Lecture 18

Ο Κανόνας του Chebychev

Αυτός ο κανόνας ισχύει για όλες τις συλλογές δεδομένων ανεξάρτητα αν αυτές είναι συμμετρικές.

“ Σε απόσταση κ τυπικών αποκλίσεων από τον μέσο όρο υπάρχουν τουλάχιστον

$$\left(1-\frac{1}{\kappa^2}\right)$$

ποσοστό τιμών του δείγματος”.

Παράδειγμα:

1). Αν κ=2

$$\left(\bar{X}-2S,\bar{X}+2S\right)\Rightarrow 1-\frac{1}{4}=\frac{3}{4}\rightarrow 75\%$$

κ=3

$$\left(\bar{X}-3S,\bar{X}+3S\right)\Rightarrow 1-\frac{1}{9}=\frac{8}{9}\rightarrow 89\%$$

κ=4

$$\left(\bar{X}-4S,\bar{X}+4S\right)\Rightarrow 1-\frac{1}{16}=\frac{5}{16}\rightarrow 94\%$$

2). Για μια συλλογή δεδομένων έχω ότι : δειγματικός μέσος όρος =11 και S=2. Ας υποθέσουμε ότι η κατανομή είναι άγνωστη. Μεταξύ ποιών ορίων υπάρχει το 85% των τιμών του δείγματος;

Έχω ότι:

$$1-\frac{1}{\kappa^2}=\frac{85}{100}\Rightarrow \kappa=2.85$$

Άρα

$$\left(\bar{X}-58S,\bar{X}+2.58S\right)=\left(11-2.58.(2),11+2.58.(2)\right)=\left(5.84,16.16\right)$$

Εύρος Τεταρτημορίων

IQR=Q3-Q1

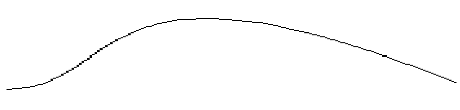
Πως περιγράφουμε το σχήμα μιας κατανομής;
Από την στατιστική συνάρτηση Λοξότητα η οποία μας δείχνει πόσο απέχει μια συλλογή δεδομένων από την συμμετρία.

Συντελεστής Λοξότητας:

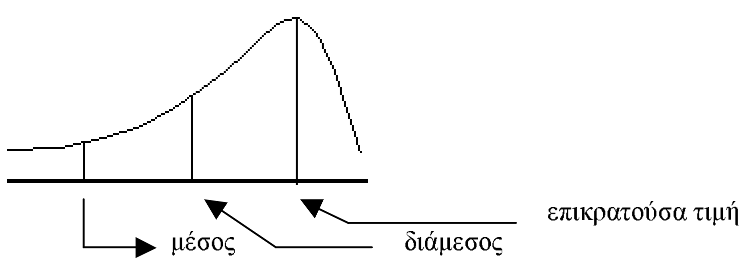
$$S_\kappa = 3\left(\bar{X} - \textit{διάμεσος}\right)\Big/ S$$

Αν
$S_\kappa=0 \Rightarrow$ συμμετρική κατανομή
$S_\kappa<0 \Rightarrow$ αρνητικά λοξή κατανομή
$S_\kappa>0 \Rightarrow$ θετικά λοξή κατανομή

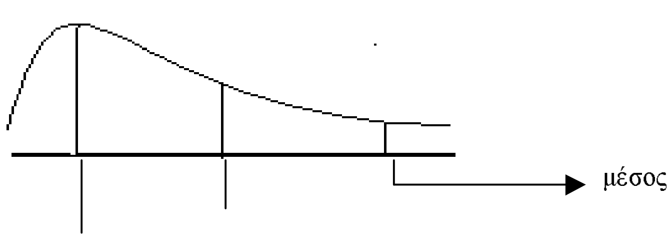
α) συμμετρική : διάμεσος = μέσο= επικρατούσα τιμή



β) αρνητικά λοξή ή λοξή προς τα αριστερά



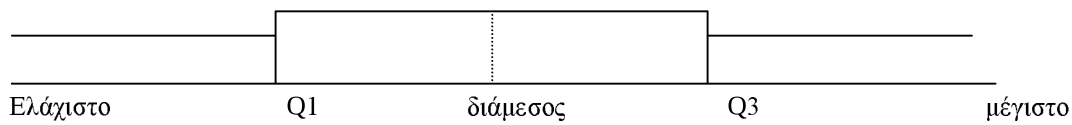
γ) θετικά λοξή ή λοξή προς τα δεξιά



διάμεσος
επικρατούσα τιμή

<u>Box-Plots:</u>

Τα Box-Plots είναι διαγράμματα τα οποία κάνουν χρήση του τεταρτημόριου για να δώσουν μια γραφική παράσταση μιας συλλογής δεδομένων.



Σημεία τα οποία μπορεί να είναι μεγαλύτερα από Q3+I.S*(IQR) ή μικρότερα από Q1-I.S*(IQR) ονομάζονται outliers.

<u>Παράδειγμα :</u>

Η ηλικία νέων διευθυντικών στελεχών:

35	37	37	39	40	40
41	41	43	43	43	43
44	44	44	44	44	45
45	46	46	46	46	48

$$Q_1 = \frac{25}{100} x 24 = 6$$

$$Q_1 = \frac{40+41}{2} = 40.5$$

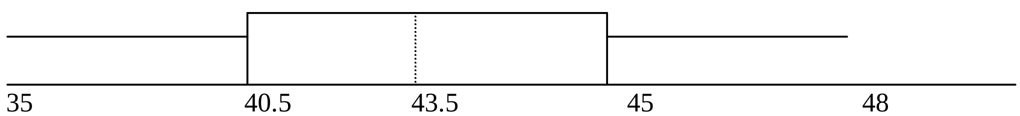
$$Q_3 = \frac{75}{100} x 24 = 18$$

$$Q_3 = \frac{45+45}{2} = 45$$

$$\Delta\iota\acute{\alpha}\mu\varepsilon\sigma o\varsigma = \frac{43+44}{2} = 43.5$$

$$\min = 35$$

$$\max = 48$$



<u>Ο μέσος όρος σαν τυχαία μεταβλητή</u>

Έστω X1, X2 ~f(X) ανεξάρτητες με :

$$\mathrm{EX}_1 = \mathrm{EX}_2 = \mu$$

$$\kappa\alpha\iota$$

$$Var\mathrm{X}_1 = Var\mathrm{X}_2 = \sigma^2$$

Ο μέσος όρος είναι:

$$\left(\bar{\mathrm{X}} = \frac{\mathrm{X}_1 + \mathrm{X}_2}{2}\right)$$

Είναι μια τυχαία μεταβλητή με :

$$\mathrm{E}\left(\bar{\mathrm{X}}\right) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

και

$$Var\mathrm{X} = \frac{\sigma^2}{2}$$

<u>Γενικά</u> αν X1,…, X2 ανεξάρτητες και ισόνομες τότε

$$\mathrm{E}\left(\bar{\mathrm{X}}\right) = \mathrm{E}(\mathrm{X}_1) = \mu$$

και

$$Var\left(\bar{\mathrm{X}}\right) = \frac{\sigma^2}{\eta}$$

Επιπλέον αν

$$\mathrm{X}_1,...,\mathrm{X}_\eta \sim \mathrm{N}(\mu,\sigma^2)$$

τότε ισχύει ότι:

$$\bar{\mathrm{X}} \sim \mathrm{N}\left(\mu,\frac{\sigma^2}{\eta}\right)$$

γιατί ο γραμμικός συνδιασμός είναι κανονική.