Μέτρα Κινδύνου για Δίτιμα – Κατηγορικά Δεδομένα

Σε αυτή την ενότητα θα ορίσουμε δείκτες μέτρησης του κινδύνου εμφάνισης μίας νόσου όταν έχουμε δίτιμες κατηγορικές μεταβλητές.

Στην πιο απλή περίπτωση μας ενδιαφέρει να συγκρίνουμε τον κίνδυνο εμφάνισης μιας νόσου σε ένα άτομο που έχει εκτεθεί στον παράγοντα κινδύνου με τον αντίστοιχο κίνδυνο ενός ατόμου που δεν έχει εκτεθεί στον ίδιο παράγοντα κινδύνου.

	Y: Νόσος		
X: Παράγοντας Κινδύνου	1 (Ασθενής)	2 (Υγιής)	Σύνολο
1 (Παρόν)	n_{11}	n_{12}	$n_{1.}$
2 (Απών)	n_{21}	n_{22}	$n_{2.}$
Σύνολο	$n_{.1}$	$n_{.2}$	$n_{..} = n$

Αποδιδόμενος Κίνδυνος (Attributable Risk, AR)

Ως αποδιδόμενο κίνδυνο ορίζουμε τη διαφορά μεταξύ των ρυθμών επίπτωσης (ή των δεικτών θνησιμότητας) των ομάδων με άτομα εκτεθειμένα και μη εκτεθειμένα σε ένα παράγοντα κινδύνου.

Όταν αναφερόμαστε σε ένα πίνακα συνάφειας 2×2 τότε ο αποδιδόμενος κίνδυνος ορίζεται ως η διαφορά των δεσμευμένων πιθανοτήτων $\pi_{j=1|i=1}$ και $\pi_{j=1|i=2}$

$$AR = \pi_E - \pi_{\bar{E}} = P(A|E) - P(A|\bar{E}) = \pi_{j=1|i=1} - \pi_{j=1|i=2} = \pi_{j=2|i=2} - \pi_{j=2|i=1}$$

$\pi_E = P(A|E)$

$\pi_{\bar{E}} = P(A|\bar{E})$

πιθανότητες εμφάνισης της νόσου αν κάποιος ασθενής (A) έχει εκτεθεί στον κίνδυνο (E) ή όχι ($\bar{E}$)

Αποδιδόμενος Κίνδυνος

Εκτίμηση

$$AR = p_E - p_{\bar{E}} = p_{j=1|i=1} - p_{j=1|i=2} = \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}} = \frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{21}}{n_{21}+n_{22}}$$

X: Παράγοντας Κινδύνου	*Y*: Νόσος 1 (Ασθενής)	2 (Υγιής)	Σύνολο
1 (Παρόν)	n_{11}	n_{12}	$n_{1.}$
2 (Απών)	n_{21}	n_{22}	$n_{2.}$
Σύνολο	$n_{.1}$	$n_{.2}$	$n_{..} = n$

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i.}}$$

Αποδιδόμενος Κίνδυνος

έλεγχος υποθέσεων (ανεξαρτησίας)

- $H_0 : AR = 0$ έναντι της εναλλακτικής
- $H_1 : AR \neq 0$.

Το πρόβλημα με τον παραπάνω δείκτη είναι ότι μετράει απόλυτες διαφορές. Έτσι οι πιθανότητες 0.10 και 0.20 θα μας δώσουν ίδιο *AR* με τις πιθανότητες 0.6 και 0.7 αγνοώντας το γεγονός ότι στην πρώτη περίπτωση η πιθανότητα της μίας ομάδας είναι διπλάσια από την άλλη.

$$p_{AR} = \frac{\pi_E - \pi_{\bar{E}}}{\pi_E} = \frac{P(A|E) - P(A|\bar{E})}{P(A|E)} = \frac{\pi_{j=1|i=1} - \pi_{j=1|i=2}}{\pi_{j=1|i=1}} = \frac{AR}{\pi_E}$$

Σχετικός Κίνδυνος (Relative Risk, RR)

ορίζεται ως ο λόγος της επίπτωσης ή της θνησιμότητας δύο ομάδων με διαφορετική έκθεση στον παράγοντα κινδύνου

Σε πίνακες 2x2 δίνεται από το λόγο των πιθανοτήτων

$\pi_{j=1|i=1}$ και $\pi_{j=1|i=2}$

$$RR = \frac{\pi_E}{\pi_{\bar{E}}} = \frac{P(A|E)}{P(A|\bar{E})} = \frac{\pi_{j=1|i=1}}{\pi_{j=1|i=2}} = \frac{1-\pi_{j=2|i=1}}{1-\pi_{j=2|i=2}}$$

$$RR = \frac{p_E}{p_{\bar{E}}} = \frac{p_{j=1|i=1}}{p_{j=1|i=2}} = \frac{\frac{n_{11}}{n_{1.}}}{\frac{n_{21}}{n_{2.}}}$$

Σχετικός Κίνδυνος (Relative Risk, RR)

έλεγχος υποθέσεων (ανεξαρτησίας)

- H_0 : RR = 1 έναντι της εναλλακτικής
- H_1 : RR ≠ 1

μοναδιαίος σχετικός κίνδυνος συνεπάγεται ανεξαρτησία μεταξύ των δύο μεταβλητών.

Σχετικός Κίνδυνος (Relative Risk, RR)

Αν RR = a τότε

- Η πιθανότητα εμφάνισης της νόσου $(Y = 1)$ όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου $(X = 1)$ είναι ίση με a φορές την ίδια πιθανότητα όταν δεν εκτεθεί στον παράγοντα κινδύνου $(X = 2)$.
- Αν a > 1 τότε: Η πιθανότητα εμφάνισης της νόσου $(Y = 1)$ όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου $(X = 1)$ είναι ίση με a − 1 φορές μεγαλύτερη (ή (a − 1)100% φορές μεγαλύτερη) από την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο $(X = 2)$.

Σχετικός Κίνδυνος (Relative Risk, RR)

- Αν a < 1 τότε: Η πιθανότητα εμφάνισης της νόσου $(Y = 1)$ όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου $(X = 1)$ είναι ίση με 1 − a φορές μικρότερη (ή (1 − a) 100% φορές μικρότερη) από την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο $(X = 2)$ (η μεταβλητή ας εδώ λέγεται προστατευτικός παράγοντας).
- Αν a = 1 τότε: Η πιθανότητα εμφάνισης της νόσου $(Y = 1)$ όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου $(X = 1)$ είναι ίση με την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο $(X = 2)$ (δηλαδή η μεταβλητή X δεν επηρεάζει την εμφάνιση της νόσου συνεπώς δεν αποτελεί παράγοντα κινδύνου).

Λόγος Σχετικών Πιθανοτήτων

- *Σχετική Πιθανότητα (Odds)*

υποκαθιστά το όρο «probability» (πιθανότητα) και συνήθως δίνει μια εκτίμηση της τύχης που έχει κάποιος να κερδίσει σε ένα αγώνα ή στοίχημα.

Μαθηματικά το odds ενός ενδεχομένου *A* δίνεται από τον τύπο:

$$Odds(A) = \frac{P(A)}{1 - P(A)}$$

- λόγος της πιθανότητας εμφάνισης ενός ενδεχομένου έναντι της πιθανότητας μη εμφάνισης του (δηλαδή η πιθανότητα του συμπληρωματικού, ως προς το δειγματικό χώρο, ενδεχομένου).

Λόγος Σχετικών Πιθανοτήτων

η σχετική πιθανότητα είναι 1 – 1 μετασχηματισμός της αρχικής πιθανότητας εφόσον:

$$odds = \frac{\pi}{1-\pi} \Leftrightarrow \pi = \frac{odds}{1-odds}$$

Η ερμηνεία της θεωρείται πιο εύκολη από την απλή πιθανότητα διότι συγκρίνει την πιθανότητα εμφάνισης ενός ενδεχομένου με την πιθανότητα μη εμφάνισης.

Έτσι όταν η Ελληνική ομάδα είχε σχετική πιθανότητα να μην κερδίσει το ευρωπαϊκό πρωτάθλημα του 2004 80: 1 αυτό σήμαινε ότι πιθανότητα να μην κερδίσει το πρωτάθλημα ήταν ίση με 80 φορές την πιθανότητα να κερδίσει. Πιο απλά η πιθανότητα να κερδίσει ήταν μόλις (1/(80 + 1) =) 0.0123 .

Λόγος Σχετικών Πιθανοτήτων

- Αν *odds* =1 (ή 1 : 1) τότε οι πιθανότητες εμφάνισης ή μη εμφάνισης του ενδεχομένου που εξετάζουμε είναι ίσες (δηλαδή 50%).
- Αν *odds* = a τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι ίση με a φορές την πιθανότητα μη εμφάνισής του.
- Αν *odds* = a και

i. a > 1 τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι a − 1 (ή (a − 1)100%) φορές μεγαλύτερη από την πιθανότητα μη εμφάνισης του.

ii. a < 1 τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι 1 − a (ή (1 − a) 100%) φορές μικρότερη από την πιθανότητα μη εμφάνισης του.

Λόγος Σχετικών Πιθανοτήτων

- odds =1 συνεπάγεται ίση πιθανότητα εμφάνισης και μη εμφάνισης της νόσου.
- odds > 1 συνεπάγεται η εμφάνιση της νόσου είναι πιο πιθανή από την πιθανότητα μη εμφάνισης της νόσου.
- odds < 1 συνεπάγεται η εμφάνιση της νόσου είναι λιγότερο πιθανή από την πιθανότητα μη εμφάνισης της νόσου

Λόγος Σχετικών Πιθανοτήτων

2×2 πίνακες συνάφειας

$$odds(X=1)=\frac{\pi_{1|i=1}}{\pi_{2|i=1}}=\frac{\pi_{1|i=1}}{1-\pi_{1|i=1}}=\frac{\pi_{11}}{\pi_{12}}$$

$$odds(X=2)=\frac{\pi_{1|i=2}}{\pi_{2|i=2}}=\frac{\pi_{1|i=2}}{1-\pi_{1|i=2}}=\frac{\pi_{21}}{\pi_{22}}$$

$$odds(X=1)=\frac{p_{1|i=1}}{p_{2|i=1}}=\frac{\frac{n_{11}}{n_{1.}}}{\frac{n_{12}}{n_{1.}}}=\frac{n_{11}}{n_{12}}$$

$$odds(X=2)=\frac{p_{1|i=2}}{p_{2|i=2}}=\frac{\frac{n_{21}}{n_{2.}}}{\frac{n_{22}}{n_{2.}}}=\frac{n_{21}}{n_{22}}$$

Λόγος Σχετικών Πιθανοτήτων

μας ενδιαφέρουν οι σχετικές πιθανότητες εμφάνισης της νόσου όταν έχουμε έκθεση ή όχι στον παράγοντα κινδύνου.

$$odds_E = \frac{P(A|E)}{1-P(A|E)} = \frac{\pi_E}{1-\pi_E}$$

$$odds_{\overline{E}} = \frac{P(A|\overline{E})}{1-P(A|\overline{E})} = \frac{\pi_{\overline{E}}}{1-\pi_{\overline{E}}}$$

$$odds_E = \frac{p_E}{1-p_E}$$

$$odds_E = \frac{p_{\overline{E}}}{1-p_{\overline{E}}}$$

Σύγκριση Σχετικών Πιθανοτήτων

λόγος σχετικών πιθανοτήτων του *X* = 1 έναντι του *X* = 2 δίνεται από τον τύπο

$$OR = \frac{odds(X=1)}{odds(X=2)} = \frac{\pi_{1|i=1} / \pi_{2|i=1}}{\pi_{1|i=2} / \pi_{2|i=2}} = \frac{\pi_{11} / \pi_{12}}{\pi_{21} / \pi_{22}} = \frac{\pi_{11} \cdot \pi_{22}}{\pi_{21} \cdot \pi_{12}}$$

λόγος σταυρωτού ή χιαστού πολλαπλάσιου

- Στους 2×2 πίνακες συνάφειας ο λόγος σχετικών πιθανοτήτων (ΛΣΠ) εκτιμάται από τον εκτιμητή:

$$OR = \frac{p_{11} \cdot p_{22}}{p_{21} \cdot p_{12}} = \frac{n_{11} \cdot n_{22}}{n_{21} \cdot n_{12}}$$

Σύγκριση Σχετικών Πιθανοτήτων

Αν θέλουμε να ορίσουμε το ΛΣΠ ως συνάρτηση των πιθανοτήτων εμφάνισης της νόσου τότε:

$$OR = \frac{\pi_E}{1-\pi_E} \times \frac{\pi_{\bar{E}}}{1-\pi_{\bar{E}}} \quad \Rightarrow \quad OR = \frac{p_E}{1-p_E} \times \frac{p_{\bar{E}}}{1-p_{\bar{E}}}$$

- Η σύνδεση των ΛΣΠ *(OR)* με την έννοια της ανεξαρτησίας σε πίνακες 2×2 είναι άμεση αφού:

«ανεξαρτησία» $\Rightarrow$ $OR = 1$

- H_0: $OR = 1$ έναντι της εναλλακτικής
- H_1: $OR \neq 1$.

Σύγκριση Σχετικών Πιθανοτήτων

- Αν OR = a τότε η σχετική πιθανότητα της νόσου (Y = 1) όταν ένα άτομο εκτεθεί στον κίνδυνο (X = 1) είναι ίση με a φορές την ίδια σχετική πιθανότητα όταν δεν εκτεθεί στον κίνδυνο (X = 2). Ή ισοδύναμα: η σχετική πιθανότητα μη εμφάνισης της νόσου (Y = 2) όταν ένα άτομο εκτεθεί στον κίνδυνο (X = 2) είναι ίση με a φορές την ίδια σχετική πιθανότητα όταν εκτεθεί στον κίνδυνο (X = 1).
- Αν OR = a > 1 τότε η σχετική πιθανότητα της νόσου (Y = 1) όταν ένα άτομο εκτεθεί στον κίνδυνο (X = 1) είναι a – 1 (ή {a – 1}100%) φορές μεγαλύτερη της ίδιας σχετικής πιθανότητας όταν δεν εκτεθεί στον κίνδυνο (X = 2).
- Αν OR = a < 1 τότε η σχετική πιθανότητα της νόσου (Y = 1) όταν ένα άτομο εκτεθεί στον κίνδυνο (X = 1) είναι 1 – a (ή {1 – a}100%) φορές μικρότερη της ίδιας σχετικής πιθανότητας όταν δεν εκτεθεί στον κίνδυνο (X = 2).

Σύγκριση Σχετικών Πιθανοτήτων

- Στην στατιστική πολλές φορές χρησιμοποιούμε και το λογάριθμο του ΛΣΠ. Ο λόγος είναι ότι μπορούμε να εξετάσουμε την κατανομή του πιο εύκολα. Επιπλέον ο λογάριθμος είναι αυτός που επίσης χρησιμοποιείται στα μοντέλα λογιστικής παλινδρόμησης.
- H_0: $\log OR = 0$ έναντι της εναλλακτικής
- H_1: $\log OR \neq 0$.

Σχέση Σχετικού Κινδύνου και ΛΣΠ

- ο ΛΣΠ είναι ασυμπτωτικά ίσος με το σχετικό κίνδυνο όταν η πιθανότητα εμφάνισης της νόσου είναι μικρή δηλαδή στις σπάνιες αρρώστιες.

$$OR = \frac{Odds_E}{Odds_{\bar{E}}} = \frac{Odds(X=1)}{Odds(X=2)} = \frac{\pi_{j=1|1} / \pi_{j=2|1}}{\pi_{j=1|2} / \pi_{j=2|2}} = \frac{\pi_{j=1|1}}{\pi_{j=1|2}} \times \frac{\pi_{j=2|2}}{\pi_{j=2|1}} = \frac{RR(Y=1)}{RR(Y=2)}.$$

- Αν το γεγονός που εξετάζουμε έχει πιθανότητα εμφάνισης πολύ μικρή (για παράδειγμα μια σπάνια ασθένεια) τότε $\pi_{j=2|1} \approx 1$ και $\pi_{j=2|2} \approx 1$ άρα και $RR(Y = 2) \approx 1$ καταλήγοντας στο αποτέλεσμα $OR \approx RR(Y = 1)$.

Σχέση Σχετικού Κινδύνου και ΛΣΠ

$$OR = \frac{\pi_{1|i=1} / \pi_{2|i=1}}{\pi_{1|i=2} / \pi_{2|i=2}} = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}} = \frac{\pi_{i=1|1}\pi_{1.}\pi_{i=2|2}\pi_{2.}}{\pi_{i=1|2}\pi_{2.}\pi_{i=2|1}\pi_{1.}} \Leftrightarrow OR = \frac{\pi_{i=1|1} / \pi_{i=2|2}}{\pi_{i=1|2} / \pi_{i=2|1}}$$

Συνεπώς ο ΛΣΠ ορίζεται επαρκώς αν αντί για τις δεσμευμένες πιθανότητες $\pi_{j=1|i}$ για i = 1, 2 γνωρίζουμε τις πιθανότητες $\pi_{i|j=1}$ για i = 1, 2 οι οποίες μπορούν να εκτιμηθούν από μια μελέτη μαρτύρων-ασθενών.

Παράδειγμα

- Σε μια προοπτική μελέτη κατά την οποία εξετάσθηκαν 368 άνδρες καπνιστές ηλικίας κάτω των 60 ετών οι οποίοι έπαθαν μια καρδιακή ανακοπή και επιβίωσαν. Μετά από 2 έτη εξετάσθηκαν πόσοι από αυτούς είχαν επιβιώσει και τους χωρίσαμε ανάλογα εάν είχαν κόψει το τσιγάρο ή όχι. Έτσι εδώ μας ενδιαφέρει να εξετάσουμε αν το σταμάτημα του καπνίσματος (X) είχε ευνοϊκή επίδραση στην επιβίωση μετά από δύο έτη (Y). Τα δεδομένα δίνονται στον 2×2 Πίνακα που ακολουθεί:

X:	Y: Επιβίωση σε 2 χρόνια		
Συνέχισαν το κάπνισμα	1: Πεθαμένος	2: Ζωντανός	Σύνολο
1: Ναι	19 (12.3%)	135 (87.7%)	154 (41.8%)
2: Όχι	15 (7.0%)	199 (93.0%)	214 (58.2%)
Σύνολο	34 (9.2%)	334 (90.8%)	368

Παράδειγμα

αποδιδόμενος κίνδυνος

$$AR = \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}} = \frac{19}{154} - \frac{15}{214} = 0.123 - 0.070 = 0.053.$$

Συνεπώς οι ασθενείς που συνέχισαν να καπνίζουν μετά το καρδιακό επεισόδιο παρουσιάζουν αυξημένη πιθανότητα θανάτου, σε σχέση με της που διέκοψαν το κάπνισμα, κατά 5.3 ποσοστιαίες μονάδες.

Αυτή η διαφορά αποδίδεται στο σταμάτημα του καπνίσματος. Το πλεονέκτημα του αποδιδόμενου κινδύνου είναι ότι δίνει τη διαφορά σε ποσοστιαίες μονάδες κάτι που δε φαίνεται στο σχετικό κίνδυνο

Παράδειγμα

σχετικός κίνδυνος

$$RR = \frac{\frac{n_{11}}{n_{1.}}}{\frac{n_{21}}{n_{2.}}} = \frac{\frac{19}{154}}{\frac{15}{214}} = \frac{0.123}{0.070} = 1.757.$$

Αυτό σημαίνει ότι ο κίνδυνος θανάτου για της ασθενείς που συνέχισαν το κάπνισμα είναι ίσος με 1.76 φορές (ή 76% μεγαλύτερος από) τον ίδιο κίνδυνο των ασθενών που σταμάτησαν το κάπνισμα.

Παράδειγμα

- Ο ποσοστιαίος αποδιδόμενος κίνδυνος είναι ίσος με 0.053/0.123 = 43.1%. Αυτό σημαίνει ότι το 43% του κινδύνου θανάτου που διατρέχει της ασθενής που συνέχισε το κάπνισμα μπορεί να αποδοθεί στο γεγονός ότι δε σταμάτησε το κάπνισμα.
- Συνεπώς, το 43% των θανάτων (2 χρόνια μετά το πρώτο καρδιακό επεισόδιο) των ασθενών που συνέχισαν να καπνίζουν θα μπορούσε να είχε αποφευχθεί αν είχαμε μπορέσει να της πείσουμε να κόψουν το κάπνισμα (δηλαδή περίπου 8 άτομα, 19 × 0.431 = 8.1).

Παράδειγμα

ΛΣΠ είναι ίσος

$$OR = \frac{n_{11} \cdot n_{22}}{n_{21} \cdot n_{12}} = \frac{19 \cdot 199}{15 \cdot 135} = \frac{3781}{2025} = 1.867.$$

Δηλαδή η σχετική πιθανότητα θανάτου για της ασθενείς που συνέχισαν το κάπνισμα είναι 86% μεγαλύτερη από την πιθανότητα θανάτου των ασθενών που σταμάτησαν το κάπνισμα.

Βλέπουμε εδώ, ότι παρόλο που η πιθανότητα θανάτου είναι λίγο μεγαλύτερη του 10% για τη μία ομάδα, ο ΛΣΠ πλησιάζει αρκετά το σχετικό κίνδυνο.

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

- **Γενική Προσέγγιση:** Της υποθέσουμε ότι έχουμε δύο ομάδες με διαφορετική έκθεση στον κίνδυνο *Ε* και $\bar{E}$ και ότι η μεταβλητή απόκρισης *Υ* είναι της δίτιμη: έχει ή όχι τη νόσο *(Α* και $\bar{A}$ *).* Σε αυτή την περίπτωση της ενδιαφέρει να ελέγξουμε κατά πόσο η πιθανότητα εμφάνισης της νόσου είναι ίδια της δύο ομάδες.

$$\left.\begin{array}{c} H_0 : P(A|E) = P(A|\bar{E}) \\ \text{έναντι της εναλλακτικής} \\ H_0 : P(A|E) \neq P(A|\bar{E}) \end{array}\right\} \Leftrightarrow \left.\begin{array}{c} H_0 : P(A|E) - P(A|\bar{E}) = 0 \\ \text{έναντι της εναλλακτικής} \\ H_0 : P(A|E) - P(A|\bar{E}) \neq 0 \end{array}\right\} \Leftrightarrow$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

$$\left.\begin{array}{c} H_0 : \pi_E = \pi_{\overline{E}} \\ \acute{\varepsilon}\nu\alpha\nu\tau\iota\ \tau\eta\varsigma\ \varepsilon\nu\nu\alpha\lambda\alpha\kappa\tau\iota\kappa\acute{\eta}\varsigma \\ H_0 : \pi_E \neq \pi_{\overline{E}} \end{array}\right\} \Leftrightarrow \qquad \left.\begin{array}{c} H_0 : \pi_E - \pi_{\overline{E}} = 0 \\ \acute{\varepsilon}\nu\alpha\nu\tau\iota\ \tau\eta\varsigma\ \varepsilon\nu\nu\alpha\lambda\alpha\kappa\tau\iota\kappa\acute{\eta}\varsigma \\ H_0 : \pi_E - \pi_{\overline{E}} \neq 0 \end{array}\right\} \Leftrightarrow$$

$$\begin{array}{c} H_0 : AR = 0 \\ \acute{\varepsilon}\nu\alpha\nu\tau\iota\ \tau\eta\varsigma\ \varepsilon\nu\nu\alpha\lambda\alpha\kappa\tau\iota\kappa\acute{\eta}\varsigma \\ H_0 : AR \neq 0 \end{array}$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

- Συνεπώς μπορούμε να υποθέσουμε ότι ο αριθμός των ατόμων που έχουν τη νόσο σε κάθε ομάδα έκθεσης ακολουθεί διωνυμική κατανομή οπότε

$$Y \mid E \sim Bin(\pi_E, n_E) \qquad Y \mid \bar{E} \sim Bin(\pi_{\bar{E}}, n_{\bar{E}})$$

- Επιπλέον μπορούμε να χρησιμοποιήσουμε το κεντρικό οριακό θεώρημα και να φτιάξουμε ένα z– *test*. Αν οι αναμενόμενες τιμές $n_E\pi_E$, $n_E(1-\pi_E)$, $n_{\bar{E}}\pi_{\bar{E}}$ και $n_{\bar{E}}(1-\pi_{\bar{E}})$ είναι μεγαλύτερες του πέντε τότε

$$p_E \sim N\left(\pi_E, \frac{\pi_E(1-\pi_E)}{n_E}\right) \qquad p_{\bar{E}} \sim N\left(\pi_{\bar{E}}, \frac{\pi_{\bar{E}}(1-\pi_{\bar{E}})}{n_{\bar{E}}}\right)$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

- Συνεπώς

$$AR = p_E - p_{\overline{E}} \sim N\left(\pi_E - \pi_{\overline{E}}, \frac{\pi_E\left(1-\pi_E\right)}{n_E} + \frac{\pi_{\overline{E}}\left(1-\pi_{\overline{E}}\right)}{n_{\overline{E}}}\right)$$

- Συνάρτηση ελέγχου

$$z = \frac{\widehat{AR} - AR}{\sqrt{\frac{\pi_E\left(1-\pi_E\right)}{n_E} + \frac{\pi_{\overline{E}}\left(1-\pi_{\overline{E}}\right)}{n_{\overline{E}}}}} \sim N\left(0,1\right)$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

- Αν ισχύει η H_0 τότε

$$AR = \pi_E - \pi_{\overline{E}} = 0$$

$$\pi_E = \pi_{\overline{E}} = \pi$$

- Συνάρτηση ελέγχου

$$z = \frac{\widehat{AR}}{\sqrt{p(1-p)\left(\frac{1}{n_E} + \frac{1}{n_{\overline{E}}}\right)}} = \frac{p_E - p_{\overline{E}}}{\sqrt{p(1-p)\left(\frac{1}{n_E} + \frac{1}{n_{\overline{E}}}\right)}} \sim N(0,1)$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος

- μπορούμε να φτιάξουμε 100(1 – α)% διαστήματα εμπιστοσύνης για τον αποδιδόμενο κίνδυνο το οποίο θα δίνεται από τον τύπο

$$AR \pm z_{1-a/2}\sqrt{\frac{p_E\left(1-p_E\right)}{n_E}+\frac{p_{\overline{E}}\left(1-p_{\overline{E}}\right)}{n_{\overline{E}}}}\,.$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος για πίνακες 2x2

- ο αποδιδόμενος κίνδυνος δίνεται από τον τύπο:

$$AR = \pi_E - \pi_{\overline{E}} = \pi_{1|1} - \pi_{1|2}$$

- και εκτιμάται

$$AR = p_E - p_{\overline{E}} = p_{1|1} - p_{1|2} = \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}} = \frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{21}}{n_{21}+n_{22}} = \frac{n_{11}n_{21} + n_{11}n_{22} - n_{21}n_{11} - n_{21}n_{12}}{(n_{11}+n_{12})\cdot(n_{21}+n_{22})}$$

$$= \frac{n_{11}n_{22} - n_{21}n_{12}}{(n_{11}+n_{12})\cdot(n_{21}+n_{22})} = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{1.}n_{2.}}.$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος για πίνακες 2x2

- τυπικό σφάλμα του αποδιδόμενου κινδύνου, εάν η H_0 ισχύει

$$se(AR)=\sqrt{p(1-p)\left(\frac{1}{n_{1.}}+\frac{1}{n_{2.}}\right)}=\sqrt{\frac{n_{.1}n_{.2}}{n^2}\left(\frac{1}{n_{1.}}+\frac{1}{n_{2.}}\right)}=\sqrt{\frac{n_{.1}n_{.2}}{n^2}\left(\frac{n_{1.}+n_{2.}}{n_{1.}n_{2.}}\right)}=\sqrt{\frac{n_{.1}n_{.2}}{n\,n_{1.}n_{2.}}}$$

$$p=\frac{n_{.1}}{n}=\frac{n_{11}+n_{21}}{n_{11}+n_{12}+n_{21}+n_{22}}.$$

Συμπερασματολογία

Αποδιδόμενος Κίνδυνος για πίνακες 2x2

- συνάρτηση ελέγχου

$$z = \frac{\dfrac{n_{11}}{n_{11}+n_{12}} - \dfrac{n_{21}}{n_{21}+n_{22}}}{\sqrt{\dfrac{n_{.1}n_{.2}}{n\,n_{1.}n_{2.}}}} = \frac{\dfrac{n_{11}n_{22}-n_{21}n_{12}}{n_{1.}n_{2.}}}{\sqrt{\dfrac{n_{.1}n_{.2}}{n\,n_{1.}n_{2.}}}} = \frac{(n_{11}n_{22}-n_{21}n_{12})\sqrt{n}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}}$$

Συμπερασματολογία

Σχετικός Κίνδυνος

Γενική Προσέγγιση: ο σχετικός κίνδυνος ορίζεται ως ο λόγος των πιθανοτήτων εμφάνισης κινδύνου για της δύο ομάδες διαφορετικής έκθεσης στον κίνδυνο. Σε αυτή την ενότητα θα ασχοληθούμε με την κατανομή δειγματοληψίας του δειγματικού σχετικού κινδύνου την οποία μπορούμε να τη χρησιμοποιήσουμε για την κατασκευή διαστημάτων εμπιστοσύνης και ελέγχων υποθέσεων.

$$RR = p_E / p_{\overline{E}} \quad \Longrightarrow \quad \log RR = \log p_E - \log p_{\overline{E}}$$

Συμπερασματολογία

Σχετικός Κίνδυνος

- Γνωρίζουμε ότι

$$n_E p_E \sim Bin(\pi_E, n_E) \qquad n_{\overline{E}} p_{\overline{E}} \sim Bin(\pi_{\overline{E}}, n_{\overline{E}})$$

$$E(p_E) = \pi_E \quad \text{και} \quad Var(p_E) = \frac{\pi_E(1-\pi_E)}{n_E}$$

$$E(p_{\overline{E}}) = \pi_{\overline{E}} \quad \text{και} \quad Var(p_{\overline{E}}) = \frac{\pi_{\overline{E}}(1-\pi_{\overline{E}})}{n_{\overline{E}}}.$$

Για αρκετά μεγάλο δείγμα (πιο συγκεκριμένα για $n_E p_E \geq 5$ και $n_E(1 - p_E) \geq 5$) μπορούμε να θεωρήσουμε ότι η κατανομή του p_E προσεγγίζεται ικανοποιητικά από την κανονική κατανομή με μέσο και διακύμανση που δίνονται από την παραπάνω εξίσωση.

Συμπερασματολογία

Σχετικός Κίνδυνος

Επιπλέον μπορούμε να χρησιμοποιήσουμε το κεντρικό οριακό θεώρημα και να φτιάξουμε ένα z– *test.*

$$p_E \sim N\left(\pi_E, \frac{\pi_E(1-\pi_E)}{n_E} \right) \qquad p_{\overline{E}} \sim N\left(\pi_{\overline{E}}, \frac{\pi_{\overline{E}}(1-\pi_{\overline{E}})}{n_{\overline{E}}} \right)$$

Συμπερασματολογία

Σχετικός Κίνδυνος

Σειρά Taylor

$$h(x)=\sum_{k=0}^{\infty}h^{(k)}(a)\frac{(x-a)^k}{k!}=h(a)+h'(a)(x-a)+h''(a)\frac{(x-a)^2}{2}+h'''(a)\frac{(x-a)^3}{6}+\dots$$

$h^{(k)}(x)$ παράγωγος *k* τάξης της συνάρτησης *h(x).*

Έστω λοιπόν ότι *X* είναι τυχαία μεταβλητή και *α* = *E(X)* = *μ* τότε μπορούμε να χρησιμοποιήσουμε την παραπάνω έκφραση για να υπολογίσουμε ασυμπτωτικά τη μέση τιμή και τη διακύμανση μιας συνάρτησης της *X*.

Συμπερασματολογία

Σχετικός Κίνδυνος

$$E(h(x)) = \sum_{k=0}^{\infty} h^{(k)}(\mu)\frac{E(X-\mu)^k}{k!} \approx h(\mu) + h'(\mu)E(X-\mu) + h''(\mu)\frac{E(X-\mu)^2}{2}$$

$$\approx h(\mu) + h''(\mu)\frac{V(X)}{2}.$$

$$V(h(x)) = V\left(\sum_{k=0}^{\infty} h^{(k)}(\mu)\frac{(X-\mu)^k}{k!}\right) \approx V\left(h(\mu) + h'(\mu)V(X-\mu)\right)$$

$$\approx \{h'(\mu)\}^2 V(X).$$

Συμπερασματολογία

Σχετικός Κίνδυνος

θέτοντας $X = p_E$, $h(X) = \log(p_E)$, $E(X) = \pi_E$ και $V(X) = \pi_E(1-\pi_E)/n_E$

$$E(\log p_E) \approx \log \pi_E - \frac{1}{\pi_E^2}\frac{\pi_E(1-\pi_E)}{2n_E} = \log \pi_E - \frac{1}{2n_E}\frac{(1-\pi_E)}{\pi_E}$$

$$V(\log p_E) \approx \frac{1}{\pi_E^2}\frac{\pi_E(1-\pi_E)}{n_E} = \frac{1}{n_E}\frac{(1-\pi_E)}{\pi_E}.$$

το οποίο για μεγάλο n_E γίνεται ίσο με $\log \pi_E$. Συνεπώς το $\log p_E$ είναι ασυμπτωτικά αμερόληπτος εκτιμητής του $\log \pi_E$.

Συμπερασματολογία

Σχετικός Κίνδυνος

$$n \to \infty$$

$$\sqrt{n_E}\left(\log p_E - \log \pi_E\right) \to \sqrt{n_E}\,\frac{p_E - \pi_E}{\pi_E} \sim N\left(0, \frac{1-\pi_E}{\pi_E}\right)$$

$$\log p_E \sim N\left(\log \pi_E, \frac{1}{n_E}\frac{1-\pi_E}{\pi_E}\right) \qquad \log p_{\bar{E}} \sim N\left(\log \pi_{\bar{E}}, \frac{1}{n_{\bar{E}}}\frac{1-\pi_{\bar{E}}}{\pi_{\bar{E}}}\right)$$

$$\log RR = \log p_E - \log p_{\bar{E}} \sim N\left(\log RR, \frac{1}{n_E}\frac{1-\pi_E}{\pi_E}\,\frac{1}{n_{\bar{E}}}\frac{1-\pi_{\bar{E}}}{\pi_{\bar{E}}}\right)$$

Συμπερασματολογία

Σχετικός Κίνδυνος

Εκτίμηση σφάλματος

$$V(\log RR) = \frac{1}{n_E}\frac{1-p_E}{p_E} + \frac{1}{n_{\overline{E}}}\frac{1-p_{\overline{E}}}{p_{\overline{E}}}$$

$$= \frac{1}{n_E}\frac{n_E - r_E}{r_E} + \frac{1}{n_{\overline{E}}}\frac{n_{\overline{E}} - r_{\overline{E}}}{r_{\overline{E}}}$$

$$= \frac{1}{r_E} - \frac{1}{n_E} + \frac{1}{r_{\overline{E}}} - \frac{1}{n_{\overline{E}}},$$

r_E και $r_{\overline{E}}$ είναι ο παρατηρούμενος αριθμός των ασθενών με ή χωρίς έκθεση στον κίνδυνο αντίστοιχα.

Συμπερασματολογία

Σχετικός Κίνδυνος

- Ένα 100(1 – α)% διάστημα εμπιστοσύνης για το λογάριθμο του σχετικού κινδύνου θα δίνεται από της τιμές:

$$log\ RR \pm z_{1-a/2}\sqrt{V\left(log\ RR\right)},$$

$$log\ RR \pm z_{1-a/2}\sqrt{\frac{1}{r_E}-\frac{1}{n_E}+\frac{1}{r_{\bar{E}}}-\frac{1}{n_{\bar{E}}}}$$

$$e^{log\ RR \pm z_{1-a/2}\sqrt{\frac{1}{r_E}-\frac{1}{n_E}+\frac{1}{r_{\bar{E}}}-\frac{1}{n_{\bar{E}}}}}$$

Συμπερασματολογία

Σχετικός Κίνδυνος

- H_0: log *RR* = 0   έναντι της εναλλακτικής υπόθεσης
- H_1: log *RR* ≠ 0.

συνάρτηση ελέγχου

$$z = \frac{\log\left(A\widehat{R}\right)}{\sqrt{\frac{1}{r_E} - \frac{1}{n_E} + \frac{1}{r_{\overline{E}}} - \frac{1}{n_{\overline{E}}}}} \sim N(0,1)$$

- Αν |z| < $Z_{1-\alpha/2}$ τότε δεν απορρίπτουμε την υπόθεση της ανεξαρτησίας σε επίπεδο σημαντικότητας 100α % αλλιώς απορρίπτουμε την H_0.

Συμπερασματολογία

Σχετικός Κίνδυνος για πίνακες 2x2

- Τυπικό σφάλμα

$$se(\log RR) = \sqrt{\frac{1}{r_E} - \frac{1}{n_E} + \frac{1}{r_{\overline{E}}} - \frac{1}{n_{\overline{E}}}}\,.$$

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Γενική Προσέγγιση: λόγος σχετικών πιθανοτήτων (ΛΣΠ ή Odds Ratio ή *OR)* δίνεται από τον τύπο:

$$OR = \frac{Odds_E}{Odds_{\overline{E}}} = \frac{\dfrac{\pi_E}{1-\pi_E}}{\dfrac{\pi_{\overline{E}}}{1-\pi_{\overline{E}}}} = \frac{\pi_E\left(1-\pi_{\overline{E}}\right)}{\pi_{\overline{E}}\left(1-\pi_E\right)}$$

και εκτιμάται από

$$OR = \frac{\dfrac{p_E}{1-p_E}}{\dfrac{p_{\overline{E}}}{1-p_{\overline{E}}}} = \frac{p_E\left(1-p_{\overline{E}}\right)}{p_{\overline{E}}\left(1-p_E\right)}$$

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Παίρνοντας τον log έχουμε:

$$log\, OR = log \frac{p_E}{1-p_E} - log \frac{p_{\overline{E}}}{1-p_{\overline{E}}}$$

Με βάση την διωνυμική κατανομή βρίσκουμε ότι:

$$\mathrm{E}(p_E) = \pi_E \quad \text{και} \quad \mathrm{Var}(p_E) = \frac{\pi_E(1-\pi_E)}{n_E}.$$

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Από τον ασυμπτωτικό τύπο

$$E(h(x)) = \sum_{k=0}^{\infty} h^{(k)}(\mu)\frac{E(X-\mu)^k}{k!} \approx h(\mu)+h''(\mu)\frac{V(X)}{2}.$$

$$E\left(log\frac{p_E}{1-p_E}\right) \approx log\frac{\pi_E}{1-\pi_E}+\left(-\pi_E^{-2}+(1-\pi_E)^{-2}\right)\frac{\pi_E(1-\pi_E)}{n_E}$$

$$\approx log\,Odds_E+\frac{1}{n_E}\left(Odds_E-\frac{1}{Odds_E}\right)$$

εφόσον $h(x)=log\,x - log(1-x)$, $h'(x)=x^{-1} + (1 - x)^{-1} = 1/[x(1-x)]$ και $h''(x) = -x^{-2} + (1 - x)^{-2}$.

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Επιπλέον για n_E μεγάλο έχουμε

$$E\left(log\frac{p_E}{1-p_E}\right) \approx log\,Odds_E.$$

Όσον αφορά τη διακύμανση έχουμε

$$V\left(log\frac{p_E}{1-p_E}\right) \approx \{h'(\mu)\}^2 V(p_E) \approx \left(\frac{1}{\pi_E(1-\pi_E)}\right)^2 \frac{\pi_E(1-\pi_E)}{n_E} \approx \frac{1}{n_E}\frac{1}{\pi_E(1-\pi_E)}$$

$$\approx \left(n_E\frac{r_E}{n_E}\frac{n_E-r_E}{n_E}\right)^{-1} = \left(\frac{r_E(n_E-r_E)}{n_E}\right)^{-1} \approx \frac{n_E}{r_E(n_E-r_E)} = \frac{n_E-r_E+r_E}{r_E(n_E-r_E)}$$

$$\approx \frac{1}{r_E}+\frac{1}{n_E-r_E}.$$

$$\pi_E = \frac{r_E}{n_E}$$

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Άρα:

$$\log OR \sim N\left(\log OR, \frac{1}{r_E} + \frac{1}{n_E - r_E} + \frac{1}{r_{\overline{E}}} + \frac{1}{n_{\overline{E}} - r_{\overline{E}}}\right)$$

Ένα 100(1 – α)% διάστημα εμπιστοσύνης για το λογάριθμο του ΛΣΠ θα δίνεται από της τιμές:

$$log\,OR \pm z_{1-a/2}\sqrt{V\left(log\,OR\right)},$$

$$log\,OR \pm z_{1-a/2}\sqrt{\frac{1}{r_E} + \frac{1}{n_E - r_E} + \frac{1}{r_{\overline{E}}} + \frac{1}{n_{\overline{E}} - r_{\overline{E}}}}$$

Συμπερασματολογία

Λόγος Σχετικών Πιθανοτήτων

Αντίστοιχα για το ΛΣΠ το αντίστοιχο διάστημα εμπιστοσύνης προκύπτει από της τιμές:

$$e^{\log OR \pm z_{1-a/2}\sqrt{\frac{1}{r_E}+\frac{1}{n_E-r_E}+\frac{1}{r_{\overline{E}}}+\frac{1}{n_{\overline{E}}-r_{\overline{E}}}}}$$

- H_0: log *OR* = υ εναντι της εναλλακτικης υπόθεσης
- H_1: log *OR* ≠ 0.

συνάρτηση ελέγχου

$$z=\frac{\log OR}{\sqrt{\frac{1}{r_E}+\frac{1}{n_E-r_E}+\frac{1}{r_{\overline{E}}}+\frac{1}{n_{\overline{E}}-r_{\overline{E}}}}}\sim N(0,1)$$

Αν $|z| < Z_{1-\alpha/2}$ τότε δεν απορρίπτουμε την υπόθεση της ανεξαρτησίας σε επίπεδο σημαντικότητας 100^{α}% αλλιώς απορρίπτουμε την H_0.

Συμπερασματολογία

Λόγος πιθανοτήτων για πίνακες 2x2

$$\log OR \sim N\left(\log OR, \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)$$

Στην περίπτωση που ένα κελί είναι ίσο με μηδέν τότε ο ΛΣΠ θα είναι ίσος με μηδέν ή θα απειρίζεται.

Παράδειγμα

- Σε μια προοπτική μελέτη κατά την οποία εξετάσθηκαν 368 άνδρες καπνιστές ηλικίας κάτω των 60 ετών οι οποίοι έπαθαν μια καρδιακή ανακοπή και επιβίωσαν. Μετά από 2 έτη εξετάσθηκαν πόσοι από αυτούς είχαν επιβιώσει και τους χωρίσαμε ανάλογα εάν είχαν κόψει το τσιγάρο ή όχι. Έτσι εδώ μας ενδιαφέρει να εξετάσουμε αν το σταμάτημα του καπνίσματος (X) είχε ευνοϊκή επίδραση στην επιβίωση μετά από δύο έτη (Y). Τα δεδομένα δίνονται στον 2×2 Πίνακα που ακολουθεί:

X:	Y: Επιβίωση σε 2 χρόνια		
Συνέχισαν το κάπνισμα	1: Πεθαμένος	2: Ζωντανός	Σύνολο
1: Ναι	19 (12.3%)	135 (87.7%)	154 (41.8%)
2: Όχι	15 (7.0%)	199 (93.0%)	214 (58.2%)
Σύνολο	34 (9.2%)	334 (90.8%)	368

Παράδειγμα

α) Αποδιδόμενος Κίνδυνος

$$AR = \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}} = \frac{19}{154} - \frac{15}{214} = 0.123 - 0.070 = 0.053.$$

Επιθυμούμε να δούμε αν αυτή η διαφορά θνησιμότητας είναι στατιστικά σημαντική, συνεπώς θέλουμε να ελέγξουμε την υπόθεση:

- H_0: AR = 0, έναντι της εναλλακτικής
- H_1: AR ≠ 0.

τυπικό σφάλμα του αποδιδόμενου κινδύνου, εάν η H_0 ισχύει

$$se(AR \mid H_0) = \sqrt{\frac{n_{.1}n_{.2}}{n\, n_{1.} n_{2.}}} = \sqrt{\frac{34 \cdot 334}{368 \cdot 154 \cdot 214}} = \sqrt{\frac{11356}{12127808}} = \sqrt{0.000936} = 0.0306.$$

Παράδειγμα

α) Αποδιδόμενος Κίνδυνος

Επιθυμούμε να δούμε αν αυτή η διαφορά θνησιμότητας είναι στατιστικά σημαντική, συνεπώς θέλουμε να ελέγξουμε την υπόθεση:

- H_0: AR = 0, έναντι της εναλλακτικής
- H_1: AR ≠ 0.

τυπικό σφάλμα του αποδιδόμενου κινδύνου, εάν η H_0 ισχύει

$$z = \frac{AR}{se\left(AR \mid H_0\right)} = \frac{0.053}{0.0306} = 1.732 < z_{0.975} = 1.96,$$

δεν απορρίπτουμε την H_0

δεν υπάρχει στατιστική διαφορά της θνησιμότητας ανάλογα αν ο ασθενής συνεχίσει ή όχι το κάπνισμα

Παράδειγμα

α) Αποδιδόμενος Κίνδυνος

95% διάστημα εμπιστοσύνης

$$se(AR) = \sqrt{\frac{p_E(1-p_E)}{n_E} + \frac{p_{\bar{E}}(1-p_{\bar{E}})}{n_{\bar{E}}}} = \sqrt{\frac{0.123(1-0.123)}{154} + \frac{0.07(1-0.07)}{214}}$$

$$= \sqrt{0.000700461 + 0.0003042056} = 0.0317.$$

$0.053 \pm 1.96 \times 0.0317 = (-0.009, 0.115).$

Παράδειγμα

α) Σχετικός Κίνδυνος

$$RR = \frac{\frac{n_{11}}{n_{1.}}}{\frac{n_{21}}{n_{2.}}} = \frac{\frac{19}{154}}{\frac{15}{214}} = \frac{0.123}{0.070} = 1.757.$$

$$\log RR = 0.5636$$

$$se(\log RR) = \sqrt{\frac{1}{19} - \frac{1}{154} + \frac{1}{15} - \frac{1}{214}} = 0.3288.$$

95% διάστημα εμπιστοσύνης του λογαρίθμου

$$0.5636 \pm 1.96 \times 0.3288 = (-0.0808, 1.2080)$$

$$\left(e^{-0.0808}, e^{1.2080}\right) = (1.0842, 3.3468).$$

Παράδειγμα

- H_0: $RR = 1$ έναντι της εναλλακτικής υπόθεσης
- H_1: $RR \neq 1$.

- H_0: $\log RR = 0$ έναντι της εναλλακτικής υπόθεσης
- H_1: $\log RR \neq 0$.

$$z = \frac{0.5636}{0.3288} = 1.7141 < z_{0.975} = 1.96,$$

δεν απορρίπτουμε την H_0

τα ποσοστά θνησιμότητας δεν αλλάζουν για τα άτομα που συνέχισαν να καπνίζουν σε σχέση με τα άτομα που σταμάτησαν το κάπνισμα

Παράδειγμα

α) Λόγος πιθανοτήτων

$$OR = \frac{n_{11} \cdot n_{22}}{n_{21} \cdot n_{12}} = \frac{19 \cdot 199}{15 \cdot 135} = \frac{3781}{2025} = 1.867.$$

$$\log OR = 0.6244$$

$$se(\log OR) = \sqrt{\frac{1}{19} - \frac{1}{135} + \frac{1}{15} - \frac{1}{199}} = 0.363.$$

95% διάστημα εμπιστοσύνης του λογαρίθμου

$$0.6244 \pm 1.96 \times 0.363 = (-0.0867, 1.3358)$$

$$\left(e^{-0.0867}, e^{1.3358}\right) = (0.9167, 3.8030).$$

Παράδειγμα

- H_0: $OR = 1$ έναντι της εναλλακτικής υπόθεσης
- H_1: $OR \neq 1$.

- H_0: $\log OR = 0$ έναντι της εναλλακτικής υπόθεσης
- H_1: $\log OR \neq 0$.

$$z = \frac{0.6244}{0.363} = 1.72 < z_{0.975} = 1.96,$$

δεν απορρίπτουμε την H_0

η σχετική πιθανότητα θανάτου δεν διαφοροποιείται σημαντικά για τα άτομα που συνέχισαν να καπνίζουν σε σχέση με τα άτομα που σταμάτησαν το κάπνισμα

Παράδειγμα

- Πριν μερικά χρόνια υπήρχαν υποψίες ότι ο καρκίνος του στήθους στις γυναίκες σχετίζεται με την ηλικία της 1ης γέννας. Η υπόθεση αυτή αναφέρει ότι ο καρκίνος αυξάνεται όσο μεγαλώνει η ηλικία που η γυναίκα γεννάει το 1ο παιδί. Άρα με βάση την θεωρία αυτή ένας σημαντικός παράγοντας κινδύνου είναι η ηλικία της γυναίκας στη 1η γέννα. Αυτή η θεωρία εξηγεί κ γιατί ο καρκίνος του στήθους είναι σημαντικά αυξημένος σε ομάδες με αυξημένο κοινωνικό, οικονομικό κ εκπαιδευτικό επίπεδο, αφού ζευγάρια που ανήκουν σε αυτές τις ομάδες γεννούν συνειδητά σε μεγάλες ηλικίες. Το 1970 έγινε μια διεθνής πολυκεντρική μελέτη με την συμμετοχική των 6 χωρών. Περιπτώσεις καρκινοπαθών επιλέχτηκαν από τα ίδια νοσοκομεία, έτσι ώστε να είναι συγκρίσιμοι ηλικίας ως προς τον καρκίνο κ άλλα νοσήματα. Όλες οι γυναίκες χωρίστηκαν σε 2 ομάδες κινδύνου ανάλογα αν η ηλικία της 1ης γέννας ήταν μικρότερη ή μεγαλύτερη από των 30 χρόνων(ασθενείς και placebo). Ο πίνακας διπλής εισόδου είναι:

Παράδειγμα

Ηλικία 1ης γέννας			
	>=30	<=29	
Ασθενής	683	2537	3220
Μάρτυρας	1498	8747	10245
	2181	11284	13465

$\pi_{11} = 683/3220 = 0.212$ $\quad$ $\pi_{12} = 2537/3220 = 0.787$

$\pi_{21} = 1498/10245 = 0.146$ $\quad$ $\pi_{22} = 8747/10245 = 0.853$

Παράδειγμα

- **Αποδιδόμενος κίνδυνος(AR)**

$$A\hat{R} = p_E - p_{\overline{E}} = 0.212 - 0.146 = 0.066$$

Άρα εκείνες οι γυναίκες που ασθενούν (έχουν παρουσιάσει καρκίνο) έχουν πιθανότητα κατά 6.6% να ήταν πάνω από 30 ετών όταν γέννησαν το 1^{o} τους παιδί

Παράδειγμα

Ho:AR=0,εναλλακτική H1: AR $\neq 0$,

$$V(A\hat{R}/H_0) = p(1-p)(\frac{1}{n_E}+\frac{1}{n_{\overline{E}}}) = \frac{2181}{13465}\frac{11284}{13465}(\frac{1}{3220}+\frac{1}{10245}) = 0.000054$$

$$z = \frac{A\hat{R}}{se(AR/H_0)} = \frac{A\hat{R}}{\sqrt{V(AR/H_0)}} = \frac{0.066}{0.0073} \square 9.041,\ z_{1-a/2} = z_{0.975} = 1.96 \text{ αν έχω α=5\%}$$

- Άρα 9.041>1.96, απορρίπτω την Ho σε α=5%. Η διαφορά ήταν στατιστικά σημαντική ο καρκίνος για την γυναίκα αν έκανε το 1^{ο} της παιδί άνω των 30 ετών, άρα αν το πρώτο παιδί γίνεται όταν είναι άνω των 30 επηρεάζει το ενδεχόμενο να εμφανίσει καρκίνο του μαστού.

Παράδειγμα

100(1-α)% διάστημα εμπιστοσύνης

$$se(A\hat{R}) = \sqrt{\frac{p_E(1-p_E)}{n_E} + \frac{p_{\overline{E}}(1-p_{\overline{E}})}{n_{\overline{E}}}} = \sqrt{\frac{0.212*0.788}{3220} + \frac{0.146*0.854}{10245}} = 0.00721$$

$$A\hat{R} \pm Z_{1-a/2} se(A\hat{R}) = 0.066 \pm 1.96*0.00721 = 0.066 \pm 0.014 = (0.08, 0.052)$$

Παράδειγμα

- **Σχετικός κίνδυνος (RR)**

$$R\hat{R} = \frac{p_{\mathrm{E}}}{p_{\overline{\mathrm{E}}}} = \frac{0.212}{0.146} = 1.452$$

άρα ο κίνδυνος εμφάνισης καρκίνου για τις γυναίκες που έκαναν το $1^{ο}$ παιδί άνω των 30 ετών είναι ίσος με 1.452 (45.2% μεγαλύτερος κίνδυνος) από εκείνες που έκαναν το πρώτο παιδί κάτω των 29 ετών.

Παράδειγμα

$$\log R\hat{R} = \log 1.452 = 0.372\,,$$

$$se(\log R\hat{R}) = \sqrt{\frac{1}{p_E} - \frac{1}{n_E} + \frac{1}{p_{\overline{E}}} - \frac{1}{n_{\overline{E}}}} = \sqrt{0.00146 - 0.00031 + 0.00066 - 0.00009} = 0.0384$$

- 95% διάστημα εμπιστοσύνης για το logRR

$$\log R\hat{R} \pm z_{1-a/2} se(\log R\hat{R}) \rightarrow 0.372 \pm 1.96 * 0.0384 \rightarrow 0.372 \pm 0.0752 \rightarrow (0.4472, 0.2968)$$

$$e^{(0.4472,0.2968)} = e^{0.4472}, e^{0.2968} = (1.563, 1.345)$$

Παράδειγμα

- έλεγχο υπόθεσης :
- H_o: RR=1, εναλλακτική
- H_1: RR≠1

$$z = \frac{\log RR}{se(\log RR)} = \frac{0.372}{0.0384} = 9.6875 \text{, άρα } z > z_{0.975} = 1.96 \text{ απορρίπτω την μηδενική}$$

άρα είναι στατιστικά σημαντικό, τα ποσοστά των γυναικών που έκαναν παιδί το 1ο άνω των 30 ετών σχετίζεται με την εμφάνιση καρκίνου μαστού

Παράδειγμα

- **Λόγος υπεροχής (OR)**

$$\text{ΛΥ=}\ \hat{OR} = \frac{683 * 8747}{1498 * 2537} = 1.571$$

άρα η σχετική πιθανότητα των γυναικών που έκαναν το 1^{ο} παιδί άνω των 30 ετών και έπαθαν καρκίνο του μαστού είναι (57.1%) μεγαλύτερη από την πιθανότητα των γυναικών που έκαναν παιδί κάτω των 29 ετών και δεν έπαθαν καρκίνο.

Παράδειγμα

$$\log OR = \log 1.571 = 0.451$$

$$se(\log OR) = \sqrt{\frac{1}{683} + \frac{1}{2537} + \frac{1}{1498} + \frac{1}{8747}} = 0.0394$$

- 95% διάστημα εμπιστοσύνης για το logOR

$$\log OR \pm 1.96 se(\log OR) \rightarrow 0.45 \pm 1.960.039 = 0.45 \pm 0.077 = (0.52, 0.37)$$

$$e^{(0.5282, 0.3738)} = (1.695, 1.453)$$

Παράδειγμα

- έλεγχο υπόθεσης :
- H_o: OR=1, εναλλακτική
- H_1: OR≠1

$$z = \frac{\log OR}{se(\log OR)} = \frac{0.451}{0.0394} = 11.46$$

- απορρίπτω την μηδενική υπόθεση
- Εδώ η σχετική πιθανότητα εμφάνισης καρκίνου για τις γυναίκες που έκαναν το 1^o παιδί άνω των 30 ετών διαφοροποιείται σημαντικά σε σχέση με τα άτομα που δεν εμφάνισαν καρκίνο και έκαναν παιδί κάτω των 29 ετών.