

ΒΙΟΣΤΑΤΙΣΤΙΚΗ

Διδάσκων: Τριανταφύλλου Ιωάννης

Τ.Ε.Ι. ΑΘΗΝΑΣ

Αιγάλεω 2014

Περιγραφική Στατιστική

Που και πως θα μας φανεί χρήσιμη???

Για να περιγράψουμε ένα δείγμα παρατηρήσεων ως προς τα χαρακτηριστικά του

Παράδειγμα

Κατά τη διόρθωση 200 γραπτών ενός μαθήματος

- Ποια είναι η μέση βαθμολογία?
- Τι ποσοστό των γραπτών βαθμολογήθηκε κάτω από τη βάση?
- Το σύνολο των γραπτών αποτελεί μία ομοιογενή ομάδα?

Περιγραφική Στατιστική

Βασικά εργαλεία

- (i) Μέτρα θέσης ή κεντρικής τάσης*
- (ii) Μέτρα διασποράς*
- (iii) Μέτρα ασυμμετρίας*
- (iv) Μέτρα κύρτωσης*
- (v) Γραφήματα*

Περιγραφική Στατιστική

Βασικές ενότητες

(A) Μελέτη μεμονωμένων παρατηρήσεων

Παράδειγμα. Έστω ότι καταγράφουμε τα κέρδη 10 τουριστικών επιχειρήσεων κατά τη θερινή περίοδο (σε χιλιάδες Ευρώ):

5, 12, 10, 20, 17, 9, 23, 8, 12, 11

(B) Μελέτη ομαδοποιημένων παρατηρήσεων

Παράδειγμα. Έστω ότι καταγράφουμε τα κέρδη 10 τουριστικών επιχειρήσεων κατά τη θερινή περίοδο (σε χιλιάδες Ευρώ):

Έσοδα	Αριθμός επιχειρήσεων
[0-8)	1
[8-16)	6
[16-24)	3

Περιγραφική Στατιστική

Πληθυσμός: περιλαμβάνει μεγάλο αριθμό παρατηρήσεων

Δείγμα: περιλαμβάνει μικρό αριθμό παρατηρήσεων X_1, X_2, \dots, X_n
αντιπροσωπεύοντας ολόκληρο τον πληθυσμό

(A) Μελέτη μεμονωμένων παρατηρήσεων

(i) Μέτρα θέσης

(α) Μέση τιμή (ή μέσος όρος)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- ✓ Δίνει την κεντρική τάση του δείγματος (είναι ο «πιο αντιπροσωπευτικός αριθμός του δείγματος»)

Περιγραφική Στατιστική

(β) Σταθμική μέση τιμή (ή σταθμικός μέσος όρος)

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_v x_v}{w_1 + w_2 + \dots + w_v}, \quad w_i \in \mathfrak{R}$$

✓ **Ερμηνεία:** Δίνει την κεντρική τάση του δείγματος δίνοντας μεγαλύτερη σημασία σε κάποιες εκ των παρατηρήσεων

(γ) Διάμεσος

1^η περίπτωση: το πλήθος των παρατηρήσεων είναι μονό (περιττό)

Παράδειγμα: 1, 4, 6, 12, 18 ($v = 5$)

$\delta = x_{((v+1)/2)}$

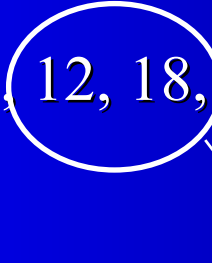
Περιγραφική Στατιστική

(γ) Διάμεσος

2^η περίπτωση: το πλήθος των παρατηρήσεων είναι ζυγό (άρτιο)

$$\delta = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

Παράδειγμα: 1, 4, 6, 8, 12, 18, 22, 30, 44, 55 ($n = 10$)


$$\delta = \frac{x_{(5)} + x_{(6)}}{2} = \frac{12 + 18}{2} = 15$$

- ✓ **Ερμηνεία:** Το 50% των παρατηρήσεων του δείγματος (δηλαδή οι μισές παρατηρήσεις) είναι μικρότερες ή ίσες από τη διάμεσο

Περιγραφική Στατιστική

(δ) Επικρατούσα τιμή M_0 (ή κορυφή)

- ✓ Είναι η παρατήρηση με τη μεγαλύτερη συχνότητα

(ii) Μέτρα διασποράς

(α) Διακύμανση

- ✓ Βασίζεται στις αποστάσεις των παρατηρήσεων του δείγματος από τη μέση τιμή του

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

Περιγραφική Στατιστική

(β) Τυπική απόκλιση

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

- ✓ Εκφράζει την κατά μέσο όρο απόσταση των παρατηρήσεων του δείγματος από τη μέση τιμή τους

(γ) Εύρος

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

- ✓ Εκφράζει τη μέγιστη απόσταση μεταξύ δύο οποιονδήποτε παρατηρήσεων του δείγματος

Περιγραφική Στατιστική

Παρατήρηση. Ο συντελεστής μεταβλητότητας ορίζεται ως εξής:

$$CV = \frac{s}{\bar{x}}$$

Αν $CV \leq 0.1$, τότε το δείγμα χαρακτηρίζεται ως ομοιογενές

Αν $CV > 0.1$, τότε το δείγμα χαρακτηρίζεται ως ανομοιογενές

(iii) Μέτρα ασυμμετρίας

Συντελεστής ασυμμετρίας κατά Pearson

$$\gamma = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3}$$

$$S.E.S = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

Περιγραφική Στατιστική

- ✓ Εκφράζει το βαθμό συμμετρίας των παρατηρήσεων του δείγματος γύρω από τη μέση τιμή τους
- **1^η περίπτωση:** Αν ο συντελεστής $\gamma > 0$, τότε λέμε ότι το δείγμα παρουσιάζει **θετική ασυμμετρία**.
- ✓ **Πρακτικά:** Οι περισσότερες παρατηρήσεις του δείγματος είναι μικρότερες από τη μέση τιμή του ($M_0 < \delta < \bar{x}$)
- **2^η περίπτωση:** Αν ο συντελεστής $\gamma < 0$, τότε λέμε ότι το δείγμα παρουσιάζει **αρνητική ασυμμετρία**. ($\bar{x} < \delta < M_0$)
- ✓ **Πρακτικά:** Οι περισσότερες παρατηρήσεις του δείγματος είναι μεγαλύτερες από τη μέση τιμή του

Περιγραφική Στατιστική

(iv) Μέτρα κύρτωσης

- ✓ Εκφράζει το βαθμό συγκέντρωσης των παρατηρήσεων του δείγματος γύρω από την κορυφή τους

Συντελεστής κύρτωσης κατά Pearson

$$S.E.K = \sqrt{\frac{\nu^2 - 1}{(\nu - 3)(\nu + 5)}}$$

$$a = \frac{\frac{1}{\nu} \sum_{i=1}^{\nu} (x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{\nu} \sum_{i=1}^{\nu} (x_i - \bar{x})^2} \right)^4} \quad a' = a - 3$$

- **1^η περίπτωση:** Αν ο συντελεστής $a > 3$ (ή ισοδύναμα $a' > 0$), τότε λέμε ότι το δείγμα παρουσιάζει **θετική κύρτωση**.
- ✓ **Πρακτικά:** Υπάρχει μικρό πλήθος παρατηρήσεων κοντά στην κορυφή του δείγματος

Περιγραφική Στατιστική

- **2^η περίπτωση:** Αν ο συντελεστής $a < 3$ (ή ισοδύναμα $a' < 0$), τότε λέμε ότι το δείγμα παρουσιάζει **αρνητική κύρτωση**.
- ✓ **Πρακτικά:** Υπάρχει μεγάλο πλήθος παρατηρήσεων κοντά στην κορυφή του δείγματος

Παράδειγμα 1. Εταιρία στατιστικών μελετών διεξήγαγε έρευνα για το μισθολογικό (μηνιαίο) καθεστώς των εργαζομένων σε λογιστικά γραφεία. Τα αποτελέσματα της έρευνας για 12 εργαζομένους δίνονται ακολούθως:

750, 500, 750, 750, 850, 1000, 600, 500, 800, 500, 700, 700.

Να περιγραφεί το δείγμα ως προς την κεντρική τάση, τη διασπορά, την ασυμμετρία και την κύρτωση που παρουσιάζει.

Περιγραφική Στατιστική

Παράδειγμα 2. Έστω δείγμα n παρατηρήσεων με μέση τιμή \bar{x} και τυπική απόκλιση s .

- (α) Αν υποθέσουμε ότι όλες οι παρατηρήσεις του δείγματος αυξάνονται κατά c μονάδες, να υπολογιστεί η μέση τιμή, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας του νέου δείγματος.
- (β) Αν υποθέσουμε ότι όλες οι παρατηρήσεις του δείγματος τριπλασιάζονται, να υπολογιστεί η μέση τιμή, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας του νέου δείγματος.

Περιγραφική Στατιστική

Παράδειγμα 3. Έστω δείγμα X_1, X_2, \dots, X_{10} που εκφράζει το πλήθος των ωρών που εργάζονται 10 υπάλληλοι μίας εταιρίας. Αν γνωρίζουμε ότι

$$\sum_{i=1}^{10} X_i = 94, \quad \sum_{i=1}^{10} X_i^2 = 900$$

να υπολογιστεί η μέση τιμή και η διακύμανση του δείγματος.

Παράδειγμα 4. Σε μία επιχείρηση, γνωρίζουμε ότι απασχολούνται 3 ανώτερα διοικητικά στελέχη με μέσο μηνιαίο μισθό $\bar{x}_1 = 3000$ Ευρώ, 6 υπάλληλοι με πτυχίο Τριτοβάθμιας Εκπαίδευσης με μέσο μηνιαίο μισθό $\bar{x}_2 = 1500$ και 8 υπάλληλοι βασικής Εκπαίδευσης με μέσο μηνιαίο μισθό $\bar{x}_3 = 900$

- (α) Να υπολογιστεί ο μέσος μηνιαίος μισθός των 17 υπαλλήλων.
(β) Αν ο μισθός των 3 ανωτέρων στελεχών μειωθεί κατά 200 Ευρώ, να υπολογιστεί ο μέσος μηνιαίος μισθός των 17 υπαλλήλων.

Περιγραφική Στατιστική

(B) Μελέτη ομαδοποιημένων παρατηρήσεων

Ομάδα	Συχνότητα	Κέντρο ομάδας
1η	v_1	x_1
2η	v_2	x_2
\vdots	\vdots	\vdots
κ-οστή	v_k	x_k

(α) Μέση τιμή (ή μέσος όρος)

$$\bar{x} = \frac{v_1 x_1 + v_2 x_2 + \dots + v_k x_k}{v_1 + v_2 + \dots + v_k}, \quad v_i \in \mathbb{Z}^+$$

Περιγραφική Στατιστική

(β) Διακύμανση

$$s^2 = \frac{\nu_1(x_1 - \bar{x})^2 + \nu_2(x_2 - \bar{x})^2 + \dots + \nu_k(x_k - \bar{x})^2}{\nu - 1}$$

(γ) Διάμεσος

$$\delta = L_i + \frac{c}{\nu_i} \left(\frac{\nu}{2} - N_{i-1} \right)$$

(δ) Επικρατούσα τιμή (ή κορυφή)

$$M_0 = L_i + c \frac{\Delta_1}{\Delta_1 + \Delta_2}, \quad \Delta_1 = \nu_i - \nu_{i-1}, \quad \Delta_2 = \nu_i - \nu_{i+1}$$

Περιγραφική Στατιστική

(ε) Ποσοστημότητα α %

$$P_a = L_i + \frac{c}{v_i} \left(\frac{av}{100} - N_{i-1} \right)$$

Παράδειγμα 5. Κατά τη θερινή περίοδο, ο αριθμός των πελατών που είχαν 20 ξενοδοχεία (ανά ημέρα) δίνονται ως εξής:

Ομάδα	Συχνότητα
[0,10)	5
[10,20]	8
[20,30)	4
[30,40)	3

Να υπολογισθούν μέτρα θέσης και διασποράς του δείγματος.

Περιγραφική Στατιστική

Παράδειγμα 6. Κατά τη θερινή περίοδο, ο αριθμός των πελατών που είχαν 20 ξενοδοχεία (ανά ημέρα) δίνονται ως εξής:

7,5,12,13,17,22,32,19,30,36,32,20,31,24,27,35,11,6,9,15

(α) Να ομαδοποιηθούν τα δεδομένα σε κατάλληλο αριθμό ομάδων.

$$q = 1 + 3,32 \log_{10} n$$

(β) Να υπολογισθούν μέτρα θέσης και διασποράς του δείγματος (πριν και μετά την ομαδοποίηση).

Περιγραφική Στατιστική

(ν) Γραφήματα

(α) Ιστόγραμμα συχνοτήτων (απλών, σχετικών και αθροιστικών)

(β) Κυκλικό διάγραμμα (ή διάγραμμα-πίτα)

(γ) Θηκόγραμμα

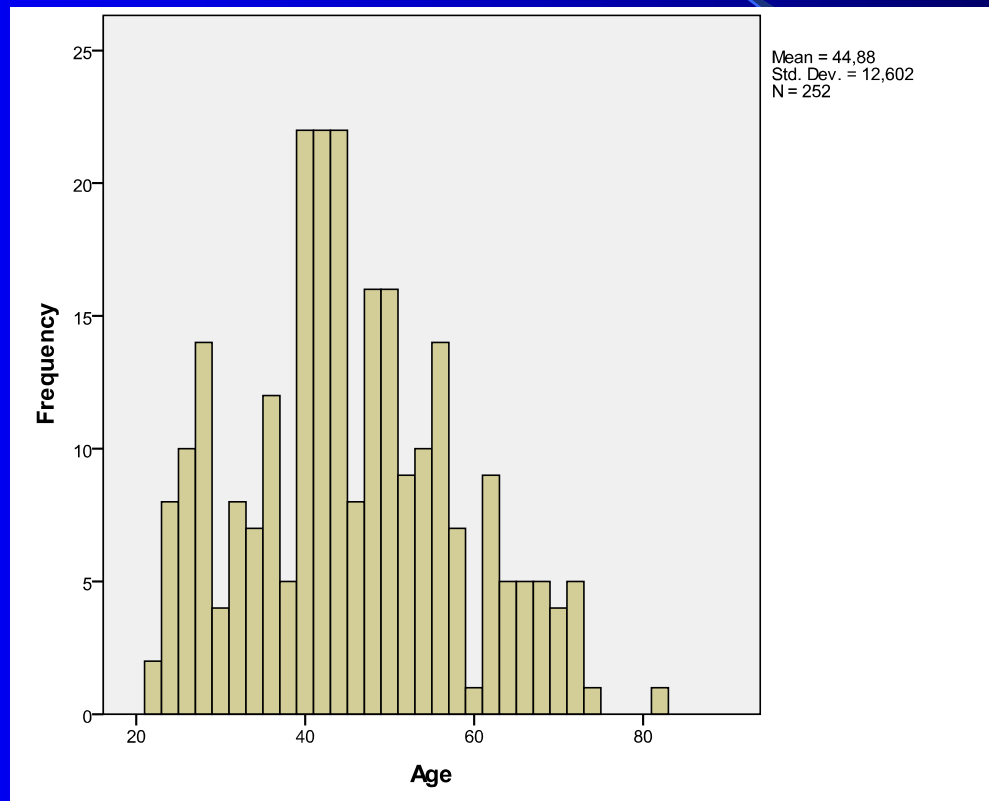
Περιγραφική Στατιστική

Πίνακας συχνοτήτων
(απλών, σχετικών και αθροιστικών)

Βαθμολογία	Συχνότητα	Σχετική Συχνότητα	Αθροιστική συχνότητα	Αθροιστική σχετική συχνότητα
[0,3)	19			
[3,5)		34%		
[5,7)				
[8,10]	3			
Σύνολο	50			

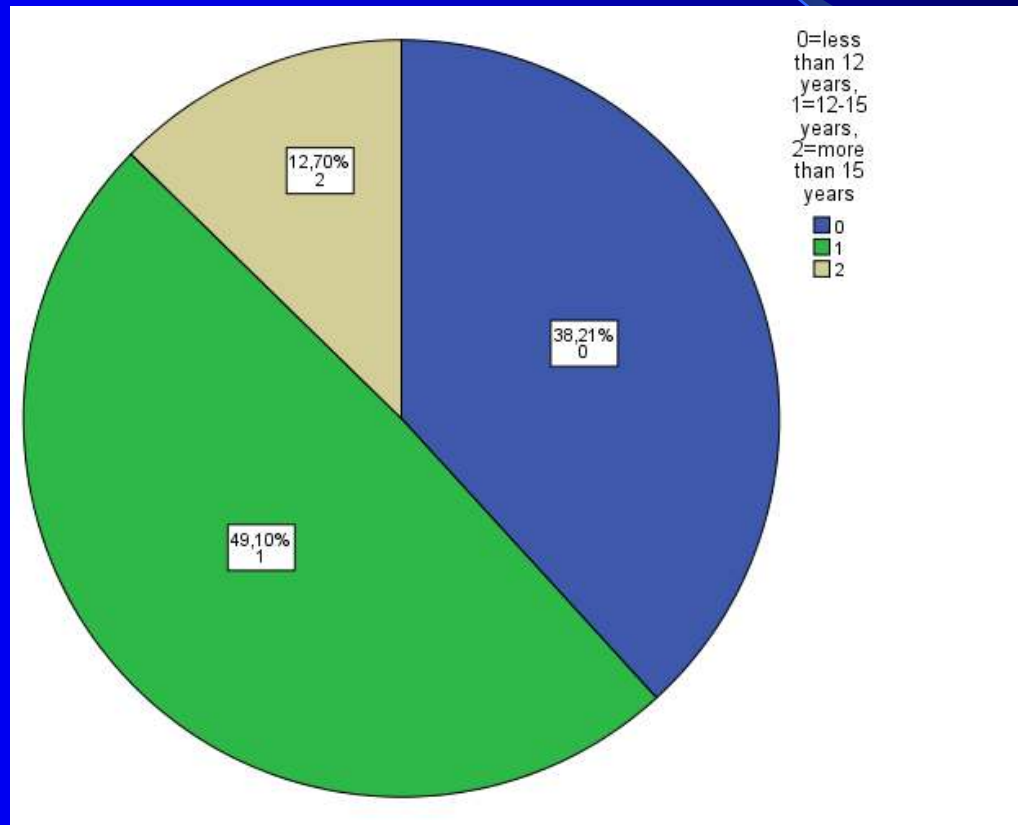
Περιγραφική Στατιστική

(α) Ιστόγραμμα συχνοτήτων



Περιγραφική Στατιστική

(β) Κυκλικό διάγραμμα (ή διάγραμμα-πίτα)



Περιγραφική Στατιστική

(γ) Θηκόγραμμα

