1

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΣΥΓΚΡΙΣΗ ΑΝΑΛΟΓΙΩΝ ΔΥΟ ΑΝΕΞΑΡΤΗΤΩΝ ΠΛΗΘΥΣΜΩΝ : $X \sim B(n, p_x)$, $Y \sim B(m, p_y)$

ΑΝ ΙΣΧΥΕΙ Ο ΚΑΝΟΝΑΣ ΤΟΥ "ΠΕΝΤΕ" ΤΟΤΕ ΚΑΤΑ ΠΡΟΣΕΓΓΙΣΗ ΕΧΟΥΜΕ ΑΝΑΦΕΡΕΙ ΟΤΙ

$$\hat{p}_x = \frac{X}{n} \sim N\left(p_x, \frac{p_x q_x}{n}\right) \quad \text{με } q_x = 1 - p_x$$

και

$$\hat{p}_y = \frac{Y}{m} \sim N\left(p_y, \frac{p_y q_y}{m}\right) \quad q_y = 1 - p_y$$

ΚΑΙ ΛΟΓΩ ΑΝΕΞΑΡΤΗΣΙΑΣ ΤΩΝ X, Y

$$VAR(\hat{p}_x - \hat{p}_y) = \frac{p_x q_x}{n} + \frac{p_y q_y}{m}$$

ΣΥΝΕΠΩΣ ΙΣΧΥΕΙ

$$\hat{p}_x - \hat{p}_y \sim N\left(p_x - p_y, \frac{p_x q_x}{n} + \frac{p_y q_y}{m}\right)$$

ΚΑΙ Η ΤΥΠΟΠΟΙΗΜΕΝΗ

$$Z = \frac{\hat{p}_x - \hat{p}_y - (p_x - p_y)}{\sqrt{\frac{p_x q_x}{n} + \frac{p_y q_y}{m}}} \sim N(0,1)$$

ΕΣΤΩ $H_0: p_x = p_y = p$

$H_1: p_x < p_y \qquad p_x > p_y \qquad p_x \neq p_y$

ΤΟΤΕ, ΥΠΟ H_0 : $p_x = p_y = p$ ΚΑΙ ΥΠΑΡΧΟΥΝ 2 ΕΚΤΙΜΗΤΡΙΕΣ (ΔΗΛΑΔΗ $\hat{p}_x$ ΚΑΙ $\hat{p}_y$) ΠΟΥ ΕΚΤΙΜΟΥΝ p. ΠΟΙΑ ΝΑ ΠΡΟΤΙΜΗΣΟΥΜΕ; ΕΠΙΛΕΓΟΥΜΕ ΩΣ "ΚΑΛΥΤΕΡΗ" ΕΚΤΙΜΗΤΡΙΑ ΤΟΥ p ΤΗΝ ΣΤΑΘΜΙΣΜΕΝΗ ΕΚΤΙΜΗΤΡΙΑ

$$\hat{p}_s = \frac{n\hat{p}_x + m\hat{p}_y}{n+m} = \frac{X+Y}{n+m}$$

②

* $\hat{p}_s$ ΕΙΝΑΙ ΑΜΕΡΟΛΗΠΤΗ ΕΚΤΙΜΗΤΡΙΑ ΤΟΥ p

ΥΠΟ $H_0: p_x = p_y$, ΙΣΧΥΕΙ ΟΤΙ

$$Z = Z_0 = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_s \hat{q}_s}{n} + \frac{\hat{p}_s \hat{q}_s}{m}}} = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}_s \hat{q}_s \left(\frac{1}{n} + \frac{1}{m}\right)}}$$

(ΕΧΟΥΜΕ ΑΝΤΙΚΑΤΑΣΤΑΣΕΙ p_s ΜΕ $\hat{p}_s$ ΚΑΙ q_s ΜΕ $\hat{q}_s$ ΣΤΟΝ ΠΑΡΟΝΟΜΑΣΤΗ ΑΦΟΥ ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΤΗΣ ΠΡΟΣΕΓΓΙΣΗΣ ΜΕ ΤΗΝ ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ ΕΙΝΑΙ ΑΣΥΜΠΤΩΤΙΚΟ)

ΚΑΝΟΝΕΣ ΑΠΟΦΑΣΗΣ

$H_1: p_x < p_y$, $\cancel{H_0}$ ΑΝ $Z_0 < -Z_{1-\alpha}$

$H_1: p_x > p_y$, $\cancel{H_0}$ ΑΝ $Z_0 > Z_{1-\alpha}$

$H_1: p_x \neq p_y$, $\cancel{H_0}$ ΑΝ $Z_0 > Z_{1-\frac{\alpha}{2}}$
Ή $Z_0 < -Z_{1-\frac{\alpha}{2}}$

③

ΠΑΡΑΔΕΙΓΜΑ

Από τυχαίο δείγμα 400 φοιτητών (ανδρών) Χ

προέκυψε ποσοστό καπνιστών 30%

ενώ από τυχαίο δείγμα 300

φοιτητριών (γυναίκες) Υ προέκυψε

ποσοστό καπνιστριών 20%.

Να ελέγξει

$H_0: p_x = p_y$ σε επίπεδο σημαντικότητας

$H_1: p_x > p_y$ $\alpha = 5\%$

1) Εφαρμόζουμε κανόνα πέντε ξεχωριστά στις 2 ομάδες

$\hat{p}_x = 0.3$

$\hat{p}_y = 0.2$

$n\hat{p}_x = 400(0.3) > 5$

$n\hat{q}_x = 400(0.7) > 5$

$m\hat{p}_y = 300(0.2) > 5$

$m(\hat{q}_y) = 300(0.8) > 5$

→ ισχύει προσέγγιση από N κατανομή

επιπλέον αφού n και m > 30

βελτιώνεται η προσέγγιση αυτή.

2) υπολογίζουμε $\hat{p}_s = \dfrac{400(0.3)+300(0.2)}{400+300}$

$= 0.26$

και υπολογίζουμε

$$Z_0 = \frac{0.3-0.2}{\sqrt{0.26(1-0.26)\left(\frac{1}{400}+\frac{1}{300}\right)}} = 2.8$$

(4)

3. ΕΦΑΡΜΟΖΟΥΜΕ ΚΑΝΟΝΑ ΑΠΟΦΑΣΗΣ

ΑΠΟ ΠΙΝΑΚΕΣ $N(0,1)$: $Z_{0.95} = 1.645$

ΑΦΟΥ $Z_0 = 2.8 > Z_{0.95} = 1.645$

$\Rightarrow$ ΑΠΟΡΡΙΠΤΕΤΑΙ H_0 ($\cancel{H_0}$).

ΔΗΛΑΔΗ ΥΠΑΡΧΟΥΝ ΑΡΚΕΤΑ ΙΣΧΥΡΕΣ ΕΝΔΕΙΞΕΙΣ ΥΠΕΡ ΤΟΥ ΟΤΙ ΤΟ ΠΟΣΟΣΤΟ ΚΑΠΝΙΣΤΩΝ ΣΤΟΥΣ ΑΝΔΡΕΣ ΕΙΝΑΙ ΣΗΜΑΝΤΙΚΑ ΜΕΓΑΛΥΤΕΡΟ ΑΠΟ ΑΥΤΟ ΣΤΙΣ ΓΥΝΑΙΚΕΣ