

12.2 Εμπειρική Εκτίμηση Ομαδοποιημένων Δεδομένων

• Εμπειρική Συνάρτηση Κατανομής : $F_n(C_j) = \frac{1}{n} \sum_{i=1}^j n_i$

• Εμπειρικός Εκτιμητής $\hat{E}_n$ της μ -τάξης : $E_n(X^k) = \sum_{j=1}^r \left[\frac{n_j}{n} \cdot \frac{C_j^{k+1} - C_{j-1}^{k+1}}{(k+1)(C_j - C_{j-1})} \right]$

Θ A_n $C_i = m$:

$$E_n[(X)_{(m)}] = \sum_{j=1}^i \left[\frac{n_j}{n} \cdot \frac{C_j + C_{j-1}}{2} \right] + m \cdot \sum_{j=i+1}^r \frac{n_j}{n}$$

Θ A_n $m \in (C_{i-1}, C_i)$

$$E_n[(X)_{(m)}] = \sum_{j=1}^{i-1} \left[\frac{n_j}{n} \cdot \frac{C_j + C_{j-1}}{2} \right] + \frac{n_i}{n} \cdot \frac{C_{i-1} + m}{2} \cdot \frac{m - C_{i-1}}{C_i - C_{i-1}} + m \cdot \frac{n_i}{n} \cdot \frac{C_i - m}{C_i - C_{i-1}} + m \cdot \sum_{j=i+1}^r \frac{n_j}{n}$$

Άσκηση / Παράδειγμα ασκ. 35 σελ 182

Έστω ότι έχουμε 51 παρατηρήσεις ομαδοποιημένες σε 4 κατηγορίες όπως παρουσιάζονται στον παρακάτω πίνακα. Να βρεθούν :

Πίνακας	
Κατανομή Κόστους Ζημιών	
Είδος Πληρωμών	Αριθ. Πληρωμών
$[0, 2]$	9
$(2, 6]$	13
$(6, 9]$	19
$(9, 15]$	10

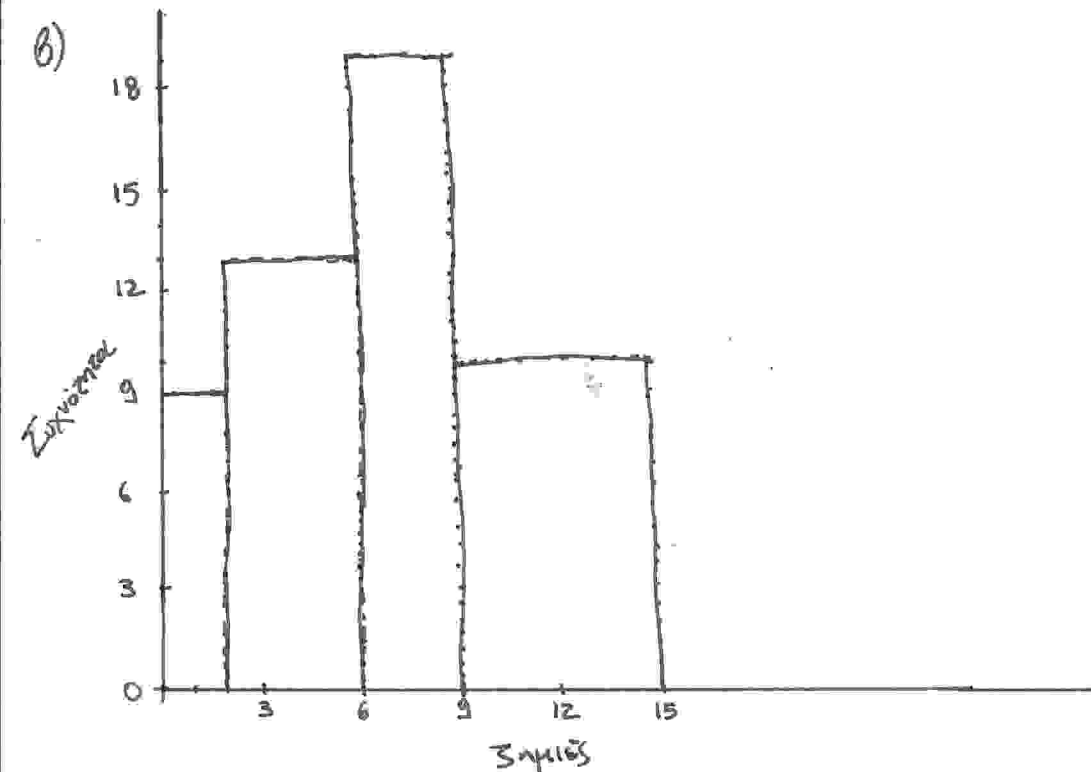
- Η εμπειρική συνάρτηση κατανομής
- Το ισόγραμμα
- Ο εμπειρικός εκτιμητής της $1^{ης}$ και $2^{ης}$ τάξης καθώς και της διασποράς
- Η εμπειρική περιορισμένη μέση τιμή για :
 - $m=9$
 - $m=8$
- Ο εμπειρικός εκτιμητής του $70^{ου}$ ποσοστημορίου.

Λύση

α) Για ομαδοποιημένα δεδομένα έχουμε $F_n(C_j) = \frac{1}{n} \sum_{i=1}^j n_i$.
Έτσι έχουμε :

$$F_{51}(0) = 0, F_{51}(2) = \frac{9}{51}, F_{51}(6) = \frac{9+13}{51} = \frac{22}{51}, F_{51}(9) = \frac{9+13+19}{51} = \frac{41}{51}, F_{51}(15) = 1$$

(2)



γ) Η εμπειρική " k -οστή ροπή" δίνεται από $E_n(X^k) = \sum_{j=1}^r \left[\frac{n_j}{n} \cdot \frac{C_j^{k+1} - C_{j-1}^{k+1}}{(k+1) \cdot (C_j - C_{j-1})} \right]$

$$\bullet E_{51}(X) = \frac{9}{51} \cdot \frac{2^2 - 0^2}{2 \cdot (2 - 0)} + \frac{13}{51} \cdot \frac{6^2 - 2^2}{2 \cdot (6 - 2)} + \frac{19}{51} \cdot \frac{9^2 - 6^2}{2 \cdot (9 - 6)} + \frac{10}{51} \cdot \frac{15^2 - 9^2}{2 \cdot (15 - 9)} =$$

$$= \frac{9}{51} \cdot 1 + \frac{13}{51} \cdot 4 + \frac{19}{51} \cdot \frac{15}{2} + \frac{10}{51} \cdot 12 = 6.343137$$

$$\bullet E_{51}(X^2) = \frac{9}{51} \cdot \frac{2^3 - 0^3}{3 \cdot (2 - 0)} + \frac{13}{51} \cdot \frac{6^3 - 2^3}{3 \cdot (6 - 2)} + \frac{19}{51} \cdot \frac{9^3 - 6^3}{3 \cdot (9 - 6)} + \frac{10}{51} \cdot \frac{15^3 - 9^3}{3 \cdot (15 - 9)} =$$

$$= \frac{9}{51} \cdot \frac{4}{3} + \frac{13}{51} \cdot \frac{52}{3} + \frac{19}{51} \cdot \frac{171}{3} + \frac{10}{51} \cdot \frac{441}{3} = 54.71242$$

Αρα $\text{Var}_{51}(X) = E_{51}(X^2) - E_{51}(X)^2 = (54.71242) - (6.343137)^2 = 14.47703$.

δ) Η εμπειρική περιορισμένη μέση τιμή για $m=9$ (συμπίπτει με το άρρο του 3ου κελιού) δίνεται από $E_n(X_{1m}) = \sum_{j=1}^i \left[\frac{n_j}{n} \cdot \frac{C_j + C_{j-1}}{2} \right] + m \cdot \sum_{j=i+1}^r \frac{n_j}{n}$

$$E_{51}(X_{19}) = \frac{9}{51} \cdot \frac{2^2 - 0^2}{2 \cdot (2 - 0)} + \frac{13}{51} \cdot \frac{6^2 - 2^2}{2 \cdot (6 - 2)} + \frac{19}{51} \cdot \frac{9^2 - 6^2}{2 \cdot (9 - 6)} + 9 \cdot \frac{10}{51}$$

$$= \frac{9}{51} \cdot 1 + \frac{13}{51} \cdot 4 + \frac{19}{51} \cdot \frac{15}{2} + \frac{10}{51} \cdot 9 = 5.754902$$

ii) Η εμπειρική περιορισμένη μέση τιμή για $m=8$ (έντος του διαστήματος) (3)
(6,9)

Δίνεται από τον τύπο:

$$E_n[X_{nm}] = \sum_{j=1}^{i-1} \left[\frac{n_j}{n} \cdot \frac{c_j + c_{j-1}}{2} \right] + \frac{n_i}{n} \cdot \frac{c_{i-1} + m}{2} \cdot \frac{m - c_{i-1}}{c_i - c_{i-1}} + m \cdot \frac{n_i}{n} \cdot \frac{c_i - m}{c_i - c_{i-1}} + m \cdot \sum_{j=i+1}^r \frac{n_j}{n}$$

$$\begin{aligned} E_{51}(X_{18}) &= \frac{9}{51} \cdot \frac{0+2}{2} + \frac{13}{51} \cdot \frac{6+2}{2} + \frac{19}{51} \cdot \frac{6+8}{2} \cdot \frac{8-6}{9-6} + 8 \cdot \frac{19}{51} \cdot \frac{9-8}{9-6} + 8 \cdot \frac{10}{51} \\ &= \frac{9}{51} \cdot 1 + \frac{13}{51} \cdot 4 + \frac{19}{51} \cdot \frac{28}{6} + \frac{19}{51} \cdot \frac{8}{3} + \frac{10}{51} \cdot 8 = \\ &= 5.496732 \end{aligned}$$

ε) Για να βρούμε το 70^ο ποσοστημώριο υπολογίζουμε τη σχέση:

$$\frac{P \cdot n}{100} = \frac{70 \cdot 51}{100} = 35.7$$

Πρέπει να βρούμε που βρίσκεται η τιμή 35.7 στην στήλη των αθροιστικών συχνοτήτων F_i . Γι αυτό κατασκευάζουμε τον πίνακα αθροιστικών συχνοτήτων

Πίνακας

Κατανομή Κοστών Ζημιών		
Είδος Πληρωμών	Αριθμός Πληρωμών	Συνολικός Αριθμός Πληρωμών
(0, 2]	9	9
(2, 6]	13	22
(6, 9]	19	41
(9, 15]	10	51

Η τιμή 35.7 βρίσκεται στο 3^ο καί (6,9]. Άρα

$$\begin{aligned} \hat{\xi}_{70} &= X_0 + \frac{C}{f_3} \cdot \left(\frac{P \cdot n}{100} - F_2 \right) = \\ &= 6 + \frac{3}{19} \cdot (35.7 - 22) = 8.163158 \end{aligned}$$

Κεφάλαιο 13 Έλεγχος υαδής Προσάρμοξης

4

13.2 Kolmogorov - Smirnov test

$H_0: F(x) = F_0(x)$, $F(x) = H$ πραγματική σ.κ του δείγματος

$H_1: F(x) \neq F_0(x)$, $F_0(x) = H$ θεωρητική κατανομή που μας ενδιαφέρει να ελεγχουμε

$$D_n = \max_x |F_n(x) - F_0(x)| , F_n(x) = \frac{1}{n} \cdot \sum_{i=1}^n \underbrace{1_{[X_i \leq x]}}_{\# \text{ των } X_i \leq x}$$

Απορρίπτουμε την H_0 σε ε.σ. α αν ισχύει ότι $D_n > D_{n;\alpha}$

όπου προσεγγιστικά έχουμε : $D_{n;0.1} = \frac{1.22}{\sqrt{n}}$, $D_{n;0.05} = \frac{1.36}{\sqrt{n}}$, $D_{n;0.01} = \frac{1.63}{\sqrt{n}}$

Άσκηση 1

Να ελεγχθεί σε επίπεδο σημαντικότητας $\alpha = 0.1$ αν οι παρατηρήσεις

1 2 3 5 13

προέρχονται από την inverse exponential με παράμετρο $\theta = 2$ και σ.η.η.

$$f(x|\theta) = \frac{\theta e^{-\theta/x}}{x^2} , x > 0$$

Λύση

Ο έλεγχος χρησιμοποιεί την σ.κ. $F_0(x)$. Γιατί θα πρέπει πρώτα να βρούμε την F_0 .

$$F_0(x) = \int_0^x \frac{\theta e^{-\theta/y}}{y^2} dy = \int_0^x (-e^{-\theta/y})' dy = [e^{-\theta/y}]_0^x = e^{-\theta/x} - \lim_{y \rightarrow 0} e^{-\theta/y} = e^{-\theta/x}$$

$\Rightarrow F_0(x) = e^{-\theta/x}$, $x > 0$. Άρα ο πίνακας είναι :

x	$F_0(x) = e^{-2/x}$	$F_n^-(x)$	$F_n^+(x)$	$ F_0(x) - F_n^-(x) $	$ F_0(x) - F_n^+(x) $	$ F_0(x) - F_n(x) $
1	$e^{-2} = 0.1353353$	0	$1/5 = 0.2$	0.1353353	0.06466472	0.1353353
2	$e^{-1} = 0.3678794$	$1/5 = 0.2$	$2/5 = 0.4$	0.1678794	0.03212056	0.1678794
3	$e^{-2/3} = 0.5134171$	$2/5 = 0.4$	$3/5 = 0.6$	0.1134171	0.08658288	0.1134171
5	$e^{-2/5} = 0.67032$	$3/5 = 0.6$	$4/5 = 0.8$	0.07032005	0.12968	0.12968
13	$e^{-2/13} = 0.8574039$	$4/5 = 0.8$	$5/5 = 1$	0.05740392	0.1425961	0.1425961

Άρα $D_n = \max_x |F_n(x) - F_0(x)| = 0.1678794$ και $D_{5;0.1} = 0.5456006$

Άρα αφού $D_n > D_{n;0.1}$ δεν ισχύει, δεν μπορούμε να απορρ. την H_0 σε ε.σ. 0.1

Άσκηση 47/Σελ 244

- Δίνονται :
- α) Ο αριθμός των αναζητήσεων για ένα σύνολο 10 ασφαλιστικών κινδύνων ακολουθεί την κατανομή Poisson με μέση τιμή Θ
 - β) η Θ ακολουθεί την κατανομή Gamma με μέση τιμή 0,15 και διασπορά 0,002
 - γ) κατά την διάρκεια του πρώτου έτους παρατήρησης βλέπουμε ότι έχουν συμβεί 5 ζημιές από το σύνολο των 10 ασφαλιστικών κινδύνων. Να βρεθεί η εν των υστέρων διασπορά της Θ .

Λύση

$$\bullet f(x_i|\theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} \Rightarrow f(\underline{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = e^{-\theta n} \cdot \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

$$\bullet u(\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}, \theta > 0, \alpha > 0, \beta > 0 \text{ ώστε } E(\theta) = 0,15 \text{ και } \text{Var}(\theta) = 0,002$$

$$\Rightarrow \frac{\alpha}{\beta} = 0,15 \text{ και } \frac{\alpha}{\beta^2} = 0,002 \Rightarrow \begin{cases} \alpha = 11,25 \\ \beta = 75 \end{cases}$$

Άρα $\Theta \sim \text{Gamma}(11,25, 75)$

Για την εν των υστέρων κατανομή έχουμε :

$$\begin{aligned} f(\theta|\underline{x}) &\propto f(\underline{x}|\theta) \cdot u(\theta) = e^{-\theta n} \cdot \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \cdot \frac{\theta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta} \\ &\propto \underbrace{\theta^{\sum_{i=1}^n x_i + \alpha - 1} \cdot e^{-\theta(\beta+n)}}_{\text{πυκνότητα Gamma}} \end{aligned}$$

$$\text{Άρα } \Theta|\underline{x} \sim \text{Gamma}(\sum x_i + \alpha, \beta + n) \Rightarrow \text{Var}(\Theta|\underline{x}) = \frac{\alpha^*}{\beta^{*2}} \Rightarrow$$

$$\Rightarrow \text{Var}(\Theta|\underline{x}) = \frac{\sum_{i=1}^n x_i + \alpha}{(\beta + n)^2} \text{ για } \alpha = 11,75, \beta = 75, n = 10, \sum_{i=1}^{10} x_i = 5$$

$$\text{έχουμε ότι } \text{Var}(\Theta|\underline{x}) = \frac{5 + 11,75}{(75 + 10)^2} = 0,0023183391$$

Για 3 έτη έχουμε 120 συμβόλαια. Αν κάθε έτος έχουμε 10 ατοχήματα τότε $\sum_{i=1}^{120} x_i = 30$

Κεφάλαιο 13 Έλεγχος Καλής Προσαρμογής.

(5)

$$H_0: F(x) = F_0(x) \quad , \quad F(x) = \text{Η πραγματική β.κ. του δείγματος}$$

$$H_a: F(x) \neq F_0(x) \quad F_0(x) = \text{Η θεωρητική κατανομή που μας ενδιαφέρει να ελέγξουμε.}$$

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^k \frac{(n_j - n \cdot p_j)^2}{n \cdot p_j} \quad , \quad \text{όπου } p_j = F_0(c_j) - F_0(c_{j-1})$$

Απορρίπτεται η H_0 σε επίπεδο σημαντικότητας α αν ισχύει $\chi^2 > \chi^2_{k-r-1; \alpha}$ με $k = \#$ κελιών και $r = \#$ των παραμέτρων της κατανομής F_0

(Αν το k δεν δίνεται τότε χρησιμοποιώ $k \approx 1 + 3.322 \log_{10} n \rightarrow \text{Sturge's Rule}$)

Άσκηση Bonus

Έστω ~~σημεία~~ μιας ασφαλιστικής εταιρίας δίνονται στον παρακάτω πίνακα:

Να ελέγξετε αν τα δεδομένα προέρχονται από την ευθετική κατανομή με μέση τιμή 50, σε επίπεδο σημαντικότητας $\alpha = 0,05\%$.

Πίνακας	
Εύρος Ομάδας j	O_j
$(0, 20]$	32
$(20, 65]$	31
$(65, 120]$	28
$(120, \infty)$	19

Λύση

Αφού οι σημεία $X \sim \text{Exp}(\theta)$ με $E(X) = 50 \Rightarrow \frac{1}{\theta} = 50 \Rightarrow \theta = \frac{1}{50} = 0.02$.

Αρα η $F_0(x) = 1 - e^{-0.02x}$, $r=1$, $k=4$, $n=110$

$$E_1 = n \cdot p_1 = n \cdot [F_0(20) - F_0(0)] = 110 \cdot [1 - e^{-0.02 \cdot 20} - (1 - 1)] = 110 \cdot 0.32968 = 36.26479$$

$$E_2 = n \cdot p_2 = n \cdot [F_0(65) - F_0(20)] = 110 \cdot [1 - e^{-0.02 \cdot 65} - 1 + e^{-0.02 \cdot 20}] = 110 \cdot 0.3977883 = 43.75671$$

$$E_3 = n \cdot p_3 = n \cdot [F_0(120) - F_0(65)] = 110 \cdot [1 - e^{-0.02 \cdot 120} - 1 + e^{-0.02 \cdot 65}] = 110 \cdot 0.1818138 = 19.99952$$

$$E_4 = n \cdot p_4 = n \cdot [F_0(\infty) - F_0(120)] = 110 \cdot [1 - 1 + e^{-0.02 \cdot 120}] = 110 \cdot 0.09071795 = 9.978975$$

Αρα

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^4 \frac{(n_j - n \cdot p_j)^2}{n \cdot p_j} = \frac{(32 - 36.26479)^2}{36.26479} + \frac{(31 - 43.75671)^2}{43.75671} + \frac{(28 - 19.99952)^2}{19.99952} + \frac{(19 - 9.978975)^2}{9.978975} = 15.5761 \quad , \quad \chi^2_{4-1-1; 0.05} = \chi^2_{2; 0.05} = 5.991465$$

Αφού ισχύει ότι $\chi^2 > \chi^2_{2; 0.05}$ τότε απορρ. η H_0 σε ε.σ. $\alpha = 5\%$.