

Μάθημα 14

§ 6.6 Συντελεστής Μεταβλητότητας

Η τυπική απόκλιση και η διακύμανση των τιμών μιας μεταβλητής εξαρτώνται από τις μονάδες μέτρησης των δεδομένων. π.χ. αν ένα σύνολο μετρήσεων βαρών έχει τυπική απόκλιση $s=4$ μονάδες η απόκλιση αυτή επιδέχεται διαφορετικές ερμηνείες ανάλογα με το αν τα βάρη είναι μετρημένα σε κιλά ή σε τόνους. Η αδυναμία αυτή διορθώνεται με τον **συντελεστή μεταβλητότητας** (coefficient of variation).

Ορίζεται ως ο λόγος της τυπικής απόκλισης προς τον αριθμητικό μέσο των μετρήσεων δηλαδή:

$$CV = \frac{s}{\mu} (\times 100) \quad \text{ή} \quad cv = \frac{s}{\bar{x}} (\times 100)$$

ανάλογα με το αν πρόκειται για δεδομένα που προέρχονται από πληθυσμό ή δείγμα, αντίστοιχα. Ο συντελεστής μεταβλητότητας συνήθως εκφράζεται επί τοις 100 δηλαδή εκφράζει την τυπική απόκλιση ως ποσοστό του μέσου των δεδομένων.

Παράδειγμα Σε τυχαία επιλεγμένο δείγμα 6 ατόμων μετρήθηκε το βάρος και το ύψος τους ως εξής:

Βάρος (σε κιλά)

Ύψος (σε εκατοστά)

72, 67, 87, 70, 85, 78

173, 167, 185, 174, 182, 179

Θα συζητήσουμε τη μεταβλητότητα των δύο ειδών μετρήσεων. Ο μέσος των βαρών είναι $\bar{x}_B = 76,5$ κιλά και η τυπική απόκλιση $s_B = 8,216$ κιλά ενώ ο μέσος των υψών είναι

$\bar{X}_0 = 176,67 \text{ cm}$ και η τυπική απόκλιση $S_0 = 6,573 \text{ cm}$. Οι

διαφορετικές μονάδες μέτρησης δεν επιτρέπουν την απευθείας σύγκριση των τυπικών αποκλίσεων τους και των διακυμάνσεων τους και επομένως οδηγούμαστε στην σύγκριση των αντίστοιχων συντελεστών μεταβλητότητας οι οποίοι είναι:

$$cv(X_B) = \left(\frac{S_B}{\bar{X}_B} \right) \times 100 = \left(\frac{8,216}{76,5} \right) \times 100 = 10,74\%$$

$$cv(X_0) = \left(\frac{S_0}{\bar{X}_0} \right) \times 100 = \left(\frac{6,573}{176,67} \right) \times 100 = 3,72\%$$

Η απευθείας σύγκριση των συντελεστών μεταβλητότητας μας οδηγεί στο συμπέρασμα μεγαλύτερης μεταβλητότητας των βαρών των ατόρων σε σχέση με τα ύψη τους. Από τον συντελεστή $cv(X_B) = 10,74\%$ προκύπτει ότι η τυπική απόκλιση είναι το $10,74\%$ του μέσου θάρους. Ο συντελεστής $cv(X_0) = 3,72\%$ δηλώνει ότι η τυπική απόκλιση είναι το $3,72\%$ του μέσου ύψους.

§6.7 Συντελεστής Συγκέντρωσης του Gini

Είναι μέτρο παρόμοιο με το συντελεστή μεταβλητότητας

και ορίζεται ως το πηγάριο της μέσης διαφοράς Δ

διά το διπλάσιο του μέσου μ δηλαδή: $G = \frac{\Delta}{2\mu}$, όπου

$$\mu = \frac{\sum_{i=1}^N x_i}{N} \quad \text{και} \quad \Delta = \frac{2}{N^2} \sum_{r=1}^{N-1} F_r(N - F_r)$$

Αποδεικνύεται ότι ο συντελεστής του Gini (G) παίρνει τιμές στο διάστημα $[0, 1]$. Η ποσότητα F_r είναι η αθροί-

στην συχνότητα του x_r .

Παράδειγμα Για το σύνολο των επιταχύνσεων 155 ανθρωπίνων υπολογίζουμε τον συντελεστή συσχέτισης του Gini:

Κλάσεις	w_i	f_r	F_r	$N - f_r$	$F_r(N - f_r)$
10-11	10,5	1	1	154	154
11-12	11,5	3	4	150	600
12-13	12,5	8	12	142	1704
13-14	13,5	13	25	130	3250
14-15	14,5	29	54	101	5454
15-16	15,5	28	82	73	5986
16-17	16,5	22	104	51	5304
17-18	17,5	17	121	34	4114
18-19	18,5	13	134	21	2814
19-20	19,5	9	143	12	1716
20-21	20,5	5	148	7	1036
21-22	21,5	3	151	4	604
22-23	22,5	2	153	2	306
23-24	23,5	1	154	1	154
24-25	24,5	1	155	-	-
	155			33196	

$$\text{Άρα } \mu = \frac{1}{155} \sum_{i=1}^{15} f_i w_i = \frac{1}{155} \times 2511,5 = 16,203$$

$$\Delta = \frac{2}{N^2} \sum_{r=1}^{N-1} F_r(N - F_r) = \frac{2}{155^2} 33196 = 2,763$$

$$G = \frac{\Delta}{2\mu} = \frac{2,763}{2 \times 16,203} = 0,0853$$

Η τιμή αυτή δείχνει πόσο διαφέρει κάθε μέτρηση από όλες τις άλλες.

Η μικρή τιμή του G δηλώνει μικρή μεταβλητότητα των δεδομένων.

§6.8 Οραδοποιημένα Δεδομένα

(α) Μέση απόκλιση

Πληθυσμός μεγέθους N

$$A_1 = \frac{\sum_{i=1}^K f_i |w_i - \mu|}{N}$$

για σύνολο συνεχών δεδομένων
σε κατανομή K κλάσεων.Δείγμα μεγέθους n

$$a_1 = \frac{\sum_{i=1}^K f_i |w_i - \bar{x}|}{n}$$

Αν οι απόλυτες αποκλίσεις
των κέντρων των τάξεων
υπολογίζονται από τη
διάμεσο οι σχέσεις

γίνονται: $A'_1 = \frac{1}{N} \sum_{i=1}^K f_i |w_i - M|$ και $a'_1 = \frac{1}{n} \sum_{i=1}^K f_i |w_i - m|$

όπου M : διάμεσος πληθυσμού, m : διάμεσος δείγματος.

(β) Διασπορά και τυπική απόκλιση

Πληθυσμός μεγέθους N

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

από όπου έχουμε:

$$\sigma^2 = \frac{\sum_{i=1}^K f_i w_i^2 - 2\mu \sum_{i=1}^K f_i w_i + \mu^2 \sum_{i=1}^K f_i}{N}$$

$$= \frac{\sum_{i=1}^K f_i w_i^2}{N} - 2\mu^2 + \mu^2 = \frac{\sum_{i=1}^K f_i w_i^2}{N} - \mu^2$$

Ο τεχνητός τύπος δεινοδώνει τη διαδικασία υπολογισμού της διασποράς του πληθυσμού.

Τυπική απόκλιση πληθυσμού

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^K f_i (w_i - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^K f_i w_i^2}{N} - \mu^2}$$

Για το δείγμα μεγέθους n , η δειγματική διακύμανση

είναι
$$s^2 = \frac{\sum_{i=1}^K f_i (w_i - \bar{x})^2}{n-1}$$
 από τον οποίο παίρνουμε έναν

τύπο που δεινοδώνει τη διαδικασία υπολογισμού της δειγματικής διακύμανσης:

$$s^2 = \frac{\sum_{i=1}^K f_i w_i^2 - 2\bar{x} \sum_{i=1}^K f_i w_i + \bar{x}^2 \sum_{i=1}^K f_i}{n-1}$$
$$= \frac{\sum_{i=1}^K f_i w_i^2 - 2\bar{x} n\bar{x} + n\bar{x}^2}{n-1} = \frac{1}{n-1} \left[\sum_{i=1}^K f_i w_i^2 - n\bar{x}^2 \right]$$

και η δειγματική τυπική απόκλιση δίνεται ως:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^K f_i (w_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^K f_i w_i^2 - n\bar{x}^2}{n-1}}$$

Παράδειγμα

Κατανομή συχνοτήτων των επιταχύνσεων από 0-200 km/h σε sec δείγματος $n=155$ αυτοκινήτων. Θέλουμε να επι-
μήσουμε τη διασπορά της κατανομής.

Μήσες	w_i	f_i	$f_i w_i$	$f_i w_i^2$
[10,11)	10,5	1	10,5	110,25
[11,12)	11,5	3	34,5	396,75
[12,13)	12,5	8	100	1250
[13,14)	13,5	13	175,5	2369,25
[14,15)	14,5	23	420,5	6097,25
[15,16)	15,5	28	434	6727
[16,17)	16,5	22	363	5989,5
[17,18)	17,5	17	297,5	5206,25
[18,19)	18,5	13	240,5	4449,25
[19,20)	19,5	9	175,5	3422,25
[20,21)	20,5	5	102,5	2101,25
[21,22)	21,5	3	64,5	1386,75
[22,23)	22,5	2	45	1012,5
[23,24)	23,5	1	23,5	552,25
[24,25)	24,5	1	24,5	600,25
Σύνολο	155		2511,5	41670,75

$$\bar{x} = \frac{2511,5}{155} = 16,203, \quad s^2 = \frac{1}{155-1} (41670,75 - 155 \times 16,203^2) = 6,347$$

$$s = \sqrt{6,347} = 2,519, \quad cv = \frac{2,519}{16,203} \times 100 = 15,5\%$$