Μαθημα 2ο: 27/2/19

χ^2 τεστ καλής προσαρμογής - Ονομαστικά Δεδομένα

Έστω $X_1, \dots, X_n$ τ.δ. από πληθυσμό με ασκ $F(x)$ και $F_0(x)$ γνωστή κατανομή. Μας ενδιαφέρει ο έλεγχος της υπόθεσης: $H_0: F(x) = F_0(x)$ v $H_a: F(x) \neq F_0(x)$.
Διαιρούμε τον δειγματικό χώρο σε k κατηγορίες (ή τμήματα) και έστω n_i οι συχνότητες των δεδομένων, $i=1,\dots,k$, $n = \sum_{i=1}^{k} n_i$ στις k κατηγορίες.
Υπολογίζουμε την πιθανότητα η τ.μ. X να πάρει τιμές σε κάθε μια από τις k κατηγορίες με την βοήθεια της κατανομής $F_0(x)$ της H_0.
Δηλαδή: $p_i = P(X \in$ i-οστή κατηγορία $, i=1,\dots,k \mid F_0(x))$
όταν δίνεται η κατανομή της H_0

και $\sum_{i=1}^{k} p_i = 1$

Αν η H_0 είναι σωστή, $n_i \sim B(n, p_i)$ $\left[(n_1,\dots,n_k) \sim M(n, p_1, \dots, p_k)\right]$

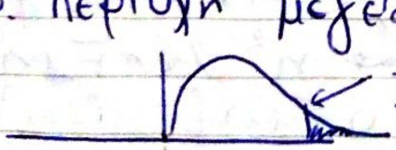
Έτσι, $e_i = n p_i$, ο αναμενόμενος αριθμός αν αληθεύει η H_0 (δηλ. η $F_0(x)$)

Κριτήριο για τον έλεγχο της H_0: $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} \overset{\text{ασυμπ}}{\approx} \chi^2_{k-1}$

$= \sum_{i=1}^{k} \frac{n_i^2}{e_i} - n \left[\sim \chi^2_{k-1}\right]$ (Θ.1.5.1 σελ.15 - Ζωγράφου)

$\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} e_i = n$.

κρ. περιοχή μεγέθους a: $\chi^2 \geq \chi^2_{a,k-1}$



χ^2_{k-1}

$\chi^2_{\alpha,k-1}$

Για S παραμετρους που εκτιμώνται τοτε Β.Ε: $k-S-1$.

παράδειγμα (7.1)

$n=90$

πλευρά:	1	2	3	4	5	6
Συχν. (n_i):	13	17	16	18	12	14
Αναμ. Συχν (e_i):	15	15	15	15	15	15

$H_0: p_i = \frac{1}{6}, i=1,\dots,6$

$H_a: p_i \neq \frac{1}{6}$

Είναι το ζάρι ισοβαρές;

λύση

$np_i = e_i = 90 \cdot \frac{1}{6} = 15$ φορές πρέπει να ρίξω το ζάρι (περιμένουμε οι αποκλίσεις από το 15 να είναι πολύ μικρές και να οφείλονται στην τύχη).

$$\chi^2 = \sum_{i=1}^{6} \frac{(n_i - e_i)^2}{e_i} = \frac{(13-15)^2}{15} + \frac{(17-15)^2}{15} + \dots + \frac{(14-15)^2}{15} = 1,87$$

$$\chi^2 > \chi^2_{0.05,5} = 11.07$$

Άρα δεν απορρ. η H_0.

Παράδειγμα

Για να διαπιστώσουμε αν το πλήθος (ο αριθμός) των μεταβολών μιας μετοχής κατά την διάρκεια μιας μέρας στο χρηματιστήριο ακολουθεί κατανομή Poisson καταγράφονται οι ημερήσιες μεταβολές της τιμής της μετοχής για ένα έτος (340 μέρες λειτουργίας). Τα δεδομένα:

μεταβλήθηκε 3 φορές, 45 μέρες

πλήθος ημερήσιων μεταβολών (μετοχής) (x)	0	1	2	3	4	5	6	≥7
Συχνότητες (n_i)	62	123	84	45	17	5	3	1

$n=340$

Ho : τα δεδομένα προέρχονται από την κατανομή Pois(λ)

Ha: Δεν

λ άγνωστο, $\hat{\lambda} = \bar{X}$.

$$\bar{X} = \frac{\Sigma x_i n_i}{n} = \frac{0\cdot 62 + \dots + 7\cdot 1}{340} = 1,6 = \hat{\lambda}$$

Άρα Ho: $F(x) = Pois\,(\lambda = 1,6)$

$$P_X(x) = \frac{e^{-1.6}(1.6)^x}{x!}, \quad x = 0, 1, \dots$$

x	$P(X=x_i)=P_i$	n_i	$e_i = n\cdot P_i$	i
0	0.2012	62	68.696	1
1	0.3230	123	109.820	2
2	0.2584	84	87.856	3
3	0.1378	45	46.852	4
4	0.0551	17	18.734	5
5	0.0176	5	5.984	6
6	0.0047	4 {3	1.972 {1.598	7
≥7	0.0011	1	0.374	8

$$X^2 = \sum_{i=1}^{7} \frac{(n_i - e_i)^2}{e_i} = \frac{(62-68.696)^2}{68.696} + \dots + \frac{(4-1.972)^2}{1.972} = 4.874$$

$$X^2 \geq X^2_{a,\,k-r-1} \left(= X^2_{0.05,\,7-1-1} = X^2_{0.05,\,5} = 11.070 \right)$$

επειδή 4.874 < 11.070 δεν απορρ Ho.

Επειδή η συχν. ω τελευταίο είναι κάτω από το 5, παίρνουμε τα δύο τελευταία μαζί