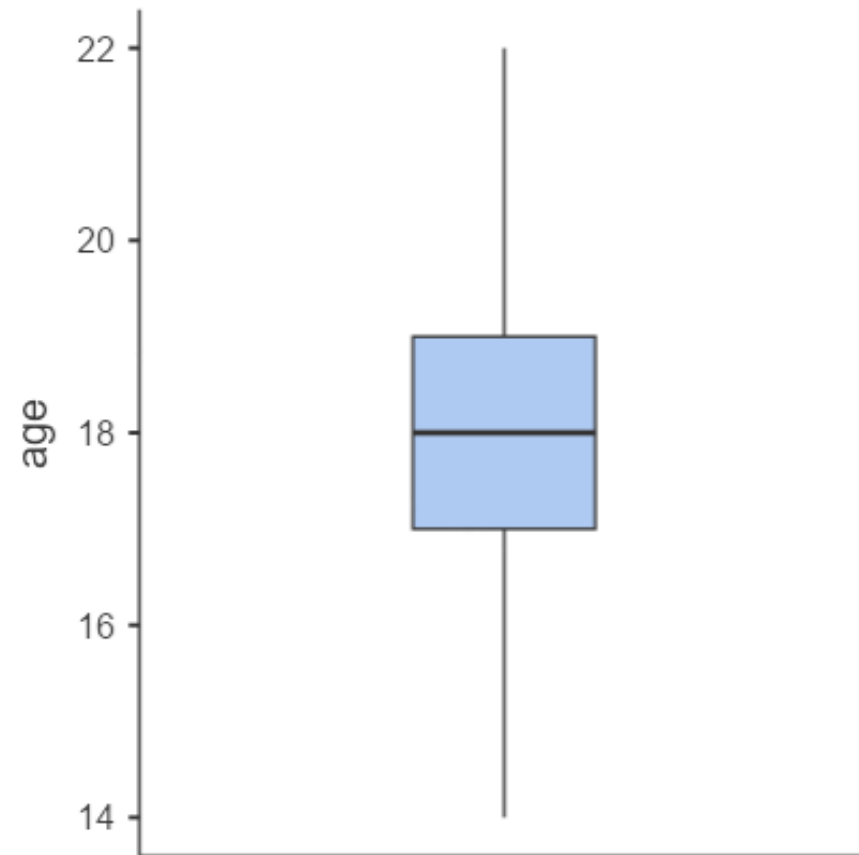


## Θηκόγραμμα (boxplot)





## Θηκογράμματα (boxplots)-Παραδείγματα

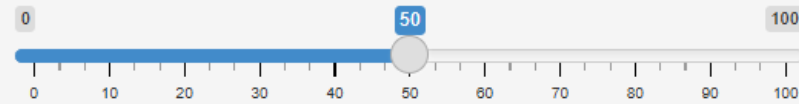




## Understanding Location and Dispersion with Dot Plots (Integer Data)

Dataset A

Mean:

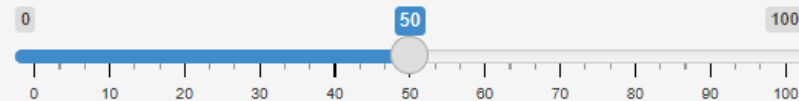


Standard Deviation:

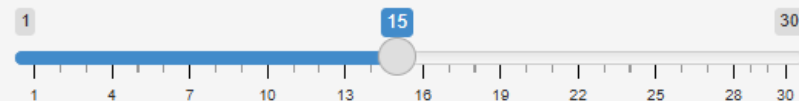


Dataset B

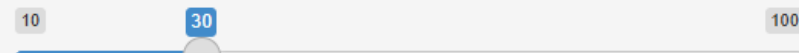
Mean:



Standard Deviation:

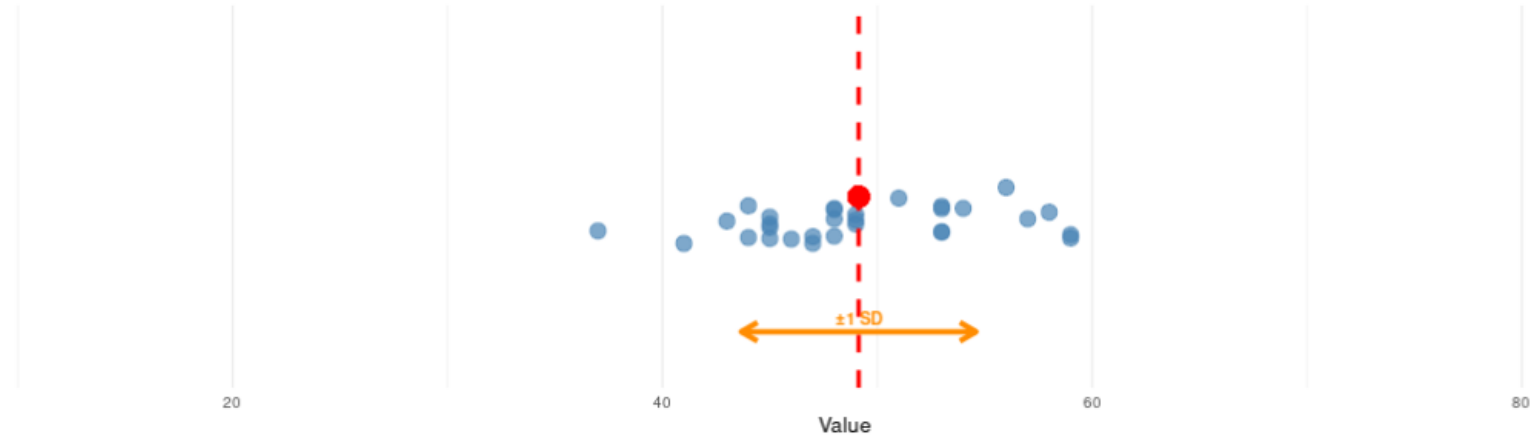


Sample Size (both datasets):



Click on the plots to add integer data points manually!

Dataset A (click to add points)



Dataset B (click to add points)

