ΣΤΑΤΙΣΤΙΚΗ (ΜΥ0202)

Μ.Ν. Ντυκέν,
Πανεπιστήμιο Θεσσαλίας
Τ.Μ.Χ.Π.Π.Α.

Βόλος, 2018-2019

ΔΙΑΛΕΞΗ 07

Περιεχόμενο της Διάλεξης
Μέτρα διασποράς , μεταβλητότητας
Μέτρα ασυμμετρίας

2. Διερευνητική Ανάλυση

Μέτρα Θέσης

Μέτρα Διασποράς

Ασυμμετρία

2. Διερευνητική Ανάλυση

Μέτρα Διασποράς - Μεταβλητότητας

Τα Μέτρα διασποράς

Τα μέτρα διασποράς για μια μεταβλητή Χ εκφράζουν τις **<u>αποκλίσεις</u>** των τιμών που παίρνει η μεταβλητή **<u>γύρω από τα μετρά θέσης</u>** και ειδικά γύρω από την μέση τιμή.

Τα μέτρα διασποράς είναι χρήσιμα διότι μας δίνουν, με τρόπο περιληπτικό και αντικειμενικό, πληροφορίες σχετικά με τη **<u>μεταβλητότητα</u>** ή την **<u>ανομοιογένει</u>**α των παρατηρήσεων.

2. ΚΥΡΙΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

1. Εύρος	$e = Max - Min$	Επηρεάζεται από τον πλήθος των παρατηρήσεων και από τις 2 ακραίες τιμές.
2. Ενδοτεταρτομοριακό διάστημα (εύρος)	$D_F = Q_3 - Q_1$	Διάστημα της κατανομής μέσα στον οποίο έχουμε το 50% των κεντρικών τιμών. [Όσο πιο μικρό είναι το D_F τόσο πιο συγκεντρωμένες είναι οι τιμές γύρω από τη Διάμεσο]
3. Σχετικό Ενδοτεταρτομοριακό διάστημα	$ED = \frac{Q_3 - Q_1}{Q_2}$	Μέτρο σχετικής συγκέντρωσης των τιμών.
4. Διάστημα του Kelley	$DK = D_9 - D_1$	Διάστημα της κατανομής μέσα στον οποίο έχουμε το 90% των τιμών.
5. Αναλόγια μεταξύ 9ου Δεκατημόριου και 1ου Δεκατημόριου (Dispersion Ratio)	$DR = \frac{D_9}{D_1}$	Μέτρο διασποράς. Χρησιμοποιείται συχνά στην ανάλυση των ανισοτήτων.

2. ΚΥΡΙΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

6. Μέση απόκλιση
   - Ατομικά Δεδομένα: $$MA = \frac{\sum_{i=1}^{n} \left| Xi - \overline{X} \right|}{n}$$
   - Ομαδοποιημένα: $$MA = \frac{\sum_{i=1}^{k} n_i \left| Xi - \overline{X} \right|}{n}$$

Μέση απόσταση όλων των τιμών σε σχέση με την μέση τιμή.

7. Διακύμανση για <u>πληθυσμό</u> ή <u>μεγάλο</u> δείγμα
   - Ατομικά Δεδομένα: $$V[X] = s^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}{n}$$
   - Ομαδοποιημένα: $$V[X] = s^2 = \frac{\sum_{i=1}^{k} n_i \left(X_i - \overline{X} \right)^2}{n}$$

8. Διακύμανση για <u>μικρό</u> δείγμα
   - Ατομικά Δεδομένα: $$V[X] = s^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}{n-1}$$
   - Ομαδοποιημένα: $$V[X] = s^2 = \frac{\sum_{i=1}^{k} n_i \left(X_i - \overline{X} \right)^2}{n-1}$$

n = πλήθος παρατηρήσεων

k = αριθμός κλάσεων όταν τα δεδομένα είναι ομαδοποιημένα

2. ΚΥΡΙΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

9. Τυπική Απόκλιση $s = \sqrt{V(X)}$

Εκφράζεται στην ίδια μονάδα μέτρησης όπως η μέση τιμή.

Αν το μέσο μηνιαίο εισόδημα των νοικοκυριών (σε €) = 870 ενώ η διακύμανση = 27225,

➔ η μονάδα μέτρησης της διακύμανσης V[X] είναι σε $€^2$
ενώ η τυπική απόκλιση s είναι σε €.

10. Δείκτης Μεταβλητότητας $CV = \frac{s}{\overline{\overline{X}}}$

Συχνά εκφράζεται σε %.

Στην χωρική ανάλυση, πρόκειται για ένα από ***τους βασικότερους δείκτες μέτρησης των ανισοτήτων***.

2. ΚΥΡΙΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Με τη διασπορά, εξετάζουμε σε ποιο βαθμό οι τιμές της μεταβλητής Χ (οι παρατηρήσεις) απέχουν από τη μέση τιμή.

- ✓Όσο πιο πολύ οι τιμές απέχουν από τη μέση τιμή, τόσο μεγαλύτερη είναι η διασπορά.
- ✓Όταν όλες οι τιμές δεν απέχουν πολύ από τη μέση τιμή και κατά συνέπεια συγκεντρώνονται γύρω από τη μέση τιμή, τότε η διασπορά είναι μικρή.

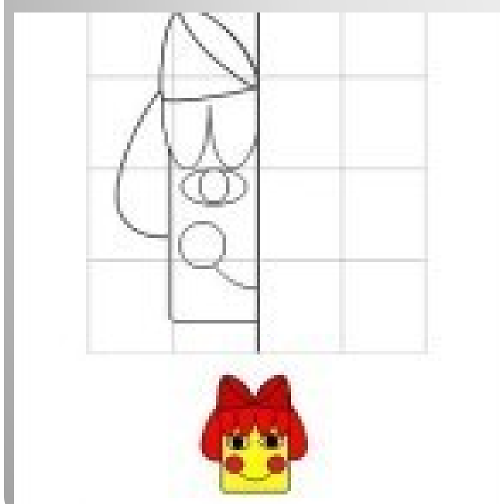
Η διασπορά είναι αξιόπιστο μέτρο διασποράς.

Μειονέκτημα:

Οι μονάδες είναι υψωμένες στο τετράγωνο. Αν η μεταβλητή Χ αφορά το μηνιαίο εισόδημα των νοικοκυριών σε €, η διασπορά εκφράζεται σε €^{2}!
Αντίθετα, η τετραγωνική ρίζα της διασποράς (η τυπική απόκλιση) εκφράζεται στην ίδια μονάδα μέτρησης όπως η μέση τιμή .

ΥΠΟΛΟΓΙΣΜΟΣ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ

ΑΣΚΗΣΗ 1



ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ [01]

Η μεταβλητή που εξετάζεται είναι η **MPP = *Ποσοστό (%) μονογονεϊκών οικογενειών με τουλάχιστον ένα παιδί στο σύνολο των οικογενειών (2011)*** *για τις 13 περιφέρειες της Ελλάδας.*

i	ΧΩΡΙΚΕΣ ΕΝΟΤΗΤΕΣ	MPP
1	Ανατ. Μακεδονία & Θράκη	12,3
2	Δυτική Μακεδονία	12,6
3	Κρήτη	13,2
4	Βόρειο Αιγαίο	13,4
5	Θεσσαλία	13,7
6	Νότιο Αιγαίο	13,8
7	Ήπειρος	14,1
8	Κεντρική Μακεδονία	14,3
9	Ιόνια Νησιά	15,0
10	Στερεά Ελλάδα	15,2
11	Πελοπόννησος	15,2
12	Δυτική Ελλάδα	15,6
13	Αττική	17,7

Πηγή:Eurostat, Census 2011

1. *Να Υπολογίσετε όλα τα μετρά διασποράς.*
2. *Ποια τα συμπεράσματα σας;*

Για σωστή επεξεργασία, είναι απαραίτητο να παρουσιάζετε τα αποτελέσματα σε μορφή πίνακας

ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ [02]

i	ΧΩΡΙΚΕΣ ΕΝΟΤΗΤΕΣ	MPP
1	Ανατ. Μακεδονία & Θράκη	12,3
2	Δυτική Μακεδονία	12,6
3	Κρήτη	13,2
4	Βόρειο Αιγαίο	13,4
5	Θεσσαλία	13,7
6	Νότιο Αιγαίο	13,8
7	Ήπειρος	14,1
8	Κεντρική Μακεδονία	14,3
9	Ιόνια Νησιά	15,0
10	Στερεά Ελλάδα	15,2
11	Πελοπόννησος	15,2
12	Δυτική Ελλάδα	15,6
13	Αττική	17,7

Εύρος = Μέγιστη τιμή – Ελάχιστη = 5,4

Md = Q_2 = Διάμεσος = X_7 = 14,1
Q_1 = 1ο τεταρτημόριο = X_4 = 13,4
Q_3 = 3ο τεταρτημόριο = X_{10} = 15,2

Ενδοτεταρτημοριακό εύρος = Q_3 - Q_1 = 1,8

Το 50% των κεντρικών παρατηρήσεων βρίσκεται σε ένα διάστημα = 1,8 μονάδες ενώ το 100% βρίσκεται σε εύρος = 5,4 μονάδες.
Το 50% των κεντρικών παρατηρήσεων βρίσκεται σε ένα διάστημα που αντιστοιχεί σε περίπου 1/3 του συνολικού εύρους (1,8/5,4 = 0,333).

Υπάρχει σχετική συγκέντρωση στο κέντρο της κατανομής.

ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ [02]

i	ΧΩΡΙΚΕΣ ΕΝΟΤΗΤΕΣ	X= MPP	$\lvert X_i - \bar{X} \rvert$	$(X_i - \bar{X})^2$
1	Ανατ. Μακεδονία & Θράκη	12,3	2,02	4,06
2	Δυτική Μακεδονία	12,6	1,72	2,94
3	Κρήτη	13,2	1,12	1,24
4	Βόρειο Αιγαίο	13,4	0,92	0,84
5	Θεσσαλία	13,7	0,62	0,38
6	Νότιο Αιγαίο	13,8	0,52	0,27
7	Ήπειρος	14,1	0,22	0,05
8	Κεντρική Μακεδονία	14,3	0,02	0,00
9	Ιόνια Νησιά	15,0	0,68	0,47
10	Στερεά Ελλάδα	15,2	0,88	0,78
11	Πελοπόννησος	15,2	0,88	0,78
12	Δυτική Ελλάδα	15,6	1,28	1,65
13	Αττική	17,7	3,38	11,46
	ΣΥΝΟΛΟ	**186,10**	**14,25**	**24,92**

| 12,3 – 14,32 | = 2,02

(12,3 – 14,32)² = 4,06

Οι υπολογισμοί πραγματοποιήθηκαν με το Excel και λαμβάνουν υπόψη όλα τα δεκαδικά.

$$\sum_{i=1}^{13} X_i = 186,1$$

$$\sum_{i-1}^{13} \lvert X_i - \bar{X} \rvert = 14,25$$

$$\sum_{i-1}^{13} (X_i - \bar{X})^2 = 24,92$$

ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ [02]

i	ΧΩΡΙΚΕΣ ΕΝΟΤΗΤΕΣ	X= MPP	$\lvert X_i - \bar{X} \rvert$	$(X_i - \bar{X})^2$
1	Ανατ. Μακεδονία & Θράκη	12,3	2,02	4,06
2	Δυτική Μακεδονία	12,6	1,72	2,94
3	Κρήτη	13,2	1,12	1,24
4	Βόρειο Αιγαίο	13,4	0,92	0,84
5	Θεσσαλία	13,7	0,62	0,38
6	Νότιο Αιγαίο	13,8	0,52	0,27
7	Ήπειρος	14,1	0,22	0,05
8	Κεντρική Μακεδονία	14,3	0,02	0,00
9	Ιόνια Νησιά	15,0	0,68	0,47
10	Στερεά Ελλάδα	15,2	0,88	0,78
11	Πελοπόννησος	15,2	0,88	0,78
12	Δυτική Ελλάδα	15,6	1,28	1,65
13	Αττική	17,7	3,38	11,46
	ΣΥΝΟΛΟ	**186,10**	**14,25**	**24,92**

$$\bar{X} = \frac{1}{13}\sum_{i=1}^{13} X_i = \frac{186{,}1}{13} = \mathbf{14,32}$$

$$MA = \frac{1}{13}\sum_{i-1}^{13} \lvert X_i - \bar{X} \rvert = \frac{14{,}25}{13} = \mathbf{1,10}$$

$$s^2 = \frac{1}{n}\sum_{i-1}^{13} (X_i - \bar{X})^2 = \frac{24{,}92}{13} = \mathbf{1,917}$$

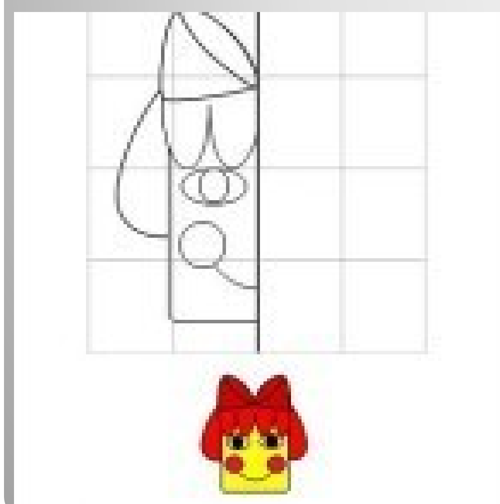
$$s = \sqrt{s^2} = \sqrt{1{,}917} = 1{,}38$$

$$CV = \frac{s}{\bar{X}} = \frac{1{,}38}{14{,}32} = \mathbf{0,097\ (9,7\%)}$$

Ο συντελεστής μεταβλητότητας (CV) είναι της τάξης του 10%, πολύ περιορισμένη τιμή. Υπάρχει σημαντική ομοιογένεια: οι περιφέρειες δεν διαφέρουν σημαντικά ως προς το ποσοστό των μονογονεϊκών οικογενειών με τουλάχιστον ένα παιδί.

ΥΠΟΛΟΓΙΣΜΟΣ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ

ΑΣΚΗΣΗ 2



ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ [01]

Σύμφωνα με την ΕΛΣΤΑΤ, η αύξηση των αφίξεων ξένων τουριστών κατά σταθμό εισόδου για τα έτη 2006 και 2007, είναι η ακόλουθα:

Αφίξεις Ξένων Τουριστών κατά Σταθμό εισόδου, 2006-2007		
Σταθμοί εισόδου ξένων τουριστών	Έτος 2006	Έτος 2007
Αθήνα	3.698.953	3.872.156
Ζάκυνθο	466.821	475.146
Ηράκλειο	2.063.030	2.087.144
Θεσσαλονίκη	736.873	770.791
Καβάλα	74.365	80.222
Κέρκυρα	828.740	839.460
Κεφαλληνία	160.902	158.729
Κω	649.526	681.313
Μύκονος	99.960	103.086
Ρόδος	1.291.429	1.384.902
Σάμος	128.888	133.290
Θήρας	178.027	192.018
Σκιάθος	116.068	120.163
Χανιά	637.073	678.510
Λοιποί σταθμοί	378.472	424.292
ΣΥΝΟΛΟ	**11.509.127**	**12.001.222**

1. *Να Υπολογίσετε τον δείκτη Μεταβλητότητας για την μεταβλητή «μεταβολή αφίξεων 2007/2006».*
2. *Να σχεδιάσετε το θηκόγραμμα.*
3. *Ποια τα συμπεράσματα σας;*

Πηγή: Ελ.Στατ, Στατιστική Επετηρίδα της Ελλάδας, 2011

ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ [01]

Σύμφωνα με την ΕΛΣΤΑΤ, η αύξηση των αφίξεων ξένων τουριστών κατά σταθμό εισόδου για τα έτη 2006 και 2007, είναι η ακόλουθα:

Αφίξεις Ξένων Τουριστών κατά Σταθμό εισόδου, 2006-2007			
Σταθμοί εισόδου ξένων τουριστών	Έτος 2006	Έτος 2007	Μεταβολή 2007/2006
Αθήνα	3.698.953	3.872.156	4,68
Ζάκυνθο	466.821	475.146	1,78
Ηράκλειο	2.063.030	2.087.144	1,17
Θεσσαλονίκη	736.873	770.791	4,60
Καβάλα	74.365	80.222	7,88
Κέρκυρα	828.740	839.460	1,29
Κεφαλληνία	160.902	158.729	-1,35
Κω	649.526	681.313	4,89
Μύκονος	99.960	103.086	3,13
Ρόδος	1.291.429	1.384.902	7,24
Σάμος	128.888	133.290	3,42
Θήρας	178.027	192.018	7,86
Σκιάθος	116.068	120.163	3,53
Χανιά	637.073	678.510	6,50
Λοιποί σταθμοί	378.472	424.292	12,11
ΣΥΝΟΛΟ	**11.509.127**	**12.001.222**	**4,28**

Πηγή: Ελ.Στατ, Στατιστική Επετηρίδα της Ελλάδας, 2011

Έστω r: ποσοστό μεταβολής της μεταβλητής X μεταξύ 2 περιόδων:
t = τελική περίοδος
και
o = αρχική περίοδος

$$r = 100 \times \frac{(X_t - X_0)}{X_0}$$

Για Αθήνα:

$$r = 100 \times \frac{(3.872.156 - 3.698.953)}{3698953}$$

➔ r = 4,68

ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ [02]

i	Αφίξεις Ξένων Τουριστών κατά Σταθμό εισόδου, 2006-2007		$(r_i - \bar{r})^2$
	Σταθμοί εισόδου ξένων τουριστών	Μεταβολή 2007/2006 **r_i**	
1	Κεφαλληνία	-1,35	
2	Ηράκλειο	1,17	
3	Κέρκυρα	1,29	
4	Ζάκυνθο	1,78	
5	Μύκονος	3,13	
6	Σάμος	3,42	
7	Σκιάθος	3,53	
8	Θεσσαλονίκη	4,60	
9	Αθήνα	4,68	
10	Κω	4,89	
11	Χανιά	6,50	
12	Ρόδος	7,24	
13	Θήρας	7,86	
14	Καβάλα	7,88	
15	Λοιποί σταθμοί	12,11	
	Άθροισμα	**68,73**	**159,23**

Εφόσον θα πρέπει να σχεδιάσουμε το θηκόγραμμα, είναι προτιμότερο να παράγουμε εκ των προτέρων τον πίνακα με τα δεδομένα σε αύξουσα σειρά, έτσι ώστε να δουλέψουμε στη συνέχεια με ένα και μοναδικό πίνακα.

Επίσης, ο υπολογισμός του συντελεστή μεταβλητότητας βασίζεται στα δεδομένα των 15 σταθμών.

➔ *Εννοείται ότι, δεν πρέπει να λαμβάνουμε υπόψη στους υπολογισμούς τη τιμή της μεταβλητής για την Χώρα (τελευταία γραμμή του αρχικού πίνακα).*

ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ [03]

i	Αφίξεις Ξένων Τουριστών κατά Σταθμό εισόδου, 2006-2007		$(r_i - \bar{r})^2$
	Σταθμοί εισόδου ξένων τουριστών	Μεταβολή 2007/2006 **r_i**	
1	Κεφαλληνία	-1,35	35,19
2	Ηράκλειο	1,17	11,64
3	Κέρκυρα	1,29	10,84
4	Ζάκυνθο	1,78	7,85
5	Μύκονος	3,13	2,11
6	Σάμος	3,42	1,35
7	Σκιάθος	3,53	1,11
8	Θεσσαλονίκη	4,60	0,00
9	Αθήνα	4,68	0,01
10	Κω	4,89	0,09
11	Χανιά	6,50	3,68
12	Ρόδος	7,24	7,06
13	Θήρας	7,86	10,75
14	Καβάλα	7,88	10,88
15	Λοιποί σταθμοί	12,11	56,67
	Άθροισμα	**68,73**	**159,23**

$$\bar{r} = \frac{\sum_{i=1}^{15} r_i}{n} = \frac{68{,}73}{15} = 4{,}58$$

$$V[r] = s^2 = \frac{\sum_{i=1}^{n} (r_i - \bar{r})^2}{n-1} = \frac{159{,}23}{14} = 11{,}37$$

$$s = \sqrt{V(r)} = \sqrt{11{,}37} = 3{,}37$$

$$CV = \frac{s}{\bar{\bar{X}}} = \frac{3{,}37}{4{,}58} = 0{,}736\,(73{,}6\%)$$

Υψηλή ανομοιογένεια και σημαντικό εύρος τιμών:
e = max-min = 13,46

ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ (04)

Σύμφωνα με τα δεδομένα, ο αριθμός αφίξεων σε επίπεδο χώρας, αυξήθηκε από 11.509.127 σε 12.001.222, δηλαδή μια αύξηση κατά +4,28% (βλέπε 1º πίνακα).

Όμως, βρήκαμε ότι, ο αριθμητικός μέσος του ποσοστού μεταβολής (μέση τιμή) ισούται με +4,58%.

Το γεγονός ότι, οι δύο τιμές διαφέρουν είναι λογικό και αναμενόμενο εφόσον:

(α) οι δύο τιμές βασίζονται σε διαφορετικό τρόπο υπολογισμού,

(β) η μέση τιμή του ποσοστού μεταβολής επηρεάζεται σημαντικά από τη διακύμανση και τις ακραίες τιμές.

Έστω r_E = ποσοστό μεταβολής σε επίπεδο χώρας

A_{2006} = Σύνολο αφίξεων 2006

A_{2007} = Σύνολο αφίξεων 2007

$$\boxed{r_E = \frac{A_{2007} - A_{2006}}{A_{2006}} \times 100}$$

Έστω r_i = ποσοστό μεταβολής στο σταθμό i (i=1,…15)

Για κάθε σταθμό:

$A_{i,2006}$ = Αφίξεις 2006

$A_{i,2007}$ = Αφίξεις 2007

$$r_i = \frac{A_{i,2007} - A_{i,2006}}{A_{i,2006}} \times 100$$

$$\boxed{\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i}$$

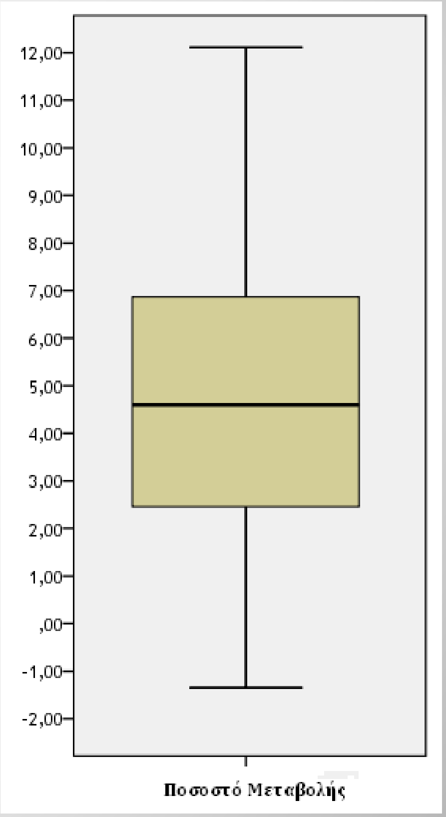
ΑΥΞΗΣΗ ΤΩΝ ΑΦΙΞΕΩΝ ΞΕΝΩΝ ΤΟΥΡΙΣΤΩΝ [05]

i	Αφίξεις Ξένων Τουριστών κατά Σταθμό εισόδου, 2006-2007		
	Σταθμοί εισόδου ξένων τουριστών	Μεταβολή 2007/2006 r_i	
1	Κεφαλληνία	-1,35	
2	Ηράκλειο	1,17	
3	Κέρκυρα	1,29	
4	Ζάκυνθο	1,78	$Q_1 = 2{,}46$
5	Μύκονος	3,13	
6	Σάμος	3,42	
7	Σκιάθος	3,53	
8	Θεσσαλονίκη	**4,60**	Διάμεσος
9	Αθήνα	4,68	
10	Κω	4,89	
11	Χανιά	6,50	$Q_3 = 6{,}87$
12	Ρόδος	7,24	
13	Θήρας	7,86	
14	Καβάλα	7,88	
15	Λοιποί σταθμοί	12,11	

$D_F = Q_3 - Q_1 = 4{,}42$

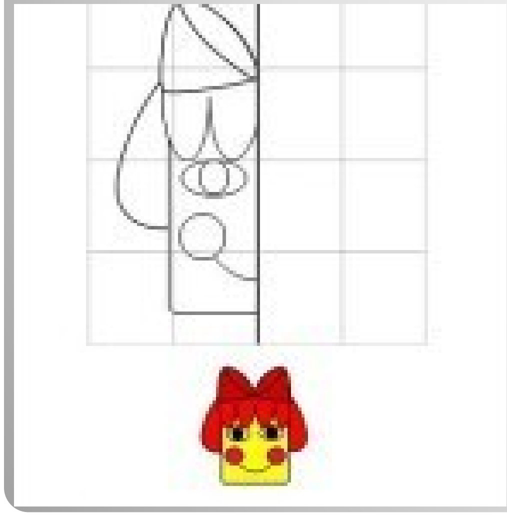
$W_1 = Q_1 - 1{,}5D_F = -4{,}17 < \min$

$W_3 = Q_3 + 1{,}5D_F = +13{,}49 > \max$



ΥΠΟΛΟΓΙΣΜΟΣ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ

ΑΣΚΗΣΗ 3



ΜΗΚΟΣ ΤΩΝ ΑΚΤΩΝ ΤΩΝ ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [01]

Ο πίνακας που ακολουθεί, μας δίνει την κατανομή των 90 μεγαλύτερων νησιών της Ελλάδας (εκτός Κρήτης) με βάση το μήκος των ακτών τους.

i	Μήκος ακτών (σε χλμ)	Αριθμός νησιών
1	10 – 30	19
2	30 – 50	27
3	50 – 100	18
4	100 – 150	16
5	150 – 200	4
6	200 - 420	6
ΣΥΝΟΛΟ		90

Πηγή ΕΛΣΤΑΤ, Στατιστική Επετηρίδα, 2011

1. ***Να υπολογίσετε τα βασικά μέτρα μεταβλητότητας.***
2. ***Συμπεράσματα***

Γνωρίζουμε ότι **Q_1 = 32,6 & Q_3 = 110,9 ενώ Q_2 = 49,3** (βλέπε Διάλεξη 04)

ΜΗΚΟΣ ΤΩΝ ΑΚΤΩΝ ΤΩΝ ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [02]

Μέτρα μεταβλητότητας:

a) Εύρος **= max – min** = 420 – 10 = **410 χλμ**

b) Σχετικό Ενδοτεταρτομοριακό διάστημα $ED = \frac{Q_3 - Q_1}{Q_2} = \frac{78{,}3}{49{,}3} = 1{,}6$

c) Διάστημα του Kelley: **DK = D_9 – D_1**

Για τον υπολογισμό των δεκατημορίων, χρησιμοποιούμε τον ακόλουθο τύπο (παρόμοιο με τον τύπο για τα τεταρτημόρια):

$$D_p = L_m + \frac{w}{n_m}(p.n - N_{m-1})$$

ή

$$D_p = L_m + \frac{w}{f_m}(p - F_{m-1})$$

D_1: p = 0,10
D_9: p = 0,90

ΜΗΚΟΣ ΤΩΝ ΑΚΤΩΝ ΤΩΝ ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Μέτρα μεταβλητότητας:

D_1:
Η θέση δίνεται από (0,1 x 90) = 9ο
➔ **D_1** βρίσκεται στην 1η τάξη ➔ **m= 1**

$$D_1 = L_1 + \frac{w}{n_1}(0{,}1\times n - N_0) = 10 + \frac{20}{19}\times(9-0) = 19{,}5$$

D_1:
Η θέση δίνεται από (0,9 x 90) = 81
➔ **D_1** βρίσκεται στην 5η τάξη ➔ **m= 5**

$$D_9 = L5 + \frac{w}{n_5}(0{,}9\times n - N_4) = 150 + \frac{50}{4}\times(81-80) = 162{,}5$$

Κατά συνέπεια το Διάστημα του Kelley = D9 – D1 = **143** σχετικά μικρό σε σχέση με το εύρος: 80% των νησιών βρίσκεται σε ένας εύρος 143 μονάδων ενώ το 100% των νησιών βρίσκεται στο συνολικό εύρος 410 μονάδων (μέγιστη – ελάχιστη).

(d) Αναλόγια μεταξύ 9ου Δεκατημόριου και 1ου Δεκατημόριου (Dispersion Ratio)

$$DR = \frac{D_9}{D_1} = \frac{162{,}5}{19{,}5} = 8{,}3$$

Το μήκος των ακτών των 10% μεγαλυτέρων νησιών (D9) είναι 8,3 φόρες μεγαλύτερο από το 10% μικρότερων νησιών (D1). Υπάρχει επομένως σημαντική διαφοροποίηση μεταξύ των μικρών και μεγάλων νησιών.

ΜΗΚΟΣ ΤΩΝ ΑΚΤΩΝ ΤΩΝ ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [04]

Μέτρα μεταβλητότητας:

(e) Δείκτης Μεταβλητότητας $CV = \frac{s}{\overline{X}}$

Για να αποφύγουμε ενδεχόμενα λάθη, είναι απαραίτητο να παρουσιάζουμε τους υπολογισμούς σε **<u>μορφή πίνακα</u>**

i	Μήκος ακτών (σε χλμ)	Κέντρο Τάξης X_i	Αριθμός νησιών n_i	$n_i X_i$	$n_i(X_i - \overline{X})^2$
1	10 – 30	20	19	380	73036
2	30 – 50	40	27	1080	47628
3	50 – 100	75	18	1350	882
4	100 – 150	125	16	2000	29584
5	150 – 200	175	4	700	34596
6	200 – 420	310	6	1860	311904
ΣΥΝΟΛΟ			90	**7370**	**497630**

$$\overline{X} = \frac{\sum_{i=1}^{6} n_i X_i}{n} = \frac{7370}{90} = 81{,}9 \Rightarrow \overline{X} = 82$$

$$s^2 = \frac{\sum_{i=1}^{6} n_i (X_i - \overline{X})^2}{n} = \frac{497630}{90} = 5529{,}2$$

$$s = \sqrt{5529{,}2} = 74{,}4 \Rightarrow S = 74$$

$$CV = \frac{s}{\overline{X}} = \frac{74}{82} = 0{,}902 \ (90\%)$$

Όπως ήταν αναμενόμενο, τα νησιά της Ελλάδας παρουσιάζουν σημαντική ανομοιογένεια: ορισμένα νησιά – λόγω του μεγέθους τους – διαθέτουν σημαντικό μήκος ακτών σε αντίθεση με άλλα. Αυτό το αποτέλεσμα αναδεικνύει μια μορφή χωρικής ανισότητας.

Χωρικές ανισότητες:

Απλός συντελεστής μεταβλητότητας (CV)
έναντι
Σταθμισμένος συντελεστής μεταβλητότητας (wCV)

ΣΥΝΤΕΛΕΣΤΕΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ: ΧΩΡΙΚΕΣ ΑΝΙΣΟΤΗΤΕΣ

Ο ***απλός συντελεστής μεταβλητότητας*** αποτελεί το 1ο δείκτη μέτρησης των χωρικών ανισοτήτων, όμως δίνει σε κάθε χωρική ενότητα, το ίδιο βάρος (1/n).

$$CV = \frac{s}{\overline{X}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\overline{X}} = \frac{\sqrt{\sum_{i=1}^{n}\frac{1}{n}(X_i - \overline{X})^2}}{\overline{X}}$$

Έχουμε δεδομένα για κάθε χωρική ενότητα

Δεν λαμβάνει υπόψη ότι, συχνά οι εξεταζόμενες χωρικές ενότητες <u>δεν έχουν το ίδιο μέγεθος</u> και επομένως την ίδια βαρύτητα. Σε αυτή την περίπτωση, είναι προτιμότερο να χρησιμοποιούμε εναλλακτικό συντελεστή μεταβλητότητας (σταθμισμένο).

$$wCV = \frac{\sqrt{\sum_{i=1}^{n}w_i(X_i - \overline{X})^2}}{\overline{X}}$$

Όπου το βάρος κάθε χωρικής ενότητας δίνεται από:

$$w_i = \frac{P_i}{P_\bullet}$$

P_i = Πληθυσμός της χωρικής ενότητας i

$P_\bullet$ = Πληθυσμός όλων των χωρικών ενοτήτων

ΧΩΡΙΚΕΣ ΑΝΙΣΟΤΗΤΕΣ : ΠΑΡΑΔΕΙΓΜΑ

Οι ***δαπάνες στην έρευνα*** στα Πανεπιστήμια (ως % του ΑΕΠ) για τις **8 χώρες της Νότιας Ευρώπης** δίνονται στον ακόλουθο πίνακα.

i	ΧΩΡΕΣ	Πληθυσμός (σε χιλιάδες)	w_i	Δαπάνες στην έρευνα
1	Βουλγαρία	7.781	0,051	0,05
2	Ελλάδα	11.056	0,072	0,27
3	Ισπανία	42.922	0,281	0,31
4	Ιταλία	57.685	0,378	0,36
5	Κύπρος	728	0,005	0,13
6	Μάλτα	401	0,003	0,16
7	Πορτογαλία	10.484	0,069	0,27
8	Ρουμανία	21.452	0,141	0,04
	Σύνολο	152.509	1,000	

Μεταβλητή προς διερεύνηση

$$w_3 = P_3 / P_{\bullet} = 49.922 / 152.509 = 0{,}281(28\%)$$

$$w_6 = P_6 / P_{\bullet} = 401 / 152.509 = 0{,}003(0{,}3\%)$$

Η Νότια Ευρώπη αποτελείται από 8 χώρες. Δεν πρόκειται για δείγμα

Πρέπει να υπολογίσετε τους δύο συντελεστές μεταβλητότητας για την μεταβλητή «Δαπάνες στην έρευνα». Ποια τα συμπεράσματα σας;

ΧΩΡΙΚΕΣ ΑΝΙΣΟΤΗΤΕΣ : ΠΑΡΑΔΕΙΓΜΑ

Προετοιμασία του Πίνακα Υπολογισμού

i	ΧΩΡΕΣ	Πληθυσμός (σε χιλιάδες)	w_i	Δαπάνες στην έρευνα Xi	?	?
1	Βουλγαρία	7.781	0,051	0,05		
2	Ελλάδα	11.056	0,072	0,27		
3	Ισπανία	42.922	0,281	0,31		
4	Ιταλία	57.685	0,378	0,36		
5	Κύπρος	728	0,005	0,13		
6	Μάλτα	401	0,003	0,16		
7	Πορτογαλία	10.484	0,069	0,27		
8	Ρουμανία	21.452	0,141	0,04		
	Σύνολο	152.509	1,000	?	?	?

n = 8, wi = βάρος κάθε χώρας και η μεταβλητή X_i = Δαπάνες στην έρευνα

ΧΩΡΙΚΕΣ ΑΝΙΣΟΤΗΤΕΣ : ΠΑΡΑΔΕΙΓΜΑ

i	ΧΩΡΕΣ	Πληθυσμός (σε χιλιάδες)	w_i	Δαπάνες στην έρευνα	$(X_i - \bar{X})^2$	$w_i(X_i - \bar{X})^2$
1	Βουλγαρία	7.781	0,051	0,05	0,0225	0,0011
2	Ελλάδα	11.056	0,072	0,27	0,0049	0,0004
3	Ισπανία	42.922	0,281	0,31	0,0121	0,0034
4	Ιταλία	57.685	0,378	0,36	0,0256	0,0097
5	Κύπρος	728	0,005	0,13	0,0049	0,0000
6	Μάλτα	401	0,003	0,16	0,0016	0,0000
7	Πορτογαλία	10.484	0,069	0,27	0,0049	0,0003
8	Ρουμανία	21.452	0,141	0,04	0,0256	0,0036
	Σύνολο	152.509	1	**1,59**	**0,1021**	**0,0186**

Άθροισμά κάθε στήλη

Μέσος όρος =	**0,20**		
Διασπορά =	0,013		
Τυπική απόκλιση =	0,1130	CV =	0,568
Σταθμισμένη Διασπορά =	0,0186	wCV =	0,685
Σταθμισμένη τυπική απόκλιση =	0,1362		

2. Διερευνητική Ανάλυση

Μέτρα Θέσης

Μέτρα Διασποράς

Ασυμμετρία

2.δ. Διερευνητική Ανάλυση
Μέτρα Ασυμμετρίας

4. ΑΣΥΜΜΕΤΡΙΑ ΤΗΣ ΚΑΤΑΝΟΜΗΣ [1/4]

Τα μέτρα ασυμμετρίας μας επιτρέπουν να εξετάζουμε:

(α) κατά πόσο η κατανομή συχνοτήτων είναι ***συμμετρική ή ασύμμετρη*** και,

(β) όταν δεν είναι συμμετρική, να διερευνήσουμε εάν το μεγαλύτερο μέρος της κατανομής βρίσκεται προς τα δεξιά ή προς τα αριστερά.

Στόχος: να εξετάζουμε σε ποιο βαθμό, *η κατανομή προσεγγίζει την Κανονική Κατανομή*.

Αν η κατανομή του πληθυσμού που εξετάζουμε είναι συμμετρική, *τότε η μέση τιμή, η διάμεσος και η κορυφή της κατανομής συμπίπτουν*.

Δύο είναι οι βασικοί συντελεστές:

α_3 = ***Συντελεστής Λοξότητας*** (coefficient of Skewness)

α_4 = ***Συντελεστής Κύρτωσης*** (coefficient of Kurtosis) → Τύπος κορυφής

4. ΑΣΥΜΜΕΤΡΙΑ ΤΗΣ ΚΑΤΑΝΟΜΗΣ [2/4]

Οι συντελεστές αυτοί ορίζονται με τη βοήθεια των **<u>κεντρικών ροπών τάξης r</u>** που υπολογίζονται ως εξής:

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r \quad \text{ή ακόμα} \quad m_r = \frac{1}{n}\sum_{i=1}^{k} n_i\,(x_i - \overline{x})^r$$

Παρατηρούμε ότι, $m_1 = 0$ και $m_2 = s^2$

α_3 = ***Συντελεστής Λοξότητας*** (coefficient of Skewness) = $\alpha_3 = \dfrac{m_3}{s^3}$

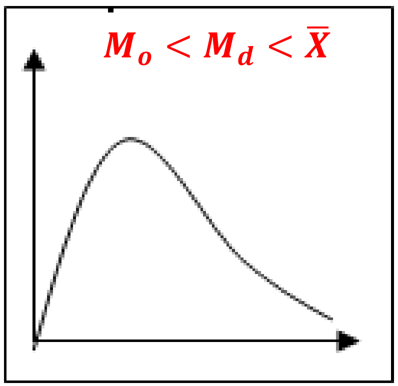
α_4 = ***Συντελεστής Κύρτωσης*** (coefficient of Kurtosis) = $\alpha_4 = \dfrac{m_4}{s^4}$

4. ΑΣΥΜΜΕΤΡΙΑ ΤΗΣ ΚΑΤΑΝΟΜΗΣ [3/4]

Ανάλυση:

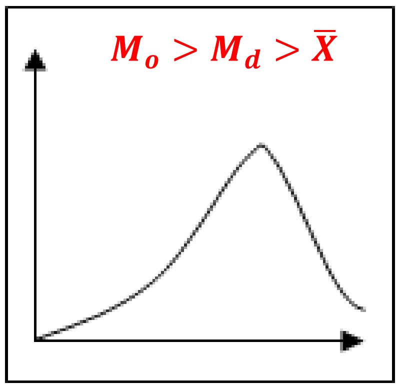
- $\alpha_3 > 0$, οι περισσότερες τιμές της μεταβλητής βρίσκονται **δεξιά** της κορυφής
- $\alpha_3 < 0$, οι περισσότερες τιμές της μεταβλητής βρίσκονται **αριστερά** της κορυφής

- $\alpha_4 < 3$, η κατανομή λέγεται πλατύκυρτη, δηλαδή δεν υπάρχει υψηλή κορυφή
- $\alpha_4 > 3$, η κατανομή λέγεται λεπτόκυρτη, υπάρχει πολύ υψηλή κορυφή



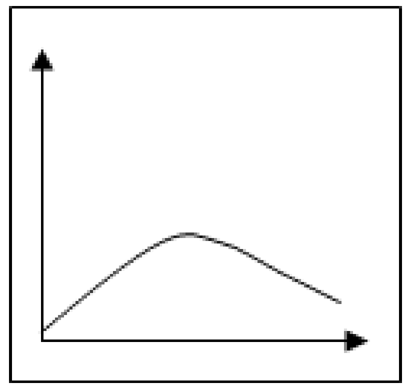
Θετική λοξότητα
$\alpha_3 > 0$

Οι περισσότερες παρατηρήσεις βρίσκονται δεξιά της κορυφής

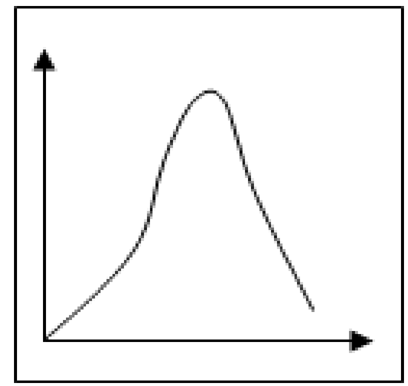


Αρνητική λοξότητα
$\alpha_3 < 0$

Οι περισσότερες παρατηρήσεις βρίσκονται αριστερά της κορυφής

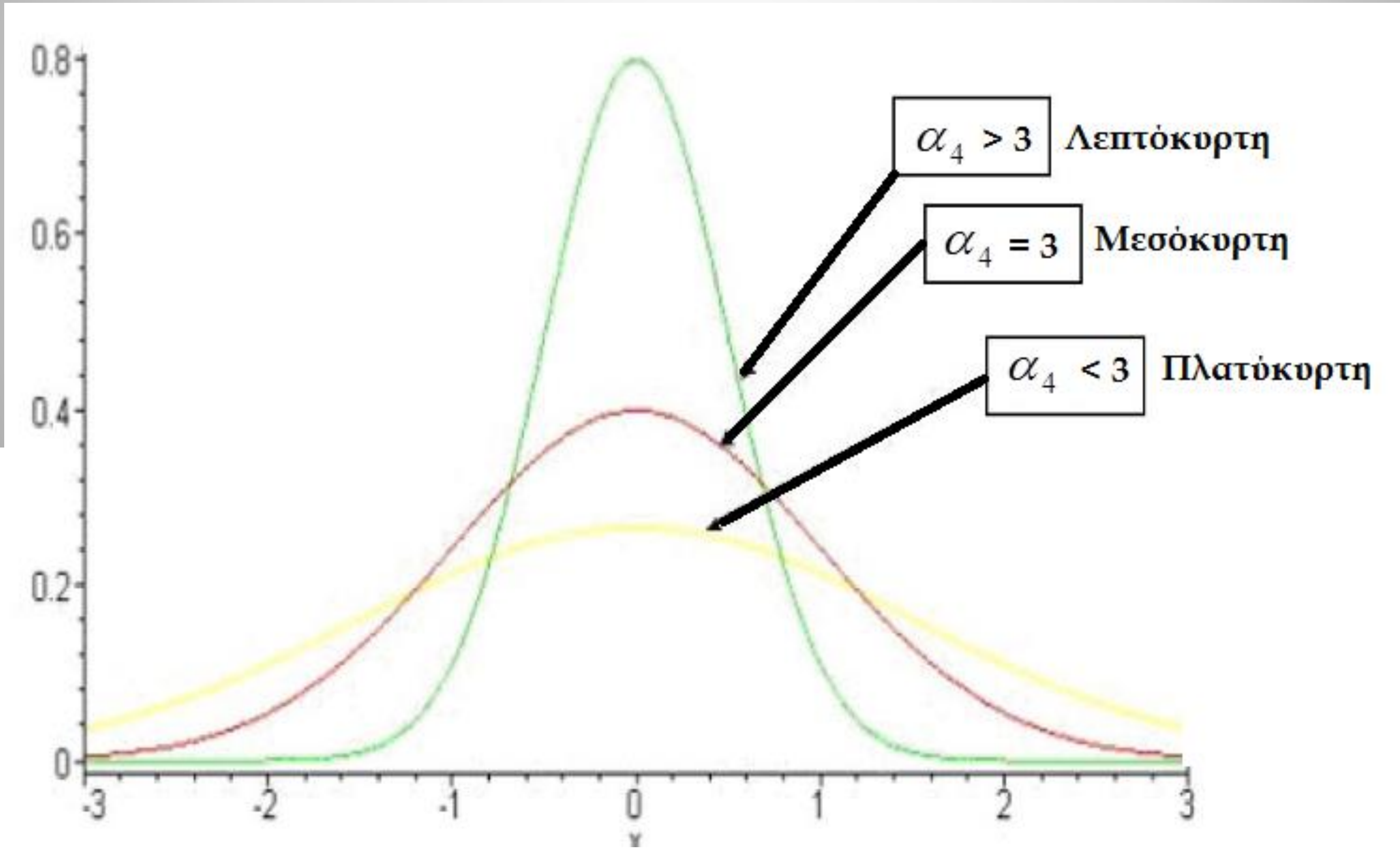


Πλατύκυρτη κατανομή
$\alpha_4 < 3$



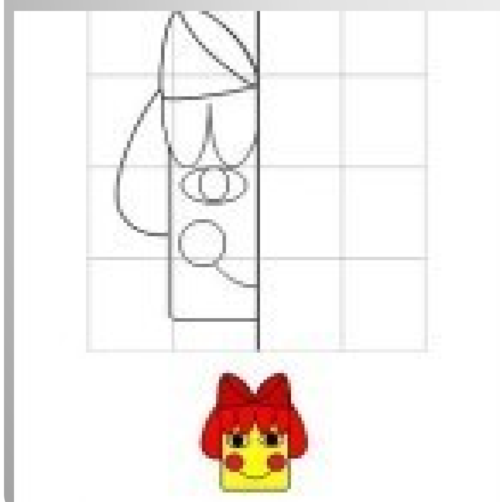
Λεπτόκυρτη κατανομή
$\alpha_4 > 3$

4. ΑΣΥΜΜΕΤΡΙΑ ΤΗΣ ΚΑΤΑΝΟΜΗΣ [4/4]



ΑΝΑΛΥΣΗ ΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ

ΑΣΚΗΣΗ 1



Ζητήσαμε σε 170 καταναλωτές να αξιολογούν την ποιότητα τριών διαφορετικών προϊόντων, χρησιμοποιώντας μια κλίμακα από 1 = ιδιαίτερα κακή ποιότητα έως 7 = ιδιαίτερα καλή ποιότητα.

Τα αποτελέσματα της έρευνας συνοψίζονται στον ακόλουθο πίνακα.

Σύνολο απαντήσεων ανά προϊόν			
Ποιότητα Προϊόντος	**Προϊόν 1**	**Προϊόν 2**	**Προϊόν 3**
1	5	18	25
2	10	22	43
3	40	30	37
4	70	36	32
5	35	30	26
6	8	20	5
7	2	14	2
Σύνολο απαντήσεων	170	170	170

1. *Να εξετάζετε σε ποιο βαθμό η κατανομή για το καθένα από τα 3 προϊόντα είναι συμμετρική ή ασύμμετρη;*
2. *Να γίνει γραφική αναπαράσταση.*
3. *Ποια τα συμπεράσματα σας;*

ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ ΓΙΑ ΤΟ 1^{ο} ΠΡΟΙΟΝ

Προετοιμασία του Πίνακα Υπολογισμού

$$m_r = \frac{1}{n} \sum_{i=1}^{k} n_i \,(x_i - \bar{x})^r$$

Ποιότητα Προϊόντος	Προϊόν 1				
	n_i	__?___	___?__	___?__	__?___
1	5				
2	10				
3	40				
4	70				
5	35				
6	8				
7	2				
Σύνολο	**170**	**?**	**?**	**?**	**?**

n = 170

$n_1 = 5$
$n_2 = 10$
…
$n_7 = 2$

$$\sum_{i=1}^{k=7} n_i = 170$$

ΥΠΟΛΟΓΙΣΜΟΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ ΓΙΑ ΤΟ 1ο ΠΡΟΙΟΝ

Ποιότητα Προϊόντος	Προϊόν 1				
	n_i	$n_i X_i$	$n_i(X_i-\bar{X})^2$	$n_i(X_i-\bar{X})^3$	$n_i(X_i-\bar{X})^4$
1	5	5	41,88	-121,20	350,78
2	10	20	35,88	-67,95	128,71
3	40	120	31,98	-28,59	25,56
4	70	280	0,78	0,08	0,01
5	35	175	42,80	47,34	52,35
6	8	48	35,48	74,71	157,34
7	2	14	19,29	59,92	186,11
Σύνολο	**170**	**662**	**208,09**	**-35,70**	**900,86**

$$\bar{X}=\frac{662}{170}=3{,}89 \quad s^2=\frac{208{,}09}{170}=1{,}224 \Rightarrow s=1{,}106 \Rightarrow CV=\frac{1{,}106}{3{,}89}=0{,}287\ (28{,}7\%)$$

Περιορισμένη μεταβλητότητα

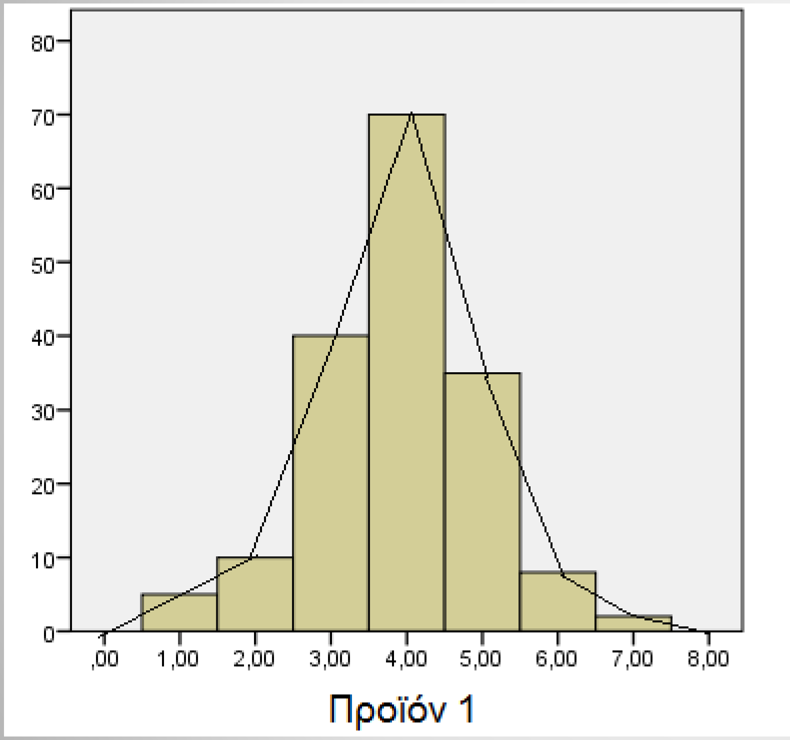
$$m_3=\frac{-35{,}70}{170}=-0{,}210 \quad s^3=1{,}106^3=1{,}353 \Rightarrow \alpha_3=\frac{m_3}{s^3}=\frac{-0{,}210}{1{,}353}=-0{,}155<0$$

Πολύ μικρή αρνητική λοξότητα (σχεδόν καμία)

$$m_4=\frac{900{,}86}{170}=5{,}299 \quad s^4=1{,}106^4=1{,}498 \Rightarrow \alpha_4=\frac{m_4}{s^4}=\frac{5{,}299}{1{,}498}=3{,}537>3$$

Λεπτόκυρτη

ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟ 1^{o} ΠΡΟΙΟΝ



Η κατανομή είναι πραγματικά Λεπτόκυρτη ενώ υπάρχει ελαχίστη ασυμμετρία.

Ποιότητα	n_i	f_i (%)
1	5	2,9%
2	10	5,9%
3	40	23,5%
4	70	41,2%
5	35	20,6%
6	8	4,7%
7	2	1,2%
Σύνολο	170	100,0%

ΑΝΑΛΥΣΗ ΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ ΓΙΑ ΤΟ 2^{o} ΚΑΙ 3^{o} ΠΡΟΙΟΝ

Όπως και για το 1^{o} προϊόν, η ανάλυση της ασυμμετρίας βασίζεται στον υπολογισμό των συντελεστών α_3 και α_4. Σύμφωνα με τον ορισμό αυτών των συντελεστών, έχουμε τα ακόλουθα αποτελέσματα:

	2^{o} Προϊόν	**3^{o} Προϊόν**
$\sum_{i=1}^{7} n_i X_i$	664	524
$\sum_{i=1}^{7} n_i (X_i - \bar{X})^2$	514,49	354,85
$\sum_{i=1}^{7} n_i (X_i - \bar{X})^3$	21,41	172,24
$\sum_{i=1}^{7} n_i (X_i - \bar{X})^4$	3304,70	1736,82

Με βάση τα ενδιάμεσα αποτελέσματα όπως προκύπτουν στο Πίνακα,

ποια τα συμπεράσματα σχετικά με την ασυμμετρία των απαντήσεων των καταναλωτών για το 2^{o} και 3^{o} προϊόν;

(*) n=170 > 120 μεγάλο δείγμα

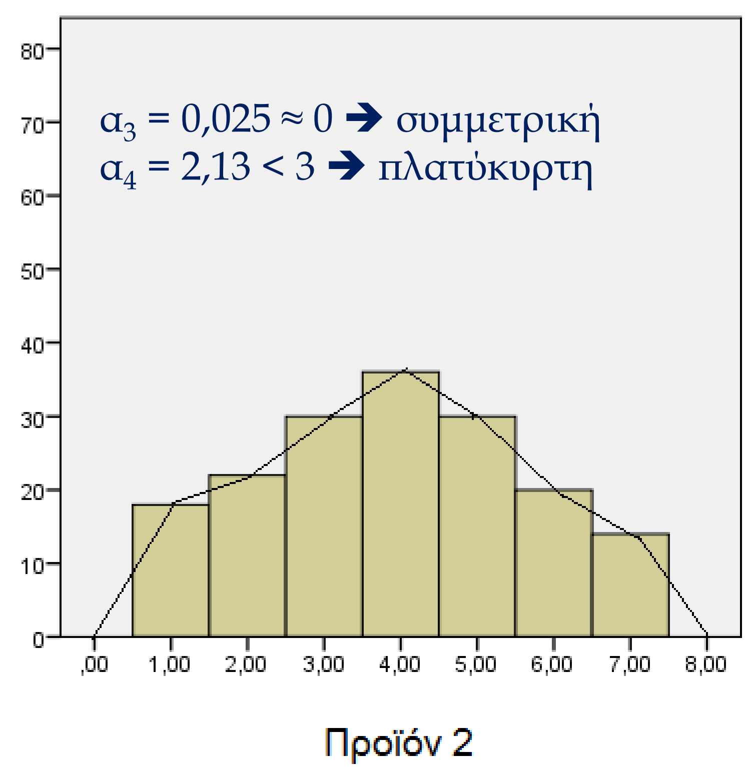
ΑΝΑΛΥΣΗ ΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ ΓΙΑ ΤΟ 2^o ΚΑΙ 3^o ΠΡΟΙΟΝ

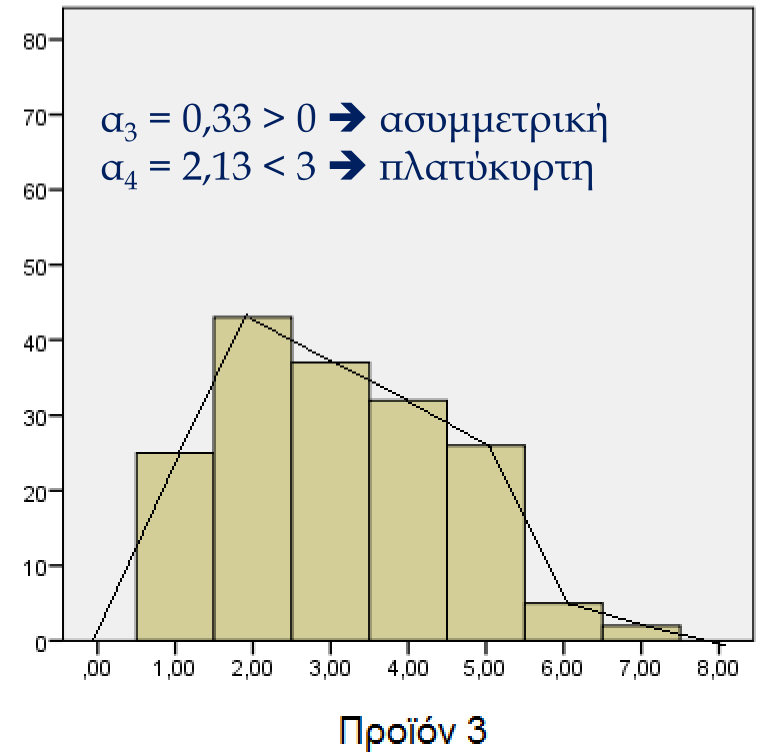
	2^o Προϊόν	**3^o Προϊόν**
$\sum_{i=1}^{7} n_i X_i$	664	524
$\sum_{i=1}^{7} n_i (X_i - \overline{X})^2$	514,49	354,85
$\sum_{i=1}^{7} n_i (X_i - \overline{X})^3$	21,41	172,24
$\sum_{i=1}^{7} n_i (X_i - \overline{X})^4$	3304,70	1736,82
$\overline{X}$	664/170 = **3,9**	524/170=**3,1**
s^2 (*)	514,49/170 =3,02	354,85/170=2,09
s	**1,74**	**1,45**
m_3	21,41/170=0,13	172,24/170=1,01
a_3	$0,13/(1,74)^3$= **0,025**	$1,01/(1,45)^3$= **0,33**
m_4	3304,7/170=19,44	1736,82/170=10,22
a_4	$19,44/(3,02)^2$= **2,13**	$10,22/(2,09)^2$= **2,34**

Συμπεράσματα:

(*) n=170 > 120 μεγάλο δείγμα

ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟ $2^{ο}$ ΚΑΙ $3^{ο}$ ΠΡΟΙΟΝ





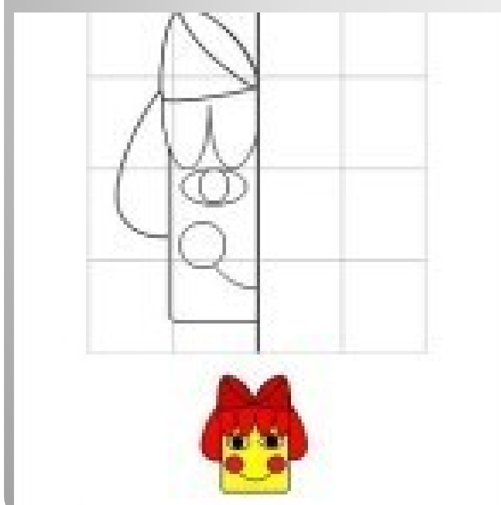
Προσοχή: τα διαγράμματα πρέπει να έχουν την ίδια κλίμακα στον κάθετο άξονα για να γίνει σωστή σύγκριση

ΠΑΡΑΡΤΗΜΑ:

ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΛΥΣΕΙΣ

ΠΑΡΑΔΕΙΓΜΑ

1



ΠΑΡΑΔΕΙΓΜΑ 1

Ο πίνακας που ακολουθεί, μας δίνει την κατανομή των ανέργων της Ελλάδας κατά τάξη ηλικίας για τα έτη 2011 και 2011 (σε χιλιάδες ατόμων).

i	Ηλικία	Αριθμός ανέργων	
		2011	***2012***
1	15 – 29	321,3	395,5
2	30 – 44	356,0	506,6
3	45 – 64	197,6	299,0
4	65+	2,0	2,7
ΣΥΝΟΛΟ		876,9	1203,8

1. Να βρείτε τη διάμεση ηλικία των ανέργων για τα δύο έτη.
2. Να βρείτε τη μέση ηλικία των ανέργων για τα δύο έτη.
3. Να υπολογίσετε τον συντελεστή μεταβλητότητα για τα δύο έτη.

ΠΑΡΑΔΕΙΓΜΑ 1

1. Υπολογισμός Διάμεσης ηλικίας

i	Ηλικία	Αριθμός ανέργων			
		2011		2012	
		ni	Ni	ni	Ni
1	15 – 29	321,3	321,3	395,5	395,5
2	30 – 44	356,0	677,3	506,6	902,1
3	45 – 64	197,6	874,9	299,0	1201,1
4	65+	2,0	876,9	2,7	1203,8
ΣΥΝΟΛΟ		876,9		1203,8	

Για να υπολογίσουμε την Q_2, πρέπει να βρούμε <u>την τάξη που περιλαμβάνει την διάμεσο</u>.
Για το 2011: n/2 (0,5xn) = 438,45 ➔ m =2
Για το 2012: n/2 (0,5xn) = 601,9 ➔ m =2

$$\mathbf{Q_2 = L_m + \frac{w}{n_m}(0{,}5.n - N_{m-1}) = L_2 + \frac{w}{n_2}(0{,}5.n - N_1)}$$

w=15, δεδομένου ότι πρόκειται για ηλικίες από 30 έως και 44, δηλαδή 15 έτη.

2011: $Q_2 = 30 + \frac{15}{356{,}0} \times (438{,}45 - 321{,}3) = 34{,}94 \Rightarrow 35$

2012: $Q_2 = 30 + \frac{15}{506{,}6} \times (601{,}9 - 395{,}5) = 36{,}11 \Rightarrow 36$

ΠΑΡΑΔΕΙΓΜΑ 1

1. Υπολογισμός Διάμεσης ηλικίας

Η διάμεση ηλικία αυξήθηκε κατά ένα έτος (από 35 στα 36). Το αποτέλεσμα αυτό είναι απόλυτα λογικό δεδομένου ότι, η σχετική συχνότητα της 2ης τάξης ηλικίας (30-44) αυξήθηκε από 40,6% στα 42,1% (βλέπε παρακάτω πίνακα). Το ποσοστό των νεαρών ανέργων (15-29) μειώθηκε παρά το γεγονός ότι, σε απόλυτες τιμές, ο αριθμός τους αυξήθηκε.

Ηλικία	2011		2012
	fi		fi
15 – 29	36,6		32,9
30 – 44	40,6		42,1
45 – 64	22,5		24,8
65+	0,2		0,2
	100,0		100,0

Ο αριθμός των νεαρών ανέργων αυξήθηκε κατά 23% (από 321,3 σε 395,5 χιλιάδες ατόμων). ενώ ο αριθμός ανέργων ηλικίας 30-44 αυξήθηκε κατά 42% (από 356 σε 506,6 χιλιάδες ατόμων).

ΠΑΡΑΔΕΙΓΜΑ 1

2. Υπολογισμός Μέσης ηλικίας

		Αριθμός ανέργων			
i	Κέντρο τάξης ηλικίας	2011		2012	
		ni	ni.Xi	ni	ni.Xi
1	22,5	321,3	7229,25	395,5	8898,75
2	37,5	356,0	13350,00	506,6	18997,50
3	55,0	197,6	10868,00	299,0	16445,00
4	70,0	2,0	140,00	2,7	189,00
ΣΥΝΟΛΟ		876,9	31587,25	1203,8	44530,25

Κέντρο 1ης τάξης = (15+30)/2 = 22,5

Κέντρο 4ης τάξης = (65+75)/2 = 70

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{4} n_i \times X_i$$

Για το 2011: $\bar{X} = \frac{1}{876,9} \times 31587,25 = 36$

Για το 2012: $\bar{X} = \frac{1}{1203,8} \times 44530,25 = 37$

Όπως και στην περίπτωση της διάμεσου, η μέση ηλικία αυξήθηκε κατά ένα έτος.
Και για τα δύο χρόνια 1011 και 2012, η διάμεσος διαφέρει από τη μέση τιμή.

ΠΑΡΑΔΕΙΓΜΑ 1

3. Υπολογισμός Συντελεστή Μεταβλητότητας

		Αριθμός ανέργων			
i	Κέντρο τάξης ηλικίας	2011		2012	
		ni	ni.(Xi-X)²	ni	ni.(Xi-X)²
1	22,5	321,3	58556,93	395,5	83153,88
2	37,5	356,0	801,00	506,6	126,65
3	55,0	197,6	71333,60	299,0	96876,00
4	70,0	2,0	2312,00	2,7	2940,30
ΣΥΝΟΛΟ		876,9	133003,53	1203,8	183096,83

$$V[X] = s^2 = \frac{1}{n}\sum_{i=1}^{4} n_i \times (X_i - \overline{X})^2$$

Για το 2011: V[X] = 151,675 ➔ s = 12,316 ➔ CV = 12,3/36 = 0,342 (34,2%)

Για το 2012: V[X] = 152,099 ➔ s = 12,333 ➔ CV = 12,3/37= 0,333 (33,3%)

Ο συντελεστής μεταβλητότητας δεν παρουσιάζει σημαντική μεταβολή και παραμένει σχετικά περιορισμένο: με βάση τις ηλικίες, δεν παρατηρείται πολύ υψηλές ανισότητες.