

Θεωρία των στατιστικών υποδειγμάτων

Με την εφαρμογή της θεωρίας των στατιστικών υποδειγμάτων στο χώρο της υγείας επιχειρείται να δοθούν απαντήσεις σε διάφορα σημαντικά προβλήματα, όπως ποια η διευκρίνιση των συνιστωσών αιτιών εμφάνιση ενός νοσήματος. Για να πραγματοποιηθούν αυτές οι επιδιώξεις, γίνονται διάφορες υποθέσεις για τις ιδιότητες που έχουν οι μονάδες ενός πληθυσμού. Το σύνολο αυτών των υποθέσεων ονομάζεται υπόδειγμα. Για να εκφραστεί ποσοτικά το σύνολο αυτών των υποθέσεων, ορίζεται μια μαθηματική εξίσωση ή ένα σύνολο εξισώσεων που ονομάζεται μαθηματικό υπόδειγμα και με το οποίο εκφράζεται ακριβώς ο τρόπος με τον οποίο μεταβάλλεται ένα χαρακτηριστικό των μονάδων του πληθυσμού, όταν μεταβάλλονται συγκεκριμένα άλλα χαρακτηριστικά τους. Στα μαθηματικά υποδείγματα υπάρχει ντετερμινιστική σχέση μεταξύ των μεταβλητών, υπάρχει δηλαδή ακριβής σχέση μεταξύ αιτιατού και αιτίων και, επομένως, αναφέρονται σε φυσικά φαινόμενα. Εκτός από τα φυσικά, υπάρχουν και τα τυχαία φαινόμενα, τα οποία δεν μπορούν να ερμηνευτούν ακριβώς με τα γνωστά μαθηματικά υποδείγματα. Η προσπάθεια της μαθηματικής επιστήμης, στην περίπτωση αυτή, είναι να εξηγήσει την κατά μέσο όρο μεταβολή ενός χαρακτηριστικού (εξαρτημένη μεταβλητή), όταν μεταβάλλονται άλλα γνωστά (ανεξάρτητες μεταβλητές), αλλά και άλλα άγνωστα. Αυτό επιτυγχάνεται με τα στατιστικά υποδείγματα, τα οποία αποτελούν ένα σύνολο υποθέσεων που θεωρούνται ότι ισχύουν κατά μέσο όρο για τις ιδιότητες κάθε μονάδας ενός πληθυσμού και οι οποίες εκφράζονται με μια εξίσωση ή ένα σύνολο εξισώσεων. Με τα στατιστικά υποδείγματα συνεξετάζεται η επίδραση επί της εξαρτημένης μεταβλητής όλων των άγνωστων ή μη μετρήσιμων μεταβλητών με μια μεταβλητή που ονομάζεται διαταρακτικός όρος ή τυχαίο σφάλμα. Βασικά στατιστικά υποδείγματα είναι: (α) *Το υπόδειγμα της γραμμικής παλινδρόμησης.* Με αυτό εκτιμάται η τιμή ενός ποσοτικού χαρακτηριστικού, από τις τιμές άλλων χαρακτηριστικών (προσδιοριστές, παράγοντες κινδύνου). Επίσης, εκτιμάται η μεταβολή ενός ποσοτικού χαρακτηριστικού, όταν μεταβάλλεται ένα μόνο άλλο ή πολλά άλλα, κυρίως ποσοτικά χαρακτηριστικά. Βασική υπόθεση στο υπόδειγμα είναι ότι η μεταβολή του ποσοτικού χαρακτηριστικού σχετίζεται ευθύγραμμα με τη μεταβολή των άλλων. (β) *Το υπόδειγμα της λογιστικής παλινδρόμησης.* Με αυτό εκτιμάται η πιθανότητα, καθώς και η μεταβολή της πιθανότητας να εμφανιστεί ένα χαρακτηριστικό (νόσημα), όταν μεταβάλλεται η τιμή άλλων χαρακτηριστικών ή όταν εμφανίζονται άλλα χαρακτηριστικά (παράγοντες κινδύνου). Πρόκειται για υπόδειγμα μη γραμμικής παλινδρόμησης. (γ) *Το υπόδειγμα της γραμμικής διακριτικής παλινδρόμησης.* Με αυτό, κατατάσσονται οι μονάδες ενός πληθυσμού σε κατηγορίες, ανάλογα με τις τιμές μερικών χαρακτηριστικών τους ή ανάλογα με το αν έχουν ή όχι κάποια χαρακτηριστικά. (δ) *Τα υποδείγματα μη γραμμικής παλινδρόμησης.* Υπάρχουν άπειρα μη γραμμικά στατιστικά υποδείγματα, που προέρχονται από αντίστοιχα μαθηματικά υποδείγματα, ανάλογα με τη μορφή με την οποία ορίζεται το τυχαίο σφάλμα. Με αυτά εκτιμάται η τιμή ενός ποσοτικού χαρακτηριστικού, από τις τιμές άλλων χαρακτηριστικών. Επίσης, εκτιμάται η μεταβολή ενός ποσοτικού χαρακτηριστικού, όταν μεταβάλλεται ένα μόνο άλλο ή πολλά άλλα, κυρίως ποσοτικά χαρακτηριστικά. Βασική υπόθεση είναι ότι τα χαρακτηριστικά συνδέονται με το υπόψη χαρακτηριστικό με μη γραμμική σχέση.

Θ. Κατοστάρας

Τμήμα Νοσηλευτικής,
Πανεπιστήμιο Αθηνών

Theory of statistical models

Abstract at the end of the article

Λέξεις ευρετηρίου

Διακριτική ανάλυση
Εξάρτηση
Λογιστική εξάρτηση
Παλινδρόμηση
Στατιστικό υπόδειγμα

1. ΕΙΣΑΓΩΓΗ

Πώς μπορεί να ερμηνευτεί η μεταβολή ενός χαρακτηριστικού; Υπάρχουν παράγοντες που συντελούν στη διαμόρφωσή της; Σε τι βαθμό την επηρεάζουν; Μπορούν να γίνουν γνωστές οι συνιστώσες αιτίες εμφάνισης ενός νοσήματος;

Αυτά είναι μέρος από ένα σύνολο ερωτημάτων, στα οποία επιχειρείται να δοθεί απάντηση με τη μέθοδο της παλινδρόμησης, που ονομάζεται και μέθοδος της εξάρτησης.

Οι βασικές επιδιώξεις με τη μέθοδο της παλινδρόμησης, όταν αυτή εφαρμόζεται στην έρευνα στο χώρο της υγείας, είναι:

- Να εκτιμηθεί η τιμή ενός ποσοτικού χαρακτηριστικού από τις τιμές άλλων χαρακτηριστικών (προσδιοριστές, παράγοντες κινδύνου)
- Να εκτιμηθεί η μεταβολή ενός ποσοτικού χαρακτηριστικού, όταν μεταβάλλεται ένα μόνο άλλο, παρουσία πολλών άλλων χαρακτηριστικών
- Να εκτιμηθεί η μεταβολή ενός ποσοτικού χαρακτηριστικού, όταν μεταβάλλονται συγχρόνως πολλά χαρακτηριστικά
- Να εκτιμηθεί η πιθανότητα εμφάνισης ενός χαρακτηριστικού (νόσημα), όταν μεταβάλλεται η τιμή άλλων χαρακτηριστικών
- Να εκτιμηθεί η πιθανότητα εμφάνισης ενός χαρακτηριστικού, όταν εμφανίζονται άλλα χαρακτηριστικά
- Να καταταγούν οι μονάδες ενός δείγματος σε ομάδες.

Για να πραγματοποιηθούν αυτές οι επιδιώξεις, γίνονται διάφορες υποθέσεις για τον τρόπο που μεταβάλλεται ένα χαρακτηριστικό και ορίζεται ένα στατιστικό υπόδειγμα.

2. ΟΡΙΣΜΟΣ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Υπόδειγμα είναι ένα σύνολο υποθέσεων, που αναφέρονται στις ιδιότητες μιας οντότητας. Οι υποθέσεις μπορεί να αναφέρονται σε άτομο ή φαινόμενο ή γενικά σε ένα σύνολο. Τα υποδείγματα προκύπτουν από τη θεωρία ή την εμπειρία. Ο ισχυρισμός ότι η αρτηριακή πίεση σχετίζεται με το σωματικό βάρος του ατόμου είναι μια υπόθεση. Για να εκφραστεί ποσοτικά η υπόθεση αυτή, θα πρέπει να γίνει μια επιπλέον υπόθεση και να οριστεί ο μαθηματικός τρόπος έκφρασης της συσχέτισης της

αρτηριακής πίεσης με το σωματικό βάρος, δηλαδή πρέπει να οριστεί ένα μαθηματικό υπόδειγμα.

Μαθηματικό υπόδειγμα είναι μια μαθηματική εξίσωση ή ένα σύστημα εξισώσεων, οι οποίες εκφράζουν ποσοτικά ένα σύνολο υποθέσεων που γίνονται για τις ιδιότητες κάθε μονάδας ενός πληθυσμού. Για παράδειγμα, είναι δυνατό να υποθεθεί ότι, σε κάθε άτομο που έχει σωματικό βάρος X , η αρτηριακή πίεσή του είναι Y και προσδιορίζεται ακριβώς με μία από τις άπειρες μαθηματικές σχέσεις, όπως την: $Y=a+bX$.

Στα μαθηματικά υποδείγματα υπάρχει πλήρης εξηγητική ικανότητα και είναι δυνατή η πρόβλεψη με ακρίβεια των τιμών της Y , όταν είναι γνωστές οι τιμές της X . Στα μαθηματικά υποδείγματα υπάρχει ντετερμινιστική σχέση μεταξύ των μεταβλητών, υπάρχει δηλαδή η μαθηματική έκφραση μεταξύ του αιτιατού και των αιτιών και, επομένως, τα μαθηματικά υποδείγματα αναφέρονται σε φυσικά φαινόμενα.

Εκτός από τα φυσικά, υπάρχουν και τα τυχαία φαινόμενα, τα οποία δεν μπορεί να ερμηνευτούν με τους γνωστούς φυσικούς νόμους. Η αρτηριακή πίεση των ατόμων δεν μπορεί να προσδιοριστεί αν είναι γνωστό μόνο το σωματικό βάρος τους και ενδεχομένως να μη σχετίζεται με το σωματικό βάρος. Η εμπειρία δείχνει ότι υπάρχει ένα πλήθος άλλων παραγόντων, εκτός του σωματικού βάρους, οι οποίοι ενδεχομένως να επιδρούν στη διαμόρφωση της αρτηριακής πίεσης και από τους οποίους πολλοί είναι άγνωστοι ή μη μετρήσιμοι, δηλαδή είναι τυχαίοι παράγοντες. Η ύπαρξη αυτών των παραγόντων δεν επιτρέπει τον προσδιορισμό της αρτηριακής πίεσης σε ένα άτομο, όταν είναι γνωστό μόνο το σωματικό βάρος του.

Κάτω από αυτή την αδυναμία, η προσπάθεια της μαθηματικής επιστήμης είναι να εξηγήσει την κατά μέσο όρο μεταβολή της Y , όταν είναι γνωστές οι τιμές της X . Αυτό επιτυγχάνεται με τη δημιουργία των στατιστικών υποδειγμάτων, στα οποία συνεξετάζεται και η επίδραση των τυχαίων παραγόντων επί της Y .

Στατιστικό υπόδειγμα είναι ένα σύνολο υποθέσεων, που θεωρούνται ότι ισχύουν κατά μέσο όρο για τις ιδιότητες κάθε μονάδας ενός πληθυσμού και οι οποίες εκφράζονται με μια μαθηματική εξίσωση ή ένα σύστημα εξισώσεων.

Το απλούστερο στατιστικό υπόδειγμα, π.χ. μεταξύ της αρτηριακής πίεσης Y και του σωματικού βάρους X , είναι το:

$$Y=a+bX+e$$

και ονομάζεται απλό γραμμικό υπόδειγμα.

Η Y ονομάζεται εξαρτημένη ή ερμηνευόμενη μεταβλητή. Η X ονομάζεται ανεξάρτητη ή ερμηνευτική μεταβλητή ή προσδιοριστής ή παράγοντας κινδύνου.

Η ε ονομάζεται διαταρακτικός όρος ή τυχαίο σφάλμα ή τυχαία απόκλιση και είναι μεταβλητή με την οποία συνεκφράζεται η επίδραση ενός πλήθους άγνωστων ή και μη μετρήσιμων παραγόντων, οι οποίοι επιδρούν και συνδιαμορφώνουν την Y . Η ε είναι τυχαία ή στοχαστική μεταβλητή, δηλαδή ακολουθεί ένα νόμο πιθανότητας και περιέχει και τα ενδεχόμενα σφάλματα και τα λάθη που γίνονται κατά τη διαδικασία των μετρήσεων. Στο χώρο της υγείας, η ε εκφράζει συνήθως τη βιολογική και κοινωνική συμπεριφορά των ατόμων, η οποία είναι αστάθμητη. Τα α και β ονομάζονται παράμετροι του υποδείγματος. Το α είναι ο σταθερός όρος και το β είναι ο συντελεστής του υποδείγματος. Οι παράμετροι του υποδείγματος δεν είναι δυνατό να υπολογιστούν, αλλά μπορεί να εκτιμηθούν με μαθηματικές μεθόδους και με μετρήσεις που γίνονται σε ένα υποσύνολο του πληθυσμού, το οποίο κατ'οικονομία ονομάζεται δείγμα (στην πραγματικότητα, δείγμα δεν υπάρχει, γιατί δεν είναι δυνατό η εύρεση υποσυνόλου του πληθυσμού που να έχει τα ακριβή χαρακτηριστικά του πληθυσμού).

Οι σημαντικότερες μέθοδοι εκτίμησης στατιστικών υποδειγμάτων είναι:

- Η μέθοδος των ελαχίστων τετραγώνων
- Η μέθοδος της μέγιστης πιθανοφάνειας.

Τα βασικά στατιστικά υποδείγματα που χρησιμοποιούνται στις μελέτες στο χώρο της υγείας και οι βασικές υποθέσεις τους είναι το υπόδειγμα της γραμμικής παλινδρόμησης (linear regression), της λογιστικής παλινδρόμησης ή λογαριθμιστικής εξάρτησης (logistic regression), της διακριτικής ανάλυσης (discriminant analysis) και τα μη γραμμικά υποδείγματα (non-linear models).

3. ΥΠΟΔΕΙΓΜΑ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (ΕΞΑΡΤΗΣΗΣ)

Το υπόδειγμα γραμμικής παλινδρόμησης είναι της μορφής:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + \varepsilon$$

Ανάλογα με τις υποθέσεις που γίνονται για το διαταρακτικό όρο, ε ορίζονται διάφορα υποδείγματα γραμμικής παλινδρόμησης.

Η συνθεότερη υπόθεση είναι ότι η Y αποτελεί συνεχή τυχαία μεταβλητή που ακολουθεί κανονική κατανο-

μή, γιατί είναι γραμμικός συνδυασμός της τυχαίας απόκλισης ε , για την οποία συνήθως γίνεται η υπόθεση ότι ακολουθεί την κανονική κατανομή με μέσο μηδέν. Κάτω από την τελευταία υπόθεση, το υπόδειγμα παίρνει την ισοδύναμη μορφή:

$$E(Y) = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k$$

όπου $E(Y)$ είναι η αναμενόμενη τιμή της Y και η οποία ταυτίζεται με το μέσο της Y στον πληθυσμό.

Εφαρμόζοντας τη μέθοδο ελαχίστων τετραγώνων ή τη μέθοδο μέγιστης πιθανοφάνειας στις μετρήσεις ενός δείγματος, λαμβάνεται εκτίμηση του υποδείγματος της μορφής:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

όπου τα $b_0, b_1, b_2, \dots, b_k$ αποτελούν κατάλληλες εκτιμήσεις των παραμέτρων του υποδείγματος $B_0, B_1, B_2, \dots, B_k$, που ονομάζονται και συντελεστές του.

Κάθε $b_i, i=1, 2, \dots, k$ δείχνει την κατά μέσο όρο μεταβολή της Y στο δείγμα, όταν η μεταβλητή X_i αυξηθεί κατά μία μονάδα και οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

Με διάφορους στατιστικούς ελέγχους, ελέγχεται η αξιοπιστία της εκτίμησης της ικανότητας του υποδείγματος να ερμηνεύει τη μεταβλητότητα της Y και να χρησιμοποιείται για προβλέψεις και βρίσκεται διάστημα εμπιστοσύνης για κάθε παράμετρο του υποδείγματος, διάστημα εμπιστοσύνης για τις διάφορες προβλεπόμενες τιμές της εξαρτημένης μεταβλητής, καθώς και διάστημα εμπιστοσύνης για τις διάφορες προβλεπόμενες μέσες τιμές της εξαρτημένης μεταβλητής, κάτω από συνθήκες σχετικής βεβαιότητας.

Η γραμμικότητα των υποδειγμάτων αναφέρεται στη γραμμικότητα των παραμέτρων του και όχι στη γραμμικότητα των ερμηνευτικών μεταβλητών του. Με αυτή την έννοια, το υπόδειγμα:

$$Y = B_0 + B_1X_1 + B_2X_2^2 + \dots + B_kX_k^k + \varepsilon$$

εκτιμάται ως γραμμικό, αν τεθεί $X = X_1, X^2 = X_2, \dots, X^k = X_k$.

4. ΥΠΟΔΕΙΓΜΑ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ Ή ΛΟΓΑΡΙΘΜΙΣΤΙΚΗΣ ΕΞΑΡΤΗΣΗΣ

Είναι της μορφής:

$$P = \exp(B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + \varepsilon) / [1 + \exp(B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k)]$$

όπου $\exp(a)$ είναι ο αριθμός 2,718 υψωμένος στη δύναμη a και P είναι το ποσοστό με το οποίο εμφανίζεται

ένα συγκεκριμένο χαρακτηριστικό (νόσημα) στον πληθυσμό, δηλαδή το P είναι η πιθανότητα επιλογής μιας μονάδας του πληθυσμού με το χαρακτηριστικό το οποίο μελετάται.

Μετά από ένα μετασχηματισμό, το υπόδειγμα παίρνει τη μορφή:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u$$

όπου $Y = \ln [P/(1-P)]$ είναι ο λογάριθμος του otz (odds) της P , δηλαδή του λόγου των συμπληρωματικών πιθανοτήτων, της πιθανότητας P να εμφανίζει μια μονάδα του πληθυσμού το χαρακτηριστικό, προς την πιθανότητα $1-P$ να μην το εμφανίζει, και $u = \ln e$ είναι ο φυσικός λογάριθμος (logarithmus natural) του διαταρακτικού όρου ε .

Εφαρμόζοντας τη μέθοδο μεγίστης πιθανοφάνειας στα στοιχεία ενός δείγματος, λαμβάνεται εκτίμηση του υποδείγματος της μορφής:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

όπου τα $b_0, b_1, b_2, \dots, b_k$ αποτελούν κατάλληλες εκτιμήσεις των συντελεστών του υποδείγματος $B_0, B_1, B_2, \dots, B_k$.

Κάθε b_j μετράει την κατά μέσο όρο μεταβολή του λογάριθμου του λόγου των συμπληρωματικών πιθανοτήτων του χαρακτηριστικού που μελετάται στο δείγμα, όταν η μεταβλητή X_j αυξηθεί κατά μία μονάδα και οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

Αν αντικατασταθούν συγκεκριμένες τιμές των ερμηνευτικών μεταβλητών στην εκτίμηση του υποδείγματος, λαμβάνεται η πιθανότητα εμφάνισης του χαρακτηριστικού.

Αν ένας προσδιοριστής X_i είναι δίτιμος, με $X_i=1$ όταν υπάρχει στη μονάδα του δείγματος ένα χαρακτηριστικό (παράγοντας κινδύνου) και με $X_i=0$ όταν δεν υπάρχει, η τιμή $\exp(b_i)$ αποτελεί εκτίμηση του σχετικού λόγου OR των συμπληρωματικών πιθανοτήτων, ο οποίος αποτελεί εκτίμηση του σχετικού κινδύνου RR να εμφανιστεί το χαρακτηριστικό (νόσημα), όταν στη μονάδα υπάρχει ο μελετούμενος παράγοντας και στην περίπτωση που το χαρακτηριστικό είναι σπάνιο.

Με διάφορους στατιστικούς ελέγχους, ελέγχεται η αξιοπιστία των συντελεστών του υποδείγματος, βρίσκεται διάστημα εμπιστοσύνης για κάθε παράμετρο του υποδείγματος και διάστημα εμπιστοσύνης για κάθε σχετικό λόγο συμπληρωματικών πιθανοτήτων του πληθυσμού, κάτω από συνθήκες σχετικής βεβαιότητας.

5. ΥΠΟΔΕΙΓΜΑ ΔΙΑΚΡΙΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Με τη μέθοδο αυτή δημιουργείται ένας κανόνας, με τον οποίο μπορεί να καταταχθεί μια μονάδα του πληθυσμού σε μια ομάδα από δύο ομάδες. Επιλέγεται ένα σχετικά μεγάλο υποσύνολο του πληθυσμού και με τις τιμές των μεταβλητών του X_1, X_2, \dots, X_k υπολογίζεται μια συνάρτηση της μορφής:

$$Y = B_1X_1 + B_2X_2 + \dots + B_kX_k$$

η οποία ονομάζεται διακριτική συνάρτηση. Οι προσδιοριστές X_1, X_2, \dots, X_k είναι δυνατό να επιλεγούν σταδιακά, ανάλογα με την «ένταση» της συνεισφοράς τους στη διάκριση της μονάδας του πληθυσμού. Ανάλογα με την τιμή που προκύπτει για την Y , η μονάδα του πληθυσμού κατατάσσεται σε μια από τις δύο κατηγορίες. Κάθε συντελεστής B_i εκτιμάει το ποσό της συμμετοχής της μεταβλητής X_i στη διάκριση της μονάδας του πληθυσμού, όταν η μεταβλητή X_i αυξηθεί κατά μία μονάδα, ενώ οι υπόλοιπες δεν μεταβληθούν. Από θεωρητική άποψη, οι μεταβλητές X_1, X_2, \dots, X_k πρέπει να ακολουθούν μια πολυμεταβλητή κανονική κατανομή, με διακυμάνσεις και συνδιακυμάνσεις αντίστοιχα ίσες στις δύο ομάδες, υπόθεση που είναι δυνατό να παραβιάζεται οριακά.

Με τη μέθοδο της διακριτικής ανάλυσης μπορεί να γίνει διάγνωση για την ύπαρξη ή όχι ενός νοσήματος, μέσω της εξέτασης διαφόρων παραγόντων κινδύνου.

6. ΥΠΟΔΕΙΓΜΑΤΑ ΜΗ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Υπάρχουν άπειρα σε πλήθος στατιστικά, μη γραμμικά υποδείγματα, των οποίων η εκτίμηση είναι από σχετικά απλή έως και αρκετά δύσκολη, ανάλογα με το αν είναι δυνατό ή όχι να εκτιμηθούν ως γραμμικά. Τα υποδείγματα τα οποία δεν είναι δυνατό να εκτιμηθούν ως γραμμικά, εκτιμώνται συνήθως με προσεγγιστικές μεθόδους και σε συνάρτηση με τη μαθηματική έκφραση του διαταρακτικού όρου. Το ίδιο μαθηματικό υπόδειγμα μπορεί να μετατραπεί σε διαφορετικά στατιστικά υποδείγματα, ανάλογα με τη μαθηματική έκφραση του τυχαίου σφάλματος. Στα περισσότερα μη γραμμικά υποδείγματα τα οποία εκτιμώνται, δεν είναι δυνατή η ερμηνεία της εκτίμησης των συντελεστών τους.

Βασικές μορφές μη γραμμικών υποδειγμάτων, τα οποία μετατρέπονται σε γραμμικά, είναι:

- Η $Y = X / (B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + \varepsilon)$.

Εκτιμάται ως γραμμικό, αν αντιστραφεί και τεθεί $Y^* = X/Y$

$$\bullet \text{ Η } Y = B_0 X_1^{B_1} X_2^{B_2} \dots X_k^{B_k} \varepsilon$$

Αν λογαριθμηθούν τα μέλη του, παίρνει τη μορφή:

$$Y^* = B_0 + B_1 X_1^* + B_2 X_2^* + \dots + B_k X_k^* + u$$

όπου $Y^* = \ln Y$, $X_1^* = \ln X_1$, $X_2^* = \ln X_2$, ..., $X_k^* = \ln X_k$, $u = \ln \varepsilon$ και εκτιμάται ως γραμμικό υπόδειγμα.

ABSTRACT

Theory of statistical models

T. KATOSTARAS

Faculty of Nursing, University of Athens, Athens, Greece

Archives of Hellenic Medicine 2003, 20(5):526-531

The application of the theory of the statistical models in the health field attempts to give answers to several important problems, one of which is that of the causes leading to the appearance of an illness. For this aim to be fulfilled, several assumptions are made about the properties of population units. In order to express quantitatively the group of these assumptions, a mathematical equation or a group of equations is defined, which is called a mathematical model. This model is used to express the mode by which a feature of the population units changes, when other specific features vary. Mathematical models specify a deterministic relation between their variables, an exact relationship between cause and outcome and consequently refer to natural phenomena. Apart from the natural, there are also incidental phenomena, which cannot be fully explained by the known mathematical models. The effort, then, of mathematics is concentrated on explaining the average variation of a feature (dependent variable), when other known (independent variables) and unknown features change. This is achieved by the use of statistical models. These models constitute a group of assumptions about the properties of the population units, which are considered to be true on average and are expressed through an equation or a group of equations. With statistical models the effects of the unknown and immeasurable variables on the dependent variable are co-examined, with the use of a variable that is called random error. *The linear regression model:* With the use of this model, the value of a quantitative feature is estimated through the values of the other features (determinants, risk factors). The variation of a quantitative characteristic can also be estimated, when only one or more of the other quantitative characteristics vary. A fundamental assumption of the model is that the variation of the quantitative characteristic is linearly related to the variation of the other characteristics. *The logistic regression model:* With this model, the likelihood and its variation can be estimated for a characteristic (illness) to show up, when the values of other characteristics change or when other characteristics show up (risk factors). This is a non-linear model. *The discriminant analysis model:* With this model, the population units are classified into categories, with the classification depending on the value of some of their characteristics or on the existence of certain features. *The non-linear regression models:* There are countless non-linear statistical models, which originate from respective mathematical models, depending on the form of the random error. With the use of these, the value of a quantitative characteristic is estimated, through the values of other characteristics. The variation of a quantitative characteristic can also be estimated when one or more, mainly quantitative, characteristics change. A fundamental assumption is that the changing characteristics relate to the characteristic concerned with a non-linear relationship.

Key words: Dependence, Discriminant analysis, Logistic regression, Regression, Statistical model

Βιβλιογραφία

1. GREENLAND S. Introduction to regression model. In: Rothmans KJ, Greenland S (eds) *Modern Epidemiology*. Lippincott-Raven, Philadelphia, 1998:319-400
2. HOSMER D, LEMESHOW S. *Applied logistic regression*. Wiley, New York, 1989
3. PAGANO M, GAUVREAU K. *Principles of Biostatistics*. Duxbury Press, An International Thomson Publ Co, Boston, 1996
4. ΓΚΑΜΑΛΕΤΣΟΣ Θ. *Θεωρητική οικονομετρία*. Τόμος β'. Εκδόσεις Καμπερόπουλος, Αθήνα, 1985

5. ΚΑΤΟΣΤΑΡΑΣ Θ. *Εισαγωγή στη στατιστική*. Έκδοση ιδίου, Αθήνα, 1993
6. ΠΑΠΑΕΥΑΓΓΕΛΟΥ Γ, ΚΑΤΟΣΤΑΡΑΣ Θ. *Βιοστατιστική και μεθοδολογία*. Έρευνα. Εκδόσεις Ζήτα, Αθήνα, 1995
7. ΣΠΑΡΟΣ Λ. Η έννοια του λόγου των οτζ στην αιτιολογική έρευνα. *Αρχ Ελλην Ιατρ* 1997, 14:373–374
8. ΣΠΑΡΟΣ Λ. *Μετα-επιδημιολογία ή εφαρμοσμένη ιατρική έρευνα*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2001
9. ΤΡΙΧΟΠΟΥΛΟΣ Δ, ΤΖΩΝΟΥ Α, ΚΑΤΣΟΥΓΙΑΝΝΗ Κ. *Βιοστατιστική*. Επιστημονικές Εκδόσεις Μαρία Γ. Παρισιάνου, Αθήνα, 2000

Corresponding author:

T. Katostaras, 123 Papadiamantopoulou street, GR-115 27
Athens, Greece
e-mail: tkatos@nurs.oua.gr

.....