MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Διδάσκων: Κώστας Σμαραγδάκης (kesmarag@uoc.gr)

Τελική εξέταση 03-07-2020 : Διάρκεια 2 ώρες

Θέμα Α (3.5 Μ)

Έστω X_1, X_2, X_3 ανεξάρτητες τυχαίες μεταβλητές τέτοιες ώστε:

$$X_i \sim \mathcal{N}(i, i) \quad (\text{μέση τιμή: } i, \text{ διασπορά: } i)$$

1. Βρείτε τη στατιστική κατανομή που ακολουθεί η τυχαία μεταβλητή Y

$$Y = \sum_{j=1}^{3} jX_j$$

2. Υπολογίστε την πιθανότητα $P(2.24 \leq Y \leq 25.76)$.

3. Έστω ένα αρκετά μεγάλο δείγμα παρατηρήσεων της Y. Σχεδιάστε το αναμενόμενο διάγραμμα Box-and-Whisker.

Λύση - Θέμα Α

1

$$Y = X_1 + 2X_2 + 3X_3$$
$$\bar{Y} = 1 + 2 * 2 + 3 * 3 = 14$$
$$\sigma_Y^2 = 1 + 2^2 * 2 + 3^2 * 3 = 36$$

Άρα $Y \sim \mathcal{N}(14, 36) = \mathcal{N}(14, 6^2)$

2

$$z_1 = \frac{y_1 - \bar{Y}}{\sigma_Y} = \frac{2.24 - 14}{6} = -1.96, \; z_2 = \frac{y_2 - \bar{Y}}{\sigma_Y} = \frac{25.76 - 14}{6} = 1.96$$
$$P(y_1 \leq Y \leq y_2) = P(z_1 \leq Z \leq z_2) = P(Z \leq 1.96) - P(Z \leq -1.96)$$
$$= P(Z \leq 1.96) - \Big(1 - P(Z \leq 1.96)\Big) = 0.975 - (1 - 0.975) = 0.95$$

3

Για $Z \sim \mathcal{N}(0, 1)$

- Για $z_1 = -0.67$ έχουμε $P(Z \leq z_1) = 1 - P(Z \leq 0.67) \approx 0.25$.
- Για $z_2 = 0.67$ έχουμε $P(Z \leq z_1) \approx 0.75$.

- Τα αντίστοιχα y_1, y_2 είναι τα αναμενώμενα πρώτο και τρίτο τετραρτημόρια για το δείγμα μας.

$$y_1 = Q_1 = 9.98,\ y_2 = Q_3 = 18.02$$

- Η διάμεσος αναμένουμε να είναι ίση με τη μέση τιμή λόγο συμμετρικότητας της κατανομής.
- Χρειαζόμαστε επίσης το IQR που υπολογίζεται απο τα προηγούμενα.

Μετά σχεδιάζουμε το Box-and-Whisker.

Θέμα Β (3.5 Μ)

Δίνεται η χρονολογική σειρά:

$$Y_t = \{0, 1, 2, 4, 4, 10, 18, 33, 56\}$$

1. Υπολογίστε το αυτοπαλινδρομικό μοντέλο 1ης τάξης.
2. Χρησιμοποιήστε το μοντέλο του πρώτου ερωτήματος και βρείτε εκτίμηση για την επόμενη χρονική στιγμή ($t = 10$).
3. Υπολογίστε το αντίστοιχο μοντέλο για την χρονολογική σειρά

$$\mathcal{Y}_t = \kappa Y_t + \lambda, \quad \kappa \in \mathbb{R} - \{0\},\ \lambda \in \mathbb{R}$$

Λύση - Θέμα Β

1

Έχουμε το σύνολο δεδομένων

$$\{(0,1), (1,2), (2,4), (4,4), (4,10), (10,18), (18,33), (33,56)\}$$

$$b = \frac{SS_{Y_{t-1},Y_t}}{SS_{Y_{t-1},Y_{t-1}}} = \frac{1536}{902} = 1.703$$

$$a = \bar{Y_t} - b\bar{Y}_{t-1} = 16 - 1.703 * 9 = 0.673$$

Άρα $\hat{y}_t = 0.673 + 1.703 y_{t-1}$

2

Η εκτίμηση για $t = 10$ είναι:

$$y_{10} = 0.673 + 1.703 * y_9 = 0.673 + 1.703 * 56 = 96.041$$

3

Έχουμε

$$SS_{\mathcal{Y}_{t-1},\mathcal{Y}_t} = k^2 * SS_{Y_{t-1},Y_t}$$

$$SS_{\mathcal{Y}_{t-1},\mathcal{Y}_{t-1}} = k^2 * SS_{Y_{t-1},Y_{t-1}}$$

άρα $b' = b$

$$a = \bar{\mathcal{Y}_t} - b'\bar{\mathcal{Y}}_{t-1} = k\bar{Y_t} + \lambda - b * (k\bar{Y}_{t-1} + \lambda)$$

και αντικαθηστούμε τις τιμές των $b, \bar{Y_t}, \bar{Y}_{t-1}$.

Θέμα Γ (3.5 Μ)

Έστω τυχαίο δείγμα 30 παρατηρήσεων (ζεύγη) της μορφής $(x$ ελεύθερη, y εξαρτημένη$)$ για το οποίο δίνονται:

$$SS_{xx} = 2247.5, \quad SS_{yy} = 8672, \quad SS_{xy} = 4263,$$

και

$$\bar{X} = 14.5, \quad \bar{Y} = 29.$$

Αφαιρώντας τις παρατηρήσεις $(0, 8)$ και $(29, 50)$ οι οποίες λόγω των συνθηκών που επικρατούσαν κατά τη μέτρησή τους κρίθηκαν αναξιόπιστες, υπολογίστε για το νέο δείγμα (28 ζεύγη παρατηρήσεων) το 95% διάστημα εμπιστοσύνης για την $\mu_{y|x^*=4.5}$.

Λύση - Θέμα Γ

Μετά την αφαίρεση των παρατηρήσεων έχουμε:

$$\bar{X}' = \frac{30 * \bar{X} - 0 - 29}{28} = 14.5 = \bar{X}$$

$$\bar{Y}' = \frac{30 * \bar{Y} - 8 - 50}{28} = 29 = \bar{Y}$$

$$SS'_{xx} = SS_{xx} - (0 - 14.5)^2 - (29 - 14.5)^2 = 1827$$

$$SS'_{yy} = SS_{yy} - (8 - 29)^2 - (50 - 29)^2 = 7790$$

$$SS'_{xy} = SS_{xy} - (0 - 14.5)(8 - 29) - (29 - 14.5)(50 - 29) = 3654$$

$$b = \frac{SS'_{xy}}{SS'_{xx}} = 2, \quad \hat{\mu}_{y|4.5} = 29 + 2(4.5 - 14.5) = 9$$

$$s_e = \sqrt{\frac{SS'_{yy} - bSS'_{xy}}{26}} = 4.31$$

$$s_{\hat{\mu}_{y|4.5}} = s_e\sqrt{1/28 + 10^2/1827} = 1.296$$

Για $df = 26$ έχουμε $t = 2.056$ και βρίσκουμε το διαστήμα εμπιστοσύνης

$$[9 - 2.056 * 1.296,\ 9 + 2.056 * 1.296] = [6.335, 11.665]$$