(27) 2/4

Συμπερασματολογία για το διωνυμικό p

n - δοκιμές Bernoulli

p = πιθανοτ. επιτυχ. $= P(E)$

X = αρ. E σε n-δοκιμές, $X \sim B(n,p)$ κ $E(X) = np$ κ $Var(X) = np(p-1)$

$\hat{p} = \frac{X}{n}$ κ $E(\hat{p}) = E\left(\frac{X}{n}\right) = p$ δηλ. αμερόλ.

$\sigma_{\hat{p}}^2 = Var\left(\frac{X}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \rightsquigarrow \hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$X \overset{\text{προσεγγ.}}{\sim} N(np, np(1-p))$ ή $\hat{p} \overset{\text{προσεγγ.}}{\sim} N\left(p, \frac{(1-p)p}{n}\right)$

$\frac{\hat{p}-p}{\sqrt{\frac{p(p-1)}{n}}} \overset{\text{προσεγγ.}}{\sim} N(0,1)$

Για να ελέγξουμε υποθέσεις της μορφής

(i) $H_0: p \leq p_0$ κ $H_\alpha: p > p_0$ (i) $X \geq k_\alpha$

(ii) $H_0: p \geq p_0$ κ $H_\alpha: p < p_0$ (ii) $X \leq k'_\alpha$

(iii) $H_0: p = p_0$ κ $H_\alpha: p \neq p_0$ (iii) $X \leq k'_{\alpha/2}$ ή $X \geq k_{\alpha/2}$

όπου k'_α κ k_α ο μεγαλύτερος κ μικρότερος αντίστοιχα ακέραιος με τον οποίο:

$$\sum_{x=0}^{k'_\alpha} \binom{n}{x} p_0^x (1-p_0)^{n-x} \leq \alpha \quad \text{κ} \quad \sum_{k=0}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-x} \leq \alpha$$

Παρ. ($B(n=10, p=\frac{1}{3})$)

$P(X=10) = 0{,}000017$

$P(X=9) = 0{,}000336$

$P(X=8) = 0{,}003027$

$P(X=7) = 0{,}0161668$

$\Rightarrow P(X \geq 7) = 0{,}019548 \leq 0{,}05\,(\alpha) \Rightarrow P(X \geq 6) = 0{,}076233 > 0{,}05$

$P(X=6) = 0{,}056678$

$H_0: p = \frac{1}{3}$ κ $H_\alpha: p > \frac{1}{3}$, κρ. περ.: $X \geq k_\alpha (= k_{0{,}05})$ 7

$X \overset{H_0}{\sim} B(10, \frac{1}{3}) \Rightarrow n = 10,\ p = \frac{1}{3}$

$(1-\alpha)100\%$ Δ.Ε για το p: $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Για τον έλεγχο των (i), (ii) κ (iii) προσεγγιστικά χρησιμοποιούμε το στατιστικό $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 - p_0^2}{n}}}$ κ κρ. περ.

$U - L = 2w = 2 z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{w}\right)^2 \Rightarrow$

$n \le \left(\frac{z_{\alpha/2}}{w}\right)^2 \frac{1}{4}$ γιατί $\hat{p}(1-\hat{p}) \le \frac{1}{4}$ γιατί $0 \le \hat{p} \le 1$

Παρ. 4.4

$n = 2789 <$ 2431, 358 πεθ.

p = πιθαν. επιβ. $\sim \hat{p} = \frac{2431}{2789} = 0{,}872$

95% ΔΕ : $[L, U] = [0.86, 0.884]$

$H_0: p = 0{,}9$ κ $H_a: p \neq 0{,}9$, $\alpha = 0{,}05$

$Z = \frac{0{,}872 - 0{,}9}{\sqrt{0{,}9(1-0{,}9)/2789}} = -4{,}91$

Κρ. περ. $|Z| \ge z_{\alpha/2} (= z_{0.025}) = 1{,}96$

άρα απορ. η H_0

	B +	B −	Σύνολο
A +	25	14	39
A −	4	17	21
Σύνολο	29	31	60

H_0: τα δύο μέσα το ίδιο αποτέλεσμα

Συμπερασματολογία για τη διαφορά δύο ποσοστών

A) Σύγκριση 2 ποσοστών και το Test - Τεστ του Mc-Nemar

Γενικά: τα μέλη ενός πληθυσμού υποβάλλονται σε 2 δοκιμασίες με σκοπό τη σύγκριση της αποτελεσματικότητας αυτών.

Έστω p_1 κ p_2 οι άγνωστες πιθανότητες επιτυχίας των δοκιμασιών 1 κ 2 αντίστοιχα. Στόχος μας ο έλεγχος της $H_0: p_1 = p_2$

N μέλη του πληθυσμού εκλέγονται τυχαία κ υποβάλλονται στη δοκ. 1 κ 2. Τα δεδομένα θα μπορούσαμε να τα συμβολίσουμε ως

	Δοκ 2 E	Δοκ 2 A	Σύνολο
Δοκ 1 E	X	Y	X+Y
Δοκ 1 A	Z	W	Z+W
Σύνολο	X+Z	Y+W	N

$\hat{p}_1 = \frac{X+Y}{N}$, $\hat{p}_2 = \frac{X+Z}{N}$

$$\left.\begin{array}{l} X+Y \sim B(n, p_1) \\ X+Z \sim B(n, p_2) \end{array}\right\} \Rightarrow \left.\begin{array}{l} E(X+Y) = Np_1 = E(X) + E(Y) \\ E(X+Z) = Np_2 = E(X) + E(Z) \end{array}\right\} \Rightarrow$$

$\xRightarrow{H_0} E(Y) = E(Z)$

Αν $n = Y + Z$ τότε όταν H_0 : αληθεύει $Y \sim B(n, \frac{1}{2})$ έχω
$Y \overset{\text{προσεγ.}}{\sim} N(\frac{n}{2}, \frac{n}{4})$ και έτσι το στατιστικό που χρησιμοποιώ
είναι $M = \dfrac{Y - \frac{n}{2} \pm \frac{1}{2}}{\sqrt{n}/2} \overset{\text{προσεγγ.}}{\sim} N(0,1)$

$$M = \frac{Y - \frac{n}{2}}{\sqrt{n}/2} \underset{\text{προσεγγ.}}{\overset{H_0}{\sim}} N(0,1) \to M = \frac{Y - \frac{n}{2} - \frac{1}{2}}{\sqrt{n}/2} \text{ αν } Y > \frac{n}{2}$$

$$M = \frac{Y - \frac{n}{2} + \frac{1}{2}}{\sqrt{n}/2} \text{ αν } Y < \frac{n}{2}$$

H_0: Α κ Β το ίδιο αποτέλ. $(p_1 = p_2)$
H_a: Α κ Β διαφορετικό $(p_1 \neq p_2)$

$$M = \frac{14 - \frac{18}{2} - \frac{1}{2}}{\sqrt{18}/2} = 2{,}12$$

Κρ. περ. $|M| \geq Z_{a/2} (= Z_{0,025} = 1{,}96)$ επειδή $2{,}12 > 1{,}96$ έχω απορρίπτω την H_0
p-τιμή $= P(|M| \geq 2{,}12) = 2 \cdot P(M \geq 2{,}12) = 2 \cdot 0{,}017 = 0{,}034 < 0{,}05$
Απορρίπτω την H_0

<u>Σύγκριση δύο ποσοστών χωρίς ζεύγη</u>
<u>παρ. 9.10</u>

Έχουμε 100 ασθενείς $<$ 50 φαρμ. $<$ 34, 16 ; 50 placebo $<$ 9, 41

H_0: το φάρμακο δεν είναι αποτελεσματικό (το ίδιο αποτελ. με το placebo)

<u>Γενικά</u>

Έστω δύο πληθ. ανεξάρτητοι τα μέλη των οποίων ανήκουν σε δύο κατηγορίες Α κ Β π.χ. Α η κατηγορία ένα μέλος να έχει κάποιο χαρακτηριστικό γνώρισμα α.
Έστω p_1, p_2 οι πιθανότητες των μελών των πληθυσμών 1 κ 2 να ανήκουν στην κατηγορία Α. Ένα τυχ. δείγμα μέγεθος n_1 επιλέγεται από τον πλ. 1 κ έστω X_1 μέλη να ανήκουν στην κατηγορία Α. Επίσης ένα τυχ. δείγμα μέγεθος n_2 από τον πλ. 2 κ έστω X_2 τα μέλη της κατηγορίας Α.

$\hat{p}_1 = \frac{X_1}{n_1}$ κ $X_1 \sim B(n_1, p_1)$ κ $\hat{p}_2 = \frac{X_2}{n_2}$ κ $X_2 \sim B(n_2, p_2)$

<u>Ενδιαφέρει η $p_1 - p_2$</u> : $X_1 \overset{\text{προσεγγ.}}{\sim} N(n_1 p_1, n_1 p_1 (1-p_1))$
$X_2 \overset{\text{προσεγγ.}}{\sim} N(n_2 p_2, n_2 p_2 (1-p_2))$ $\Big\} \Rightarrow$

$$\Rightarrow \hat{p}_1 - \hat{p}_2 \overset{\text{προσεγγ.}}{\sim} N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

$$\frac{\hat{p}_1-\hat{p}_2-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \rightsquigarrow$$

$\rightsquigarrow (1-\alpha)100\%$ ΔΕ για p_1-p_2: $\hat{p}_1-\hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Αν αληθεύει η $H_0: p_1=p_2=p \Rightarrow \hat{p}_1=\hat{p}_2=\hat{p}=\frac{x_1+x_2}{n_1+n_2}$

Για τον έλεγχο της H_0 χρησιμοποιούμε το στατιστικό

$$Z=\frac{\hat{p}_1-\hat{p}_2-0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \underset{\text{προσεγγ.}}{\overset{H_0}{\sim}} N(0,1)$$ κ Κρ. περιοχ. ανάλογα

Πχ (με το φάρμακο)

$\hat{p}_1=\frac{34}{50}$, $\hat{p}_2=\frac{9}{50}$, $\hat{p}=\frac{43}{100}$

$H_0: p_1=p_2$ κ $H_\alpha: p_1>p_2$

$$z=\frac{\frac{34}{50}-\frac{9}{50}}{\sqrt{\frac{43}{100}\left(1-\frac{43}{100}\right)\left(\frac{1}{50}+\frac{1}{50}\right)}}=5,05$$

$Z>Z_\alpha(=Z_{0,01}=2,326)$

$5,05>2,326$ απορρίπτεται η H_0

$P(Z\geq 5,05)\cong 0,00<0,01$ άρα απορρίπτεται η H_0