ΠΜΣ ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗ ΥΓΕΙΑ, ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΚΗ ΑΠΟΤΙΜΗΣΗ

Βιοστατική ΙΙ

Ενότητα 3
Δείκτες Νοσηρότητας, Μέτρα Κινδύνου και Διαγνωστικού Ελέγχου

Ιωάννης Ντζούφρας
Τμήμα Στατιστικής
Οικονομικό Πανεπιστήμιο Αθηνών

Αθήνα, 4 Νοεμβρίου 2006

3.1 Μέτρα Νοσηρότητας - Εμφάνισης μιας Νόσου

3.1.1 Επιπολασμός

ΟΡΙΣΜΟΣ : Στιγμιαίος Επιπολασμός

Ως στιγμιαίο επιπολασμό (prevalence) μιας ασθένειας ονομάζουμε το ποσοστό (ή αναλογία) του πληθυσμού που έχει την νόσο τη χρονική στιγμή t .

ΟΡΙΣΜΟΣ : Επιπολασμός Περιόδου

Ως επιπολασμό περιόδου $[t_0, t_1)$ μιας ασθένειας ονομάζουμε το ποσοστό (ή αναλογία) του πληθυσμού που έχει την νόσο σε μία οποιαδήποτε χρονική στιγμή t μέσα στο διάστημα $[t_0, t_1)$ (δηλαδή $t_0 \le t < t_1$).

- Πιο απλά : (στιγμιαίος) επιπολασμός ⇒ πιθανότητα να έχουμε την υπό εξέταση νόσο d τη συγκεκριμένη χρονική στιγμή t. Δίνεται από τον τύπο

$$\text{Prevalence}_t^d = \frac{N_t^d}{N_t},$$

  - ◇ N_t^d είναι ο αριθμός των ατόμων του πληθυσμού με την ασθένεια d τη χρονική στιγμή t και
  - ◇ N_t είναι το μέγεθος του συνολικού πληθυσμού τη χρονική στιγμή t.

- Θεωρείται μέτρο συγχρονικών-διατμηματικών μελετών (cross-sectional measure)
- Δε λαμβάνει υπόψη του το χρόνο
- Εκτιμάται μόνο από μελέτες όπου η δειγματοληψία είναι τυχαία και αντιπροσωπευτική του γενικού πληθυσμού.
- Η εκτίμηση του επιπολασμού γίνεται με τον τύπο :

$$\widehat{\text{Prevalence}}^d = \frac{1}{n}\sum_{i=1}^{n} Y_i^d = \frac{n^d}{n},$$

  όπου Y_i^d είναι μια δίτιμη-δείκτρια μεταβλητή που παίρνει τιμές ένα (1) και μηδέν (0) αν το i άτομο έχει ή όχι τη νόσο d αντίστοιχα και n^d είναι ο συνολικός αριθμός των ατόμων στο δείγμα που έχουν τη νόσο d.
- Οι παραπάνω τύποι αναφέρονται στο στιγμιαίο επιπολασμό.
- Για λεπτομέρειες βλ. Daly *et al.* (1991, σελ. 281-282) και Pereira-Maxwell (1998, σελ. 62).

3.1.2 Αθροιστική Επίπτωση

Επιπολασμός ⇒ στιγμιαίος δείκτης της έκτασης μιας νόσου στον πληθυσμό.

Εδώ θα ορίσουμε δείκτες που λαμβάνουν υπόψη τους και το <u>χρόνο</u>.

ΟΡΙΣΜΟΣ : Επεισόδιο ή περιστατικό μίας νόσου

Μια περίπτωση (case) ονομάζεται « επεισόδιο » ή « περιστατικό » (incident) τη στιγμή ακριβώς που εμφανίζει για πρώτη φορά τα συμπτώματα της νόσου.

ΟΡΙΣΜΟΣ : Αθροιστική επίπτωση

« Αθροιστική επίπτωση » (cummulative incidence) μίας νόσου για το χρονικό διάστημα $[t_0, t_1)$ είναι η πιθανότητα ένα άτομο που δεν έχει εκδηλώσει προηγουμένως την ίδια νόσο να εμφανίσει για πρώτη φορά τα συμπτώματα της νόσου μέσα σε αυτό το διάστημα (Rosner, 1994, σελ. 61).

Δίνεται από τη πιθανότητα

$$I^d_{t_0,t_1} = P(t_0 < D^d < t_1 | D^d > t_0) = \frac{NI^d_{t_0,t_1}}{\overline{NI}^d_{t_0}}$$

- D^d είναι ο χρόνος εμφάνισης της νόσου d,
- $NI^d_{t_0,t_1}$ είναι ο αριθμός των νέων επεισοδίων της νόσου d στο χρονικό διάστημα $[t_0, t_1)$ και
- $\overline{NI}(t_0)$ είναι αριθμός των ατόμων του πληθυσμού χωρίς να έχει εκδηλώσει τη νόσο d πρίν τον χρόνο t_0.

Εκτιμάται από προοπτικές μελέτες (σε αντίθεση με τον επιπολασμό) από το τύπο :

$$\hat{I}^d_{t_0,t_1} = \frac{\sum_{i=1}^n Y_i^{Id(t_0,t_1)}}{\sum_{i=1}^n Y_i^{\overline{Id}(t_0)}} = \frac{ni^d_{t_0,t_1}}{\overline{ni}^d_{t_0}}$$

- $ni^d_{t_0,t_1}$ και $\overline{ni}^d_{t_0}$ είναι οι αντίστοιχες δειγματικές ποσότητες των $NI^d_{t_0,t_1}$ και $\overline{NI}(t_0)$
- $Y_i^{Id(t_0,t_1)}$ είναι δίτιμη-δείκτρια $(0-1)$ μεταβλητή με τιμή ένα αν το i άτομο εμφανίσει επεισόδιο μίας νόσου στο χρονικό διάστημα $[t_0, t_1)$
- $Y_i^{\overline{Id}(t_0)}$ είναι δίτιμη-δείκτρια $(0-1)$ μεταβλητή με τιμή ένα αν το i άτομο δεν έχει προϊστορία της νόσου μέχρι το χρόνο t_0 αντίστοιχα).

ΟΡΙΣΜΟΣ : Συνάρτηση Επιβίωσης

Συνάρτηση ή πιθανότητα αποφυγής της νόσου (ή επιβίωσης) μέχρι τον χρόνο t είναι η πιθανότητα να ένα άτομο να μην εμφανίσει την νόσο πρίν τη χρονική στιγμή t (ή να επιβιώσει τουλάχιστον μέχρι τη χρονική στιγμή t).

Δίνεται από τον τύπο

$$S(t) = P(D^d > t).$$

Χρησιμοποιώντας τη συνάρτηση επιβίωσης μπορούμε να γράψουμε :

$$I^d_{t_0,t_1} = \frac{S(t_0) - S(t_1)}{S(t_0)} = 1 - \frac{S(t_1)}{S(t_0)} .$$

3.2 Ρυθμός Επίπτωσης

Η αθροιστική επίπτωση

- . . . λαμβάνει υπόψη του το χρόνο ως πρός τα επεισόδια της νόσου που εμφανίζονται μέσα σε ένα χρονικό διάστημα.
- . . . δε λαμβάνει υπόψη του το χρόνο διάρκειας μιας νόσου ο οποίος είναι σημαντικός παράγοντας για τη κατανόηση της επίδρασης της νόσου στο πληθυσμό.

Παράδειγμα

$$\text{Νόσος με} \left\{ \begin{array}{l} \text{μικρή πιθανότητα εμφάνισης} \\ \text{αλλά μεγάλη διάρκεια} \end{array} \right\} \Rightarrow \text{εμφανίζει υψηλό επιπολασμό.}$$

Δηλαδή υπάρχουν πολλά άτομα στο πληθυσμό που υποφέρουν από τη νόσο αυτή και για το λόγο αυτό να είναι σημαντική η επίδρασή της στο σύνολο του πληθυσμού.

ΟΡΙΣΜΟΣ : Ρυθμός επίπτωσης

Έτσι ορίζουμε το «ρυθμό επίπτωσης» (incidence rate) ως το λόγο του αριθμού των νέων επεισοδίων μιας νόσου σε ένα χρονικό διάστημα $[t_0, t_1)$ προς «ανθρωποχρόνο κινδύνου» για την ίδια περίοδο (Pereira-Maxwell, 1998, σελ. 29).

ΟΡΙΣΜΟΣ : Ανθρωποχρόνος κινδύνου

Ο «ανθρωποχρόνος κινδύνου» δεν είναι τίποτα άλλο από το χρόνο παρακολούθησης αθροιστικά για όλα τα άτομα του πληθυσμού. Έτσι, εδώ ο παρανομαστής είναι χρόνος σε αντίθεση με την αθροιστική επίπτωση όπου παρανομαστής είναι ο αριθμός των ατόμων του πληθυσμού.

Όσον αφορά το πληθυσμό, ο ρυθμός επίπτωσης δίνεται από τον τύπο :

$$IR^d_{t_0,t_1} = \frac{NI^d_{t_0,t_1}}{T_{(t_0,t_1)}} = \frac{NI^d_{t_0,t_1}}{\sum_{i=1}^{N_{t_0}} T_{i,(t_0,t_1)}}$$

- N_{t_0} είναι όλα τα άτομα στο πληθυσμό το χρόνο t_0,
- $T_{(t_0,t_1)}$ είναι ο συνολικός ανθρωποχρόνος παρακολούθησης και
- $T_{i,(t_0,t_1)}$ είναι ο χρόνος παρακολούθησης μέσα στο χρονικό διάστημα $[t_0, t_1)$ για το άτομο i.

Αν δεν υπάρχουν απώλειες στο χρόνο παρακολούθησης και ως συνέπεια όλα τα άτομα παρακολουθούνται για χρόνο $t_1 - t_0$ τότε ο ρυθμός επίπτωσης δίνεται ως

$$IR^d_{t_0,t_1} = \frac{NI^d_{t_0,t_1}}{N_{t_0}(t_1 - t_0)} \ .$$

3.3 Διάρκεια Ασθένειας

ΟΡΙΣΜΟΣ

Ως «διάρκεια ασθένειας» (duration) ορίζουμε το χρονικό διάστημα από τη πρώτη εμφάνιση της νόσου ως την τελική θεραπεία και τη συμβολίζουμε ως D (ή D^d).

- Όταν μια ασθένεια είναι θανατηφόρα τότε ως διάρκεια ζωής ονομάζεται ο χρόνος μέχρι την κατάληξη του ασθενή.
- Όταν αναφερόμαστε σε ένα πληθυσμό ή ένα δείγμα τότε μας ενδιαφέρει η μέση διάρκεια της νόσου για τα άτομα που έχουν εμφανίσει τη νόσο και τη συμβολίζουμε ως $\overline{D}$ (ή $\overline{D}^d$).
- Η διάρκεια είναι σημαντικό μέτρο της επίδρασης μίας ασθένειας σε ένα πληθυσμό : Ασθένειες με μεγάλη μέση διάρκεια ⇒ υψηλό επιπολασμό ακόμα και αν ο ρυθμός επίπτωσης είναι μικρός.

3.4 Μέτρα Κινδύνου για Δίτιμα - Κατηγορικά Δεδομένα.

Στις Ιατρικές μελέτες

$\Rightarrow$ η κύρια μεταβλητή απόκρισης είναι δίτιμη (η ασθένεια είναι παρούσα ή όχι).

$\Rightarrow$ συνήθως και ο παράγοντας κινδύνου είναι επίσης κατηγορική μεταβλητή.

- Στην πιο απλή περίπτωση $\Rightarrow$ συγκρίνουμε ...
  1. τον κίνδυνο εμφάνισης μιας νόσου σε ένα άτομο που έχει εκτεθεί στον παράγοντα κινδύνου
  2. με τον αντίστοιχο κίνδυνο ενός ατόμου που δεν έχει εκτεθεί στον ίδιο παράγοντα κινδύνου.

Για το λόγο αυτό έμφαση θα δώσουμε στους δείκτες μέτρησης του κινδύνου εμφάνισης μίας νόσου όταν έχουμε δίτιμες κατηγορικές μεταβλητές.

- Σύγκριση ομάδων ως προς ποσοτικές μεταβλητές $\Rightarrow$ σύγκριση μέτρων κεντρικής θέσης (μέση τιμή και/ή διάμεσος).
- Σύγκριση ομάδων ως προς κατηγορικές μεταβλητές $\Rightarrow$ σύγκριση των δεσμευμένων κατανομών ή πιθανοτήτων.
- Δηλαδή σύγκριση των ποσοστών (ή πιθανοτήτων ή αναλογιών) για κάθε διαφορετική ομάδα έκθεσης σε ένα παράγοντα κινδύνου.

Στην ενότητα αυτή θα ασχοληθούμε με την πιο απλή περίπτωση της σύγκρισης δίτιμων (Bernoulli) παραγόντων κινδύνου X σε δύο ομάδες μιας δεύτερης κατηγορικής μεταβλητής Y που συνήθως είναι η εμφάνιση της νόσου. Δηλαδή θα επικεντρωθούμε στη σύγκριση κατηγορικών μεταβλητών που περιγράφονται από 2×2 πίνακες συνάφειας.

3.4.1 Εισαγωγικές Έννοιες για 2×2 Πίνακες Συνάφειας.

3.4.1.1 Πίνακες Συνάφειας.

Συνδιασμένη περιγραφή 2 κατηγορικών μεταβλητών

- Ομαδοποιημένα ραβδοδιαγράμματα (barchart) και
- Πίνακες συνάφειας διπλής εισόδου (two - way contingency tables).

ΟΡΙΣΜΟΣ

Οι πίνακες συνάφειας διπλής εισόδου είναι η κατανομή συχνοτήτων του συνδιασμού των επιπέδων των 2 υπό εξέταση κατηγορικών μεταβλητών.

- Αν οι δύο μεταβλητές X και Y έχουν I και J επίπεδα τότε έχουμε $I \times J$ συνδυασμούς.
- **Εκτιμάει** την από κοινού συνάρτηση πιθανότητας των δύο μεταβλητών που εξετάζουμε.
- **Απεικόνιση** $\Rightarrow$ με ένα ορθογώνιο πίνακα με I γραμμές και J στήλες.

- Κάθε συνδυασμός εμφανίζεται στον πίνακα διπλής εισόδου ως ένα κουτί το οποίο ονομάζεται **κελί** (cell).
- Ο αριθμός των εμφανίσεων κάθε συνδυασμού $(X, Y) = (i, j)$ στο δείγμα ονομάζεται συχνότητα (count) και συμβολίζεται με n_{ij} .
- Το συνολικό μέγεθος του δείγματος συμβολίζεται με n και είναι ίσο με $n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$.
- Η αντίστοιχη πιθανότητα $P(X = i, Y = j)$ συμβολίζεται ως π_{ij} και εκτιμάται από το $p_{ij} = n_{ij}/n$.

Ένας $I \times J$ πίνακα $\Rightarrow$ πίνακα συνάφειας διπλής εισόδου με I γραμμές και J στήλες.
Ένας 2×2 πίνακας διπλής εισόδου θα έχει την ακόλουθη μορφή:

	Y: Νόσος		Περ.Κατ.
X:Παρ. Κινδύνου	1 (Ασθενής)	2 (Υγιής)	X
1 (Παρόν)	n_{11}	n_{12}	$n_{1\bullet}$
2 (Απών)	n_{21}	n_{22}	$n_{2\bullet}$
Περ.Κατ. Y	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet} = n$

- Οι **μονομεταβλητές κατανομές** των X και Y δίνονται στην τελευταία στήλη και γραμμή του παραπάνω πίνακα και ονομάζονται περιθωριακές κατανομές (marginal distributions).
- Η περιθωριακή συχνότητα μιας κατηγορίας i της μεταβλητής X συμβολίζεται ως $n_{i\bullet}$ (ή εναλλακτικά n_{i+}): $n_{i\bullet} = \sum_{j=1}^{J} n_{ij}$
- η περιθωριακή συχνότητα μιας κατηγορίας j της μεταβλητής Y συμβολίζεται ως $n_{\bullet j}$ (ή εναλλακτικά n_{+j}) : $n_{\bullet j} = \sum_{i=1}^{I} n_{ij}$.

- Η τελεία « $\bullet$ » ή εναλλακτικά το σύμβολο της πρόσθεσης « + » στους δείκτες υποδηλώνει άθροισμα.
- Με $n_{\bullet\bullet}$ ή n_{++} πολύ συχνά συμβολίζεται το συνολικό άθροισμα του δείγματος.
- Με παρόμοιο τρόπο ορίζονται και οι περιθωριακές πιθανότητες $\pi_{i\bullet}$ και $\pi_{\bullet j}$ των X και Y και οι αντίστοιχες δειγματικές τους εκτιμήσεις $p_{i\bullet}$ και $p_{\bullet j}$.
- Οι παραπάνω πίνακες διπλής εισόδου ονομάζονται πίνακες συνάφειας (contingency tables) όταν τα κελιά περιέχουν συχνότητες (counts).
- Εναλλακτική ονομασία είναι « πίνακες σταυρωτής καταχώρησης » (cross-classified tables).

3.4.1.2 Δεσμευμένες Πιθανότητες.

Σε κάθε πίνακα συνάφειας συνήθως

- η μία μεταβλητή (π.χ. Y) είναι το λογικό αποτέλεσμα - απόκριση (response) και
- η άλλη (π.χ. X) είναι η ανεξάρτητη ή επεξηγηματική μεταβλητή (explanatory variable).

Σε αυτή την περίπτωση η ανεξάρτητη μεταβλητή X μπορεί να είναι σταθερή ή προκαθορισμένη από το σχεδιασμό του πειράματος και όχι τυχαία μεταβλητή.

Τότε η πιθανότητα $\pi_{ij} = P(X = i, Y = j)$ δεν έχει νόημα (αφού η X δεν είναι τυχαία μεταβλητή).

Μας ενδιαφέρει ...

- ... η **κατανομή της μεταβλητής Y για δεδομένο (σταθερό) X**.
- Δηλαδή θέλουμε να εξετάσουμε πως η μεταβλητή Y αλλάζει για διαφορετικές τιμές του X.
- Έτσι έχει νόημα η ανάλυση των υπό-συνθήκη πιθανοτήτων

$$\pi_{j|i} = P(Y = j | X = i)$$

με $\sum_{j=1}^{J} \pi_{j|i} = 1$. Οι πιθανότητες $(\pi_{1|i}, \ldots, \pi_{J|i})$ δίνουν τη δεσμευμένη (ή υπό συνθήκη) κατανομή (conditional distribution) της Y για το i επίπεδο της X.
- Σε 2×2 πίνακες: έλεγχος ανεξαρτησίας $\Rightarrow$ έλεγχος της ισότητας δύο και μόνο δεσμευμένων πιθανοτήτων
- Μπορούμε να φτιάξουμε διάφορα μέτρα βασισμένα στις πιθανότητες

$$P(\text{Ασθενής}|\text{Κίνδυνος Παρόν}) = \pi_{j=1|i=1} \;\&\; P(\text{Ασθενής}|\text{Κίνδυνος Απών}) = \pi_{j=1|i=2}$$

και στην ουσία θα μετρούν την σχέση και απόκλιση των δύο αυτών πιθανοτήτων.

3.4.1.3 Ανεξαρτησία Δύο Κατηγορικών Μεταβλητών.

ΟΡΙΣΜΟΣ

Η ανεξαρτησία μεταξύ δύο μεταβλητών ορίζεται στατιστικά ως

$$\pi_{ij} = P(X = i, Y = j) = P(X = i)P(Y = j) = \pi_{i\bullet}\, \pi_{\bullet j}\ .$$

Ανεξαρτησία των δύο μεταβλητών $\Leftrightarrow$ δεσμευμένες κατανομές $P(Y|X = i)$ είναι ίσες για όλα τα i εφόσον

$$\pi_{j|i} \quad = \quad P(Y = j|X = i) = \frac{P(Y = j, X = i)}{P(X = i)} = \frac{P(Y = j)P(X = i)}{P(X = i)} = P(Y = j) = \pi_{\bullet j}\ .$$

Η εμφάνιση της νόσου (Y) παραμένει σταθερή ανεξάρτητα της έκθεσης ή όχι στον παράγοντα κινδύνου X.

Με τον ίδιο τρόπο μπορούμε να βρούμε τα ίδια αποτελέσματα για τις δεσμευμένες κατανομές $\pi_{i|j} = P(X = i|Y = j)$.

- Συνεπώς ο έλεγχος της ανεξαρτησίας είναι ισοδύναμος με τον έλεγχο

$$H_0 : \pi_{j|i} = \pi_{\bullet j}\ \forall\ i = 1, \ldots, I \text{ και } j = 1, \ldots, J\ .$$

- Ο περιορισμός $\sum_{j=1}^{J} \pi_{j|i} = 1 \Rightarrow$ ο δείκτης j μπορεί να πάρει τιμές μέχρι το $J - 1$ εφόσον ισότητα των πιθανοτήτων των πρώτων $J - 1$ επιπέδων θα συνεπάγεται αυτόματα και ισότητα των πιθανοτήτων στο J επίπεδο.
- Για 2×2 πίνακες ο έλεγχος μπορεί να περιοριστεί στην απλή σχέση

$$H_0 : \pi_{j=1|i=1} = \pi_{j=1|i=2} \text{ έναντι της εναλλακτικής } H_1 : \pi_{j=1|i=1} \neq \pi_{j=1|i=2}\ , \qquad (1)$$

- που για τη περίπτωση νόσου και παράγοντα κινδύνου μπορεί να γραφτεί ως

$$H_0 : P(A|E) = P(A|\overline{E}) \text{ έναντι της εναλλακτικής } H_1 : P(A|E) \neq P(A|\overline{E}),$$

  - ◇ A είναι το ενδεχόμενο να είναι ένα άτομο ασθενής,
  - ◇ E και $\overline{E}$ είναι τα ενδεχόμενα ένα άτομο να έχει εκτεθεί ή όχι στον παράγοντα κινδύνου αντιστοίχα.

- Όλα τα μέτρα κινδύνου που θα περιγράψουμε πιο κάτω συγκρίνουν τον κίνδυνο, ο οποίος μετριέται με την πιθανότητα εμφάνισης της νόσου, σε κάθε ομάδα έκθεσης σε ένα παράγοντα κινδύνου.
- Τα μέτρα κινδύνου στην ουσία μετράνε τις διαφορές στον κίνδυνο (πιθανότητα εμφάνισης της νόσου) και μας δίνουν ενα συνολικό δείκτη εξάρτησης μεταξύ του παράγοντα κινδύνου που εξετάζουμε και της νόσου.

3.4.2 Οφειλόμενος ή Αποδιδόμενος Κίνδυνος

ΟΡΙΣΜΟΣ

Ως αποδιδόμενο κίνδυνο (Δαφνή, 1997 - attributable risk) ή διαφορά κινδύνου (risk difference, Rosner, 1995, σελ. 362-363) ορίζουμε τη διαφορά μεταξύ των ρυθμών επίπτωσης (ή των δεικτών θνησιμότητας) των ομάδων με άτομα εκτεθειμένα και μη εκτεθειμένα σε ένα παράγοντα κινδύνου.

Εναλλακτικά ονομάζεται και οφειλόμενος ή αποδοτέος κίνδυνος (Τριχόπουλος, 1982, σελ. 189).

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ : Πληθυσμός

Σε 2×2 πίνακα $\Rightarrow$ αποδιδόμενος κίνδυνος = διαφορά των δεσμευμένων πιθανοτήτων $\pi_{j=1|i=1}$ και $\pi_{j=1|i=2}$:

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

- $\pi_E = P(A|E)$: πιθανότητα εμφάνισης της νόσου των εκτεθημένων στον κίνδυνο (E)
- $\pi_{\overline{E}} = P(A|E)$: πιθανότητα εμφάνισης της νόσου των μη εκτεθημένων στον κίνδυνο ($\overline{E}$).

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ : Εκτίμηση από δείγμα

$$\widehat{AR} = p_E - p_{\overline{E}} = p_{j=1|i=1} - p_{j=1|i=2} = \frac{n_{11}}{n_{1\bullet}} - \frac{n_{21}}{n_{2\bullet}} = \frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{21}}{n_{21}+n_{22}},$$

- p_E και $p_{\overline{E}}$ αντιστοιχούν στις δειγματικές αναλογίες εμφάνισης της νόσου ανάλογα αν κάποιος έχει εκτεθεί στον κίνδυνο ή όχι.

Ο έλεγχος ανεξαρτησίας (1) $\Leftrightarrow H_0 : AR = 0$ versus $H_1 : AR \neq 0$.

Πρόβλημα: AR μετράει απόλυτες διαφορές.

Π.χ.: Πιθανότητες $(0.1, 0.2)$ & $(0.6, 0.7) \Rightarrow$ ίδιο $AR = 0.1$.

Γιατί πρόβλημα;: Αγνοεί το γεγονός ότι στην 1η περίπτωση η πιθανότητα της μίας ομάδας είναι 2πλάσια από την άλλη.

ΟΡΙΣΜΟΣ : Ποσοστιαίος Αποδιδόμενος Κίνδυνος

Αυτό διορθώνεται με τον ποσοστιαίο αποδιδόμενο κίνδυνο (Τριχόπουλος, 1982, σελ. 190):

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

ΕΡΜΗΝΕΙΑ : Αποδιδόμενος Κίνδυνος

Έστω $AR = a > 0 \Rightarrow$

- Από το σύνολο των ατόμων που έχουν εκτεθεί σε κίνδυνο, το $100a\%$ οφείλεται (θα μπορούσε να αποφευχθεί) στην έκθεση στον παράγοντα κινδύνου.
- $100a\%$ των εκτεθειμένων σε κίνδυνο ατόμων θα είχε αποφύγει την υπό εξέταση νόσο εάν δεν είχε εκτεθεί στον παράγοντα κινδύνου (ή $a \times n_E$ άτομα σε απόλυτους αριθμούς).

Π.χ. Έστω $AR = 0.1 \Rightarrow 10\%$ των εκτεθειμένων σε κίνδυνο ατόμων θα είχε αποφύγει την υπό εξέταση νόσο αν δεν είχε εκτεθεί στον παράγοντα κινδύνου. Αν $n_E = 143$ τότε σημαίνει ότι θα απέφευγαν τη νόσο (0.1×143) κατά μέσο όρο 14 άτομα.

ΕΡΜΗΝΕΙΑ : Ποσοστιαίος Αποδιδόμενος Κίνδυνος

Έστω $PAR = a > 0 \Rightarrow$

- Από το σύνολο νοσούντων που έχουν εκτεθεί σε κίνδυνο, το $100a\%$ οφείλεται (θα μπορούσε να αποφευχθεί) στην έκθεση στον παράγοντα κινδύνου.
- $100a\%$ των νοσούντων από την εκτεθειμένη σε κίνδυνο ομάδα ατόμων, θα είχε αποφύγει την υπό εξέταση νόσο εάν δεν είχε εκτεθεί στον παράγοντα κινδύνου (ή $a \times r_E$ άτομα σε απόλυτους αριθμούς).

Π.χ. Έστω $PAR = 0.1/0.2 = 0.5 \Rightarrow 50\%$ των νοσούντων στην ομάδα που έχει εκτεθεί στον παράγοντα κινδύνου, θα είχε αποφύγει την υπό εξέταση νόσο αν δεν είχε εκτεθεί στον παράγοντα κινδύνου. Αν $r_E = 0.2 \times 143 \approx 29$ τότε σημαίνει ότι θα απέφευγαν τη νόσο (0.5×29) κατά μέσο όρο 14 άτομα.

3.4.3 Σχετικός Κίνδυνος (Relative Risk, RR)

ΟΡΙΣΜΟΣ

Ως « σχετικός κίνδυνος » (relative risk) ορίζεται ο λόγος της επίπτωσης ή της θνησιμότητας δύο ομάδων με διαφορετική έκθεση στον παράγοντα κινδύνου.

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ : Πληθυσμός

Σε 2×2 πίνακες $\Rightarrow$ λόγος των πιθανοτήτων $\pi_{j=1|i=1}$ προς $\pi_{j=1|i=2}$:

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

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ : Εκτίμηση από δείγμα

$$\widehat{RR} = \frac{p_E}{p_{\overline{E}}} = \frac{p_{j=1|i=1}}{p_{j=1|i=2}} = \frac{n_{11}/n_{1\bullet}}{n_{21}/n_{2\bullet}} .$$

Ο έλεγχος ανεξαρτησίας (1) $\Leftrightarrow H_0 : RR = 1$ **versus** $H_1 : RR \neq 1$**.**

Συνεπώς, στην πράξη, αν ο σχετικός κίνδυνος είναι κοντά στο ένα $\Rightarrow$ η πιθανότητα εμφάνισης της νόσου δεν επηρεάζεται από τη διαφορετική έκθεση στον παράγοντα κινδύνου X (ή δεν αλλάζει για τις διαφορετικές κατηγορίες της μεταβλητής X).

ΕΡΜΗΝΕΙΑ RR 1

Αν $RR = a$ τότε Η πιθανότητα εμφάνισης της νόσου ($Y = 1$) όταν ένα άτομο έχει εκτεθεί στον παράγοντα κινδύνου ($X = 1$) είναι ίση με a φορές την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο ($X = 2$).

ΕΡΜΗΝΕΙΑ RR 2

Αν $RR = a$, τότε η ερμηνεία του σχετικού κινδύνου μπορεί να γίνει με οποιαδήποτε από τις παρακάτω προτάσεις:

1. 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 ($X = 2$).
2. 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 εκτεθεί στον κίνδυνο ($X = 2$) (η μεταβλητή μας εδώ λέγεται προστατευτικός παράγοντας).
3. 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 έκθεση στον κίνδυνο ($X = 2$) (δηλαδή η μεταβλητή X δεν επηρεάζει την εμφάνιση της νόσου συνεπώς δεν αποτελεί παράγοντα κινδύνου).

3.4.4 Λόγος Σχετικών Πιθανοτήτων (Odds Ratio, OR)

3.4.4.1 Σχετική Πιθανότητα (odds)

ΟΡΙΣΜΟΣ

Ο όρος "odds" χρησιμοποιείται πολύ συχνά στην Αγγλική γλώσσα υποκαθιστώντας το όρο "probability" (πιθανότητα) και συνήθως δίνει μια εκτίμηση της τύχης που έχει κάποιος να κερδίσει σε ένα αγώνα ή στοίχημα.

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ

Μαθηματικά το odds ενός ενδεχομένου A δίνεται από τον τύπο

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

και δεν είναι τίποτα άλλο από τον λόγο της πιθανότητας εμφάνισης ενός ενδεχομένου έναντι της πιθανότητας μη εμφάνισης του (δηλαδή τη πιθανότητα του συμπληρωματικού, ως προς το δειγματικό χώρο, ενδεχομένου).

Απόδοση στα Ελληνικά

- Δεν ήταν εύκολη υπόθεση.
- Στο λεξικό Αγγλικής γλώσσας του Longman (1987) βρίσκουμε τις ακόλουθες δύο ερμηνείες (σχετικές με πιθανότητες):
  - «Η πιθανότητα ότι κάτι θα συμβεί ή όχι»
  - «Πιθανότητα εκφρασμένη σε αριθμούς για χρήση σε στοιχήματα».
- Ο όρος χρησιμοποιείται κυρίως για στοιχήματα: Δίνει τις «ευκαιρίες» νίκης μιας ομάδας ή ενός διαγωνιζόμενου.
  Π.χ.: Η Ελληνική ομάδα ποδοσφαίρου πριν το Ευρωπαϊκό πρωτάθλημα ποδοσφαίρου του 2004 είχε, σύμφωνα με τα διεθνή γραφεία στοιχημάτων, πιθανότητα κατάκτησης (ή πιο σωστά αποτυχίας κατάκτησης) του πρωταθλήματος 80 προς 1 $(80:1)$ δηλαδή odds =1/80.
- Συνεπώς odds ⇒ ο λόγος ή η αναλογία του αριθμού εμφάνισης ενός ενδεχομένου (εκφρασμένο είτε σε απόλυτους αριθμούς είτε ως πιθανότητες) προς τον αριθμό μη εμφάνισης του ενδεχομένου αυτού (Pereira-Maxwell, 1998, σελ. 49).

Απόδοση στα Ελληνικά (2)

- Αγγλο-ελληνικό λεξικό ⇒ το odds = «πιθανότητα» (δεν αποδίδει την ακριβή έννοια του όρου).
- Από τα παραπάνω odds = το λόγο της πιθανότητας ενός ενδεχομένου έναντι της πιθανότητας του συμπληρωματικού ενδεχομένου.
- Πληρέστερη και ακριβέστερη απόδοση του στα ελληνικά θα μπορούσε να γίνει με τον όρο «Λόγος πιθανοτήτων συμπληρωματικών ενδεχομένων».
- Χρησιμοποιούμε τον πιο σύντομο όρο **«συμπληρωματική πιθανότητα»**
  - πιο εύχρηστος
  - αποδίδει τις δύο πιο σημαντικές έννοιες του odds : την σχέση με πιθανότητα και τη συμπληρωματικότητα των ενδεχομένων που συγκρίνουμε.
  - Ο ελληνικός όρος «Συμπληρωματική πιθανότητα» εμφανίστηκε στην Ελληνική βιβλιογραφία από τη συνάδελφο Ουρανία Δαφνή (βλέπε Δαφνή, 1997, σελ. 30).
  - Στην Ιατρική μερικές φορές «Λόγος Συμπληρωματικών Πιθανοτήτων».

Απόδοση στα Ελληνικά (3)

Πρόβλημα με τον όρο συμπληρωματική πιθανότητα

- Όταν έχουμε πάνω από 2 ενδεχόμενα ⇒ odds ονομάζεται ο λόγος 2 ενδεχομένων ακόμα και αν δεν είναι συμπληρωματικά
- Συνεπώς $odds_{12} = P(A_1)/P(A_2) = \pi_1/\pi_2$ με $\pi_1 + \pi_2 \leq 1$.
- Στην περίπτωση αυτή ο παραπάνω όρος δεν είναι σωστός.
- Για το λόγο αυτό ονομάζουμε το odds **«σχετική πιθανότητα»** που μπορεί να ερμηνευτεί ως η πιθανότητα ενός ενδεχομένου σε σχέση (αναλογικά) με την πιθανότητα ενός άλλου ενδεχομένου.

- Πρακτικά η συμπληρωματική πιθανότητα είναι $1-1$ μετασχηματισμός της αρχικής πιθανότητας.
- Όταν έχουμε δίτιμες μεταβλητές τότε
$$odds = \frac{\pi}{1-\pi} \Leftrightarrow \pi = \frac{odds}{1+odds} .$$
  που σημαίνει ότι μεταφέρει ακριβώς την ίδια πληροφορία με διαφορετικό τρόπο.
- Η ερμηνεία της θεωρείται πιο εύκολη από την απλή πιθανότητα διότι συγκρίνει την πιθανότητα εμφάνισης ενός ενδεχομένου με την πιθανότητα μη εμφάνισης.
- Π.χ. Όταν η Ελληνική ομάδα είχε σχετική πιθανότητα να μην κερδίσει το ευρωπαϊκό πρωτάθλημα του 2004 $80:1$
  - ⇒ 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 .

$\boxed{\text{Στην Ιατρική}}$: Σχετική πιθανότητα = πιθανότητα εμφάνισης της νόσου σε σχέση με τη μη εμφάνιση της.

ΕΡΜΗΝΕΙΑ (για δίτιμες μεταβλητές)

1. Αν odds =1 (ή $1:1$) τότε οι πιθανότητες εμφάνισης ή μη εμφάνισης του ενδεχομένου που εξεταζουμε είναι ίσες (δηλαδή 50%).
2. Αν $odds = \alpha$ τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι ίση με α φορές την πιθανότητα μη εμφάνισης του.
3. Αν $odds = \alpha$ και
   (α') $\alpha > 1$ τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι $\alpha - 1$ (ή $(\alpha-1)100\%$) φορές μεγαλύτερη από την πιθανότητα μη εμφάνισης του.
   (β') $\alpha < 1$ τότε η πιθανότητα εμφάνισης του ενδεχομένου που εξετάζουμε είναι $1-\alpha$ (ή $(1-\alpha)100\%$) φορές μικρότερη από την πιθανότητα μη εμφάνισης του.

ΕΡΜΗΝΕΙΑ 2 (για δίτιμες μεταβλητές)

Συνεπώς όταν μιλάμε για την συμπληρωματική πιθανότητα εμφάνισης της νόσου τότε:

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

ΣΤΑΤΙΣΤΙΚΟΣ ΟΡΙΣΜΟΣ: odds σε 2×2 Πίνακες (Πληθυσμός)

Ορίζουμε δύο σχετικές πιθανότητες ή odds (X) (ένα για κάθε επίπεδο της μεταβλητής X) από τις σχέσεις:

$$odds(X=1) = \frac{\pi_{1|i=1}}{\pi_{2|i=1}} = \frac{\pi_{1|i=1}}{1-\pi_{1|i=1}} = \frac{\pi_{11}}{\pi_{12}} \;\&\; odds(X=2) = \frac{\pi_{1|i=2}}{\pi_{2|i=2}} = \frac{\pi_{1|i=2}}{1-\pi_{1|i=2}} = \frac{\pi_{21}}{\pi_{22}}.$$

Δειγματική Εκτίμηση odds από 2×2 Πίνακες

$$\widehat{odds}(X=1) = \frac{p_{1|i=1}}{p_{2|i=1}} = \frac{n_{11}/n_{1\bullet}}{n_{12}/n_{1\bullet}} = \frac{n_{11}}{n_{12}} \;\&\; \widehat{odds}(X=2) = \frac{p_{1|i=2}}{p_{2|i=2}} = \frac{n_{21}/n_{2\bullet}}{n_{22}/n_{2\bullet}} = \frac{n_{21}}{n_{22}}.$$

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

Συνεπώς μας ενδιαφέρουν οι συμπληρωματικές πιθανότητες

$$odds_E = \frac{P(A|E)}{1-P(A|E)} = \frac{\pi_E}{1-\pi_E} \quad \text{και} \quad odds_{\overline{E}} = \frac{P(A|\overline{E})}{1-P(A|\overline{E})} = \frac{\pi_{\overline{E}}}{1-\pi_{\overline{E}}}$$

οι οποίες εκτιμούνται από τους τύπους

$$\widehat{odds}_E = \frac{p_E}{1-p_{\overline{E}}} \quad \text{και} \quad \widehat{odds}_{\overline{E}} = \frac{p_E}{1-p_{\overline{E}}}.$$