MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

07-05-2020

Άσκηση 1

Υπολογίστε τη διάμεσο, το πρώτο και το τρίτο τεταρτημόριο για το σύνολο δεδομένων:

$$\{-2,\ 4,\ 8,\ -4,\ 6,\ -3,\ 5,\ 8,\ 10,\ 8,\ -4,\ 6,\ -10\}$$

Επιπλέον χρησιμοποιήστε το κριτήριο 1.5x(IQR) για να εξετάσετε εάν υπάρχουν ακραίες τιμές και σχεδιάστε πρόχειρα το διάγραμμα Box-and-Whisker.

Q_1, M, Q_3

$\{-10, -4, -4, -3, -2, 4, 5, 6, 6, 8, 8, 8, 10\}$ $\quad N = 13$

$\frac{N+1}{2} = 7$

if $P_1 \cdot (N-1) = 3.2$

$Q_1 \in [x_4, x_5]$

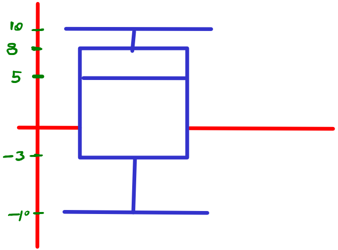
$Q_1 = x_4 + 0.2 * (x_5 - x_4)$

$P_1 \cdot (N-1) = 0.25 * 12 = 3 \in \mathbb{Z}$ $\quad Q_1 = x_4$

$P_3 \cdot (N-1) = 0.75 * 12 = 9 \in \mathbb{Z}$ $\quad Q_3 = x_{10}$

$IQR = Q_3 - Q_1 = 8 - (-3) = 11$

$[Q_1 - 1.5 * IQR,\ Q_3 + 1.5 * IQR] = [-3 - 1.5 * 11,\ 8 + 1.5 * 11] = [-19.5, 24.5]$



Άσκηση 2

Να δείξετε ότι το άθροισμα των τετραγώνων των διαφορών N μετρήσεων $\{x_1, \ldots x_N\}$ από μια τιμή $\alpha \in \mathbb{R}$ γίνεται ελάχιστο αν και μόνο αν $\alpha = \bar{X}$.

Εάν επιπλέον θεωρήσουμε τους συντελεστές $\alpha_n = \left(\frac{1}{16}\right)^n$, για ποία τιμή του α θα γίνει ελάχιστο το παρακάτω άθροισμα;

$$\sum_{n=1}^{N} \alpha_n (x_n - \alpha)^2$$

Υπολογίστε αυτή τη τιμή του α αναλυτικά για $x_n = \left(\frac{1}{2}\right)^n, \quad n = 1, \ldots, 10$

$$f(a) = \sum_{n=1}^{N} \alpha_n (x_n - \alpha)^2$$

$$f'(\alpha) = -2 \sum_{n=1}^{N} \alpha_n (x_n - \alpha) = 0 \Leftrightarrow \sum_{n=1}^{N} \alpha_n x_n - \alpha \sum_{n=1}^{N} \alpha_n = 0$$

$$\Leftrightarrow \alpha = \frac{\sum_{n=1}^{N} \alpha_n x_n}{\sum_{n=1}^{N} \alpha_n}$$

$$\alpha = \frac{\sum_{n=0}^{10} \left(\overbrace{\frac{1}{16} \cdot \frac{1}{2}}^{\frac{1}{32}}\right)^n - 1}{\sum_{n=0}^{10} \left(\frac{1}{16}\right)^n - 1}$$

$r < 1$

$$\sum_{n=0}^{N} r^n = \frac{1 - r^{N+1}}{1 - r}$$

Άσκηση 3

Έστω οι παρακάτω ομαδοποιημένες παρατηρήσεις μιας μεταβλητής X

	x	m	f	F	mf
	$[0,1)$	$1/2$	8	8	4
$d=1$	$[1,2)$	$3/2$	4	12	6
	$[\overset{\alpha_3}{2},3)$	$5/2$	6	18	15
	$[\overset{\alpha_4}{3},4)$	$7/2$	10	28	35
total			28	$\overset{\shortparallel}{N}$	60

Υπολογίστε τη μέση τιμή, τη διάμεσο, το πρώτο και τρίτο τεταρτημόριο καθώς και την επικρατέστερη τιμή (εφόσον υπάρχει).

$$\bar{X} = \frac{\Sigma mf}{\Sigma f} = \frac{60}{28} = \frac{15}{7}$$

$$N/2 = 14 \qquad F_2 < N/2 < F_3 \qquad M \in [\alpha_3, \alpha_4)$$

$$M = \alpha_3 + d \cdot \frac{N/2 - F_2}{f_3} = 2 + \frac{14-12}{6} = 2 + 1/3$$

$$N/4 = 7 \qquad \underset{0}{F_0} < N/4 < \underset{8}{F_1} \qquad Q_1 \in [\alpha_1, \alpha_2)$$

$$Q_1 = \alpha_1 + d \frac{N/4 - F_0}{f_1} = 0 + \frac{7}{8} = \frac{7}{8}$$

$$\frac{3N}{4} = 21 \qquad F_3 < 3N/4 < F_4 \qquad Q_3 = \alpha_4 + d \frac{3N/4 - F_3}{f_4} = 3 + \frac{3}{10} = 3.3$$

Άσκηση 4

Έστω το σύνολο δεδομένων

$$\{4, 6, 8, x, 10, 12, 18\}$$

Επιλέξτε κατάλληλα την τιμ του x έτσι ώστε να αποτελεί διάμεσο του δείγματος και να έχει τιμή όσον το δυνατόν πλησιέστερα στη δειγματική μέση τιμή.

$x \in [8, 10]$

$\bar{X} = \frac{58 + x}{7}$ $\quad |\bar{X} - x| = \left| \frac{58 + x - 7x}{7} \right| = \frac{1}{7} |58 - 6x|$ $\quad x = \frac{58}{6} \in [8, 10]$

Άσκηση 5

Μια επιχείρηση απασχολεί 48 εργαζόμενους.

- Οι 10 έχουν μηνιαίο εισόδημα 3600 €
- Οι 6 έχουν μηνιαίο εισόδημα 2000 €
- Οι 8 έχουν μηνιαίο εισόδημα 1000 €
- Οι 24 έχουν μηνιαίο εισόδημα 600 €

Υπολογίστε το συντελεστή Gini για τα μηνιαία εισοδήματα των εργαζομένων της επιχείρησης.

$$\varphi_i = \frac{m_i f_i}{\sum_{j=1}^{N} m_j f_j} \qquad \phi_i = \sum_{j=1}^{i} \varphi_j$$

m	f	mf	φ	Φ	F	RF	$\Sigma\Phi$	ΔRF	$\Sigma\phi \cdot \Delta RF$
600	24	14900	0.205	0.205	24	0.5	0.205	0.5	0.1025
1000	8	8000	0.114	0.319	32	0.67	0.524	0.17	0.089
2000	6	12000	0.17	0.489	38	0.79	0.808	0.12	0.09696
3600	10	36000	0.511	1	48	1	1,489	0.21	0.31269
total		70400							✓

✓

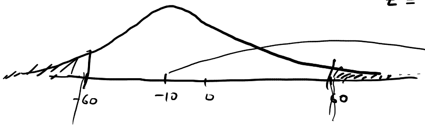
$$Gini = 1 - \sum_{n} (\Sigma\phi_n \cdot \Delta RF_n)$$

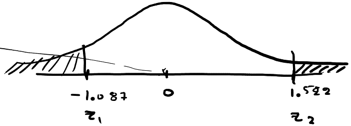
$$\Sigma\phi_n = \Phi_{n-1} + \Phi_n$$

$$\Delta RF_n = RF_n - RF_{n-1}$$

Το σφάλμα ενός μοντέλου μηχανικού ρολογιού στη διάρκεια ενός έτους ακολουθεί κανονική κατανομή με μέση τιμή -10 s και τυπική απόκλιση 46 s. Υπολογίστε το ποσοστό των ρολογιών του συγκεκριμένου μοντέλου τα οποία αναμένεται να παρουσιάσουν μετά από ένα χρόνο συνεχούς λειτουργίας σφάλμα μεγαλύτερο του ενός λεπτού (είτε προς τα επάνω είτε προς τα κάτω).

$$z = \frac{x-\mu}{\sigma}$$





$$z_1 = \frac{-60-(-10)}{46} = \frac{-50}{46} = -1.087$$

$$z_2 = \frac{60-(-10)}{46} = \frac{70}{46} = 1.522$$

$$P(Z > z_2) + P(Z < z_1) =$$

$$= 1 - P(Z < z_2) + P(Z < z_1) =$$

$$= 1 - 0.9357 + 0.1401 = 0.2044$$

20.44 %

Σε δείγμα 2000 ανδρών μιας χώρας το 30 % μετρήθηκε να έχει ύψος μικρότερο από 164 *cm*. Βρείτε το 95% διάστημα εμπιστοσύνης για το ποσοστό των ανδρών της συγκεκριμένης χώρας με ύψος μικρότερο από 164 *cm*.

$\hat{p} = 0.3$ $\qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{N}} = \sqrt{\frac{0.3*0.7}{2000}} = 0.0102$

$95\% = (1-\alpha)\cdot 100\% \qquad \alpha = 0.05$

$P(Z<z) = 1 - \alpha/2 = 0.975 \longrightarrow z = 1.96$

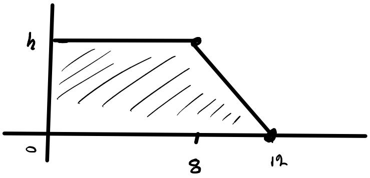
$p \in [\hat{p} - z\sigma_{\hat{p}}, \hat{p} + z\sigma_{\hat{p}}] = [0.3 - 0.0102*1.96, 0.3 + 0.0102*1.96]$

Άσκηση 8

Η συνάρτηση πυκνότητας πιθανότητας μιας συνεχούς τυχαίας μεταβλητής X δίνεται ως

$$p(x) = \begin{cases} 0 & , x < 0 \\ h & , 0 \le x < 8 \\ \left(\dfrac{8-x}{4} + 1\right)h & , 8 \le x < 12 \\ 0 & , x \ge 12 \end{cases}$$

Υπολογίστε την τιμή του h και στη συνέχεια τη μέση τιμή της τυχαίας μεταβλητής ($\mathbb{E}\{X\}$).



$$E = \frac{8+12}{2} h = 1 \Leftrightarrow h = \frac{1}{10}$$

$$\mathbb{E}\{x\} = \int_{-\infty}^{+\infty} x \, p(x) \, dx$$

$$\frac{1}{h} \mathbb{E}\{x\} = \int_0^8 x \, dx + 3\int_8^{12} x \, dx - \frac{1}{4}\int_8^{12} x^2 \, dx$$

Άσκηση 9

Έστω ανελκυστήρας με μέγιστο επιτρεπτό φορτίο $3000\ kg$. Υποθέστε ότι ο πληθυσμός των ατόμων που χρησιμοποιεί τον ανελκυστήρα έχει μέσο βάρος $84\ kg$ και τυπική απόκλιση $15\ kg$. Για 38 άτομα υπολογίστε την πιθανότητα το συνολικό τους βάρος να ξεπερνά τα $3000\ kg$. Υποθέστε ότι τα 38 άτομα αποτελούν τυχαίο δείγμα του πληθυσμού.

$\mu = 84\ kg \qquad \sigma = 15\ kg$

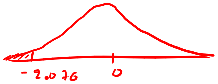
$Y = X_1 + X_2 + \dots + X_{38}$

$$\mu_Y = \mathbb{E}\Big\{ \sum_{n=1}^{38} X_n \Big\} = \sum_{n=1}^{38} \mu_{X_n} = 38\,\mu_{X_1} = 38 \cdot 84 = 3192\ kg$$

$$\sigma_Y^2 = \sum \sigma_{X_n}^2 = N \cdot \sigma_{X_1}^2$$

$$\sigma_Y = \sqrt{N} \cdot \sigma_{X_1} = 92.46\ kg$$

$Y \sim \mathcal{N}(3192,\ 92.46^2)$



$P(Y > 3000)$

$$z = \frac{3000 - \mu_Y}{\sigma_Y} = \frac{3000 - 3192}{92.46} = -2.076$$

$$P(Z > z) = 1 - \overbrace{P(Z \leq z)}^{0.0192} = 0.9808$$

$98.08\ \%$

Άσκηση 10

Τα ύψη των 5 βασικών παικτών μιας ομάδας μπάσκετ έχουν:

- Μέση τιμή 186 *cm*
- Διάμεσο 196 *cm*
- Εύρος 27 *cm*

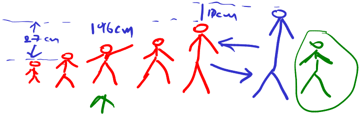
a) Εάν ο ψηλότερος από τους 5 παίκτες αντικατασταθεί από κάποιον που είναι 10 *cm* ψηλότερος, να βρεθούν η μέση τιμή, η διάμεσος και το εύρος για τη νέα πεντάδα.
b) Εάν ο ψηλότερος παίκτης αντικατασταθεί από κάποιον που είναι 11 *cm* κοντύτερος, ποιες από τις νέες τιμές (μέση τιμή, διάμεσο και εύρος) μπορείτε να υπολογίσετε και ποιες θα είναι αυτές οι νέες τιμές;

α)

$\bar{x} = 186$ cm

$M = 196$ cm

$R = 27$ cm



$M' = 196$ cm

$R' = R + 10\text{cm} = 37$ cm

$x_5' = x_5 + 10$

$$\bar{x}' = \frac{x_1+x_2+x_3+x_4+x_5'}{5} = \bar{x} + \frac{10}{5} = \bar{x} + 2\text{cm} = 188\text{ cm}$$

b)

$$\bar{x}'' = \frac{x_1+x_2+x_3+x_4+x_5-11}{5} = \bar{x} - \frac{11}{5} = 186 - 2.2 \text{ cm}$$