

Γεωργικός Πειραματισμός

4ο Εργαστήριο

Αναστάσιος Κατσιλέρος

Γεωπονικό Πανεπιστήμιο Αθηνών
Εργαστήριο Βελτίωσης Φυτών και Γεωργικού Πειραματισμού

katsileros@aua.gr

Αθήνα 2020

Το φαινόμενο εμφάνισης διαφορών στις παρατηρήσεις ή μετρήσεις μεταξύ μονάδων που ανήκουν στην ίδια ομάδα ή κατηγορία, ονομάζεται **παραλλακτικότητα** (variation).

Το χαρακτηριστικό ή ιδιότητα των μονάδων από τις οποίες λαμβάνονται αυτές οι παρατηρήσεις ή μετρήσεις αναφέρεται ως **μεταβλητή** (variable). Οι μεταβλητές διακρίνονται σε δύο κατηγορίες, τις **ποσοτικές** (quantitative) και τις **ποιοτικές** (qualitative).

Οι ποσοτικές μεταβλητές μπορούν να μετρηθούν και διαχωρίζονται σε **συνεχείς** (continuous), οι οποίες μπορούν να πάρουν αριθμητικές τιμές που καλύπτουν ένα διάστημα τιμών (π.χ. βάρος, ύψος, απόδοση κ.α.) και σε **διακριτές** (discrete), οι οποίες μπορούν να πάρουν τιμές από ένα αριθμήσιμο σύνολο (π.χ αριθμός παιδιών ανά οικογένεια, αριθμός στάσεων ανά φυτό κ.α.).

Οι ποιοτικές μεταβλητές δεν μπορούν να μετρηθούν και διακρίνονται σε **κατηγορικές - ονομαστικές** (nominal), οι οποίες επιδέχονται μόνο αυθαίρετη κατάταξη (π.χ φύλο ανθρώπων, χρώμα ματιών, απουσία-παρουσία κ.α.) και οι **διατακτικές** (ordinal), οι οποίες κατατάσσονται σε ιεραρχική σειρά (π.χ. Καλώς, Λίαν Καλώς, Άριστα).

Το σύνολο όλων των δυνατών ή υποθετικών παρατηρήσεων μίας μεταβλητής ονομάζεται **στατιστικός πληθυσμός** (statistical population). Ο πληθυσμός μπορεί να είναι υπαρκτός ή μη και το μέγεθος του μπορεί να είναι πεπερασμένο ή άπειρο. Η μελέτη όλων των μονάδων ενός στατιστικού πληθυσμού είναι επιθυμητό, αλλά στις περισσότερες περιπτώσεις ακόμη και στους πεπερασμένους πληθυσμούς είναι αδύνατο και για αυτό το λόγο ο ερευνητής επιλέγει ένα τμήμα του πληθυσμού για να τον μελετήσει. Για να είναι το τμήμα αντιπροσωπευτικό του πληθυσμού αλλά και για να διασφαλίζεται η αμεροληψία, θα πρέπει η επιλογή των μονάδων του τμήματος να γίνεται τυχαία από τον ερευνητή.

Ο τρόπος συλλογής των παρατηρήσεων από έναν στατιστικό πληθυσμό μπορεί να γίνει μέσω μιας **μεθόδου δειγματοληψίας** (sampling), οπότε το τμήμα του πληθυσμού αναφέρεται ως **δείγμα** (sample) ή μετά από μία **επέμβαση** ή **μεταχείριση** (treatment) σύμφωνα με κάποιο **πειραματικό σχέδιο** (experimental design) ή από άμεση παρατήρηση ενός φαινομένου. Οι μετρήσεις ή οι παρατηρήσεις οι οποίες συλλέγονται από τις μονάδες με κάποια μέθοδο, αναφέρονται και ως **πρωτογενή δεδομένα** (raw data).

Παρόλο που ο ερευνητής χρησιμοποιεί ένα τμήμα του πληθυσμού για τη μελέτη του, τα πρωτογενή δεδομένα που προκύπτουν συνήθως είναι μεγάλα σε μέγεθος. Προτού προχωρήσει στην ανάλυση των δεδομένων θα πρέπει να οργανώσει, να ταξινομήσει και να παρουσιάσει τα δεδομένα απλά και κατανοητά. Για το σκοπό αυτό χρησιμοποιούνται πίνακες συχνότητας, γραφικές μέθοδοι και αριθμητικά περιγραφικά μέτρα.

Πίνακες κατανομής συχνότητων

Ένας αποτελεσματικός τρόπος για να περιγραφεί η κατανομή των δεδομένων είναι να δοθεί η συχνότητα τους, δηλαδή πόσες φορές εμφανίζεται η κάθε τιμή στα δεδομένα που έχουν συλλεχτεί. Η κατανομή αυτή των δεδομένων μπορεί να αποδοθεί υπό μορφή πίνακα συχνότητων ή και διαγραμμάτων.

Όταν οι τιμές προέρχονται από μία διακριτή μεταβλητή, στην πρώτη στήλη του πίνακα τοποθετούνται οι τιμές σε αύξουσα σειρά και στις επόμενες στήλες, η **απόλυτη συχνότητα** ή **συχνότητα f_i** (frequency) των τιμών, η **σχετική συχνότητα** εμφάνισης των τιμών **p_i** (relative frequency), η οποία ορίζεται ως ο λόγος της συχνότητας εμφάνισης f_i μιας τιμής προς το σύνολο των παρατηρήσεων n , η **αθροιστική συχνότητα F_i** (cumulative frequency) κάθε τιμής, που ορίζεται ως το άθροισμα των συχνοτήτων όλων των τιμών που είναι μικρότερες ή ίσες από μία συγκεκριμένη τιμή και η **αθροιστική σχετική συχνότητα P_i** (cumulative relative frequency), που ορίζεται ως το άθροισμα των σχετικών συχνοτήτων όλων των τιμών που είναι μικρότερες ή ίσες από μία συγκεκριμένη τιμή.

Παράδειγμα 1. Τα δεδομένα (7, 4, 5, 4, 6, 5, 7, 6, 5, 6, 8, 6, 4, 6, 5) αφορούν τη διάρκεια σπουδων δεκαπέντε φοιτητών, σε μία τετραετή σχολή.

Τιμές (Έτη)	Απόλυτη Συχνότητα f_i	Σχετική Συχνότητα p_i	Αθροιστική Συχνότητα F_i	Σχετική Αθροιστική Συχνότητα P_i
4	3	0,20	3	0,20
5	4	0,27	7	0,47
6	5	0,33	12	0,80
7	2	0,13	14	0,93
8	1	0,07	15	1
		1		

Εισαγωγή δεδομένων στο φύλλο εργασίας του στατιστικού πακέτου Jamovi.

The screenshot displays the Jamovi software interface. The window title is "jamovi - Untitled". The top menu bar includes "Data" and "Analyses". Below the menu bar is a toolbar with icons for various statistical functions: Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. A "Modules" button with a plus sign is also present.

The main workspace is divided into two panes. The left pane shows a spreadsheet with columns labeled A, B, and C. The data is as follows:

	A	B	C
1	7		
2	4		
3	5		
4	4		
5	6		
6	5		
7	7		
8	6		
9	5		
10	6		
11	8		
12	6		
13	4		
14	6		
15	5		
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			

The right pane is currently blank, displaying the Jamovi logo and the text "version 1.2.17".

At the bottom of the window, the status bar shows: "Ready", "Filters 0", "Row count 15", "Filtered 0", "Deleted 0", "Added 15", and "Cells edited 45".

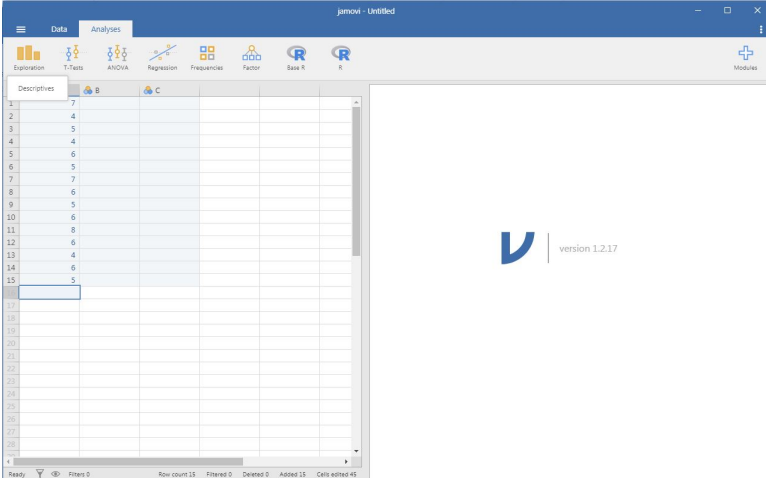
Μετανομασία της στήλης των δεδομένων σε Y και χαρακτηρισμός τους ως Nominal.

The screenshot shows the JAMOVI software interface. The main window is titled "jamovi - Untitled". The "Analyses" tab is active, showing various statistical analysis options like Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The "DATA VARIABLE" configuration panel is open for a variable named "Y". The "Measure type" is set to "Nominal", and the "Data type" is set to "Integer (auto)". The "Levels" list contains the values 4, 5, 6, and 7. The "Missing values" field is empty. The "Retain unused levels" option is turned off. Below the configuration panel, a data table is visible with columns Y, B, and C. The data for column Y is as follows:

	Y	B	C
1	7		
2	4		
3	5		
4	4		
5	6		
6	5		
7	7		
8	6		
9	5		
10	6		
11	8		
12	6		
13	4		
14	6		

The status bar at the bottom indicates "Ready", "Filters 0", "Row count 15", "Filtered 0", "Deleted 0", "Added 15", and "Cells edited 45". The version number "version 1.2.17" is displayed in the bottom right corner of the main window.

Επιλογή Descriptives από το μενού Analyses-Exploration.



The screenshot displays the JAMOVI software interface. The 'Analyses' menu is open, and 'Descriptives' is selected. The main window shows a data grid with 15 rows and 2 columns (B and C). The status bar at the bottom indicates 'Ready', 'Filters 0', 'Row count 15', 'Filtered 0', 'Deleted 0', 'Added 15', and 'Cells edited 45'. The version number 'version 1.2.17' is visible in the bottom right corner of the main window.

	B	C
1	7	
2	4	
3	5	
4	4	
5	6	
6	5	
7	7	
8	6	
9	5	
10	6	
11	8	
12	6	
13	4	
14	6	
15	5	

Εισαγωγή στήλης δεδομένων Y στο πλαίσιο Variables.

The screenshot shows the jamovi software interface. The top menu bar includes 'Data' and 'Analyses'. The toolbar contains icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The main window is titled 'jamovi - Untitled'. The 'Descriptives' analysis window is open, showing a list of variables on the left with 'B' and 'C' selected. The 'Variables' list on the right contains 'Y'. Below the variable lists are options for 'Split by' and 'Frequency tables'. The right panel displays the 'Descriptives' table and a 'References' section.

Descriptives	Y
N	15
Missing	0
Mean	5.60
Median	6
Minimum	4
Maximum	8

References

- [1] The jamovi project (2020). *jamovi*. (Version 1.2) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2019). *R: A Language and environment for statistical computing*. (Version 3.6) [Computer software]. Retrieved from <https://cran.r-project.org/>.

Επιλογή δημιουργίας πίνακα συχνотήτων (Frequency tables).

The screenshot shows the jamovi software interface. The top menu bar includes 'Data' and 'Analyses'. The 'Analyses' menu is open, showing options like 'Exploration', 'T-Tests', 'ANOVA', 'Regression', 'Frequencies', 'Factor', 'Base R', and 'R'. The 'Descriptives' panel is active, showing a list of variables 'B' and 'C' on the left, and 'Y' in the 'Variables' box on the right. The 'Split by' box is empty. The 'Frequency tables' checkbox is checked. Below the checkbox are expandable sections for 'Statistics' and 'Plots'. The right pane displays the results for 'Descriptives' for variable 'Y'.

Descriptives

Descriptives	Y
N	15
Missing	0
Mean	5.60
Median	6
Minimum	4
Maximum	8

Frequencies

Frequencies of Y

Levels	Counts	% of Total	Cumulative %
4	3	20.0 %	20.0 %
5	4	26.7 %	46.7 %
6	5	33.3 %	80.0 %
7	2	13.3 %	93.3 %
8	1	6.7 %	100.0 %

References

Όταν οι τιμές προέρχονται από μία συνεχή ή διακριτή μεταβλητή με μεγάλο πλήθος τιμών, τότε δεν είναι δυνατή η εμφάνιση όλων των τιμών με τις αντίστοιχες συχνότητες στον πίνακα, οπότε είναι απαραίτητο η ομαδοποίηση των τιμών σε μια σειρά μη επικαλυπτόμενων διαστημάτων. Τα διαστήματα αυτά ονομάζονται **κλάσεις** ή **τάξεις** (classes). Ο αριθμός και το εύρος των κλάσεων επιλέγονται από τον ερευνητή, έτσι ώστε η κατανομή να περιγράφει ικανοποιητικά τα δεδομένα.

Για την επιλογή του αριθμού των κλάσεων μπορούν να χρησιμοποιηθούν και διάφοροι εμπειρικοί τύποι, όπως ο τύπος του Sturges στον οποίο ο αριθμός των κλάσεων υπολογίζεται ως εξής:

$$k = 1 + 3,322 \log_{10} n$$

όπου k είναι ο αριθμός των κλάσεων και n το πλήθος των παρατηρήσεων.

Το πλάτος των κλάσεων είναι συνήθως σταθερό και προκύπτει διαιρώντας το εύρος των παρατηρήσεων με τον επιθυμητό αριθμό των κλάσεων. Στη συνέχεια το πλάτος προσαρμόζεται και στρογγυλοποιείται έτσι ώστε τα άκρα του πρώτου και τελευταίου διαστήματος να περιέχουν τη μικρότερη και τη μεγαλύτερη τιμή αντίστοιχα.

Τα κέντρα των κλάσεων m_i δίνονται από τον τύπο:

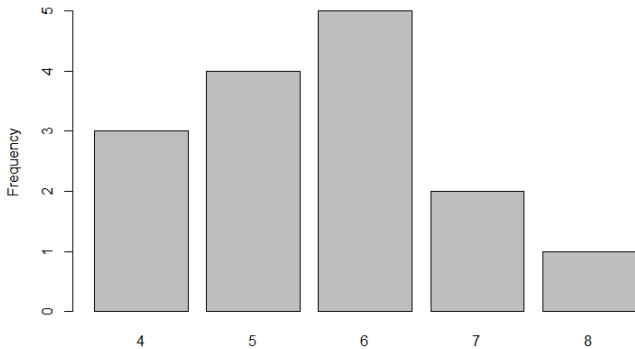
$$m_i = \frac{a_{(i-1)} + a_i}{2}$$

όπου a_i τα άκρα του διαστήματος

Παράδειγμα 2. Τα δεδομένα (2.8, 3.4, 2.6, 3.2, 2.8, 2.5, 2.8, 3.0, 2.9, 2.8, 2.7, 2.7, 2.8, 3.0, 2.4, 2.9, 3.0, 3.1, 2.5, 2.3) αφορούν τον βάρος σε γραμμάρια των σπόρων του κύριου στάχυ, ανά φυτό σιταριού.

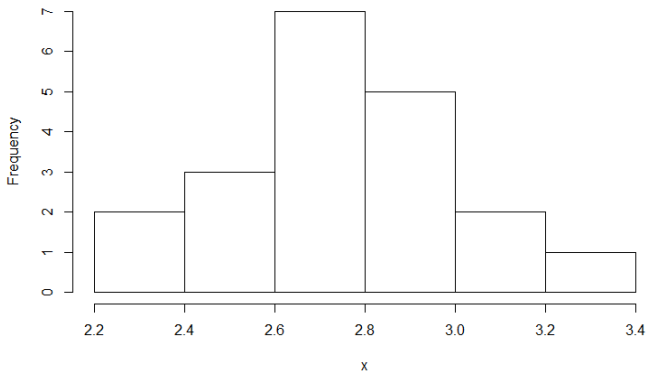
Κλάσεις	Κέντρο Κλάσης m_i	Απόλυτη Συχνότητα f_i	Αθροιστική Συχνότητα F_i	Σχετική Συχνότητα p_i	Σχετική Αθροιστική Συχνότητα P_i
(2,2-2,4]	2,3	2	2	0,10	0,10
(2,4-2,6]	2,5	3	5	0,15	0,25
(2,6-2,8]	2,7	7	15	0,35	0,60
(2,8-3,0]	2,9	5	17	0,25	0,85
(3,0-3,2]	3,1	2	19	0,10	0,95
(3,2-3,4]	3,3	1	20	0,05	1
		20		1	

Τα δεδομένα του πίνακα συχνοτήτων μπορούν να αποδοθούν γραφικά σε ένα **ραβδόγραμμα** (bar chart), το οποίο χρησιμοποιείται συνήθως για κατηγοριοποιημένα ή διακριτά δεδομένα. Αποτελείται από ορθογώνιες στήλες των οποίων η απόσταση μεταξύ τους και το πλάτος των βάσεων τους καθορίζονται αυθαίρετα, ενώ το ύψος κάθε ορθογώνιας στήλης είναι ίσο με τη συχνότητα f_i ή τη σχετική συχνότητα p_i της αντίστοιχης τιμής.



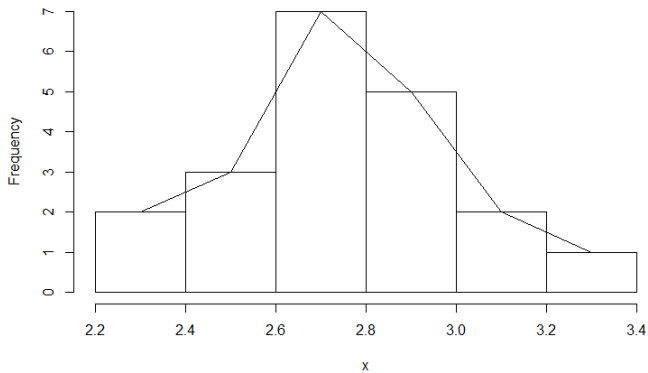
Το **ιστόγραμμα** (histogram) χρησιμοποιείται για τη γραφική αναπαράσταση δεδομένων από διακριτή με μεγάλο πλήθος τιμών ή συνεχή μεταβλητή, στις οποίες οι τιμές έχουν ομαδοποιηθεί σε κλάσεις. Αποτελείται από ορθογώνιες στήλες οι οποίες είναι ενωμένες μεταξύ τους, το πλάτος των βάσεων τους είναι ίσο με το εύρος της κλάσης και το ύψος κάθε ορθογώνιας στήλης είναι ίσο με την συχνότητα f_i ή τη σχετική συχνότητα p_i ή την αθροιστική συχνότητα F_i της αντίστοιχης κλάσης.

Histogram of x



Το **πολύγωνο συχνοτήτων** (frequency polygon) είναι ένας άλλος τρόπος παρουσίασης της κατανομής συχνοτήτων και κατασκευάζεται όταν ενώνονται τα κέντρα των κλάσεων. Για πολύ μεγάλο μέγεθος παρατηρήσεων, το πλάτος των κλάσεων μειώνεται στο ελάχιστο και το πολύγωνο συχνοτήτων προσεγγίζει μία ομαλή καμπύλη, η οποία αποδίδει τη θεωρητική κατανομή των συχνοτήτων και ονομάζεται **καμπύλη συχνοτήτων** (frequency curve).

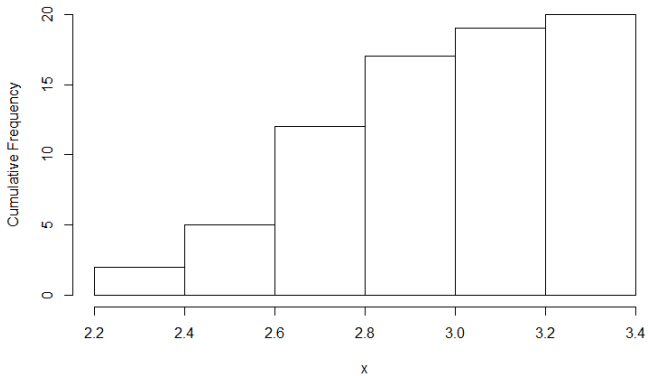
Histogram of x



Histogram of x



Histogram of x



Τα μέτρα **κεντρικής τάσης** ή **θέσης** (central tendency measures) προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα.

Τα κυριότερα μέτρα κεντρικής τάσης είναι η μέση τιμή, η διάμεσος, η επικρατούσα τιμή και τα ποσοστιαία σημεία.

Η **μέση τιμή πληθυσμού** (population mean) μεγέθους N , συμβολίζεται με το μ και ορίζεται ως:

$$\mu = \frac{\sum_{i=1}^N y_i}{N}$$

Η **μέση τιμή δείγματος** (sample mean) αποτελεί το σημαντικότερο μέτρο θέσης και συμβολίζεται με \bar{Y} . Αν y_1, y_2, \dots, y_n οι τιμές των παρατηρήσεων του δείγματος από έναν πληθυσμό, η μέση τιμή ορίζεται ως:

$$\bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$$

Η **διάμεσος** (median) ορίζεται ως η κεντρική τιμή η οποία χωρίζει ένα σύνολο διατεταγμένων σε αύξουσα σειρά παρατηρήσεων σε δύο ίσα μέρη και συμβολίζεται με \tilde{Y} . Αν ο αριθμός n των παρατηρήσεων είναι περιττός τότε η διάμεσος είναι η μεσαία παρατήρηση, ενώ αν είναι άρτιος είναι το ημιάθροισμα των δύο μεσαίων παρατηρήσεων.

Η **επικρατούσα τιμή** ή **κορυφή** (mode) ορίζεται ως η τιμή ή τιμές που εμφανίζονται με τη μεγαλύτερη συχνότητα στα δεδομένα.

Το **p-ποσοστιαίο σημείο** (quantiles) ενός συνόλου παρατηρήσεων είναι η τιμή εκείνη για την οποία το $p\%$ των παρατηρήσεων είναι μικρότερες από αυτή την τιμή ($0 \leq p < 1$). Τα πιο χρησιμοποιούμενα ποσοστιαία σημεία είναι τα **τεταρτομόρια** (quartiles) και τα **εκατοστημόρια** (percentiles).

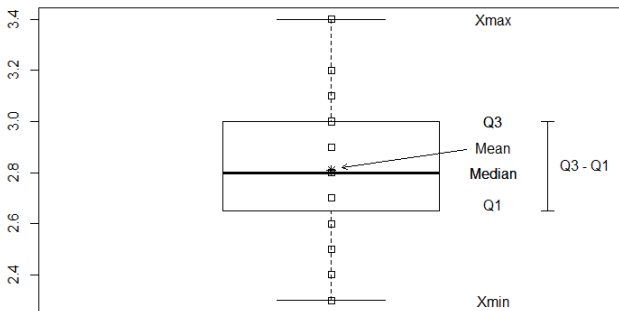
Το **πρώτο τεταρτημόριο** (first quartile) είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 25% των διατεταγμένων παρατηρήσεων του δείγματος και συμβολίζεται ως Q_1 .

Το **δεύτερο τεταρτημόριο** (second quartile) είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 50% των διατεταγμένων παρατηρήσεων του δείγματος και συμβολίζεται ως Q_2 το οποίο είναι και η διάμεσος.

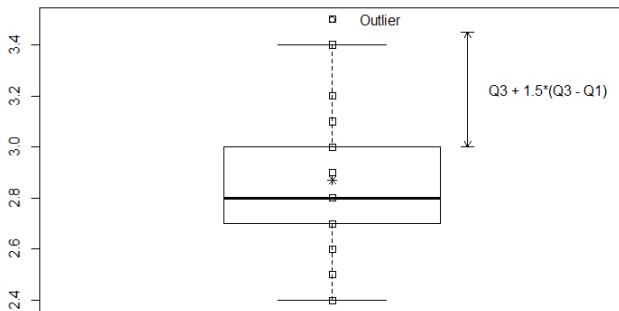
Το **τρίτο τεταρτημόριο** (third quartile) είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 75% των διατεταγμένων παρατηρήσεων του δείγματος και συμβολίζεται ως Q_3 .

Για τον υπολογισμό των τεταρτημορίων χρησιμοποιούνται διάφοροι αλγόριθμοι, στην πιο απλή περίπτωση τα τεταρτημόρια υπολογίζονται όπως η διάμεσος.

Η διάμεσος, το πρώτο και τρίτο τεταρτημόριο καθώς και η ελάχιστη και μέγιστη τιμή των δεδομένων, αποδίδονται γραφικά με το **θηκόγραμμα** (box plot). Για τη κατασκευή του θηκογράμματος δημιουργείται ένα ορθογώνιο πλαίσιο με βάσεις το πρώτο και το τρίτο τεταρτημόριο και εντός του πλαισίου τοποθετείται η διάμεσος. Από την μέση των βάσεων αναπτύσσονται γραμμές ή **κεραίες** (whiskers) που εκτείνονται μέχρι τις οριακές άνω και κάτω τιμές. Η άνω οριακή τιμή ορίζεται η μεγαλύτερη παρατήρηση η οποία είναι μικρότερη ή ίση από την τιμή $Q3 + 1.5 \cdot (Q3 - Q1)$, ενώ η κάτω οριακή τιμή, η μικρότερη παρατήρηση η οποία είναι μεγαλύτερη ή ίση από την τιμή $Q1 - 1.5 \cdot (Q3 - Q1)$.



Οι παρατηρήσεις που βρίσκονται έξω από τις οριακές τιμές ονομάζονται **έκτροπες** ή **ακραίες τιμές** (outlier), ενώ αν απέχουν από τις βάσεις περισσότερο από $\pm 3 \cdot (Q3 - Q1)$, ονομάζονται **εξαιρετικά έκτροπες τιμές** (extreme outlier).



Τα **μέτρα μεταβλητότητας** (variability measures) είναι αριθμητικά μεγέθη που δίνουν την **διασπορά – διασκόρπιση** (dispersion) των παρατηρήσεων γύρω από τις κεντρικές τιμές της κατανομής.

Τα κυριότερα μέτρα είναι το εύρος, το ενδοτεταρτομοριακό εύρος, η διακύμανση, η τυπική απόκλιση και ο συντελεστής παραλλακτικότητας.

Η **διακύμανση του πληθυσμού** σ^2 (population variance) μετράει τη διασπορά N παρατηρήσεων γύρω από τη μέση τιμή μ του πληθυσμού και είναι ο λόγος του αθροίσματος τετραγώνων των διαφορών ή αποκλίσεων $(Y_i - \mu)$ προς το σύνολο των παρατηρήσεων N :

$$\sigma^2 = \frac{\sum_{i=1}^N (Y_i - \mu)^2}{N}$$

Η **διακύμανση του δείγματος** s^2 (sample variance) μετράει τη διασπορά n παρατηρήσεων γύρω από τη μέση τιμή του δείγματος ενός πληθυσμού. Η δειγματική διακύμανση ορίζεται ως ο λόγος του αθροίσματος τετραγώνων των αποκλίσεων ($Y_i - \bar{Y}$) προς τον αριθμό των παρατηρήσεων του δείγματος μείον ένα. Ο όρος του παρανομαστή ονομάζεται βαθμοί ελευθερίας.

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

Βαθμοί ελευθερίας (degrees of freedom): Χρησιμοποιώντας την μέση τιμή \bar{Y} από τις παρατηρήσεις, αντί την μέση τιμή του μ πληθυσμού για τον υπολογισμό του αθροίσματος τετραγώνων των αποκλίσεων, υποεκτιμάται η διακύμανση του πληθυσμού σ^2 , γιατί οι παρατηρήσεις Y_i τείνουν να είναι πιο κοντά στη μέση τιμή των παρατηρήσεων \bar{Y} παρά στην πραγματική μέση τιμή του πληθυσμού μ . Επομένως για την καλύτερη εκτίμηση της παραμέτρου της διακύμανσης του πληθυσμού σ^2 , το άθροισμα τετραγώνων διαιρείται με $n - 1$ αντί με το n .

Η **τυπική απόκλιση** (standard deviation) αποτελεί ένα μέτρο διασποράς των τιμών σε σχέση με την μέση τιμή, που εκφράζεται στην ίδια μονάδα μέτρησης.

Ορίζεται ως η τετραγωνική ρίζα της διακύμανσης του πληθυσμού:

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

ή του δείγματος:

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

Το **εύρος** (range) είναι η διαφορά της ελάχιστης από τη μέγιστη τιμή των δεδομένων και συμβολίζεται με R . Το εύρος αν και απλό στον υπολογισμό του δεν θεωρείται αξιόπιστο μέτρο διασποράς επειδή εξαρτάται από ακραίες τιμές.

Το **ενδοτεταρτομοριακό εύρος** (interquartile range) είναι η διαφορά του τρίτου με το πρώτο τεταρτημόριο ($Q_3 - Q_1$) η οποία περιλαμβάνει το 50% των τιμών της κατανομής και συμβολίζεται ως IQR.

Ο **συντελεστής παραλλακτικότητας** (coefficient of variation) ορίζεται ως ο λόγος της τυπικής απόκλισης s προς την μέση τιμή \bar{Y} των παρατηρήσεων και συμβολίζεται ως $CV\%$.

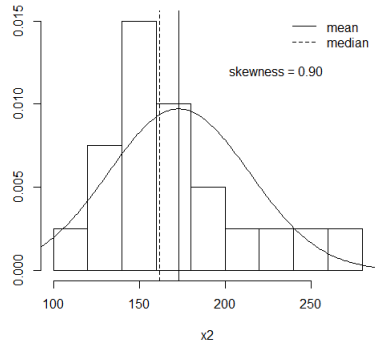
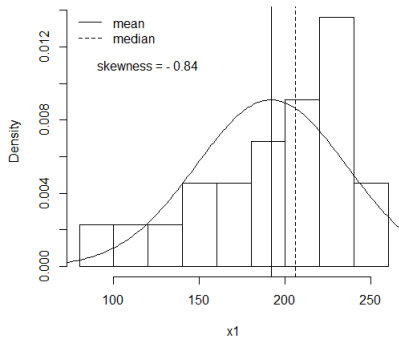
$$CV\% = \frac{s}{\bar{Y}}\%$$

Ο συντελεστής παραλλακτικότητας είναι ανεξάρτητος από τις μονάδες μέτρησης και επιτρέπει τη σύγκριση της μεταβλητότητας διαφορετικών δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής. Το μέγεθος του συντελεστή παραλλακτικότητας εξαρτάται από το είδος της έρευνας.

Το μέτρο ή συντελεστής ασυμμετρίας ή λοξότητας (coefficient of skewness) μετράει το βαθμό της ασυμμετρίας ή απόκλισης από τη συμμετρία της κατανομής των τιμών μιας μεταβλητής. Όταν η κατανομή είναι συμμετρική, η μέση τιμή, η διάμεσος και η επικρατούσα τιμή συμπίπτουν. Αν και υπάρχουν διάφοροι συντελεστές ασυμμετρίας, από πιο διαδεδομένους είναι συντελεστής με βάση τις ροπές (moment coefficient of skewness), ο οποίος ορίζεται από τον τύπο:

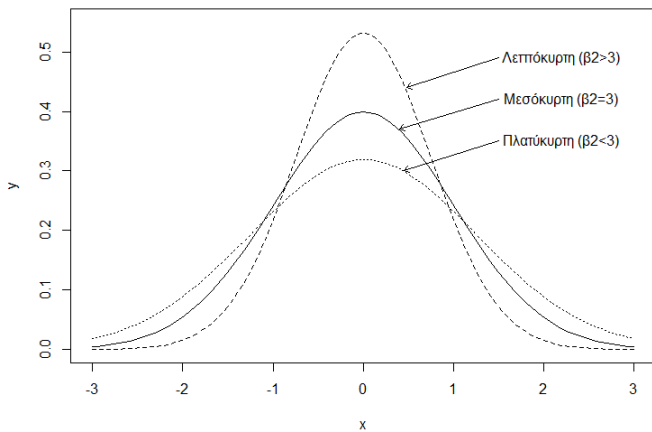
$$\beta_1 = \frac{m_3}{m_2^{3/2}}$$

$$\text{όπου } m_k = \frac{\sum (Y_i - \bar{Y})^k}{n}$$



Το **μέτρο** ή **συντελεστής κύρτωσης** (coefficient of kurtosis) μετράει το βαθμό συγκέντρωσης των τιμών γύρω από τη μέση τιμή. Μια κατανομή η οποία έχει μεγάλη συχνότητα ή συγκέντρωση τιμών γύρω από τη μέση τιμή ονομάζεται **λεπτόκυρτη** (leptokurtic), αν η συγκέντρωση τιμών είναι μικρή ονομάζεται **πλατύκυρτη** (platykurtic), ενώ οι κατανομές που προσεγγίζονται από την κανονική κατανομή λέγονται **μεσόκυρτες** (mesokurtic). Για τον υπολογισμό του συντελεστή χρησιμοποιείται ο συντελεστής του κύρτωσης με βάση τις ροπές, ο οποίος ορίζεται από τον τύπο:

$$\beta_2 = \frac{m_4}{m_2^2}$$



Παράδειγμα 2. Τα δεδομένα (2.8, 3.4, 2.6, 3.2, 2.8, 2.5, 2.8, 3.0, 2.9, 2.8, 2.7, 2.7, 2.8, 3.0, 2.4, 2.9, 3.0, 3.1, 2.5, 2.3) αφορούν τον βάρος σε γραμμάρια των σπόρων του κύριου στάχυ, 20 φυτών σιταριού.

Μέση τιμή δείγματος

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{2,8 + 3,4 + \dots + 2,5 + 2,3}{20} = 2,81$$

Διάμεσος

$$\tilde{Y} = \frac{Y_{10} - Y_{11}}{2} = \frac{2,8 + 2,8}{2} = 2,8$$

Επικρατούσα τιμή : 2,8

Διακύμανση του δείγματος

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1} = \frac{(2,8 - 2,81)^2 + \dots + (2,3 - 2,81)^2}{20 - 1} = 0,074$$

Τυπική απόκλιση

$$s = \sqrt{s^2} = \sqrt{0,074} = 0,27$$

Τυπικό σφάλμα του μέσου

$$s.e.m. = \frac{s}{\sqrt{n}} = \frac{0,27}{\sqrt{20}} = 0,06$$

Συντελεστής παραλλακτικότητας

$$CV\% = \left(\frac{s}{\bar{Y}}\right) * 100 = \left(\frac{0,27}{2,81}\right) * 100 = 9,6\%$$

Εισαγωγή δεδομένων στο φύλλο εργασίας.

The screenshot displays the jamovi software interface. The title bar reads "jamovi - Untitled". The main menu is divided into "Data" and "Analyses". The "Analyses" menu is open, showing options: Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The spreadsheet area shows columns A, B, and C. Column A contains numerical data from row 1 to 20. Column B and C are empty. The status bar at the bottom indicates "Ready", "Filters 0", "Row count 20", "Filtered 0", "Deleted 0", "Added 20", and "Cells edited 0".

	A	B	C
1	2.8		
2	3.4		
3	2.6		
4	3.2		
5	2.8		
6	2.5		
7	2.8		
8	3.0		
9	2.9		
10	2.8		
11	2.7		
12	2.7		
13	2.8		
14	3.0		
15	2.4		
16	2.9		
17	3.0		
18	3.1		
19	2.5		
20	2.3		

Χαρακτηρισμός της στήλης δεδομένων ως Continuous (Decimal).

The screenshot shows the Jamovi software interface. The top menu bar includes 'Data' and 'Analyses'. The main toolbar contains icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The central panel is titled 'DATA VARIABLE' and shows the configuration for variable 'A'. The 'Measure type' is set to 'Continuous' and the 'Data type' is set to 'Decimal (auto)'. The 'Levels' panel is empty. The bottom-left panel shows a data table with 14 rows and 3 columns (A, B, C). Column A contains the following values: 2.8, 3.4, 2.6, 3.2, 2.8, 2.5, 2.8, 3.0, 2.9, 2.8, 2.7, 2.7, 2.8, 3.0.

	A	B	C
1	2.8		
2	3.4		
3	2.6		
4	3.2		
5	2.8		
6	2.5		
7	2.8		
8	3.0		
9	2.9		
10	2.8		
11	2.7		
12	2.7		
13	2.8		
14	3.0		

Εισαγωγής της στήλης δεδομένων A, στο πλαίσιο Variables.

The screenshot shows the JAMOVI software interface. The top menu bar includes 'Data' and 'Analyses'. The toolbar contains icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The main window is titled 'Descriptives' and contains a search bar, a list of variables (B and C), and a 'Variables' list containing 'A'. Below the variable lists are options for 'Frequency tables', 'Statistics', and 'Plots'. The right-hand pane displays the 'Descriptives' results for variable 'A' and a 'References' section with two citations.

Descriptives	
	A
N	20
Missing	0
Mean	2.81
Median	2.80
Minimum	2.30
Maximum	3.40

References

- [1] The Jamovi project (2020). *Jamovi*. (Version 1.2) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2019). *R: A Language and environment for statistical computing*. (Version 3.6) [Computer software]. Retrieved from <https://cran.r-project.org/>.

Επιλογή εμφάνισης μέτρων κεντρικής τάσης και μεταβλητότητας από το Statistics.

The screenshot shows the jamovi software interface. The top menu bar includes 'Data' and 'Analyses'. Below the menu bar are icons for various statistical tests: Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The main window is titled 'jamovi - Untitled'. The 'Descriptives' configuration window is open, showing a list of variables (B and C) on the left and a 'Variables' list containing 'A'. The 'Statistics' section is expanded, showing options for Sample Size, Percentile Values, Dispersion, Central Tendency, Distribution, and Normality. The 'Descriptives' table on the right displays the following data:

Descriptives	A
N	20
Missing	0
Mean	2.81
Std. error mean	0.0607
Median	2.80
Mode	2.80
Sum	56.2
Standard deviation	0.271
Variance	0.0736
Range	1.10
Minimum	2.30
Maximum	3.40
Skewness	0.111
Std. error skewness	0.512
Kurtosis	0.0907
Std. error kurtosis	0.992
25th percentile	2.68
50th percentile	2.80
75th percentile	3.00

Επιλογή εμφάνισης γραφημάτων από το Plots.

The screenshot shows the jamovi software interface. The top menu bar includes 'Data' and 'Analyses'. The main toolbar contains icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, and R. The 'Descriptives' panel on the left shows variables B and C, with A selected in the 'Variables' list. The 'Plots' section is expanded, showing options for Histograms (Histogram and Density checked), Box Plots (Box plot checked, Violin and Data unchecked), and Bar Plots (Bar plot checked). A 'Jittered' dropdown is visible under the Box Plots section. The main workspace displays two plots for variable A: a histogram with a density curve overlaid, and a box plot below it. The histogram x-axis ranges from 2.25 to 3.50, and the box plot y-axis ranges from 2.25 to 3.25.