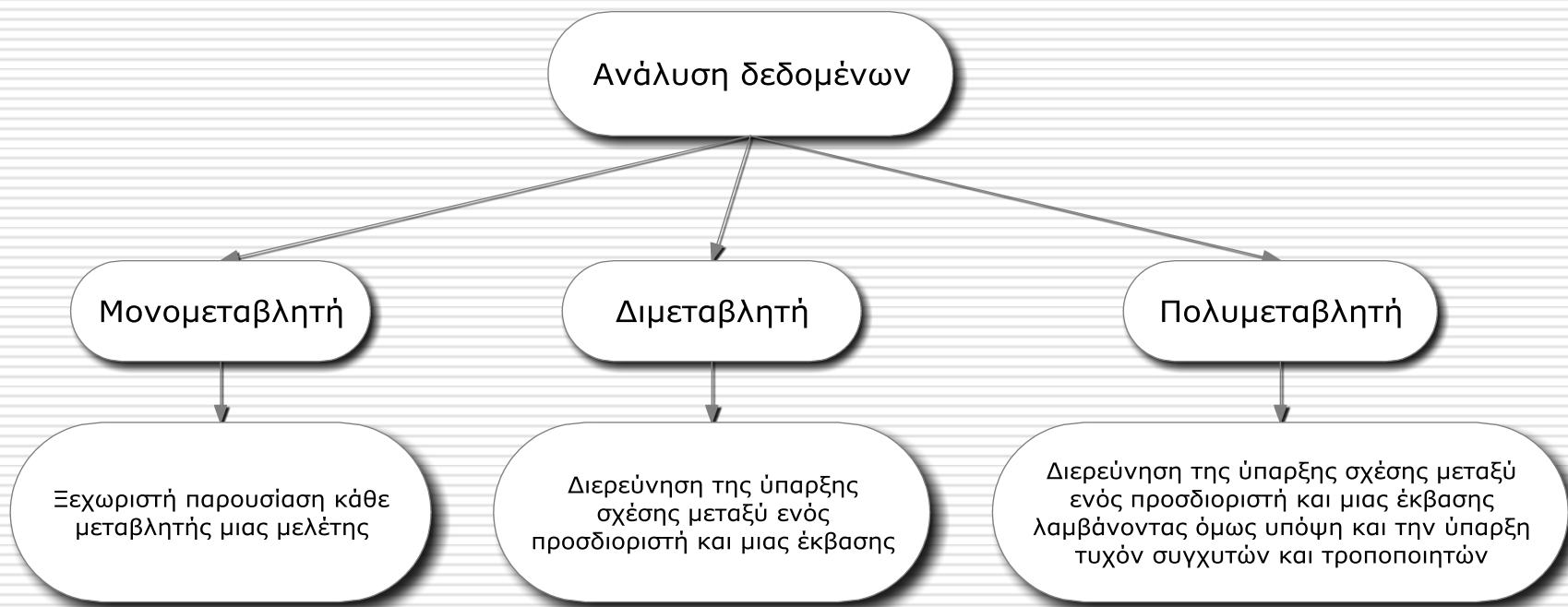

Μέτρα θέσης και διασποράς Μονομεταβλητή ανάλυση



Μονομεταβλητή ανάλυση

- Στην περίπτωση των ονομαστικών μεταβλητών αφορά την παράθεση των απόλυτων και των σχετικών συχνοτήτων
 - Στην περίπτωση των ποσοτικών μεταβλητών αφορά την παράθεση των κατάλληλων μέτρων θέσης και διασποράς
-

Μονομεταβλητή ανάλυση

- Ιδιαίτερη περίσκεψη απαιτείται στην περίπτωση των διατάξιμων μεταβλητών όπου είναι δυνατή η παράθεση τόσο των απόλυτων και των σχετικών συχνοτήτων όσο και των μέτρων θέσης και διασποράς
 - Σημειώνεται ότι στην περίπτωση των διατάξιμων μεταβλητών, η παράθεση των μέτρων θέσης και διασποράς έχει νόημα μόνο εφόσον οι μεταβλητές αυτές αντιμετωπιστούν από τους ερευνητές ως ποσοτικές μεταβλητές. Από στατιστικής απόψεως, η προσέγγιση αυτή δεν είναι η πλέον ενδεδειγμένη, αλλά εφόσον επιλεγεί πρέπει να αναφέρονται με σαφήνεια τα κριτήρια και οι προϋποθέσεις εφαρμογής της
-

Συχνότητες

- ❑ Η **απόλυτη συχνότητα** (frequency) εκφράζεται ως απόλυτος αριθμός και αφορά τον αριθμό των παρατηρήσεων για μια συγκεκριμένη κατηγορία ή τάξη της μελετώμενης μεταβλητής
 - ❑ Η **σχετική συχνότητα** (percent) εκφράζεται ως ποσοστό και προκύπτει από τη διαίρεση των παρατηρήσεων σε μια συγκεκριμένη κατηγορία με τον συνολικό αριθμό των παρατηρήσεων που αφορούν τη μελετώμενη μεταβλητή
 - ❑ Η **αθροιστική σχετική συχνότητα** (cumulative percent) εκφράζεται ως ποσοστό και για μια συγκεκριμένη κατηγορία της μελετώμενης μεταβλητής προκύπτει από την πρόσθεση των σχετικών συχνοτήτων για τη συγκεκριμένη κατηγορία και όλες τις προηγούμενες κατηγορίες
-

Πίνακας. Δημογραφικά χαρακτηριστικά υγιών (n=100) και ασθενών (n=100)

Χαρακτηριστικό	Υγιείς	Ασθενείς
Φύλο		
Άντρες	51 (51,0)	55 (55,0)
Γυναίκες	49 (49,0)	45 (45,0)
Οικογενειακή κατάσταση		
Έγγαμοι-ες	55 (55,0)	58 (58,0)
Άγαμοι-ες	38 (38,0)	32 (32,0)
Χήροι-ες	7 (7,0)	10 (10,0)
Συμβίωση		
Ναι	83 (83,0)	72 (72,0)
Όχι	17 (17,0)	28 (28,0)

Οι τιμές εκφράζονται ως n (%)

Πίνακας. Δημογραφικά χαρακτηριστικά υγιών (n=100) και ασθενών (n=100)

Χαρακτηριστικό	Υγιείς	Ασθενείς
Τόπος γέννησης		
Αθήνα	55 (55,0)	51 (51,0)
Θεσσαλονίκη	2 (2,0)	7 (7,0)
Επαρχία	43 (43,0)	42 (42,0)
Τόπος διαμονής		
Αθήνα	85 (85,0)	67 (67,0)
Θεσσαλονίκη	3 (3,0)	2 (2,0)
Επαρχία	12 (12,0)	31 (31,0)
Εκπαιδευτικό επίπεδο		
Πρωτοβάθμια εκπαίδευση (νηπιαγωγείο-δημοτικό)	53 (53,0)	53 (53,0)
Δευτεροβάθμια εκπαίδευση (γυμνάσιο-λύκειο)	31 (31,0)	37 (37,0)
Τριτοβάθμια εκπαίδευση (ανώτερες και ανώτατες σχολές)	16 (16,0)	10 (10,0)

Οι τιμές εκφράζονται ως n (%)

Μέτρα θέσης

- Τα μέτρα θέσης (measures of location) ή, αλλιώς, μέτρα κεντρικής τάσης (measures of central tendency) αφορούν τις τιμές αυτές γύρω από τις οποίες οι παρατηρήσεις τείνουν να συγκεντρώνονται σε μεγαλύτερο βαθμό
 - Τα σημαντικότερα μέτρα θέσης είναι ο μέσος, η διάμεσος και η επικρατούσα τιμή
-

Μέτρα διασποράς

- Τα μέτρα διασποράς (measures of dispersion) ή, αλλιώς, μέτρα μεταβλητότητας (measures of variability) δηλώνουν τον βαθμό στον οποίο διασπείρονται οι παρατηρήσεις
 - Τα σημαντικότερα μέτρα διασποράς είναι το εύρος, το ενδοτεταρτημοριακό εύρος, η διασπορά, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας
-

Μέτρα θέσης και διασποράς

- Σημειώνεται πως όταν τα μέτρα θέσης και διασποράς αφορούν πληθυσμούς συμβολίζονται με ελληνικούς χαρακτήρες (π.χ., ο μέσος ενός πληθυσμού συμβολίζεται με «μ»), ενώ όταν αφορούν «δείγματα» συμβολίζονται με λατινικούς χαρακτήρες (π.χ., ο μέσος ενός «δείγματος» συμβολίζεται με «x»).
-

Ποσοτικές μεταβλητές

- Στην περίπτωση των ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή θα πρέπει να παρουσιάζονται ο μέσος και η τυπική απόκλιση, ενώ στην περίπτωση που δεν ακολουθούν την κανονική κατανομή θα πρέπει να παρουσιάζονται η διάμεσος και το ενδοτεταρτημοριακό εύρος (ή το εύρος)
-

Μέσος

- ❑ Ο αριθμητικός μέσος (arithmetic mean) ή μέσος όρος (average) ή απλά μέσος (mean) αποτελεί το συνηθέστερο μέτρο θέσης
 - ❑ Ο μέσος αποτελεί ουσιαστικά τη μέση τιμή των παρατηρήσεων αναφορικά με μια μεταβλητή και προκύπτει από τη διαίρεση του αθροίσματος των τιμών όλων των παρατηρήσεων ενός «δείγματος» με τον συνολικό αριθμό των παρατηρήσεων
-

Μέσος

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- n : ο συνολικός αριθμός των παρατηρήσεων του δείγματος
 - Σ : το άθροισμα των τιμών των παρατηρήσεων
 - x_i : η τιμή κάθε παρατήρησης
-

Μέσος

- Εάν οι βαθμολογίες 5 μαθητών σ' ένα μάθημα είναι 18, 19, 19, 19 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε η μέση βαθμολογία στο συγκεκριμένο «δείγμα», σύμφωνα με την ισότητα 10-1, είναι ίση με:

$$\bar{x} = \frac{1}{5}(18 + 19 + 19 + 19 + 20) = 19$$

Μέσος

- Σημειώνεται ότι ο μέσος, σε αντίθεση με τη διάμεσο, επηρεάζεται σημαντικά από τις τιμές των απομακρυσμένων παρατηρήσεων. Εάν στο παραπάνω παράδειγμα, η βαθμολογία του πέμπτου μαθητή δεν ήταν 20, αλλά ήταν 5, τότε η μέση βαθμολογία θα ήταν ίση με:

$$\bar{x} = \frac{1}{5}(18 + 19 + 19 + 19 + 5) = 16$$

Μέσος

- ❑ Ο μέσος είναι το κατάλληλο μέτρο θέσης στην περίπτωση των ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή
 - ❑ Εάν οι ποσοτικές μεταβλητές δεν ακολουθούν την κανονική κατανομή, τότε το κατάλληλο μέτρο θέσης είναι η διάμεσος.
 - ❑ Στην περίπτωση των διατάξιμων μεταβλητών που αντιμετωπίζονται ως ποσοτικές μεταβλητές, το κατάλληλο μέτρο θέσης είναι η διάμεσος και όχι ο μέσος, ανεξάρτητα από το αν οι μεταβλητές ακολουθούν ή όχι την κανονική κατανομή
-

Διάμεσος

- Η διάμεσος (median) αποτελεί το 50^ο εκατοστιαίο σημείο μιας ομάδας παρατηρήσεων
 - Τοποθετώντας μια ομάδα παρατηρήσεων στη σειρά από τη μικρότερη στη μεγαλύτερη τιμή, η διάμεσος αποτελεί τη μεσαία τιμή
 - Έτσι, οι μισές τιμές των παρατηρήσεων θα είναι μεγαλύτερες ή ίσες με τη διάμεσο και οι άλλες μισές θα είναι μικρότερες ή ίσες με τη διάμεσο
-

Διάμεσος

- Όταν υπάρχουν n παρατηρήσεις και το n είναι περιττός ή, αλλιώς, μονός αριθμός, τότε η διάμεσος είναι η μεσαία παρατήρηση, δηλαδή η παρατήρηση:

$$\frac{n+1}{2}$$

Διάμεσος

- Για παράδειγμα, εάν οι βαθμολογίες 5 μαθητών σ' ένα μάθημα είναι 18, 19, 19, 19 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε η διάμεσος είναι η τρίτη παρατήρηση, δηλαδή είναι η τιμή 19:

$$\left(\frac{n+1}{2} = \frac{5+1}{2} = 3 \right)$$

Διάμεσος

- Στην περίπτωση εξάλλου που υπάρχουν n παρατηρήσεις, αλλά το n είναι ζυγός ή, αλλιώς άρτιος αριθμός, τότε η διάμεσος είναι ο αριθμητικός μέσος των δυο μεσαίων παρατηρήσεων, δηλαδή των παρατηρήσεων:

$$\frac{n}{2}$$

$$\frac{n+2}{2}$$

Διάμεσος

- Για παράδειγμα, εάν οι βαθμολογίες 6 μαθητών σ' ένα μάθημα είναι 16, 17, 18, 19, 20 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε η διάμεσος είναι ο αριθμητικός μέσος της τρίτης και της τέταρτης παρατήρησης, δηλαδή είναι η τιμή $18,5 = (18 + 19) / 2$:

$$\left(\frac{n}{2} = \frac{6}{2} = 3 \text{ και } \frac{n+2}{2} = \frac{6+2}{2} = 4 \right)$$

Διάμεσος

- Σημειώνεται ότι η διάμεσος επηρεάζεται πολύ λιγότερο από τις τιμές των απομακρυσμένων παρατηρήσεων σε σχέση με τον μέσο
 - Για παράδειγμα, εάν οι βαθμολογίες 5 μαθητών σ' ένα μάθημα είναι 18, 19, 19, 19 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε τόσο ο μέσος όσο και η διάμεσος ισούνται με 19
 - Εάν όμως η βαθμολογία του πέμπτου μαθητή δεν ήταν 20, αλλά ήταν 5, τότε ο μέσος θα ήταν ίσος με 16, θα μειώνονταν δηλαδή κατά 3 μονάδες, ενώ η διάμεσος θα ήταν και πάλι ίση με 19
-

Διάμεσος

- Η διάμεσος είναι το κατάλληλο μέτρο θέσης στην περίπτωση των ποσοτικών μεταβλητών που δεν ακολουθούν την κανονική κατανομή, καθώς και στην περίπτωση των διατάξιμων μεταβλητών
-

Επικρατούσα τιμή

- Η επικρατούσα τιμή (mode) είναι η τιμή που συμβαίνει πιο συχνά σε μια ομάδα παρατηρήσεων και χρησιμοποιείται συνήθως στην περίπτωση των διατάξιμων και των ονομαστικών μεταβλητών και σπανιότερα στην περίπτωση των ποσοτικών μεταβλητών
 - Για παράδειγμα, εάν οι βαθμολογίες 5 μαθητών σ' ένα μάθημα είναι 18, 19, 19, 19 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε η επικρατούσα τιμή είναι η τιμή 19, καθώς εμφανίζεται τις περισσότερες φορές
-

Εύρος

- ❑ Το εύρος (range) αποτελεί το απλούστερο μέτρο διασποράς και είναι ίσο με τη διαφορά μεταξύ της ελάχιστης και της μέγιστης τιμής μιας ομάδας παρατηρήσεων
 - ❑ Το εύρος υπολογίζεται εξαιρετικά απλά, αλλά δεν χρησιμοποιείται ιδιαίτερα, καθώς λαμβάνει υπόψη του μόνο την ελάχιστη και τη μέγιστη τιμή μιας ομάδας παρατηρήσεων και επιπλέον επηρεάζεται σε μεγάλο βαθμό από τις τιμές των απομακρυσμένων παρατηρήσεων
 - ❑ Για παράδειγμα, εάν οι τιμές 5 παρατηρήσεων είναι 80, 82, 85, 87 και 90, τότε το εύρος είναι ίσο με 10 ($=90-80$)
-

Ενδοτεταρτημοριακό εύρος

- ❑ Το ενδοτεταρτημοριακό εύρος (interquartile range) επηρεάζεται πολύ λιγότερο από τις τιμές των απομακρυσμένων παρατηρήσεων σε σχέση με το εύρος και προκύπτει από την αφαίρεση του 25^{ου} εκατοστιαίου σημείου των παρατηρήσεων από το 75^ο
 - ❑ Περιλαμβάνει το κεντρικό 50% των παρατηρήσεων
 - ❑ Είναι το κατάλληλο μέτρο διασποράς στην περίπτωση των ποσοτικών μεταβλητών που δεν ακολουθούν την κανονική κατανομή, καθώς και στην περίπτωση των διατάξιμων μεταβλητών. Στην περίπτωση αυτή, το ενδοτεταρτημοριακό εύρος πρέπει να παρουσιάζεται σε συνδυασμό με τη διάμεσο
-

Διασπορά

- Η διασπορά (variance) και η τυπική απόκλιση (standard deviation) αποτελούν τα συνηθέστερα μέτρα διασποράς και ποσοτικοποιούν τη μεταβλητότητα των παρατηρήσεων γύρω από τον μέσο του δείγματος
 - Στην περίπτωση των πληθυσμών, η διασπορά και η τυπική απόκλιση συμβολίζονται με « σ^2 » και « σ » αντιστοίχως, ενώ στην περίπτωση των δειγμάτων συμβολίζονται με s^2 και s αντιστοίχως
-

Διασπορά

- Η διασπορά σ' ένα δείγμα προκύπτει από την αφαίρεση του μέσου μιας ομάδας παρατηρήσεων από κάθε μια από τις παρατηρήσεις, υψώνοντας στο τετράγωνο τις διαφορές αυτές, πραγματοποιώντας έπειτα την πρόσθεσή τους και διαιρώντας το αποτέλεσμα με τον συνολικό αριθμό των παρατηρήσεων μειωμένο κατά μια:

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

- n : ο συνολικός αριθμός των παρατηρήσεων του δείγματος
- Σ : το άθροισμα των τετραγώνων των διαφορών
- x_i : η τιμή κάθε παρατήρησης
- $\bar{x}$: μέσος του δείγματος

Διασπορά

- Για παράδειγμα, εάν οι βαθμολογίες 5 μαθητών σ' ένα μάθημα είναι 18, 19, 19, 19 και 20 και η κλίμακα βαθμολόγησης είναι 0-20, τότε η μέση βαθμολογία στο συγκεκριμένο «δείγμα», σύμφωνα με την ισότητα 10-1, είναι ίση με 19, ενώ η διασπορά, σύμφωνα με την ισότητα 10-2, είναι ίση με:

$$s^2 = \frac{1}{(5-1)} \left[(18-19)^2 + (19-19)^2 + (19-19)^2 + (19-19)^2 + (20-19)^2 \right]$$

$$s^2 = \frac{1}{4} \left[(-1)^2 + 0^2 + 0^2 + 0^2 + 1^2 \right]$$

$$s^2 = 0,5$$

Τυπική απόκλιση

- Η τυπική απόκλιση σ' ένα «δείγμα» ισούται με την τετραγωνική ρίζα της διασποράς:

$$s = \sqrt{s^2}$$

Τυπική απόκλιση

- ❑ Στην πράξη, η τυπική απόκλιση χρησιμοποιείται συχνότερα σε σχέση με τη διασπορά και αυτό οφείλεται στο γεγονός ότι η τυπική απόκλιση έχει τις ίδιες μονάδες μέτρησης με τον μέσο
 - ❑ Συγκρίνοντας δυο ομάδες παρατηρήσεων, η ομάδα με τη μικρότερη τυπική απόκλιση έχει πιο ομοιογενείς παρατηρήσεις και παρουσιάζει μικρότερη μεταβλητότητα
 - ❑ Η τυπική απόκλιση είναι το κατάλληλο μέτρο διασποράς στην περίπτωση των ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή. Στην περίπτωση αυτή, η τυπική απόκλιση πρέπει να παρουσιάζεται σε συνδυασμό με τον μέσο
-

Συντελεστής μεταβλητότητας

- Ο συντελεστής μεταβλητότητας (coefficient of variation) είναι ένα μέτρο σχετικής μεταβλητότητας και εκφράζεται ως ποσοστό. Προκύπτει από τη διαίρεση της τυπικής απόκλισης με τον μέσο και τον πολλαπλασιασμό του προϊόντος της διαίρεσης με το 100:

$$\text{Συντελεστής μεταβλητότητας} = \frac{s}{\bar{x}} \times 100\%$$

Συντελεστής μεταβλητότητας

- Η τυπική απόκλιση και ο μέσος έχουν τις ίδιες μονάδες μέτρησης, οπότε ο συντελεστής μεταβλητότητας είναι καθαρός αριθμός, δεν έχει δηλαδή μονάδες μέτρησης, και για τον λόγο αυτόν μπορεί να χρησιμοποιηθεί για τη σύγκριση μεταξύ μεταβλητών με διαφορετικές μονάδες μέτρησης
-