ΑΠΑΝΤΗΣΕΙΣ ΘΕΜΑΤΩΝ 2

ΕΞΕΤΑΣΤΙΚΗΣ ΦΕΒΡΟΥΑΡΙΟΥ 2016

ΑΝΑΛΥΣΗ ΚΑΤΗΓΟΡΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Σε έλεγχο μεταξύ δύο φαρμάκων π είναι η πιθανότητα το ένα εκ των δύο να είναι καλύτερο. Αν θέλουμε να εκτιμήσουμε το π κάτω από την υπόθεση H_0 :π=0.5 έναντι της εναλλακτικής H_1 :π≠0.5, τότε με μέγεθος δείγματος 20 παρατηρήσεις υπολογίστε τα παρακάτω: α. Βρέστε και σκιαγραφήστε την πιθανοφάνεια, Δώστε την εκτίμηση μέγιστης πιθανοφάνειας για το π, β. υπολογίστε το 95% δ.ε του π μαζί με τον υπολογισμό της αντίστοιχης ελεγχοσυνάρτησης, γ. υπολογίστε τον λόγο πιθανοφάνειας καθώς και το 95% δ.ε.

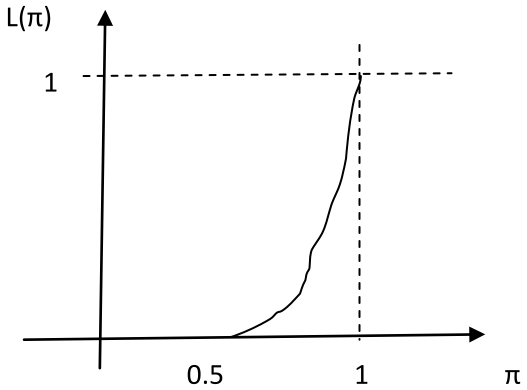
Απάντηση

α. Με βάση το γεγονός ότι η κατανομή των φαρμάκων ακολουθεί δυωνυμική κατανομή, η πιθανοφάνεια δίνεται από την σχέση

$$L(\pi) = \binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

Στο συγκεκριμένο παράδειγμα n=20 και x=20 (αφού κάθε φορά το νέο φάρμακο είναι καλύτερο), άρα

$$L(\pi) = \binom{20}{20}\pi^{20}(1-\pi)^{20-20} = \pi^{20}$$



Εκτίμηση μέγιστης πιθανοφάνειας

$$\hat{\pi}_{\text{MLE}} = \frac{x}{n} = \frac{20}{20} = 1$$

β. Υπολογισμός της ελεγχοσυνάρτησης με βάση τα ποσοστά

$$z^{*} = \frac{\hat{\pi}-\pi_0}{\sqrt{\dfrac{\hat{\pi}\left(1-\hat{\pi}\right)}{n}}} = \frac{1-0.5}{\sqrt{\dfrac{1\left(1-1\right)}{20}}} = \infty$$

με $p\left(z > z^{*}\right) = 0$ και 95% διάστημα εμπιστοσύνης

$$\left[\hat{\pi} - z_{a/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{a/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right] = [1,1]$$

γ. Για τον λόγο πιθανοφανειών έχουμε:

Η πιθανοφάνεια της δυωνυμικής κατανομής δίνεται από την σχέση:

$$L(\pi) = \log\left[\pi^x (1-\pi)^{n-x}\right] = x\log(\pi) + (n-x)\log(1-\pi)$$

Κάτω από την H_0 η αντίστοιχη πιθανοφάνεια δίνεται από την σχέση:

$$L_0(\pi) = x\log(\pi_0) + (n-x)\log(1-\pi_0)$$

και κάτω από την H_0 η αντίστοιχη πιθανοφάνεια δίνεται από την σχέση:

$$L_1(\hat{\pi}) = x\log(\hat{\pi}) + (n-x)\log(1-\hat{\pi})$$

Επομένως ο λόγος πιθανοφανειών δίνεται από την σχέση:

$$-2(L_0 - L_1) = 2\left[x\log\frac{x}{n\pi_0} + (n-x)\log\frac{n-x}{n-n\pi_0}\right]$$

Άρα

$$G^2 = -2(L_0 - L_1) = 2\left[20\log\frac{20}{10} + (20-20)\log\frac{20-20}{20-10}\right] = 27.7$$

Το $G^2 \rightarrow x^2_{1,0.005}$ άρα 27.7>3.84 και επομένως η H_0 απορρίπτεται

Μελέτη έδειξε ότι η πιθανότητα ένα νεογέννητο παιδί να πέσει θύμα φόνου είναι 0.0263 για τα μη λευκά αγόρια και 0.0049 για τα λευκά αγόρια, ενώ 0.0072 για τα μη λευκά κορίτσια και 0.0023 για τα λευκά κορίτσια. Να βρεθούν: α. ο λόγος υπεροχής μεταξύ των δύο φυλών καθώς και το 95% δ.ε. β. αν μισά από τα νεογνά ανήκουν σε διαφορετικό φύλο για κάθε φυλή να υπολογισθεί ο λόγος περιθωρίων σχετικών πιθανοτήτων μεταξύ των φυλών για θύματα φόνου.

Απάντηση

α. Για τους άνδρες ο λόγος υπεροχής δίνεται από την σχέση:

$$\widehat{OR} = \frac{0.0049 / (1-0.0049)}{0.0263 / (1-0.0263)} = \frac{0.0049 / 0.9951}{0.0263 / 0.9737} = 0.18$$

που σημαίνει ότι ο λόγος υπεροχής λευκών ανδρών να βρεθούν δολοφονημένοι είναι 0.18 φορές το λόγο υπεροχής μη λευκών ανδρών να βρεθούν δολοφονημένοι.

Για τις γυναίκες ο λόγος υπεροχής δίνεται από την σχέση:

$$\widehat{OR} = \frac{0.0023 / (1-0.0023)}{0.0072 / (1-0.0072)} = \frac{0.0023 / 0.9977}{0.0072 / 0.9928} = 0.32$$

που σημαίνει ότι ο λόγος υπεροχής λευκών γυναικών να βρεθούν δολοφονημένες είναι 0.32 φορές το λόγο υπεροχής μη λευκών γυναικών να βρεθούν δολοφονημένες.

β. Αν ορίσουμε P(M)=0.5 την πιθανότητα το φύλο να είναι άνδρας, P(F)=0.5 το φύλο να είναι γυναίκα και V={θύμα δολοφονίας}, W={λευκός}, τότε έχουμε:

$$P(V \mid \bar{W}) = P(V \mid \bar{W}, M)P(M) + P(V \mid \bar{W}, \bar{M})P(\bar{M}) =$$
$$= 0.0263 * 0.5 + 0.0072 * 0.5 = 0.01675$$
$$P(V \mid W) = 0.049 * 0.5 + 0.0023 * 0.5 = 0.0036$$

Ο λόγος των περιθωρίων σχετικών πιθανοτήτων (λόγος υπεροχής) δίνεται από την σχέση:

$$\widehat{OR} = \frac{0.0036 / (1-0.0036)}{0.01675 / (1-0.01675)} = \frac{0.0036 / 0.9964}{0.01675 / 0.8325} = 0.179$$

Αν P(E')= 1-P(E), να αποδείξετε

α.AR=[P(D)-P(D|E')]/P(D) και

β. AR=[P(E)(RR-1)]/[1+P(E)(RR-1)]

Απάντηση

α. Ο αριθμητής ερμηνεύει πόσοι παραπάνω είχαν την ασθένεια από αυτούς που δεν είχαν εκτεθεί. Άρα

$$AR=\frac{P(D)-P(D|E')}{P(D)}$$

β.

$$AR=\frac{P(E)(RR-1)}{1+P(E)(RR-1)} \Rightarrow AR+AR*P(E)*(RR-1)=P(E)*(RR-1) \Rightarrow$$

$$\Rightarrow RR*\left[AR*P(E)-P(E)\right]=AR*P(E)-AR-P(E) \Rightarrow$$

$$\Rightarrow RR=\frac{AR*P(E)-AR-P(E)}{AR*P(E)-P(E)}=\frac{AR*P(E)-AR-P(E)}{P(E)*(AR-1)} \Rightarrow$$

$$\Rightarrow RR=\frac{\frac{P(D)-P(D|E')}{P(D)}*\left[P(E)-1\right]-P(E)}{P(E)*\left[-\frac{P(D|E')}{P(D)}\right]} \quad \text{(με βάση την αντικατάσταση από το (α))} \Rightarrow$$

$$\Rightarrow RR=\frac{\left[P(D)-P(D|E')\right]*\left[P(E)-1\right]-P(E)*P(D)}{-P(E)*P(D|E')} \Rightarrow$$

$$\Rightarrow RR=\frac{P(E)*P(D)-P(D)-P(D|E')*P(E)+P(D|E')-P(E)*P(D)}{-P(E)*P(D|E')} \Rightarrow$$

$$\Rightarrow RR=\frac{-P(D)+P(D|E')*\left[1-P(E)\right]}{-P(E)*P(D|E')} \Rightarrow$$

$$\Rightarrow RR=\frac{P(D)-P(D|E')*P(E')}{P(E)*P(D|E')} \Rightarrow RR=\frac{P(E)*P(D|E)}{P(E)*P(D|E')} \Rightarrow RR=\frac{P(D|E)}{P(D|E')}$$

άρα ισχύει