MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

19-05-2021

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;

$z = 1.96$

$\overset{\leq 0.01}{[\quad]}$ 95% $N = ;$

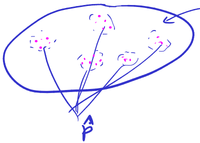
$\xleftrightarrow{2 z S_{\hat p}}$

$[\hat p - z S_{\hat p}, \hat p + z S_{\hat p}]$

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

σταθερά

Περιθώριο σφάλματος $z S_{\hat p}$

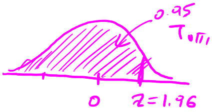


$p = 0.3$

$[0.29, 0.31]$

95%

95%



Τυπική κανονική κατανομή $N(0,1^2)$

Θέλουμε $1.96 \cdot S_{\hat p} \leq 0.01$

$1.96 \max_{\hat p} S_{\hat p} \leq 0.01$

$f(\hat p) = \sqrt{\hat p (1-\hat p)}$ $\quad \hat p \in (0,1)$

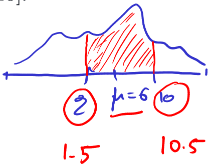
$f'(\hat p) = \frac{1}{2\sqrt{\hat p (1-\hat p)}}(1-2\hat p) = 0 \Rightarrow \hat p = \frac{1}{2}$

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

$1.96 \frac{1}{2\sqrt N} \leq 0.01 \Rightarrow N \geq 2500$

Άσκηση

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



Για $k \geq 1$ τουλάχιστον το $\left(1 - \frac{1}{k^2}\right)100\%$

βρίσκ. στο διάστημα $[\mu - k\sigma, \mu + k\sigma]$

$$\mu - k\sigma = 2$$

$$6 - k \cdot 2 = 2 \Rightarrow k = 2$$

$$\left(1 - \frac{1}{4}\right) \cdot 100\% = \frac{3}{4} \cdot 100\% = 75\%$$

$$\in [2, 10]$$

Άσκηση

$X_1 \sim N(1,1)$, $X_2 \sim N(2,4)$, $X_3 \sim N(3,9)$

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

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

β) ▶ Έστω $y_1, y_2, \ldots, y_{100}$ παρατηρήσεις της Y_j. Βρείτε το 95 % διάστημα εμπιστοσύνης για τη δειγματική μέση τιμή της Y_j.

$Y_j = X_j - \bar{X}$, το $\bar{X}$ είναι τυχαία μεταβλητή $\bar{X} = \frac{X_1+X_2+X_3}{3}$

α) Γραμμ. συνδυασμός κανονικών κατανομών είναι κανονική κατανομή !!!

$$Y_j = X_j - \frac{X_1+X_2+X_3}{3} = \frac{3X_j - X_1 - X_2 - X_3}{3}$$

$(j=1)$ $Y_1 = \frac{2X_1 - X_2 - X_3}{3} = \frac{2}{3}X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3$

$(j=2)$ $Y_2 = \frac{-X_1 + 2X_2 - X_3}{3}$

$(j=3)$ $Y_3 = \frac{-X_1 - X_2 + 2X_3}{3}$

$$\mu_{Y_1} = \frac{2}{3}\mu_{X_1} - \frac{1}{3}\mu_{X_2} - \frac{1}{3}\mu_{X_3} = \cancel{\frac{2}{3}} - \cancel{\frac{1}{3}}2 - \frac{1}{3}3 = -1$$

$$\sigma_{Y_1}^2 = \left(\frac{2}{3}\right)^2 \sigma_{X_1}^2 + \left(-\frac{1}{3}\right)^2 \sigma_{X_2}^2 + \left(-\frac{1}{3}\right)\sigma_{X_3}^2 = \frac{4}{9}\cdot 1 + \frac{1}{9}\cdot 4 + \frac{1}{9}\cdot 9 =$$

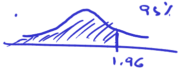
$$= 1 + \frac{8}{9}$$

$Y_1 \sim N\left(-1, 1 + \frac{8}{9}\right)$

$Y_1 \equiv Y^{(1)}$

β) $j = 1$ $Y^{(1)} \sim N\left(-1, 1 + \frac{8}{9}\right)$

$y_1^{(1)}, y_2^{(1)}, \dots, y_{100}^{(1)}$ $\qquad \mu_{y^{(1)}}$



$$\sigma^2 = 1 + \frac{8}{9} \Rightarrow \sigma = \sqrt{1+\frac{8}{9}} =$$

$$= \sqrt{\frac{17}{9}} = \frac{\sqrt{17}}{3}$$

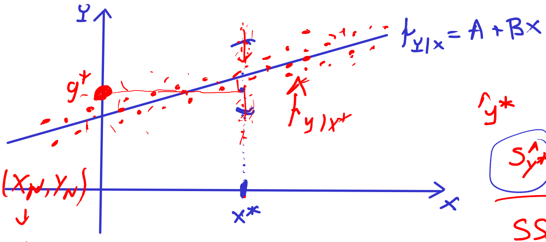
$$\mu_{y_j^{(1)}} \in \left[\bar{y}^{(1)} - 1.96\,\frac{\sqrt{17}}{3},\ \bar{y}^{(1)} + 1.96\cdot\frac{\sqrt{17}}{3}\right]$$ όπου $$\bar{y}^{(1)} = \frac{\sum_{j=1}^{100} y_j^{(1)}}{100}$$

Γραφική Παλινδρόμηση και Διαστήματα Εμπιστοσύνης (Απλή)

$$Y = \underbrace{A + BX}_{\mu_{Y|X}} + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

$$\hat{y} = \alpha + bx$$

$$e = y - \hat{y}$$



$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$$

$$e_1 = y_1 - \alpha - bx_1 \qquad e_N = y_N - \alpha - bx_N$$

$$\{e_1, e_2, \ldots, e_N\}$$

$$e_n = y_n - \alpha - b x_n$$

$$\bar{e} = \bar{Y} - \alpha - b\bar{X}$$

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

$$\hat{y}^*$$

$$S_{\hat{y}^*} = S_e \sqrt{1 + \frac{1}{N} + \frac{(x^* - \bar{X})^2}{SS_{xx}}}$$

$$SS_{xx} = \sum (x_n - \bar{X})^2 = \sum x_n^2 - \frac{(\sum x_n)^2}{N}$$

$$S_e = \sqrt{\frac{\sum e_n^2}{N-2}}$$

$$S_{\hat{\mu}_{y|x^*}} = \sqrt{\frac{\ldots}{98}} \sqrt{\frac{1}{100} + \frac{(x^* - \bar{X})^2}{SS_{xx}}}$$

$$\sum e_n^2 = SS_{yy} - bSS_{xy}$$

$$SS_{yy} = \sum (Y_n - \bar{Y})^2$$

$$SS_{xy} = \sum (x_n - \bar{X})(y_n - \bar{Y}) = \sum x_n y_n - \frac{\sum x_n \sum y_n}{N}$$

$$s_{\hat{\mu}_{y|x^*}} = S_e \sqrt{\frac{1}{N} + \frac{(x^* - \bar{x})^2}{SS_{xx}}}$$

$\hat{y} = \alpha + bx$

$\hat{y}^* = \alpha + bx^*$

Διάστημα Εμπιστοσύνης : Τα $\hat{y}^*$, $\hat{\mu}_{y|x^*}$ έχουν (ίδια τιμή $\alpha + bx^*$)

95%

$$\left[\alpha + bx^* - \overset{z}{1.96} \begin{cases} \cdot\, S_{\hat{\mu}_{y|x^*}} \\ \cdot\, S_{\hat{y}^*} \end{cases},\ \alpha + bx^* + 1.96 \begin{cases} \cdot\, S_{\hat{\mu}_{y|x^*}} \\ S_{\hat{y}^*} \end{cases}\right]$$

—

Συνάντηση : 11-06 19:00