Ανάλυση Επιβίωσης

Επικ. Καθ. Σ. Ζημερας
Τμήμα Στατιστικής και Αναλογιστικών – Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου
Σάμος
2020

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Η εύρεση της σχέσης μεταξύ μιας μεταβλητής που δηλώνει το χρόνο επιβίωσης ενός ατόμου και άλλων συμμεταβλητών, επιτυγχάνεται συνήθως μέσω ενός μοντέλου παλινδρόμησης. Όταν έχουμε αποκομμένα δεδομένα επιβίωσης, χρησιμοποιείται συνήθως το *μοντέλο παλινδρόμησης του Cox* (Cox regression model) ή διαφορετικά το *μοντέλο αναλογικού διακινδύνευσης του Cox* (Cox proportional hazard model).

Το μοντέλο αναλογικής διακινδύνευσης του Cox, έχει γίνει το πιο ευρέως διαδεδομένο εργαλείο για τη μοντελοποίηση της σχέσης των συμμεταβλητών με το χρόνο επιβίωσης ή άλλων αποκομμένων αποτελεσμάτων. Αυτό το μοντέλο δεν απαιτεί γνώση της κατανομής που ακολουθεί το εκάστοτε στατιστικό όπως

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

προηγουμένως. Η συνάρτηση κινδύνου σε αυτό το μοντέλο μπορούν να πάρει οποιαδήποτε μορφή, μέχρι και τη μορφή συνάρτησης βήματος (step-function), αλλ σε διαφορετικά άτομα θεωρείται ανάλογη και ανεξάρτητη του χρόνου. Η συνήθης συνάρτηση πιθανοφάνειας αντικαθίσταται από τη μερική συνάρτηση πιθανοφάνειας. Το σημαντικό γεγονός είναι ότι η στατιστική συμπερασματολογία με βάση τη μερική συνάρτηση πιθανοφάνειας είναι παρόμοια με εκείνη που βασίζεται στη συνάρτηση πιθανοφάνειας.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Ορισμός του Μοντέλου

Θεωρούμε ότι έχουμε ένα συγκεκριμένο αριθμό ατόμων προς μελέτη π.χ. $\boldsymbol{n}$ και ότι το $\boldsymbol{x}' = \left(x_1, x_2, \dots x_p\right)$ είναι το διάνυσμα των μεταβλητών που πιστεύουμε ότι επηρεάζουν το χρόνο ζωής αυτών των ατόμων. Οι μεταβλητές αυτές μπορεί να παριστάνουν διάφορα χαρακτηριστικά όπως:

- Θεραπείες
- Φυσικές ιδιότητες των ατόμων (όπως π.χ. φύλο, ηλικία)
- Εξωγενείς παράγοντες

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Επιπλέον ορίζουμε το $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots x_{ip}), i = 1, 2, \ldots, n$, να είναι το διάνυσμα με τις τιμές των συμμεταβλητών που αντιστοιχεί στο i άτομο.

Η γενική μορφή ενός αναλογικού κινδύνου μοντέλου είναι:

$$h(t; \boldsymbol{x}) = h_0(t) g(\boldsymbol{x})$$

όπου $g(\boldsymbol{x})$ είναι μια συνάρτηση του διανύσματος $\boldsymbol{x}$. Το μοντέλο αναλογικού κινδύνου του Cox υποθέτει ότι η $g(\boldsymbol{x})$ είναι μια εκθετική συνάρτηση των συμμεταβλητών και ισούται με

$$g(\boldsymbol{x}) = \sum_{j=1}^{p} \beta_j x_j = exp[\beta' \boldsymbol{x}]$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Αρχικά υποθέτουμε ότι οι συμμεταβλητές δεν εξαρτώνται από το χρόνο, ότι δηλαδή οι τιμές των συμμεταβλητών $\boldsymbol{x_i}$ καταγράφηκαν στην αρχή της μελέτης, στο $t = 0$, και ότι οι τιμές αυτές είναι σταθερές καθ'όλη τη διάρκεια της μελέτης. Άρα σύμφωνα με τα παραπάνω, στο μοντέλο αναλογικής διακινδύνευσης Cox, οι συμμεταβλητές $\boldsymbol{x}$ δρουν στη συνάρτηση διακινδύνευσης μέσω της σχέσης

$$h(t; \boldsymbol{x}) = h_0(t)e^{\beta' \boldsymbol{x}}$$

όπου $h_0(t)$ είναι μια βασική συνάρτηση διακινδύνευσης (ή *αναφορική συνάρτηση κινδύνου)* και β' ένα διάνυσμα p συντελεστών, οι οποίοι εκφράζουν ποσοτικά την επίδραση της καθεμιάς των συμμεταβλητών $\boldsymbol{x}$. Η ανεξαρτησία της διακινδύνευσης και κατά συνέπεια και της επιβίωσης από τη συμμεταβλητή $\boldsymbol{x_i}$ σημαίνει ότι $\beta_i' = 0$.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Ας θεωρήσουμε την κινδυνότητα (*Hazard Ratio* – HR) ως το λόγο των συναρτήσεων διακινδύνευσης δύο ατόμων. Στο μοντέλο αναλογικής διακινδύνευσης του Cox, παρατηρείται η εξής ιδιότητα: οι συναρτήσεις διακινδύνευσης των εκάστοτε ατόμων είναι ανάλογες μεταξύ τους. Δηλαδή, έστω ότι $[h(t|\boldsymbol{x_1})/h(t|\boldsymbol{x_2})]$ ο λόγος των συναρτήσεων διακινδύνευσης δύο ατόμων και $\boldsymbol{x_1} = (x_{11}, x_{12}, \ldots x_{1p})$, $\boldsymbol{x_2} = (x_{21}, x_{22}, \ldots x_{2p})$ να είναι τα διανύσματα των συμμεταβλητών. Τότε ο λόγος αυτός τότε είναι σταθερός (ανεξάρτητος του χρόνου):

$$HR(t) = \frac{h(t|\boldsymbol{x_1})}{h(t|\boldsymbol{x_2})} = \frac{h_0(t)e^{\beta'\boldsymbol{x_1}}}{h_0(t)e^{\beta'\boldsymbol{x_2}}} = e^{\beta'(\boldsymbol{x_1}-\boldsymbol{x_2})}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έχουμε δει ότι η συνάρτηση διακινδύνευσης είναι μαθηματικά ισοδύναμη με τη συνάρτηση επιβίωσης και συνδέονται μέσω της σχέσης $\boldsymbol{h(t) = -\frac{d}{dx} ln\, S(t)}$ ή της $\boldsymbol{S(t)} = exp\,[-\boldsymbol{H(t)}]$ όπου$\boldsymbol{H(t)}$ η σωρευτική συνάρτηση διακινδύνευσης.

Επομένως έχουμε:

$$H(t;\boldsymbol{x}) = \int_0^t h_0(u) e^{\beta'\boldsymbol{x}} du = H_0(t) e^{\beta'\boldsymbol{x}}$$

και άρα

$$S(t;\boldsymbol{x}) = exp[-H(t;\boldsymbol{x})] = exp\left[-H_0(t)e^{\beta'\boldsymbol{x}}\right] = [S_0(t)]^{e^{\beta'x}}$$

όπου $S_0(t) = exp[-H_0(t)]$ η αναφορική συνάρτηση επιβίωσης (*baseline survival function*).

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Παρατηρήσεις:

- Ο όρος *αναλογικοί κίνδυνοι* (proportional hazards), προέρχεται από το γεγονός ότι οποιαδήποτε δύο άτομα έχουν συναρτήσεις κινδύνου που είναι η μία (σταθερό) πολλαπλάσιο της άλλης.
- Στην περίπτωση του PH μοντέλου του Cox, η $g(\mathbf{x})$ είναι η συνάρτηση $e^{\boldsymbol{\beta}'\mathbf{x}} = e^{\beta_1 x_1 + \ldots + \beta_p x_p}$.
- Όταν θεωρούμε για το $h_0(t)$ κάποια κατανομή, τότε έχουμε την *παραμετρική μορφή* του μοντέλου αναλογικού κινδύνου.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Αν λογαριθμήσουμε τη σχέση

$$HR(t) = \frac{h(t,\mathbf{x}_1)}{h(t,\mathbf{x}_2)} = \frac{\cancel{h_0(t)}e^{\boldsymbol{\beta}'\mathbf{x}_1}}{\cancel{h_0(t)}e^{\boldsymbol{\beta}'\mathbf{x}_2}} = e^{\boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)}$$

$$\ln[h(t,\mathbf{x}_1)] - \ln[h(t,\mathbf{x}_2)] = \boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)$$

που δείχνει ότι το μοντέλο θεωρεί μια σταθερή διαφορά μεταξύ των λογάριθμων των κινδύνων των δύο ατόμων.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Θεωρούμε τώρα, ότι έχουμε μόνο μία μεταβλητή, τη X, που αντιπροσωπεύει το είδος της θεραπείας και θεωρούμε επίσης ότι παίρνει την τιμή 1 (x_1=1) αν το άτομο λαμβάνει τη θεραπεία A και 0 (x_2=0) αν λαμβάνει τη θεραπεία B. Τότε, η συνάρτηση κινδύνου για τα άτομα που ανήκουν στην πρώτη ομάδα (A), είναι $h(t,1) = h_0(t)e^{\beta}$, ενώ για τα άτομα της ομάδας B θα είναι $h(t,0) = h_0(t)e^{0} = h_0(t)$. Τότε η κινδυνότητα θα είναι:

$$HR = \frac{h(t,1)}{h(t,0)} = e^{\beta}$$

$$S_1(t) = \left[S_0(t)\right]^{e^{\beta}}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έτσι, αν:

- β>0 ή $e^{\beta} > 1$ και ο κίνδυνος ενός ατόμου που λαμβάνει τη θεραπεία Α θα είναι μεγαλύτερος από τον κίνδυνο ενός ατόμου που λαμβάνει τη θεραπεία Β ενώ η πιθανότητα επιβίωσης ενός ατόμου της ομάδας Α θα είναι μικρότερη από την πιθανότητα επιβίωσης ενός ατόμου της ομάδας Β, όπως προκύπτει από την (4.6΄) (αφού $S_0(t)<1$ ή $S_1(t) = [S_0(t)]^{e^{\beta}} < S_0(t)$)

- β=0 ή h(t,1)=h(t,0) και $S_1(t)=S_0(t)$, δηλαδή οι δύο θεραπείες θεωρούνται ισοδύναμες.

- β<0 ή $0 < e^{\beta} < 1$ και ο κίνδυνος ενός ατόμου που λαμβάνει τη θεραπεία Α θα είναι μικρότερος από τον κίνδυνο ενός ατόμου που λαμβάνει τη θεραπεία Β ενώ η πιθανότητα επιβίωσης ενός ατόμου της ομάδας Α θα είναι μεγαλύτερη από την πιθανότητα επιβίωσης ενός ατόμου της ομάδας Β.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Εκτίμηση των παραμέτρων του μοντέλου Cox

Θεωρούμε ένα σύνολο n ατόμων και υποθέτουμε ότι υπάρχουν συνολικά k πλήρεις, διακεκριμένοι χρόνοι, και $n-k$ δεξιά αποκομμένοι χρόνοι επιβίωσης. Έστω $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ οι k ταξινομημένοι πλήρεις χρόνοι και $R_{t_{(i)}}$ ή $\boldsymbol{R_i}$ το σύνολο των ατόμων που βρίσκονται σε κίνδυνο στο χρόνο $t_{(i)}$, δηλαδή το σύνολο των ατόμων που είναι υπό παρακολούθηση τη χρονική στιγμή $t_{(i)}$. Συμβολίζουμε με $\boldsymbol{x}_{(\boldsymbol{i})} = \left(x_{(i)1}, x_{(i)2}, \dots x_{(i)p}\right), i = 1, 2, \dots, k$ το διάνυσμα των συμμεταβλητών που αντιστοιχεί στο άτομο με πλήρη χρόνο ζωής $t_{(i)}$, $1 \leq i \leq k$. Από τη βασική θεωρία

πιθανοτήτων, η πιθανότητα αποτυχίας (θανάτου) ενός ατόμου του συνόλου $\boldsymbol{R_i}$ είναι:

$$\frac{exp\,[\sum_{j=1}^{p} \beta_j x_{j(i)}]}{\sum_{l \in R_i} exp\,[\sum_{j=1}^{p} \beta_j x_{jl}]} = \frac{exp\,(\boldsymbol{\beta}'\boldsymbol{x}_{(\boldsymbol{i})})}{\sum_{l \in R_i} exp\,(\boldsymbol{\beta}'\boldsymbol{x_l})}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Κάθε αποτυχία συμβάλλει κατά παράγοντα και ως εκ τούτου η μερική συνάρτηση πιθανοφάνειας για το σύνολο των δεδομένων είναι:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \left\{ \frac{exp\,(\boldsymbol{\beta}'\boldsymbol{x}_{(i)})}{\sum_{l \in R_i} exp\,(\boldsymbol{\beta}'\boldsymbol{x}_l)} \right\}$$

από την οποία προκύπτει η εκτιμήτρια μεγίστης πιθανοφάνειας $\hat{\boldsymbol{\beta}}$ του $\boldsymbol{\beta}$.

Οι συντελεστές παλινδρόμησης $\boldsymbol{\beta}$, εκτιμώνται από τις τιμές $\hat{\boldsymbol{\beta}}$ που μεγιστοποιούν τη μερική πιθανοφάνεια $L(\boldsymbol{\beta})$ ή ισοδύναμα το λογάριθμο της. Ο λογάριθμος της πιθανοφάνειας είναι:

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^{k} \left\{ \boldsymbol{\beta}'\boldsymbol{x}_{(i)} - \log\left[\sum_{l \in R_i} exp\,(\boldsymbol{\beta}'\boldsymbol{x}_l) \right] \right\}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Οι πρώτες μερικές παράγωγοι είναι

$$\frac{\partial l}{\partial \beta_r} = \sum_{j=1}^{k} \underset{\sim}{x}_{(j)r} - \sum_{j=1}^{k} \left[\sum_{i \in R_j} x_{(i)r} e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{(i)}} \Big/ \sum_{i \in R_j} e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{(i)}} \right], \ 1 \leq r \leq k.$$

Η παραπάνω σχέση αποτελεί σύστημα p-εξισώσεων, το οποίο λύνεται με επαναληπτικές μεθόδους όπως την Newton-Raphson. Επισημαίνεται ότι στην παραπάνω διαδικασία δεν προσδιορίζεται η $h_0(t)$, εξού και ο όρος μερική πιθανοφάνεια.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έλεγχος του λόγου πιθανοφανειών (Likelihood Ratio Test)

Αποτελεί το πιο σύνηθες τρόπο για έλεγχο υποθέσεων. Μια τέτοια υπόθεση μπορεί να είναι ότι $\beta_i = 0$ κάτι που σημαίνει ότι η διακινδύνευση και η διάρκεια ζωής, εξαρτάται από τη συμμεταβλητή x_i. Το μοντέλο προσαρμόζεται με και χωρίς τις συμμεταβλητές, η αφαίρεση των οποίων ισοδυναμεί με την επιβολή του πιο πάνω περιορισμού. Έστω $\hat{l}_1$ η μεγιστοποιημένη τιμή του λογαρίθμου της μερικής πιθανοφάνειας του μοντέλου για $\beta_i \neq 0$ και $\hat{l}_0$ για $\beta_i = 0$. Έτσι γίνεται σύγκριση κατά τον έλεγχο αυτό, της τιμής του $-2(\hat{l}_0 - \hat{l}_1)$ με την χ_1^2 κατανομή.

Κάτι τέτοιο συνεχίζεται για όλο τον αριθμό των συμμεταβλητών, όπου για τον έλεγχο της μηδενικής υπόθεσης, συγκρίνεται η μεταβολή της τιμής $-2(\hat{l}_0 - \hat{l}_p)$ με την τιμή της χ_p^2 κατανομής.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έλεγχος του Wald

Εναλλακτικά, χρησιμοποιείται ο έλεγχος Wald. Η ελεγχοσυνάρτηση Wald για κάθε μεταβλητή j είναι

$$W = \left\{\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right\}^2$$

Για τον έλεγχο πάλι της μηδενικής υπόθεσης ότι, $\beta_i = 0$ η τιμή W συγκρίνεται με την κατανομή χ_1^2 οπότε και ελέγχεται η μηδενική υπόθεση ότι. Ισοδύναμα μπορούμε να χρησιμοποιήσουμε την τιμή της

$$\left\{\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right\}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Είναι συχνό φαινόμενο οι τιμές που προκύπτουν από τους παραπάνω ελέγχους να βρίσκονται πολύ κοντά. Αυτό παρατηρείται κυρίως όταν έχουμε μεγάλα δείγματα. Σε μικρά δείγματα διαφέρουν και πιο αξιόπιστος θεωρείται ο έλεγχος του λόγου των πιθανοφανειών.

Η επιλογή των σημαντικών μεταβλητών από ένα σύνολο, γίνονται με τις γνωστές, από την ανάλυση παλινδρόμησης, διαδικασίες κατά βήματα (*stepwise*) όπως την προς τα εμπρός επιλογή (*forward selection*) και την προς τα πίσω απαλοιφή (*backward elimination*). Συγκεκριμένα, δεδομένου ότι έχουμε απορρίψει τη μηδενική υπόθεση ότι όλοι οι συντελεστές , με $\beta_j \neq 0,\ j\ =\ 1,2, \dots, p$, ελέγχουμε ποιες από τις υποψήφιες συμμεταβλητές πρέπει να ληφθούν υπόψη στο μοντέλο. Οπότε, είτε προστίθενται διαδοχικά στο αρχικό μοντέλο (χωρίς τις συμμεταβλητές), μεταβλητές με στατιστικά σημαντικούς συντελεστές, είτε αφαιρούνται διαδοχικά μη σημαντικές μεταβλητές από το μοντέλο που περιέχει όλες τις υποψήφιες μεταβλητές.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Στρωματοποίηση (Stratification)

Όταν χρειάζεται να μελετηθεί η επίδραση ενός επιπέδου μιας κατηγορικής μεταβλητής Z, σε σχέση με άλλες μεταβλητές χωρίς να ενδιαφέρει η επίδραση της Z στο αποτέλεσμα, τότε χρησιμοποιείται μια επέκταση του μοντέλου αναλογικού κινδύνου του Cox. Επίσης, όταν μια μεταβλητή έχει επίπεδα που δημιουργούν συναρτήσεις κινδύνου οι οποίες δεν ικανοποιούν την υπόθεση της αναλογικότητας, τότε στρωματοποιούμε ως προς τη μεταβλητή αυτή. Το μοντέλο που προκύπτει ονομάζεται στρωματοποιημένο μοντέλο του Cox (*stratified Cox model*) και εφαρμόζεται αφού θεωρήσουμε τη στρωματοποίηση των δεδομένων της Z σε υποομάδες, κάθε μία από τις οποίες χαρακτηρίζεται από ένα επίπεδο του παράγοντα. Το στρωματοποιημένο μοντέλο επιτρέπει στη μορφή της συνάρτησης κινδύνου να αλλάζει ανάμεσα στα επίπεδα της στρωματοποιημένης μεταβλητής. Η μεταβλητή Z μπορεί εκτός από κατηγορική να είναι το αποτέλεσμα χωρισμού μίας ποσοτικής μεταβλητής σε ομάδες.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Η συνάρτηση κινδύνου ενός ατόμου που ανήκει στο στρώμα i με διάνυσμα μεταβλητών $\boldsymbol{x}$, είναι:

$$h_i(t; \boldsymbol{x}) = h_{0i}(t)e^{\beta' \boldsymbol{x}}, i = 1, \dots, I$$

όπου

i: δηλώνει το στρώμα του παράγοντα

I: το πλήθος των επιπέδων του παράγοντα

$h_{0i}(t)$: η αναφορική συνάρτηση κινδύνου

Από το στρωματοποιημένο μοντέλο, φαίνεται ότι τα άτομα που ανήκουν στο ίδιο στρώμα, έχουν τις ίδιες αναφορικές συναρτήσεις κινδύνου, ενώ αντίθετα τα άτομα που ανήκουν σε διαφορετικά στρώματα έχουν διαφορετικές αναφορικές συναρτήσεις κινδύνου. Επίσης, τα άτομα που ανήκουν στο ίδιο στρώμα έχουν συναρτήσεις κινδύνου ανάλογες μεταξύ τους, αφού για παράδειγμα για δύο άτομα με μεταβλητές $\boldsymbol{x_1}$ και $\boldsymbol{x_2}$, που ανήκουν στο στρώμα $i, i = 1, \dots, I$, ισχύει:

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

$$\frac{h_i(t|\boldsymbol{x_1})}{h_i(t|\boldsymbol{x_2})} = \frac{h_{0i}(t)e^{\beta'\boldsymbol{x_1}}}{h_{0i}(t)e^{\beta'\boldsymbol{x_2}}} = e^{\beta'(\boldsymbol{x_1}-\boldsymbol{x_2})}$$

Αντίθετα, άτομα που ανήκουν σε διαφορετικά στρώματα δεν έχουν ανάλογες συναρτήσεις κινδύνου, αφού οι αναφορικές συναρτήσεις κινδύνου $h_{0i}(t)$ κάθε στρώματος είναι αυθαίρετες συναρτήσεις του χρόνου και αφήνονται ασυσχέτιστες.

Επιπλέον, από το στρωματοποιημένο μοντέλο φαίνεται ακόμη ότι οι συντελεστές παλινδρόμησης $\boldsymbol{\beta}$ είναι οι ίδιοι σε κάθε στρώμα. Σε αντίθετη περίπτωση, τα δεδομένα κάθε στρώματος θα θεωρούνταν ως διαφορετικά σύνολα δεδομένων και θα αναλύονταν ξεχωριστά.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Η εκτίμηση των συντελεστών παλινδρόμησης $\boldsymbol{\beta}$ προκύπτει από τη μεγιστοποίηση της συνάρτησης μερικής πιθανοφάνειας που γενικεύει την $L(\boldsymbol{\beta})$ και δίνεται από τη σχέση

$$L_s(\boldsymbol{\beta}) = \prod_{i=1}^{k} L_i(\boldsymbol{\beta})$$

Κάθε παράγοντας $L_i(\boldsymbol{\beta})$είναι η μερική πιθανοφάνεια που υπολογίζεται από τη σχέση $L(\boldsymbol{\beta})$ που ορίσαμε προηγουμένως για το στρώμα i και υπολογίζεται σε κάθε διακεκριμένο χρόνο αποτυχίας που παρατηρείται στο συγκεκριμένο στρώμα.

Οι έλεγχοι υποθέσεων στους συντελεστές παλινδρόμησης μπορούν να γίνουν με τους συνηθισμένους ελέγχους που βασίζονται στις υποθέσεις της συνάρτησης πιθανοφάνειας, τροποποιώντας όμως αναλόγως για κάθε στρώμα το $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Παράδειγμα - *Υπολογισμός της μερικής πιθανοφάνειας σε στρωματοποιημένα δεδομένα*

Θεωρούμε ότι έχουμε μία ποσοτική μεταβλητή $\boldsymbol{x}$ και τη μεταβλητή $\boldsymbol{Z}$, ως προς την οποία θέλουμε να στρωματοποιήσουμε. Η $\boldsymbol{Z}$ έχει δύο επίπεδα (1 και 2) και θέλουμε να μελετηθεί η επίδραση των επιπέδων της μεταβλητής $\boldsymbol{Z}$, σε σχέση με τη μεταβλητή $\boldsymbol{x}$, χωρίς να ενδιαφέρει η επίδραση της $\boldsymbol{Z}$ στο αποτέλεσμα. Να βρεθεί η μερική πιθανοφάνεια των δεδομένων που βρίσκονται στον πίνακα.

$\boldsymbol{t_i}$	Κατάσταση	$\boldsymbol{z_i}$	$\boldsymbol{X}$
4	0	1	3
5	1	1	4
2	1	1	2
6	0	2	1
3	1	2	8
1	1	1	7

1 Κατάσταση = 1: Πλήρης Χρόνος, Κατάσταση = 0: Αποκομμένος Χρόνος

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Στρωματοποιώντας ως προς τη $\boldsymbol{Z}$, το αντίστοιχο στρωματοποιημένο μοντέλο είναι το

$$h_i(t;\boldsymbol{x}) = h_{0i}(t)e^{\beta'\boldsymbol{x}}, i = 1,2$$

Για το πρώτο επίπεδο, $z = 1$, έχουμε 3 πλήρεις, διακεκριμένους χρόνους, τους $t_1{}^1 = 1$, $t_2{}^1 = 2$, $t_3{}^1 = 5$ με αντίστοιχες τιμές συμμεταβλητών, $x_1{}^1 = 7$, $x_2{}^1 = 2$, $x_3{}^1 = 4$. Αντίστοιχα στο δεύτερο επίπεδο υπάρχουν 2 παρατηρήσεις, από τις οποίες η μία μόνο είναι πλήρης η $t_1{}^2 = 3$ με $x_1{}^2 = 8$.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Από τη σχέση $L_i(\boldsymbol{\beta})$ παίρνουμε

$$L_1(\boldsymbol{\beta}) = \prod_{i=1}^{3} \frac{e^{\beta x_k{}^i}}{\sum_{j \in xR_k} e^{\beta x_j{}^i}} = \frac{e^{7\beta}}{e^{7\beta} + e^{2\beta} + e^{3\beta} + e^{4\beta}} * \frac{e^{2\beta}}{e^{2\beta} + e^{3\beta} + e^{4\beta}} * \frac{e^{4\beta}}{e^{4\beta}}$$

$$L_2(\boldsymbol{\beta}) = \prod_{i=1}^{3} \frac{e^{\beta x_k{}^i}}{\sum_{j \in xR_k} e^{\beta x_j{}^i}} = \frac{e^{8\beta}}{e^{8\beta} + e^{\beta}}$$

Έτσι, η μερική πιθανοφάνεια είναι:

$$L_s(\boldsymbol{\beta}) = \prod_{i=1}^{k} L_i(\boldsymbol{\beta}) =$$

$$= L_1(\boldsymbol{\beta}) * L_2(\boldsymbol{\beta}) = \frac{e^{7\beta}}{e^{7\beta} + e^{2\beta} + e^{3\beta} + e^{4\beta}} * \frac{e^{2\beta}}{e^{2\beta} + e^{3\beta} + e^{4\beta}} * \frac{e^{8\beta}}{e^{8\beta} + e^{\beta}}$$

Θεωρώντας $l(\boldsymbol{\beta}) = ln\, L(\boldsymbol{\beta})$ τότε το $l(\boldsymbol{\beta})$ ισούται με

$$l(\boldsymbol{\beta}) = 17\beta - \left[ln\left(e^{7\beta} + e^{2\beta} + e^{3\beta} + e^{4\beta}\right) + ln\left(e^{2\beta} + e^{3\beta} + e^{4\beta}\right) + ln\left(e^{8\beta} + e^{\beta}\right)\right]$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Χρονοεξαρτώμενες Μεταβλητές

Μια δεύτερη επέκταση του μοντέλου αναλογικού κινδύνου του Cox προκύπτει όταν έχουμε μεταβλητές εξαρτώμενες από το χρόνο. Μέχρι τώρα θεωρούσαμε ότι οι μεταβλητές x_i ήταν σταθερές στο χρόνο. Υπάρχουν όμως περιπτώσεις στις οποίες οι τιμές κάποιων μεταβλητών μεταβάλλονται με το χρόνο. Έτσι, το άτομο i θα έχει τιμή $x_i(t)$ στο χρόνο t. Το μοντέλο αναλογικού κινδύνου του Cox μπορεί να επεκταθεί έτσι ώστε να ενσωματώνει τέτοιες μεταβλητές. Ο πιο συνηθισμένος τύπος εξαρτώμενης από το χρόνο μεταβλητής είναι μια επαναλαμβανόμενη μέτρηση σε ένα άτομο ή μια αλλαγή στη θεραπεία ενός ατόμου.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Οι εξαρτώμενες από το χρόνο μεταβλητές χωρίζονται σε δύο μεγάλες κατηγορίες, στις εξωτερικές μεταβλητές (external covariates) και τις εσωτερικές μεταβλητές (internal covariates). Οι τιμές που παίρνει μια εξωτερική μεταβλητή στο χρόνο, δεν επηρεάζονται από την εμπειρία ζωής του ατόμου μέσα στη μελέτη, αφού οι τιμές που παίρνει δημιουργούνται από ένα ‘μηχανισμό’ που είναι εξωτερικός του ατόμου. Η ηλικία ενός ατόμου θεωρείται συνήθως ως σταθερή μεταβλητή σε μελέτες όπου η διάρκεια τους είναι μικρή. Όταν όμως η διάρκεια ενός πειράματος είναι μεγάλη, τότε απαιτείται ο ορισμός της ηλικίας ως μεταβλητής εξαρτώμενης από το χρόνο. Η ηλικία θεωρείται εξωτερική μεταβλητή, αφού σε οποιοδήποτε χρόνο παρακολούθησης του ατόμου, μπορεί να βρεθεί από την ηλικία του ατόμου στην είσοδο της μελέτης. Ένα άλλο παράδειγμα εξωτερικής μεταβλητής είναι η δοσολογία του φαρμάκου που θα χορηγηθεί σε ένα ασθενή, η οποία καθορίζεται από την αρχή της μελέτης και μεταβάλλεται με το χρόνο, με τρόπο προκαθορισμένο από την αρχή της θεραπευτικής αγωγής.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Μια ειδική κατηγορία εξωτερικών μεταβλητών είναι οι βοηθητικές (ancillary) μεταβλητές, των οποίων οι τιμές είναι το αποτέλεσμα μιας στοχαστικής διαδικασίας εξωτερικής του ατόμου και δεν έχουν καμία σχέση με τα ‘εσωτερικά’ χαρακτηριστικά της μελέτης. Ένα παράδειγμα βοηθητικών μεταβλητών είναι μια μεταβλητή που περιγράφει την αλλαγή ενός ασθενή από την ομάδα ελέγχου στην ομάδα θεραπείας, αν η αλλαγή καθορίζεται τυχαία από κάποιο εξωτερικό παράγοντα. Στην περίπτωση αυτή η μεταβλητή είναι ένας δείκτης που ορίζεται για κάθε άτομο i, ως:

$$x_i(t) = \begin{cases} 0, & \alpha\nu\ \tau o\ \acute{\alpha}\tau o\mu o\ i\ \varepsilon\acute{\iota}\nu\alpha\iota\ \sigma\tau\eta\nu\ o\mu\acute{\alpha}\delta\alpha\ \varepsilon\lambda\acute{\varepsilon}\gamma\chi o\upsilon\ \sigma\tau o\ \chi\rho\acute{o}\nu o\ t \\ 1, & \alpha\nu\ \tau o\ \acute{\alpha}\tau o\mu o\ i\ \varepsilon\acute{\iota}\nu\alpha\iota\ \sigma\tau\eta\nu\ o\mu\acute{\alpha}\delta\alpha\ \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\varsigma\ \sigma\tau o\ \chi\rho\acute{o}\nu o\ t \end{cases}$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έλεγχοι της Υπόθεσης Αναλογικότητας Κινδύνων

Γραφικός Έλεγχος

Σύμφωνα με την υπόθεση του αναλογικού κινδύνου, για τη συνάρτηση επιβίωσης ενός ατόμου με διάνυσμα συμμεταβλητών $\boldsymbol{x} = (x_1, x_2, \dots x_p)$, ισχύει

$$S(t; \boldsymbol{x}) = [S_0(t)]^{e^{\beta' \boldsymbol{x}}}$$

Αν λογαριθμίσουμε και τα δύο μέλη της εξίσωσης προκύπτει

$$\log\big(S(t; \boldsymbol{x})\big) = e^{\beta' \boldsymbol{x}} * \log\big(S_0(t)\big)$$

και ισοδύναμα έχουμε

$$\log\Big(\log\big(S(t; \boldsymbol{x})\big)\Big) = \beta' \boldsymbol{x} + \log\Big(\log\big(S_0(t)\big)\Big)$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Για $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ διανύσματα συμμεταβλητών δύο ατόμων, κάτω από την υπόθεση της αναλογικότητας, παρατηρούμε ότι:

$$\log\big(\log\big(S(t;\boldsymbol{x}_1)\big)\big)-\log\big(\log\big(S(t;\boldsymbol{x}_2)\big)\big)=\beta'\boldsymbol{x}_1-\beta'\boldsymbol{x}_2$$

Δηλαδή οι συναρτήσεις επιβίωσης των δύο ατόμων έχουν μια σταθερή απόσταση ίση με $\beta'(\boldsymbol{x}_1-\boldsymbol{x}_2)$ και ισχύει ότι

$$\log\big(\log\big(S(t;\boldsymbol{x}_1)\big)\big)=\log\big(\log\big(S(t;\boldsymbol{x}_2)\big)\big)+\beta'(\boldsymbol{x}_1-\boldsymbol{x}_2)$$

Η σταθερή απόσταση $\beta'(\boldsymbol{x}_1-\boldsymbol{x}_2)$ θα είναι αναμενόμενη και υπολογίσιμη εάν σχεδιάσουμε τις γραφικές παραστάσεις των $\log\left(\log\left(S(t;\boldsymbol{x}_{1,2})\right)\right)$ συναρτήσει του χρόνου. Οι δύο καμπύλες που θα προκύψουν θα είναι παράλληλες και αυτός είναι ένας πρώτος έλεγχος που θα έλεγχε την αναλογικότητα των κινδύνων.

Συμπερασματικά αν οι καμπύλες που θα προκύψουν είναι παράλληλες ή σχεδόν παράλληλες, τότε θεωρούμε ότι ισχύει η υπόθεση της αναλογικότητας.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Σχεδιάζονται ξεχωριστά, στο ίδιο γράφημα, οι καμπύλες για κάθε επίπεδο της μεταβλητής. Πρέπει οι καμπύλες που θα προκύψουν για κάθε επίπεδο της μεταβλητής να είναι παράλληλες. Όταν μια μεταβλητή είναι ποσοτική, τότε την κατηγοριοποιούμε και δημιουργούμε τις γραφικές παραστάσεις για κάθε κατηγορία αυτής. Αν θέλουμε να συμπεριλάβουμε στο μοντέλο αλληλεπιδράσεις κάποιων μεταβλητών, τότε σχεδιάζουμε στο ίδιο γράφημα τις γραφικές παραστάσεις των $\log\left(\log\left(S(t;\boldsymbol{x}_{1,2})\right)\right)$ συναρτήσει του χρόνου, για κάθε συνδυασμό των επιπέδων των μεταβλητών.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Έλεγχος Βάσει των Ορισμένων Χρονοεξαρτόμενων Μεταβλητών

Έστω ότι έχουμε τις $\boldsymbol{x} = (x_1, x_2, \dots x_k)$μεταβλητές, και ότι θέλουμε να εξετάσουμε αν η σταθερή μεταβλητή x_k ικανοποιεί την υπόθεση της αναλογικότητας των κινδύνων, στην παρουσία των υπόλοιπων $k-1$ μεταβλητών. Έτσι, ορίζουμε ένα μετασχηματισμό της x_k, πολλαπλασιάζοντας την με μία συνάρτηση του χρόνου, $g(t)$, άρα $x_k(t) = x_k * g(t)$ και επομένως γίνεται μεταβλητή εξαρτημένη του χρόνου. Συμβολίζουμε με $\boldsymbol{x}^{-}$ το διάνυσμα των υπόλοιπων $k-1$ μεταβλητών και θεωρούμε το ακόλουθο μοντέλο του Cox:

$$h(t;\boldsymbol{x}) = h_0(t) exp(\beta_k x_k + \gamma x_k(t) + \boldsymbol{\beta}^{-}\boldsymbol{x}^{-})$$

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Κάνουμε έλεγχο της μηδενικής υπόθεσης $H_0\colon \gamma = 0$. Αν η υπόθεση αυτή γίνει δεκτή, τότε συμπεραίνουμε ότι η μεταβλητή x_k ικανοποιεί την υπόθεση της αναλογικότητας των κινδύνων. Γενικά, μια μη-μηδενική τιμή του γ, θα σήμαινε μια αλλαγή της κινδυνότητας μεταξύ δύο ατόμων με διαφορετική τιμή του x_k στο χρόνο. Η μορφή της αλλαγής αυτής εξαρτάται από τη μορφή που θα επιλεγεί για τη συνάρτηση $g(t)$ και επειδή ο σκοπός είναι να εξετάσουμε την υπόθεση της αναλογικότητας των κινδύνων και όχι να μοντελοποιήσουμε την επίδραση του x_k στο χρόνο, επιλέγουμε συνήθως απλές συναρτήσεις του χρόνου όπως $g(t) = t$ ή $g(t) = lnt$.

ΜΟΝΤΕΛΑ ΑΝΑΛΟΓΙΣΤΙΚΟΥ ΚΙΝΔΥΝΟΥ COX

Στην περίπτωση που θέλουμε να εξετάσουμε αν οι μεταβλητές $x_1, x_2, \dots x_k$ ικανοποιούν ταυτόχρονα την υπόθεση της αναλογικότητας των κινδύνων, τότε σχηματίζουμε το γενικευμένο μοντέλο

$$h(t;\boldsymbol{x}) = h_0(t)exp\left(\sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \gamma_i \left[x_\iota g(t)\right]\right)$$

και κάνουμε τον έλεγχο της υπόθεσης: $H_0: \gamma = 0$. Αν η μηδενική υπόθεση γίνει δεκτή, συμπεραίνουμε ότι οι μεταβλητές$x_1, x_2, \dots x_k$ ικανοποιούν την υπόθεση της αναλογικότητας των κινδύνων.