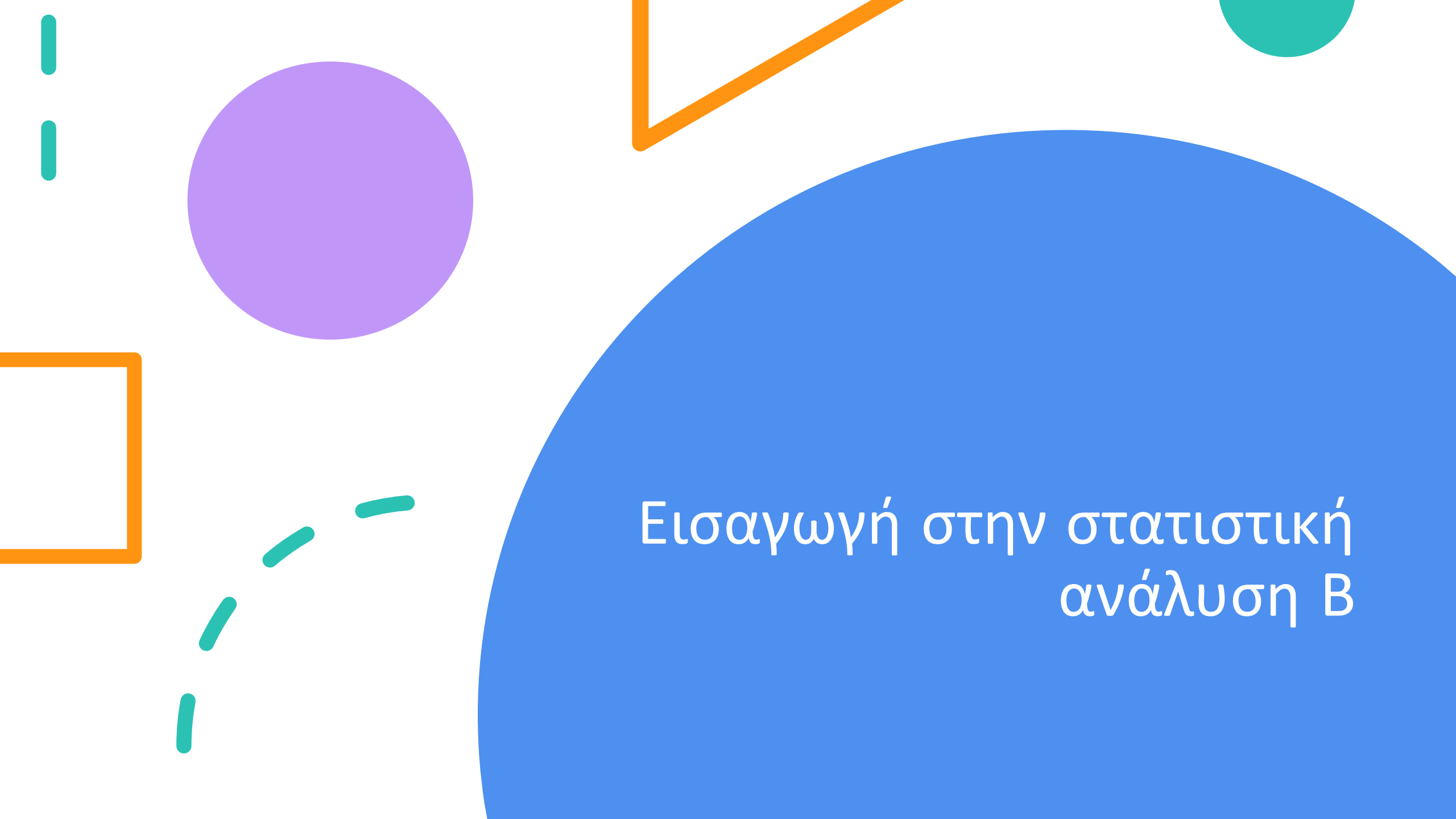




Εισαγωγή στην στατιστική ανάλυση Β

Περιγραφική Στατιστική

Είναι μία ομάδα στατιστικών μεθόδων για οργάνωση και περιγραφή των δεδομένων μας

Η παρουσίαση μπορεί να γίνει με μορφή:

- Πινάκων
- Διαγραμμάτων
- Μέτρων θέσης και διασποράς



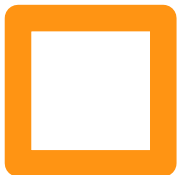
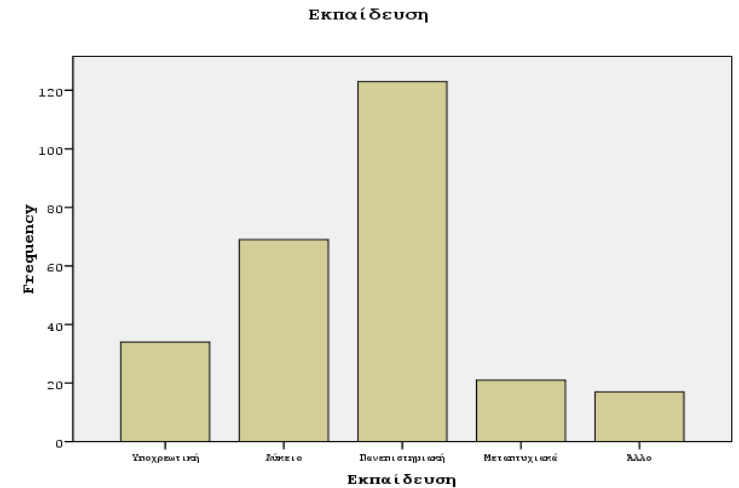
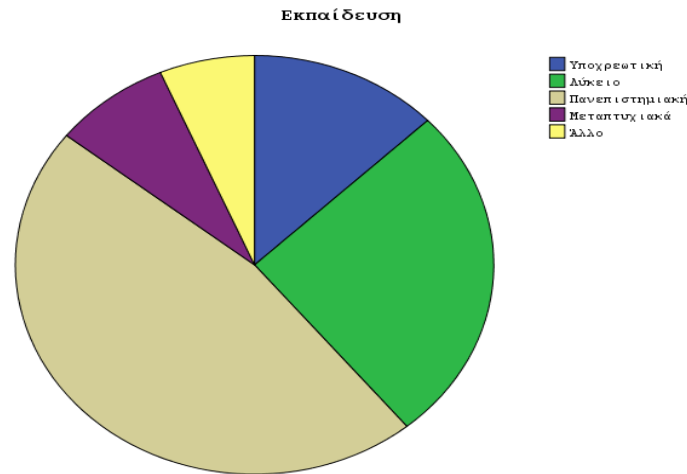
Περιγραφική Στατιστική – Κατηγορικά δεδομένα – απεικόνιση συχνοτήτων

Πίνακας συχνοτήτων

Κυκλικό διάγραμμα (pie chart)

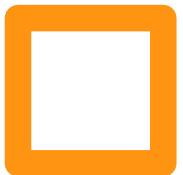
Ακιδωτό διάγραμμα (bar chart)

Εκπαίδευση					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Υποχρεωτική	35	13,1	13,1	13,1
	Λύκειο	72	27,0	27,0	40,1
	Πανεπιστήμιο	122	45,7	45,7	85,8
	Μεταπτυχιακό	20	7,5	7,5	93,3
	Άλλο	18	6,7	6,7	100,0
Total		267	100,0	100,0	



Περιγραφική Στατιστική – Κατηγορικά δεδομένα – μέτρα θέσης

- ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ (mode)
 - Είναι η τιμή που συναντάται τις περισσότερες φορές σε μία κατανομή, δηλαδή η τιμή με τη μεγαλύτερη συχνότητα (frequency)

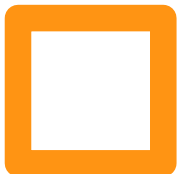


Περιγραφική Στατιστική – Αριθμητικά δεδομένα: Πίνακας Κατανομής Συχνοτήτων

Είναι ο πίνακας που δείχνει
πόσο συχνά συναντάται κάθε
τιμή μιας μεταβλητής

*Πως σας φαίνεται ως
περιγραφή των δεδομένων
μας;*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18,00	1	1,0	1,0	1,0
	19,00	3	3,1	3,1	4,1
	20,00	1	1,0	1,0	5,2
	21,00	2	2,1	2,1	7,2
	22,00	3	3,1	3,1	10,3
	23,00	2	2,1	2,1	12,4
	24,00	3	3,1	3,1	15,5
	25,00	5	5,2	5,2	20,6
	26,00	2	2,1	2,1	22,7
	27,00	3	3,1	3,1	25,8
	28,00	1	1,0	1,0	26,8
	29,00	3	3,1	3,1	29,9
	31,00	3	3,1	3,1	33,0
	32,00	3	3,1	3,1	36,1
	33,00	3	3,1	3,1	39,2
	34,00	2	2,1	2,1	41,2
	35,00	1	1,0	1,0	42,3
	36,00	2	2,1	2,1	44,3
	37,00	2	2,1	2,1	46,4
	38,00	2	2,1	2,1	48,5
	39,00	1	1,0	1,0	49,5
	40,00	2	2,1	2,1	51,5
	41,00	2	2,1	2,1	53,6
	42,00	2	2,1	2,1	55,7
	43,00	2	2,1	2,1	57,7
	44,00	1	1,0	1,0	58,8
	45,00	2	2,1	2,1	60,8
	46,00	1	1,0	1,0	61,9
	47,00	2	2,1	2,1	63,9
	48,00	1	1,0	1,0	64,9
	49,00	2	2,1	2,1	67,0
	50,00	1	1,0	1,0	68,0
	51,00	2	2,1	2,1	70,1
	52,00	1	1,0	1,0	71,1
	53,00	1	1,0	1,0	72,2
	54,00	1	1,0	1,0	73,2
	55,00	1	1,0	1,0	74,2
	56,00	1	1,0	1,0	75,3
	57,00	1	1,0	1,0	76,3
	58,00	1	1,0	1,0	77,3
	61,00	1	1,0	1,0	78,4
	62,00	2	2,1	2,1	80,4
	64,00	2	2,1	2,1	82,5
	65,00	1	1,0	1,0	83,5
	66,00	2	2,1	2,1	85,6
	68,00	1	1,0	1,0	86,6
	70,00	2	2,1	2,1	88,7
	71,00	1	1,0	1,0	89,7
	73,00	2	2,1	2,1	91,8
	75,00	2	2,1	2,1	93,8
	76,00	1	1,0	1,0	94,8
	77,00	2	2,1	2,1	96,9
	79,00	3	3,1	3,1	100,0
	Total	97	100,0	100,0	



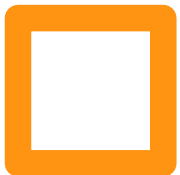
Περιγραφική Στατιστική – Αριθμητικά δεδομένα:

Πίνακας Ομαδοποιημένων Συχνοτήτων

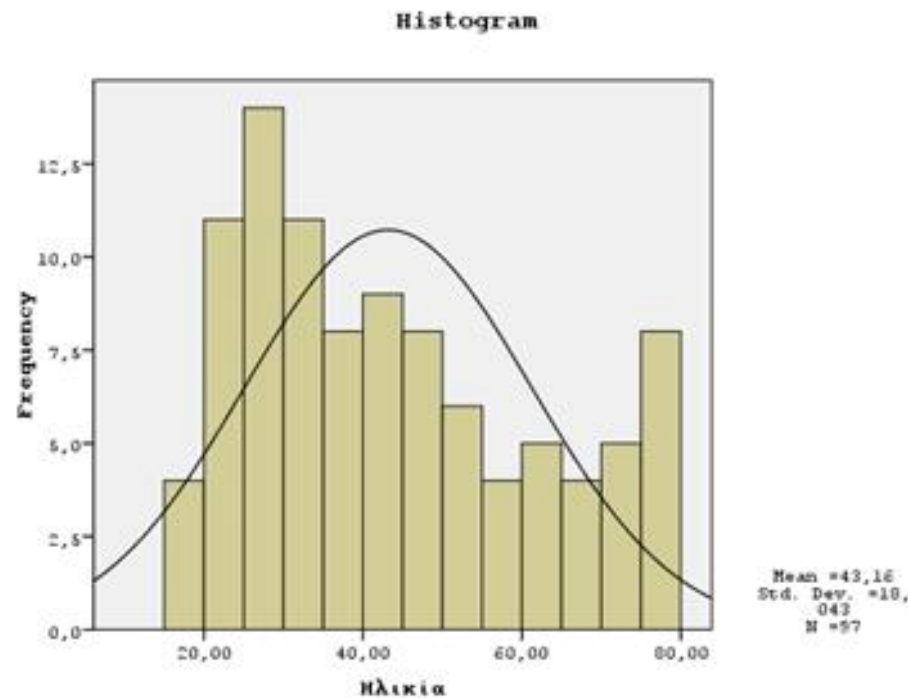
Είναι ο πίνακας στον οποίο παρουσιάζεται ο αριθμός των περιπτώσεων που εμπίπτουν σε κάθε **διάστημα τιμών**

Πως σας φαίνεται ως περιγραφή των δεδομένων μας;

Διαστήματα τιμών	Μέση τιμή	Συχνότητα	Σχετική συχνότητα
18-22	20	10	10,3
23-27	25	15	15,5
28-32	30	11	11,3
33-37	35	10	10,3
38-42	40	8	8,2
43-47	45	8	8,2
48-62	50	7	7,2
53-57	55	5	5,2
58-62	60	4	4,1
63-67	65	5	5,2
68-72	70	4	4,1
73-77	75	7	7,2
78-82	80	3	3,1
ΣΥΝΟΛΟ		97	99,9

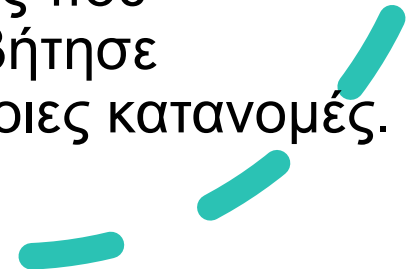
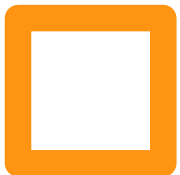


Περιγραφική Στατιστική – Αριθμητικά δεδομένα: Ιστόγραμμα



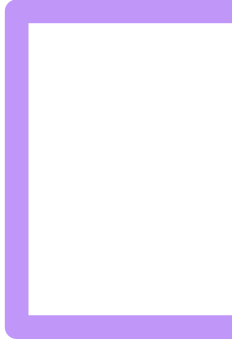
Ένα ιστόγραμμα απαντά σε μία απλή ερώτηση: **πόσο συχνά εμφανίζεται κάθε τιμή** σε ένα σύνολο αριθμητικών δεδομένων; Στον οριζόντιο άξονα βλέπουμε τις τιμές (π.χ. αριθμός συλλαβών μιας λέξης), στον κατακόρυφο πόσες φορές εμφανίστηκαν. Κάθε μπάρα λέγεται **κλάση**.

Ιστογράμματα χρησιμοποιούνται στη **υφολογία (stylistics)** — τη στατιστική ανάλυση ύφους. Π.χ. το μήκος λέξεων και προτάσεων είναι αξιόπιστοι δείκτες συγγραφικής ταυτότητας, ρυθμού και ρητορικής στρατηγικής. Ο αλγόριθμος που «αναγνώρισε» τον Σαίξπηρ ή αμφισβήτησε αποδόσεις κειμένων βασίζεται σε τέτοιες κατανομές.





Κατανομή δείγματος (sample distribution)

- Είναι ο **τρόπος με τον οποίο απλώνονται/κατανέμονται** οι τιμές μέσα στο δείγμα μου. Μπορώ να δω:
 - ποιες τιμές εμφανίζονται συχνά και ποιες σπάνια.
 - πού βρίσκεται η μέση.
 - πόσο διαφέρουν οι τιμές μεταξύ τους.
- 
- 

$$\bar{v} = \frac{\sum x_i v_i}{V}$$

$$= \frac{\sum (t_i - \bar{x})^2}{V} = \frac{1}{V} \left(\sum t_i^2 - \frac{(\sum t_i)^2}{V} \right)$$

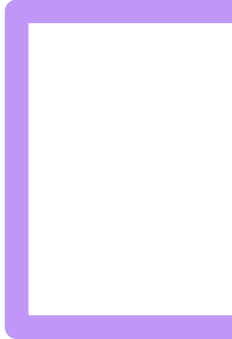
$$\frac{\sum v_i (x_i - \bar{x})^2}{V} = \frac{1}{V} \left(\sum v_i x_i^2 - \frac{(\sum v_i x_i)^2}{V} \right)$$

Μέτρα Θέσης και Μέτρα Διασποράς

Για την περιγραφή αριθμητικών μεταβλητών



Περιγραφική Στατιστική – Αριθμητικά δεδομένα: Μέτρα Θέσης και Διασποράς

- **ΔΕΙΚΤΕΣ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ (central tendency measures) – ΜΕΤΡΑ ΘΕΣΗΣ**
 - Αριθμητικές τιμές που δείχνουν το κέντρο της κατανομής
 - **ΔΕΙΚΤΕΣ ΔΙΑΣΠΟΡΑΣ (dispersion or variability) – ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ**
 - Στατιστικοί όροι για κάθε μέτρηση οι οποίοι αναφέρονται στη διακύμανση των τιμών μιας κατανομής
- 
- 

Μέτρα Θέσης (πού βρίσκεται η μέση;)

- **ΜΕΣΟΣ ΟΡΟΣ (mean, μέση τιμή, $\bar{M}$ ή $\bar{x}$)**
 - το άθροισμα όλων των τιμών διά το πλήθος τους.
Μειονέκτημα: επηρεάζεται από ακραίες τιμές.
- **ΔΙΑΜΕΣΟΣ (median, M_d)**
 - η τιμή που βρίσκεται στη μέση, αφού τακτοποιήσουμε τα δεδομένα από το μικρότερο στο μεγαλύτερο. Ανθεκτική στις ακραίες τιμές. Χωρίζει την κατανομή σε 2 ίσα τμήματα.
- **ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ (mode, M_o)**
 - Είναι η τιμή που εμφανίζεται πιο συχνά. Στο ιστόγραμμα αντιστοιχεί στην ψηλότερη μπάρα.

Περιγραφική Στατιστική – Αριθμητικά δεδομένα: Μέτρα Θέσης

Δείγμα 7 μαθητών για μέτρηση ύψους (cm):

160, 155, 165, 170, 160, 158, 162

Τα βάζουμε σε σειρά:

155, 158, 160, 160, 162, 165, 170

Μέσος όρος = (άθροισμα όλων των τιμών)/(πλήθος τιμών) = $1130/7 = 161,43$

Διάμεσος = η τιμή στο κέντρο της κατανομής = ??

Δεσπόζουσα = η τιμή με τη μεγαλύτερη συχνότητα = ??



Κατανομές τιμών δείγματος

Κανονική Κατανομή

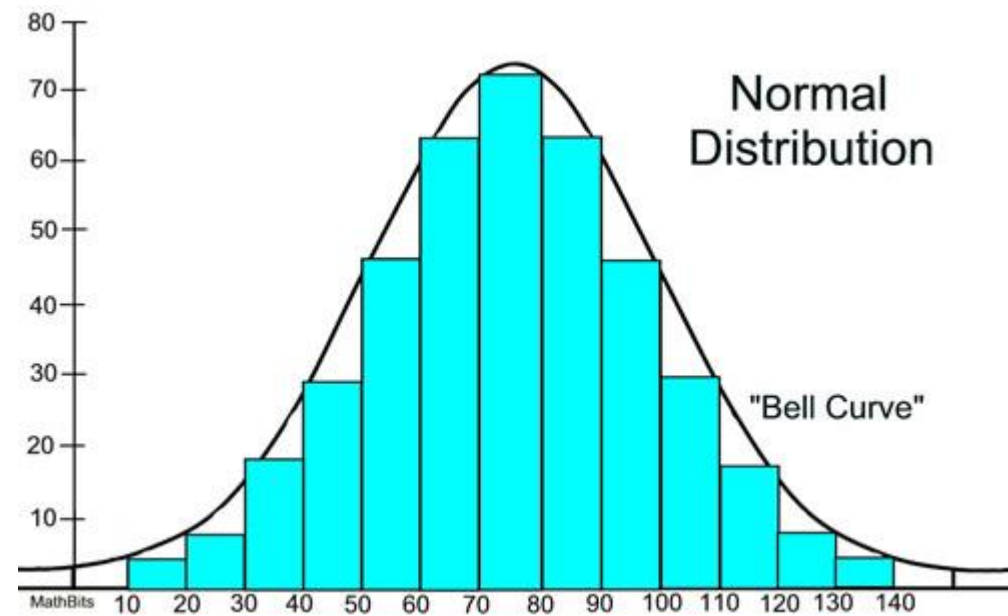
Κανονική κατανομή

Είναι συμμετρική (τόσα δείγματα κάτω από τον μ.ο. τόσα και επάνω)

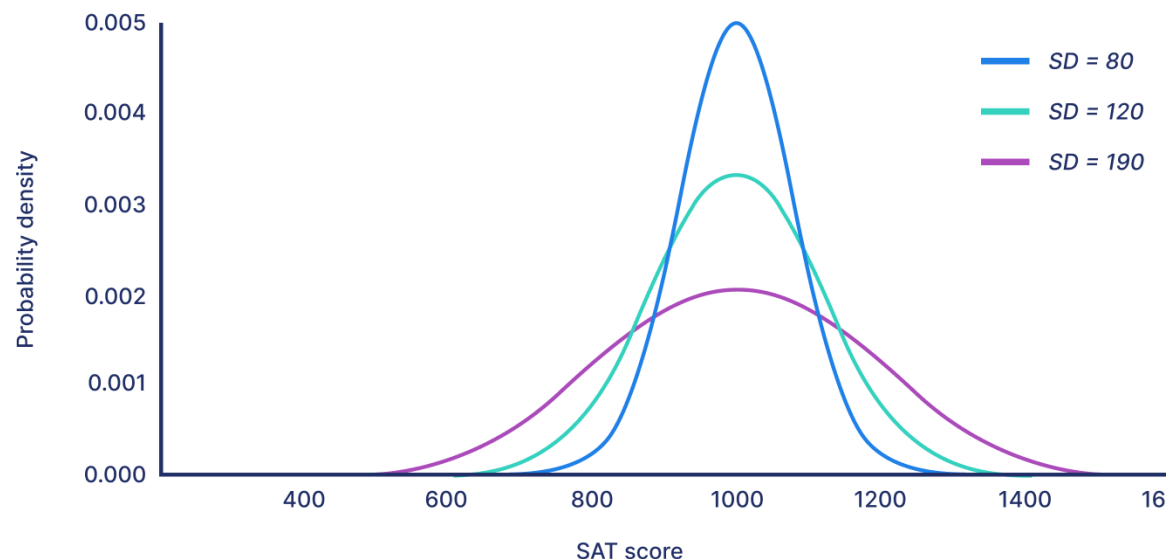
Σχήμα καμπάνας (το μέγιστο ύψος (mode) βρίσκεται στον μ.ο.)

Mean, Mode, και Median βρίσκονται στο κέντρο

Είναι ασυμπτωτική

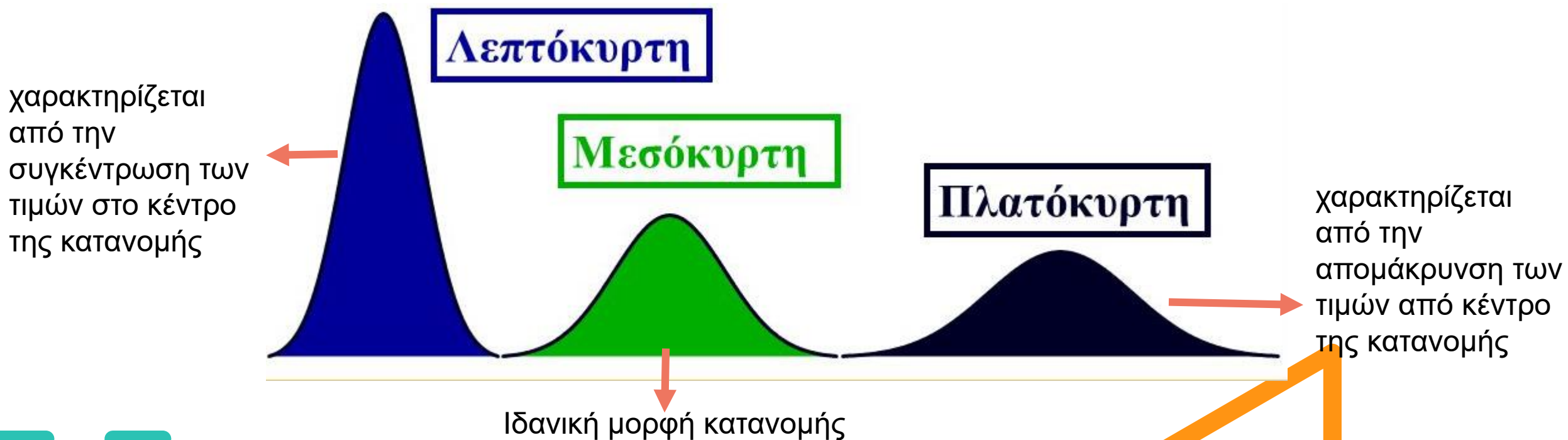


Normal distributions with different standard deviations



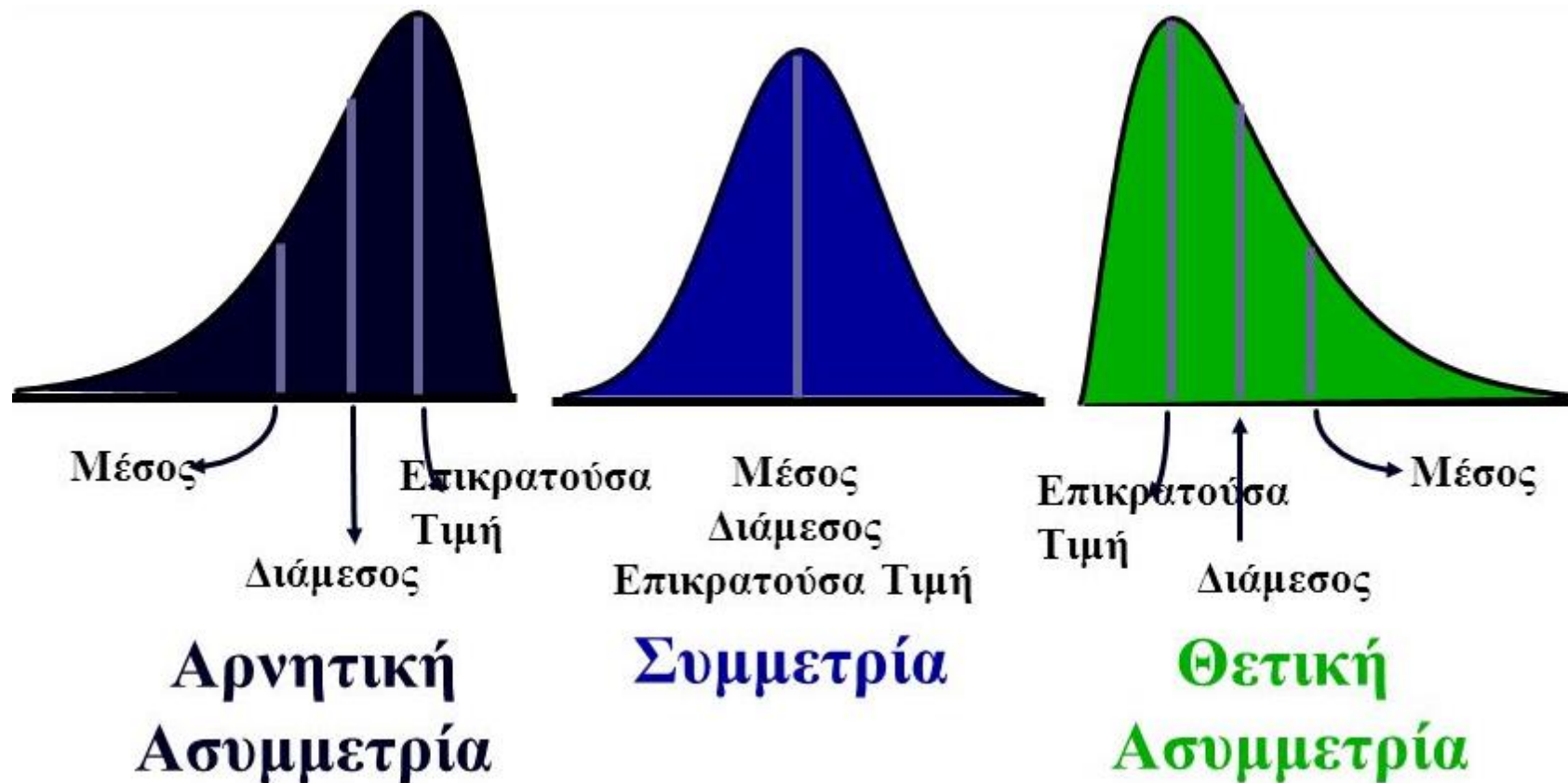
Κύρτωση (kurtosis)

Είναι ο βαθμός συγκέντρωσης των τιμών γύρω από το κέντρο της κατανομής



Ασυμετρία (skewness)

Ασύμμετρη κατανομή: το σημείο συγκέντρωσης των περισσότερων τιμών βρίσκεται δεξιά ή αριστερά στον άξονα τιμών



Διασπορά

Δύο κατανομές εντελώς διαφορετικές, οι οποίες έχουν ίδια:

- μέση τιμή,
- διάμεσο και
- επικρατούσα τιμή



Διασπορά: γενικός στατιστικός όρος για κάθε μέτρηση ο οποίος αναφέρεται στη διακύμανση των τιμών μιας κατανομής, δηλαδή πόσο μακριά από τη μέση τιμή είναι οι παρατηρήσεις μας

Δείκτες διασποράς

ΕΥΡΟΣ (range)

- Η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής της κατανομής
 - Μειονέκτημα: Η ύπαρξη ακραίων τιμών

ΕΝΔΟΤΕΤΑΡΤΗΜΟΤΙΑΚΟ ΕΥΡΟΣ (interquartile range)

- Είναι το εύρος του κεντρικού 50% των τιμών μιας κατανομής (οι πιο αντιπροσωπευτικές τμές). Το υπόλοιπο 50% των τιμών δεν λαμβάνεται υπ όψιν κατά τον υπολογισμό του.

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ (standard deviation, sd, σ)

- Είναι η μέση απόσταση των τιμών από τη μέση τιμή. Η ζώνη $\pm 1\sigma$ γύρω από τη μέση τιμή περιέχει συνήθως γύρω στο 68% των τιμών.

Τυπική Απόκλιση (Standard Deviation) — Παράδειγμα

- Δύο τάξεις έκαναν τεστ κατανόησης κειμένου (βαθμολογία 0-100):
- **Τάξη Α:** 68, 70, 71, 69, 72, 70, 71, 69 → $MO = 70$, $SD = 1.2$
- **Τάξη Β:** 40, 55, 70, 85, 95, 60, 75, 80 → $MO = 70$, $SD = 17.4$
- Ο MO είναι ίδιος (70) — αλλά οι τάξεις είναι πολύ διαφορετικές!
 - Τάξη Α: Αρκετά ομοιόγενης, μικρή διασπορά βαθμολογίας
 - Τάξη Β: Μεγάλη ανομοιογένεια
- Η τυπική απόκλιση μας λέει πόσο “σκορπισμένες” είναι οι τιμές γύρω από τον MO .

Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

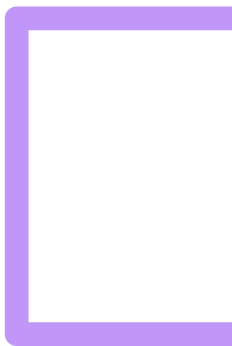
1. Τι ερευνά το άρθρο;

- Ξεκινάμε με την ερευνητική ερώτηση ή υπόθεση.
- Τι θέλουν να εξετάσουν οι συγγραφείς; Ποιες μεταβλητές τους ενδιαφέρουν;



Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

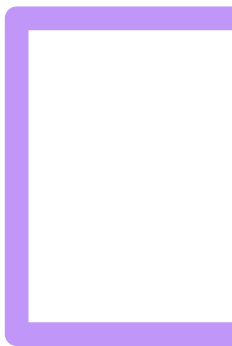
2. Τι είδους δεδομένα χρησιμοποιούνται;

- Ποιο είναι το δείγμα (μέγεθος, χαρακτηριστικά);
 - Τι μεταβλητές υπάρχουν (ποσοτικές, κατηγορικές);
 - Υπάρχει ομάδα σύγκρισης ή πείραμα;
- 
- 



Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

3. Ποια στατιστική μέθοδος εφαρμόζεται;

- Εντοπίζουμε τη στατιστική τεχνική (π.χ. t-test, ANOVA, συσχέτιση).
 - Αν δεν την ξέρουμε, επικοινωνούμε με το διδακτικό προσωπικό.
- 
- 

Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

4. Ποιο είναι το αποτέλεσμα της ανάλυσης;

- Εντοπίζουμε τα στατιστικά ευρήματα: t , F , r , p -value κ.λπ.
- Σημαίνει αυτό που βρήκαν στατιστικά σημαντική διαφορά ή σχέση; (Συνήθως όταν $p < 0,05$)

Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

5. Είναι το αποτέλεσμα "μεγάλο" ή "σημαντικό";

- Κοίταζουμε το μέγεθος επίδρασης (effect size) αν δίνεται (π.χ. Cohen's d , η^2).
- Είναι σημαντικό και πρακτικά, όχι μόνο στατιστικά;

Οδηγός ανάγνωσης άρθρου με στατιστική ανάλυση

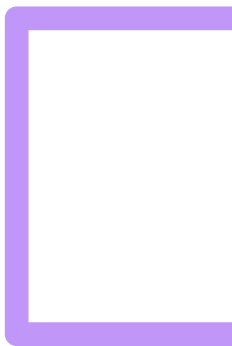
6. Πώς το ερμηνεύουν οι συγγραφείς;

- Μελετούμε προσεκτικά την ερμηνεία: τι συμπεράσματα βγάζουν;
- Συμφωνεί η ερμηνεία με τα ευρήματα;
- Υπάρχουν περιορισμοί που αναγνωρίζουν οι ίδιοι οι συγγραφείς;



Οδηγός ανάγνωσης άρθρου με στατιστική
ανάλυση

Μην προσπαθείτε να κατανοήσετε
κάθε αριθμό· επικεντρωθείτε στο
πλαίσιο, τη λογική της ανάλυσης και
στα συμπεράσματα.



Στατιστικά Τεστ

Πώς να διαβάζετε στατιστικά αποτελέσματα σε γλωσσολογικά άρθρα

p-value · t-test · ANOVA · Chi-square · Συσχέτιση · Mann-Whitney · Παλινδρόμηση

Χάρτης στατιστικών τεστ

p-value

Βάση όλων

t-test

2 ομάδες

ANOVA

3+ ομάδες

Mann-Whitney

2 ομάδες, μη-
παραμετρικό

Συσχέτιση

2 μεταβλητές

Chi-square

Κατηγορίες

Παλινδρόμηση

Πρόβλεψη

Κοινό σε όλα: καταλήγουν σε p-value

Τι είναι η τιμή p ;

ΒΑΣΗ

Κάθε τεστ απαντά: «Πόσο βέβαιοι μπορούμε να είμαστε ότι η διαφορά που βρήκαμε είναι αληθινή και όχι τυχαία;»

$p < .01$
Πολύ σημαντικό

$p < .05$ — Στατιστικά σημαντικό

$p > .05$ — Μη σημαντικό

$p < .05$
Στατιστικά σημαντικό — μικρή
πιθανότητα τυχαίου ευρήματος.

$p > .05$
Δεν βρέθηκε σημαντική διαφορά
(δεν σημαίνει ότι δεν υπάρχει).

n.s. = not significant
Συντομογραφία που εμφανίζεται
συχνά στα άρθρα.

Στα άρθρα μπορεί να εμφανίζονται: $t(58) = 3.42, p = .001$ · $F(2,147) = 12.3, p < .001$ · $r = -.18, p < .001$ · $p = .48$ (n.s.)

effect size (η^2)

Το p -value μας λέει αν υπάρχει διαφορά

Το effect size (η^2) μας λέει πόσο μεγάλη είναι αυτή η διαφορά

Και τα δύο είναι σημαντικά για να καταλάβουμε τι σημαίνουν τα αποτελέσματά μας στην πράξη.



Σχέσεις μεταξύ Μεταβλητών

Επίδραση μίας μεταβλητής σε άλλη (Διαφορές): Όταν θέλουμε να δούμε αν μια ανεξάρτητη μεταβλητή (κατηγορική) επηρεάζει μια εξαρτημένη (αριθμητική), χρησιμοποιούμε μεθόδους όπως το *t-test* ή η *ANOVA*.

Συσχέτιση (Correlation): Δείχνει αν και σε ποιο βαθμό δύο αριθμητικές μεταβλητές σχετίζονται μεταξύ τους. Η συσχέτιση δεν σημαίνει αιτιακή σχέση· απλώς δείχνει ότι οι μεταβολές της μιας σχετίζονται με τις μεταβολές της άλλης.

Σχέση μεταξύ κατηγορικών μεταβλητών: Για να εξετάσουμε αν υπάρχει σύνδεση μεταξύ δύο κατηγορικών μεταβλητών (π.χ. φύλο και τύπος απάντησης), χρησιμοποιείται το τεστ *chi-square*.

Πρόβλεψη (Predictive relationship): Όταν μας ενδιαφέρει να προβλέψουμε την τιμή μιας μεταβλητής με βάση άλλες, χρησιμοποιούμε μεθόδους όπως η *παλινδρόμηση* (regression). Σε αυτή την περίπτωση, η μία μεταβλητή θεωρείται ανεξάρτητη (predictor) και η άλλη εξαρτημένη (outcome).

Επίδραση πολλών μεταβλητών σε άλλη (Διαφορές): Όταν θέλουμε να δούμε αν πάνω από μια ανεξάρτητη μεταβλητή (κατηγορική ή αριθμητική) επηρεάζουν μια εξαρτημένη (αριθμητική), χρησιμοποιούμε μεθόδους όπως το *MANOVA* και *ANCOVA*.

t-test — Σύγκριση δύο ομάδων

ΠΑΡΑΜΕΤΡΙΚΟ

Ερώτημα: Διαφέρει σημαντικά ο μέσος όρος δύο ομάδων;

$$t(58) = 3.42, \quad p = .001$$

t(58)
Τιμή t — όσο
μεγαλύτερη, τόσο πιο
διαφορετικές οι ομάδες.
58 = βαθμοί ελευθερίας.

p = .001
0.1% πιθανότητα τυχαίου
αποτελέσματος.

Independent
Δύο ξεχωριστές ομάδες

Paired
Ίδια ομάδα, δύο χρόνοι

Χρειάζεται: κανονική κατανομή δεδομένων (ιστόγραμμα
καμπάνα).
Αν όχι → Mann-Whitney

M = μέση τιμή, SD = τυπική απόκλιση

Σε μια κοινότητα, κάποια παιδιά εκπαιδεύονται ταυτόχρονα σε δύο γλώσσες, τη μητρική και μια ξένη γλώσσα, ενώ τα υπόλοιπα μόνο στη μητρική. Οι εκπαιδευτικοί θέλουν να μελετήσουν αν υπάρχουν διαφορές στην ανάπτυξη λεξιλογίου.

Μεταβλητές:

- Ανεξάρτητη: Τύπος εκπαίδευσης (μονόγλωσση, δίγλωσση) – **Κατηγορική**.
- Εξαρτημένη: Ανάπτυξη λεξιλογίου στη μητρική γλώσσα – **Αριθμητική**.

Ερευνητική Υπόθεση (H1):

Η δίγλωσση εκπαίδευση επηρεάζει την ανάπτυξη του λεξιλογίου στη μητρική γλώσσα.

Mann-Whitney U — Μη παραμετρικό εναλλακτικό

ΜΗ-ΠΑΡΑΜ.

Ερώτημα: Διαφέρουν δύο ομάδες, όταν τα δεδομένα ΔΕΝ ακολουθούν κανονική κατανομή;

$$U = 1842, \quad z = -3.21, \quad p = .001$$

Ranks αντί μέσων
Κατατάσσει τις τιμές από τη μικρότερη στη μεγαλύτερη — δεν επηρεάζεται από ακραίες τιμές.

Πότε:
Συχνότητες λέξεων, μήκη με ακραίες τιμές, μικρά δείγματα, πολλά μηδενικά στα δεδομένα.

Παράδειγμα από άρθρο (Κακαβούλια κ.α./Glossologia 22 (2014), 41-60)

Παραγωγή λέξεων ανά λεπτό (άτομα με αφασία και μη)

Πίνακας 1. Λεκτική παραγωγή – Λεκτική ροή

	Άτομα με αφασία		Ομάδα ελέγχου	
	Λέξεις	Λέξεις/λεπτό	Λέξεις	Λέξεις/λεπτό
Προσωπική αφήγηση	158,5 (87,4)	59,3 (21,7)	173,7 (86,4)	103,7 (28)
Πάρτι	78,8 (40,3)	51,3 (14,2)	93,9 (31,1)	89 (34,1)
Δαχτυλίδι	156,4 (57,3)	55,1 (21,2)	233,6 (105,5)	124,6 (44,7)
Λαγός & χελώνα	141,6 (37,9)	49,5 (17,9)	180,7 (62,6)	110,5 (35,5)

Όπως φαίνεται στον Πίνακα 1, ως προς τη λεκτική ροή, διαπιστώθηκε ότι τα άτομα με αφασία παρήγαγαν λιγότερες λέξεις ανά λεπτό σε σχέση με τα άτομα της ομάδας ελέγχου. Η διαφορά ήταν συστηματική σε όλες τις ιστορίες και στατιστικά σημαντική (προσωπική αφήγηση: Mann-Whitney $U=5,000$, $p=.008$, πάρτι: $U=6,000$, $p=.011$, δαχτυλίδι: $U=2,000$, $p=.003$, λαγός & χελώνα: $U=.000$, $p=.001$). Όσον αφορά τον αριθμό των παραγόμενων λέξεων, δεν βρέθηκε στατιστικά σημαντική διαφορά μεταξύ των δύο ομάδων.

Md = διάμεσος (median) — αναφέρεται αντί μέσου όρου γιατί είναι πιο αντιπροσωπευτικός για μη κανονικά δεδομένα.

Στο corpus linguistics το Mann-Whitney είναι εξαιρετικά συχνό γιατί οι συχνότητες λέξεων σχεδόν ποτέ δεν κατανέμονται κανονικά.

ANOVA (MANOVA, ANCOVA, Two Way)— Σύγκριση 3+ ομάδων

ΠΑΡΑΜΕΤΡΙΚΟ

Ερώτημα: Διαφέρουν σημαντικά οι μέσοι όροι τριών ή περισσότερων ομάδων;

$$F(2, 147) = 12.34, \quad p < .001$$

$F(2, 147)$
2 = ομάδες-1
147 = N-ομάδες
(βαθμοί ελευθερίας)

Post-hoc (π.χ. Tukey)
Το ANOVA λέει μόνο
«κάπου υπάρχει
διαφορά». Το post-hoc
δείχνει ποιες ζεύγη
διαφέρουν.

Συμπέρασμα: τα γλωσσολογικά άρθρα γίνονται συστηματικά
δυσκολότερα να διαβαστούν.

Elsafoury & Λέγγερης - Γλωσσολογία/Glossologia 31
(2024), 1-13

*Σύγκριση διάρκειας φωνηέντων από φυσικούς ομιλητές
αραβικής. Έλεγχος επίδρασης φύλου και είδους φωνήεντος στη
διάρκεια.*

Τα αποτελέσματα έδειξαν στατιστικά σημαντική
επίδραση του είδους του φωνήεντος, $F(4, 152) = [23.23]$, $p < .001$, partial $\eta^2 = .379$ στη
διάρκεια του φωνήεντος (η επίδραση του φύλου
και η αλληλεπίδραση φωνήεντος και φύλου δεν
ήταν στατιστικά σημαντικές). Post-hoc
συγκρίσεις (Bonferroni) για την περαιτέρω
διερεύνηση της επίδρασης του φωνήεντος στη
διάρκεια κατέδειξαν την ακόλουθη σειρά (από τη
μικρότερη προς μεγαλύτερη διάρκεια): /i, u/ < /e,
o/ < /a/, $p < .05$.

Χρήση: σύγκριση ειδών κειμένου, επιπέδων γλώσσας, χρονικών περιόδων, συγγραφέων...

Επίδραση μίας μεταβλητής σε άλλη

Μέθοδος	Πότε χρησιμοποιείται	Δείκτες αναφοράς	Μέγεθος επίδρασης
t-test ανεξάρτητων δειγμάτων	Σύγκριση μέσων όρων ανάμεσα σε δύο ομάδες	t, p-value	Cohen's d
t-test συσχετισμένων δειγμάτων	Σύγκριση ίδιων ατόμων σε δύο συνθήκες (π.χ. πριν-μετά)	t, p-value	Cohen's d (paired)
ANOVA	Σύγκριση μέσων όρων μεταξύ 3+ ομάδων	F, p-value	η^2 (eta squared) ή partial η^2
Welch's ANOVA	Όταν παραβιάζεται η υπόθεση της ομοιογένειας διακυμάνσεων στην ANOVA	F, p-value	ω^2 (omega squared)
Mann–Whitney U	Μη παραμετρική εναλλακτική του t-test για δύο ομάδες	U, p-value	r
Kruskal–Wallis H	Μη παραμετρική εναλλακτική της ANOVA για 3+ ομάδες	H, p-value	— (ερμηνεία μέσω κατατάξεων)
ANCOVA (Analysis of Covariance)	Ελέγχει τη διαφορά μέσων όρων μεταξύ ομάδων λαμβάνοντας υπόψη μία ή περισσότερες συγκατανεμητές (covariates)	F, p-value, έλεγχος επίδρασης covariate(s)	partial η^2 , adjusted means
MANOVA (Multivariate ANOVA)	Χρησιμοποιείται όταν υπάρχουν πολλαπλές εξαρτημένες μεταβλητές ταυτόχρονα	Wilks' Lambda, Pillai's Trace, F, p-value	partial η^2 ή multivariate η^2



Βιβλιογραφία

Field A. (2000). *Discovering Statistics using SPSS*. SAGE.

Ρούσσος Π., Τσαούσης Γ. (2010). *Στατιστική εφαρμοσμένη στις κοινωνικές επιστήμες*. Αθηνά: ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ.

<https://www.jamovi.org/user-manual.html>