Ανάλυση Επιβίωσης

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

Εισαγωγή

- Σκοπός του μαθήματος είναι η ανάλυση δεδομένων που αφορούν την θνησιμότητα ζώντων οργανισμών. Τα δεδομένα θνησιμότητας προέρχονται από τις καταγραφές των χρόνων θανάτου κάποιων ατόμων που ανήκουν σε κάποια καθορισμένη ομάδα. Επίσης έχουμε πληροφορίες για συμπληρωματικά χαρακτηριστικά όπως αίτια θανάτου, διάρκεια της ασθένειας, φυσικά χαρακτηριστικά καταγραμμένα λίγο πριν την στιγμή του θανάτου ή ακομα φύλο, ηλικία, οικογενειακή ιστορία.

Έκθεση στον κίνδυνο

- Στις αναλύσεις δεδομένων επιβίωσης, μας ενδιαφέρει, η μελέτη της αναλογίας των θανάτων μεταξύ ομάδων ατόμων κάτω από συγκεκριμένες συνθήκες.
- Όσο μεγαλύτερη είναι η περίοδος παρακολούθησης κάπου ατόμου τόσο μεγαλύτερη είναι η δυνατότητα να καταγράψουμε κάποτε τον θάνατό του.
- Η συγκρισιμότητα των αριθμών θανάτων απαιτεί την χρήση κάποιας μοναδιαίας χρονικής περιόδου παρακολούθησης. Επομένως πρέπει να ξέρουμε για κάθε άτομο την περίοδο έκθεσης στο κίνδυνο δηλαδή την χρονική περίοδο κατά την διάρκεια της οποίας θα συμβεί ο θάνατος και θα καταγραφεί στους παραχωρηθέντους θανάτους.

Έκθεση στον κίνδυνο

- Είναι αναγκαίο να περιγράψουμε την ταχύτητα των αλλαγών σε μια κοινωνία ζώντων οργανισμών γεγονός που μας οδηγεί στην χρήση μέτρων αναφοράς (π.χ. θέλουμε να μελετήσουμε πόσο γρήγορα εξαπλώνεται μια επιδημία και πόσο γρήγορα κοπάζει. Κατάλληλα μέτρα είναι οι ρυθμοί.

Λόγος

Δύο είδη λόγων χρησιμοποιούνται:

1. Για να συγκρίνει τις συχνότητες δύο αμοιβαίων αποκλειόμενων κλάσεων

$$\text{λόγος φύλου} = \frac{\text{αριθμός ανδρών}}{\text{αριθμός γυναικών}}$$

Έκθεση στον κίνδυνο

- Παράδειγμα :

$$\text{λόγος εμβρυικών θανάτων} = \frac{\text{αριθμός εμβρυικών θανάτων}}{\text{αριθμός ζώντων νεογνών}}$$

Ο λόγος εκφράζει τον αριθμών των εμβρυικών θανάτων σε σύγκριση με τις γεννήσεις ζώντων νεογνών.

Έκθεση στον κίνδυνο

- **Αναλογία**

Είναι είδος λόγου όπου ο αριθμητής είναι μέρος του παρονομαστή.

$$p = \frac{\alpha}{\alpha+\beta}$$

1. Αν α, β είναι ακέραιοι και παριστάνουν συχνότητες τότε το p είναι σχετικές συχνότητες, π.χ.

$$\frac{\text{αριθμός ανδρών}}{\text{αριθμός ανδρών} + \text{αριθμός γυναικών}}$$

δίνει την αναλογία των ανδρών σε συνολικό πληθυσμό

Έκθεση στον κίνδυνο

Σε μεγάλους πληθυσμούς η αναλογία μπορεί να προσδιορίσει την πιθανότητα ενός συγκεκριμένου ενδεχομένου. Σε δείγμα η αναλογία μπορεί να χρησιμοποιηθεί σαν εκτίμηση της πιθανότητας ενός ενδεχομένου.

π.χ. Η ποσότητα

$$\frac{\text{αριθμός εμβρυικών θανάτων}}{\text{αριθμός εμβρυικών θανάτων} + \text{αριθμός ζώντων εμβρύων}}$$

Εκτιμά την πιθανότητα θανάτου ενός εμβρύου πριν την γέννηση του – ρυθμός εμβρυικών θανάτων

Ρυθμός Συνεχών διαδικασιών

- **Ρυθμός**

Ορίζεται ως το μέτρο της μεταβολής της μεταβλητής y όταν η μεταβλητή x από την οποία εξαρτάται μεταβάλλεται κατά μονάδα.

$$p = \frac{\alpha}{\alpha+\beta}$$

- **Απόλυτος Ρυθμός**

Έστω $y = y(t)$ συνεχής, αυστηρά, μονότονη συνάρτηση του t, που περιγράφει μια χρονικά εξαρτημένη διαδικασία.

Έστω $\Delta t\,(t > 0)$ μικρό χρονικό διάστημα και

$$\Delta y = y(t + \Delta t) - y(t)$$

η μεταβολή του y στο Δt.

Ρυθμός Συνεχών διαδικασιών

Τότε το

$$a(t, t+\Delta t) = \frac{\Delta y}{\Delta t} = \frac{dy}{dt}\Big|_{t=t'}$$

για κάποιο t' (t,t+Δt) καλείται μέσος απόλυτος ρυθμός.

Αν

$$\lim_{\Delta t \to 0} \frac{y(t+\Delta t) - y(t)}{\Delta t} = \frac{dy}{dt} = \beta(t)$$

τότε το β(t) καλείται απόλυτος στιγμιαίος ρυθμός

α(t): μέση ταχύτητα της διαδικασίας στο (t,t+Δt)

β(t): στιγμιαία ταχύτητα σε οποιοδήποτε σημείο t. Περιγράφει τον τρόπο μεταβολής του y(t) ως προς τον χρόνο.

Ρυθμός Συνεχών διαδικασιών

- **Σχετικός ρυθμός**

  Εκθετικό αυξητικό μοντέλο

Έστω N(t) το μέγεθος του πληθυσμού στο χρόνο t. Υποθέτουμε ότι N(t) είναι συνεχής συνάρτηση του t. Αν

dN(t) η μεταβολή του πληθυσμού στο διάστημα dt είναι ανάλογη του N(t) και του dt τότε

$$dN(t) = kN(t)dt$$

Τότε $k = \frac{1}{N(t)} \frac{dN(t)}{dt}$ είναι η σταθερά ρυθμού ανάπτυξης

και $N(t) = N(0)\exp\left[\int_0^t kdt\right] = N(0)e^{k-t}$ όπου N(0) πληθυσμός στο χρόνο t=0.

Ρυθμός Συνεχών διαδικασιών

- **Συνάρτηση ρυθμού κινδύνου.**

Για κάθε συνεχής συνάρτηση κατανομής P{X≤x}=F(x), ο στιγμιαίος ρυθμός που ονομάζεται και συνάρτηση ρυθμού κινδύνου είναι:

$$\mu(x) = \frac{1}{1-F(x)}\frac{dF(x)}{dx} = -\frac{d\ln(1-F(x))}{dx}$$

- **Μέσος κεντρικός ρυθμός**

Αν y(t) άγνωστο, τότε υπολογισμός μέσου ρυθμού b(t,t+Δt)

$$b(t,t+\Delta t) = \frac{1}{y(t')}\frac{\Delta y}{\Delta t} = \frac{\Delta y}{y(t')\Delta t}$$

Ρυθμός Συνεχών διαδικασιών

- **Εκτίμηση του y(t᾿)**

Έστω y(t) περιγράφει διαδικασία χημικής αντίδρασης

Η y(t᾿) είναι η διαθέσιμη μάζα στο σημείο t᾿$\in$(t,t+Δt). Αν υποθέσουμε ότι y(t) είναι κατά προσέγγιση γραμμική στο διάστημα (t,t+Δt), τότε το t᾿ μπορεί να θεωρηθεί μέσο του διαστήματος. Άρα αν $t' = t + \frac{1}{2}\Delta t$ τότε

$$y(t') = y\left(t + \frac{1}{2}\Delta t\right)$$

Ρυθμός Συνεχών διαδικασιών

Σε κάποιες περιπτώσεις η y(t᾿) προσεγγίζεται από την αρχική τιμή y(t). Π.χ

αν N(t) και N(t+Δt) είναι τα μεγέθη ενός πληθυσμού στους χρόνους t και t+Δt, τότε ο μέσος ρυθμός ανάπτυξης ανά μονάδα χρόνου και 100 άτομα είναι

$$r\left(t, t+\Delta t\right) = \frac{N\left(t+\Delta t\right) - N\left(t\right)}{N\left(t\right)\Delta t} 100$$

Πληθυσμός σε κίνδυνο

- **Λογοκρισία στα δεδομένα (censored data)** προκύπτει εάν έχουμε κάποια πληροφορία για τον χρόνο στον οποίο συνέβη το γεγονός (πρίν ή μετά από κάποιον γνωστό χρόνο) αλλά δεν γνωρίζουμε τον ακριβή χρόνο επιβίωσης.

Η λογοκρισία μπορεί να συμβεί εάν :

- Το άτομο δεν βιώνει το γεγονός πρίν **τελειώσει η μελέτη (the study ends).**
- Το άτομο χάνεται από τη μελέτη **πρίν την ολοκλήρωσή της (lost to follow-up).**
- Το άτομο **αποχωρεί από τη μελέτη (withdraws from the study)** λόγω θανάτου ή άλλου λόγου που μπορεί να σχετίζεται με την ασθένειά του.

Πληθυσμός σε κίνδυνο

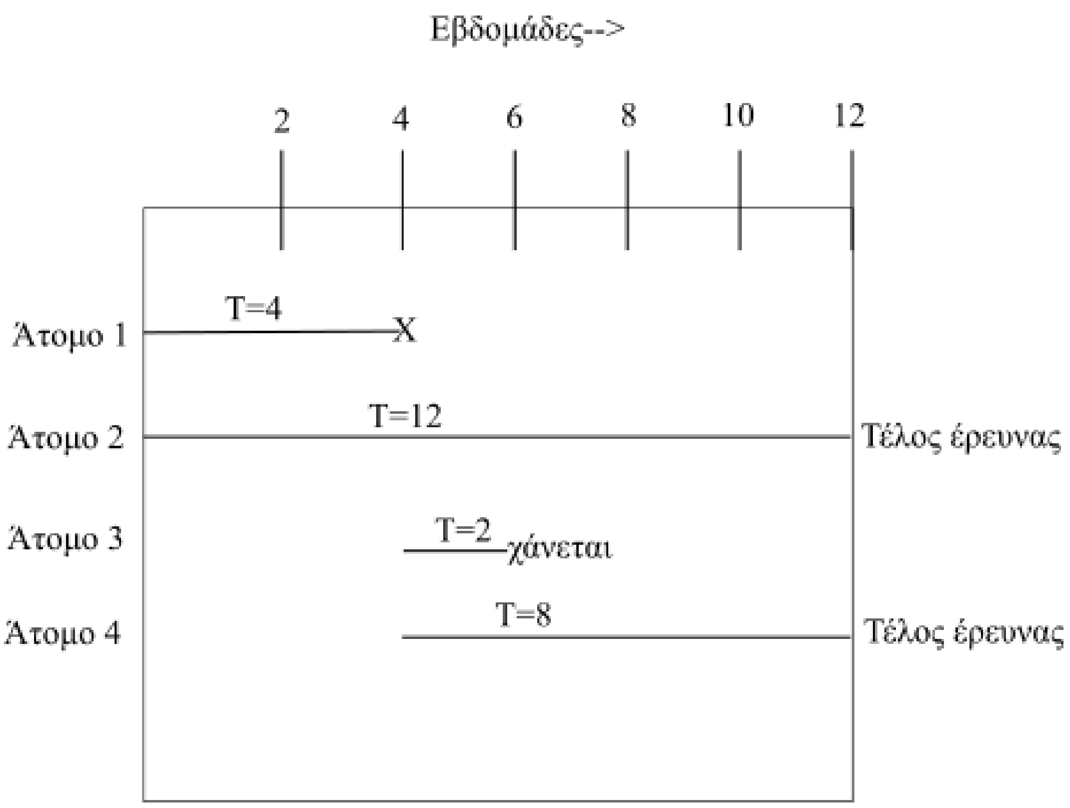
- Επιλέγοντας σαν χρονική μονάδα το έτος, θα αναφερθούμε στους ετήσιους ρυθμούς θανάτου.

Έστω N_T αριθμός ατόμων που έχουν παρακολουθηθεί κατά την χρονική περίοδο T ετών. Έστω τ_j η διάρκεια της χρονικής περιόδου (σε έτη) κατά την διάρκεια της οποίας το άτομο j ήταν υπό παρακολούθηση – εκτεθειμένο στο κίνδυνο θανάτου. Το άθροισμα της χρονικής περιόδου έκθεσης στον κίνδυνο είναι

$$A_T = \sum_{j=1}^{N_T} 1.\tau_j$$

και δίνει τον συνολικό αριθμό ατόμων – ετών εκτεθειμένων στο κύνδυνο

Πληθυσμός σε κίνδυνο



Πληθυσμός σε κίνδυνο

- **Το Άτομο 1** επιβιώνει 4 εβδομάδες όπου και συμβαίνει το γεγονός.
- **Το Άτομο 2** είναι λογοκριμένο (δεξιά) εφόσον επιβιώνει μέχρι το τέλος του χρόνου παρατήρησης της έρευνας και το μόνο που γνωρίζουμε είναι ότι έχει επιβιώσει τουλάχιστον 12 εβδομάδες.
- **Το Άτομο 3** εισέρχεται στην έρευνα την $4^{η}$ εβδομάδα και χάνεται την $6^{η}$ εβδομάδα. Κατά συνέπεια ο χρόνος επιβίωσης του συγκεκριμένου ατόμου είναι λογοκριμένος (δεξιά) μετά τις 2 αυτές εβδομάδες.
- **Το Άτομο 4** εισέρχεται στην έρευνα την $4^{η}$ εβδομάδα και επιβιώνει μέχρι το τέλος. Ο λογοκριμένος (δεξιά) χρόνος επιβίωσης εδώ είναι 8 εβδομάδες.

Πληθυσμός σε κίνδυνο

- Αν N(t) το πληθυσμιακό μέγεθος την χρονική στιγμή t, και υποθέσουμε ότι η N(t) είναι συνεχής συνάρτηση του t τότε

$$A_T = \int_0^T N(t)dt = TN(t')$$

για κάποιο t’$\in$(0,T). Άρα

$$N(t') = \frac{A_T}{T}$$

και καλείται μέσος πληθυσμός εκτεθειμένος στον κίνδυνο.

Πληθυσμός σε κίνδυνο

- Παράδειγμα

Έστω T=3, περίοδος παρακολούθησης. Αν $N_T = 10$, αριθμός ατόμων υπό παρακολούθηση κατά την περίοδο T και τ_j η διάρκεια παρακολούθησης του ατόμου j, με

τ_j : 2.3, 1.5, 2.8, 2.5, 3, 1.8, 2.7, 2.5, 3, 3 έτη

Ποιος είναι ο αριθμός των ατόμων – ετών

$$A_T = 2.3 + 1.5 + ... + 3 = 24.3$$

Ποιο είναι το μέσο μέγεθος του πληθυσμού εκτεθειμένο στον κίνδυνο

$$N\left(t'\right) = \frac{24.3}{3} = 8.1 \approx 8$$