

Ανάλυση Επιβίωσης

Επικ. Καθ. Σ. Ζημερας

Τμήμα Στατιστικής και Αναλογιστικών – Χρηματοοικονομικών
Μαθηματικών

Πανεπιστήμιο Αιγαίου

Σάμος

2020

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- Η εκτίμηση της συνάρτησης επιβίωσης (ή αντίστοιχα της συνάρτησης κινδύνου) μπορεί να γίνει είτε με μη παραμετρικό είτε με παραμετρικό τρόπο
- Παρόλο που η μη παραμετρική εκτίμηση χρησιμοποιείται περισσότερο από την παραμετρική κρίνεται απαραίτητο να αναφέρουμε την περίπτωση της παραμετρικής εκτίμησης στην οποία η κατανομή των δεδομένων επιβίωσης θεωρείται γνωστή

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- **Παραμετρική εκτίμηση**

Στην παραμετρική εκτίμηση η θεωρητική κατανομή των χρόνων επιβίωσης είναι δεδομένη (είναι γνωστή από πληροφορίες που βασίζονται σε παλαιότερες έρευνες) οπότε απλά ορίζεται το κατάλληλο παραμετρικό μοντέλο.

- **Μη παραμετρική εκτίμηση**

Στην περίπτωση των μη παραμετρικών μεθόδων δεν είναι γνωστό ποιά συγκεκριμένη θεωρητική κατανομή ακολουθούν οι κατανομές των χρόνων επιβίωσης.

Οι μη παραμετρικές μέθοδοι εκτίμησης είναι λιγότερο αποτελεσματικές από τις παραμετρικές μεθόδους, όταν οι χρόνοι επιβίωσης είναι γνωστό ότι ακολουθούν μία συγκεκριμένη κατανομή, αλλά καταλληλότερες όταν η θεωρητική κατανομή των χρόνων επιβίωσης δεν είναι γνωστή, η οποία είναι και η συνηθέστερη περίπτωση.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- **Μη παραμετρική εκτίμηση**

Αρχικά παραστούμε γραφικά τα δεδομένα με μία καμπύλη επιβίωσης (survival curve) έτσι ώστε να εντοπιστεί τυχόν μοτίβο κάποιας κατανομής.

Εάν τα δεδομένα συνάδουν με κάποια παραμετρική κατανομή τότε η στατιστική συμπερασματολογία μπορεί να βασιστεί στην συγκεκριμένη κατανομή και οι παράμετροι μπορούν να εκτιμηθούν. Εάν όμως καμία θεωρητική κατανομή δεν προσαρμόζει επαρκώς στα δεδομένα χρησιμοποιούνται μη παραμετρικές μέθοδοι.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- **Εμπειρική συνάρτηση επιβίωσης**

Η εκτίμηση της $S(t)$ μπορεί να γίνει μέσω της εμπειρικής συνάρτησης επιβίωσης (empirical survival function) :

$$\tilde{S}(t) = \frac{\text{Αριθμός παρατηρήσεων που επιβίωσαν πέραν του χρόνου } t}{\text{Συνολικός αριθμός παρατηρήσεων}}$$
$$= \frac{\#t_i \geq t}{n} = \frac{1}{n} \sum_{i=1}^n I\{t_i \geq t\}$$

Πρόκειται για μία μη αρνητική, φθίνουσα, βηματική συνάρτηση για την οποία ισχύει :

$$\tilde{S}(0) = 1 \text{ και } \tilde{S}(\infty) = 0.$$

Μέσω της εμπειρικής συνάρτησης επιβίωσης, μπορούμε να ορίσουμε την εμπειρική

συνάρτηση κατανομής, $\tilde{F}(t) : \tilde{S}(t) = 1 - \tilde{F}(t)$

$$\text{όπου } \tilde{F}(t) = \frac{\text{Αριθμός παρατηρήσεων που επιβίωσαν μέχρι τον χρόνο } t}{\text{Συνολικός αριθμός παρατηρήσεων}} = \frac{\#t_i \leq t}{n}.$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- **Εμπειρική συνάρτηση επιβίωσης**

Για την εύρεση ενός διαστήματος εμπιστοσύνης για την $S(t)$ αρκεί να ορίσουμε

$$B(t) = \sum_{i=1}^n I(t_i \geq t) \text{ μία διωνυμική μεταβλητή,}$$

τότε $B(t) \sim \text{Binomial}(n, p = S(t))$.

$$\text{Επομένως : } E[\tilde{S}(t)] = \frac{1}{n} np = p = S(t)$$

$$\text{Var}[\tilde{S}(t)] = \frac{1}{n^2} \text{Var}(B(t)) = \frac{1}{n^2} npq = \frac{S(t)(1 - S(t))}{n}.$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- Εμπειρική συνάρτηση επιβίωσης

Όταν n μεγάλο, $\tilde{S}(t) \stackrel{\text{προσεγγ.}}{\sim} \text{Normal}\left(S(t), \frac{S(t)(1-S(t))}{n}\right)$.

Τότε, ένα 95% διάστημα εμπιστοσύνης για την $S(t)$ είναι το εξής :

$$\left(\tilde{S}(t) - 1.96 \sqrt{\frac{\tilde{S}(t)(1-\tilde{S}(t))}{n}}, \tilde{S}(t) + 1.96 \sqrt{\frac{\tilde{S}(t)(1-\tilde{S}(t))}{n}} \right)$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

- **Εμπειρική συνάρτηση επιβίωσης**

Σημειώσεις :

- Εάν το n είναι μικρό ($n < 20$), τότε είναι καταλληλότερο να βρεθούν τα διαστήματα εμπιστοσύνης χρησιμοποιώντας τους πίνακες της διωνυμικής κατανομής.
- Εάν το n είναι μεγάλο ($n \geq 30$), χρησιμοποιούμε την προσέγγιση της κανονικής κατανομής για να εξάγουμε διαστήματα εμπιστοσύνης.
- Η κανονική προσέγγιση δουλεύει καλύτερα όταν $0 \ll \mathcal{S}(t) \ll 1$ (το $\mathcal{S}(t)$ δεν είναι κοντά στο 0 ή το 1). Όταν το $\mathcal{S}(t)$ είναι κοντά στο 0 ή 1, η τεχνική Poisson προσέγγισης είναι καλύτερη.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή εμπειρικής συνάρτησης επιβίωσης

Έχουμε τα εξής δεδομένα 21 ασθενών που πάσχουν από λευχαιμία,

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.

Τα δεδομένα αφορούν (μη λογοκριμένους) χρόνους υποχώρησης (σε εβδομάδες,) (times of remission) ασθενών που ανήκουν στο control group.

Σύμφωνα με τα δεδομένα εκτιμούμε την συνάρτηση επιβίωσης και ένα 95% δ.ε. ($\alpha=0.05$) μέσω της εμπειρικής συνάρτησης επιβίωσης $S(t)$

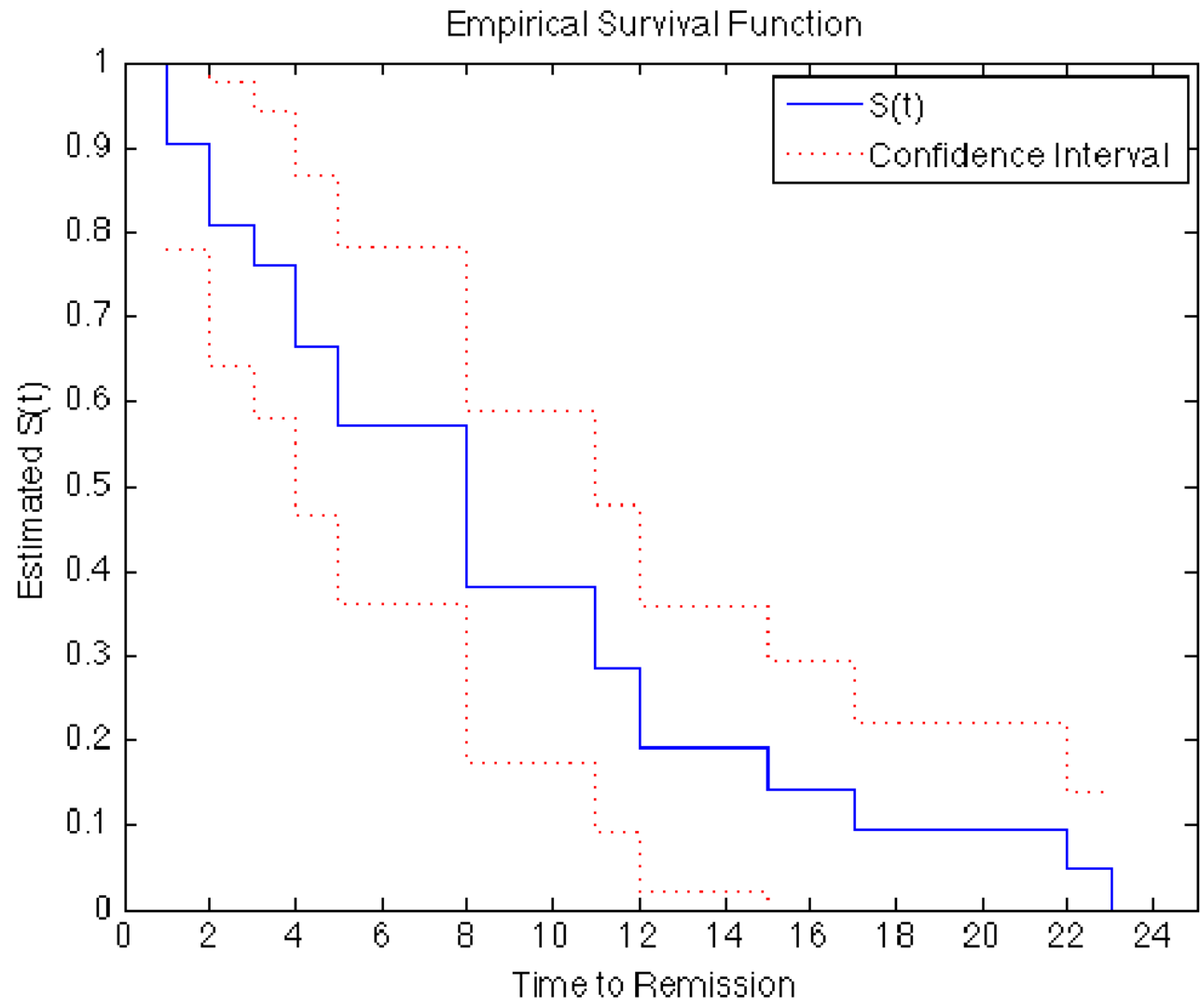
Στον παρακάτω πίνακα παρουσιάζονται οι τιμές της συνάρτησης επιβίωσης για κάθε χρονική στιγμή t , καθώς και τα κάτω και άνω άκρα του δ.ε. σε κάθε περίπτωση.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή εμπειρικής συνάρτησης επιβίωσης

t	$\tilde{\xi}(t)$	Κάτω άκρο δ.ε. (Lower bound)	Άνω άκρο δ.ε. (Upper bound)
1	0.9523	0.7792	1.0000
2	0.8095	0.6416	0.9775
3	0.7619	0.5797	0.9441
4	0.6667	0.4650	0.8683
5	0.5714	0.3598	0.7831
8	0.3810	0.1733	0.5887
11	0.2857	0.0925	0.4789
12	0.1905	0.0225	0.3584
15	0.1429	0	0.2925
17	0.0952	0	0.2208
22	0.0476	0	0.1387
23	0.0000	-	-

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ



ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Η αναλογιστική μέθοδος, ή αλλιώς life-table μέθοδος αποτελεί μία προσέγγιση της μεθόδου Kaplan-Meier που θα παρουσιαστεί παρακάτω και εν αντιθέσει με την εμπειρική συνάρτηση μπορεί να χρησιμοποιηθεί σε **λογοκριμένα δεδομένα**.

Βασίζεται σε ομαδοποιημένους χρόνους επιβίωσης και είναι κατάλληλη για μεγάλο αριθμό δεδομένων.

Οι **πίνακες επιβίωσης (life tables)** είναι μία από τις παλαιότερες τεχνικές για τη μέτρηση της θνησιμότητας και της επιβίωσης ενός πληθυσμού. Αποτελούνται από πίνακες, οι οποίοι συνοψίζουν τις προηγούμενες εμπειρίες ενός πληθυσμού για μία συγκεκριμένη περίοδο και εκδίδονται από τις κρατικές υπηρεσίες, βασισμένοι σε δεδομένα από απογραφές.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Cohort life table : Περιγράφει την εμπειρία θνησιμότητας από τη γέννηση μέχρι τον θάνατο μίας καθορισμένης κοορτής (cohort) ατόμων που γεννιούνται περίπου την ίδια χρονική στιγμή. Τα άτομα που βρίσκονται σε κίνδυνο στην αρχή του διαστήματος είναι εκείνα που επιβίωσαν στο αμέσως προηγούμενο διάστημα.

Current life table : Κατασκευάζεται από απογραφικές πληροφορίες όσον αφορά τον αριθμό των ατόμων που βρίσκονται εν ζωή σε κάθε ηλικία για μία δεδομένη χρονιά και από ζωτικά στοιχεία όσον αφορά τον αριθμό των θανάτων ή αποτυχιών (σε μία δεδομένη χρονιά) με βάση την ηλικία. Ο συγκεκριμένος τύπος life table συνήθως εντοπίζεται σε υποθετικές κοορτές 100.000 ατόμων.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Στους πληθυσμιακούς πίνακες life table η λογοκρισία δεν είναι απαραίτητο να επιτρέπεται, εν αντιθέσει με τους κλινικούς πίνακες life table όπου και η λογοκρισία είναι απαραίτητο να επιτρέπεται, λόγω του ότι οι ασθενείς μπορεί να εισέλθουν στην μελέτη σε διαφορετικές χρονικές στιγμές ή να χαθούν πριν την ολοκλήρωσή της :

Clinical life table : Χρησιμοποιείται σε γκρουπαρισμένα δεδομένα επιβίωσης από μελέτες σε ασθενείς με συγκεκριμένες ασθένειες.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Αρχικά χωρίζουμε το διάστημα παρατήρησης σε k υποδιαστήματα, όχι απαραίτητα ισομήκη αν και τις περισσότερες φορές επιλέγονται ώστε να είναι.

Έστω λοιπόν ότι τα k υποδιαστήματα είναι τα $(t'_1, t'_2), (t'_3, t'_4), \dots, (t'_{k-1}, t'_k)$ και στο j -οστό υποδιάστημα (t'_j, t'_{j+1}) , όπου $j = 1, 2, \dots, k$, παρατηρούμε d_j αποτυχίες και c_j λογοκριμένες παρατηρήσεις αντίστοιχα ενώ n_j είναι ο αριθμός των παρατηρήσεων που έχουν επιβιώσει ή αλλιώς βρίσκονται σε κίνδυνο στην αρχή του j διαστήματος.

Η υπόθεση που χαρακτηρίζει την συγκεκριμένη μέθοδο είναι ότι ο αριθμός των λογοκριμένων παρατηρήσεων σε κάθε ένα από τα k υποδιαστήματα είναι ομοιόμορφα κατανομημένος, έτσι ώστε ο κατά μέσο όρο αριθμός ατόμων σε κίνδυνο μέσα στο

$$\text{υποδιάστημα είναι : } n'_j = n_j - \frac{c_j}{2}.$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Η παραπάνω υπόθεση είναι γνωστή και ως **αναλογιστική υπόθεση (actuarial assumption)**.

Είναι προφανές ότι η πιθανότητα αποβίωσης ενός ατόμου στο j υποδιάστημα είναι : $\frac{d_j}{n_j'}$,

ενώ η αντίστοιχη πιθανότητα επιβίωσης είναι : $\frac{n_j' - d_j}{n_j'}$.

Επομένως, εάν θέλουμε να υπολογίσουμε την πιθανότητα ένα άτομο να επιβιώσει πέραν του υποδιαστήματος r τότε θα πρέπει να επιβιώσει αρχικά σε όλα τα προηγούμενα $r-1$ υποδιαστήματα.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Αναλογιστική μέθοδος

Η εκτιμήτρια της πιθανότητας επιβίωσης λοιπόν, θα προκύψει από το γινόμενο όλων των πιθανοτήτων επιβίωσης και δίνεται από τον τύπο :

$$S^*(t) = \prod_{i=1}^r \frac{n_j' - d_j}{n_i'} = \prod_{i=1}^r \left(1 - \frac{d_j}{n_i'} \right) \quad \text{για } t_r \leq t < t_{r+1}, r = 1, 2, \dots, k$$

Η πιθανότητα επιβίωσης πριν τον χρόνο t_1' είναι ίσος με την μονάδα, ενώ η πιθανότητα επιβίωσης πέραν του χρόνου t_{k+1}' είναι ίση με το μηδέν.

Η αναλογιστική μέθοδος είναι ιδιαίτερα ευαίσθητη στην επιλογή του αριθμού των υποδιαστημάτων ενώ χρησιμοποιείται σε περιπτώσεις όπου δεν είναι γνωστοί οι μεμονωμένοι χρόνοι λογοκρισίας παρά μόνο οι τιμές d_j και c_j .

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Μέθοδος Kaplan-Meier

Μία ακόμη μη παραμετρική μέθοδος είναι η μέθοδος Kaplan-Meier, της οποίας η διαφορά με την αναλογιστική μέθοδο είναι η κατασκευή των υποδιαστημάτων. Οι **Kaplan και Meier (1958)** επέκτειναν την εκτίμηση μέσω της εμπειρικής συνάρτησης κατανομής έτσι ώστε να περιλαμβάνει και λογοκριμένα δεδομένα (συγκεκριμένα χρησιμοποιείται για δεξιά λογοκριμένα δεδομένα όπως και οι προηγούμενες μέθοδοι). Εδώ, κάθε υποδιάστημα κατασκευάζεται έτσι ώστε να αρχίζει από έναν παρατηρούμενο χρόνο επιβίωσης και περιλαμβάνει μόνο μία αποτυχία έτσι ώστε όλοι οι υπόλοιποι χρόνοι που περιλαμβάνονται στο υποδιάστημα να είναι **λογοκριμένοι**.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Μέθοδος Kaplan-Meier

Περισσότερες από μία αποτυχίες μπορούν να συμπεριληφθούν σε ένα υποδιάστημα μόνο αν υπάρχουν άτομα με τον ίδιο χρόνο αποβίωσης (σε αυτή την περίπτωση λέμε ότι οι υπάρχουν κοινοί χρόνοι ή αλλιώς tied deaths). Είναι προφανές ότι οι λογοκριμένες παρατηρήσεις δεν “συνεισφέρουν” (αλλά λαμβάνονται υπ’όψιν) στην συγκεκριμένη διαδικασία.

Αν για παράδειγμα έχουν παρατηρηθεί οι χρόνοι 5,10,10,15,20⁺,30 τότε τα υποδιαστήματα που θα δημιουργηθούν είναι τα : $(0, 5), [5, 10), [10, 15), [15, 30), [30, \infty)$.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Μέθοδος Kaplan-Meier

Έστω ότι είναι γνωστοί οι πλήρεις και οι λογοκριμένοι χρόνοι n ατόμων ($\{t_1, t_2, \dots, t_n\}$) και $t_{(1)} < t_{(2)} < \dots < t_{(m)}$, οι διακεκριμένοι, διατεταγμένοι χρόνοι αποτυχίας (δεν περιλαμβάνονται οι λογοκριμένοι χρόνοι).

Για $j = 1, 2, \dots, m$, αν ορίσουμε n_j τον αριθμό των παρατηρήσεων που έχουν επιβιώσει ή αλλιώς βρίσκονται σε κίνδυνο ακριβώς πριν τη χρονική στιγμή $t_{(j)}$ και d_j τον αριθμό των αποτυχιών τη χρονική στιγμή $t_{(j)}$ τότε :

Η πιθανότητα αποβίωσης στο j υποδιάστημα είναι $\frac{d_j}{n_j}$,

ενώ η πιθανότητα επιβίωσης στο j υποδιάστημα είναι $\frac{n_j - d_j}{n_j}$.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Μέθοδος Kaplan-Meier

Όπως και στην αναλογιστική μέθοδο για να υπολογίσουμε την πιθανότητα ένα άτομο να επιβιώσει στο υποδιάστημα (t_j, t_{j+1}) πρέπει πρώτα το άτομο να επιβιώσει σε όλα τα προηγούμενα υποδιαστήματα. Αυτή η πιθανότητα εκφράζεται ως :

$$\hat{S}_{KM}(t) = \prod_{j: t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j} \right) \quad \text{για } t_j \leq t < t_{j+1}, j = 1, 2, \dots, m$$

Για την εκτιμήτρια Kaplan-Meier ισχύουν : $\hat{S}_{KM}(t) = 1, \forall t < t_{(1)}$ και $\hat{S}_{KM}(t) = 0, \forall t \geq t_{(m)}$.

Σε περίπτωση που δεν υπάρχουν λογοκριμένες παρατηρήσεις, ισχύει ότι $n_j - d_j = n_{j+1}$ και η εκτιμήτρια $\hat{S}_{KM}(t)$ ταυτίζεται με την εμπειρική συνάρτηση $\tilde{S}(t)$.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Μέθοδος Kaplan-Meier

Ο παραπάνω εκτιμητής ονομάζεται και **product limit estimator** γιατί προκύπτει ως το γινόμενο πιθανοτήτων επιβίωσης και πρόκειται για μία δεσμευμένη πιθανότητα αφού προϋποθέτει να έχει επιβιώσει το άτομο όλες τις προηγούμενες στιγμές από αυτήν που ψάχνουμε αν έχει επιβιώσει. Οι τιμές της συνάρτησης παραμένουν σταθερές μεταξύ δύο διαδοχικών χρόνων και παρουσιάζουν άλματα ακριβώς τη στιγμή που θα συμβεί το επόμενο γεγονός (η $\hat{S}_{KM}(t)$ αλλάζει λοιπόν μόνο στους χρόνους αποβίωσης).

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εναλλακτικός τρόπος υπολογισμού του εκτιμητή Kaplan-Meier

Ένας εναλλακτικός τρόπος για τον υπολογισμό του εκτιμητή Kaplan-Meier είναι μέσω του αλγορίθμου ανακατανομής προς τα δεξιά (redistribution to the right algorithm, Efron 1967). Η προσέγγιση του Efron, ανακατανέμει τις συνεισφορές των λογοκριμένων παρατηρήσεων σε όλους τους υπόλοιπους χρόνους στα δεξιά.

Ο αλγόριθμος έχει ως εξής :

Βήμα 1 : Διατάσσουμε τους χρόνους αποτυχίας (ή λογοκρισίας) σε αύξουσα σειρά. Εάν υπάρχουν χρόνοι που ταυτίζονται, βάζουμε τους λογοκριμένους χρόνους στα δεξιά των αποτυχιών (μη λογοκριμένοι).

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εναλλακτικός τρόπος υπολογισμού του εκτιμητή Kaplan-Meier

Βήμα 2 : Τοποθετούμε βάρος $\left(\frac{1}{n}\right)$ σε κάθε χρόνο.

Βήμα 3 : Κινούμενοι από τα αριστερά προς τα δεξιά, κάθε φορά που συναντάμε λογοκριμένη παρατήρηση, κατανέμουμε ομοιόμορφα την μάζα της σε όλους τους χρόνους που βρίσκονται στα δεξιά της.

Βήμα 4 : Υπολογίζουμε το $\hat{S}_{KM}(j)$ αφαιρώντας το τελικό βάρος για τον χρόνο j από το $\hat{S}_{KM}(j-1)$.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εναλλακτικός τρόπος υπολογισμού του εκτιμητή Kaplan-Meier

Παράδειγμα :

Έστω οι χρόνοι :

1, 2, 3⁺, 5, 5, 6⁺, 7, 8

Τότε :

Χρόνοι (Βήμα 1)	Βήμα 2	Βήμα 3α	Βήμα 3β	$\hat{S}_{KM}(t_j)$ (Βήμα 4)
1	$\frac{1}{8} = 0.125$			0.875
2	$\frac{1}{8} = 0.125$			0.75
3 ⁺	$\frac{1}{8} = 0.125$	0		0.75
5	$\frac{2}{8} = 0.25$	$2 \cdot \left[\frac{1}{8} + \left(\frac{1}{8} \cdot \frac{1}{5} \right) \right] = \frac{6}{40} \cdot 2 = 0.3$		0.45
6 ⁺	$\frac{1}{8} = 0.125$	0.15	0	0.45
7	$\frac{1}{8} = 0.125$	0.15	$\frac{6}{40} + \left(\frac{6}{40} \cdot \frac{1}{2} \right) = \frac{18}{80} = 0.225$	0.225
8	$\frac{1}{8} = 0.125$	0.15	0.225	0

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου

Χρησιμοποιούμε τα δεδομένα που περιέχουν λογοκριμένους χρόνους και αφορούν τους ασθενείς που ανήκουν στο drug group

$6^+, 6, 6, 6, 7, 9^+, 10^+, 10, 11^+, 13, 16, 17^+, 19^+, 20^+, 22, 23, 25^+, 32^+, 32^+, 34^+, 35^+$.

Στον πίνακα παρουσιάζονται οι τιμές της συνάρτησης επιβίωσης (εκτιμώμενης μέσω της μεθόδου KM) $S_{KM}(t)$ για κάθε χρονική στιγμή t , καθώς και το δ.ε. (95%) σε κάθε περίπτωση.

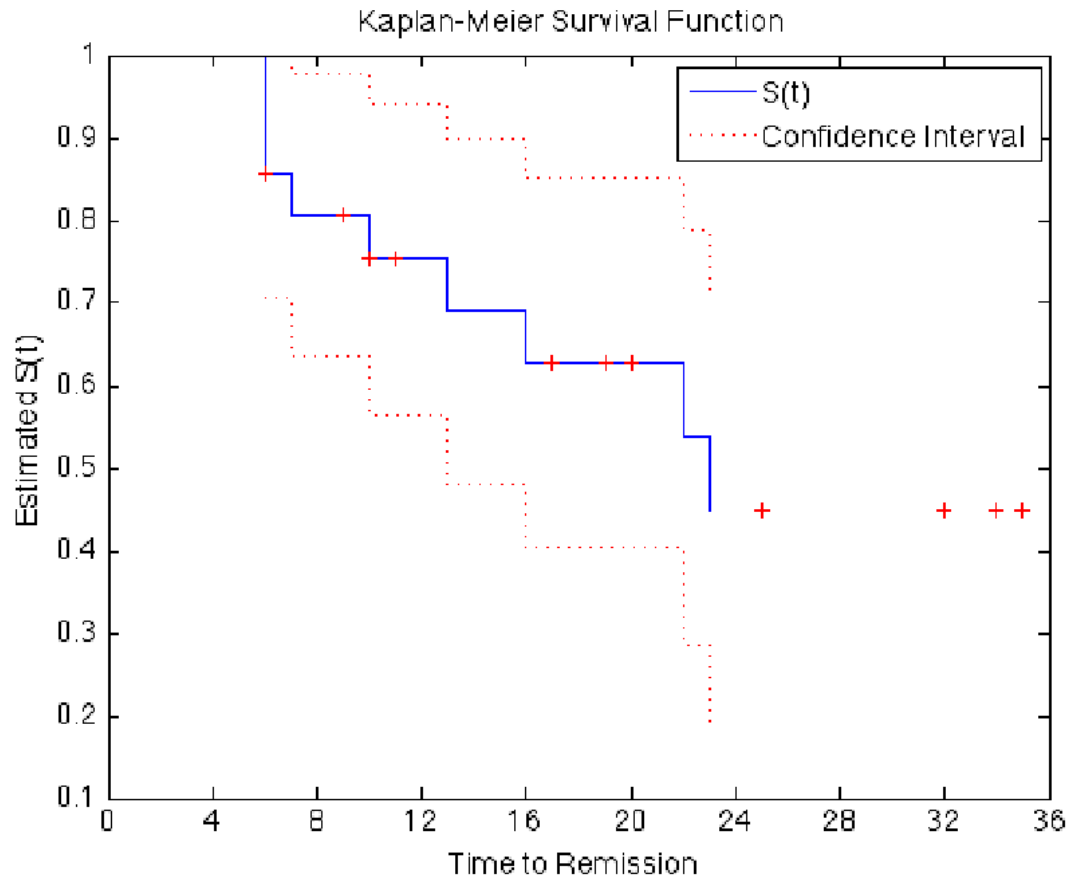
ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου

t	$\hat{S}_{KM}(t)$	Κάτω άκρο δ.ε.	Άνω άκρο δ.ε.
6	0.8571	0.7075	1.0000
7	0.8067	0.6363	0.9771
9	0.8067	0.6363	0.9771
10	0.7529	0.5641	0.9418
11	0.7529	0.5641	0.9418
13	0.6902	0.4808	0.8995
16	0.6275	0.4039	0.8510
17	0.6275	0.4039	0.8510
19	0.6275	0.4039	0.8510
20	0.6275	0.4039	0.8510
22	0.5378	0.2865	0.7891
23	0.4482	0.1844	0.7120
25	0.4482	0.1844	0.7120
32	0.4482	0.1844	0.7120
34	0.4482	0.1844	0.7120
35	0.4482	0.1844	0.7120

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου



Όσοι χρόνοι είναι λογοκριμένοι έχουν συμβολιστεί με “+” στο σχήμα ενώ παρατηρούμε ότι η συνάρτηση επιβίωσης είναι μία βηματική συνάρτηση η οποία φθίνει ως προς t και εμφανίζει ασυνέχειες ή άλματα στους παρατηρούμενους χρόνους επιβίωσης.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Συγκεντρωτικά

Όταν οι εφαρμογές εμπεριέχουν από δεξιά αποκομμένες παρατηρήσεις χρησιμοποιείται η εκτιμήτρια Kaplan-Meier της συνάρτησης επιβίωσης, η οποία έχει αποκτήσει πολύ μεγάλη χρηστική αξία. Έστω τυχαίο δείγμα N μονάδων, μερικές εκ των οποίων καταστρέφονται κατά τις διακεκριμένες χρονικές στιγμές $t_{(1)} < t_{(2)} < \dots < t_{(k)}$, $k \leq N$. Έστω ότι κατά τη χρονική στιγμή $t_{(j)}$ καταστρέφονται d_j μονάδες, ενώ αμέσως πριν από τη στιγμή αυτή λειτουργούσαν N_j μονάδες. Ο αριθμός N_j περιλαμβάνει όλες τις μονάδες που γνωρίζουμε ότι λειτουργούν εκείνη τη στιγμή, δηλαδή αυτές με χρόνο λειτουργίας $t \geq t_{(j)}$, ανεξάρτητα από το αν στη συνέχεια θα διακοπεί η λειτουργία τους ή θα συνεχίζουν να λειτουργούν μετά το πέρας του πειράματος (αποκομμένες τιμές). Δεν περιλαμβάνει τις μονάδες που έχουν ήδη καταστραφεί, ούτε τις μονάδες με αποκομμένες τιμές πριν τη στιγμή $t_{(j)}$.

Λογοκρισίες

Τα *λογοκριμένα δεδομένα* είναι αυτά για τα οποία δεν είναι γνωστός ο χρόνος επιβίωσης. Το μόνο που μπορεί να σημειωθεί είναι ότι ο χρόνος επιβίωσης είναι μεγαλύτερος από την τιμή που έχει καταγραφεί.

Τα δεδομένα αυτά εμφανίζονται σε περίπτωση που ο ασθενής είναι ακόμη ζωντανός στο τέλος της κλινικής μελέτης ή για διάφορους λόγους έχει απομακρυνθεί από αυτή.

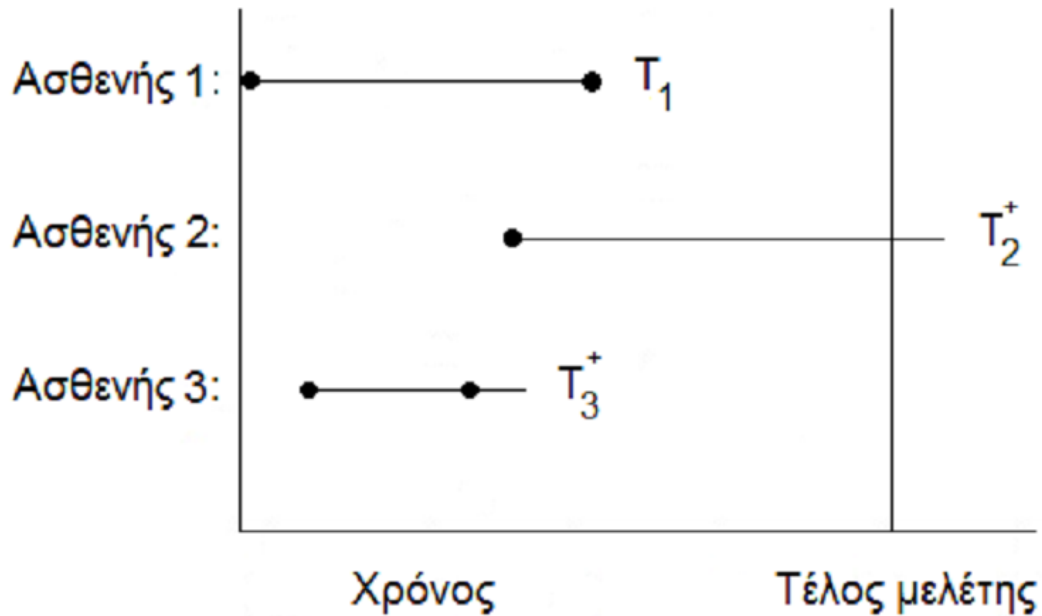
Σε μια έρευνα για την μελέτη της αποτελεσματικότητας μιας νέας θεραπείας για μια ασθένεια, η μεταβλητή που μας ενδιαφέρει είναι ο χρόνος που θα επιζήσει ο ασθενής, δηλαδή ο χρόνος επιβίωσης κάθε ατόμου. Οι ασθενείς μπορεί να εισέρχονται στη μελέτη σε διαφορετικούς χρόνους, ενώ η διάρκεια της μελέτης είναι προκαθορισμένη. Επομένως, για κάθε ασθενή καταγράφεται ο χρόνος από την είσοδό του στη μελέτη μέχρι τον θάνατό του. Στο τέλος της μελέτης είναι πιθανό να υπάρχουν ασθενείς που ζουν ακόμη ενώ θα υπάρχουν ασθενείς με τους οποίους χάθηκε η επαφή. Τον ακριβή χρόνο επιβίωσης των ατόμων αυτών δεν τον ξέρουμε, ξέρουμε όμως ότι είναι μεγαλύτερος από το χρόνο που παρεμβάλλεται από την είσοδό τους στη μελέτη μέχρι την ολοκλήρωση της μελέτης (στην πρώτη περίπτωση) και μέχρι την στιγμή που χάθηκε η επαφή (στην δεύτερη). Αυτές οι παρατηρήσεις είναι λογοκριμένες.

Λογοκρισίες

Σίγουρα, δεν πρέπει να αποκλείσουμε αυτά τα δεδομένα από την μελέτη θεωρώντας τα ως ελλιπή. Κάτι τέτοιο θα επηρέαζε τόσο την ανάλυση όσο και τα αποτελέσματα τα οποία δεν θα ήταν σωστά, αφού οι περισσότεροι από τους ασθενείς αυτούς έχουν ξεπεράσει τον χρόνο μελέτης και επομένως μας οδηγούν στο συμπέρασμα της αποτελεσματικότητας της θεραπείας.

Όπως ειπώθηκε και πιο πάνω οι λογοκριμένες παρατηρήσεις δεν προκύπτουν μόνο λόγω του χρόνου λήξης της μελέτης αλλά μπορεί να προκύψουν και όταν ο ασθενής χάνεται από την παρακολούθηση (loss to follow-up) (ο ασθενής μπορεί να αποφάσισε να μετακομίσει ή να αλλάξει γιατρό ή νοσοκομείο) ή αποσύρεται από την παρακολούθηση (drop-out) (η θεραπεία έχει πολύ κακές επιδράσεις και ο ασθενής είναι αναγκαίο να σταματήσει ή δεν θέλει να λάβει μέρος σε μια τέτοια διαδικασία μετά από ένα χρονικό διάστημα).

Λογοκρισίες



- Ο ασθενής 1 εισέρχεται στην μελέτη τον χρόνο 0 και πεθαίνει στο T_1 , δίνοντας έτσι μια πλήρη παρατήρηση.
- Ο ασθενής 2 εισέρχεται αργότερα και μέχρι το τέλος της μελέτης παραμένει ζωντανός μέχρι το T_2 τουλάχιστον, δίνοντας έτσι μια λογοκριμένη παρατήρηση που συμβολίζεται με T_2^+ .
- Ο ασθενής 3 εισέρχεται κι αυτός αργότερα αλλά χάνεται από την παρακολούθηση πριν από το τέλος της μελέτης, δίνοντας έτσι την λογοκριμένη παρατήρηση T_3^+ .

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Συγκεντρωτικά

Η εκτιμήτρια Kaplan-Meier ορίζεται ως εξής:

$$\hat{S}(t) = \frac{N_1 - d_1}{N_1} \frac{N_2 - d_2}{N_2} \dots \frac{N_i - d_i}{N_i}$$

όπου n_i ο αριθμός των μονάδων που ήταν σε λειτουργία ακριβώς πριν από τη χρονική στιγμή t_i και $i : t_i \leq t < t_{i+1}$.

Επομένως

$$\hat{S}(t) = \begin{cases} \prod_{j: t_{(j)} \leq t} \frac{N_j - d_j}{N_j}, & \text{όταν } t \geq t_{(1)} \\ 1, & \text{όταν } t \leq t_{(1)} \end{cases}$$

και

$$se(\hat{S}(t)) = \hat{S}(t) \left\{ \sum_{t_{(j)} \leq t} \frac{d_j}{N_j(N_j - d_j)} \right\}^{1/2}$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Συγκεντρωτικά

Διακόμανση Kaplan-Meier εκτιμητή (Greenwood):

$$\text{Var} \left\{ \hat{S}(t) \right\} = \left(\hat{S}(t) \right)^2 \sum_{j=1}^k \frac{d_j}{n_j (n_j - d_j)} \quad t \in [t_k, t_{k+1})$$

Τυπική Απόκλιση Kaplan-Meier εκτιμητή:

$$s.e \left\{ \hat{S}(t) \right\} = \left(\hat{S}(t) \right) \left\{ \sum_{j=1}^k \frac{d_j}{n_j (n_j - d_j)} \right\}^{1/2}$$

95% Διάστημα Εμπιστοσύνης για τη συνάρτηση επιβίωσης:

$$\hat{S}(t) \pm 1.96 \cdot s.e \left\{ \hat{S}(t) \right\}$$

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Παράδειγμα

Έστω ότι παρατηρούνται οι παρακάτω χρόνοι (διάρκεια) ύφεσης δέκα ασθενών με συμπαγής όγκους (οι καρκίνοι που αναπτύσσονται από συμπαγή όργανα και ιστούς). Έξι (6) ασθενείς υποτροπιάζουν στους 3.0, 6.5, 6.5, 10, 12 και 15 μήνες αντίστοιχα. Επιπλέον ένας (1) ασθενής παύει να παρακολουθείται στους 8.4 μήνες, ενώ τρεις (3) εξακολουθούν να υποτροπιάζουν στο τέλος της μελέτης, μετά από 4.0, 5.7 και 10 μήνες.

Στον πίνακα που ακολουθεί καταγράφονται η διάρκεια ύφεσης των ασθενών καθώς και ο υπολογισμός της συνάρτησης επιβίωσης $S(t)$.

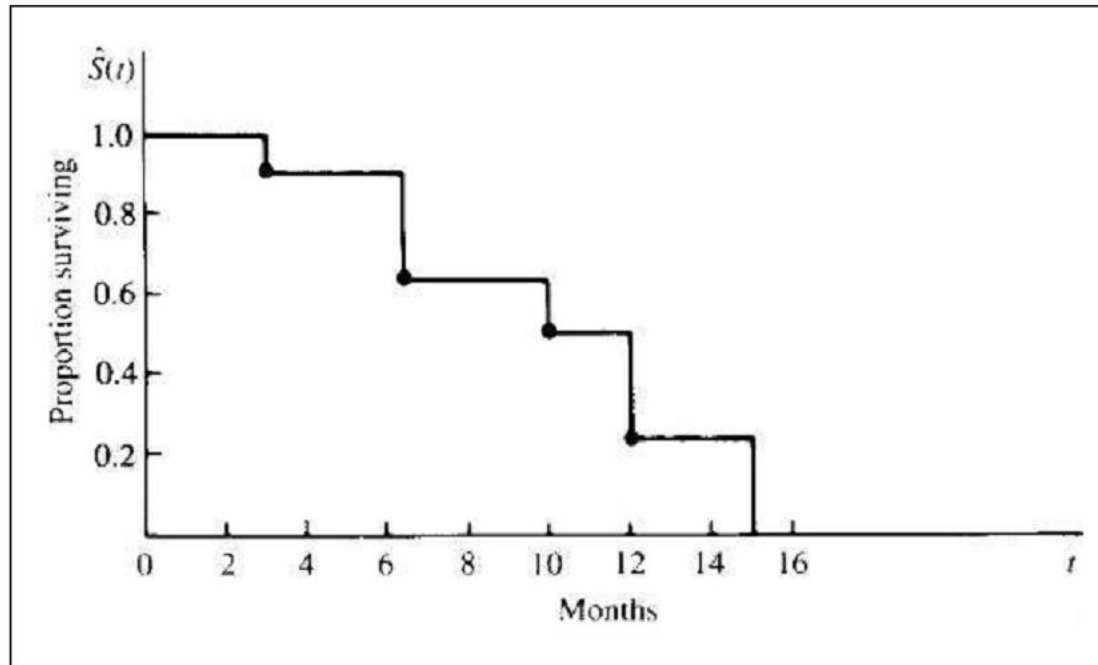
ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Παράδειγμα

Διάρκεια ύφεσης T	# Ασθενών	Μη-Αποκομμένες Παρατηρήσεις	Υπολογισμός της p_j	$\widehat{S}(t) = \prod \frac{n_j - d_j}{n_j}$
3.0	10	1	$9/10$	$9/10 = 0.9$
4.0+	9	-	-	-
5.7+	8	-	-	-
6.5	7	4	$6/7$	0.771
6.5	6	5	$5/6$	0.643
8.4+	5	-	-	-
10	4	7	$3/4$	0.482
10+	3	-	-	-
12	2	9	$1/2$	0.241
15	1	10	0	0

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

Εφαρμογή Kaplan-Meier μεθόδου - Παράδειγμα



ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

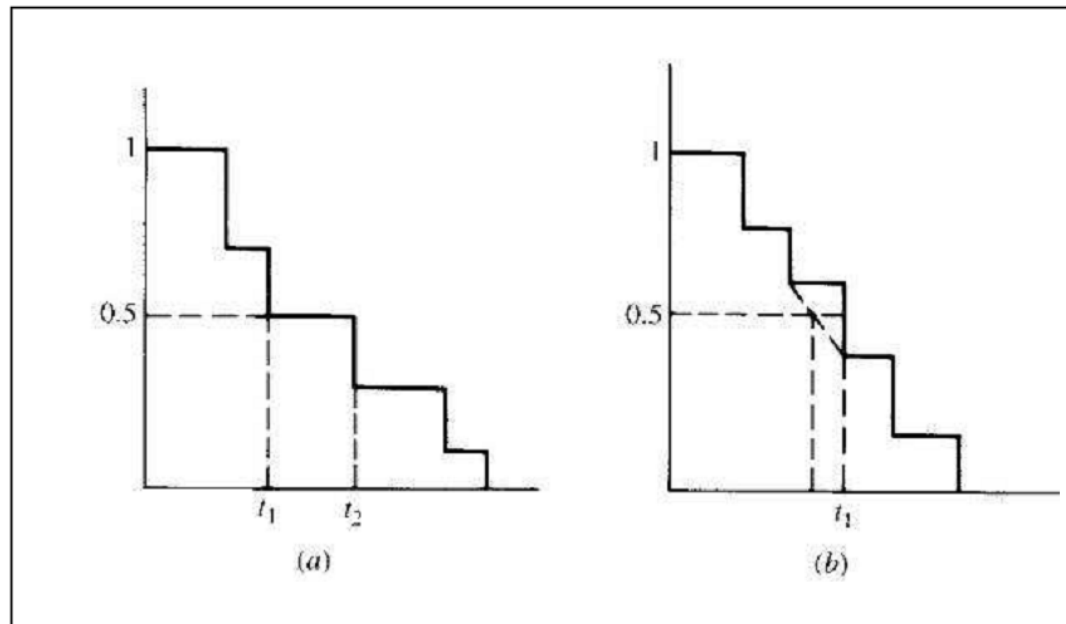
Από τη μέθοδο Kaplan – Meier, προκύπτουν χρήσιμες εκτιμήσεις της πιθανότητας επιβίωσης και της γραφικής παράστασης της συνάρτησης επιβίωσης, κάνοντας την να είναι η πλέον διαδεδομένη μέθοδος για ανάλυση δεδομένων επιβίωσης.

Καταλήγοντας, σημαντικό είναι να αναφερθούν κάποια κύρια χαρακτηριστικά της μεθόδου αυτής.

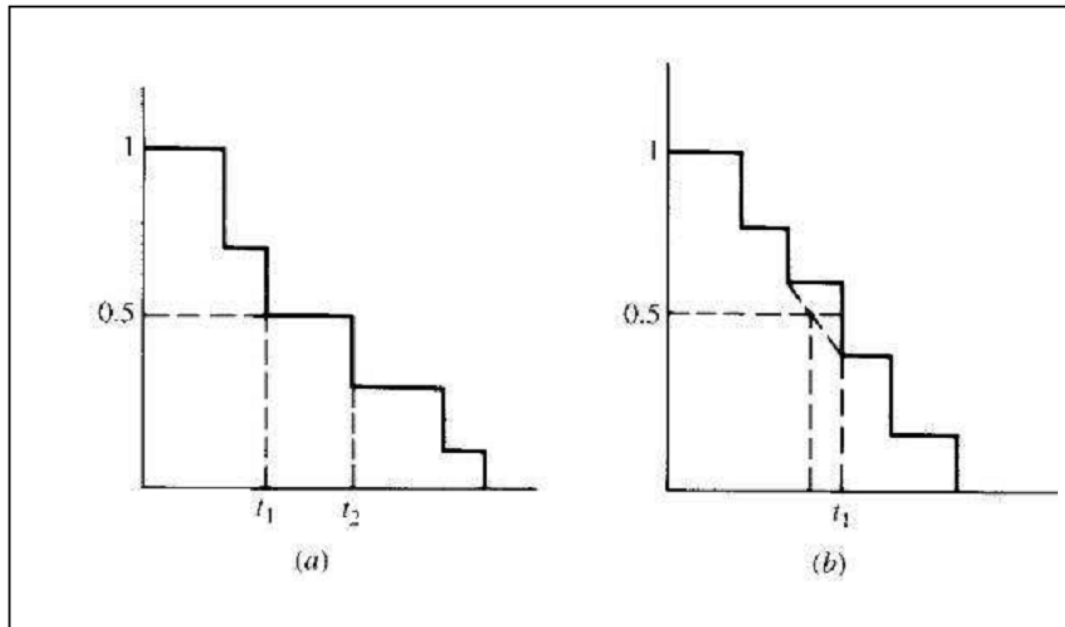
1. Οι εκτιμήσεις της μεθόδου Kaplan – Meier περιορίζονται στο χρονικό διάστημα στο οποίο συμβαίνουν. Αν η μέγιστη παρατήρηση είναι μη αποκομμένη η εκτιμήτρια γινομένου-ορίου ισούται με μηδέν, για τον λόγο ότι κανείς από το δείγμα δεν ζει πλέον. Αντίθετα, αν η μέγιστη παρατήρηση είναι αποκομμένη τότε η εκτιμήτρια δεν μπορεί να είναι μηδέν και δεν μπορεί πλέον να οριστεί πέραν της παρατήρησης αυτής.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

2. Ο διάμεσος χρόνος επιβίωσης είναι ο χρόνος επιβίωσης, στον οποίο το 50% των υπό μελέτη ατόμων επιβιώνει, και από αυτόν έχουμε μια περιληπτική στατιστική εικόνα στην ανάλυση επιβίωσης. Μια απλή εκτίμηση της διάμεσου προκύπτει από την καμπύλη επιβίωσης, όσο ο χρόνος t για τον οποίο ισχύει $S(t) = 0.5$, κάτι όμως που μπορεί να μην έχει μοναδική λύση.



ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ



Από το πιο πάνω σχήμα παρατηρούμε ότι για $S(t) = 0.5$, έχουμε διάφορες τιμές του διάμεσου χρόνου t . Έτσι για τον υπολογισμό του διάμεσου χρόνου, στην πρώτη περίπτωση, (α), παίρνουμε τη μέση τιμή των t_1, t_2 , ενώ στην περίπτωση (β), ενώνουμε τα δύο σημεία και εντοπίζουμε τον διάμεσο.

ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΡΤΗΣΗΣ ΕΠΙΒΙΩΣΗΣ

3. Αν λιγότερες από τις μισές παρατηρήσεις είναι μη αποκομμένες, ενώ η μέγιστη παρατήρηση είναι, τότε, ο διάμεσος χρόνος επιβίωσης δεν μπορεί να υπολογιστεί. Σε αυτή την περίπτωση, χρησιμοποιούμε τις πιθανότητες επιβίωσης για ένα συγκεκριμένο χρόνο π.χ. 3 χρόνια ή τον μέσο χρόνο επιβίωσης για δοσμένο χρόνο t .