MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

Άσκηση

Σε μια έρευνα θέλουμε να υπολογίσουμε μια αναλογία στο πληθυσμό (πχ. το ποσοστό των πολιτών που συμφωνούν με μια απόφαση της κυβέρνησης) χρησιμοποιώντας ένα αμερόληπτο δείγμα του πληθυσμού. Ποιο είναι το μικρότερο δυνατό δείγμα που χρειαζόμαστε ώστε το περιθώριο σφάλματος για το 95 % διάστημα εμπιστοσύνης να είναι το πολύ 0.01;

$z = 1.96$ 95%

p 95% $S_{\hat p} = \sqrt{\frac{\hat p(1-\hat p)}{N}} = \frac{\sqrt{\hat p(1-\hat p)}}{\sqrt{N}}$

$$p \in \left[\hat p - zS_{\hat p}, \hat p + \underset{\uparrow}{\overbrace{zS_{\hat p}}}\right]$$

θέλουμε $zS_{\hat p} \le 0.01$

$1.96\, S_{\hat p} \le 0.01$

$\boxed{S_{\hat p} \le \frac{0.01}{1.96}}$ $\forall \hat p$

ορίζουμε $f(x) = \sqrt{x(1-x)}$, $x \in [0,1]$

$$f'(x) = \frac{1-2x}{2\sqrt{x(1-x)}} = 0 \Rightarrow x = \frac{1}{2}$$

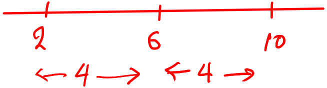
Έστω $\hat p = 1/2$

$$S_{\hat p} = \frac{1}{2\sqrt{N}} \le \frac{0.01}{1.96}$$

$$\sqrt{N} \ge \frac{1.96}{0.02} = 98 \Rightarrow N \ge 9604$$

Άσκηση

Για ένα στατιστικό πληθυσμό έχουμε $\mu = 6$ και $\sigma = 2$. Υπολογίστε το ελάχιστο ποσοστό των παρατηρήσεων στο πληθυσμό με τιμές στο διάστημα $[2, 10]$.



$$[2,10] = \left[6 - \overset{k}{2} \cdot 2\ ,\ 6 + \overset{k}{2} \cdot 2\right]$$

$$\left(1 - \frac{1}{k^2}\right) \cdot 100\% = \left(1 - \frac{1}{4}\right) \cdot 100\%$$

$$= \frac{3}{4} \cdot 100\% = 75\%$$

Άσκηση

Έστω ανεξάρτητες τυχαίες μεταβλητές $\{X_j \sim \mathcal{N}(j, j^2)\}_{j=1}^{3}$. $\rightarrow$ $X_1 \sim \mathcal{N}(1, 1^2)$, $X_2 \sim \mathcal{N}(2, 2^2)$, $X_3 \sim \mathcal{N}(3, 3^2)$

- Τι κατανομές ακολουθούν οι τυχαίες μεταβλητές $Y_j = X_j - \bar{X}$; $\quad (\bar{X} = \frac{X_1+X_2+X_3}{3})$

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$$Y_1 = X_1 - \frac{X_1+X_2+X_3}{3} = \frac{2}{3}X_1 - \frac{X_2}{3} - \frac{X_3}{3} =$$

$$\left.\begin{aligned} &\frac{2}{3}X_1 \sim \mathcal{N}\left(\frac{2}{3}\mu_1, \left(\frac{2}{3}\sigma_1\right)^2\right) \\ &-\frac{1}{3}X_2 \sim \mathcal{N}\left(-\frac{1}{3}\mu_2, \left(\frac{1}{3}\sigma_2\right)^2\right) \\ &-\frac{1}{3}X_3 \sim \mathcal{N}\left(-\frac{1}{3}\mu_3, \left(\frac{1}{3}\sigma_3\right)^2\right) \end{aligned}\right\} \; Y_1 \sim \mathcal{N}\left(\frac{2}{3}\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3, \left(\frac{2}{3}\sigma_1\right)^2 + \left(\frac{1}{3}\sigma_2\right)^2 + \left(\frac{1}{3}\sigma_3\right)^2\right)$$

Υποδ:

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$-X_1 \sim \mathcal{N}(-\mu_1, \sigma_1^2)$$

$$\alpha X_1 \sim \mathcal{N}(\alpha\mu_1, \alpha^2\sigma_1^2)$$

Άσκηση

γραμμική

Για τα ζεύγη παρατηρήσεων 2 μεταβλητών υπολογίστε το συντελεστή συσχέτισης

x y

$$\{(1,1),(2,2),(3,4),(4,3)\}$$

Πιστευετε θα αυξηθεί ή θα μειωθεί εαν προσθέσουμε στο δείγμα το ζευγός $(5,5)$

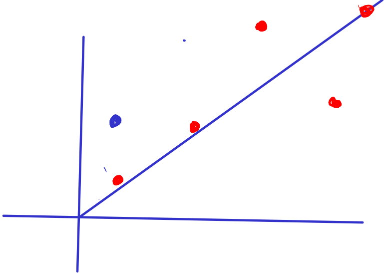
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}}$$

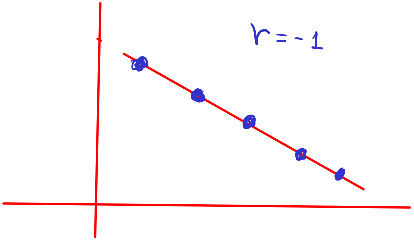
x	y	x^2	y^2	xy
1	1	1	1	1
2	2	4	4	4
3	4	9	16	12
4	3	16	9	12
10	10	30	30	29

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{N} = 29 - \frac{100}{4} = 4$$

$$r = \frac{4}{\sqrt{5^2}} = \frac{4}{5} = 0.8$$

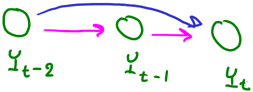
$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{N} = 30 - \frac{100}{4} = 5 = SS_{yy}$$




r = - 1

Άσκηση: Να υπολογιστεί ο συντελεστής PACF(2)

$\{1, 2, 4, 3, 5\}$



① : $Y_{t-1} \rightarrow Y_t$ $\quad \{(\overset{y_{t-1}}{1}, \overset{y_t}{\underset{e_2}{2}}), (2, \underset{e_3}{4}), (4, \underset{e_4}{3}), (3, \underset{e_5}{5})\}$ $\quad y_t = \alpha_1 + b_1 y_{t-1}$

② $Y_{t-1} \rightarrow Y_{t-2}$ $\quad \{(2,1), (4,2), (3,4), (5,3)\}$ $\quad y_{t-2} = \alpha_2 + b_2 y_{t-1}$

① $b_1 = \dfrac{SS_{y_{t-1}, y_t}}{SS_{y_{t-1}, y_{t-1}}}$

y_{t-2} y_{t-1}	y_{t-1} y_t	y_{t-2}^2 y_{t-1}^2	y_{t-1}^2 y_t^2	$y_{t-2}y_{t-1}$ $y_{t-1}y_t$
1	2	1	4	2
2	4	4	16	8
4	3	16	9	12
3	5	9	25	15
10	14	30	54	37

$$\boxed{\hat{y}_t = 2.5 + 0.4 y_{t-1}}$$

$$\alpha_1 = \bar{Y}_t - b \bar{Y}_{t-1} = \frac{14}{4} - \frac{2}{5} \cdot \frac{10}{4} = \frac{10}{4} = 2.5$$

$$b = \frac{2}{5}$$

$$SS_{y_{t-1} y_t} = \sum y_{t-1} y_t - \frac{\sum y_{t-1} \sum y_t}{4} = 37 - \frac{140}{4} = 37 - 35 = 2$$

$$SS_{y_{t-1} y_{t-1}} = \sum y_{t-1}^2 - \frac{(\sum y_{t-1})^2}{4} = 30 - \frac{100}{4} = 5$$

$(e_1)_t = y_t - \hat{y}_t$ $\quad (e_1)_2 = y_2 - \hat{y}_2 = -0.9$ $\quad (e_1)_3 , (e_1)_4 , (e_1)_5$

$\hat{y}_2 = 2.5 + 0.4 \cdot y_1 = 2.5 + 0.4 = 2.9$

$y_2 = 2$

②

$$SS_{y_{t-2} y_{t-1}} = \sum y_{t-2} y_{t-1} - \frac{\sum y_{t-2} \sum y_{t-1}}{4} = 37 - \frac{140}{4} = 2$$

$$SS_{y_{t-1} y_{t-1}} = \sum y^2_{t-1} - \frac{(\sum y_{t-1})^2}{4} = 54 - \frac{14^2}{4} = 54 - 49 = 5$$

$b_2 = \frac{2}{5}$ $\quad \alpha_2 = \bar{Y}_{t-2} - b \bar{Y}_{t-1} = \frac{10}{4} - \frac{2}{5} \cdot \frac{14}{4} =$

$\hat{y}_{t-2} = \alpha_2 + b_2 y_{t-1}$ $\quad (e_2)_{t-2} = y_{t-2} - \hat{y}_{t-2} = y_{t-2} - \alpha_2 - b_2 y_{t-1}$

$$(\hat{e}_1)_t = \alpha_3 + b_3 (\hat{e}_2)_{t-2}$$