

Περιγραφική Στατιστική

με το Excel Analysis ToolPak

Εργαστηριακή Άσκηση Στατιστικής

Τμήμα Ναυτιλιακών Σπουδών | Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Στόχοι Μαθήματος:

Κατανόηση εννοιών Περιγραφικής Στατιστικής • Χρήση Data Analysis ToolPak στο Excel • Ερμηνεία αποτελεσμάτων με ναυτιλιακά παραδείγματα

Τι είναι η Περιγραφική Στατιστική;

Βασικές Έννοιες & Ορισμοί

Ορισμός

Η Περιγραφική Στατιστική είναι ο κλάδος της Στατιστικής που συγκεντρώνει, οργανώνει, συνοψίζει και παρουσιάζει τα δεδομένα με τρόπο κατανοητό, χωρίς να βγάζουμε συμπεράσματα για μεγαλύτερο πληθυσμό.



Μέτρα Κεντρικής Τάσης

Περιγράφουν το 'κέντρο' των δεδομένων
→ Μέσος, Διάμεσος, Επικρατούσα



Μέτρα Διακύμανσης

Περιγράφουν πόσο 'απλώνονται' τα δεδομένα
→ Εύρος, Διακύμανση, Τυπική Απόκλιση



Μέτρα Σχήματος

Περιγράφουν τη μορφή της κατανομής
→ Ασυμμετρία (Skewness), Κύρτωση



Ναυτιλιακό Παράδειγμα

Σε μια εταιρεία shipping, έχουμε δεδομένα για 50 πλοία: ταχύτητα, κατανάλωση καυσίμου, ημέρες ταξιδιού. Η Περιγραφική Στατιστική μάς βοηθά να συνοψίσουμε αυτά τα δεδομένα και να εντοπίσουμε μοτίβα — π.χ. ποια είναι η μέση ταχύτητα, πόσο μεταβάλλεται η κατανάλωση, κ.λπ.

Τα Δεδομένα μας – Στόλος Εμπορικών Πλοίων

Παράδειγμα: 15 Πλοία Εταιρείας ShipData S.A.

Δεδομένα 15 πλοίων – Χρησιμοποιήστε αυτά στο Excel σας:

Πλοίο	Ταχύτητα (κν)	Κατ. Καυσίμου (t/ημ)	Ημ. Ταξιδιού	Καθυστέρηση (ώρες)
Ηρακλής	14.2	22.5	12	3
Αθηνά	16.8	26.1	8	0
Ποσειδών	12.5	19.8	15	7
Κρόνος	17.1	27.3	6	1
Πηνελόπη	13.9	21.4	14	5
Θησέας	15.5	24.2	10	2
Ελένη	16.2	25.7	9	0
Αγαμέμνων	11.8	18.9	18	12
Πατρόκλος	14.7	23.1	11	4
Αχιλλέας	18.3	29.5	5	0
Οδυσσέας	13.2	20.8	16	8
Περσεφόνη	15.9	25.1	10	3
Μίνωας	17.4	27.8	7	1
Νηρέας	12.1	19.2	17	9
Χάρων	16.5	26.3	8	2

🔔 **Ανοίξτε το Excel** → εισάγετε τα παραπάνω δεδομένα σε 5 στήλες (Α-Ε) με επικεφαλίδες στη γραμμή 1

Μέτρα Κεντρικής Τάσης

Μέσος Όρος • Διάμεσος • Επικρατούσα Τιμή

Μέσος Όρος (Mean)

\bar{x}

$$\bar{x} = (x_1 + x_2 + \dots + x_n) / n$$

Αθροίζουμε όλες τις τιμές και διαιρούμε με το πλήθος τους. Επηρεάζεται από ακραίες τιμές (outliers).

 Ταχύτητα πλοίων: $(14.2+16.8+\dots+16.5)/15 = 15.07$ κν

Διάμεσος (Median)

Me

Η μεσαία τιμή σε ταξινομημένη σειρά

Ταξινομούμε τα δεδομένα από μικρό σε μεγάλο. Για $n=15$ (περιττό), η διάμεσος είναι η 8η τιμή. ΔΕΝ επηρεάζεται από ακραίες τιμές.

 Ταχύτητες ταξινομημένες: ...14.7, 15.5, 15.9...
→ Διάμεσος = 15.5 κν

Επικρατούσα (Mode)

Mo

Η πιο συχνά εμφανιζόμενη τιμή

Η τιμή που εμφανίζεται περισσότερες φορές. Μπορεί να υπάρχουν πολλαπλές επικρατούσες ή καμία (όταν όλες εμφανίζονται μία φορά).

 Καθυστέρηση 0 ώρες εμφανίζεται 3 φορές →
Mo = 0 ώρες

Μέτρα Διακύμανσης

Εύρος • Διακύμανση • Τυπική Απόκλιση

Τα μέτρα διακύμανσης μάς λένε πόσο 'σκορπισμένα' είναι τα δεδομένα γύρω από το κέντρο τους. Μικρή διακύμανση = δεδομένα κοντά στο μέσο. Μεγάλη διακύμανση = δεδομένα πολύ διασκορπισμένα.

Εύρος (Range)

$$R = X_{\max} - X_{\min}$$

Η διαφορά μεταξύ της μέγιστης και ελάχιστης τιμής. Απλό αλλά ευαίσθητο σε ακραίες τιμές.

⚓ Ταχύτητα: $18.3 - 11.8 = 6.5$ κν
Καθυστέρηση: $12 - 0 = 12$ ώρες

Διακύμανση (Variance)

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

Μέσος τετράγωνος αποκλισμός από το μέσο. Χρησιμοποιεί $(n-1)$ για δείγμα (μη μεροληψία).

⚓ Μεγάλη διακύμανση = πλοία με πολύ διαφορετικές ταχύτητες → δύσκολος σχεδιασμός δρομολογίου

Τυπική Απόκλιση (Std Dev)

$$s = \sqrt{\text{Διακύμανση}}$$

Η τετραγωνική ρίζα της διακύμανσης. Στις ίδιες μονάδες με τα δεδομένα — πιο εύκολη ερμηνεία.

⚓ Αν $s = 1.8$ κν για ταχύτητα: τα πλοία κυμαίνονται κατά μέσο ± 1.8 κν γύρω από τη μέση ταχύτητα

Ασυμμετρία (Skewness) & Κύρτωση (Kurtosis)

Το Σχήμα της Κατανομής

Ασυμμετρία (Skewness)

Skewness = 0

Συμμετρική κατανομή — Mean = Median = Mode

Skewness > 0

Θετική ασυμμετρία (δεξιά ουρά) — Mean > Median. Π.χ. λίγα πλοία με πολύ μεγάλες καθυστερήσεις.

Skewness < 0

Αρνητική ασυμμετρία (αριστερή ουρά) — Mean < Median. Π.χ. λίγα πλοία με πολύ χαμηλή ταχύτητα.

 Κανόνας: $|Skewness| < 0.5$ = ουσιαστικά συμμετρικό
 $|Skewness| 0.5-1.0$ = μέτρια ασύμμετρο $|Skewness| > 1.0$ = έντονη ασυμμετρία

Κύρτωση (Kurtosis)

Η κύρτωση μετρά πόσο 'αιχμηρή' ή 'επίπεδη' είναι η κατανομή σε σχέση με την κανονική.

Mesokurtic ($K \approx 0$)

Κανονική κατανομή — βασική αναφορά

Leptokurtic ($K > 0$)

Αιχμηρή κορυφή, βαριές ουρές — πολλές ακραίες τιμές. Π.χ. σπάνιες αλλά μεγάλες καθυστερήσεις.

Platykurtic ($K < 0$)

 Το Excel υπολογίζει Excess Kurtosis (αφαιρεί 3 από την τιμή)
Για κανονική κατανομή: Excess Kurtosis = 0

Βήμα 1: Ενεργοποίηση Analysis ToolPak

Μόνο μία φορά ανά υπολογιστή

Το Analysis ToolPak είναι ένα πρόσθετο (add-in) του Excel που πρέπει να ενεργοποιηθεί:

1

Αρχείο → Επιλογές

Κάντε κλικ στο tab 'Αρχείο' (File) → 'Επιλογές' (Options) στο κάτω μέρος

2

Επιλογή 'Πρόσθετα'

Στο παράθυρο 'Επιλογές Excel' επιλέξτε 'Πρόσθετα' (Add-ins) στο αριστερό μενού

3

Κουμπί 'Μετάβαση...'

Στο κάτω μέρος, δίπλα στο 'Διαχείριση: Πρόσθετα Excel', κάντε κλικ 'Μετάβαση...' (Go...)

4

Τσεκάρετε το ToolPak

Τσεκάρετε το 'Analysis ToolPak' ✓ → OK. Τώρα εμφανίζεται στο tab 'Δεδομένα' (Data)

επαλήθευση: Πηγαίνετε στο tab 'Δεδομένα' — στα δεξιά θα εμφανιστεί το κουμπί 'Ανάλυση Δεδομένων' (Data Analysis)

Βήμα 2: Εκτέλεση Descriptive Statistics

Data → Data Analysis → Descriptive Statistics

1

Tab 'Δεδομένα' → κουμπί 'Ανάλυση Δεδομένων'

2

Επιλέξτε 'Descriptive Statistics' → OK

3

Input Range: επιλέξτε τα δεδομένα ΜΕ επικεφαλίδα π.χ. \$A\$1:\$E\$16

4

Grouped by: Columns (τα δεδομένα είναι σε στήλες)

5

Τσεκάρτε 'Labels in First Row' (υπάρχουν επικεφαλίδες)

6

Output Range: επιλέξτε κελί π.χ. \$G\$1 για αποτελέσματα

7

Τσεκάρτε 'Summary statistics' → OK

Ρυθμίσεις Παραθύρου

Input Range: \$A\$1:\$E\$16

Grouped By: Columns ✓

Labels in First Row: ✓ (τσεκαρισμένο)

Output Range: \$G\$1

Summary statistics: ✓ (τσεκαρισμένο)

Confidence Level: 95% (προαιρετικό)

 Τα αποτελέσματα εμφανίζονται αυτόματα στο \$G\$1!

Αποτελέσματα: Descriptive Statistics – Ταχύτητα Πλοίων

Ανάλυση της μεταβλητής 'Ταχύτητα (κν)'

Στατιστικό	Τιμή	Ερμηνεία
Mean (Μέσος Όρος)	15.07 κν	Η μέση ταχύτητα του στόλου
Standard Error	0.49	Αβεβαιότητα εκτίμησης μέσου
Median (Διάμεσος)	15.50 κν	Το 50% πλοίων: ταχύτερα, 50%: πιο αργά
Mode (Επικρατούσα)	N/A	Δεν υπάρχει επαναλαμβανόμενη τιμή
Std Deviation (Τυπ. Απόκλ.)	1.87 κν	Κατά μέσο, τα πλοία αποκλίνουν ± 1.87 κν
Sample Variance (Διακύμανση)	3.49	Τετράγωνο Τυπ. Απόκλισης
Kurtosis (Κύρτωση)	-0.82	Πλατύκυρτη (platykurtic) — επίπεδη κατανομή
Skewness (Ασυμμετρία)	0.15	Σχεδόν συμμετρική ($ 0.15 < 0.5$)
Range (Εύρος)	6.5 κν	11.8 έως 18.3 κν
Minimum	11.8 κν	Χαμηλότερη ταχύτητα (Χάρων)
Maximum	18.3 κν	Υψηλότερη ταχύτητα (Αχιλλέας)
Sum	226.1 κν	Άθροισμα όλων των ταχυτήτων
Count	15	Αριθμός πλοίων

Ερμηνεία: Ο στόλος έχει μέση ταχύτητα ~ 15 κν με μικρή διακύμανση — ομοιόμορφη απόδοση πλοίων. Η κατανομή είναι σχεδόν συμμετρική.

Ιστόγραμμα (Histogram) με Excel

Οπτική Αναπαράσταση Κατανομής

Τι είναι το Ιστόγραμμα;

Γραφική αναπαράσταση της κατανομής συχνοτήτων. Ο άξονας Χ δείχνει τις τιμές (bins) και ο άξονας Υ τη συχνότητα εμφάνισης.

1 Ορίστε τα bins (κατηγορίες): π.χ. 11-13, 13-15, 15-17, 17-19 κν

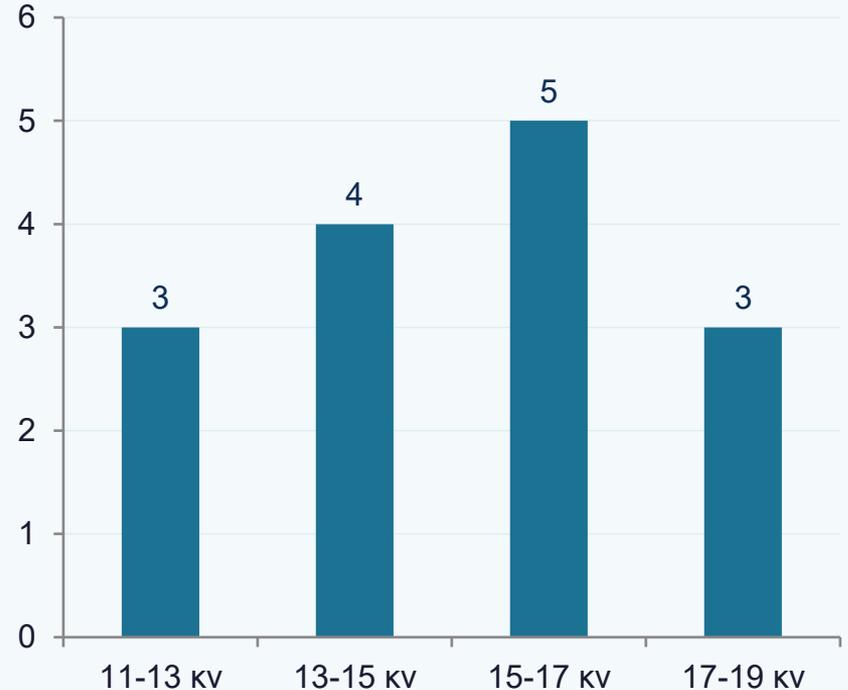
2 Data → Data Analysis → Histogram → OK

3 Input Range: τα δεδομένα ταχύτητας (B2:B16)

4 Bin Range: τα bins που ορίσατε

5 Τσεκάρετε 'Chart Output' → OK

Κατανομή Ταχύτητας Πλοίων



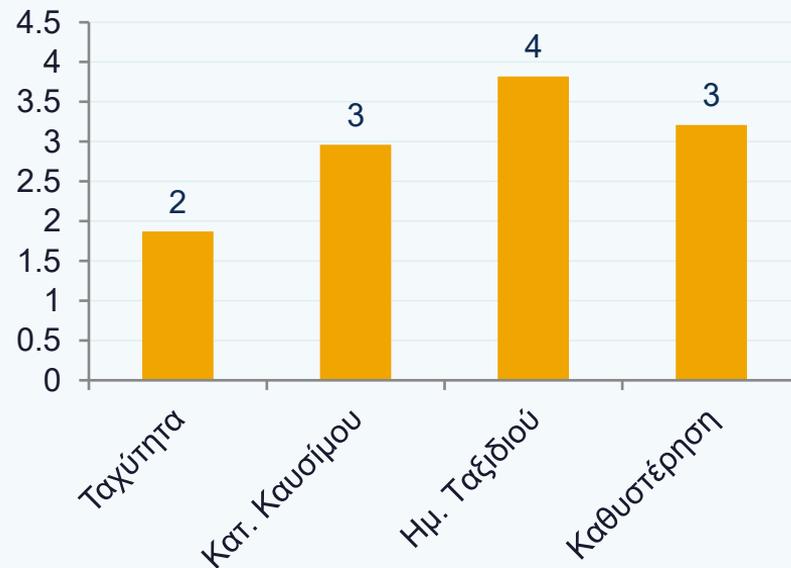
Σύγκριση: Μέσος vs Τυπική Απόκλιση ανά Μεταβλητή

Συνοπτική εικόνα του στόλου

Μέσος Όρος ανά Μεταβλητή



Τυπική Απόκλιση ανά Μεταβλητή



⚓ Συμπέρασμα: Η ταχύτητα έχει μικρή τυπ. απόκλιση (σχετικά ομοιόμορφη) ενώ οι καθυστερήσεις έχουν υψηλή (πολύ μεταβλητές). Τα ημ. ταξιδιού επίσης δείχνουν μεγάλη διακύμανση — σημαντικές διαφορές στα δρομολόγια.

Ερμηνεία Αποτελεσμάτων – Ο Οδηγός σας

Πώς διαβάζουμε τα στατιστικά;

Μέσος vs Διάμεσος

Αν $Mean \approx Median$ → συμμετρική κατανομή
Αν $Mean > Median$ → θετική ασυμμετρία (outliers υψηλά)
Αν $Mean < Median$ → αρνητική ασυμμετρία (outliers χαμηλά)

 Καθυστερήσεις: $Mean=3.8$, $Median=3$ → ελαφρά θετική ασυμμετρία (κάποια πλοία με μεγάλες καθυστερήσεις τραβούν τον μέσο πάνω)

Τυπική Απόκλιση

Κανόνας 68-95-99.7 (για κανονική κατανομή):

- 68% δεδομένων εντός $\pm 1s$ από μέσο
- 95% δεδομένων εντός $\pm 2s$
- 99.7% δεδομένων εντός $\pm 3s$

 Ταχύτητα: $Mean=15.07$, $s=1.87$ → 68% πλοίων: 13.2 έως 16.94 κν

Ασυμμετρία & Κύρτωση

Skewness:
 $|s| < 0.5$: συμμετρικό | $0.5-1$: μέτριο | > 1 : έντονο
Kurtosis (excess):
 ≈ 0 : κανονικό | > 0 : αιχμηρό | < 0 : επίπεδο

 Εάν καθυστέρηση $Skewness=1.8$ → έντονη θετική ασυμμετρία. Ελάχιστα πλοία με τεράστιες καθυστερήσεις επηρεάζουν τη μέση τιμή.

Συχνά Λάθη & Σημεία Προσοχής

Τι πρέπει να αποφύγετε

✘ Σύγκριση Μέσου-Διαμέσου

Ο μέσος όρος δεν αντιπροσωπεύει πάντα 'τυπικό' πλοίο. Αν υπάρχουν ακραίες τιμές, η διάμεσος είναι πιο αντιπροσωπευτική.

✔ Συγκρίνετε πάντα Mean και Median

✘ Άθροιση ποσοτών

ΔΕΝ μπορείτε να αθροίσετε ποσοστά (π.χ. % καθυστέρηση) στα PivotTables — δίνει λανθασμένα αποτελέσματα.

✔ Χρησιμοποιείτε τα πρωτογενή δεδομένα

✘ Διακύμανση δείγματος vs πληθυσμού

Το Excel δίνει διακύμανση δείγματος (n-1). Για τον πλήρη πληθυσμό χρησιμοποιείτε VAR.P (γραφείο) αντί VAR.S.

✔ Η DescStat δίνει πάντα δείγμα (n-1)

✘ Ερμηνεία χωρίς σύμφραζόμενα

Μια τυπική απόκλιση 1.87 κν μπορεί να σημαίνει διαφορετικά πράγματα ανάλογα με το πεδίο. Εξετάζετε και τον Συντελεστή Μεταβλητότητας ($CV = s/\bar{x}$).

✔ $CV = 1.87/15.07 = 12.4\%$ — μέτρια μεταβλητότητα

Εργαστηριακή Άσκηση

Εφαρμόστε όσα μάθατε στα δεδομένα ShipData S.A.

 Οδηγίες: Χρησιμοποιείτε τα δεδομένα 15 πλοίων από τη Διαφάνεια 3

A

Descriptive Statistics

- Εκτελέστε Descriptive Statistics για ΟΛΕΣ τις αριθμητικές μεταβλητές
- Καταγράψτε: Μέσο, Διάμεσο, Τυπ. Απόκλιση, Skewness, Kurtosis
- Ποια μεταβλητή έχει τη μεγαλύτερη σχετική διακύμανση ($CV = s/\bar{x} \times 100\%$);

B

Ιστόγραμμα Καθυστερήσεων

- Ορίστε bins: 0-2, 2-4, 4-6, 6-8, 8-10, 10-12+ ώρες
- Δημιουργήστε ιστόγραμμα για τη μεταβλητή 'Καθυστερήση'
- Τι σχήμα έχει η κατανομή; Επαληθεύστε με το Skewness

Γ

Ερώτηση Ανάλυσης

- Ποιο πλοίο θεωρείτε 'outlier' στην ταχύτητα; (Hint: $|x - \bar{x}| > 2s$)
- Το Αγαμέμνων (11.8 κν, 18 ημέρες) — τι συμπεραίνετε;
- Αν αφαιρέσετε το Αγαμέμνων, πώς αλλάζει ο μέσος ταχύτητας;

Ανακεφαλαίωση



Περιγραφική Στατιστική: συνοψίζει δεδομένα χωρίς συμπεράσματα για πληθυσμό



Μέτρα Κεντρικής Τάσης: Μέσος (επηρεάζεται από outliers) & Διάμεσος (ανθεκτικός)



Μέτρα Διακύμανσης: Τυπ. Απόκλιση σε ίδιες μονάδες με τα δεδομένα — πιο εύκολη ερμηνεία



Skewness & Kurtosis: αποκαλύπτουν ασυμμετρία & ακραίες τιμές στη ναυτιλία



Excel ToolPak: Data → Data Analysis → Descriptive Statistics — σε 7 βήματα



Εφαρμογή: Ανάλυση ταχύτητας, κατανάλωσης, καθυστερήσεων εμπορικού στόλου

 **Επόμενο Βήμα: Συμπερασματική Στατιστική — T-Test, ANOVA, Παλινδρόμηση με το Excel**