MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

13-05-2020

Πρόγραμμα για την ολοκλήρωση του μαθήματος

Επόμενες συναντήσεις από το zoom

- Επίλυση ασκήσεων φυλλαδίου (μέρη 2/3 και 3/3).
  - Δευτέρα 18-05-2020, 15:00
  - Τετάρτη 20-05-2020, 11:15
- Απορίες, συζήτηση και ολοκλήρωση των διαλέξεων (δεν θα πραγματοποιηθεί καταγραφή)
  - Πέμπτη 21-05-2020, 10:15

Τελική Εξέταση ?

- 60 % (ή 75 % για όσους/όσες δεν δώσουν το φυλλάδιο ασκήσεων).
- Το πιθανότερο σενάριο με τα σημερινά δεδομένα είναι να γίνει δια ζώσης εξέταση με κλειστές σημειώσεις και θέματα ανάπτυξης. Τιμές που πιθανόν να χρειαστείτε θα είναι διαθέσιμες.
- Για την εξέταση θα μπορείτε να χρησιμοποιήσετε αριθμομηχανή.

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 $\underbrace{95\ \%}_{1.96}$ διάστημα εμπιστοσύνης να είναι το πολύ 0.01;

p

$\hat{p}$

$$p \to [\hat{p} - zS_{\hat{p}},\ \hat{p} + \overbrace{zS_{\hat{p}}}]$$

Περιθώριο σφάλματος $\underline{zS_{\hat{p}}}$ $\qquad z = 1.96$

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

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

$$x \neq \{0,1\} \quad f'(x) = \frac{1}{2\sqrt{x(1-x)}}[1-2x] \qquad f'(x) = 0 \Leftrightarrow x = \frac{1}{2}$$

$$\boxed{\hat{P} = \frac{1}{2}}$$

$$zS_{1/2} \leq 0.01$$

$$1.96\sqrt{\frac{1}{4N}} \leq 0.01 \Leftrightarrow \sqrt{N} \geq \frac{1.96}{2 \cdot 0.01} = 50 \Leftrightarrow \boxed{N \geq 2500}$$

Άσκηση

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

$6-2\cdot 2 \leftarrow 6 \rightarrow 6+2\cdot 2$

Για κάθε $k>1$ τουλάχιστον $(1-1/k^2)\cdot 100\%$ των παρατηρήσεων ανήκουν στο διάστημα

$$[\mu - k\sigma, \mu + k\sigma]$$

$k=2$ τουλάχιστον $\left(1-\frac{1}{4}\right)\cdot 100\% = \frac{3}{4}100\% = 75\%$ των παρατηρήσεων

$$[2, 10]$$

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

(α) ▶ Τι κατανομές ακολουθούν οι τυχαίες μεταβλητές $Y_j = X_j - \bar{X}$; $\quad (\bar{X} = \frac{X_1+X_2+X_3}{3})$ $\quad j = 1, 2, 3$

(β) ▶ Έστω $y_1, y_2, \ldots, y_{100}$ παρατηρήσεις της Y_j. Βρείτε το 95 % διάστημα εμπιστοσύνης για τη ~~δειγματική~~ μέση τιμή ~~των παρατηρήσεων~~. της Y_j $\quad j=1 \quad \bar{Y} = \frac{y_1+y_2+\ldots+y_{100}}{100}$

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

$j=1$

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

$$\mathbb{E}\{Y_1\} = \frac{2}{3}\mathbb{E}\{X_1\} - \frac{1}{3}\mathbb{E}\{X_2\} - \frac{1}{3}\mathbb{E}\{X_3\} = \frac{2}{3} - \frac{1}{3}\cdot 2 - \frac{1}{3}3 = -1$$

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

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

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

$j=2$

$$Y_2 = X_2 - \frac{X_1+X_2+X_3}{3} = -\frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3$$

$$\boxed{Y_2 \sim \mathcal{N}\left(0, 1+\frac{17}{9}\right)}$$

$$\mathbb{E}\{Y_2\} = -\frac{1}{3}\cdot 1 + \frac{2}{3}2 - \frac{1}{3}3 = 0$$

$$\mathrm{Var}\{Y_2\} = \frac{1}{9}\mathrm{Var}\{X_1\} + \frac{4}{9}\mathrm{Var}\{X_2\} + \frac{1}{9}\mathrm{Var}\{X_3\} = \frac{1}{9} + \frac{4}{9}\cdot 4 + \frac{1}{9}\cdot 9 = 1 + \frac{17}{9}$$

$Y_1 \sim N\left(-1,\ 1+\frac{8}{9}\right)$ $\qquad N=100$

$95\% \leftrightarrow z=1.96$

$$\sigma^2 = 1+\frac{8}{9} \Leftrightarrow \sigma = \sqrt{1+\frac{8}{9}} = \frac{\sqrt{17}}{3}$$

$$S_{\bar{Y}} = \frac{\sigma}{\sqrt{N}} = \frac{\sqrt{17}/3}{10} = \frac{\sqrt{17}}{30}$$

$z \cdot S_{\bar{Y}}$

$$\mu_{Y_j} \in \left[\bar{Y} - z\frac{\sqrt{17}}{30},\ \bar{Y} + z\frac{\sqrt{17}}{30}\right]$$ με πιθανότητα $\underline{0.95}$

$$\parallel$$

$$\mu_{Y_j} \in \left[\bar{Y} - 1.96\frac{\sqrt{17}}{30},\ \bar{Y} + 1.96\frac{\sqrt{17}}{30}\right]$$