

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ



---

## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΝΑΛΟΓΙΣΤΙΚΩΝ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ ΜΑΘΗΜΑΤΙΚΩΝ

---

### «CORULAS ΚΑΙ ΠΟΛΛΑΠΛΑ ΑΠΑΥΞΗΜΑΤΑ ΓΙΑ ΒΑΣΙΚΕΣ ΑΙΤΙΕΣ ΘΑΝΑΤΩΝ»

---

---

**ΛΥΜΠΕΡΟΠΟΥΛΟΥ Ι. ΔΙΟΝΥΣΙΑ ΜΑΡΙΑ**

A.M.: 331/13092

Επιβλέπων καθηγητής: Χατζόπουλος Π.

Επιτηρητές καθηγητές: Ζήμερας Σ. – Τσιμήκας Ι.

ΙΟΥΝΙΟΣ 2017

*Στους γονείς μου*

## Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του Τμήματος Στατιστικής και Αναλογιστικών Χρηματοοικονομικών Μαθηματικών του Πανεπιστημίου Αιγαίου και επιβλέποντα κ. Πέτρο Χατζόπουλο για την πολύτιμη βοήθεια του, την καθοδήγηση αλλά και την κατανόηση που έδειξε σε όλο το διάστημα εκπόνησης της παρούσας εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τους καθηγητές του τμήματος της Στατιστικής για τις γνώσεις που μου προσέφεραν. Τέλος, θα ήθελα να ευχαριστήσω όλους εκείνους που ήταν δίπλα μου και με στήριξαν, μα πάνω από όλους την οικογένειά μου και ιδιαίτερα τους γονείς μου που μου συμπαραστέκονται και με στηρίζουν σε κάθε περίοδο της ζωής μου.



## ΠΕΡΙΛΗΨΗ

Τις τελευταίες δεκαετίες, οι συζεύξεις (copulas) έχουν αναγνωρισθεί ως ένα πολύ σημαντικό εργαλείο στα Αναλογιστικά - Χρηματοοικονομικά και στη Βιοστατιστική. Όλο και περισσότεροι επιστήμονες σε διαφορετικούς επιστημονικούς κλάδους ασχολούνται με τα Copulas καθώς είναι πολλές οι περιπτώσεις στις οποίες τα δεδομένα που μελετώνται παρουσιάζουν συσχέτιση. Η έννοια των Copulas εισήχθη πρώτη φορά το 1959 από τον A. Sklar. Μια σύζευξη (Copula) είναι μια συνάρτηση που συνδέει μια πολυδιάστατη συνάρτηση κατανομής με τις περιθώριες μονοδιάστατες συναρτήσεις κατανομών της.

Σε αυτή την εργασία θα παρουσιάσουμε τους γενικούς ορισμούς των πολυδιάστατων και δισδιάστατων copulas και τις ιδιότητές τους. Θα συζητήσουμε για τις σημαντικότερες οικογένειες των δισδιάστατων συζεύξεων, όπως είναι οι ελλειπτικές και οι Αρχιμήδειες συζεύξεις και κάποιες εφαρμογές τους.

Επιπλέον, θα δούμε κάποια καθεστάτα πολλαπλών απαυξημάτων, όπου υπάρχουν πολλές καταστάσεις που μπορεί να μεταβεί κανείς και όλες είναι απορροφητικές.

Τέλος, περιγράφουμε τις δυο βασικότερες αιτίες θανάτου, τα κακοήθη νεοπλάσματα και την ισχαιμική καρδιοπάθεια. Μέσα από τα μοντέλα Copulas βρίσκουμε την από κοινού συνάρτηση επιβίωσης και τη συσχέτιση της συνάρτησης επιβίωσης από κακοήθη νεοπλάσματα με τη συνάρτηση επιβίωσης από όλες τις άλλες αιτίες. Προσομοιώνουμε δεδομένα για να τα εφαρμόσουμε σε συγκεκριμένο Copulas.

## **ABSTRACT**

In recent decades, copulas have been identified as a very important tool in Actuarial- Finance and Biostatistics. An increasing number of scientists in different scientific fields are getting interested in copulas, as many cases where the data studied are correlated. The notion of Copula was introduced by A. Sklar in 1959. A Copula is a function that links univariate marginals to their full multivariate distribution.

In this paper work we will present the general definitions of multivariate and bivariate copulas and their properties. We will discuss for the most important families of bivariate copulas as are the Elliptical and Archimedean Copulas and some applications of copulas.

Additionally, we will see some multidimensional regimes where there are many situations that one can go through and all of them are absorbent.

Finally, describing the two most important causes of death, cancer and ischemic heart disease. We will end up through the copulas models in the common survival function and in the correlation of the survival function with cancer with all other causes. We simulate data to apply them above to a specific copula.

# ΠΕΡΙΕΧΟΜΕΝΑ

## **1. ΣΥΝΑΡΤΗΣΕΙΣ ΘΝΗΣΙΜΟΤΗΤΑΣ-ΕΙΣΑΓΩΓΗ**

- 1.1 Διάρκεια ζωής και Απομένουσα Ζωή (Lifetime)
- 1.2 Ακέραια Χρόνια Απομένουσας Ζωής (Curtate Future Lifetime)
- 1.3 Συνάρτηση Επιβίωσης Διάρκειας Ζωής (Survival Distribution Function)
- 1.4 Συνάρτηση Κατανομής Διάρκειας Ζωής (Lifetime Distribution Function)
- 1.5 Συνάρτηση Πυκνότητας της Διάρκειας Ζωής
- 1.6 Συνάρτηση Κινδύνου Διάρκειας Ζωής
- 1.7 Δεσμευμένη Συνάρτηση Επιβίωσης Διάρκειας Ζωής
- 1.8 Δεσμευμένη Συνάρτηση Κατανομής Διάρκειας Ζωής
- 1.9 Δεσμευμένη Συνάρτηση Πυκνότητας Διάρκειας Ζωής
- 1.10 Ακέραια Προσδοκώμενη Ζωή

### **Α. ΠΙΝΑΚΕΣ ΘΝΗΣΙΜΟΤΗΤΑΣ**

- A.1 Εισαγωγή
- A.2 Κατασκευή ενός πίνακα ζωής

## **2. ΜΟΝΤΕΛΑ ΠΟΛΛΑΠΛΩΝ ΑΠΑΥΞΗΜΑΤΩΝ**

- 2.1. Εισαγωγή
- 2.2. Ορισμοί
- 2.3. Κατασκευή πίνακα πολλαπλών απαυξήματων
- 2.4. Συνδεδόμενοι πίνακες μοναδιαίου απαυξήματος

## **3. ΚΕΝΤΡΙΚΟΣ ΡΥΘΜΟΣ ΘΝΗΣΙΜΟΤΗΤΑΣ**

## **4. COPULAS**

### **ΕΙΣΑΓΩΓΗ**

- 4.1 Ορισμοί
- 4.2 Θεώρημα του Sklar
- 4.3 Survival Copulas
- 4.4 Πολυδιάστατα Copulas
- 4.5 The Elliptical Copulas
  - 4.5.1 The Gaussian Copula
  - 4.5.2 T-Student Copula
- 4.6 Archimedean Copulas

## **5. ΕΙΣΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ – ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ**

- A.** Ισχαιμική Καρδιοπάθεια
- B.** Καρκίνος

### **2. Ορισμοί**

### **3. Κατασκευή Πινάκων**





## Κεφάλαιο 1

### 1.Συναρτήσεις Θνησιμότητας

#### Εισαγωγή

#### 1.1 Διάρκεια ζωής και Απομένουσα Ζωή(Lifetime)

Έστω η μεταβλητή  $x$  να δηλώνει την ηλικία ενός ατόμου. Θεωρητικά, η μεταβλητή  $x$  μπορεί να πάρει τιμές από το μηδέν μέχρι το άπειρο. Στην πράξη όμως θεωρούμε ότι η μεταβλητή μπορεί να πάρει οποιαδήποτε τιμή από μηδέν μέχρι το ανώτατο όριο ζωής, που το συμβολίζουμε με  $\omega$ , και ονομάζεται έσχατη ή **οριακή ηλικία** (the limiting age) και παριστάνει την ηλικία στην οποία θεωρούμε ότι ο θάνατος είναι σίγουρο γεγονός. Συνήθως στην αναλογιστική πρακτική θεωρούμε ότι  $\omega \cong 110$  ή  $\omega \cong 120$ . Επίσης ως  $(x)$  θα συμβολίσουμε ένα άτομο που έχει ηλικία  $x$ .

Έστω  $T$  η συνεχής τυχαία μεταβλητή να δηλώνει τη **διάρκεια ζωής ενός ατόμου (lifetime)** ή την μελλοντική ζωή ενός νεογέννητου ή ισοδύναμα να εκφράζει την ανθρώπινη ηλικία στον θάνατο (age at death). Έτσι η  $T$  έχει σύνολο τιμών  $\mathfrak{R}_T = [0, \infty]$  ή στην πράξη η  $T$  έχει

σύνολο τιμών  $\mathfrak{R}_T = [0, \omega]$ . Μπορούμε τώρα ως  $T_x$  να ορίσουμε την συνεχή (μη αρνητική) τυχαία μεταβλητή να δηλώνει την **απομένουσα(residual of life) ή μελλοντική ζωή (future lifetime)** του  $(x)$  ή ισοδύναμα να δηλώνει το χρόνο που απομένει μέχρι το θάνατό του  $(x)$ .

Για  $x=0$  έχουμε την τυχαία μεταβλητή  $T$ . Έτσι μπορούμε να δηλώσουμε ως

$T_x = T - x | T > x$  με σύνολο τιμών  $\mathfrak{R}_{T_x} = [0, \infty]$  ή στην πράξη η  $T_x$  έχει σύνολο τιμών

$\mathfrak{R}_{T_x} = [0, \omega - x]$ .

#### 1.2 Ακέραια χρόνια απομένουσας ζωής (Curtate future lifetime)

Μπορούμε ως  $K_x$  να ορίσουμε τον ακέραιο αριθμό ετών που θα ζήσει ο  $(x)$ . Τα **ακέραια χρόνια απομένουσας ζωής** (curtate future lifetime) είναι μια διακριτή (μη αρνητική) τυχαία μεταβλητή και διαφέρει από την  $T_x$  κατά το ότι η  $K_x$  αγνοεί οποιοδήποτε κλάσμα έτους ζήσει πριν από το θάνατο ο  $(x)$ . Έτσι έχουμε ότι  $K_x = [T_x]$  και  $T_x = K_x + S_x$  όπου η διακριτή τυχαία μεταβλητή  $K_x$  παίρνει ακέραιες τιμές από το σύνολο  $\{0, 1, 2, \dots, \omega - x - 1\}$  και η συνεχής τυχαία μεταβλητή  $S_x$  παίρνει τιμές στο διάστημα  $(0, 1]$ , δηλαδή  $0 \leq S_x \leq 1$  και η  $S_x$  αντιπροσωπεύει το κλάσμα έτους που ζει ο  $(x)$  κατά τη χρονιά του θανάτου του.

#### 1.3 Συνάρτηση επιβίωσης διάρκειας ζωής (Survival Distribution Function)

Ορίζουμε την συνάρτηση επιβίωσης της διάρκειας ζωής (Survival Distribution Function)  $T$  ως  $S_T(x) = P(T > x)$  για κάθε  $0 \leq x \leq \omega$ , όπου  $S_T(x)$  να δηλώνει την πιθανότητα ένα

νεογέννητο να επιβιώσει πέρα από την ηλικία  $x$ , είτε ισοδύναμα να δηλώνει την αναμενόμενη αναλογία των μελών ενός πληθυσμού νεογέννητων που επιβιώνουν πέραν της ηλικίας  $x$ . Παρατηρούμε ότι η  $S_T(x)$  είναι φθίνουσα συνάρτηση καθώς το  $x$  αυξάνει, καθώς και ότι η συνάρτηση  $S_T(x)$  είναι συνεχής συνάρτηση της  $x$ . Επίσης γνωρίζουμε, εκ των προτέρων, δυο τιμές της συνάρτησης αυτής. Ισχύει ότι  $S_T(0) = 1$  και  $S_T(\omega) = 0$ . Το διάγραμμα της συνάρτησης επιβίωσης ως προς το χρόνο είναι γνωστό ως καμπύλη επιβίωσης (survival curve).

#### **1.4 Συνάρτηση κατανομής Διάρκειας Ζωής (Lifetime Distribution Function)**

Ορίζουμε την συνάρτηση κατανομής της διάρκειας ζωής (Lifetime distribution function or failure distribution) (της τυχαίας μεταβλητής  $T$ ) ως  $F_T(x) = P(T \leq x)$  για κάθε  $0 \leq x \leq \omega$  να δηλώνει την πιθανότητα ένα νεογέννητο να αποβιώσει έως την ηλικία  $x$ . Ισχύει ότι  $F_T(x) = P(T \leq x) = 1 - S_T(x)$  καθώς και ότι  $S_T(x) = 1 - F_T(x)$ .

#### **1.5 Συνάρτηση πυκνότητας της διάρκειας ζωής (Lifetime Probability Density Function)**

Η συνάρτηση πυκνότητας της διάρκειας ζωής (lifetime probability density function) (της τυχαίας μεταβλητής  $T$ ) ορίζεται ως

$$f_T(x) = \frac{d}{dx} F_T(x) = \lim_{dx \rightarrow 0} \frac{F_T(x+dx) - F_T(x)}{dx} = \lim_{dx \rightarrow 0} \frac{P(x < T \leq x+dx)}{dx} \quad \forall 0 \leq x \leq \omega \quad \text{με}$$

$$f_T(x) \cdot dx \cong F_T(x+dx) - F_T(x) = P(x < T \leq x+dx)$$

να δηλώνει την πιθανότητα ότι ο θάνατος ενός νεογέννητου θα συμβεί στο απειροελάχιστο χρονικό-ηλικιακό διάστημα  $(x, x+dx)$ . Από τη θεωρία των πιθανοτήτων είναι γνωστό ότι το γινόμενο  $f_T(x) \cdot dx = dF_T(x)$  δηλώνει την γνωστή απειροστή πιθανότητα ή την στοιχειώδη πιθανότητα που στην περίπτωση της ανάλυσης επιβίωσης δίνει τη μη δεσμευμένη πιθανότητα ή "αποτυχία", δηλαδή εδώ ο θάνατος, να συμβεί στο απειροστό διάστημα  $(x, x+dx)$  και

$f(x) = \frac{dF(x)}{dx}$  μπορεί να θεωρηθεί ότι αποτελεί το στιγμιαίο ρυθμό θανάτων, γι'αυτό και το

διάγραμμα της συνάρτησης πυκνότητας διάρκειας ζωής  $f_T(x)$  είναι γνωστό και ως καμπύλη θανάτων.

Ανάλογες ιδιότητες από τη θεωρία πιθανοτήτων ισχύουν

$$F_T(x) = \int_0^x f_T(u) du \quad \text{και} \quad S_T(x) = \int_x^{\omega} f_T(u) du$$

Επίσης, η πιθανότητα ότι ένα νεογέννητο θα αποβιώσει μεταξύ ηλικιών  $x_1$  και  $x_2$  είναι

$$P(x_1 \leq T \leq x_2) = \int_{x_1}^{x_2} f_T(u) du = F_T(x_2) - F_T(x_1) = S_T(x_1) - S_T(x_2) \quad \forall x_1 < x_2$$

Και η δεσμευμένη πιθανότητα είναι ότι ένα νεογέννητο θα αποβιώσει μεταξύ των ηλικιών  $x_1$  και  $x_2$ , δεδομένου επιβίωσης στην ηλικία  $x_1$ , είναι

$$P(x_1 \leq T \leq x_2 | T > x_1) = \frac{F_T(x_2) - F_T(x_1)}{1 - F_T(x_1)} = \frac{S_T(x_1) - S_T(x_2)}{S_T(x_1)} \quad \forall x_1 < x_2.$$

### 1.6 Συνάρτηση Κινδύνου Διάρκειας Ζωής (Hazard Function)

Έστω  $dx$  ένα πολύ μικρό διάστημα ζωής – τιμών της συνεχούς τυχαίας μεταβλητής  $T$ . Τότε η πιθανότητα να πάρει τιμές στο διάστημα  $(x, x + dx)$ , ή ισοδύναμα, η πιθανότητα το ενδεχόμενο του θανάτου να εμφανιστεί στο διάστημα  $(x, x + dx)$ , είναι  $P(x < T \leq x + dx)$  και μπορεί να υπολογιστεί όταν είναι γνωστή η συνάρτηση πυκνότητας διάρκειας ζωής ή η συνάρτηση κατανομής ή η συνάρτηση επιβίωσης :

$$P(x < T \leq x + dx) = \int_x^{x+dx} f_T(u) du = F_T(x + dx) - F_T(x) = S_T(x) - S_T(x + dx)$$

Αν υποθέσουμε ότι το άτομο έχει επιβιώσει (βρίσκεται εν ζωή) σε ηλικία ακριβώς  $x$  τότε η πιθανότητα το ενδεχόμενο του θανάτου να εμφανιστεί στο διάστημα  $(x, x + dx)$  αντιστοιχεί στη δεσμευμένη πιθανότητα

$$P(x < T \leq x + dx | T > x) = \frac{P(x < T \leq x + dx)}{P(T > x)} = \frac{F_T(x + dx) - F_T(x)}{1 - F_T(x)} \cong \frac{f_T(x) \cdot dx}{1 - F_T(x)}$$

Διαιρώντας την παραπάνω δεσμευμένη πιθανότητα με το μήκος του διαστήματος  $(x, x + dx)$

δηλαδή με το  $dx$  παίρνουμε τον **ρυθμό**  $\frac{P(x < T \leq x + dx | T > x)}{dx}$ . Αν στη συνέχεια

παίρνουμε το όριο του παραπάνω ρυθμού, καθώς το μήκος του διαστήματος  $dx$  τείνει στο μηδέν, τότε ορίζουμε την **συνάρτηση κινδύνου** (hazard function) ή **συνάρτηση διακινδύνευσης**. Έτσι ορίζουμε ως συνάρτηση κινδύνου της διάρκειας ζωής  $T$ , και συμβολίζεται με  $h_T(x)$  το όριο :

$$h_T(x) = \lim_{dx \rightarrow 0} \frac{P(x < T \leq x + dx | T > x)}{dx} \quad \forall x \geq 0$$

Στην αναλογιστική επιστήμη, την συνάρτηση κινδύνου την ονομάζουμε **ένταση ή ισχύς θνησιμότητας** (*force of mortality*), για άτομα ηλικίας ακριβώς  $x$ , και την συμβολίζουμε ως

$h_T(x) \equiv \mu_T(x) \equiv \mu_x$ . Από τον παραπάνω ορισμό προκύπτει ότι  $\mu_x = \lim_{dx \rightarrow 0} \frac{P(x < T \leq x + dx)}{P(T > x) \cdot dx}$

. Επειδή όμως  $f_T(x) = \frac{dF_T(x)}{dx} = \lim_{dx \rightarrow 0} \frac{F_T(x+dx) - F_T(x)}{dx} = \lim_{dx \rightarrow 0} \frac{P(x < T \leq x + dx)}{dx}$

καταλήγουμε ότι  $\mu_x = \frac{f_T(x)}{P(T > x)} = \frac{f_T(x)}{S_T(x)}$ . Δηλαδή  $\mu_x = -\frac{d}{dx} \ln S_T(x)$  ή

$\mu_x = -\frac{d}{dx} \ln(S_T(x))$  και ολοκληρώνοντας έχουμε ότι

$$S_T(x) = e^{-\int_0^x \mu_u \cdot dx}$$

Παρατηρούμε ότι η ένταση θνησιμότητας  $\mu_x = \frac{f_T(x)}{1 - F_T(x)} = \frac{f_T(x)}{S_T(x)}$  έχει ερμηνεία

**δεσμευμένης μάζας πιθανότητας**. Για κάθε ηλικία  $x$ , η ένταση θνησιμότητας δίνει την τιμή της δεσμευμένης μάζας πιθανότητας της διάρκειας ζωής στην ακριβή ηλικία  $x$ , δεδομένου επιβίωσης ως την ηλικία αυτή και έτσι το γινόμενο  $\mu_x \cdot dx$  δηλώνει την πιθανότητα ότι άτομο ηλικίας ακριβώς  $x$  (δηλαδή άτομο που έχει ζήσει ως την ηλικία  $x$ ) θα αποβιώσει αμέσως μετά στο διάστημα  $dx$ , δηλαδή μεταξύ των ηλικιών  $(x, x+dx)$ . Η συνάρτηση κινδύνου είναι ρυθμός και όχι πιθανότητα για αυτό μπορεί να πάρει τιμές μεγαλύτερες της μονάδας.

Η **συσωρευτική συνάρτηση κινδύνου** (Cumulative hazard function) ορίζεται ως

$$\Lambda_T(x) = \int_0^x \mu_u \cdot du = -\ln S_T(x) \quad \text{ή} \quad S_T(x) = e^{-\Lambda_T(x)}.$$

### **1.7 Δεσμευμένη Συνάρτηση Επιβίωσης Διάρκειας Ζωής (Conditional Lifetime Survival Function)**

Ορίζουμε ως δεσμευμένη συνάρτηση επιβίωσης διάρκειας ζωής την πιθανότητα επιβίωσης έως ένα χρονικό σημείο, έστω  $x + t$ , δοθέντος ότι υπάρχει επιβίωση έως ένα χρονικό σημείο που προηγείται, έστω  $x$ , (π.χ. δοθέντος ότι ένα άτομο έχει επιβιώσει στην ηλικία των 40 ποια είναι η πιθανότητα να επιβιώσει ως την ηλικία των 70 ετών). Οι τιμές της συνάρτησης αυτής είναι προφανώς πιθανότητες (ή ισοδύναμα μπορεί να ερμηνευτεί ως η αναμενόμενη αναλογία των μελών ενός πληθυσμού ατόμων οι οποίοι θα επιβιώσουν έως την ηλικία, έστω  $x+t$ , δοθέντος ότι επιβιώνουν στην ηλικία  $x$ ). Στην αναλογιστική επιστήμη την δεσμευμένη πιθανότητα αυτή ή την δεσμευμένη συνάρτηση επιβίωσης τη συμβολίζουμε ως  ${}_t p_x$  που ερμηνεύεται ως η πιθανότητα ότι άτομο ηλικίας  $x$  θα επιβιώσει  $t$  χρόνια.

Οι πιθανότητες  ${}_t p_x$  έχουν να κάνουν με κατανομές ενός υποσυνόλου του δειγματικού χώρου της τυχαίας μεταβλητής  $T$ , έτσι ώστε οι τιμές της  $T$  να είναι μεγαλύτερες του  $x$ . Τέτοιου είδους κατανομές ονομάζονται **περικομμένες κατανομές** της  $T$  κάτωθεν της  $x$  (truncated below  $x$ ). Ισοδύναμα, ως η δεσμευμένη συνάρτηση επιβίωσης διάρκειας ζωής ( της τυχαίας μεταβλητής  $T$ ) ή ως περικομμένη συνάρτηση επιβίωσης της  $T$  κάτωθεν της  $x$

μπορούμε να ορίσουμε την **συνάρτηση επιβίωσης απομένουσας ζωής** ( της τυχαίας μεταβλητής  $T_x$  ) ως

$${}_t p_x = S_{T_x}(t) = P(T_x > t) \quad \forall 0 \leq t \leq \omega - x$$

δηλαδή την πιθανότητα άτομο ηλικίας  $x$  να επιβιώσει τουλάχιστον  $t$  χρόνια. Φυσικά  $S_{T_x}(0) = 1$  και  $S_{T_x}(\omega - x) = 0$ .

Σύμφωνα με τον ορισμό της απομένουσας ζωής ( $T_x = T - x | T > x$ )

$$S_{T_x}(t) = P(T_x > t) = P(T > x+t | T > x) = \frac{P(T > x+t)}{P(T > x)} = \frac{S_T(x+t)}{S_T(x)} \quad \forall t \geq 0$$

$$\text{ή } {}_t p_x = S_{T_x}(t) = \frac{S_T(x+t)}{S_T(x)}$$

Έτσι, για έναν πληθυσμό  $l_0$  που υπακούει σε συνάρτηση επιβίωσης  $S_{T_x}(t)$ , η γνώση της  $S_T(t)$  είναι αρκετή για την εύρεση των συναρτήσεων επιβίωσης  $S_{T_x}(t)$  που αντιστοιχούν σε

κάθε ηλικία  $x$ ,  $\forall x \geq 0$ . Εύκολα διαπιστώνουμε, από τη σχέση  ${}_t p_x = \frac{S_T(x+t)}{S_T(x)}$  ότι

${}_t p_x = \frac{l_{x+t}}{l_x}$ . Προφανώς  ${}_1 p_x = \frac{l_{x+1}}{l_x}$ , να δηλώνει την πιθανότητα άτομο ηλικίας  $x$  θα επιβιώσει

ως την ηλικία  $x+1$ . Επίσης από την σχέση  $S_T(x) = e^{-\int_0^x \mu_u du}$  έχουμε

$${}_n p_x = S_{T_x}(n) = \frac{S_T(x+n)}{S_T(x)} = \frac{e^{-\int_0^{x+n} \mu_u du}}{e^{-\int_0^x \mu_u du}} = e^{-\int_x^{x+n} \mu_u du}$$

Ακόμη, από την ιδιότητα ότι  $\lim_{n \rightarrow \infty} {}_n p_x = 0$  έχουμε ότι

$$\lim_{n \rightarrow \infty} (-\log {}_n p_x) = \infty \Rightarrow \lim_{n \rightarrow \infty} \int_x^{x+n} \mu_s \cdot ds = \infty$$

### **1.8 Δεσμευμένη Συνάρτηση Κατανομής Διάρκειας Ζωής (Conditional Lifetime Distribution Function)**

Μπορούμε να ορίσουμε τη δεσμευμένη συνάρτηση κατανομής διάρκειας ζωής, συμβολίζεται

ως  ${}_t q_x$ , δηλαδή την πιθανότητα ότι άτομο ηλικίας  $x$  θα πεθάνει μέσα σε  $t$  χρόνια (θα πεθάνει πριν φτάσει στην ηλικία  $x+t$ ) ή η περικομμένη συνάρτηση κατανομής της  $T$  κάτωθεν του  $x$ .

Ισοδύναμα, ως δεσμευμένη συνάρτηση κατανομής διάρκειας ζωής (της τυχαίας μεταβλητής  $T$ ) ή ως περικομμένη συνάρτηση κατανομής της  $T$  κάτωθεν του  $x$  μπορούμε να ορίσουμε την **συνάρτηση κατανομής απομένουσας ζωής** (της τυχαίας μεταβλητής  $T_x$ ) ως

$${}_t q_x = F_{T_x}(t) = P(T_x < t) = P(T < x+t | T > x) = \frac{P(x < T \leq x+t)}{P(T > x)} =$$

$$\frac{F_T(x+t) - F_T(x)}{1 - F_T(x)}$$

να δηλώνει την πιθανότητα ότι άτομο ηλικίας  $x$  θα αποβιώσει μέσα σε  $t$  έτη ( $\forall 0 \leq t \leq \omega - x$ ). Φυσικά  $F_{T_x}(0) = 0$  και  $F_{T_x}(\omega - x) = 1$ . Εύκολα διαπιστώνουμε ότι

$${}_t q_x = \frac{S_T(x) - S_T(x+t)}{S_T(x)} = \frac{l_x - l_{x+t}}{l_x} = {}_t d_x$$

Προφανώς  $q_x = \frac{l_x - l_{x+1}}{l_x} = \frac{d_x}{l_x}$  δηλώνει την πιθανότητα άτομο ηλικίας  $x$  θα αποβιώσει

μέσα στο επόμενο έτος ζωής. Επίσης ισχύει  ${}_t q_x + {}_t p_x = 1 \quad \forall t \geq 0$ .

Μια άλλη χρήσιμη δεσμευμένη συνάρτηση πιθανότητας ορίζεται αν συμβολίσουμε με  ${}_{m|n} q_x$  την πιθανότητα άτομο ηλικίας  $x$  να ζήσει  $m$  χρόνια αλλά να αποβιώσει στα επόμενα  $n$  χρόνια, ή ισοδύναμα άτομο ηλικίας  $x$  να πεθάνει μεταξύ των ηλικιών  $x+m$  και  $x+m+n$ . Έτσι, έχουμε ότι  ${}_{m|n} q_x = P(m < T_x \leq m+n) = F_{T_x}(m+n) - F_{T_x}(m) = {}_{m+n} q_x - {}_m q_x$  ή ισοδύναμα

$${}_{m|n} q_x = S_{T_x}(m) - S_{T_x}(m+n) = {}_m p_x - {}_{m+n} p_x \Rightarrow$$

$${}_{m|n} q_x = \frac{l_{x+m}}{l_x} - \frac{l_{x+m+n}}{l_x} = \frac{l_{x+m} - l_{x+m+n}}{l_x} = \frac{l_{x+m}}{l_x} \cdot \frac{l_{x+m} - l_{x+m+n}}{l_{x+m}} = {}_m p_x \cdot {}_n q_{x+m}$$

Η

$${}_{m|n} q_x = {}_m p_x \cdot {}_n q_{x+m} = {}_{m+n} q_x - {}_m q_x = {}_m p_x - {}_{m+n} p_x$$

Στην ειδική περίπτωση όπου  $n=1$ , συμβολίζουμε ως  ${}_m q_x$  να δηλώνει την πιθανότητα άτομο ηλικίας  $x$  θα ζήσει  $m$  χρόνια αλλά θα αποβιώσει μεταξύ των ηλικιών  $x+m$  και  $x+m+1$ .

### 1.9. Δεσμευμένη Συνάρτηση Πυκνότητας Διάρκειας Ζωής (Conditional Lifetime Density

## Function)

Η αντίστοιχη **συνάρτηση πυκνότητας της απομένουσας ζωής** ( της τυχαίας μεταβλητής  $T_x$  ) ή η περικομμένη κάτωθεν του  $x$  συνάρτηση πυκνότητας της τυχαίας μεταβλητής  $T$  ορίζεται ως  $f_{T_x}(t) = \frac{d}{dt} F_{T_x}(t)$  και από τη σχέση  $F_{T_x}(t) = \frac{F_T(x+t) - F_T(x)}{1 - F_T(x)}$  έχουμε

$$f_{T_x}(t) = \frac{d}{dt} \frac{F_T(x+t) - F_T(x)}{S_T(x)} = \frac{f_T(x+t)}{S_T(x)} \quad \forall 0 \leq t \leq \omega - x$$

Λόγω της σχέσης  $S_T(x) = \frac{S_T(x+t)}{S_T(x)}$ , παίρνουμε

$$f_{T_x}(t) = \frac{f_T(x+t)}{S_T(x)} = \frac{f_T(x+t)}{S_T(x+t)} \cdot S_{T_x}(t) \quad \text{και από την σχέση } \mu_x = \frac{f_T(x)}{S_T(x)} \text{ καταλήγουμε}$$

$$f_{T_x}(t) = {}_t p_x \cdot \mu_{x+t} \quad \forall 0 \leq t \leq \omega - x$$

όπου η  $\mu_{x+t}$  είναι η ένταση θνησιμότητας στην ηλικία  $x+t$ .

### 1.10. Ακέραια Προσδοκώμενη Ζωή (Curtate expectation of life)

Ανάλογα μπορούμε να ορίσουμε και την ακέραια προσδοκώμενη ζωή, ατόμου ηλικίας  $x$ , που είναι η μαθηματική ελπίδα της ακεραίας απομένουσας ζωής  $K_x$ , ατόμου ηλικίας  $x$  δηλαδή τα αναμενόμενα μελλοντικά γενέθλια, ατόμου ηλικίας  $x$ . Συμβολίζεται με  $e_x$  και έχουμε

$$e_x = E(K_x) = \sum_{k=1}^{\omega-x-1} k \cdot {}_k p_x \cdot q_{x+k} = \sum_{k=1}^{\omega-x-1} k \cdot \frac{d_{x+k}}{l_x} = \sum_{k=1}^{\omega-x-1} k \cdot \frac{l_{x+k} - l_{x+k+1}}{l_x} = \frac{1}{l_x} \sum_{k=1}^{\omega-x-1} l_{x+k} \Rightarrow$$
$$e_x = \sum_{k=1}^{\omega-x-1} {}_k p_x$$

Η σχέση αυτή μπορεί να ερμηνευτεί πιο άμεσα ως εξής : η πιθανότητα ότι η χρονιά από την ηλικία  $x+k$  έως την ηλικία  $x+k+1$  προστίθεται στο χρόνο ζωής του ατόμου ηλικίας  $x$  είναι  ${}_k p_x$  (δηλαδή η πιθανότητα ότι άτομο ηλικίας  $x$  επιβιώνει στην ηλικία  $x+k+1$ ). Τότε ο συνολικός ακέραιος αναμενόμενος μελλοντικός χρόνος ζωής, ατόμου ηλικίας  $x$  θα είναι

$$e_x = \sum_{k=0}^{\omega-x-2} {}_{k+1}p_x = \sum_{k=1}^{\omega-x-1} {}_k p_x$$

Μια αναδρομική χρήσιμη σχέση της ακεραίας προσδοκώμενης ζωής είναι η

$$e_x = E(K_x) = P(T_x \geq 1) \cdot E(K_x | T_x \geq 1) + P(T_x < 1) \cdot E(K_x | T_x < 1) = p_x \cdot (1 + e_{x+1}) + q_x \cdot 0 \Rightarrow$$

$$e_x = p_x \cdot (1 + e_{x+1})$$

## 2. Πίνακες Θνησιμότητας (Mortality Tables)

### 2.1. Εισαγωγή

Ένας **πίνακας θνησιμότητας** είναι μια στατιστική - αναλογιστική μέθοδος για να εκφράσουμε τις ακέριες πιθανότητες επιβίωσης ή αποβίωσης καθώς επίσης για να εκτιμήσουμε τον υπολειπόμενο χρόνο ζωής, ενός συγκεκριμένου συνόλου ανθρώπων και για δεδομένο χρονικό διάστημα. Στην αναλογιστική πρακτική ο πίνακας θνησιμότητας ή ισοδύναμα ένας **πίνακας ζωής** (life table) κατασκευάζεται αφού εκτιμηθούν πρώτα οι δεσμευμένες πιθανότητες  $q_x$ , για ακέριες ηλικίες αρχίζοντας από μια συγκεκριμένη ελάχιστη ηλικία (συνήθως από την ηλικία 0), με συνέπεια ένας πίνακας ζωής να ορίζει πλήρως την κατανομή της  $K_x$ . Οι πιθανότητες αυτές διακρίνονται για κάθε ακέρια ηλικία  $x$ . Ο πίνακας θνησιμότητας αποτελεί μια στατιστική τεχνική για την παρουσίαση και τη σύνοψη των στοιχείων θνησιμότητας ενός πληθυσμού σε μια φόρμα που επιτρέπει απαντήσεις σε ερωτήματα όπως : ποια είναι η πιθανότητα ένας άντρας ηλικίας  $x$  να επιβιώσει (ή αποβιώσει) μέχρι την ηλικία  $y$ ; Ή ποιος είναι ο μέσος αριθμός ετών απομένουσας ζωής για άτομο που έχει φτάσει στην ηλικία  $x$ ;

Οι πίνακες θνησιμότητας χρησιμοποιούνται για διάφορους σκοπούς , όπως στην δημογραφία για την πρόβλεψη πληθυσμών , στις ασφαλίσσεις ζωής για τον υπολογισμό των μαθηματικών ασφαλίσεων καθώς και στον προσδιορισμό των εισφορών σε ένα συνταξιοδοτικό πρόγραμμα . Η μέθοδος ενός πίνακα θνησιμότητας είναι εφαρμόσιμη όχι μόνο στην ανάλυση της θνησιμότητας αλλά και πολλών μετρήσιμων διαδικασιών, όπως στην κλινική μελέτη των ανθρώπων ή στις εργαστηριακές μελέτες των ζώων. Η εφαρμογή της μεθόδου μπορεί να γενικευτεί , όπως για παράδειγμα στο να περιγράψεις το ιστορικό «ζωής» και «θανάτου» των αυτοκινήτων ή στην μελέτη της διάρκειας ζωής ενός λαμπτήρα. Συνεπώς ο πίνακας θνησιμότητας εξελίχθηκε σε ένα πολύ χρήσιμο εργαλείο για τους αναλογιστές, τους βιολόγους, τους φυσικούς, τους δημογράφους , τους κατασκευαστές , τους ερευνητές δημόσιας υγείας και τους ερευνητές σε πολλά ακόμη πεδία.

Ας υποθέσουμε ότι, σε μια δεδομένη χρονική στιγμή , έχουμε έναν **ιδεατό** πληθυσμό που τον παρατηρούμε από την στιγμή της γέννησης του, και έστω ότι έχουμε  $l_0$  νεογέννητα άτομα, καθώς και ότι οι  $l_0$  ζωές υπόκεινται στον ίδιο νόμο θνησιμότητας. Με την πάροδο του χρόνου (ζωής) κάποιοι θα επιβιώσουν και κάποιοι όχι. Έστω  $l_x$  να αντιπροσωπεύει τον



(αναμενόμενο) αριθμό επιζώντων της ομάδας που φτάνουν στην ακριβής ηλικία  $x$ , από τους  $l_0$  νεογέννητους. Έτσι έχουμε  $l_0$  νεογέννητα άτομα, και υποθέτουμε ότι οι  $l_0$  ζωές υπόκεινται στον ίδιο νόμο θνησιμότητας, έτσι ώστε  $P(T^j > x) = P(T > x)$ , όπου η τυχαία μεταβλητή  $T^j$  δηλώνει την διάρκεια ζωής του  $j$ -ατόμου, για  $j=1, 2, \dots, l_0$ . Με την πάροδο του χρόνου (ζωής) κάποιιοι θα επιβιώσουν και κάποιιοι όχι.

Έστω η τυχαία μεταβλητή  $S(x) | \{l_0\}$  να δηλώνει τον αριθμό των επιζώντων στην ηλικία  $x$ ,

από την γενιά των νεογέννητων  $l_0$ . Τότε παρατηρούμε ότι  $S(x) | \{l_0\} = \sum_{j=1}^{l_0} I_j$  όπου  $I_j$  η

δείκτρια τυχαία μεταβλητή για την επιβίωση της ζωής  $j$ , για  $j=1, 2, \dots, l_0$ , έτσι ώστε η  $I_j$  θα πάρει την τιμή 1 αν η  $j$ -ζωή θα επιβιώσει στην ηλικία  $x$  ή την τιμή 0 αν η  $j$ -ζωή δεν επιβιώσει

στην ηλικία  $x$ , δηλαδή  $I_j = \begin{cases} 1 & T^j > x \\ 0 & T^j \leq x \end{cases}$ . Επειδή όμως

$$E(I_j) = P(T^j > x) = P(T > x) = S_T(x) \text{ έχουμε}$$

$$l_x = E(S(x) | \{l_0\}) = \sum_{j=1}^{l_0} E(I_j) = l_0 \cdot S_T(x) = l_0 \cdot {}_xP_0$$

Όπου το  $l_x$  αντιπροσωπεύει τον αναμενόμενο αριθμό επιζώντων της ομάδας που φτάνουν στην ακριβής ηλικία  $x$ , από τους  $l_0$  νεογέννητους και έτσι έχουμε ότι  $l_x = l_0 \cdot S_T(x)$  ή

$$S_T(x) = \frac{l_x}{l_0}.$$

Είδαμε ότι η συνάρτηση επιβίωσης  $S_T(x) = \frac{l_x}{l_0}$ , εφόσον όπως έχουμε ορίσει η  $S_T(x)$

δηλώνει την πιθανότητα ένα νεογέννητο να επιβιώσει πέρα από την ηλικία  $x$ , ή ισοδύναμα να δηλώνει την αναμενόμενη αναλογία των μελών ενός πληθυσμού νεογέννητων που επιβιώνουν πέραν της ηλικίας  $x$ . Επειδή η συνάρτηση επιβίωσης  $S_T(x)$  είναι η πιθανότητα, τότε το  $l_x$  είναι ο αναμενόμενος αριθμός επιζώντων της ομάδας που φτάνουν στην ακριβής ηλικία  $x$  από τους  $l_0$  νεογέννητους. Η διαφορά μεταξύ των δυο αυτών συναρτήσεων είναι ότι, παρόλο που μαθηματικά είναι ταυτόσημοι και περιέχουν την ίδια πληροφορία, η συνάρτηση επιβίωσης  $S_T(x)$  είναι πιθανότητα και έτσι έχει πιθανοθεωρητική ερμηνεία, ενώ η συνάρτηση  $l_x$  έχει η πιθανοθεωρητική ερμηνεία (αναμενόμενος αριθμός επιζώντων ηλικίας  $x$  από τους  $l_0$  νεογέννητους) καθώς και ντετερμινιστική (*deterministic*) ερμηνεία (αριθμός επιζώντων ηλικίας  $x$  από μια γενιά με  $l_0$  νεογέννητους).

Σύμφωνα με τα παραπάνω, για την συνάρτηση επιβίωσης της απομένουσας ζωής, ισχύει ότι

$$S_T(x) = \frac{S_T(x+t)}{S_T(x)} = \frac{l_{x+t}}{l_x}$$

Μπορούμε εναλλακτικά να ορίσουμε την ένταση θνησιμότητας και ως εξής. Εφόσον

$$l_x = l_0 \cdot S_T(x) \Rightarrow \frac{dl_x}{dx} = l_0 \cdot \frac{dS_T(x)}{dx} \Rightarrow -\frac{1}{l_x} \cdot \frac{dl_x}{dx} = -\frac{1}{S_T(x)} \cdot \frac{dS_T(x)}{dx} = \mu_x$$

Καταλήγουμε στην σχέση

$$\mu_x = -\frac{1}{l_x} \cdot \frac{dl_x}{dx} = -\frac{d}{dx} \ln(l_x)$$

Έτσι, έχουμε ότι  $\mu_x$  είναι ο **στιγμιαίος σχετικός ρυθμός** (instantaneous relative rate)

**θανάτων, ανά ζωή, ανά έτος, για άτομα με ακριβή ηλικία  $x$** , διότι η ποσότητα

$$\frac{d}{dx} l_x = \lim_{dx \rightarrow 0} \frac{l_{x+dx} - l_x}{dx}$$

αντιπροσωπεύει τον *στιγμιαίο απόλυτο ρυθμό θανάτων* (*absolute instantaneous rate*), ανά έτος, για άτομα με ακριβή ηλικία  $x$ . Από τα παραπάνω προκύπτει

πάλι ότι η συνάρτηση κινδύνου είναι *ρυθμός* και όχι *πιθανότητα* για αυτό μπορεί και να πάρει τιμές μεγαλύτερες της μονάδας.

Επίσης, επειδή  $l_x \cdot \mu_x = (l_0 \cdot S_T(x)) \cdot \mu_x = l_0 \cdot f_T(x)$  η συνάρτηση  $l_x \cdot \mu_x$  μπορεί να ερμηνευτεί ως η αναμενόμενη πυκνότητα θανάτων στο διάστημα ηλικιών  $(x, x+dx)$ , καλείται **καμπύλη θανάτων** (*curve of deaths*) και συχνά η γραφική της παράσταση χρησιμοποιείται ως μέτρο σύγκρισης με άλλες κατανομές θανάτων. Τοπικά ακρότατα σημεία της συνάρτησης  $l_x \cdot \mu_x$  αντιστοιχούν σε σημεία καμπής της συνάρτησης  $l_x$  εφόσον

$$\frac{d}{dx} l_x \cdot \mu_x = \frac{d}{dx} \left( -\frac{d}{dx} l_x \right) = -\frac{d^2}{dx^2} l_x$$

Ομοίως, αν με την τυχαία μεταβλητή  $\Delta(x) | \{l_0\}$  δηλώνουμε τον αριθμό των θανάτων μεταξύ ηλικίας  $x$  και  $x+1$ , από τις  $l_0$  ζωές και με  $d_x = E(\Delta(x) | \{l_0\})$  τον αναμενόμενο αριθμό

θανάτων στο διάστημα μεταξύ ηλικιών  $x$  και  $x+1$ , τότε  $\Delta(x) | \{l_0\} = \sum_{j=1}^{l_0} C_j$ , όπου  $C_j$  η

δείκτρια τυχαία μεταβλητή η οποία παίρνει την τιμή 1 αν η  $j$ -ζωή αποβιώσει μεταξύ ηλικιών  $x$  και  $x+1$  ή την τιμή 0 αν η  $j$ -ζωή δεν αποβιώσει μεταξύ των ηλικιών  $x$  και  $x+1$ , δηλαδή

$$C_j = \begin{cases} 1 & x < T^j \leq x+1 \\ 0 & \text{διαφορετικά} \end{cases} . \text{ Έτσι, ο αναμενόμενος αριθμός θανάτων στο διάστημα μεταξύ}$$

ηλικιών  $x$  και  $x+1$  γίνεται

$$d_x = E(\Delta(x) | \{l_0\}) = \sum_{j=1}^{l_0} E(C_j) = l_0 \cdot P(x < T_j \leq x+1) = l_0 \cdot P(x < T \leq x+1) = l_0 \cdot [S_T(x) - S_T(x+1)]$$

$= l_0 \cdot {}_x q_0$ , εφόσον η ποσότητα

$$S_T(x) - S_T(x+1) = {}_x q_0 = \frac{l_x - l_{x+1}}{l_0}$$

Δηλώνει την αδέσμευτη πιθανότητα ότι ένα νεογέννητο θα αποβιώσει μεταξύ ηλικίας  $x$  και  $x+1$  (ή την αναλογία των νεογέννητων που θα αποβιώσουν μεταξύ ηλικίας  $x$  και  $x+1$ )

Στην *πρώτη στήλη* του πίνακα θνησιμότητας έχουμε τις  $l_x$  τιμές, όπου  $l_x$  είναι ο αριθμός των επιζώντων σε ακριβή ηλικία  $x$ . Ο πίνακας αρχίζει από ένα αυθαίρετο πλήθος  $l_0$  ζώντων σε ηλικία 0 και μηδενίζεται σε μία τερματική ηλικία  $\omega$  ( $l_\omega = 0$ ). Υπάρχουν όμως και περιπτώσεις όπου ένας πίνακας ζωής αρχίζει από κάποια ηλικία  $x$  διαφορετική του 0 (όπως για παράδειγμα πίνακες ζωής που αφορούν το συντάξιμο ή ασφαλισμένο πληθυσμό). Σε κάθε περίπτωση η αντίστοιχη τιμή  $l_a$  της νεότερης ηλικίας στον πίνακα ζωής είναι γνωστή ως βάση ή ρίζα (radix) του πίνακα. Συνήθως η τιμή που επιλέγουμε για την βάση είναι μια βολική στρογγυλεμένη τιμή όπως 1.000.000 ή 500.000 ή είναι ένας αρκετά μεγάλος αριθμός κι είναι μια δύναμη του 10, με εκθέτη ακέραιο και θετικό δηλαδή 100.000, 1.000.000 κ.ο.κ. Έτσι υποθέτουμε μια γενιά (cohort) από  $l_0$  νεογέννητα άτομα καθώς και ότι δεν παρατηρούμε καινούργιες γεννήσεις ούτε μετανάστευση ή αποχώρηση από αυτό τον ιδεατό πληθυσμό με την πάροδο του χρόνου. Έτσι, στην διάρκεια του χρόνου η γενιά αυτή σταδιακά ελαττώνεται στο μέγεθος, εφόσον κάποια μέλη της πεθαίνουν, μέχρι την ολική εξαφάνιση της γενιάς αυτής σε κάποια ηλικία  $\omega$ :  $l_\omega = 0$ .

Στη *δεύτερη στήλη* του πίνακα ζωής έχουμε τις  $d_x$  τιμές, που αντιπροσωπεύουν τους θανάτους κατά τη διάρκεια του  $x$ -έτους ηλικίας, δηλαδή από την αρχή της ηλικίας  $x$  και  $x+1$  ακριβώς. Προφανώς  $d_x = l_x - l_{x+1}$  όπου  $l_x$  οι επιζώντες ηλικίας  $x$  από τους αρχικούς  $l_0$ .

Θα πρέπει να επισημάνουμε ξανά ότι οι τιμές  $l_x$  και  $d_x$  που δίνονται σε ένα πίνακα ζωής δεν έχουν καμία απόλυτη έννοια, εφόσον εξαρτιούνται από την τιμή που εκλέγουμε για βάση του πίνακα. Όμως οι τιμές αυτές αποκτούν κάποια έννοια όταν σχετίζονται μεταξύ τους. Έτσι, εφόσον όλες οι τιμές των  $l_x$  και  $d_x$  έχουν έννοια μόνο όταν διαιρούνται μεταξύ τους, όλες οι τιμές σε έναν πίνακα θνησιμότητας μπορούν να πολλαπλασιαστούν με ένα παράγοντα και να μην απέλθει αλλαγή στην θνησιμότητα που παρίσταται από τον πίνακα.

Στην *τρίτη στήλη* του πίνακα θνησιμότητας έχουμε τις  $q_x$  τιμές, δηλαδή την αναλογία των

ατόμων ηλικίας  $x$  που θα πεθάνουν μεταξύ ηλικίας  $x$  και  $x+1$ . Προφανώς  $d_x = \frac{d_x}{l_x} = \frac{l_x - l_{x+1}}{l_x}$ .

Συνήθως δημιουργούμε και μια *τέταρτη στήλη* η οποία αποτελείται από τιμές  $e_x^0$ , δηλαδή από αναμενόμενους μελλοντικούς χρόνους μέχρι το θάνατο, ατόμων ηλικίας  $x$ . Για  $x=0$  παίρνουμε τον αναμενόμενο μέσο όρο ζωής του συγκεκριμένου πληθυσμού από την γέννηση.

Η παραπάνω περιγραφή ενός πίνακα ζωής είναι μια προσδιοριστική (deterministic)

προσέγγιση ως προς την θνησιμότητα της ιδεατής γενιάς. Ωστόσο οι παραπάνω έννοιες μπορούν να οριστούν και με πιθανοθεωρητικές συναρτήσεις, όπως έχουμε περιγράψει αναλυτικά στα προηγούμενα. Από αυτή τη σκοπιά για παράδειγμα το  $d_x$ , είναι ο αναμενόμενος αριθμός θανάτων στο διάστημα ηλικιών  $(x, x + 1)$  από ένα αρχικό πλήθος  $l_0$ . Φυσικά η προσέγγιση αυτή μας δίνει την ευκαιρία να κάνουμε χρήση της θεωρίας πιθανοτήτων παρόλο που και οι δυο προσεγγίσεις έχουν τις ίδιες μαθηματικές ιδιότητες για συναρτήσεις όπως  $l_x$  και  $q_x$ .

## 2.2 Κατασκευή ενός πίνακα ζωής

Η θνησιμότητα σύμφωνα με τον παραπάνω ιδεατό πληθυσμό  $l_0$  αντικαθίσταται στην πράξη από την εκτίμηση της θνησιμότητας ενός πληθυσμού σε ένα συγκεκριμένο χρονικό διάστημα μιας έρευνας που αφορά όλες τις ηλικίες, μέσα από διάφορες μεθόδους εκτίμησης των  $q_x$  τιμών (βάση του πληθυσμού σε κάθε διάστημα ηλικίας και των αντίστοιχων θανάτων, σε κάθε διάστημα ηλικίας, που συνέβησαν σε αυτό το χρονικό διάστημα της έρευνας).

Έτσι, η υπό εξέταση ομάδα ατόμων, όπως παρουσιάζεται σε ένα πίνακα ζωής έχει τα ακόλουθα χαρακτηριστικά :

- (I) Ο πληθυσμός αυτός αποτελείται αρχικά από  $l_0$  ζωές ηλικίας 0.
- (II) Τα μέλη του πληθυσμού υπόκεινται σε κάθε ηλικία της ζωής τους σε ένα πραγματικό ετήσιο ρυθμό θνησιμότητας που προσδιορίζεται από τις (γνωστές)  $q_x$  τιμές του πίνακα.
- (III) Ο πληθυσμός αυτός είναι κλειστός. Δεν επιτρέπονται καινούργιοι να μπουν στον πληθυσμό, η μόνη έξοδος από τον πληθυσμό είναι ως αποτέλεσμα του ετήσιου ρυθμού θνησιμότητας.

Με αυτά τα χαρακτηριστικά έχουμε ότι η πρόοδος του πληθυσμού καθώς και ένας πίνακας ζωής καθορίζεται από τα ακόλουθα. Δεδομένου ενός  $l_0$  βρίσκουμε  $l_0 \cdot q_0 = d_0$  τους θανάτους για το 1<sup>ο</sup> έτος ηλικίας, έτσι  $l_0 - l_1 = d_0 \Rightarrow l_1 = l_0 - d_0$ . Με την ίδια λογική έχουμε  $l_1 \cdot q_1 = l_1$  οπότε και  $l_2 = l_1 - d_1$ , και επαναλαμβάνοντας την ίδια αναδρομική διαδικασία μέχρι και την τελευταία γραμμή του πίνακα όπου  $l_{\omega-1} \cdot q_{\omega-1} = d_{\omega-1} = l_{\omega-1}$  διότι  $q_{\omega-1} = \frac{l_{\omega-1} - l_{\omega}}{l_{\omega-1}} = 1$ .

Για την εκτίμηση των  $e_x^0$  τιμών μια προσέγγιση, κάτω από την υπόθεση της ομοιόμορφης κατανομής των θανάτων σε κάθε έτος ζωής και τις σχέσεις  $e_x = p_x \cdot (1 + e_{x+1})$  και  $e_x^0 = e_x + \frac{1}{2}$ , είναι η αναδρομική σχέση  $e_x^0 = p_x \cdot (1 + e_{x+1}) + \frac{1}{2}$  με  $e_{\omega}^0 = 0$  για  $x = 0, 1, 2, \dots, \omega - 1$ .

## Κεφάλαιο 2

### Μοντέλα Πολλαπλών Απαυξημάτων (Multiple Decrement Models)

#### 2.1 Εισαγωγή

Οι πίνακες στη θεωρία πολλαπλών απαυξημάτων στην Αναλογιστική ορολογία ονομάζονται πίνακες με πολλαπλά αίτια εξόδου ή **πίνακες με πολλαπλά απαυξήματα**, και στην βιοστατιστική αναφέρονται ως η **θεωρία των ανταγωνιστικών κινδύνων** (theory of competing risks).

Οι συνήθεις δυο υποθέσεις που έχουμε σε μοντέλα πολλαπλών απαυξημάτων είναι:

**ΥΠΟΘΕΣΗ 1.** Κάθε έξοδος από έναν πίνακα πολλαπλών απαυξημάτων οφείλεται μόνο σε μια αιτία. Η υπόθεση αυτή μας εξασφαλίζει, όπως θα δούμε αμέσως μετά, ότι ο τερματισμός του καθεστώτος από διαφορετικές αιτίες είναι ξένα μεταξύ τους γεγονότα.

**ΥΠΟΘΕΣΗ 2.** Κάθε άτομο σε έναν δεδομένο πληθυσμό μπορεί να αποχωρήσει από οποιαδήποτε αιτία που συνυπάρχει στον πληθυσμό αυτό. Δηλαδή υποθέτουμε ότι έχουμε έναν ομοιογενή πληθυσμό σε σχέση με τα αίτια εξόδου που ενεργούν πάνω στον πληθυσμό αυτό.

#### 2.2 Ορισμοί

Όπως έχουμε δει, έχουμε ορίσει ως  $T_x$  τον χρόνο μέχρι τον θάνατο του ( $x$ ) και έχουμε δει μεθόδους για να καθορίσουμε την κατανομή της τυχαίας μεταβλητής  $T_x$ . Οι ίδιες μέθοδοι μπορούν να χρησιμοποιηθούν για να μελετήσουμε τον χρόνο μέχρι τον τερματισμό από ένα συγκεκριμένο καθεστώς δεδομένου ότι αυτό οφείλεται σε ένα από τα απαυξήματα που συνυπάρχουν ταυτόχρονα.

Ως εφαρμογή, στο καθεστώς « ζώντες ικανοί » η τυχαία μεταβλητή  $T_x$  θα δηλώνει τον χρόνο μέχρι τον τερματισμό του καθεστώτος δεδομένου ότι είναι θάνατος ή ολική ανικανότητα.

Αν  $\{X_j\}$  δηλώνει τον υπολειπόμενο χρόνο ζωής του συστήματος που οφείλεται στο  $j$ -αίτιο, για  $j=1,2,\dots,m$  (όταν έχουμε  $m$ -αιτία), τότε κάτω από την υπόθεση ότι κάθε τερματισμός του καθεστώτος οφείλεται σε μια μόνο αιτία ο χρόνος μέχρι τον τερματισμό του καθεστώτος είναι  $T_x = \min(X_1, X_2, \dots, X_m)$ . Σύμφωνα με τα παραπάνω, ο  $T_x$  είναι ο μόνος χρόνος που παρατηρούμε. Στην πραγματικότητα οι χρόνοι  $\{X_j\}$  δεν υπάρχουν – δεν πραγματοποιούνται (εκτός από το ελάχιστο) αλλά ο ορισμός τους εξυπηρετεί στην ανάπτυξη της θεωρίας.

Η **από κοινού (πολυμεταβλητή) συνάρτηση επιβίωσης** ορίζεται ως

$$S_{X_1 \dots X_m}(x_1, \dots, x_m) = P\left\{\bigcap_{j=1}^m (X_j > x_j)\right\}$$

Έτσι μπορούμε να ορίσουμε και τις **περιθώριες συναρτήσεις επιβίωσης** ως

$$S_{X_j}(x_j) = S_{X_1 \dots X_m}(0, \dots, x_j, \dots, 0) \text{ για } j=1, 2, \dots, m.$$

Κάτω από την **ΥΠΟΘΕΣΗ 2** συμπεραίνουμε ότι οι  $S_{X_j}(x_j)$  ορίζουν συναρτήσεις επιβίωσης όπως τις έχουμε σχολιάσει στις συναρτήσεις θνησιμότητας.

Σύμφωνα με τον παραπάνω ορισμό, μπορούμε τώρα να ορίσουμε ένα **στοιχείο της πολυμεταβλητής έντασης απαυξήματος** σε σχέση με το  $X_j$  στο σημείο  $\underline{x} = (x_1, \dots, x_m)$  ως

$$\mu_j(x_1, \dots, x_m) = \mu_j(\underline{x}) = \lim_{h \rightarrow 0} \frac{1}{h} \cdot P\{x_j < X_j \leq x_j + h \mid \bigcap_{j=1}^m (X_j > x_j)\} = -\frac{\partial \log S_{X_1 \dots X_m}(x_1, \dots, x_m)}{\partial x_j}$$

Οπότε το διάνυσμα  $\underline{\mu}(x_1, \dots, x_m) = [\mu_1(x_1, \dots, x_m), \dots, \mu_m(x_1, \dots, x_m)]$  δηλώνει την **πολυμεταβλητή ένταση απαυξήματος**.

Όπως σχολιάσαμε και πριν, κάτω από την **ΥΠΟΘΕΣΗ 1** δεν είναι δυνατόν να παρατηρήσουμε το τυχαίο διάνυσμα  $(X_1, \dots, X_m)$  από κοινού. Το μόνο που μπορούμε να παρατηρήσουμε είναι ο χρόνος μέχρι τον τερματισμό του καθεστώτος, δηλαδή την τυχαία μεταβλητή  $T_x = \min(X_1, \dots, X_m)$ . Έτσι, ορίζουμε ως **ολική συνάρτηση επιβίωσης από όλες τις m-αιτίες** την συνάρτηση

$$S_{T_x}(t) = {}_t(ap)_x^T := P(T_x > t) = P\left\{\bigcap_{j=1}^m (X_j > t)\right\} = S_{X_1 \dots X_m}(t, \dots, t)$$

Παρατηρούμε ότι η ολική συνάρτηση επιβίωσης από όλες τις  $m$ -αιτίες είναι μια συνάρτηση μιας μεταβλητής του χρόνου  $t$ . Μπορούμε έτσι να ορίσουμε και την αντίστοιχη **ολική ένταση απαυξήματος** ως μια δεσμευμένη μάζα πιθανότητας:

$$(a\mu)_{x+t}^T = -\frac{d \log S_{T_x}(t)}{dt} = -\frac{d}{dt} \log [{}_t(ap)_x^T] \Rightarrow$$

$${}_t(ap)_x^T = \exp\left\{-\int_0^t (a\mu)_{x+s}^T \cdot ds\right\}$$

Διαφορετικά,

$$(a\mu)_{x+t}^T = -\frac{d \log S_{T_x}(t)}{dt} = -\frac{d \log S_{X_1 \dots X_m}(t, \dots, t)}{dt} = \lim_{dt \rightarrow 0} \frac{1}{dt} \cdot P\{t < T_x \leq t + dt \mid \bigcap_{j=1}^m (X_j > t)\} \Rightarrow$$

$$(a\mu)_{x+t}^T \cdot dt \cong P\{t < T_x \leq t + dt \mid \bigcap_{j=1}^m (X_j > t)\}$$

Να δηλώνει την (δεσμευμένη) πιθανότητα ότι το καθεστώς που είναι σε ισχύ στην ηλικία  $x$  θα τερματίσει αμέσως μετά από χρόνο  $t$  (στην ηλικία  $x+t$ ), δεδομένου ότι είναι σε ισχύ στην ηλικία  $x+t$ .

Επεκτείνουμε το βασικό μοντέλο εισάγοντας μια δεύτερη τυχαία μεταβλητή, το αίτιο εξόδου. Θα το συμβολίζουμε ως  $J$  και θα θεωρούμε ότι η  $J$  είναι μια διακριτή τυχαία μεταβλητή τέτοια ώστε η  $J$  θα παίρνει την  $j$  αν ο τερματισμός του καθεστώτος οφείλεται στην  $j$ -αιτία ή την τιμή 0 διαφορετικά. Έτσι

$$J = \begin{cases} j & \text{αν } T_x = X_j \\ 0 & \text{διαφορετικά} \end{cases}$$

Ως εφαρμογή, η  $J$  στο καθεστώς «ζώντες ικανοί» μπορεί να πάρει την τιμή 1 ή 2 αντίστοιχα αν το αίτιο εξόδου ή ο τερματισμός του καθεστώτος είναι θάνατος ή ολική ανικανότητα.

Ορίζουμε την **από κοινού συνάρτηση μάζας πιθανότητας** της διανυσματικής τυχαίας μεταβλητής  $(T_x, J)$  ως  $f_{T_x, J}(t, j)$  με πεδίο ορισμού το σύνολο  $\mathfrak{R}_{T_x, J} \subseteq [0, \infty] \times \{1, 2, \dots, m\}$ , όπου  $m$  ο αριθμός των απευξήματων, την περιθώρια συνάρτηση μάζας πιθανότητας της  $J$  ως  $p_J(j) = f_J(j)$  και την περιθώρια συνάρτηση μάζας πιθανότητας της  $T_x$  ως  $f_{T_x}(t)$ . Έτσι έχουμε, ότι

$$\sum_{j=1}^m f_J(j) = 1, \int_0^{\infty} f_{T_x}(t) \cdot dt = 1 \text{ και } \sum_{j=1}^m \int_0^{\infty} f_{T_x, J}(s, j) \cdot ds = 1$$

Η από κοινού κατανομή  $f_{T_x, J}(t, j)$  μπορεί να αποδοθεί γραφικά σε  $m$  διακριτά παράλληλα επίπεδα, έτσι ώστε το κάθε ένα επίπεδο να προσδιορίζει τις πιθανότητες  $f_{T_x|J}(t | J = j) \cdot dt$  για  $j = 1, 2, \dots, m$ .

Επίσης, σε αναλογία με την σχέση  $f_{T_x}(t) \cdot dt \cong P(t \leq T_x < t + dt)$  που ισχύει για οποιαδήποτε συνεχή τυχαία μεταβλητή, έχουμε ότι

$$f_{T_x, J}(t, j) \cdot dt \cong P[t \leq T_x < t + dt, J = j]$$

Οπότε, η ποσότητα  $f_{T_x, J}(t, j) \cdot dt$  εκφράζει προσεγγιστικά την (αδέσμευτη) **πιθανότητα τερματισμού του καθεστώτος, λόγω του  $j$ -αιτίου, αμέσως μετά από χρόνο  $t$ , ή ισοδύναμα την (αδέσμευτη) πιθανότητα εξόδου του  $(x)$  από την αρχική του κατάσταση λόγω του  $j$ -αιτίου στην ηλικία  $(x + t, x + t + dt)$ . Επίσης, ισχύουν οι σχέσεις**

$$P[a < T_x \leq b] = \sum_{j=1}^m \int_a^b f_{T_x, J}(t, j) \cdot dt \text{ καθώς και}$$

$$P[0 < T_x \leq t, J = j] = \int_0^t f_{T_x, J}(s, j) \cdot ds =: {}_t(aq)_x^j$$

Για  $\forall t \geq 0, j = 1, 2, \dots, m$  και έτσι το  ${}_t(aq)_x^j$  συμβολίζει την **πιθανότητα τερματισμού του καθεστώτος πριν τον χρόνο  $t$  που οφείλεται στην αιτία  $j$**  ή ισοδύναμα την πιθανότητα εξόδου του  $(x)$  από την αρχική του κατάσταση λόγω του  $j$ -αιτίου μέχρι την ηλικία  $(x+t)$ .

Ανάλογα έχουμε ότι η πιθανότητα ότι το καθεστώς **θα τερματίσει πέρα από τον χρόνο  $t$  και θα οφείλεται στην  $j$ -αιτία**, ή ισοδύναμα ότι η πιθανότητα εξόδου του  $(x)$  από την αρχική του κατάσταση λόγω του  $j$ -αιτίου μετά την ηλικία  $(x+t)$  είναι

$$P[T_x > t, J = j] = \int_t^{\infty} f_{T_x, J}(s, j) \cdot ds =: {}_t(ap)_x^j$$

Η περιθώρια συνάρτηση πιθανότητας της **J** γίνεται

$$p_j(j) = P(J = j) = P\{(T_x < \infty) \cap (J = j)\} = \int_0^{\infty} f_{T_x, J}(s, j) \cdot ds =: {}_{\infty}(aq)_x^j$$

(για  $j = 1, 2, \dots, m$ ) και δηλώνει την **αναμενόμενη αναλογία τερματισμών που οφείλεται στην  $j$ -αιτία** ή ισοδύναμα την πιθανότητα εξόδου του  $(x)$  από την κατάσταση που βρισκόταν αρχικά λόγω του  $j$ -αιτίου εξόδου, από την ηλικία  $(x)$  και μετά, με  $\sum_{j=1}^m {}_{\infty}(aq)_x^j = 1$ .

Παρατηρούμε ότι  ${}_t(ap)_x^j + {}_t(aq)_x^j = P(J = j) = {}_{\infty}(aq)_x^j \neq 1$ , δηλαδή οι συναρτήσεις πιθανοτήτων  ${}_t(aq)_x^j$  και  ${}_t(ap)_x^j$  δεν είναι καλά ορισμένες κατανομές και έτσι *δεν μπορούν να θεωρηθούν ως συναρτήσεις κατανομής και επιβίωσης* αντίστοιχα μιας τυχαίας μεταβλητής. Ωστόσο μπορούμε να ορίσουμε κατανομές με όρους τις συναρτήσεις αυτές αν ορίσουμε ως

$$S_j^*(t) := P[T_x > t | J = j] = \frac{{}_t(ap)_x^j}{{}_{\infty}(aq)_x^j}$$

Οπότε η  $F_j^*(t) := P[T_x \leq t | J = j] = \frac{{}_t(aq)_x^j}{{}_{\infty}(aq)_x^j} = 1 - S_j^*(t)$  δηλώνει την αντίστοιχη συνάρτηση

κατανομής και αντιπροσωπεύει την πραγματική κατανομή τερματισμού του καθεστώτος όταν αυτό οφείλεται στο  $j$ -αίτιο με την ταυτόχρονη λειτουργία όλων των αιτιών. Η αντίστοιχη συνάρτηση μάζας πιθανότητας γίνεται

$$f_j^*(t) := -\frac{1}{{}_{\infty}(aq)_x^j} \cdot \frac{d}{dt} {}_t(ap)_x^j$$



Η περιθώρια συνάρτηση μάζας πιθανότητας της  $T_x$  είναι  $f_{T_x}(t) = \sum_{j=1}^m f_{T_x,j}(t, j)$  με την

αντίστοιχη περιθώρια συνάρτηση κατανομής της  $T_x$   $F_{T_x}(t) = \int_0^t f_{T_x}(s) \cdot ds$  .

Αν χρησιμοποιήσουμε ως εκθέτη το σύμβολο ( $T$ ) να δηλώνει ότι η αντίστοιχη συνάρτηση αναφέρεται σε όλες τις αιτίες συνολικά , παίρνουμε ότι η συνάρτηση κατανομής της  $T_x$  δίνει

$$F_{T_x}(t) = {}_t(aq)_x^T := P(T_x \leq t) = \int_0^t f_{T_x}(s) \cdot ds = \int_0^t \sum_{j=1}^m f_{T_x,j}(s, j) \cdot ds = \sum_{j=1}^m \int_0^t f_{T_x,j}(s, j) \cdot ds \Rightarrow$$

$${}_t(aq)_x^T = \sum_{j=1}^m {}_t(aq)_x^j$$

Δηλαδή η πιθανότητα απαυξήματος (ή η πιθανότητα τερματισμού του καθεστώτος) από οποιαδήποτε αιτία μεταξύ της ηλικίας  $x$  και  $x+t$ , δεδομένης επιβίωσης του καθεστώτος στην ηλικία  $x$ , ισούται με το άθροισμα όλων των  $m$  επιμέρους πιθανοτήτων απαυξημάτων μεταξύ ηλικίας  $x$  και  $x+t$ , δεδομένης επιβίωσης του κάθε καθεστώτος στην ηλικία  $x$ .

Επίσης από την σχέση  $f_{T_x,j}(t, j) \cdot dt \cong P[t < T_x \leq t + dt, J = j]$  έχουμε ότι για  $t \geq 0$  και  $j = 1, 2, \dots, m$ .

$$f_{T_x,j}(t, j) \cdot dt = P[T_x > t] \cdot P\{t < T_x \leq t + dt, J = j \mid T_x > t\} , \text{ όπου}$$

$$P\{t < T_x \leq t + dt, J = j \mid T_x > t\} = P\{t < T_x \leq t + dt \mid \bigcap_{j=1}^m X_j > t\}$$

Και  ${}_t(ap)_x^T := P(T_x > t) = 1 - {}_t(aq)_x^T$  , οπότε καταλήγουμε ότι

$$f_{T_x,j}(t, j) \cdot dt = {}_t(ap)_x^T \cdot (a\mu)_{x+t}^j \cdot dt$$

Όπου

$$(a\mu)_{x+t}^j = \lim_{dt \rightarrow 0} \frac{1}{dt} \cdot P\{t < T_x \leq t + dt \mid \bigcap_{j=1}^m (X_j > t)\}$$

δηλώνει την ένταση του  $j$ - απαυξήματος η οποία περιγράφει τον στιγμιαίο ρυθμό απαυξήματος από την  $j$ - αιτία , στην ηλικία  $x+t$  , όταν όλες οι αιτίες λειτουργούν ταυτόχρονα.

Δηλαδή, η (από κοινού) αδέσμευτη πιθανότητα το απάυξημα να λάβει χώρα αμέσως μετά από χρόνο  $t$  (μεταξύ  $t$  και  $t + dt$ ) εξαιτίας του  $j$ - αιτίου ( $f_{T_x,j}(t, j) \cdot dt$ ), ισούται με την πιθανότητα,  ${}_t(ap)_x^T$  ότι ο ( $x$ ) θα παραμείνει στο αρχικό καθεστώς για χρόνο  $t$  (ή ισοδύναμα έως την ηλικία  $x+t$ ) επί την δεσμευμένη πιθανότητα,  $(a\mu)_{x+t}^j \cdot dt$  , ότι το

απαύξημα θα συμβεί αμέσως μετά από χρόνο  $t$  (μεταξύ  $t$  και  $t+dt$ ) εξαιτίας του  $j$  – αιτίου δεδομένου ότι το απαύξημα (η λήξη του καθεστώτος) δεν έχει συμβεί μέχρι τον χρόνο  $t$ . Έτσι, από την παραπάνω σχέση καταλήγουμε ότι ισχύει

$$(a\mu)_{x+t}^j = \frac{f_{T_x, J}(t, j)}{{}_t(ap)_x^T}$$

ή ισοδύναμα

$$(a\mu)_{x+t}^j = \frac{1}{{}_t(ap)_x^T} \cdot \frac{d}{dt} {}_t(aq)_x^j$$

Εφόσον  ${}_t(aq)_x^j = \int_0^t f_{T_x, J}(s, j) \cdot ds$

Ακόμη από την σχέση  $(a\mu)_{x+t}^T = -\frac{d \log S_{T_x}(t)}{dt} = \frac{f_{T_x}(t)}{{}_t(ap)_x^T} = \frac{\sum_{j=1}^m f_{T_x, J}(t, j)}{{}_t(ap)_x^T}$  συμπεραίνουμε ότι

$$(a\mu)_{x+t}^T = \sum_{j=1}^m (a\mu)_{x+t}^j$$

Δηλαδή η συνολική ένταση του απαυξήματος είναι το άθροισμα των επιμέρους εντάσεων εξαιτίας των  $m$  – αιτιών, κάτω από την υπόθεση **ΥΠΟΘΕΣΗ 1** ότι κάθε τερματισμός του καθεστώτος οφείλεται μόνο σε μια αιτία.

Μπορούμε τώρα να συνοψίσουμε τους ορισμούς που δώσαμε για τις από κοινού, δεσμευμένες και περιθώριες συναρτήσεις με αναλογιστικούς όρους :

$$f_{T_x, J}(t, j) = {}_t(ap)_x^T \cdot (a\mu)_{x+t}^T, \quad p_J(j) = {}_\infty(aq)_x^j$$

Και επειδή  $\sum_{j=1}^m f_{T_x, J}(t, j) = {}_t(ap)_x^T \cdot \sum_{j=1}^m (a\mu)_{x+t}^j$  έχουμε ότι

$$f_{T_x}(t) = {}_t(ap)_x^T \cdot (a\mu)_{x+t}^T$$

Επίσης, μπορούμε να ορίσουμε την δεσμευμένη συνάρτηση πιθανότητας της  $J$ , δεδομένου απαυξήματος στον χρόνο  $t$  :

$$f_{J|T_x}(j | T_x = t) = \frac{f_{T_x, J}(t, j)}{f_{T_x}(t)} = \frac{(a\mu)_{x+t}^j}{{}_t(ap)_x^T}$$

Η πιθανότητα  ${}_t(aq)_x^j = \int_0^t f_{T_x, J}(s, j) \cdot ds$  μπορεί να γραφτεί

$${}_t(aq)_x^j = \int_0^t {}_s(ap)_x^T \cdot (a\mu)_{x+s}^j \cdot ds$$

καθώς και η ολική δεσμευμένη πιθανότητα  ${}_t(aq)_x^T = \sum_{j=1}^m {}_t(aq)_x^j = \sum_{j=1}^m \int_0^t {}_s(ap)_x^T \cdot (a\mu)_{x+s}^j \cdot ds$ ,

δηλαδή

$${}_t(aq)_x^T = \int_0^t {}_s(ap)_x^T \cdot (a\mu)_{x+s}^T \cdot ds$$

Ανάλογα μπορούμε να ορίσουμε και την διακριτή τυχαία μεταβλητή  $K_x$  να δηλώνει τα **ακέραια μελλοντικά χρόνια μέχρι τον τερματισμό του καθεστώτος** (ή μέχρι το απαύξημα). Έτσι, η **από κοινού συνάρτηση πιθανότητας** της διανυσματικής τυχαίας μεταβλητής  $(K_x, J)$ , έστω  $p_{K_x, J}(k, j)$  με πεδίο ορισμού το σύνολο  $\mathfrak{R}_{T_x, J} \subseteq \{0, 1, 2, \dots\} \times \{1, 2, \dots, m\}$  δίνεται ως

$$p_{K_x, J}(k, j) = P[K_x = k, J = j] = P[k \leq T_x < k+1, J = j] = \int_k^{k+1} f_{T_x, J}(t, j) \cdot dt =$$

$$\int_k^{k+1} {}_t(ap)_x^T \cdot (a\mu)_{x+t}^j \cdot dt \Rightarrow p_{K_x, J}(k, j) = {}_k(ap)_x^T \cdot \int_0^1 {}_t(ap)_{x+k}^T \cdot (a\mu)_{x+k+t}^j \cdot dt \Rightarrow$$

$$p_{K_x, J}(k, j) = {}_k(ap)_x^T \cdot (aq)_{x+k}^j$$

$$\text{Όπου } (aq)_{x+k}^j = \int_0^1 {}_t(ap)_{x+k}^T \cdot (a\mu)_{x+k+t}^j \cdot dt$$

Ανάλογα έχουμε ότι η **συνάρτηση πιθανότητας της  $J$**  είναι

$$f_J(j) = \sum_{k=0}^{\infty} p_{K_x, J}(k, j) = {}_{\infty}(aq)_x^j$$

Η **συνάρτηση πιθανότητας της  $K_x$**

$$p_{K_x}(k) = \sum_{j=1}^m p_{K_x, J}(k, j) = P(K_x = k) = {}_k(aq)_x^T = {}_k(ap)_x^T \cdot (aq)_{x+k}^T = {}_k(ap)_x^T \cdot \sum_{j=1}^m (aq)_{x+k}^j$$

Και η **περιθώρια συνάρτηση κατανομής της  $K_x$**

$$F_{K_x}(x) = \sum_{s=0}^k p_{K_x}(s) = P(K_x \leq k) = {}_{k+1}(aq)_x^T$$

### 2.3 Κατασκευή πίνακα πολλαπλών απαυξημάτων

Από τα προηγούμενα παρατηρούμε ότι σε ένα μοντέλο πολλαπλών απαυξημάτων οι μόνες παρατηρούμενες συναρτήσεις επιβίωσης είναι η  $S_{Tx}(t) = {}_t(ap)_x^T$ , η οποία δηλώνει την ολική συνάρτηση επιβίωσης από όλες τις αιτίες μαζί, και η  ${}_t(ap)_x^j$ , η οποία δηλώνει την συνάρτηση πιθανότητας τερματισμού του καθεστώτος μετά την ηλικία  $x+t$  από την  $j$ -αιτία, για κάθε  $j=1,2,\dots,m$ . Η εκτίμηση αυτών των δυο συναρτήσεων επιβίωσης από στατιστικά δεδομένα μας επιτρέπει την κατασκευή ενός πίνακα με πολλαπλά απαυξήματα.

Έτσι θεωρούμε ομάδα ατόμων, έστω  $(al)_a^T$ , ηλικίας  $a$  ο καθένας. Για κάθε ζωή υποθέτουμε ότι η από κοινού κατανομή του χρόνου ως το απαύζημα και η αιτία απαυξήματος προσδιορίζεται από την συνάρτηση μάζας πιθανότητας  $f_{Ta,j}(t,j) = {}_t(ap)_a^T \cdot (\alpha\mu)_{a+t}^j$  για  $t \geq 0$  και  $j=1,2,\dots,m$  όπως έχουν οριστεί. Θα δηλώνουμε με  ${}_n(ad)_x^j$  την τυχαία μεταβλητή που είναι ίση με τον αριθμό των ατόμων οι οποίες θα φύγουν από την αρχική ομάδα των  $(al)_a^T$  μεταξύ ηλικίας  $x$  και  $x+n$  (θα αφήσουν το καθεστώς), για  $x \geq a$ , από την  $j$ -αιτία. Αν η μέση τιμή της δηλώνεται ως  ${}_n(ad)_x^j$ , δηλαδή  ${}_n(ad)_x^j = E[{}_n(ad)_x^j]$  να δηλώνει τον αναμενόμενο αριθμό ατόμων από την αρχική ομάδα των  $(al)_a^T$  που θα αποχωρήσουν από το καθεστώς μέσα στο ηλικιακό διάστημα  $[x, x+n]$  λόγω της  $j$ -αιτίας, τότε έχουμε

$${}_n(ad)_x^j = (al)_a^T \cdot P\{x-a < T_a < x+n-a, J = j\} =$$

$$(al)_a^T \cdot \int_{x-a}^{x+n-a} f_{Ta}(t) \cdot dt = (al)_a^T \cdot \int_{x-a}^{x+n-a} {}_t(ap)_a^T \cdot (\alpha\mu)_{a+t}^j \cdot dt = (al)_a^T \cdot {}_{x-a}(ap)_a^T \cdot \int_0^n {}_t(ap)_x^T \cdot (\alpha\mu)_{x+t}^j \cdot dt$$

Αν τώρα ορίσουμε την τυχαία μεταβλητή  $(as)_x^T$  να δηλώνει τον αριθμό των ατόμων που παραμένουν στο αρχικό καθεστώς, ηλικίας  $x$ , από την αρχική ομάδα των  $(al)_a^T$  τότε έχουμε

$$(al)_x^T = E[(as)_x^T] = (al)_a^T \cdot P\{T_a > x-a\} = (al)_a^T \cdot {}_{x-a}(ap)_a^T$$

Όπου  $(al)_x^T$  να δηλώνει τον αναμενόμενο πλήθος των ατόμων που παραμένουν στο αρχικό καθεστώς στην ηλικία ακριβώς  $x$  από το αρχικό πλήθος των  $(al)_a^T$ . Έτσι, καταλήγουμε ότι

$${}_n(ad)_x^j = (al)_x^T \cdot {}_n(aq)_x^j \quad \text{ή} \quad {}_n(aq)_x^j = \frac{{}_n(ad)_x^j}{(al)_x^T}$$

Όπου  ${}_n(aq)_x^j$  δηλώνει την δεσμευμένη πιθανότητα αποχώρησης ατόμων ηλικίας  $(x)$  από την

αρχική κατάσταση λόγω του  $j$ -αιτίου μέχρι την ηλικία  $(x+t)$ , από την αρχική ομάδα των  $(al)_a^T$ .

Παρατηρούμε ότι η τυχαία μεταβλητή  ${}_n(ad)_x^T = \sum_{j=1}^m {}_n(ad)_x^j$ , δηλώνει τον συνολικό αριθμό των ατόμων οι οποίες θα φύγουν από το αρχικό καθεστώς, από την αρχική ομάδα των  $(al)_a^T$  ατόμων, μεταξύ ηλικίας  $x$  και  $x+n$ , από οποιαδήποτε αιτία. Ανάλογα έτσι ορίζουμε τον αναμενόμενο συνολικό αριθμό ατόμων, από την αρχική ομάδα των  $(al)_a^T$ , που θα αποχωρήσουν από το αρχικό καθεστώς μέσα στο ηλικιακό διάστημα  $[x, x+n]$ , από οποιαδήποτε αιτία ως

$${}_n(ad)_x^T = E[{}_n(ad)_x^T] = \sum_{j=1}^m {}_n(ad)_x^j$$

Τότε

$${}_n(ad)_x^T = (al)_a^T \cdot \sum_{j=1}^m \int_{x-a}^{x+n-a} {}_t(ap)_a^T \cdot (\alpha\mu)_{a+t}^j \cdot dt = (al)_a^T \cdot \int_{x-a}^{x+n-a} {}_t(ap)_a^T \cdot (\alpha\mu)_{a+t}^T \cdot dt$$

Και έτσι καταλήγουμε ότι

$${}_n(ad)_x^T = (al)_x^T \cdot {}_n(aq)_x^T \quad \text{ή} \quad {}_n(aq)_x^T = \frac{{}_n(ad)_x^T}{(al)_x^T}$$

Όπου  ${}_n(aq)_x^T$  δηλώνει την δεσμευμένη πιθανότητα απουξήματος από οποιαδήποτε αιτία στο διάστημα  $[x, x+n]$  δεδομένου μη λήξης του καθεστώτος στην ηλικία  $x$ .

Αν τώρα θεωρήσουμε την ολική ένταση απουξήματος ως ένα προσδιοριστικό (*deterministic*) ρυθμό απουξήματος, παρά ως μια δεσμευμένη μάζα πιθανότητας, τότε ο αριθμός των επιζώντων στην ηλικία  $x$  από την αρχική ομάδα  $(al)_a^T$  ηλικίας  $a$  δίνεται από την σχέση

$$(al)_x^T = (al)_a^T \cdot {}_{x-a}(ap)_a^T$$

Όπου  ${}_{x-a}(ap)_a^T \equiv S_{T_a}(x-a)$  η ολική συνάρτηση επιβίωσης και έτσι ισχύει ότι

$${}_{x-a}(ap)_a^T \equiv S_{T_a}(x-a) = \frac{(al)_x^T}{(al)_a^T} \Rightarrow (al)_x^T = (al)_a^T \cdot S_{T_a}(x-a) \Rightarrow \frac{d}{dx} (al)_x^T = (al)_a^T \cdot \frac{d}{dt} S_{T_a}(x-a) \Rightarrow$$

$$-\frac{1}{(al)_x^T} \cdot \frac{d}{dx} (al)_x^T = -\frac{1}{S_{T_a}(x-a)} \cdot \frac{d}{dx} S_{T_a}(x-a) = (\alpha\mu)_x^T$$

Δηλαδή

$$(\alpha\mu)_x^T = \frac{-d \log(al)_x^T}{dx} = -\frac{1}{(al)_x^T} \cdot \frac{d}{dx} (al)_x^T$$

Το ολικό απαύξημα μεταξύ ηλικίας  $x$  και  $x+n$  είναι

$${}_n(ad)_x^T = (al)_x^T - (al)_{x+n}^T = (al)_x^T \cdot \left[1 - \frac{(al)_{x+n}^T}{(al)_x^T}\right] \Rightarrow$$

$${}_n(ad)_x^T = (al)_x^T \cdot [1 - {}_n(ap)_x^T] = (al)_x^T \cdot {}_n(aq)_x^T$$

όπου  ${}_n(aq)_x^T$  εδώ είναι το ποσοστό του απαυξήματος στο διάστημα  $[x, x+n]$  και  ${}_n(ad)_x^T$  το συνολικό πλήθος όσων εξέρχονται από το σύνολο των ζώων ηλικίας  $x$ , ανάμεσα σε ηλικίες  $x$  και  $x+n$ , από όλα τα αίτια.

Ανάλογα μπορούμε να ορίσουμε και ότι

$${}_n(ad)_x^j = (al)_x^j - (al)_{x+n}^j$$

Για  ${}_n(ad)_x^j = (al)_x^j \cdot {}_n(aq)_x^j$  όπως ορίστηκε στα προηγούμενα, όπου  $(al)_x^j$  να δηλώνει το πλήθος των ατόμων που παραμένουν στο αρχικό καθεστώς, από τους αρχικούς  $(al)_x^j$ , οι οποίοι θα υποκύψουν στο  $j$ -αίτιο από την ηλικία  $x$  και μετά:

$$(al)_x^j = \sum_{k=x}^{\infty} (ad)_k^j$$

Αν τα  $m$ -αίτια είναι ξένα μεταξύ τους είναι σαφές ότι :

$$(al)_x^T = \sum_{j=1}^m (al)_x^j$$

Η  ${}_n(aq)_x^j$  δηλώνει την αναλογία από τους  $(al)_x^T$  (ηλικίας  $x$ ) οι οποίοι θα φύγουν από το καθεστώς λόγω της  $j$ -αιτίας πριν την ηλικία  $x+n$  όταν όλες οι  $m$ -αιτίες απαυξήματος λειτουργούν ταυτόχρονα, και η  ${}_n(ad)_x^j$  δηλώνει το πλήθος όσων εξέρχονται από το σύνολο των ζώων ηλικίας  $x$ , ανάμεσα στις ηλικίες  $x$  και  $x+n$ , υποκύπτοντας στο  $j$ -αίτιο, όταν όλες οι  $m$ -αιτίες απαυξήματος λειτουργούν ταυτόχρονα.

Έτσι έχουμε τις σχέσεις

$${}_n(aq)_x^T = \frac{{}_n(ad)_x^T}{(al)_x^T} \quad {}_n(aq)_x^j = \frac{{}_n(ad)_x^j}{(al)_x^T}$$

Και με αυτό τον τρόπο μπορούμε τώρα να κατασκευάσουμε έναν πίνακα με πολλαπλά απαυξήματα από  $(al)_x^j$ ,  ${}_n(ad)_x^j$  και  ${}_n(aq)_x^j$  τιμές για  $j=1,2,\dots,m$ . Παρατηρούμε πάλι ότι

$${}_n(ap)_x^j + {}_n(aq)_x^j = \frac{(al)_{x+n}^j}{(al)_x^T} + \frac{(al)_x^j - (al)_{x+n}^j}{(al)_x^T} = \frac{(al)_x^j}{(al)_x^T}$$

Δηλαδή

$${}_n(ap)_x^j + {}_n(aq)_x^j \neq 1$$

Επίσης, αν τα  $m$ -αίτια είναι ξένα μεταξύ τους τότε ισχύει ότι

$${}_n(ad)_x^T = \sum_{j=1}^m {}_n(ad)_x^j$$

$${}_n(al)_x^T = \sum_{j=1}^m {}_n(al)_x^j$$

$${}_n(aq)_x^j = \frac{{}_n(ad)_x^j}{{}_n(ad)_x^T} \cdot {}_n(aq)_x^T$$

Η **ολική μέση ένταση** (ο ολικός κεντρικός ρυθμός) από όλες τις αιτίες για το ηλικιακό διάστημα  $(x, x+n)$  του απουζήματος, ανά έτος, ένα άτομο, ορίζεται ως

$${}_n(am)_x^T = \frac{\int_0^n \int_t (ap)_x^T \cdot (a\mu)_{x+t}^T dt}{\int_0^n \int_t (ap)_x^T dt} = \frac{\int_0^n \int_t (al)_{x+t}^T \cdot (a\mu)_{x+t}^T dt}{\int_0^n \int_t (al)_{x+t}^T dt} = \frac{{}_n(ad)_x^T}{{}_n(al)_x}$$

Όπου  ${}_n(al)_x$  να δηλώνει την κεντρική έκθεση στον κίνδυνο για άτομα ηλικιών  $(x, x+n)$ .

Παρατηρούμε ότι η ολική μέση ένταση είναι ένας σταθμισμένος μέσος των  $(a\mu)_{x+t}^T$ , με βάρη την ολική συνάρτηση επιβίωσης.

Αναλόγως, η **μέση ένταση του απουζήματος από την  $j$ -αιτία**, ανά έτος, ανά ζωή, γίνεται

$${}_n(am)_x^j = \frac{\int_0^n \int_t (ap)_x^T \cdot (a\mu)_{x+t}^j dt}{\int_0^n \int_t (ap)_x^T dt} = \frac{\int_0^n \int_t (al)_{x+t}^T \cdot (a\mu)_{x+t}^j dt}{\int_0^n \int_t (al)_{x+t}^T dt} = \frac{{}_n(ad)_x^j}{{}_n(al)_x}$$

Η οποία είναι ένας σταθμισμένος μέσος των  $(a\mu)_{x+t}^j$ , με βάρη την ολική συνάρτηση επιβίωσης.

Από την σχέση  ${}_n(ad)_x^T = \sum_{j=1}^m {}_n(ad)_x^j$  έχουμε

$${}_n(am)_x^T = \sum_{j=1}^m {}_n(am)_x^j$$

Καθώς και ότι

$${}_n(am)_x^j = \frac{{}_n(ad)_x^j}{{}_n(ad)_x^T} \cdot {}_n(am)_x^T$$

Για να προσεγγίσουμε  ${}_n(am)_x^j$  τιμές από έναν πίνακα πολλαπλών απαυξημάτων αν υποθέσουμε ότι το **ολικό απαύζημα είναι ομοιόμορφα κατανεμημένο σε κάθε διάστημα ζωής  $[x, x+n)$**  (Uniform Distribution of Total Decrement) έχουμε, με γραμμική παρεμβολή της  $(al)_{x+t}^T$  στο διάστημα  $[x, x+n)$ , ότι

$$(al)_{x+t}^T \cong \left(1 - \frac{t}{n}\right) \cdot (al)_x^T + \frac{t}{n} \cdot (al)_{x+n}^T = (al)_x^T - \frac{t}{n} \cdot {}_n(ad)_x^T, \text{ για } 0 \leq t \leq n$$

Όπου  ${}_n(ad)_x^T = (al)_x^T - (al)_{x+n}^T$ . Οπότε  ${}_n(aL)_x \cong \int_0^n \left[ (al)_x^T - \frac{t}{n} \cdot {}_n(ad)_x^T \right] \cdot dt$  δηλαδή

$${}_n(aL)_x \cong n \cdot \left[ (al)_x^T - \frac{1}{2} \cdot {}_n(ad)_x^T \right]$$

$${}_n(am)_x^j = \frac{{}_n(ad)_x^j}{{}_n(aL)_x} \cong \frac{{}_n(ad)_x^j}{n \cdot \left[ (al)_x^T - \frac{1}{2} \cdot {}_n(ad)_x^T \right]}$$

Και

$${}_n(am)_x^j \cong \frac{{}_n(aq)_x^j}{n \cdot \left[ 1 - \frac{1}{2} \cdot {}_n(aq)_x^T \right]}$$

Αντίστροφα, για να προσεγγίσουμε  ${}_n(aq)_x^j$  τιμές συναρτήσει των  ${}_n(am)_x^j$  τιμών κάτω από

την ίδια υπόθεση έχουμε ότι  ${}_n(aq)_x^j \cong \frac{{}_n(ad)_x^j}{(al)_x^T} \cong \frac{{}_n(am)_x^j \cdot {}_n(aL)_x}{\frac{1}{n} \cdot \left[ {}_n(aL)_x + \frac{n}{2} \cdot {}_n(ad)_x^j \right]} \Rightarrow$

$${}_n(aq)_x^j \cong \frac{{}_n(am)_x^j}{\frac{1}{n} \cdot \left[ 1 + \frac{n}{2} \cdot {}_n(am)_x^T \right]}$$

Επίσης, μπορούμε να αντικαταστήσουμε το  $\frac{n}{2}$  με τον όρο  $n \cdot (1 - {}_n f_x)$ , όπου  ${}_n f_x$  το αναμενόμενο κλάσμα (fraction) του διαστήματος ζωής που αφορά τα τελευταία  $n$ - έτη ζωής



του  $(x)$  και να πάρουμε τη βελτιωμένη σχέση

$${}_n(aq)_x^j = \frac{{}_n(am)_x^j}{\frac{1}{n} \cdot [1 + n \cdot (1 - {}_n f_x) \cdot {}_n(am)_x^T]}$$

Για την κατασκευή ενός πίνακα πολλαπλών απαυξημάτων εφόσον έχουν εκτιμηθεί οι πιθανότητες  ${}_n(aq)_x^T$  και  ${}_n(aq)_x^j$  ακολουθούμε τα εξής βήματα:

- Πρώτα κατασκευάζουμε έναν πίνακα με όλες τις αιτίες μαζί, όπως σαν να είχαμε έναν απλό πίνακα με ένα απαύζημα. Έτσι, διαλέγουμε μια βάση (*radix*)  $(al)_0^T$  και σταδιακά υπολογίζουμε τις ποσότητες

$$(al)_{x+n}^T = (al)_x^T \cdot (1 - {}_n(aq)_x^T) \text{ και } {}_n(ad)_x^T = (al)_x^T \cdot {}_n(aq)_x^T$$

- Το επόμενο βήμα είναι ο υπολογισμός των  ${}_n(ad)_x^j$ , τα οποία μπορούν να υπολογιστούν από την σχέση

$${}_n(ad)_x^j = (al)_x^T \cdot {}_n(aq)_x^j$$

- Τέλος, οι τιμές  $(al)_x^j$  μπορούν να υπολογιστούν από την σχέση

$$(al)_x^j = {}_n(ad)_x^j + {}_n(ad)_{x+n}^j + \dots + {}_n(ad)_\omega^j = {}_n(aq)_x^j \cdot (al)_x^T + {}_n(aq)_{x+n}^j \cdot (al)_{x+n}^T + \dots + (al)_\omega^j$$

Εφόσον οι τιμές  $(al)_x^j$  εκφράζουν τα απαυζήματα από την  $j$ -αίτια τα οποία θα συμβούν μετά την ηλικία  $x$ .

## **2.4 Συνδεόμενοι Πίνακες Μοναδιαίου Απαυζήματος (Associated Single Decrement Tables)**

Για κάθε ένα αίτιο εξόδου σε ένα πίνακα πολλαπλών απαυξημάτων μπορούμε να ορίσουμε ένα μοντέλο μοναδιαίου απαυζήματος το οποίο εξαρτάται μόνο από το συγκεκριμένο απαύζημα και συνεπώς είναι ανεξάρτητο από τα άλλα απαυζήματα που είχαμε στον πίνακα πολλαπλών απαυξημάτων. Σε πίνακα πολλαπλών εξόδων το κάθε ένα απαύζημα δεν είναι ανεξάρτητο από το άλλο, τα αίτια εξόδου λειτουργούν ταυτόχρονα και *ανταγωνιστικά* και κάθε πιθανότητα εξαρτάται από την πιθανότητα των υπολοίπων απαυξημάτων. Έτσι, αν μας ενδιαφέρει ένα συγκεκριμένο  $j$ -αίτιο και αν οι πιθανότητες των υπολοίπων απαυξημάτων αυξηθούν, τότε, λιγότερες ζωές θα ήταν εκτεθειμένες στο συγκεκριμένο απαύζημα και η πιθανότητα θα μειωνόταν. Δηλαδή, η «πραγματική» (ανεξάρτητη) πιθανότητα υπόκυψης στο  $j$ -αίτιο αν δεν παρεμβάλλονταν τα υπόλοιπα αίτια, θα ήταν μεγαλύτερη από το  $(aq)_x^j$ .

Άρα, το  $(aq)_x^j = \frac{(ad)_x^j}{(al)_x^T}$  (ποσοστό του  $(al)_x^T$  που εξέρχεται κάτω από τη δράση του  $j$ -αιτίου)

υποεκτιμά το μέγεθος της «γνήσιας» επίδρασης του αιτίου  $j$ . Για να εκτιμήσουμε την πραγματική (ή απόλυτη ή καθαρή) επίδραση του αιτίου  $j$  θα έπρεπε όλη η ομάδα των  $(al)_x^T$

ατόμων να παραμείνει εκτεθειμένη στο αίτιο  $j$  για ένα ολόκληρο έτος . Είναι όμως προφανές ότι κατά τη διάρκεια του έτους κάποια μέλη της ομάδας αποσπώνται από την ομάδα από τα άλλα αίτια.

Έτσι, θεωρούμε ένα συνδεδεμένο πίνακα μοναδιαίου απανξήματος με απανξήμα  $j$  και ορίζουμε ως

- $q_x^j$  τον **απόλυτο ρυθμό απανξήματος** (absolute rate of decrement) ή **ανεξάρτητο ρυθμό απανξήματος** (independent rate of decrement),

- $m_x^j = \frac{\int_0^1 {}_tP_x^j \cdot \mu_{x+t}^j dt}{\int_0^1 {}_tP_x^j dt}$  τον αντίστοιχο **ανεξάρτητο κεντρικό ρυθμό**, ανά έτος, ανά

ζωή, για έναν πίνακα μοναδιαίου απανξήματος. Παρατηρούμε ότι και αυτός ο κεντρικός ρυθμός είναι σταθμισμένος μέσος των  $\mu_{x+t}^j$  στο έτος ζωής με βάρη τώρα τα  ${}_tP_x^j$ , όπου  ${}_tP_x^j$  την πιθανότητα ότι ο  $(x)$  θα βρίσκεται στο ίδιο καθεστώς (δεν θα υποκύψει στο  $j$ -αίτιο ) μεταξύ ηλικίας  $x$  και  $x+t$  σε έναν πίνακα μοναδιαίου απανξήματος και

- $l_x^j$  τον αριθμό ατόμων σε έναν πίνακα μοναδιαίου απανξήματος ηλικίας  $x$  ακριβώς.

Ο **θεμελιώδης σύνδεσμος** μεταξύ ενός πίνακα πολλαπλών απανξημάτων και ενός πίνακα μοναδιαίου απανξήματος είναι η σχέση

$$(\alpha\mu)_{x+t}^j = \mu_{x+t}^j$$

Η ισότητα αυτή ισχύει κάτω από την **ΥΠΟΘΕΣΗ 3**, όπου

**ΥΠΟΘΕΣΗ 3:** οι τυχαίες μεταβλητές  $X_1, \dots, X_n$  είναι αμοιβαία ανεξάρτητες , γιατί τότε  $S_{X_1, \dots, X_m}(x_1, \dots, x_m) = S_{X_1}(x_1) \dots S_{X_m}(x_m)$  , όπου  $S_{X_j}(x_j)$  δηλώνει την περιθώρια συνάρτηση επιβίωσης της  $S_{X_1, \dots, X_m}(x_1, \dots, x_m)$  σε σχέση με το  $X_j$  και η αντίστοιχη απόλυτη ένταση ενός

απανξήματος γίνεται  $\mu_{x+t}^j = -\frac{d \log S_{X_j}(t)}{dt}$  η οποία είναι ίδια με την ένταση  $(\alpha\mu)_{x+t}^j$  διότι

$$(\alpha\mu)_{x+t}^j = \lim_{h \rightarrow 0} \frac{1}{h} \cdot P\{t < X_j \leq t+h / \bigcap_{j=1}^m (X_j > t)\} = \lim_{h \rightarrow 0} \frac{1}{h} \cdot P\{t < X_j \leq t+h / (X_j > t)\} \Rightarrow$$

$$(\alpha\mu)_{x+t}^j = -\frac{d \log S_{X_j}(t)}{dt}$$

Σχολιάζοντας πιθανές ερμηνείες μιας συνάρτησης επιβίωσης εξαρτώμενη μόνο από την  $j$ -αίτια , αποκλείουμε πρώτα το ενδεχόμενο να μιλάμε για έναν πληθυσμό που υπόκειται μόνο στο  $j$ -αίτιο διότι ένας τέτοιος πληθυσμός δεν υπάρχει ούτε και μπορεί να κατασκευαστεί και επίσης αναιρεί την **ΥΠΟΘΕΣΗ 2**. Μια προσέγγιση θα μπορούσε να ήταν αν υποθέσουμε ότι ο πληθυσμός αποτελείται από ομοιογενείς υπό-πληθυσμούς καθένας από τον οποίο είναι

υποκείμενος σε διαφορετικό απαύξημα και χειρίζοντας την περιθώρια συνάρτηση επιβίωσης ως μια πιθανή δομή για την δεδομένη αιτία ξεχωριστά. Έτσι, αν η από κοινού συνάρτηση επιβίωσης είναι  $S_{X_1, \dots, X_m}(x_1, \dots, x_m)$  και αν το καθεστώς τερματίζει μόνο από την  $j$ -αιτία τότε όλοι οι άλλοι χρόνοι απαυξήματων εξέρχονται θέτοντάς τους ίσους με το μηδέν. Έτσι, η περιθώρια συνάρτηση επιβίωσης  $S_{X_j}(x_j) = S_{X_1, \dots, X_m}(0, \dots, x_j, \dots, 0)$  θα μπορούσε να θεωρηθεί ότι αντιστοιχεί στον χρόνο κατανομής του  $j$ -απαυξήματος ανεξάρτητα.

Η χρησιμότητα των απόλυτων ή ανεξάρτητων ρυθμών απαυξήματων έγκειται στην κατασκευή μοντέλων και πινάκων πολλαπλών απαυξήματων εφόσον συνήθως για την κατασκευή ενός πίνακα πολλαπλών απαυξήματων επιλέγουμε πρώτα κατάλληλους απόλυτους ρυθμούς απαυξήματων και κάτω από κάποιες υποθέσεις εκτιμούμε τις αντίστοιχες  ${}_t(aq)_x^j$  τιμές. Επίσης ισχύει και το αντίστροφο, δηλαδή από  ${}_t(aq)_x^j$  τιμές να εκτιμούμε απόλυτες τιμές.

Έτσι, ορίζουμε τον **ανεξάρτητο ρυθμό του  $j$ -απαυξήματος** ως

$${}_t p_x^j = e^{-\int_0^t \mu_{x+s}^j ds}$$

Με  ${}_t q_x^j = 1 - {}_t p_x^j$ . Σε αντίθεση με την συνάρτηση επιβίωσης  ${}_t(ap)_x^T$ , ο ρυθμός  ${}_t p_x^j$  δεν ορίζει απαραίτητα συνάρτηση επιβίωσης και δεν απαιτείται ότι  $\lim_{t \rightarrow \infty} {}_t p_x^j = 0$ , δηλαδή μπορεί

να υπάρχουν απαυξήματα για τα οποία το ολοκλήρωμα  $\int_0^{\infty} \mu_{x+t}^j dt \neq 0$ . Παρατηρούμε επίσης

ότι  $\int_0^{\infty} \mu_{x+t}^T dt = \infty$ , οπότε πρέπει για τουλάχιστον ένα απαύξημα  $j$  να ισχύει  $\int_0^{\infty} \mu_{x+t}^j dt = \infty$ .

Επίσης, από την σχέση  ${}_t(ap)_x^T = e^{-\int_0^t \sum_{j=1}^m \mu_{x+s}^j ds}$  έχουμε ότι

$${}_t(ap)_x^T = e^{-\int_0^t \sum_{j=1}^m \mu_{x+s}^j ds} = \prod_{j=1}^m {}_t p_x^j$$

Το αποτέλεσμα αυτό δεν περιέχει κάποια προσέγγιση, αλλά είναι ακριβές σχέση κάτω από την **ΥΠΟΘΕΣΗ 3** ότι η απόλυτη ένταση του  $j$ -απαυξήματος,  $\mu_{x+t}^j$ , ταυτίζεται με την ένταση του  $j$ -απαυξήματος,  $(\alpha\mu)_{x+t}^j$ , σε ένα μοντέλο πολλαπλών απαυξήματων.

Τέλος, παρατηρούμε ότι

$${}_t(ap)_x^T = \prod_{j=1}^m {}_t p_x^j \Rightarrow {}_t p_x^j \geq {}_t(ap)_x^T \Rightarrow {}_t p_x^j \cdot \mu_{x+t}^j \geq {}_t(ap)_x^T \cdot \mu_{x+t}^j \Rightarrow$$

$$q_x^j = \int_0^1 p_x^j \cdot \mu_{x+t}^j dt \geq \int_0^1 (ap)_x^j \cdot \mu_{x+t}^j dt = (aq)_x^j, \text{ δηλαδή}$$

$$q_x^j \geq (aq)_x^j, \forall j, x$$

## Κεφάλαιο 3

### Κεντρικός Ρυθμός Θνησιμότητας (Central Rate of Mortality)

Ένας άλλος ετήσιος δείκτης μέτρησης της θνησιμότητας είναι ο *κεντρικός ρυθμός θανάτου* (*central death rate*) ή ο **κεντρικός ρυθμός θνησιμότητας (Central rate of mortality)** ή η **μέση ένταση**, συμβολίζεται ως  $m_x$  και ορίζεται για κάθε έτος ζωής  $[x, x+1)$  ως

$$m_x = \frac{\int_0^1 S_{T_x}(t) \cdot \mu_{x+t} dt}{\int_0^1 S_{T_x}(t) dt} = \frac{\int_0^1 l_{x+t} \cdot \mu_{x+t} dt}{\int_0^1 l_{x+t} dt} = -\frac{\int_0^1 dl_{x+t} \cdot dt}{\int_0^1 l_{x+t} dt} = \frac{d_x}{L_x} \quad \text{για } x=0, 1, 2, \dots, \omega-1$$

Από τον ορισμό, το  $m_x$ , μπορεί να ερμηνευτεί ως ο σταθμισμένος μέσος της ποσότητας  $\mu_{x+t}$  στο διάστημα  $x$  ως  $x+1$ , με τα βάρη να αναφέρονται σε πληθυσμό που είναι παρών σε ηλικίες  $x+t$ , δηλαδή  $l_{x+t}$  ή ισοδύναμα με βάρη την συνάρτηση επιβίωσης της απομένουσας ζωής.

Ο κεντρικός ρυθμός θνησιμότητας,  $m_x = \frac{d_x}{L_x}$ , είναι ο πιο σημαντικός δείκτης θνησιμότητας

σε σχέση με την πιθανότητα θνησιμότητας,  $q_x = \frac{d_x}{l_x}$ , διότι οι αναμενόμενες τιμές θανάτων,

που προέρχεται από μια έρευνα ενός πραγματικού πληθυσμού ατόμων, έχουν καλύτερες στατιστικές ιδιότητες όταν εκτιμούμε  $m_x$  τιμές παρά τις  $q_x$  ως προς την αποδοτικότητα της εκτίμησης και των ασυμπτωτικών καλών ιδιοτήτων τους (όπως αναλύεται περαιτέρω), καθώς και ότι η  $q_x$  έχει την αδυναμία ότι εκφράζει μόνο το συνολικό αποτέλεσμα της θνησιμότητας μέσα στο έτος ηλικίας  $x$  και  $x+1$  δηλαδή μας δίνει τελικά πόσοι πέθαναν στο τέλος του έτους και δεν επηρεάζεται από τον τρόπο που αυτοί οι θάνατοι είναι κατανεμημένοι στο αντίστοιχο έτος ζωής.

Επίσης, ο κεντρικός ρυθμός θνησιμότητας  $m_x$  είναι ο λόγος δύο ποσοτήτων. Στον αριθμητή έχουμε τους θανάτους που αναμένουμε μεταξύ ηλικιών  $x$  και  $x+1$  να συμβούν από τις  $l_x$  ζωές

που είναι παρόντες σε ηλικία  $x$ . Στον παρανομαστή η ποσότητα  $L_x := \int_0^1 l_{x+t} \cdot dt$  μπορεί να ερμηνευτεί με δυο τρόπους:

- Ο μέσος πληθυσμός που είναι παρών μεταξύ των ηλικιών  $x$  και  $x+1$  σε έναν πίνακα

$$\text{ζωής, διότι } \int_0^1 l_{x+t} dt = \frac{\int_0^1 l_{x+t} dt}{\int_0^1 dt}$$

- Ο συνολικός χρόνος σε έτη που αναμένεται να ζήσουν οι  $l_x$  ζωές μεταξύ ηλικίας  $x$  και  $x+1$ , δηλαδή η (αναμενόμενη) *κεντρική έκθεση στον κίνδυνο*, εφόσον

$$L_x = \int_0^1 l_{x+t} dt = l_x \cdot \int_0^1 \frac{l_{x+t}}{l_x} dt = l_x \cdot \int_0^1 {}_t p_x dt = l_x \cdot e_{x:1}^0, \text{ όπου } \int_0^1 {}_t p_x dt = e_{x:1}^0 \text{ ορίζεται ως ο}$$

αναμενόμενος μέσος χρόνος ζωής στο ηλικιακό διάστημα  $[x, x+1)$  για άτομο ηλικίας  $x$ . Πράγματι, αν ορίσουμε ως τυχαία μεταβλητή  $S$  την απομένουσα ζωή του ( $x$ ) στο

$$\text{επόμενο έτος ζωής: } R_S = [0, 1] \text{ με } S = \begin{cases} 1 & \alpha v T_x \geq 1 \\ a(x) & \alpha v T_x < 1 \end{cases} \text{ τότε}$$

$$e_{x:1}^0 = E(S) = E(S | T_x < 1) \cdot P(T_x < 1) + E(S | T_x \geq 1) \cdot P(T_x \geq 1) \Rightarrow$$

$$e_{x:1}^0 = a(x) \cdot q_x + p_x$$

Όπου

$$E(S | T_x < 1) = \alpha(x) = \frac{\int_0^1 t \cdot {}_t p_x \cdot \mu_{x+t} \cdot dt}{\int_0^1 {}_t p_x \cdot \mu_{x+t} \cdot dt} = \frac{\int_0^1 t \cdot {}_t p_x \cdot \mu_{x+t} \cdot dt}{q_x} = E[T_x / T_x < 1]$$

Άρα

$$L_x = l_x \cdot e_{x:1}^0 = a(x) \cdot d_x + l_{x+1}$$

Και

$$m_x = \frac{d_x / l_x}{L_x / l_x} = \frac{q_x}{e_{x:1}^0} \Rightarrow$$

$$m_x = \frac{q_x}{p_x + q_x \cdot a(x)}$$

Ερμηνεύοντας την ποσότητα  $a(x)$ , έχουμε ότι  $a(x)$  είναι ο μέσος χρόνος ζωής μέσα στο  $[x, x+1)$ , από άτομα ηλικίας  $x$ , για αυτούς που θα αποβιώσουν μέσα στο διάστημα  $[x, x+1)$  ή ισοδύναμα το  $a(x)$  *συμβολίζει την δεσμευμένη προσδόκιμη ζωή του ( $x$ ) στο επόμενο έτος ζωής του, δεδομένου αποβίωσής του ( $x$ ) στο έτος αυτό.* Έτσι,

η σχέση  $e_{x:1}^0 = p_x \cdot 1 + q_x \cdot a(x)$  δηλώνει ότι η προσδόκιμη ζωή του ( $x$ ) στο επόμενο έτος ζωής του ( $x, x+1$ ) είναι 1 αν επιζήσει και  $a(x)$  αν αποβιώσει, οπότε ο  $e_{x:1}^0$  είναι ο

αδέσμευτος μέσος ή η μέση απομένουσα ζωή του ( $x$ ) στο επόμενο έτος ζωής του.

Για την εκτίμηση του  $a(x)$ , **συνήθεις υποθέσεις** που έχουμε είναι οι εξής:

- Αν θεωρήσουμε ότι οι θάνατοι είναι ομοιόμορφα κατανομημένοι στο έτος ζωής από  $[x, x+1)$  ή ισοδύναμα, όπως έχουμε εξετάσει στην **UDD μέθοδο**, τότε γνωρίζουμε ότι  ${}_t p_x \cdot \mu_{x+t} = q_x$ , οπότε

$$\alpha(x) = \frac{\int_0^1 t \cdot {}_t p_x \cdot \mu_{x+t} \cdot dt}{q_x} = \int_0^1 t \cdot dt = \frac{1}{2}$$

Δηλαδή

$$m_x = \frac{q_x}{p_x + q_x \cdot a(x)} = \frac{q_x}{1 - \frac{1}{2} q_x}$$

Η

$$q_x = \frac{m_x}{1 + \frac{1}{2} m_x}$$

Διαφορετικά, όπως έχουμε εξετάσει στην UDD μέθοδο, η συνάρτηση  $l_{x+t}$  είναι γραμμική συνάρτηση του  $t$ , για  $0 \leq t \leq 1$ . Τότε, όπως έχουμε δει,

$$\int_0^1 l_{x+t} \cdot dt = l_x - \frac{1}{2} \cdot d_x, \text{ οπότε } m_x = \frac{d_x}{\int_0^1 l_{x+t} \cdot dt} = \frac{d_x}{l_x - \frac{1}{2} d_x} = \frac{q_x}{1 - \frac{1}{2} q_x}.$$

Επίσης, σύμφωνα με την UDD μέθοδο, από τις σχέσεις  $m_x = \frac{q_x}{1 - \frac{1}{2} q_x}$  και

$\mu_{x+t} \cong \frac{q_x}{1-t \cdot q_x}$  παίρνουμε ότι

$$\mu_{x+\frac{1}{2}} \cong \frac{q_x}{1 - \frac{1}{2} q_x}$$

Δηλαδή, κάτω από την UDD καταλήγουμε  $m_x \cong \mu_{x+\frac{1}{2}}$ . Το αποτέλεσμα αυτό

συμφωνεί με την ερμηνεία που δώσαμε στο  $m_x$  ως ο σταθμισμένος μέσος της  $\mu_{x+t}$  με βάρη τον πληθυσμό που είναι παρών, διότι αν η συνάρτηση  $l_{x+t}$ , που αντιπροσωπεύει τον παρόντα πληθυσμό, είναι γραμμική, οπότε ο σταθμισμένος μέσος του  $\mu_{x+t}$  θα πρέπει προσεγγιστικά να είναι ίσος με το  $\mu_{x+\frac{1}{2}}$ .

- Στην περίπτωση τώρα που θεωρήσουμε ότι η ένταση θνησιμότητας είναι σταθερή σε

κάθε έτος ηλικίας (**CFM μέθοδος**), από την σχέση  $m_x = \frac{\int_0^1 l_{x+t} \cdot \mu_{x+t} dt}{\int_0^1 l_{x+t} dt}$  έχουμε

$m_x \cong \mu_x$  και λόγω της σχέσης  $1 - q_x = e^{-\int_0^1 \mu_{x+t} dt} = e^{-\mu_x}$ , καταλήγουμε ότι

$$m_x = -\log(1 - q_x)$$

Επίσης, έχουμε ότι  $a(x) = \frac{\int_0^1 t \cdot {}_tP_x \cdot \mu_{x+t} \cdot dt}{\int_0^1 {}_tP_x \cdot \mu_{x+t} \cdot dt} = \frac{\int_0^1 t \cdot e^{-t \cdot \mu_x} \cdot \mu_x \cdot dt}{q_x}$ , οπότε

$$a(x) = \frac{\frac{1 - e^{-\mu_x}}{\mu_x} - e^{-\mu_x}}{1 - e^{-\mu_x}} = \frac{q_x - p_x \cdot \mu_x}{q_x \cdot \mu_x} = -\frac{q_x - p_x \cdot \log(p_x)}{q_x \cdot \log(p_x)}$$

- Στην περίπτωση τώρα που θεωρήσαμε την υπόθεση την **μέθοδο Balducci** έχουμε

$$m_x = \frac{d_x}{\int_0^1 l_{x+t} dt} = \frac{q_x}{\int_0^1 {}_tP_x \cdot dt} = \frac{q_x}{\int_0^1 \frac{p_x}{p_x + t \cdot q_x} \cdot dt}$$
 και λύνοντας

$$m_x = \frac{-q_x^2}{(1 - q_x) \cdot \log(1 - q_x)}$$

Επίσης ισχύει ότι  $\alpha(x) = \frac{\int_0^1 t \cdot {}_tP_x \cdot \mu_{x+t} \cdot dt}{\int_0^1 {}_tP_x \cdot \mu_{x+t} \cdot dt} = \frac{\int_0^1 t \cdot \frac{p_x \cdot q_x}{[1 - (1-t) \cdot q_x]^2} \cdot dt}{q_x}$  ή

$$\alpha(x) = -\frac{p_x}{q_x^2} \cdot [q_x + \log p_x]$$

- Τέλος, αν υποθέσουμε ότι η συνάρτηση επιβίωσης της απομένουσας ζωής,  $T_x$ , ανήκει στην **οικογένεια των κλασματικών ηλικιών (FAA)**, δηλαδή ότι η συνάρτηση επιβίωσης μπορεί να περιγραφεί από την σχέση

$${}_tP_x = \begin{cases} [1 - t + t \cdot p_x^{a_x}]^{1/a_x} & a_x \neq 0 \\ p_x^t & a_x = 0 \end{cases}, \text{ καταλήγουμε ότι ο κεντρικός ρυθμός}$$

θνησιμότητας δίνεται, για διάφορες τιμές της ποσότητας  $a_x$ , από τον τύπο

$$m_x = \frac{q_x \cdot (1 + a_x) \cdot (p_x^{a_x} - 1)}{a_x \cdot (p_x^{a_x} - 1)} \text{ όταν } a_x \neq 0 \text{ (πράγματι όταν } a_x = 1 \text{ ισοδυναμεί με την}$$

UDD μέθοδο και όταν  $a_x = -1$  ισοδυναμεί με την μέθοδο *Baltucci*). Για τον υπολογισμό της ποσότητας  $a(x) = E[T_x / T_x < 1]$ , σε αυτή την γενική περίπτωση έχουμε

$$\alpha(x) = \frac{\int_0^1 t \cdot {}_t p_x \cdot \mu_{x+t} \cdot dt}{\int_0^1 {}_t p_x \cdot \mu_{x+t} \cdot dt} = \frac{-p_x + \frac{(p_x^{a_x})^{\frac{1}{a_x} + 1} - 1}{(\frac{1}{a_x} + 1) \cdot (p_x^{a_x} - 1)}}{q_x}, \text{ για } a_x \neq 0$$

Και

$$a_x = \frac{-\int_0^1 t \cdot p_x^t \cdot \ln(p_x) \cdot dt}{q_x} = \frac{1 + p_x - p_x \cdot \ln(p_x)}{q_x \cdot \ln(p_x)}$$

- Αν δεν θεωρήσουμε **καμία υπόθεση** για την συνάρτηση επιβίωσης  $S_T(x+t)$ , σε κάθε έτος ηλικίας, τότε στην γενική περίπτωση θα έχουμε, σύμφωνα με τα προηγούμενα, ότι

$$L_x = l_x \cdot e_{x|}^0 = l_{x+1} + d_x \cdot a(x) = (l_x - d_x) + d_x \cdot a(x) = l_x - \{1 - a(x)\} \cdot d_x \Rightarrow$$

$$l_x = L_x + \{1 - a(x)\} \cdot d_x$$

Οπότε

$$m_x = \frac{d_x}{l_x - \{1 - a(x)\} \cdot d_x} = \frac{q_x}{1 - \{1 - a(x)\} \cdot q_x}$$

Η

$$q_x = \frac{d_x}{l_x} = \frac{d_x}{L_x + \{1 - a(x)\} \cdot d_x} = \frac{m_x}{1 + \{1 - a(x)\} \cdot m_x}$$

Γενικεύοντας, ανάλογα αποτελέσματα έχουμε και για **ομαδοποιημένες ηλικίες** ορίζοντας ως  ${}_n a(x)$  τον αναμενόμενο (μέσο) αριθμό ετών ζωής για άτομο ηλικίας  $x$

που αποβιώνει στο διάστημα ηλικίας  $[x, x+n)$ . Επίσης, ορίζουμε ως  ${}_n f_x = \frac{1}{n} \cdot {}_n a(x)$  το αναμενόμενο κλάσμα (function) του διαστήματος ζωής που αφορά τα τελευταία  $n$ -έτη ζωής του ( $x$ ). Τα αναμενόμενα έτη ζωής των  $l_x$  ατόμων, ηλικίας  $x$ , για τα

επόμενα  $n$ -έτη το ορίζουμε ως  ${}_n L_x$ , το οποίο αποτελείται από

(α)  $n$ -έτη για κάθε επιζώντα  $l_{x+n}$  από τα  $l_x$  άτομα και

(β)  ${}_n a(x) \cdot {}_n d_x$  έτη για άτομα που αποβιώνουν στο διάστημα  $[x, x+n)$ .

Έτσι έχουμε ότι

$${}_n L_x = n \cdot l_{x+n} + {}_n a(x) \cdot {}_n d_x = n \cdot (l_x - {}_n d_x) + n \cdot {}_n f_x \cdot {}_n d_x = n \cdot [l_x - (1 - {}_n f_x) \cdot {}_n d_x]$$

Η λύνοντας ως προς  $l_x$  έχουμε  $l_x = \frac{1}{n} \cdot [{}_n L_x + n \cdot (1 - {}_n f_x) \cdot {}_n d_x]$ . Οπότε, ορίζοντας

ως  ${}_n m_x$  τον κεντρικό ρυθμό θνησιμότητας για τα έτη ηλικιών  $[x, x+n)$ , καταλήγουμε ότι

$${}_n m_x = \frac{{}_n d_x}{{}_n L_x} = \frac{{}_n d_x}{n \cdot [l_x - (1 - {}_n f_x) \cdot {}_n d_x]} \Rightarrow {}_n m_x = \frac{{}_n q_x}{n \cdot [1 - (1 - {}_n f_x) \cdot {}_n q_x]}$$



Καθώς και ότι

$${}_n q_x = \frac{{}_n d_x}{l_x} = \frac{{}_n d_x}{\frac{1}{n} \cdot [{}_n L_x + n \cdot (1 - {}_n f_x) \cdot {}_n d_x]} \Rightarrow {}_n q_x = \frac{{}_n m_x}{\frac{1}{n} \cdot [1 + n \cdot (1 - {}_n f_x) \cdot {}_n m_x]}$$

Αποτελέσματα σημαντικά για την κατασκευή ενός **συντετμημένου** (*abridged*) **πίνακα θνησιμότητας**, δηλαδή πίνακα θνησιμότητας για ομαδοποιημένες ηλικίες.

## Κεφάλαιο 4

### COPULAS

#### 4.1 Εισαγωγή

Τα Copulas παρέχουν έναν κατάλληλο τρόπο για να διαμορφώσουν και να μετρήσουν την εξάρτηση μεταξύ του χρόνου αποτυχίας τυχαίων μεταβλητών, δεδομένου ότι δίνουν τη δομή της εξάρτησης που αφορά τις γνωστές περιθώριες κατανομές του χρόνου της πολυμεταβλητής από κοινού κατανομής τους. Για να δούμε αυτό πρέπει πρώτα να παρέχουμε μια σύντομη εισαγωγή των Copulas.

Αν υποθέσουμε ότι  $u = (u_1, \dots, u_m)'$ ,  $u_j \in [0, 1]$ , ένα  $m$  copula  $C(u)$  είναι συμβατικά ορισμένο σαν μια πολυμεταβλητή σωρευτική συνάρτηση κατανομής με ομοιόμορφα περιθώρια. Ένας πιθανοθεωρητικός τρόπος να ορίσουμε το Copula παρέχεται στο θεώρημα του Sklar (1959). Έστω,  $X_1, \dots, X_m$  είναι τυχαίες μεταβλητές με συνεχείς συναρτήσεις κατανομής  $F_1, \dots, F_m$  και συναρτήσεις επιβίωσης  $S_1 = 1 - F_1, \dots, S_m = 1 - F_m$  αντίστοιχα, και από κοινού κατανομή και συνάρτηση επιβίωσης  $H(x_1, \dots, x_m)$  και  $S(x_1, \dots, x_m)$ . Το θεώρημα του Sklar μας λέει ότι αν  $H$  είναι μια  $m$ -διάστασης συνάρτηση κατανομής του τυχαίου διανύσματος  $(X_1, \dots, X_m)$  με συνεχή περιθώρια  $F_1, \dots, F_m$ , τότε υπάρχει ένα μοναδικό  $m$  copula  $C$ , τέτοιο ώστε για όλα τα  $x$  στο  $\mathbb{R}^m$

$$H(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m))$$

Και αντίστροφα, αν  $C$  είναι ένα  $m$  copula και  $F_1, \dots, F_m$  είναι κατανομή συνάρτησης επιβίωσης, τότε  $H$  είναι  $m$ -διάστασης συνάρτηση κατανομής με περιθώρια  $F_1, \dots, F_m$ . Για αυτό το λόγο το τυχαίο διάνυσμα  $(X_1, \dots, X_m) \sim H$  είναι συνάρτηση κατανομής του τυχαίου διανύσματος  $(F_1(x_1), \dots, F_m(x_m))$ . Έτσι, από τη σχέση

$H(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m))$  μπορεί κανείς να κατασκευάσει μια δομή εξάρτησης, για παράδειγμα, αν  $m$ -διάστατη συνάρτηση κατανομής  $H$  με κατάλληλη επιλογή ενός συνόλου των περιθωρίων  $F_1, \dots, F_m$  και μια συνάρτηση copula  $C$ . Προκειμένου να κατασκευάσουμε μια συνάρτηση Copula, μια συνέπεια του θεωρήματος του Sklar μπορεί να εφαρμοστεί, σύμφωνα με το ποιο Copula μπορεί να παρασταθεί σαν ένα  $m$ -διάστασης

συνάρτηση κατανομής με συνεχή περιθώρια, αξιολογούνται ως οι αντίστροφες συναρτήσεις  $F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)$  ορισμένες ως  $F_i[F_i^{-1}(x_i)] = x_i$  για παράδειγμα

$$C(u_1, \dots, u_m) = H(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m))$$

Χρησιμοποιώντας την πιθανότητα αναπόσπαστης μεταμόρφωσης

$X_j \mapsto F_j(X_j) = 1 - S_j(X_j)$  ως αποτέλεσμα της οποίας  $S_j(X_j)$  έχει μια ομοιόμορφη κατανομή στο  $[0,1]$ , εύκολα επαληθεύεται (βλ. Sklar 1996) το θεώρημα του Sklar, από τη σχέση  $H(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m))$  μπορεί να επαναδιατυπωθεί για να εκφράσει τη λειτουργία της πολυμεταβλητής συνάρτησης επιβίωσης  $S(x_1, \dots, x_m)$  μέσω ενός κατάλληλου Copula  $\bar{C}$  καλείται Copula επιβίωσης της  $(X_1, \dots, X_m)$ . Έτσι,

$$S(x_1, \dots, x_m) = \bar{C}(S_1(x_1), \dots, S_m(x_m))$$

Η Copula επιβίωσης, συνδέεται με την περιθώρια συνάρτηση επιβίωσης  $S_1(x_1), \dots, S_m(x_m)$  στην πολυμεταβλητή από κοινού συνάρτηση κατανομής  $S(t_1, \dots, t_m)$  με τον ίδιο τρόπο όπως η Copula C αναφέρεται η περιθώρια συνάρτηση κατανομής στην πολυμεταβλητή συνάρτηση κατανομής.

Ας σημειώσουμε ότι η Copula C και η Copula επιβίωσης  $\bar{C}$  ενός τυχαίου διανύσματος  $(X_1, \dots, X_m)$  δεν συμπίπτουν γενικά. Για να δούμε πως η Copula επιβίωσης εκφράζεται μέσω του αντίστοιχου Copula, αναφερόμαστε στους Nelsen (1999) και Georges et al. (2001). Η Copula επιβίωσης,  $\bar{C}$ , στο  $S(x_1, \dots, x_m) = \bar{C}(S_1(x_1), \dots, S_m(x_m))$  μπορεί να εκφραστεί μέσω του Copula C, προκύπτουν με βάση συγκεκριμένη κατανομή  $F_1, \dots, F_m$ , χρησιμοποιείται η  $C(u_1, \dots, u_m) = H(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m))$ . Ωστόσο, δεδομένου ότι  $\bar{C}(u)$  είναι ένα Copula, μπορεί κανείς να χρησιμοποιήσει άμεσα οποιαδήποτε Copula  $C(u)$  για να συνδέσει τη λειτουργία κοινής συνάρτησης επιβίωσης στην  $S(x_1, \dots, x_m) = \bar{C}(S_1(x_1), \dots, S_m(x_m))$  στα περιθώρια  $S_1(x_1), \dots, S_m(x_m)$ .

Αυτή θα είναι η προσέγγιση που υιοθετείται εδώ στην μοντελοποίηση της λειτουργίας της από κοινού συνάρτησης επιβίωσης των ανταγωνιστικών κινδύνων του χρόνου επιβίωσης.

Συμπεραίνουμε, την περιγραφή του γενικού υπόβαθρου των Copulas υπενθυμίζοντας τα θεμελιώδη όρια ανισότητας Fréchet-Hoeffding που στηρίζονται σε σχέση με κάθε συνάρτηση Copula  $m$ -διαστάσεων. Για να συλλάβουμε τη σημασία της, θα δώσουμε πρώτα τη διμεταβλητή έκδοση, δηλαδή,

$$\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2)$$

Όπου το μέγιστο και το ελάχιστο στην παραπάνω εξίσωση είναι αντίστοιχα τα άνω και κάτω όρια του Fréchet – Hoeffdings τα οποία είναι τα ίδια τα Copulas. Αν ακολουθήσουμε την συνάρτηση  $\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2)$ , έτσι ώστε οι τυχαίες μεταβλητές

$X_1, X_2$  με την συνάρτηση Copula  $C(u_1, u_2)$  είναι απολύτως αρνητικά (θετικά) εξαρτημένες, η  $C(u_1, u_2)$  πρέπει να συμπίπτει με το κάτω (άνω) όριο Frechet – Hoeffding. Ο τυχαίος μεταβλητές  $X_1, X_2$  είναι ανεξάρτητες αν και μόνο αν το Copula είναι ίσο με το προϊόν Copula, δηλαδή, αν και μόνο αν  $C(u_1, u_2) = u_1 u_2$ . Μπορεί στη συνέχεια να αποδειχθεί ότι η συνάρτηση  $\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2)$  γενικεύει την πολυμεταβλητή περίπτωση, όπως

$$\max(u_1 + \dots + u_m - m + 1, 0) \leq C(u_1, \dots, u_m) \leq \min(u_1, \dots, u_m)$$

Πολλές δημοφιλείς οικογένειες των Copulas εξαρτώνται από ένα σύνολο παραμέτρων, καθόλου άσχετο με τις παραμέτρους των περιθώριων κατανομών. Αυτό οφείλεται στο γεγονός ότι, στην πραγματικότητα τα Copulas είναι αμετάβλητα κάτω από την αυξανόμενη μετατροπή των αντίστοιχων τυχαίων μεταβλητών τους, ως εκ τούτου είναι «αμετάβλητη κλίμακα».

Το θεώρημα του Sklar και η συνέπεια του, δίνεται από την συνάρτηση  $C(u_1, \dots, u_m) = H(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m))$ , η οποία παρέχει ένα κατάλληλο εργαλείο για την κατασκευή των Copulas. Παραδείγματα τέτοιων Copulas είναι το πολυμεταβλητό Gaussian και Student-t Copula που ανήκουν στην ευρύτερη κατηγορία των Elliptical Copulas. Ωστόσο, υπάρχουν και άλλοι τρόποι κατασκευής των Copulas. Για παράδειγμα, τα δημοφιλή Archimedean Copulas κατασκευάζονται ως,

$$C^A(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$$

Όπου το  $\phi$  είναι συνεχής, κυρτή συνάρτηση που ονομάζεται γεννήτρια, έτσι ώστε  $\phi(1) = 0$  και  $\phi(0) = \infty$  (Nelsen, 1999). Σε ό,τι ακολουθεί, θα εισάγουμε και θα χρησιμοποιήσουμε το Frank Copula ως ένα από τα δυο μόνο γνωστά ολοκληρωμένα, δισδιάστατα Archimedean Copulas. Επιπλέον, θα χρησιμοποιήσουμε επίσης το Plackett Copula, το οποίο δεν ανήκει στην Elliptical ή στην Archimedean οικογένεια αλλά είναι ολοκληρωμένο και ως εκ τούτου, κατάλληλο για χρήση στο μοντέλο ανταγωνιστικών κινδύνων. Για αυτές και άλλες ιδιότητες του Frank και Plackett Copulas αναφέρεται ο Nelsen (1999).

#### **4.1 Ορισμοί**

Έστω δυο τυχαίες μεταβλητές  $X, Y$  με συναρτήσεις κατανομής  $F, G$ , αντίστοιχα, όπου  $F(x) = P(X \leq x)$  και  $G(y) = P(Y \leq y)$  και η από κοινού συνάρτηση κατανομής  $H(x, y) = P(X \leq x, Y \leq y)$ . Επίσης, έστω  $h(x, y)$  η από κοινού συνάρτηση πυκνότητας των παραπάνω τυχαίων μεταβλητών. Στη συνέχεια θα συμβολίσουμε με  $\bar{R}$  το εκτεταμένο σύνολο

των πραγματικών αριθμών το οποίο ορίζεται ως:  $\bar{R} = R \cup \{-\infty, +\infty\}$ . Δίνουμε τους παρακάτω ορισμούς:

**ΟΡΙΣΜΟΣ 4.1.1 :** (H- όγκος). Έστω  $S_1, S_2$  υποσύνολα του  $\bar{R}$  και  $H$  μια δισδιάστατη πραγματική συνάρτηση με πεδίο ορισμού  $S_1 \times S_2$ . Θεωρούμε το ορθογώνιο παραλληλόγραμμο  $B = [x_1, x_2] \times [y_1, y_2]$  του οποίου οι κορυφές ανήκουν στο  $S_1 \times S_2$ . Ο H-όγκος του  $B$  ορίζεται ως εξής :

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

Με λίγα λόγια, όταν  $H$  είναι η από κοινού συνάρτηση κατανομής δυο συνεχών τυχαίων μεταβλητών,  $X, Y$  ο H- όγκος είναι εύκολο να διαπιστωθεί ότι ισούται με την πιθανότητα :

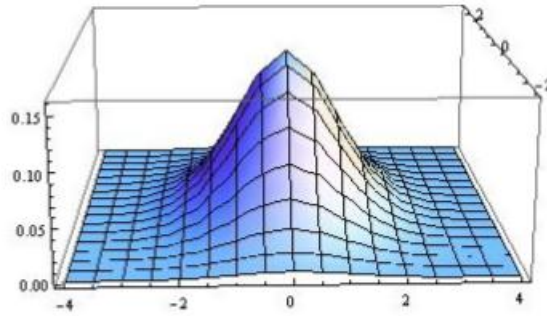
$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} h(x, y) dx dy$$

Πράγματι, χρησιμοποιώντας τον τύπο  $P(AB^C) = P(A) - P(AB)$  μπορούμε να γράψουμε

$$\begin{aligned} P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) &= P(X \leq x_2, y_1 \leq Y \leq y_2) - P(X < x_1, y_1 \leq Y \leq y_2) = \\ &= P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) - P(X < x_1, y_1 \leq Y \leq y_2) = \\ &= P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) - P(X \leq x_1, Y \leq y_2) + P(X \leq x_1, Y \leq y_1) = \\ &= H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \end{aligned}$$

Για παράδειγμα, έστω μια δισδιάστατη Τυποποιημένη Κανονική Κατανομή με  $\rho = 0$ . Για  $x_1 = -4, x_2 = 4, y_1 = -2, y_2 = 2$  ο H-όγκος είναι ο όγκος που βρίσκεται πάνω από το ορθογώνιο με κορυφές τα  $(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)$  και κάτω από την τρισδιάστατη επιφάνεια που φαίνεται στο παρακάτω σχήμα.

Σχήμα 1.1: Δισδιάστατη Τυπική Κανονική Κατανομή



**ΟΡΙΣΜΟΣ 4.1.2** (2-αύξουσα συνάρτηση) : Μια δισδιάστατη πραγματική συνάρτηση καλείται 2-αύξουσα αν  $V_H(B) \geq 0$  για όλα τα ορθογώνια παραλληλόγραμμα  $B$  των οποίων οι κορυφές ανήκουν στο  $S_1 \times S_2$ .

Να τονίσουμε σε αυτό το σημείο ότι η συνθήκη  $H(x, y) \geq 0$  είναι ικανή αλλά όχι αναγκαία συνθήκη για να έχουμε μια 2-αύξουσα συνάρτηση. Ο  $H$ -όγκος μπορεί να βγει μη αρνητικός ακόμα και αν η συνάρτηση που ολοκληρώνεται δεν είναι θετική σε όλο το πεδίο ορισμού της. Όταν όμως έχουμε συνάρτηση πυκνότητας, και επομένως, συνάρτηση κατανομής, τότε ο  $H$ -όγκος είναι μη αρνητικός. Επίσης, η μονοτονία αυτού του τύπου δεν συνεπάγεται μονοτονία κατά συντεταγμένες ούτε αντίστροφα.

**ΟΡΙΣΜΟΣ 4.1.3** (Grounded συνάρτηση): Η από κοινού συνάρτηση  $H$  με πεδίο ορισμού το  $S_1 \times S_2$  ονομάζεται grounded, αν  $H(x, b) = 0 = H(a, y)$ , όπου  $a, b$  είναι τα ελάχιστα στοιχεία του συνόλου  $S_1, S_2$ , αντίστοιχα.

Προφανώς, αν είμαστε στη συνεχή περίπτωση, το παραπάνω αποτέλεσμα ισχύει καθώς έχουμε:

$$H(x, y_{\min}) = P(X \leq x, Y \leq y_{\min}) = P(X \leq x, Y = y_{\min}) = 0$$

$$H(x_{\min}, y) = P(X \leq x_{\min}, Y \leq y) = P(X = x_{\min}, Y \leq y) = 0$$

Παρόλα αυτά δεν ισχύει πάντα για τη διακριτή περίπτωση για αυτό και δόθηκε ο παραπάνω ορισμός.

Θεωρούμε ότι η (αθροιστική) συνάρτηση κατανομής  $F$  είναι μια συνάρτηση με πεδίο ορισμού το  $\bar{R}$ , η οποία είναι αύξουσα και έχει τις παραπάνω οριακές ιδιότητες:

$$F(\infty) = \lim_{t \rightarrow \infty} F(t) = 1 \text{ και } F(-\infty) = \lim_{t \rightarrow -\infty} F(t) = 0$$

Επίσης, η από κοινού (αθροιστική) συνάρτηση κατανομής  $H$  είναι μια συνάρτηση με πεδίο ορισμού το  $\bar{R}^2$ , η οποία είναι 2-αύξουσα και έχει τις παρακάτω ιδιότητες :

$$H(x, -\infty) = \lim_{y \rightarrow -\infty} H(x, y) = 0$$

$$H(-\infty, y) = \lim_{x \rightarrow -\infty} H(x, y) = 0 \quad \text{και} \quad H(\infty, \infty) = \lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} H(x, y) = 1$$

**ΟΡΙΣΜΟΣ 4.1.4** (Γενικευμένη Αντίστροφη): Έστω  $F$  μια συνάρτηση κατανομής. Τότε ορίζεται ως γενικευμένη αντίστροφη, (generalized inverse), της  $F$  κάθε συνάρτηση  $F^{(-1)}$  με  $Dom F = [0, 1]$  για την οποία ισχύει :

$$F^{(-1)}(t) = \inf\{x \mid F(x) \geq t\} = \sup\{x \mid F(x) \leq t\}.$$

**ΟΡΙΣΜΟΣ 4.1.5** (Subcopulas και ιδιότητές τους) : Ένα δισδιάστατο Subcopula (2-Subcopula) είναι μια συνάρτηση  $C'$  η οποία έχει τις παρακάτω ιδιότητες:

1.  $Dom C' = S_1 \times S_2$ , όπου  $S_1, S_2$  είναι υποσύνολα του  $I = [0, 1]$  τα οποία περιέχουν το 0 και το 1.
2. Η  $C'$  είναι 2-αύξουσα και grounded.
3. Για κάθε  $u$  στο  $S_1$  και για κάθε  $v$  στο  $S_2$  ισχύει

$$C'(u, 1) = u \quad \text{και} \quad C'(1, v) = v$$

Το δισδιάστατο Copula είναι μια γενίκευση του 2-Subcopula με τη διαφορά ότι το πεδίο ορισμού είναι το  $I^2$ . Έτσι, έχουμε:

**ΟΡΙΣΜΟΣ 4.1.6** (Copulas και ιδιότητές τους) : Ένα δισδιάστατο Copula (2-Copula) είναι μια συνάρτηση  $C : I^2 \rightarrow I$  με τις εξής ιδιότητες:

1. Για κάθε  $u, v \in I$  ισχύουν

$$C(u, 0) = 0 = C(0, v) \quad \text{και} \quad C(u, 1) = u, C(1, v) = v$$

2. Για κάθε  $u_1, u_2, v_1, v_2 \in I$  τέτοια ώστε  $u_1 \leq u_2$  και  $v_1 \leq v_2$  ισχύει

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Με απλά λόγια η  $C$  είναι μια συνάρτηση που παίρνει τιμές στο  $[0, 1]$ . Επίσης, είναι μια grounded συνάρτηση καθώς για την ελάχιστη τιμή της μιας περιθώριας και για όλες τις τιμές της άλλης η από κοινού μηδενίζεται. Η ελάχιστη τιμή της πρώτης επιτυγχάνεται για  $u=0$  και αντίστοιχα της δεύτερης για  $v=0$ . Επιπρόσθετα, η  $C$  είναι 2-αύξουσα καθώς ο  $C$ -όγκος είναι μεγαλύτερος ή ίσος από το 0. Αν η  $C(u, v)$  είναι από κοινού συνάρτηση κατανομής μιας δισδιάστατης τυχαίας μεταβλητής, τότε η ιδιότητα  $C(u, 1) = u, C(1, v) = v$ , προκύπτει άμεσα όταν τα  $u, v$  είναι οι περιθώριες συναρτήσεις κατανομής.

Για παράδειγμα:

$$C(u,1) = \lim_{v \rightarrow 1} P(U \leq u, V \leq v) = P(U \leq u, V \leq 1) = P(U \leq u) = u$$

Και ουσιαστικά εξασφαλίζει την ύπαρξη των περιθωρίων. Βέβαια το  $C$  δεν ορίστηκε σαν από κοινού συνάρτηση κατανομής, αλλά αυτός είναι ο επιθυμητός στόχος.

Ας δούμε όμως ένα παράδειγμα:

**ΠΑΡΑΔΕΙΓΜΑ:** Δίνεται η συνάρτηση  $C(u,v) = \frac{uv}{u+v-uv}$  με  $(u,v) \in [0,1]^2$ . Θα αποδείξω

ότι το  $C$  είναι Copula. (Το συγκεκριμένο Copula είναι το Copula της δισδιάστατης λογιστικής κατανομής του Gumbel).

Έχουμε ότι το  $C(u,v) \geq 0$  καθώς  $uv \geq 0$  και  $u+v-uv = u(1-v)+v \geq 0$ . Επίσης  $C(u,v) \leq 1$  καθώς αν  $uv \leq u+v-uv$  τότε,

$$u+v-uv-uv = u(1-v)+v(1-u) \geq 0$$

Επιπλέον,

$$C(u,0) = 0 = C(0,v) \text{ και } C(u,1) = u, C(1,v) = v$$

Τέλος, για κάθε  $u_1, u_2, v_1, v_2 \in [0,1]$  με  $u_1 \leq u_2$  και  $v_1 \leq v_2$  ο  $C$ -όγκος υπολογίζεται από τον τύπο:

$$\begin{aligned} V_C([u_1, v_1] \times [u_2, v_2]) &= C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) = \\ &= \frac{u_2 v_2}{u_2 + v_2 - u_2 v_2} - \frac{u_2 v_1}{u_2 + v_1 - u_2 v_1} - \frac{u_1 v_2}{u_1 + v_2 - u_1 v_2} + \frac{u_1 v_1}{u_1 + v_1 - u_1 v_1} \end{aligned}$$

Και είναι μη αρνητικό, αφού οι παρανομαστές είναι θετικοί άρα και το γινόμενο τους θα είναι θετικό, ενώ ο αριθμητής που προκύπτει είναι:

$$v_1 v_2 (u_1 + u_2 - u_1 u_2) + u_1 u_2 (v_1 + v_2 - v_1 v_2) \geq 0$$

Επομένως το  $C$  είναι ένα Copula (σύμφωνα με τον παραπάνω ορισμό για το Copula και τις ιδιότητές του).

Ας δούμε όμως και τι έχουμε αν πάρουμε τα άκρα του διαστήματος:

$$\begin{aligned} V_C(I^2) &= C(1,1) - C(1,0) - C(0,1) + C(0,0) = \\ &= 1 - 0 - 0 + \lim_{\substack{u \rightarrow 0 \\ v \rightarrow 0}} \frac{uv}{u+v-uv} = 1 + \lim_{\substack{u \rightarrow 0 \\ v \rightarrow 0}} \frac{uv}{uv(\frac{1}{v} + \frac{1}{u} - 1)} = \end{aligned}$$

$$1 + \lim_{\substack{u \rightarrow 0 \\ v \rightarrow 0}} \frac{1}{\left(\frac{1}{v} + \frac{1}{u} - 1\right)} = 1 + 0 = 1$$

Η παραπάνω πιθανότητα που υπολογίστηκε στα άκρα του διαστήματος  $[0,1]$  ήταν αναμενόμενο να είναι 1 καθώς είναι ουσιαστικά η πιθανότητα που αντιστοιχεί σε ολόκληρο το δειγματικό χώρο.

Μία ακόμη ιδιότητα των Copula είναι ότι είναι ομοιόμορφα συνεχείς συναρτήσεις. Πιο συγκεκριμένα, για κάθε  $(u_1, v_1), (u_2, v_2) \in \text{Dom}C$  ισχύει:

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

(Η απόδειξη βρίσκεται στο βιβλίο Copula Method in Finance των Cherubini, Luciano και Vechiato, στο παράρτημα του 2<sup>ο</sup> Κεφαλαίου).

## **4.2 ΘΕΩΡΗΜΑ ΤΟΥ SKLAR**

Στην ενότητα αυτή θα παρουσιάσουμε το Θεώρημα του Sklar, το οποίο αποτελεί τη βάση της Θεωρίας του Copulas, και θα δούμε τι αλλαγές θα γίνουν στα Copulas όταν χρησιμοποιούμε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών.

### **Θεώρημα του Sklar:**

Έστω  $H$  η από κοινού συνάρτηση κατανομής δύο τυχαίων μεταβλητών  $X$  και  $Y$  με περιθώριες  $F, G$ , αντιστοίχως. Τότε υπάρχει Copula  $C$  τέτοιο, ώστε για όλα τα  $x, y$ ,  $x, y \in \bar{R}$  να ισχύει:

$$H(x, y) = C(F(x), G(y))$$

Αν οι  $F, G$  είναι συνεχείς τότε εξασφαλίζεται η μοναδικότητα του  $C$ , αλλιώς το  $C$  είναι μοναδικώς ορισμένο στο  $\mathfrak{R}(F) \times \mathfrak{R}(G)$ .

Αντίστροφα, αν  $C$  είναι ένα Copula και  $F, G$  είναι συναρτήσεις κατανομής, τότε η συνάρτηση  $H$ , που δίνεται από τον προηγούμενο τύπο, είναι μια από κοινού συνάρτηση κατανομής με περιθώριες  $F, G$ .

Στο θεώρημα αυτό θα εξηγήσουμε την ουσία του θεωρήματος. Το σημαντικό σε αυτό το θεώρημα είναι ότι ορίζει το Copula ως από κοινού συνάρτηση κατανομής και το συνδέει με τις περιθώριες. Επομένως, αν η επιλογή του Copula είναι σωστή, και εννοείται ότι οι περιθώριες είναι γνωστές, τότε έχουμε λύσει το πρόβλημα εύρεσης από κοινού συνάρτησης κατανομής δυο τυχαίων μεταβλητών. Επίσης, από το παραπάνω θεώρημα διαχωρίζει τη μελέτη της δομής εξάρτησης των δεδομένων από τη μελέτη των περιθωρίων, καθώς η



επιλογή των περιθώριων συναρτήσεων κατανομής δεν επηρεάζει την επιλογή του Copula. Για τη διαδικασία της επιλογής του Copula θα αναφερθούμε αργότερα , ας δούμε τώρα το αντίστροφο του θεωρήματος το οποίο λέει ότι μπορούμε να κατασκευάσουμε το Copula αν έχουμε τις περιθώριες και την από κοινού. Συγκεκριμένα , ισχύει:

**Πόρισμα:** Έστω  $X, Y$  δυο συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής  $F, G$  , από κοινού συνάρτηση κατανομής  $H$  και Copula  $C$ . Επίσης, έστω  $F^{(-1)}, G^{(-1)}$  οι γενικευμένες αντίστροφες των  $F, G$  , αντιστοίχως. Τότε, για κάθε  $(u, v) \in I^2$  ισχύει:

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v))$$

**Θεώρημα:** Έστω  $X, Y$  δύο συνεχείς τυχαίες μεταβλητές με Copula  $C_{XY}$  και  $\alpha(\cdot), \beta(\cdot)$  γνησίως μονότονες συναρτήσεις στο  $\mathcal{R}(X)$  και  $\mathcal{R}(Y)$  , αντιστοίχως. Τότε:

1. Αν  $\alpha, \beta$  είναι γνησίως αύξουσες , τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = C_{XY}(u, v)$$

2. Αν  $\alpha$  γνησίως αύξουσα,  $\beta$  γνησίως φθίνουσα τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$$

3. Αν  $\alpha$  γνησίως φθίνουσα,  $\beta$  γνησίως αύξουσα τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = v - C_{XY}(1 - u, v)$$

4. Αν  $\alpha, \beta$  γνησίως φθίνουσες τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$$

Με απλά λόγια , σε γνησίως αύξοντες μετασχηματισμούς των  $X, Y$  το Copula παραμένει αμετάβλητο αλλιώς στους άλλους γνησίως μονότονους μετασχηματισμούς το Copula προκύπτει με απλό τρόπο. Σε κάθε περίπτωση πάντως η επιλογή των  $\alpha, \beta$  δεν παίζει ρόλο, όπως φαίνεται από το παραπάνω θεώρημα.

### 4.3 SURVIVAL COPULAS

Στην ενότητα αυτή θα ασχοληθούμε με τα Copulas επιβίωσης, στα οποία θεωρούμε ότι οι τυχαίες μεταβλητές αναπαριστούν τους χρόνους ζωής των μονάδων , ανθρώπων κ.α. σε κάποιο πληθυσμό. Ας υπενθυμίσουμε, αρχικά, κάποια στοιχεία που αφορούν την περίπτωση αυτή : Όταν μια τυχαία μεταβλητή εκφράζει το χρόνο ζωής μονάδων, τότε ορίζεται στο

$[0, \infty]$ . Η συνάρτηση επιβίωσης,  $\bar{F}(x)$ , (αναφέρεται και ως συνάρτηση αξιοπιστίας), εκφράζει την πιθανότητα να επιβιώσει η μονάδα πέραν κάποιου χρονικού σημείου  $x$ . Επομένως,

$$\bar{F}(x) = P(X > x) = 1 - F(x),$$

Όπου η  $F$  είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής  $X$ . Στη διδιάστατη περίπτωση, όμως, τα πράγματα διαφέρουν λίγο.

Για τις ανάγκες των Copulas θα θεωρήσουμε αρχικά ότι οι χρόνοι ζωής μπορούν να πάρουν τιμές στο  $\bar{R}$ . Έτσι, για ένα ζεύγος τυχαίων μεταβλητών,  $(X, Y)$ , με από κοινού συνάρτηση κατανομής  $H$ , η από κοινού συνάρτηση επιβίωσης ορίζεται ως

$$\bar{H}(x, y) = P(X > x, Y > y)$$

Και εκφράζει την πιθανότητα και οι δυο μονάδες να έχουν επιβιώσει πέραν κάποιων χρονικών σημείων, έστω  $x$  και  $y$  αντίστοιχα. Οι περιθώριες της  $\bar{H}$  είναι οι

$$\bar{F}(x) = \bar{H}(x, -\infty) \text{ και } \bar{G}(y) = \bar{H}(-\infty, y)$$

Και θέτουμε λοιπόν ένα ερώτημα: Αφού το Θεώρημα του Sklar συνδέει την από κοινού συνάρτηση κατανομής με τις περιθώριες της μέσω του Copula, μήπως συμβαίνει κάτι αντίστοιχο και στην περίπτωση των συναρτήσεων επιβίωσης; Ας το υπολογίσουμε:

$$\begin{aligned} \bar{H}(x, y) &= 1 - F(x) - G(y) + H(x, y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)) \end{aligned}$$

Επομένως, αν ορίσουμε τη συνάρτηση  $\hat{C}: I^2 \rightarrow I$  με τύπο:

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

Θα έχουμε:

$$\bar{H}(x, y) = C(\bar{F}(x), \bar{G}(y))$$

Αποδεικνύεται ότι το  $\hat{C}$  είναι Copula, και συγκεκριμένα είναι το Copula επιβίωσης. Παρατηρούμε από την παραπάνω σχέση ότι συνδέει την από κοινού συνάρτηση επιβίωσης με τις περιθώριες με τρόπο ανάλογο με το Θεώρημα του Sklar. Θα πρέπει όμως να τονίσουμε το εξής: το Copula επιβίωσης  $\hat{C}$  δεν είναι η από κοινού συνάρτηση επιβίωσης  $\bar{C}$  δύο τυχαίων μεταβλητών που κατανομούνται ομοιόμορφα στο  $(0, 1)$  και με από κοινού συνάρτηση κατανομής το Copula  $C$ . Και αυτό γιατί:

$$\bar{C}(u, v) = P(U > u, V > v) = 1 - u - v + C(u, v) = \hat{C}(1 - u, 1 - v)$$

Τελειώνοντας την ενότητα αυτή, θα αναφέρουμε ότι υπάρχουν περιπτώσεις στις οποίες το Copula επιβίωσης έχει πιο απλό τύπο και άρα πιο εύχρηστο σε σχέση με την από κοινού συνάρτηση κατανομής και προτιμάται.

## 4.4 ΠΟΛΥΔΙΑΣΤΑΤΑ COPULAS

Στην ενότητα αυτή θα παρουσιάσουμε την επέκταση των αποτελεσμάτων που δόθηκαν στις προηγούμενες παραγράφους, για την περίπτωση που έχουμε διάσταση  $n \geq 3$ .

**ΟΡΙΣΜΟΣ 4.4.1** (H-όγκος): Έστω  $S_1, S_2, \dots, S_n$  μη κενά υποσύνολα του  $\bar{R}$  και  $H$  μια  $n$ -διάστατη πραγματική συνάρτηση με πεδίο ορισμού το  $S_1 \times S_2 \times \dots \times S_n$  (δηλαδή το πεδίο ορισμού είναι υποσύνολο του εκτεταμένου πραγματικού  $n$ -χώρου  $\bar{R}^n$ ). Έστω ακόμη  $B = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \dots \times [\alpha_n, \beta_n]$  με  $\alpha \leq \beta$  ένα  $n$ -διάστατο παραλληλεπίπεδο (n-box) με τις κορυφές του στο  $DomH$ . Τότε, ο  $H$ -όγκος του  $B$  δίνεται από τον τύπο:

$$V_H(B) = \sum \text{sgn}(c) H(c)$$

Όπου το άθροισμα υπολογίζεται για όλες τις κορυφές  $c$  του  $B$  και το  $\text{sgn}(c)$  υπολογίζεται ως εξής:

$$\text{sgn}(c) = \begin{cases} 1 & \text{αν } c_k = a_k \text{ για } k \text{ αρτιο} \\ -1 & \text{αν } c_k = a_k \text{ για } k \text{ περιττο} \end{cases}$$

Εναλλακτικά, ο  $H$ -όγκος ενός  $n$ -διάστατου παραλληλεπιπέδου

$B = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \dots \times [\alpha_n, \beta_n]$  είναι η  $n$ -οστής τάξης διαφορά της  $H$  πάνω στο  $B$ :

$$V_H(B) = \Delta_{\alpha}^{\beta} H(t) = \Delta_{\alpha_n}^{\beta_n} \Delta_{\alpha_{n-1}}^{\beta_{n-1}} \dots \Delta_{\alpha_1}^{\beta_1} H(t)$$

Ο τελεστής διαφορών  $\Delta$  υπολογίζεται ως εξής:

$$\Delta_{\alpha}^b F(x) = F(b) - F(a)$$

Στις δύο διαστάσεις :

$$\Delta_{\alpha_2}^{\beta_2} \Delta_{\alpha_1}^{\beta_1} H(x, y) = \Delta_{\alpha_2}^{\beta_2} (H(b_1, y) - H(a_1, y)) = H(b_1, b_2) - H(a_1, b_2) - H(b_1, a_2) + H(a_1, a_2)$$

Στις τρεις διαστάσεις:

$$\begin{aligned} \Delta_{a_3}^{b_3} \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(x, y) &= \Delta_{a_3}^{b_3} \Delta_{a_2}^{b_2} (H(b_1, y, z) - H(a_1, y, z)) = \\ &= \Delta_{a_3}^{b_3} (H(b_1, b_2, z) - H(a_1, b_2, z) - H(b_1, a_2, z) + H(a_1, a_2, z)) = \\ &= H(b_1, b_2, b_3) - H(a_1, b_2, b_3) - H(b_1, a_2, b_3) + H(a_1, a_2, b_3) - H(b_1, b_2, a_3) + \\ &\quad + H(a_1, b_2, a_3) + H(b_1, a_2, a_3) - H(a_1, a_2, a_3). \end{aligned}$$

Κ.ο.κ

Με λίγα λόγια , ο  $H$ -όγκος ισούται με την εξής πιθανότητα:

$$P(a_1 \leq X_1 \leq \beta_1, a_2 \leq X_2 \leq \beta_2, \dots, a_n \leq X_n \leq \beta_n) = \int_{a_n}^{\beta_n} \dots \int_{a_2}^{\beta_2} \int_{a_1}^{\beta_1} h(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Την οποία και προφανώς δεν μπορούμε να την δούμε οπτικά.

**ΟΡΙΣΜΟΣ 4.4.2.** ( $n$ - αύξουσα συνάρτηση): Μια  $n$ -διάστατη πραγματική συνάρτηση  $H$  ονομάζεται  $n$ -αύξουσα αν  $V_H(B) \geq 0$  για όλα τα  $n$ -διάστατα παραλληλεπίπεδα  $B$  των οποίων οι κορυφές ανήκουν στο  $DomH$ .

**ΟΡΙΣΜΟΣ 4.4.3** ( $n$ -grounded συνάρτηση): Μια  $n$ -διάστατη πραγματική συνάρτηση  $H$  ονομάζεται grounded όταν  $H(t)=0$  για όλα τα  $t \in DomH$  τέτοια, ώστε ένα τουλάχιστον από τα  $t_k$  ισούται με την ελάχιστη τιμή του.

Να σημειώσουμε σε αυτό το σημείο ότι για να βρούμε τις μονοδιάστατες περιθώριες βάζουμε σε όλες τις άλλες μεταβλητές τη μέγιστη τιμή τους , ενώ για τις  $k$ -διάστατες περιθώριες βάζουμε τη μέγιστη τιμή στις υπόλοιπες  $n-k$ . Προφανώς, ισχύουν τα παραπάνω αν τα σύνολα  $S_1, S_2, \dots, S_n$  είναι μη κενά και έχουν ελάχιστο και μέγιστο στοιχείο.

**ΟΡΙΣΜΟΣ 4.4.4** ( $n$ -διάστατο Copula): Ένα  $n$ -διάστατο Copula , ( $n$ -Copula) , είναι μια συνάρτηση  $C: I^n \rightarrow I$  με τις εξής ιδιότητες:

1. Για κάθε  $u \in I^n$ ,

$$C(u) = 0$$

Αν τουλάχιστον μια συνιστώσα του  $u$  είναι μηδενική (δηλαδή η  $C$  είναι grounded συνάρτηση ) και

2. Αν όλες οι συνιστώσες του  $u$  είναι μονάδες εκτός της  $u_k$ , τότε

$$C(u) = u_k$$

δηλαδή ορίζονται οι μονοδιάστατες περιθώριες.

3. Για κάθε  $a, b \in I^n$  με  $a \leq b$  ισχύει

$$V_C([a, b]) \geq 0$$

Δηλαδή είναι  $n$ -αύξουσα.

Αποδεικνύεται ότι για οποιαδήποτε  $n$ -Copula ,  $n \geq 3$ , κάθε  $k$ -διάστατη περιθώρια της  $C$  είναι μια  $k$ -Copula,  $2 \leq k \leq n$ .

Μια ακόμα ιδιότητα του  $n$ -Copula είναι ότι είναι ομοιόμορφα συνεχής συνάρτηση καθώς για κάθε  $u, v \in \text{Dom}C$  ισχύει:

$$|C'(v) - C'(u)| \leq \sum_{k=1}^n |v_k - u_k|$$

**ΟΡΙΣΜΟΣ 4.4.5** ( $n$ -διάστατη συνάρτηση κατανομής): Μια  $n$ -διάστατη συνάρτηση κατανομής είναι μια συνάρτηση  $H$  με πεδίο ορισμού το  $\bar{R}^n$  έτσι ώστε:

1. Η  $H$  είναι  $n$ -αύξουσα.
2.  $H(t) = 0$  για όλα τα  $t \in \bar{R}^n$  ώστε  $t_k = -\infty$  για τουλάχιστον ένα  $k$  και  $H(\infty, \infty, \dots, \infty) = 1$

**ΘΕΩΡΗΜΑ (Sklar):** Έστω  $H$  είναι μια  $n$ -διάστατη συνάρτηση κατανομής με περιθώριες  $F_1, F_2, \dots, F_n$ . Τότε, υπάρχει ένα  $n$ -Copula τέτοιο, ώστε για κάθε  $x \in \bar{R}^n$ , ισχύει:

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$$

Αν οι  $F_1, F_2, \dots, F_n$  είναι όλες συνεχείς, τότε το  $C$  είναι μοναδικό, αλλιώς το  $C$  είναι μοναδικώς ορισμένο στο  $\mathfrak{R}(F_1) \times \mathfrak{R}(F_2) \times \dots \times \mathfrak{R}(F_n)$ .

Αντίστροφα, αν  $C$  είναι ένα  $n$ -Copula και  $F_1, F_2, \dots, F_n$  είναι συναρτήσεις κατανομής, τότε η  $H$  όπως ορίζεται παραπάνω είναι μια  $n$ -διάστατη συνάρτηση κατανομής με περιθώριες τις  $F_1, F_2, \dots, F_n$ . (Το θεώρημα είναι αντίστοιχο και όταν έχουμε  $n$  τυχαίες μεταβλητές).

**ΠΟΡΙΣΜΑ:** Έστω  $H, F_1, F_2, \dots, F_n, C$  όπως ορίστηκαν παραπάνω και  $F_1^{(-1)}, F_2^{(-1)}, \dots, F_n^{(-1)}$ , οι γενικευμένες αντίστροφες συναρτήσεις των  $F_1, F_2, \dots, F_n$  αντίστοιχα. Τότε, για κάθε  $u \in I^n$ ,

$$C(u_1, u_2, \dots, u_n) = H(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \dots, F_n^{(-1)}(u_n))$$

### Φράγματα Frechet-Hoeffding για n-Copulas

Όπως είδαμε στις δυο διαστάσεις υπάρχουν φράγματα, τα οποία αποδεικνύεται ότι είναι και τα ίδια Copula. Η επέκτασή τους στις  $n$ -διαστάσεις είναι:

$$\text{Το κάτω όριο : } W^n(u) = \max(u_1 + u_2 + \dots + u_n - n + 1, 0)$$

Το άνω όριο:  $M^n(u) = \min(u_1, u_2, \dots, u_n)$

Επίσης το Copula –γινόμενο δίνεται από τον τύπο :

$$\Pi^n(u) = u_1 u_2 \cdots u_n$$

Να τονίσουμε σε αυτό το σημείο ότι στις n-διαστάσεις το κάτω όριο δεν είναι Copula για διάσταση  $n \geq 3$ . Παρ' όλα αυτά χρησιμοποιείται η αριστερή ανισότητα , καθώς αποδεικνύεται ότι για διάσταση  $n \geq 3$ , υπάρχει Copula C τέτοιο ώστε:

$$C(u) = W^n(u), \text{ για κάθε } u \in I^n$$

Επίσης, για το Copula Γινόμενο και το άνω φράγμα ισχύει το εξής:

**Θεώρημα:** Για  $n \geq 2$  και για  $X_1, X_2, \dots, X_n$  συνεχείς τυχαίες μεταβλητές, ισχύει:

1. Οι  $X_1, X_2, \dots, X_n$  είναι ανεξάρτητες αν και μόνο αν το Copula είναι αυτού του γινομένου και
2. Κάθε μια ξεχωριστά από τις τυχαίες μεταβλητές  $X_1, X_2, \dots, X_n$  είναι μια γνησίως αύξουσα συνάρτηση οποιασδήποτε άλλης αν και μόνο αν το Copula είναι ίσο με το άνω φράγμα.

Όταν είμαστε στις δύο διαστάσεις, αν επιτυγχάνεται το άνω φράγμα, θα λέμε ότι οι τυχαίες μεταβλητές έχουν τέλεια θετική συσχέτιση ενώ όταν επιτυγχάνεται το κάτω φράγμα θα λέμε ότι έχουν τέλεια αρνητική συσχέτιση. Στις n-διαστάσεις, όπου  $n \geq 3$ , όπως αναφέραμε, δεν ορίζεται η τέλεια αρνητική συσχέτιση.

## **4.5 The Elliptical Copulas**

Η κλάση των ελλειπτικών συζεύξεων παρουσιάζει χρήσιμα παραδείγματα πολυδιάστατων κατανομών. Σε αυτές εμφανίζονται πολλές από τις ιδιότητες της πολυδιάστατης κανονικής κατανομής. Τα Elliptical Copulas μας βοηθάνε στην μοντελοποίηση πολυδιάστατων ακραίων γεγονότων και στη διαμόρφωση μη παραμετρικών μέτρων εξάρτησης μεταξύ τυχαίων μεταβλητών. Οι ελλειπτικές συζεύξεις είναι οι συζεύξεις των ελλειπτικών κατανομών. Συνεπώς, η προσομοίωση των ελλειπτικών συζεύξεων είναι εξίσου εύκολη με την προσομοίωση των ελλειπτικών κατανομών.

Γενικά μιλώντας, όταν βλέπουμε μια ελλειπτική σύζευξη ή μια από κοινού ελλειπτική κατανομή από πάνω, τα περιγράμματα αυτής της κατανομής έχουν ελλειψοειδή μορφή. Μια από τις ιδιότητες των Elliptical Copulas είναι η ακτινική συμμετρία, ενώ το κύριο πλεονέκτημά τους είναι ότι μπορούμε εύκολα να πάρουμε παραδείγματα από αυτά. Από την άλλη, ένα σημαντικό μειονέκτημά τους, είναι ότι δεν μπορούν εύκολα να γραφτούν σε κλειστές φόρμες.

### **4.5.1 The Gaussian Copula**

Η κανονική σύζευξη ή Gaussian Copula είναι η σύζευξη της πολυδιάστατης κανονικής κατανομής. Έστω R συμμετρικός, θετικά ορισμένος πίνακας με  $\text{diag}R=1$  και  $\Phi_R$  η

τυποποιημένη πολυδιάστατη κανονική κατανομή με πίνακα συσχετίσεων  $R$ .

Η πολυδιάστατη κανονική σύζευξη ορίζεται ως εξής:

$$C_R^{Ga}(u_1, \dots, u_n) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

Για  $n=2$  η κανονική σύζευξη γράφεται ως εξής:

$$C_R^{Ga}(u, v) = \Phi_R^2(\Phi^{-1}(u), \dots, \Phi^{-1}(v))$$

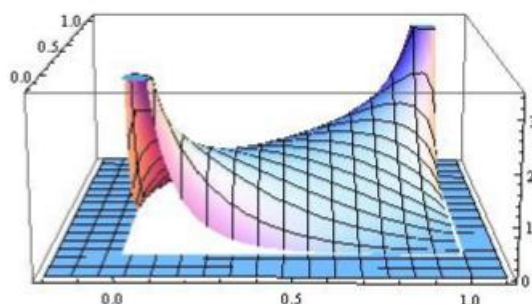
Έχουμε:

$$C_R^{Ga}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-R_{12}^2)^{1/2}} \exp\left\{-\frac{s^2 - 2R_{12}st + t^2}{2(1-R_{12}^2)}\right\} ds dt$$

Στην περίπτωση που οι δυο τυχαίες μεταβλητές  $u$  και  $v$  είναι ασυσχέτιστες, η κανονική σύζευξη παίρνει την παρακάτω μορφή:

$$\begin{aligned} C_0^{Ga}(u, v) &= \int_{-\infty}^{\Phi^{-1}(u)} f_1(s) ds \int_{-\infty}^{\Phi^{-1}(v)} f_2(t) dt \\ &= uv \\ &= \Pi(u, v) \end{aligned}$$

Σχήμα 1.2: Η συνάρτηση πυκνότητας ενός Gaussian Copula με περιθώριες ομοιόμορφες στο  $(0,1)$



#### **4.5.2 T-Student Copula**

Η T-Student σύζευξη είναι η σύζευξη της t-student κατανομής. Έστω  $X$  ένα διάνυσμα με  $n$ -διάστατη t-student κατανομή με  $\nu$  βαθμούς ελευθερίας, διάνυσμα μέσης τιμής  $\mu$  (για  $\nu > 1$ ) και

πίνακα συνδιασπορών  $\frac{V}{\nu-2} \Sigma$  (για  $\nu > 2$ ). Έχουμε:

$$X^d = \mu + \frac{\sqrt{v}}{\sqrt{S}} Z$$

Όπου  $\mu \in R^n$ ,  $S \sim \chi_v^2$  και το τυχαίο διάνυσμα  $Z \sim N(0, \Sigma)$ . Τα  $\mu, S, Z$  είναι ανεξάρτητα.

Η σύζευξη του διανύσματος  $X$  είναι η t-student σύζευξη με  $v$  βαθμούς ελευθερίας. Το t-student copula γράφεται αναλυτικά ως εξής:

$$C_{v,R}^t(u) = t_{v,R}^n(t_v^{-1}(u_1), \dots, t_v^{-1}(u_n))$$

Όπου :  $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$ , για  $i, j \in \{1, \dots, n\}$

: Η  $t_{v,R}^n$  είναι η πολυδιάστατη συνάρτηση κατανομής του τυχαίου διανύσματος  $\frac{\sqrt{v}Y}{\sqrt{S}}$ ,

$$Y \sim N_n(0, \Sigma)$$

: Οι τυχαίες μεταβλητές  $S \sim \chi_v^2$  και το τυχαίο διάνυσμα  $Y$  είναι ανεξάρτητα.

: Οι  $t_v$  είναι οι περιθώριες συναρτήσεις της  $t_{v,R}^n$ .

Για  $n=2$ , η t-student σύζευξη παίρνει την παρακάτω μορφή:

$$C_{v,R}^t(u, v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-R_{12}^2)^{1/2}} \left\{ 1 + \frac{s^2 - 2R_{12}st + t^2}{v(1-R_{12}^2)} \right\}^{-(v+2)/2} ds dt$$

Όπου  $R_{12}$  είναι ο γραμμικός συντελεστής συσχέτισης της δισδιάστατης t-student κατανομής με  $v$  βαθμούς ελευθερίας, αν  $v > 2$ .

Παρατηρούμε ότι ο τύπος της t-student σύζευξης μοιάζει πολύ με αυτόν της Gaussian αλλά έχει μια επιπλέον παράμετρο, το  $v$ . Η παράμετρος αυτή ελέγχει την εξάρτηση στα άκρα της κατανομής. Όταν η παράμετρος  $v$  είναι μικρή τότε έχουμε εξάρτηση στα άκρα και ανεξαρτησία στο κέντρο της κατανομής. Αντίθετα, όταν το  $v$  είναι μεγάλο τότε έχουμε εξάρτηση στο κέντρο και η t-student σύζευξη συμπεριφέρεται ανάλογα με την Gaussian σύζευξη.

### ΘΕΩΡΗΜΑ

Έστω  $X$  ένα διάνυσμα τυχαίων μεταβλητών που ακολουθούν μια ελλειπτική κατανομή και έχουν πίνακα συνδιασπορών  $R$  τότε οι συντελεστές συσχέτισης του Kendall και του Spearman δίνονται από τους τύπους:

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(R_{ij})$$



$$\rho(X_i, X_j) = \frac{6}{\pi} \arcsin\left(\frac{R_{ij}}{2}\right)$$

Τα όρια  $\lambda_U$  και  $\lambda_L$  της άνω και κάτω εξάρτησης ουράς διαμορφώνονται ως εξής:

$$\text{Έχουμε ότι } P\{V \leq v/U = u\} = \frac{\partial C(u, v)}{\partial u} \text{ και } P\{V > v/U = u\} = 1 - \frac{\partial C(u, v)}{\partial u}.$$

Τότε:

$$\begin{aligned} \lambda_U &= -\lim_{u \rightarrow 1^-} \left[ -2 + \frac{\partial C(u, v)}{\partial u} \Big|_{u=v} + \frac{\partial C(u, v)}{\partial v} \Big|_{v=u} \right] \\ \lambda_U &= -\lim_{u \rightarrow 1^-} [P\{V \geq v/U = u\} + P\{U \geq v/V = u\}] \end{aligned}$$

Για  $C(u, v) = C(v, u)$ , έχουμε

$$\lambda_U = -\lim_{u \rightarrow 1^-} P\{V \geq u/U = u\}$$

Εάν τώρα ορίσουμε  $x = F^{-1}(u)$  και  $y = F^{-1}(v)$  όπου  $x, y \in R$  με  $F$  και  $G$  τις περιθώριες συναρτήσεις των  $X$  και  $Y$  αντίστοιχα. Τότε, το παραπάνω όριο γράφεται ως εξής:

$$\lambda_U = -2 \lim_{x \rightarrow +\infty} [P\{F^{-1}(V) \geq x/F^{-1}(U) = x\}] = -2 \lim_{x \rightarrow +\infty} P\{X > x/Y = x\}$$

- Αν  $F = \Phi$ , η κανονική κατανομή και γνωρίζοντας ότι για να την δισδιάστατη κανονική κατανομή  $Y/X = x \sim N(\rho x, 1 - \rho^2)$ , γράφουμε:

$$\lambda_U = -2 \lim_{x \rightarrow +\infty} \left[ 1 - \Phi\left(\frac{x - \rho x}{\sqrt{1 - \rho^2}}\right) \right] = -2 \lim_{x \rightarrow +\infty} \left[ 1 - \Phi\left(x \frac{\sqrt{1 - \rho}}{\sqrt{1 + \rho}}\right) \right]$$

- Όταν  $\rho < 1$ , στην κανονική σύζευξη δεν υπάρχει το  $\lambda_U$ .
- Λόγω συμμετρίας, τα παραπάνω ισχύουν και για το όριο  $\lambda_L$ .

- Αν  $F = t_\nu$ , η student κατανομή με  $\nu$  βαθμούς ελευθερίας. Χρησιμοποιώντας την σχέση (Dematra and McNeil (2004))

$$P\{X > x/X = x\} = 1 - t_{\nu+1} \left[ \left( \frac{\nu + x^2}{\nu + 1} \right)^{-1/2} \frac{x - rx}{\sqrt{1 - r^2}} \right], \text{ έχουμε:}$$

$$\lambda_U = 2t_{\nu+1} \left[ \frac{(-(\nu + 1)(1 - r))^{1/2}}{\sqrt{1 + r}} \right]$$

$$\text{Επίσης έχουμε: } \lambda_L = 2t_{\nu+1} \left[ \frac{(-(\nu + 1)(1 - r))^{1/2}}{\sqrt{1 + r}} \right]$$

Παρατηρούμε ότι οι συντελεστές εξάρτησης  $\lambda_U$  και  $\lambda_L$  είναι ίδιοι λόγω ακτινικής

συμμετρίας και είναι συναρτήσεις των βαθμών ελευθερίας και της γραμμικής συσχέτισης.

## 4.6 Archimedean Copulas

Είδαμε προηγουμένως την οικογένεια των ελλειπτικών συζεύξεων και αναφέραμε ότι η προσομοίωσή τους είναι εξίσου εύκολη με την προσομοίωση των ελλειπτικών κατανομών. Παρόλα αυτά όμως τα Elliptical Copulas, παρουσιάζουν κάποια μειονεκτήματα. Για παράδειγμα, δεν μπορούν να παρουσιαστούν σε κλειστές φόρμες και περιορίζονται να έχουν ακτινική συμμετρία. Σε πολλές χρηματοοικονομικές και ασφαλιστικές εφαρμογές παρουσιάζονται ισχυρότερες εξαρτήσεις μεταξύ μεγάλων απωλειών απ' ό,τι μεταξύ μεγάλων κερδών. Τέτοιες ασυμμετρίες δεν μπορούν να μελετηθούν με την βοήθεια των ελλειπτικών συζεύξεων.

Στην παράγραφο αυτή θα παρουσιάσουμε μια άλλη οικογένεια συζεύξεων που καλούνται Archimedean Copulas. Σε αντίθεση με τα Elliptical Copulas τα Archimedean μπορούν να εκφραστούν σε κλειστές φόρμες (τύπους). Παρουσιάζουν όμως και αυτά κάποια μειονεκτήματα. Επειδή οι Αρχιμήδειες συζεύξεις δεν προέρχονται από τις πολυδιάστατες συναρτήσεις κατανομών που παίρνουμε χρησιμοποιώντας το θεώρημα του Sklar, για να επιβεβαιώσουμε την πολυδιάστατη μορφή των δισδιάστατων Archimedean Copulas χρησιμοποιούμε συγκεκριμένες τεχνικές. Επίσης, στις αρχιμήδειες συζεύξεις παρουσιάζεται έλλειψη ελευθέρων παραμέτρων στους πίνακες συσχετίσεων γιατί κάποιες τιμές εισάγονται εσκεμμένα ως ίσες.

Παρόλα αυτά η χρησιμότητα και οι εφαρμογές των Archimedean Copulas παρουσιάζουν μεγάλο ενδιαφέρον λόγω της ευκολίας με την οποία κατασκευάζονται, των πολλών οικογενειών συζεύξεων που προέρχονται από τα Archimedean Copulas και λόγω των ιδιοτήτων αυτών των οικογενειών.

**ΟΡΙΣΜΟΣ** (Archimedean Copulas): Τα Copulas της μορφής:

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

ονομάζονται Archimedean Copulas και η συνάρτηση  $\varphi$  είναι ο χαρακτηριστικός γεννήτοράς τους. Αν  $\varphi(0) = \infty$ , τότε έχουμε Archimedean Copula με αυστηρό χαρακτηριστικό γεννήτορα  $\varphi$ .

**Παράδειγμα:** Τα Archimedean Copulas τα οποία παράγονται από τους χαρακτηριστικούς γεννήτορες  $\varphi(t)=1-t$  και  $\varphi(t)=-\log t$  για κάθε  $t \in [0,1]$  είναι το κάτω φράγμα και το Copula ανεξαρτησίας αντίστοιχα.

Πράγματι, έχουμε:

- 1) Για  $\varphi(t)=1-t$ , η ψευδοαντίστροφη είναι:

$$\phi^{[-1]}(t) = \begin{cases} 1-t, & \text{αν } 0 \leq t \leq 1 \\ 0, & \text{αν } t > 1 \end{cases} = \max(1-t, 0)$$

Και επομένως το Copula που παράγεται είναι:

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)) = \max(1 - (1-u) - (1-v), 0) = \max(u + v - 1, 0) = W(u, v)$$

2) Για  $\phi(t) = -\log t$ , η ψευδοαντίστροφη είναι:

$$\phi^{[-1]}(t) = \phi^{-1}(t) = \exp(-t).$$

Αυτό ισχύει γιατί  $\phi(0) = \infty$  και άρα ο χαρακτηριστικός γεννήτορας είναι αυστηρός και το Copula που παράγεται από αυτόν είναι:

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)) = \exp(-(-\log u - \log v)) = uv = \prod(u, v)$$

Μία εύλογη απορία θα ήταν, ποιος είναι ο λόγος που ονομάστηκαν Archimedean τα συγκεκριμένα Copulas. Σύμφωνα με την Αρχιμήδεια Ιδιότητα για τους πραγματικούς αριθμούς έχουμε ότι για  $a, b \in \mathbb{R}$  όπου  $a, b > 0$  υπάρχει  $n \in \mathbb{Z}$  ώστε  $na > b$ . Κάτι αντίστοιχο ισχύει και για τα συγκεκριμένα Copulas.

**ΟΡΙΣΜΟΣ:** Έστω  $\phi: [0, 1] \rightarrow [0, \infty)$  μια συνεχής και αυστηρά φθίνουσα συνάρτηση με  $\phi(1) = 0$ . Η ψευδο-αντίστροφη συνάρτηση της  $\phi$  είναι η συνάρτηση  $\phi^{[-1]}: [0, \infty) \rightarrow [0, 1]$  με τύπο:

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0) \\ 0 & \phi(0) \leq t \leq \infty \end{cases}$$

Σημειώνουμε ότι η  $\phi^{[-1]}$  είναι συνεχής και φθίνουσα στο διάστημα  $[0, \infty)$  και αυστηρά φθίνουσα στο  $[0, \phi(0)]$ . Επιπλέον,  $\phi^{[-1]}(\phi(u)) = u$  στο διάστημα  $[0, 1]$  και

$$\phi^{[-1]}(\phi(t)) = \begin{cases} t & , 0 \leq t \leq \phi(0) \\ \phi(0) & , \phi(0) \leq t \leq \infty \end{cases}$$

Τέλος, εάν  $\phi(0) = \infty$  τότε έχουμε ότι  $\phi^{[-1]} = \phi^{-1}$ .

**ΘΕΩΡΗΜΑ:** Έστω συνάρτηση  $\phi: [0, 1] \rightarrow [0, \infty)$  συνεχής και αυστηρά φθίνουσα τέτοια ώστε  $\phi$  και έστω  $\phi^{[-1]}$  η ψευδο-αντίστροφη της  $\phi$ . Τότε η συνάρτηση  $C: [0, 1]^2 \rightarrow [0, 1]$  που δίνεται από τον τύπο:

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v))$$

Είναι μια σύζευξη αν και μόνο αν η συνάρτηση  $\phi$  είναι κυρτή.

Οι συζεύξεις αυτές καλούνται Archimedean Copulas. Η συνάρτηση  $\phi$  καλείται γεννήτορας της σύζευξης. Εάν  $\phi(0) = \infty$  τότε η  $\phi$  είναι ένας αυστηρός γεννήτορας της σύζευξης. Σε αυτή την περίπτωση,  $\phi^{[-1]} = \phi^{-1}$  και η  $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$  καλείται αυστηρή Archimedean σύζευξη.

**Παράδειγμα:** Έστω συνάρτηση  $\phi(t) = 1 - t$  για  $t \in [0, 1]$ . Τότε η ψευδο-αντίστροφη της  $\phi$ ,  $\phi^{[-1]}(t) = 1 - t$  για  $t \in [0, 1]$  και  $\phi^{[-1]}(t) = 0$  για  $t > 1$ . Άρα,  $\phi^{[-1]}(t) = \max(1 - t, 0)$ . Επίσης, όπως έχουμε αναφέρει, ισχύει ότι:  $C(u, v) = \max(u + v - 1, 0) = W(u, v)$ . Συνεπώς συμπεραίνουμε ότι το διδιάστατο Frechet-Hoeffding κάτω όριο είναι ένα Archimedean Copula.

## ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΕΞΑΡΤΗΣΗ ΣΤΑ ARCHIMEDEAN COPULAS

Στη συνέχεια θα δείξουμε πως ορίζεται ο συντελεστής συσχέτισης του Kendall καθώς επίσης και οι συντελεστές εξάρτησης ουράς για κάθε μια από τις οικογένειες των Archimedean Copulas.

### ΘΕΩΡΗΜΑ

Έστω  $X$  και  $Y$  τυχαίες μεταβλητές και  $C$  μια Αρχιμήδεια Σύζευξη που «γεννιέται» από την συνάρτηση  $\phi$ . Τότε, ο συντελεστής συσχέτισης του Kendall,  $\tau$ , για τις  $X$  και  $Y$  δίνεται από τον τύπο:

$$\tau_c = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt$$

### Οικογένεια Gumbell

Θεωρούμε την γεννήτρια συνάρτηση  $\phi(t) = (-\ln t)^\theta$ , για  $\theta \geq 1$ . Έχουμε:  $\frac{\phi(t)}{\phi'(t)} = \frac{t \ln t}{\theta}$ .

Τότε ο συντελεστής συσχέτισης του Kendall παίρνει τη μορφή:

$$\tau_\theta = 1 + \int_0^1 \frac{t \ln t}{\theta} dt = 1 + \frac{4}{\theta} \left( \left[ \frac{t^2}{2} \ln t \right]_0^1 - \int_0^1 \frac{t}{2} dt \right) = 1 + \frac{4}{\theta} \left( 0 - \frac{1}{4} \right) = 1 - \frac{1}{\theta}$$

### Οικογένεια Clayton

Θεωρούμε την γεννήτρια συνάρτηση  $\phi(t) = \frac{t^{-\theta} - 1}{\theta}$ , με  $\theta \in [-1, \infty) \setminus \{0\}$ . Έχουμε:

$\frac{\phi(t)}{\phi'(t)} = \frac{t^{\theta+1} - t}{\theta}$ . Ο συντελεστής συσχέτισης του Kendall είναι:

$$\tau_{\theta} = 1 + \int_0^1 \frac{t^{\theta+1} - t}{\theta} dt = 1 + \frac{4}{\theta} \left( \frac{1}{\theta+2} - \frac{1}{2} \right) = \frac{\theta}{\theta+2}$$

## **ΚΕΦΑΛΑΙΟ 5: ΕΙΣΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ , ΠΕΡΙΓΡΑΦΙΚΗ**

### **ΣΤΑΤΙΣΤΙΚΗ**

Η εργασία έχει ως στόχο την μελέτη ασθενειών, για το έτος 2013. Μετά από ιατρική συμβουλή, επιλέχθηκαν οι παρακάτω ασθένειες, ώστε να έχει ενδιαφέροντα αποτελέσματα και μεγάλο αριθμό θνησιμότητας:

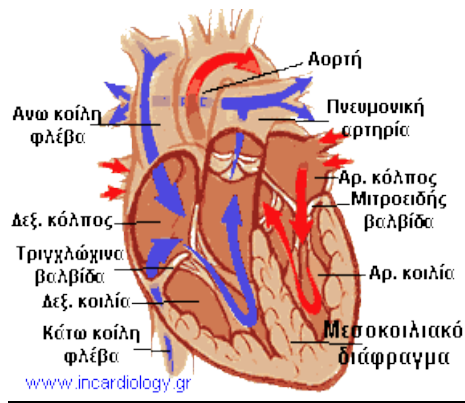
A. Ισχαιμική καρδιοπάθεια

B. Κακοήθη νεοπλασμάτα των:

- i. οργάνων του αναπνευστικού συστήματος και ενδοθωρακικών οργάνων
- ii. οργάνων του πεπτικού συστήματος και περιτοναίου
- iii. ουροποιητικών οργάνων

### **A. ΙΣΧΑΙΜΙΚΗ ΚΑΡΔΙΟΠΑΘΕΙΑ**

Ισχαιμική καρδιοπάθεια ονομάζουμε την έλλειψη ροής αίματος και οξυγόνου προς τον καρδιακό μυ. Είναι ο όρος που χρησιμοποιούμε για να χαρακτηρίσουμε τα καρδιακά προβλήματα που προκαλούνται από την στένωση των καρδιακών αρτηριών. Όταν στενεύουν οι αρτηρίες, λιγότερο αίμα και οξυγόνο φθάνουν στον καρδιακό μυ. Αυτή η πάθηση ονομάζεται πάθηση στεφανιαίας αρτηρίας ή καρδιοπάθεια. Οι συγκεκριμένες παθήσεις οδηγούν σε καρδιακή ανακοπή. Η ισχαιμική καρδιοπάθεια προκαλεί πόνο στο θώρακα και δυσφορία.



**Εικόνα 1: Ισχαιμική Καρδιοπάθεια**

Η στεφανιαία νόσος είναι επίσης γνωστή ως ισχαιμική καρδιοπάθεια, η οποία αποτελείται από μια ομάδα παθήσεων που συμπεριλαμβάνουν σταθερή στηθάγχη, ασταθή στηθάγχη, εγκεφαλικό επεισόδιο, έμφραγμα, αιφνίδιο θάνατο με ισχαιμικό χαρακτήρα κλπ.

Οι παράγοντες κινδύνου που ευνοούν την εμφάνιση της ισχαιμικής καρδιακής νόσου ορίζονται ως τροποποιούμενοι και μη τροποποιούμενοι. Η πρώτη κατηγορία περιλαμβάνει παράγοντες όπως η αρτηριακή υπέρταση, το κάπνισμα (παρόν περίπου στο 36% των περιπτώσεων), ο διαβήτης, η έλλειψη σωματικής δραστηριότητας (στο 10% των περιπτώσεων), η παχυσαρκία (περίπου στο 20% των περιπτώσεων), η δυσλιπιδαιμία, η υπερβολική κατανάλωση αλκοόλ κ.α. Στη δεύτερη κατηγορία περιλαμβάνονται παράγοντες όπως η ηλικία, το φύλο και το οικογενειακό ιστορικό.

*Οι αιτίες της ισχαιμικής καρδιακής νόσου:*

Ο κύριος μηχανισμός της νόσου, που δεν είναι ο μόνος, είναι η αθηροσκλήρωση των στεφανιαίων αρτηριών, η οποία είναι κοινή στα άτομα που παρουσιάζουν τους περισσότερους παράγοντες που αναφέρθηκαν παραπάνω.

Σε σπάνιες περιπτώσεις, η στεφανιαία νόσος μπορεί να εμφανιστεί χωρίς να υπάρχει στεφανιαία αθηροσκλήρωση. Μια τέτοια περίπτωση είναι το Σύνδρομο X, που εκδηλώνεται με τυπικό ισχαιμικό πόνο στο στήθος, χωρίς παρουσία αθηρωματικής στένωσης στις μεγαλύτερες στεφανιαίες αρτηρίες (όπως φαίνεται μετά τη διενέργεια στεφανιογραφίας). Το Σύνδρομο X πλήττει συχνότερα γυναίκες και η ακριβής του αιτιολογία είναι άγνωστη. Μια πιθανή επεξήγηση, αν και δεν έχει επιβεβαιωθεί, έχει να κάνει με τα προβλήματα της μικροκυκλοφορίας και της μικροαγγείωσης.

Παράγοντες πρόληψης της ισχαιμικής καρδιοπάθειας είναι η υγιεινή διατροφή, η τακτική σωματική δραστηριότητα, η διακοπή του καπνίσματος και η αποτελεσματική θεραπεία παραγόντων κινδύνου όπως η αρτηριακή υπέρταση, ο διαβήτης ή η δυσλιπιδαιμία. Μια σειρά από εξετάσεις, όπως το ηλεκτροκαρδιογράφημα, ο υπέρηχος καρδιάς, το τεστ κοπώσεως και η στεφανιογραφία, βοηθούν στον ορισμό της ιατρικής διάγνωσης και παρακολούθησης της νόσου.

#### Τα συμπτώματα της ισχαιμικής καρδιακής νόσου:

Τα πιο συνηθισμένα συμπτώματα είναι ο πόνος στο στήθος, ο οποίος εντοπίζεται κυρίως οπισθοστερνικά και εξαπλώνεται στα χέρια, τους ώμους, την πλάτη, το σαγόι κλπ. Ο πόνος παίρνει διάφορες μορφές, όπως καυστικός, συσφιγκτικός, βάρους, διαρκεί λίγα λεπτά και εμφανίζεται ή εντείνεται κυρίως κατά τη διάρκεια συναισθηματικής και σωματικής κόπωσης.

Οι άνδρες είναι πιο πιθανό να παρουσιάσουν απότομο, επίμονο πόνο στο στήθος και στο χέρι.

Τα συμπτώματα των γυναικών περιλαμβάνουν ανεξήγητη κόπωση, δύσπνοια, ζάλη, ναυτία και μια δυσάρεστη αίσθηση σφιξίματος στο στήθος, την πλάτη και τα χέρια.

Δυστυχώς, ένας μεγάλος αριθμός των ισχαιμικών καρδιοπαθειών δεν παρουσιάζουν σαφώς αναγνωρίσιμα συμπτώματα.

#### Οι κύριοι παράγοντες κινδύνου για την ισχαιμική καρδιοπάθεια είναι:

- Η διατροφή
- Το κάπνισμα
- Η αυξημένη αρτηριακή πίεση
- Ο σακχαρώδης διαβήτης
- Η ελλιπής άσκηση

Το κάπνισμα θεωρείται ο σημαντικότερος παράγοντας πρόκλησης ισχαιμικού καρδιακού επεισοδίου. Οι τοξικές συνέπειες του καπνού είναι σημαντικές λόγω της τοξικής βλάβης που αυτός προκαλεί σε όλο το αγγειακό σύστημα με αποτέλεσμα στυτική δυσλειτουργία, ανεπάρκεια καρδιακών αγγείων και κακή κυκλοφορία αίματος.

Τα ισχαιμικά καρδιακά επεισόδια μπορεί να είναι παροδικά ή να προέρχονται από σιωπηρή ισχαιμία.

- i. Το παροδικό ισχαιμικό επεισόδιο είναι μια προσβολή διάρκειας από λίγα δευτερόλεπτα μέχρι αρκετές ώρες. Το αίτιο είναι η μείωση της ροής του αίματος σε κάποια περιοχή της καρδιάς, η οποία δυσλειτουργεί. Η μείωση αυτή μπορεί να οφείλεται σε πτώση της αρτηριακής πίεσης (υπόταση) ή σε στένωση ενός αγγείου.

- ii. Πολλοί άνθρωποι μπορεί να έχουν υποστεί ισχαιμικά καρδιακά επεισόδια χωρίς να το γνωρίζουν. Αυτοί οι άνθρωποι πάσχουν από ισχαιμία χωρίς πόνο (σιωπηρή ισχαιμία). Σε αυτή την περίπτωση μπορούν να υποστούν καρδιακή ανακοπή χωρίς προηγούμενη προειδοποίηση. Οι συνηθέστεροι τρόποι ελέγχου για τη διάγνωση της σιωπηρής ισχαιμίας είναι το τεστ κοπώσεως, ή η εφαρμογή της συσκευής Holter.

Οι γυναίκες έχουν μεγαλύτερη πιθανότητα από τους άντρες να προσβληθούν από έναν τύπο κρυφής στεφανιαίας πάθησης, κατά την οποία ο καρδιακός μυς στερείται οξυγόνου έστω κι αν οι στεφανιαίες αρτηρίες τους φαίνονται καθαρές και χωρίς φραγμούς στις ακτινογραφικές εξετάσεις, σύμφωνα με ανακοινώσεις των γιατρών.

## **B. ΚΑΡΚΙΝΟΣ**

Είναι ένας από τους μεγαλύτερους εχθρούς της υγείας του παγκόσμιου πληθυσμού και παρότι ολόκληρες «στρατιές» επιστημόνων έχουν ριχθεί στη μάχη εναντίον του, εκείνος συνεχίζει να μην παραδίδει (σε μεγάλο βαθμό) τα «όπλα» του.

Το νεόπλασμα (από την αρχαία λέξη νέο + πλάσμα, "δημιούργημα") μια ανώμαλη μάζα ιστού που είναι αποτέλεσμα της νεοπλασίας. Ένα νεόπλασμα μπορεί να είναι καλοήθες, δυνητικά κακόηθες, ή κακόηθες.

- Τα καλοήθη νεοπλάσματα περιλαμβάνουν ινομώματα της μήτρας και μελαγχρωματικοί σπίλοι (κρεατοελιές). Είναι οριοθετημένα και εντοπισμένα και δεν μετατρέπονται σε καρκίνο.
- Τα δυνητικά κακοήθη νεοπλάσματα περιλαμβάνουν τα καρκινώματα in situ. Δεν εισβάλλουν ούτε καταστρέφουν, αλλά αν τους δοθεί χρόνος μετατρέπονται σε καρκίνο.
- Τα κακοήθη νεοπλάσματα κοινώς αποκαλούνται καρκίνος. Εισβάλλουν και καταστρέφουν τους περιβάλλοντες ιστούς, μπορεί να κάνουν μεταστάσεις και τελικά σκοτώνουν τον ξενιστή.
- Το δευτερογενές νεόπλασμα αναφέρεται σε οποιοδήποτε είδος καρκινικού όγκου που αποτελεί είτε μεταστατικό παρακλάδι κυρίως όγκου, ή έναν άσχετο όγκο που αυξάνεται ως αποτέλεσμα συγκεκριμένων θεραπειών καρκίνου, όπως χημειοθεραπεία ή ραδιοθεραπεία.

Ο καρκίνος είναι ένα από τα σοβαρότερα προβλήματα υγείας που παρατηρούνται σήμερα στις αναπτυγμένες χώρες. Οι στατιστικές δείχνουν ότι αποτελεί τη δεύτερη πιο συχνή αιτία θανάτου μετά τις καρδιοπάθειες. Συνήθως προσβάλλει ανθρώπους μεγάλης ηλικίας, υπάρχουν όμως και μορφές καρκίνου που εμφανίζονται σε νεαρής ηλικίας άτομα, ακόμη και σε παιδιά.

Ο όρος «καρκίνος» δεν αποδίδεται σε μια και μόνη ασθένεια, αλλά σε μια ομάδα ασθενειών που χαρακτηρίζονται από τον ανεξέλεγκτο πολλαπλασιασμό των κυττάρων. Σε αντίθεση με τα φυσιολογικά κύτταρα στο σώμα μας, τα οποία αυξάνονται, διαιρούνται και πεθαίνουν με έναν αυστηρά ελεγχόμενο τρόπο, τα καρκινικά κύτταρα διαφέρουν διότι συνεχίζουν να



διαίρονται ανεξέλεγκτα. Αυτό έχει ως αποτέλεσμα την ανάπτυξη μιας μάζας κυττάρων, που ονομάζεται όγκος. Οι όγκοι μπορεί να είναι καλοήθεις ή κακοήθεις.



Γενικά, τα χαρακτηριστικά των καρκινικών κυττάρων είναι ότι:

- Ανθίστανται στην απόπτωση
- Πολλαπλασιάζονται παρουσία ή μη αυξητικών παραγόντων
- Ανθίστανται σε σήματα που σταματούν τον κυτταρικό πολλαπλασιασμό
- Ανθίστανται στον μηχανισμό γήρανσης
- Κάνουν μεταστάσεις
- Δημιουργούν αγγεία για την αιμάτωσή τους

*i. Κακοήθη νεοπλάσματα των οργάνων του αναπνευστικού συστήματος και ενδοθωρακικών οργάνων*



Εικόνα 2: Κακοήθη νεοπλάσματα αναπνευστικού συστήματος

Ο καρκίνος του πνεύμονα εμφανίζεται πολύ συχνά. Μάλιστα, κατά τις τελευταίες δεκαετίες, η συχνότητά του έχει αυξηθεί, κυρίως στις γυναίκες. Ο καρκίνος του πνεύμονα χωρίζεται σε δυο μεγάλες ομάδες:

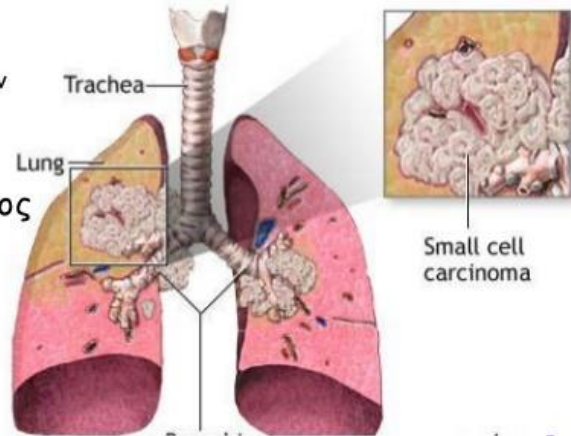
- a. Ο μικροκυτταρικός τύπος καρκίνου του πνεύμονα έχει άμεση σχέση με το κάπνισμα, αναπτύσσεται και διασπείρεται ταχέως και για το λόγο αυτό θεωρείται ότι δεν αντιμετωπίζεται με επέμβαση. Θεωρητικά έχει κάνει μεταστάσεις την ώρα της διάγνωσης ακόμα κι αν ο εργαστηριακός έλεγχος δεν έχει διαπιστώσει κάποια μετάσταση. Ο τύπος αυτός ανταποκρίνεται πολύ καλά στην χημειοθεραπεία και την ακτινοθεραπεία.
- b. Ο μη μικροκυτταρικός τύπος καρκίνου πνεύμονα διαιρείται σε τρεις υποομάδες: τον πλακώδη, το αδenoκαρκίνωμα και το μεγαλοκυτταρικό τύπο. Το αδenoκαρκίνωμα είναι το μοναδικό που δεν έχει άμεση σχέση με το κάπνισμα. Ο μη μικροκυτταρικού τύπου καρκίνος του πνεύμονα εξαπλώνεται τοπικά και αργά να δώσει απομακρυσμένες μεταστάσεις. Η χειρουργική θεραπεία παίζει σημαντικό ρόλο στην πλήρη θεραπεία. Όταν η νόσος βρίσκεται σε προχωρημένα στάδια, η αντιμετώπισή της περιλαμβάνει χημειοθεραπεία και ακτινοβολία.

## καρκίνος του πνεύμονα 1/2

- ✓ κάπνισμα
- ✓ η έκθεση στον αμίαντο και σε άλλα βιομηχανικά υλικά
- ✓ η ρύπανση της ατμόσφαιρας
- ✓ οι χρόνιες φλεγμονώδεις βλάβες των πνευμόνων

### κυριότερα συμπτώματα

- ✓ βήχας, έντονος και επίμονος
- ✓ αιμόπτυση
- ✓ απώλεια βάρους
- ✓ δύσπνοια
- ✓ θωρακικός πόνος
- ✓ πληκτροδακτυλία



παθήσεις αναπνευστικού συστήματος, Α. Καραματσούκη

15

Τα κυριότερα συμπτώματα από τον καρκίνο του πνεύμονα είναι τα εξής:

- Συνεχιζόμενος βήχας ή αλλαγή προϋπάρχοντος χρόνιου βήχα
- Λοίμωξη του πνεύμονα που δεν περνάει ή υποτροπιάζει συχνά
- Αποβολή αίματος με τα πτύελα
- Βράχνιασμα της φωνής που δεν περνάει
- Λαχάνιασμα και αναπνευστικός συριγμός
- Πόνος του θώρακα κατά το βήχα ή τη βαθιά εισπνοή
- Απώλεια όρεξης και βάρους
- Δυσκολία στην κατάποση, μεγάλη κόπωση και υπνηλία

Οι κύριοι παράγοντες κινδύνου του καρκίνου του πνεύμονα είναι:

- Κάπνισμα
- Ηλικία
- Φύλο
- Κληρονομικότητα
- Δίαιτα

- Ακτινοβολία
- Επάγγελμα
- Ατμοσφαιρική ρύπανση

Ο καρκίνος του πνεύμονα είναι 10 φορές συχνότερος στους καπνιστές απ' ό,τι στους μη καπνιστές. Ο κίνδυνος αυξάνει όσο αυξάνει η ποσότητα και η διάρκεια του καπνίσματος και όσο μικρότερη είναι η ηλικία έναρξης, ενώ ελαττώνεται με τη διακοπή του καπνίσματος. Επίσης, η καθημερινή έκθεση των μη-καπνιστών επί πολλά χρόνια στο κάπνισμα άλλων φαίνεται ότι επηρεάζει την αναπνευστική λειτουργία και αυξάνει την πιθανότητα προσβολής από καρκίνο του πνεύμονα. Ο καρκίνος του πνεύμονα μπορεί να προληφθεί. Σχεδόν όλοι οι καρκίνοι του πνεύμονα οφείλονται στο κάπνισμα. Έτσι, έχει μεγάλη σημασία ιδίως για τους νέους να μην αρχίσουν το κάπνισμα, εκθέτοντας έτσι τον οργανισμό τους σε ένα παράγοντα με σημαντικές βλαπτικές επιδράσεις για την υγεία τους. Επίσης είναι πολύ σημαντικό οι καπνιστές να σταματήσουν το κάπνισμα.

Παρακάτω είναι οι κωδικοί με τα κακοήθη νεοπλάσματα του αναπνευστικού συστήματος και των ενδοθωρακικών οργάνων:

C30: Κακοήθη νεοπλάσματα των ρινικών κοιλοτήτων και του μέσου ωτός

C31: Κακοήθη νεοπλάσματα των επικουρικών κόλπων

C32: Κακοήθη νεοπλάσματα του λάρυγγα

C33: Κακοήθη νεοπλάσματα της τραχείας

C34: Κακοήθη νεοπλάσματα των βρόγχων και του πνεύμονα

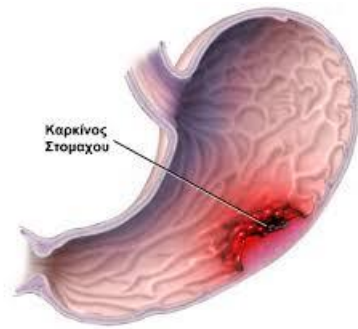
C37: Κακοήθη νεοπλάσματα του θύμου αδένος

C38: Κακοήθη νεοπλάσματα της καρδιάς, του μεσοθωρακίου και του υπεζωκότα

C39: Κακοήθη νεοπλάσματα σε άλλες και ασαφείς εντοπίσεις στο αναπνευστικό σύστημα και σε ενδοθωρακικά όργανα.

ii) Κακοήθη νεοπλάσματα των οργάνων του πεπτικού συστήματος και περιτοναίου

Ο καρκίνος του στομάχου είναι κακοήθης νόσος που προσβάλλει συνήθως άτομα άνω των 55 ετών. Μπορεί όμως να εμφανιστεί και σε άτομα μικρότερης ηλικίας. Υπολογίζεται ότι η συχνότητα στους άνδρες είναι διπλάσια απ' ό,τι στις γυναίκες. Ο καρκίνος του στομάχου, στα αρχικά του στάδια, δεν δημιουργεί συμπτώματα που να οδηγούν τον ασθενή στο γιατρό. Για το λόγο αυτό η ασθένεια βρίσκεται σε προχωρημένα στάδια όταν γίνεται η διάγνωση. Ο καρκίνος αυτός, εκτός από το γεγονός ότι μπορεί να επεκταθεί τοπικά σε σημαντικό βαθμό προτού γίνει η διάγνωση, παράλληλα είναι δυνατόν να κάνει μεταστάσεις σε άλλα όργανα.



Εικόνα 3: Κακοήθη νεοπλάσματα του στομάχου

Η αντιμετώπιση του καρκίνου του στομάχου μπορεί να είναι αποτελεσματικότερη με καλύτερη πρόγνωση εάν η διάγνωση γίνει έγκαιρα με ανίχνευση της νόσου προτού ακόμη εξαπλωθεί.

Ποια είναι τα συχνότερα σημεία και συμπτώματα του καρκίνου του στομάχου;

- Ανορεξία
- Απροσδόκητη απώλεια βάρους
- Αποστροφή από το κρέας
- Πόνος στην κοιλιά
- Αίσθημα πληρότητας στο στομάχι έστω και με τη λήψη μικρών ποσοτήτων φαγητού
- Δυσπεψία, αναγούλες και εμετοί
- Αιματέμεση (εμετοί που περιέχουν αίμα)
- Μέλαινα (αίμα στα κόπρανα από το άνω πεπτικό σύστημα που έχει πάρει κατά τη διόδό του στο πεπτικό σωλήνα μαύρο χρώμα)
- Αδυναμία, κούραση

Οι κύριοι παράγοντες κινδύνου του καρκίνου του στομάχου είναι:

- Διατροφή
- Κάπνισμα και αλκοόλ
- Κληρονομικότητα
- Άλλοι παράγοντες κινδύνου (πχ. Κακοήθης αναιμία, χρόνια ατροφική γαστρίτιδα, εντερική μετάπλαση, παχυσαρκία)

Παρακάτω είναι οι κωδικοί με τα κακοήθη νεοπλάσματα των οργάνων του πεπτικού συστήματος και περιτοναίου.

C15: Κακοήθη νεοπλάσματα του οισοφάγου

- C16: Κακοήθη νεοπλάσματα του στομάχου
- C17: Κακοήθη νεοπλάσματα του λεπτού εντέρου
- C18: Κακοήθη νεοπλάσματα που εντοπίζεται στο κόλον
- C19: Κακοήθη νεοπλάσματα της ορθοσιγμοειδικής συμβολής
- C20: Κακοήθη νεοπλάσματα του ορθού
- C21: Κακοήθη νεοπλάσματα του πρωκτού και του πρωκτικού σωλήνα
- C22: Κακοήθη νεοπλάσματα του ήπατος και των ενδοηπατικών χολαγγείων
- C23: Κακοήθη νεοπλάσματα της χοληδόχου κύστεως
- C24: Κακοήθη νεοπλάσματα σε διάφορα άλλα τμήματα της χοληδόχου οδού
- C25: Κακοήθη νεοπλάσματα του παγκρέατος
- C26: Κακοήθη νεοπλάσματα σε άλλα και ill-defined όργανα του πεπτικού συστήματος

iii) Κακοήθη νεοπλάσματα του ουροποιητικού συστήματος

Το ουροποιητικό σύστημα αποτελείται από τα νεφρά, τους ουρητήρες, την ουροδόχο κύστη και την ουρήθρα.

Ορισμένοι αδένες, όπως οι όρχεις και ο προστάτης στους άνδρες και τα επινεφρίδια και στα δύο φύλα, είναι λειτουργικά και ανατομικά άρρηκτα συνδεδεμένα με τα υπόλοιπα ουρολογικά όργανα.

Το όργανο-κλειδί σ' αυτό το σύστημα είναι ο νεφρός. Τα νεφρά φιλτράρουν και αφαιρούν με την μορφή των ούρων το πλεόνασμα των υγρών και των άχρηστων και βλαβερών ουσιών από το αίμα. Ένα ζεύγος οργάνων που βρίσκονται δεξιά και αριστερά από την σπονδυλική στήλη κάτω από τις πλευρές στο μέσο της πλάτης.

Οι ουρητήρες είναι λεπτά σωληνάκια που μεταφέρουν τα ούρα από τους νεφρούς στην ουροδόχο κύστη, η οποία λειτουργεί σαν ρεζερβουάρ στην κάτω κοιλία για την αποθήκευση των ούρων.

Με την ουρήθρα, ένα λεπτό κανάλι που οδηγεί τα ούρα από την κύστη προς τα έξω, αδειάζει η ουροδόχος κύστη και αποβάλλονται τα ούρα από το σώμα.

Ο προστάτης είναι ένας αδένας του γεννητικού συστήματος του άνδρα που βρίσκεται μπροστά από την ουροδόχο κύστη και συμβάλλει στην ρευστοποίηση του σπέρματος άρα και στην βελτιστοποίηση της γονιμότητάς του.

Τα επινεφρίδια είναι ένα ζεύγος αδένων που βρίσκονται σαν σκούφοι πάνω από κάθε νεφρό

και συμβάλλουν , εκκρίνοντας διάφορες ουσίες , στην ρύθμιση της αρτηριακής πίεσεως και του ισοζυγίου των υγρών του ανθρώπινου σώματος. Εάν ένα τμήμα του ουροποιητικού συστήματος δεν λειτουργεί όπως πρέπει παρουσιάζονται διάφορα συμπτώματα που διαφέρουν κατά περίπτωση. Τα συμπτώματα μπορεί να είναι ήπια , μέσης βαρύτητας ή και σοβαρά , μπορεί να είναι εμφανή ή και λανθάνοντα (να μην υπάρχουν καθόλου δηλαδή). Πρέπει να υπογραμμισθεί ότι εάν τα συμπτώματα που υπήρχαν, εξαλειφθούν ή βελτιωθούν από μόνα τους δεν σημαίνει ότι και η πάθηση, η οποία τα προκάλεσε, θεραπεύτηκε ή πέρασε από μόνη της.

Μερικά από τα τυπικά σημεία των ουρολογικών παθήσεων:

- Του καρκίνου της ουροδόχου κύστης,
- Του νεφρού ή του προστάτη ή της απόφραξης της πυελοουρητικής συμβολής

Τα συμπτώματα είναι τα παρακάτω:

- Αίμα στα ούρα, αίμα στο σπέρμα
- Ψηλαφητός όγκος στην κοιλιά
- Ανεξήγητη απώλεια βάρους ή ορέξεως
- Αίσθημα αδυναμίας ή εύκολης κοπώσεως
- Πόνος στην οσφυϊκή περιοχή ή στην μέση , διαταραχές της ούρησης

Παρακάτω είναι οι κωδικοί με τα κακοήθη νεοπλάσματα των ουροποιητικών οργάνων:

C51-C58: Κακοήθη νεοπλάσματα των γυναικείων γεννητικών οργάνων

C51: Κακοήθη νεοπλάσματα του αιδοίου

C52: Κακοήθη νεοπλάσματα του κόλπου

C53: Κακοήθη νεοπλάσματα του τραχήλου της μήτρας

C54: Κακοήθη νεοπλάσματα του σώματος της μήτρας

C55: Κακοήθη νεοπλάσματα της μήτρας (ανεξαρτήτως τμήματος στο οποίο αναφέρονται)

C56: Κακοήθη νεοπλάσματα των ωοθηκών

C57: Κακοήθη νεοπλάσματα διαφόρων άλλων τμημάτων γυναικείων γεννητικών οργάνων

C58: Κακοήθη νεοπλάσματα του πλακούντα

C60-C63: Κακοήθη νεοπλάσματα των ανδρικών γεννητικών οργάνων



C60: Κακοήθη νεοπλάσματα του πέους

C61: Κακοήθη νεοπλάσματα του προστάτη

C62: Κακοήθη νεοπλάσματα του όρχεως

C63: Κακοήθη νεοπλάσματα διαφόρων άλλων τμημάτων των ανδρικών γεννητικών οργάνων

## 2.ΟΡΙΣΜΟΙ

- **Πληθυσμός:**

1. *Πραγματικός πληθυσμός* είναι ο αριθμός των ατόμων που βρέθηκαν παρόντα κατά την ημέρα της απογραφής, σε κάθε περιφέρεια, νομό, δήμο/κοινότητα, δημοτικό/κοινοτικό διαμέρισμα και αυτοτελή οικισμό.
2. *Μόνιμος πληθυσμός* είναι ο αριθμός των ατόμων που έχουν τη συνήθη διαμονή τους σε κάθε περιφέρεια, νομό, δήμο/κοινότητα, δημοτικό/κοινοτικό διαμέρισμα και αυτοτελή οικισμό.
3. *Νόμιμος πληθυσμός* κάθε δήμου ή κοινότητας είναι ο αριθμός των ατόμων κάθε ηλικίας και φύλου, τα οποία κατά την ημέρα της απογραφής δήλωσαν ότι είναι καταχωρημένα στα αντίστοιχα δημοτολόγια και εφόσον κατά την ημέρα της απογραφής βρίσκονταν στη χώρα και απογράφηκαν ως παρόντα ή απουσίαζαν προσωρινά στο εξωτερικό.

- **Θάνατος:**

Θάνατος είναι η οριστική παύση όλων των βιολογικών λειτουργιών που υποστηρίζουν τη διαβίωση ενός οργανισμού. Φαινόμενα που συνήθως επιφέρουν το θάνατο περιλαμβάνουν τα γηρατειά, τον υποσιτισμό, την ασθένεια, και το τραύμα το οποίο περιλαμβάνει την αυτοκτονία, την δολοφονία και το ατύχημα. Όλοι οι γνωστοί οργανισμοί αντιμετωπίζουν αναπόφευκτα το θάνατο. Τα σώματα των ζωντανών οργανισμών αρχίζουν να αποσυντίθενται λίγο μετά το θάνατο.

Ο θάνατος διακρίνεται στον φυσιολογικό, ο οποίος έρχεται με τα βαθιά γεράματα και τον ξαφνικό που οφείλεται σε ασθένεια, φόνο ή δυστύχημα. Με το θάνατο η θερμοκρασία του σώματος πέφτει στους 20°C, η αναπνοή μαζί με τους χτύπους της καρδιάς σταματά και το πρόσωπο κιτρινίζει. Ορισμένα κύτταρα του ατόμου ζουν και δυο μέρες μετά το θάνατό του. Τα αίτια του θανάτου, στην περίπτωση που δεν προϋπήρξε σοβαρό πρόβλημα υγείας, εξακριβώνει ο ιατροδικαστής.



### ΚΕΦΑΛΑΙΟ 3: Κατασκευή πινάκων

Σε αυτό το κεφάλαιο, θα εφαρμόσουμε τη θεωρία των μοντέλων πολλαπλών απαντημάτων αλλά και των απλών συνδεδεμένων πινάκων, με στόχο την εξαγωγή και την ερμηνεία των αποτελεσμάτων.

Τα πρώτα βήματα για την κατασκευή των πινάκων είναι τα ακόλουθα:

- Αναζητήσαμε στην Ελληνική Στατιστική Υπηρεσία και από την Eurostat, τον αριθμό θανάτων για τη χρονολογία 2013. Έπειτα, επιλέχθηκαν οι 4 αιτίες για τις οποίες έγινε η μελέτη με στόχο την πρόβλεψη.
- Στη συνέχεια, ήταν απαραίτητο να έχουμε τον μόνιμο πληθυσμό της κάθε χρονιάς, στο μέσο του έτους (αυτό γιατί ο κεντρικός ρυθμός θνησιμότητας ορίζεται ως ο αριθμός των θανάτων στην ηλικία  $x$  δια τον μέσο πληθυσμό). Επειδή ο πληθυσμός που χρησιμοποιήσαμε από την Eurostat και την Ελληνική Στατιστική Υπηρεσία ήταν στην αρχή του έτους, δηλαδή 1<sup>η</sup> Ιανουαρίου, κάναμε την παρακάτω προσέγγιση:

**Παράδειγμα:** Αν θέλαμε να βρούμε τον πληθυσμό αντρών ηλικίας 0-4 για το 1981:

$$\text{Πληθυσμός } 81 = \frac{\text{πληθυσμός αρχή έτους } 80 + \text{πληθυσμός αρχή έτους } 81}{2}$$

- Σύμφωνα με τον Johnson, και την αντίστοιχη κατασκευή πινάκων, ακολουθήσαμε τα βήματά του, χρησιμοποιώντας τα δικά μας στατιστικά στοιχεία, για το σύνολο των ετών 2010, 2011, 2012, για όλες τις ηλικιακές κλάσεις.

Παρακάτω παρουσιάζονται οι πίνακες 1 και 2, οι οποίοι αποτελούνται από τις ηλικιακές κλάσεις, τον πληθυσμό, το σύνολο θανάτων από όλες τις αιτίες, τις επιλεγμένες αιτίες θανάτου, για άντρες και γυναίκες. Αυτές είναι:

- a. Ισχαιμική καρδιοπάθεια
- b. Κακοήθη νεοπλάσματα του αναπνευστικού συστήματος και ενδοθωρακικών οργάνων
- c. Κακοήθη νεοπλάσματα του πεπτικού συστήματος και περιτοναίου και
- d. Κακοήθη νεοπλάσματα του ουροποιητικού συστήματος
- e. Υπόλοιπες αιτίες θανάτου (δηλαδή, όλες οι αιτίες εξαιρουμένων των προαναφερόμενων). Δεν θα σχολιάζαμε τα αποτελέσματα σε αυτήν την περίπτωση, αφού μας ενδιέφερε να επικεντρωθούμε στις παραπάνω 4 αιτίες.

Α	B	C	D	E	F	G	H	I	J	K	
1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΙ	ΙΣΧΑΙΜΙΚΗ ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	ΚΕΝΤΡΙΚΗ ΕΚΘΕΣΗ	ΕΝΤΑΣΗ ΘΝ/ΤΗΤΑΣ ΚΑΡΚΙΝΟ	ΕΝΤΑΣΗ ΘΝ/ΤΗΤΑΣ ΙΣΧΑΙΜΙΚΗΣ ΚΑΡΔ	ΕΝΤΑΣΗ ΘΝ/ΤΗΤΑΣ ΑΝΑΠΝΕΥΣΤΙΚΑ	ΕΝΤΑΣΗ ΘΝ/ΤΗΤΑΣ ΑΛΛΩΝ ΑΙΤΙΩΝ
2	2013	30	8	9	1	44	77294	0,000102	0,000113	1,60E-05	0,000565
3	2013	31	11	11	2	47	80137	0,000137	0,000131	2,66E-05	0,000584
4	2013	32	13	12	3	48	82203	0,000159	0,000149	3,30E-05	0,000587
5	2013	33	14	14	3	48	83164	0,000169	0,000167	3,59E-05	0,000578
6	2013	34	14	16	3	46	83121	0,000169	0,000187	3,54E-05	0,000557
7	2013	35	14	18	3	43	82434	0,000166	0,000221	3,53E-05	0,000525
8	2013	36	14	22	3	41	83107	0,00017	0,000262	3,76E-05	0,000499
9	2013	37	15	25	3	41	83005	0,000183	0,000301	4,06E-05	0,000499
10	2013	38	17	28	4	43	82086	0,000206	0,000339	4,44E-05	0,000526
11	2013	39	19	30	4	47	81747	0,000235	0,00037	4,83E-05	0,000571
12	2013	40	22	33	4	50	81490	0,000272	0,000409	5,22E-05	0,000616
13	2013	41	26	38	5	53	82267	0,00031	0,000459	5,61E-05	0,000642
14	2013	42	29	43	5	56	82983	0,000353	0,000514	6,13E-05	0,000676
15	2013	43	33	48	6	60	83334	0,0004	0,000577	6,61E-05	0,000718
16	2013	44	38	54	6	64	83097	0,000454	0,000651	7,66E-05	0,000771
17	2013	45	44	65	8	71	82461	0,000538	0,000785	9,45E-05	0,00086
18	2013	46	53	78	9	78	80565	0,000659	0,000965	0,000116726	0,000971
19	2013	47	62	88	10	83	77789	0,000795	0,001128	0,000132528	0,001066
20	2013	48	71	95	11	85	74908	0,000941	0,001266	0,000140241	0,001137
21	2013	49	79	99	10	85	72267	0,001096	0,00137	0,000138265	0,001173
22	2013	50	95	104	9	84	71453	0,001325	0,001458	0,000124324	0,001171
23	2013	51	115	113	8	84	71732	0,001602	0,001571	0,000114983	0,001176
24	2013	52	132	122	9	86	71847	0,001836	0,001692	0,000118394	0,0012
25	2013	53	146	131	10	89	71502	0,00204	0,001831	0,000125077	0,001248
26	2013	54	157	141	12	93	70299	0,002228	0,002001	0,000166491	0,00133
27	2013	55	173	155	15	99	68409	0,002528	0,002265	0,000223653	0,001446
28	2013	56	196	171	19	104	67600	0,002894	0,002533	0,00028417	0,001537
29	2013	57	216	184	22	108	67715	0,003196	0,00272	0,000327254	0,00159
30	2013	58	235	194	24	110	65615	0,003588	0,002952	0,000368059	0,001679
31	2013	59	253	200	25	111	62617	0,004034	0,003192	0,000402134	0,001779
32	2013	60	278	207	26	111	61686	0,004511	0,003352	0,000427548	0,001796
33	2013	61	310	217	29	111	61765	0,005022	0,003514	0,000463311	0,001799
34	2013	62	337	228	31	114	61292	0,005496	0,003718	0,000509519	0,001867
35	2013	63	358	239	34	121	60372	0,005934	0,003962	0,000566702	0,001999
36	2013	64	374	251	38	130	59539	0,006289	0,004218	0,000630976	0,002183
37	2013	65	403	265	42	141	59167	0,006809	0,004476	0,000711771	0,002388
38	2013	66	438	280	47	152	58644	0,007472	0,004773	0,000808704	0,002592
39	2013	67	460	295	52	162	57076	0,008066	0,005163	0,000919287	0,00283
40	2013	68	469	309	57	170	53389	0,008791	0,005791	0,001072189	0,003183
41	2013	69	465	323	62	177	48543	0,009584	0,006663	0,001272041	0,00365
42	2013	70	452	330	66	182	45299	0,009983	0,007294	0,001458185	0,004008
43	2013	71	448	340	72	188	44629	0,010041	0,00761	0,001610892	0,004208
44	2013	72	458	363	80	200	45652	0,010022	0,007945	0,00175385	0,004384
45	2013	73	480	400	91	219	46853	0,010253	0,008531	0,001933145	0,004664
46	2013	74	517	451	103	243	47285	0,010929	0,009529	0,002187016	0,005139
47	2013	75	574	530	125	280	46653	0,012305	0,011361	0,002678282	0,006004
48	2013	76	627	616	151	320	45209	0,013871	0,013627	0,003335776	0,007075
49	2013	77	656	683	172	351	43321	0,015134	0,015177	0,003980384	0,008098
50	2013	78	660	731	190	373	41142	0,016036	0,017775	0,004613883	0,009065
51	2013	79	639	760	203	386	38858	0,016456	0,019157	0,005223745	0,009941
52	2013	80	619	801	223	404	36409	0,017008	0,021987	0,006128071	0,011098
53	2013	81	615	861	251	431	33964	0,018118	0,023351	0,007391426	0,012684
54	2013	82	611	917	276	456	31572	0,019366	0,0259041	0,008741413	0,014433
55	2013	83	607	968	298	479	28702	0,021166	0,033728	0,010380655	0,016677
56	2013	84	604	1015	317	500	24936	0,024202	0,040684	0,012708791	0,020042

Εικόνα 4: (am)x κεντρικός ρυθμός θνησιμότητας για το έτος 2013

Στον πίνακα παρατηρούμε για το έτος 2013, τον αριθμό θανάτων από τον καρκίνο, την ισχαιμική καρδιοπάθεια, τα αναπνευστικά προβλήματα και όλες τις υπόλοιπες αιτίες. Βλέπουμε, ότι είναι πολύ μικρός ο αριθμός θανάτων από αναπνευστικά προβλήματα απ' ότι ο θάνατος από καρκίνο ή από ισχαιμική καρδιοπάθεια. Επίσης, έχουμε βρει την ένταση θνησιμότητας για κάθε μια από τις τέσσερις αιτίες θανάτου, η οποία υπολογίζεται από τον τύπο:

$$\hat{\mu}_x = \frac{\Delta_x}{R_x^c}$$

	A	B	C	D	E	F	G	H	I	J
1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	txx1	txx2	txx3	txx4
2	2013	30	8	9	1	44	0,0001	0,00011	1,6E-05	0,00057
3	2013	31	11	11	2	47	0,00014	0,00013	2,66E-05	0,00058
4	2013	32	13	12	3	48	0,00016	0,00015	3,3E-05	0,00059
5	2013	33	14	14	3	48	0,00017	0,00017	3,58E-05	0,00058
6	2013	34	14	16	3	46	0,00017	0,00019	3,53E-05	0,00056
7	2013	35	14	18	3	43	0,00017	0,00022	3,53E-05	0,00053
8	2013	36	14	22	3	41	0,00017	0,00026	3,76E-05	0,0005
9	2013	37	15	25	3	41	0,00018	0,0003	4,06E-05	0,0005
10	2013	38	17	28	4	43	0,00021	0,00034	4,44E-05	0,00053
11	2013	39	19	30	4	47	0,00023	0,00037	4,83E-05	0,00057
12	2013	40	22	33	4	50	0,00027	0,00041	5,22E-05	0,00062
13	2013	41	26	38	5	53	0,00031	0,00046	5,61E-05	0,00064
14	2013	42	29	43	5	56	0,00035	0,00051	6,13E-05	0,00068
15	2013	43	33	48	6	60	0,0004	0,00058	6,8E-05	0,00072
16	2013	44	38	54	6	64	0,00045	0,00065	7,65E-05	0,00077
17	2013	45	44	65	8	71	0,00054	0,00078	9,44E-05	0,00086
18	2013	46	53	78	9	78	0,00066	0,00096	0,000117	0,00097
19	2013	47	62	88	10	83	0,00079	0,00113	0,000132	0,00106
20	2013	48	71	95	11	85	0,00094	0,00126	0,00014	0,00113
21	2013	49	79	99	10	85	0,00109	0,00137	0,000138	0,00117
22	2013	50	95	104	9	84	0,00132	0,00145	0,000124	0,00117
23	2013	51	115	113	8	84	0,0016	0,00157	0,000115	0,00117
24	2013	52	132	122	9	86	0,00183	0,00169	0,000118	0,0012
25	2013	53	146	131	10	89	0,00203	0,00183	0,000135	0,00124
26	2013	54	157	141	12	93	0,00222	0,002	0,000166	0,00133
27	2013	55	173	155	15	99	0,00252	0,00226	0,000223	0,00144
28	2013	56	196	171	19	104	0,00288	0,00252	0,000283	0,00153
29	2013	57	216	184	22	108	0,00318	0,00271	0,000326	0,00158
30	2013	58	235	194	24	110	0,00357	0,00294	0,000366	0,00167
31	2013	59	253	200	25	111	0,00402	0,00318	0,0004	0,00177
32	2013	60	278	207	26	111	0,00449	0,00334	0,000425	0,00179
33	2013	61	310	217	29	111	0,00499	0,0035	0,000461	0,00179
34	2013	62	337	228	31	114	0,00546	0,0037	0,000507	0,00186
35	2013	63	358	239	34	121	0,0059	0,00394	0,000563	0,00199
36	2013	64	374	251	38	130	0,00625	0,00419	0,000627	0,00217
37	2013	65	403	265	42	141	0,00676	0,00444	0,000707	0,00237
38	2013	66	438	280	47	152	0,00741	0,00474	0,000802	0,00257
39	2013	67	460	295	52	162	0,008	0,00512	0,000912	0,00281
40	2013	68	469	309	57	170	0,00871	0,00574	0,001062	0,00315
41	2013	69	465	323	62	177	0,00948	0,00659	0,001259	0,00361
42	2013	70	452	330	66	182	0,00987	0,00721	0,001442	0,00396
43	2013	71	448	340	72	188	0,00992	0,00752	0,001592	0,00416
44	2013	72	458	363	80	200	0,0099	0,00785	0,001733	0,00433
45	2013	73	480	400	91	219	0,01012	0,00842	0,001909	0,00461
46	2013	74	517	451	103	243	0,01078	0,0094	0,002157	0,00507
47	2013	75	574	530	125	280	0,01211	0,01118	0,002636	0,00591
48	2013	76	627	616	151	320	0,01361	0,01337	0,003274	0,00694
49	2013	77	656	683	172	351	0,01482	0,01544	0,003897	0,00793
50	2013	78	660	731	190	373	0,01566	0,01736	0,004507	0,00885
51	2013	79	639	760	203	386	0,01605	0,01908	0,005093	0,00969
52	2013	80	619	801	223	404	0,01654	0,02139	0,005961	0,01079
53	2013	81	615	861	251	431	0,01756	0,02457	0,007164	0,01229
54	2013	82	611	917	276	456	0,0187	0,02804	0,008439	0,01393
55	2013	83	607	968	298	479	0,02033	0,0324	0,009972	0,01602
56	2013	84	604	1015	317	500	0,02308	0,03879	0,012117	0,01911
57							1	1	1	1

Εικόνα 5: (α)χ πιθανότητα αποβίωσης για το έτος 2013

Στον παραπάνω πίνακα, έχουμε υπολογίσει την συνάρτηση κατανομής της απομένουσας ζωής (δηλαδή την πιθανότητα ότι άτομο ηλικίας  $x$  θα πεθάνει τα επόμενα  $t$  χρόνια) για τα κακοήγη νεοπλασμάτα, την ισχαιμική καρδιοπάθεια, τα αναπνευστικά προβλήματα και τέλος για όλες τις υπόλοιπες αιτίες μαζί. Παρατηρούμε ότι και στις 4 αιτίες θανάτων υπάρχει μεγαλύτερη πιθανότητα αποβίωσης στις ηλικίες (74+). Η πιθανότητα αποβίωσης υπολογίζεται από τον τύπο:

$$(aq)_x^j \cong \frac{1}{1} \cdot \left[ 1 + \frac{1}{2} \cdot {}_n(am)_x^T \right]$$

1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	t(aq)xT	(ad)x1	(ad)x2	(ad)x3	(ad)x4	(ad)xT
2	2013	30	8	9	1	44	0,0008	10,2135	11,3367	1,59929	56,5076	79,6571
3	2013	31	11	11	2	47	0,00088	13,7034	13,1331	2,65521	58,2854	87,7771
4	2013	32	13	12	3	48	0,00093	15,8369	14,8756	3,29223	58,5327	92,5375
5	2013	33	14	14	3	48	0,00095	16,8439	16,6829	3,57428	57,6032	94,7043
6	2013	34	14	16	3	46	0,00095	16,787	18,6025	3,52174	55,4395	94,3507
7	2013	35	14	18	3	43	0,00095	16,4757	21,987	3,51374	52,2888	94,2653
8	2013	36	14	22	3	41	0,00097	16,8542	26,0489	3,73843	49,61	96,2516
9	2013	37	15	25	3	41	0,00102	18,1446	29,8744	4,03183	49,5503	101,601
0	2013	38	17	28	4	43	0,00111	20,3944	33,5881	4,40452	52,137	110,524
1	2013	39	19	30	4	47	0,00122	23,2962	36,6613	4,78698	56,5465	121,291
2	2013	40	22	33	4	50	0,00135	26,9268	40,5224	5,16852	60,9171	133,535
3	2013	41	26	38	5	53	0,00147	30,683	45,3569	5,5475	63,479	145,066
4	2013	42	29	43	5	56	0,0016	34,7868	50,7413	6,05235	66,6615	158,242
5	2013	43	33	48	6	60	0,00176	39,3774	56,8741	6,70442	70,7073	173,663
6	2013	44	38	54	6	64	0,00195	44,6229	63,99	7,5294	75,8541	191,996
7	2013	45	44	65	8	71	0,00228	52,8159	77,0058	9,26814	84,3867	223,477
8	2013	46	53	78	9	78	0,00271	64,5127	94,4051	11,4238	95,007	265,349
9	2013	47	62	88	10	83	0,00312	77,54	110,083	12,9325	104,054	304,609
0	2013	48	71	95	11	85	0,00348	91,5442	123,14	13,64	110,544	338,868
1	2013	49	79	99	10	85	0,00377	106,189	132,721	13,3992	113,691	366
2	2013	50	95	104	9	84	0,00407	127,9	140,72	12,0009	113,068	393,689
3	2013	51	115	113	8	84	0,00445	153,934	150,964	11,0519	113,047	428,997
4	2013	52	132	122	9	86	0,00484	175,695	161,898	11,327	114,809	463,728
5	2013	53	146	131	10	89	0,00524	194,185	174,289	12,858	118,82	500,152
6	2013	54	157	141	12	93	0,00571	210,942	189,44	15,7615	125,869	542,013
7	2013	55	173	155	15	99	0,00644	237,858	213,149	21,0443	136,016	608,068
8	2013	56	196	171	19	104	0,00722	270,402	236,755	26,5559	143,622	677,335
9	2013	57	216	184	22	108	0,0078	296,453	252,27	30,3524	147,446	726,522
0	2013	58	235	194	24	110	0,00855	330,076	271,588	33,858	154,437	789,959
1	2013	59	253	200	25	111	0,00936	367,769	291,01	36,6613	162,215	857,655
2	2013	60	278	207	26	111	0,01004	407,267	302,646	38,6002	162,165	910,678
3	2013	61	310	217	29	111	0,01074	448,668	313,967	41,3945	160,744	964,774
4	2013	62	337	228	31	114	0,01152	485,536	328,448	45,0163	164,938	1023,94
5	2013	63	358	239	34	121	0,01239	518,032	345,881	49,4701	174,525	1087,91
6	2013	64	374	251	38	130	0,01323	542,007	363,452	54,3755	188,096	1147,93
7	2013	65	403	265	42	141	0,01428	578,696	380,437	60,4946	202,978	1222,61
8	2013	66	438	280	47	152	0,01552	625,586	399,59	67,709	217,011	1309,9
9	2013	67	460	295	52	162	0,01684	664,374	425,258	75,7227	233,147	1398,5
0	2013	68	469	309	57	170	0,01866	711,316	468,552	86,7506	257,552	1524,17
1	2013	69	465	323	62	177	0,02095	760,094	528,408	100,883	289,5	1678,89
2	2013	70	452	330	66	182	0,02249	774,576	565,942	113,136	310,938	1764,59
3	2013	71	448	340	72	188	0,0232	761,243	576,913	122,129	319,016	1779,3
4	2013	72	458	363	80	200	0,02382	741,921	588,18	129,842	324,537	1784,48
5	2013	73	480	400	91	219	0,02506	740,515	616,123	139,619	336,866	1833,12
6	2013	74	517	451	103	243	0,0274	768,618	670,179	153,814	361,448	1954,06
7	2013	75	574	530	125	280	0,03183	839,843	775,364	182,791	409,746	2207,74
8	2013	76	627	616	151	320	0,0372	914,024	897,991	219,816	466,206	2498,04
9	2013	77	656	683	172	351	0,04208	957,819	998,05	251,908	512,489	2720,27
0	2013	78	660	731	190	373	0,04639	970,031	1075,21	279,098	548,353	2872,69
1	2013	79	639	760	203	386	0,04991	947,541	1126,84	300,787	572,437	2947,61
2	2013	80	619	801	223	404	0,05468	928,15	1199,89	334,426	605,638	3068,11
3	2013	81	615	861	251	431	0,06159	931,34	1303,2	379,96	652,042	3266,54
4	2013	82	611	917	276	456	0,06911	930,595	1395,47	420,046	693,542	3439,66
5	2013	83	607	968	298	479	0,07873	942,057	1501,18	462,031	742,296	3647,56
6	2013	84	604	1015	317	500	0,09309	984,97	1655,78	517,225	815,676	3973,65
7												

Εικόνα 6: (ad)x αιτίες θανάτων για το έτος 2013

Σε αυτόν τον πίνακα τετραπλού απαιξήματος υπολογίσαμε τον αναμενόμενο αριθμό των ατόμων από την αρχική ομάδα των  $(al)_a^T$  που θα αποβιώσουν μέσα στο ηλικιακό διάστημα  $[x, x+n]$ . Παρατηρούμε ότι στην πρώτη αιτία το μέγιστο είναι 984,97, στη δεύτερη αιτία το μέγιστο είναι 1655,78, στην Τρίτη αιτία το μέγιστο είναι 517,225 και στην τέταρτη το

μέγιστο είναι 815,676. Υπολογίζεται από τον παρακάτω τύπο:

$${}_n(ad)_x^j = (ad)_x^T \cdot {}_n(aq)_x^j$$

Η

$${}_n(aq)_x^j = \frac{{}_n(ad)_x^j}{(ad)_x^T}$$

Όπου  ${}_n(aq)_x^j$  δηλώνει τη δεσμευμένη πιθανότητα αποχώρησης ατόμων ηλικίας ( $x$ ) από την αρχική κατάσταση λόγω του  $j$ -αιτίου μέχρι την ηλικία ( $x+t$ ) από την αρχική ομάδα των  $(ad)_x^T$ .

Α	Β	Γ	Δ	Ε	Ζ	Η	Θ	Ι	Κ	
1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	(αl)x1	(αl)x2	(αl)x3	(αl)x4
2	2013	30	8	9	1	44	21977,6	21902,7	4965,3	12443
3	2013	31	11	11	2	47	21967,4	21891,3	4963,7	12387
4	2013	32	13	12	3	48	21953,7	21878,2	4961,05	12328
5	2013	33	14	14	3	48	21937,8	21863,3	4957,75	12270
6	2013	34	14	16	3	46	21921	21846,6	4954,18	12212
7	2013	35	14	18	3	43	21904,2	21828	4950,66	12157
8	2013	36	14	22	3	41	21887,7	21806,1	4947,14	12104
9	2013	37	15	25	3	41	21870,9	21780	4943,41	12055
10	2013	38	17	28	4	43	21852,7	21750,1	4939,37	12005
11	2013	39	19	30	4	47	21832,3	21716,5	4934,97	11953
12	2013	40	22	33	4	50	21809	21679,9	4930,18	11897
13	2013	41	26	38	5	53	21782,1	21639,4	4925,01	11836
14	2013	42	29	43	5	56	21751,4	21594	4919,47	11772
15	2013	43	33	48	6	60	21716,6	21543,3	4913,41	11705
16	2013	44	38	54	6	64	21677,3	21486,4	4906,71	11635
17	2013	45	44	65	8	71	21632,6	21422,4	4899,18	11559
18	2013	46	53	78	9	78	21579,8	21345,4	4889,91	11475
19	2013	47	62	88	10	83	21515,3	21251	4878,49	11380
20	2013	48	71	95	11	85	21437,8	21140,9	4865,56	11275
21	2013	49	79	99	10	85	21346,2	21017,8	4851,92	11165
22	2013	50	95	104	9	84	21240	20885	4838,52	11051
23	2013	51	115	113	8	84	21112,1	20744,3	4826,52	10938
24	2013	52	132	122	9	86	20958,2	20593,4	4815,46	10825
25	2013	53	146	131	10	89	20782,5	20431,5	4804,14	10710
26	2013	54	157	141	12	93	20588,3	20257,2	4791,28	10591
27	2013	55	173	155	15	99	20377,4	20067,7	4775,52	10466
28	2013	56	196	171	19	104	20139,5	19854,6	4754,47	10330
29	2013	57	216	184	22	108	19869,1	19617,8	4727,92	10186
30	2013	58	235	194	24	110	19572,7	19365,6	4697,57	10039
31	2013	59	253	200	25	111	19242,6	19094	4663,71	9884
32	2013	60	278	207	26	111	18874,8	18803	4627,05	9722
33	2013	61	310	217	29	111	18467,6	18500,3	4588,45	9560
34	2013	62	337	228	31	114	18018,9	18186,3	4547,05	9399
35	2013	63	358	239	34	121	17533,3	17857,9	4502,03	9234
36	2013	64	374	251	38	130	17015,3	17512	4452,56	9060
37	2013	65	403	265	42	141	16473,3	17148,6	4398,19	8871
38	2013	66	438	280	47	152	15894,6	16768,1	4337,69	8668
39	2013	67	460	295	52	162	15269	16368,5	4269,99	8451
40	2013	68	469	309	57	170	14604,7	15943,3	4194,26	8218
41	2013	69	465	323	62	177	13893,3	15474,7	4107,51	7961
42	2013	70	452	330	66	182	13133,2	14946,3	4006,63	7671
43	2013	71	448	340	72	188	12358,7	14380,4	3893,49	7360
44	2013	72	458	363	80	200	11597,4	13803,5	3771,36	7041
45	2013	73	480	400	91	219	10855,5	13215,3	3641,52	6717
46	2013	74	517	451	103	243	10115	12599,2	3501,9	6380
47	2013	75	574	530	125	280	9346,37	11929	3348,09	6018
48	2013	76	627	616	151	320	8506,53	11153,6	3165,3	5609
49	2013	77	656	683	172	351	7592,5	10255,6	2945,48	5142
50	2013	78	660	731	190	373	6634,68	9257,58	2693,57	4630
51	2013	79	639	760	203	386	5664,65	8182,37	2414,48	4082
52	2013	80	619	801	223	404	4717,11	7055,53	2113,69	3509
53	2013	81	615	861	251	431	3788,96	5855,63	1779,26	2904
54	2013	82	611	917	276	456	2857,62	4552,43	1399,3	2252
55	2013	83	607	968	298	479	1927,03	3156,96	979,256	1558
56	2013	84	604	1015	317	500	984,97	1655,78	517,225	816
57										

Εικόνα 7: (αl)x για το έτος 2013

Στο συγκεκριμένο πίνακα υπολογίσαμε το  $(al)_x^j$  το οποίο δηλώνει το αναμενόμενο πλήθος των ατόμων από τους αρχικούς  $(al)_0^T$  που θα αποβιώσουν λόγω του j-αιτίου από την ηλικία x και μετά. Για την πρώτη αιτία στην ηλικία 30 έχουμε 21977,6, για την δεύτερη αιτία έχουμε 21902,7, για την Τρίτη αιτία έχουμε 4965,3 και για την τέταρτη αιτία 12443. Υπολογίζεται από τον τύπο:

$$(al)_x^j = {}_n(ad)_x^j + {}_n(ad)_{x+n}^j + \dots + {}_n(ad)_\omega^j$$



#	A	B	C	D	E	F	G	H	I	J	K
1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	πχ1	πχ2	πχ3	πχ4	πχT
2	2013	30	8	9	1	44	0,21978	0,21903	0,04965	0,12443	0,39311
3	2013	31	11	11	2	47	0,21985	0,21909	0,04968	0,12396	0,39273
4	2013	32	13	12	3	48	0,2199	0,21915	0,04969	0,12349	0,39233
5	2013	33	14	14	3	48	0,21995	0,2192	0,04971	0,12302	0,39193
6	2013	34	14	16	3	46	0,21999	0,21924	0,04972	0,12256	0,39152
7	2013	35	14	18	3	43	0,22003	0,21927	0,04973	0,12212	0,39111
8	2013	36	14	22	3	41	0,22007	0,21925	0,04974	0,12171	0,3907
9	2013	37	15	25	3	41	0,22012	0,2192	0,04975	0,12132	0,39028
10	2013	38	17	28	4	43	0,22016	0,21913	0,04976	0,12095	0,38984
11	2013	39	19	30	4	47	0,2202	0,21903	0,04977	0,12056	0,38936
12	2013	40	22	33	4	50	0,22023	0,21893	0,04979	0,12013	0,38885
13	2013	41	26	38	5	53	0,22026	0,21881	0,0498	0,11968	0,3883
14	2013	42	29	43	5	56	0,22027	0,21868	0,04982	0,11921	0,38771
15	2013	43	33	48	6	60	0,22027	0,21851	0,04984	0,11873	0,38708
16	2013	44	38	54	6	64	0,22026	0,21832	0,04986	0,11822	0,3864
17	2013	45	44	65	8	71	0,22024	0,2181	0,04988	0,11768	0,38565
18	2013	46	53	78	9	78	0,2202	0,21781	0,0499	0,11709	0,38479
19	2013	47	62	88	10	83	0,22014	0,21743	0,04992	0,11643	0,38378
20	2013	48	71	95	11	85	0,22003	0,21698	0,04994	0,11573	0,38265
21	2013	49	79	99	10	85	0,21986	0,21647	0,04997	0,11499	0,38144
22	2013	50	95	104	9	84	0,21959	0,21592	0,05002	0,11425	0,38019
23	2013	51	115	113	8	84	0,21916	0,21534	0,0501	0,11355	0,37899
24	2013	52	132	122	9	86	0,21853	0,21473	0,05021	0,11288	0,37782
25	2013	53	146	131	10	89	0,21776	0,21408	0,05034	0,11222	0,37663
26	2013	54	157	141	12	93	0,21686	0,21337	0,05047	0,11156	0,3754
27	2013	55	173	155	15	99	0,21587	0,21259	0,05059	0,11087	0,37404
28	2013	56	196	171	19	104	0,21473	0,21169	0,05069	0,11014	0,37252
29	2013	57	216	184	22	108	0,21339	0,21069	0,05078	0,10939	0,37086
30	2013	58	235	194	24	110	0,21186	0,20962	0,05085	0,10866	0,36912
31	2013	59	253	200	25	111	0,21008	0,20846	0,05092	0,10791	0,36729
32	2013	60	278	207	26	111	0,20801	0,20722	0,05099	0,10714	0,36536
33	2013	61	310	217	29	111	0,20559	0,20595	0,05108	0,10642	0,36346
34	2013	62	337	228	31	114	0,20277	0,20466	0,05117	0,10577	0,3616
35	2013	63	358	239	34	121	0,19961	0,2033	0,05125	0,10512	0,35968
36	2013	64	374	251	38	130	0,19614	0,20187	0,05133	0,10443	0,35762
37	2013	65	403	265	42	141	0,19244	0,20033	0,05138	0,10363	0,35534
38	2013	66	438	280	47	152	0,18837	0,19872	0,05141	0,10273	0,35286
39	2013	67	460	295	52	162	0,18381	0,19704	0,0514	0,10174	0,35018
40	2013	68	469	309	57	170	0,17882	0,19521	0,05136	0,10063	0,34719
41	2013	69	465	323	62	177	0,17335	0,19308	0,05125	0,09933	0,34365
42	2013	70	452	330	66	182	0,16737	0,19047	0,05106	0,09776	0,3393
43	2013	71	448	340	72	188	0,16112	0,18748	0,05076	0,09596	0,33419
44	2013	72	458	363	80	200	0,15479	0,18423	0,05034	0,09398	0,32854
45	2013	73	480	400	91	219	0,14842	0,18068	0,04979	0,09183	0,32231
46	2013	74	517	451	103	243	0,14185	0,17669	0,04911	0,08947	0,31527
47	2013	75	574	530	125	280	0,13476	0,172	0,04828	0,08678	0,30706
48	2013	76	627	616	151	320	0,12669	0,16611	0,04714	0,08353	0,29678
49	2013	77	656	683	172	351	0,11744	0,15864	0,04556	0,07955	0,28375
50	2013	78	660	731	190	373	0,10714	0,14949	0,0435	0,07476	0,26775
51	2013	79	639	760	203	386	0,09592	0,13856	0,04089	0,06912	0,24856
52	2013	80	619	801	223	404	0,08407	0,12575	0,03767	0,06254	0,22597
53	2013	81	615	861	251	431	0,07144	0,1104	0,03355	0,05474	0,19869
54	2013	82	611	917	276	456	0,05741	0,09147	0,02811	0,04524	0,16482
55	2013	83	607	968	298	479	0,04159	0,06814	0,02114	0,03363	0,1229
56	2013	84	604	1015	317	500	0,02308	0,03879	0,01212	0,01911	0,07002
57											

Εικόνα 8: (πχ) πιθανότητα αποβίωσης εξαρτημένη

Σε αυτό το σημείο, έχουμε υπολογίσει τη δεσμευμένη πιθανότητα αποβίωσης, την εξαρτημένη μεταβλητή, μετά την ηλικία  $x$  και αυτό οφείλεται στο  $j$ -αίτιο, δεδομένου επιβίωσης μέχρι την ηλικία  $x$ . Παρατηρούμε και στις 4 αιτίες θανάτων ότι μικραίνει η πιθανότητα αποβίωσης στις μεγαλύτερες ηλικίες (79+). Το υπολογίσαμε με βάση τον τύπο:

$$\pi_x^j = \frac{(al)_x^j}{(al)_x^T}$$

	A	B	C	D	E	F	G	H	I	J
1	ΕΤΟΣ	ΗΛΙΚΙΕΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	qx1	qx2	qx3	qx4
2	2013	30	8	9	1	44	0,0001	0,00011	1,6E-05	0,00057
3	2013	31	11	11	2	47	0,00014	0,00013	2,7E-05	0,00058
4	2013	32	13	12	3	48	0,00016	0,00015	3,3E-05	0,00059
5	2013	33	14	14	3	48	0,00017	0,00017	3,6E-05	0,00058
6	2013	34	14	16	3	46	0,00017	0,00019	3,5E-05	0,00056
7	2013	35	14	18	3	43	0,00017	0,00022	3,5E-05	0,00053
8	2013	36	14	22	3	41	0,00017	0,00026	3,8E-05	0,0005
9	2013	37	15	25	3	41	0,00018	0,0003	4,1E-05	0,0005
0	2013	38	17	28	4	43	0,00021	0,00034	4,4E-05	0,00053
1	2013	39	19	30	4	47	0,00024	0,00037	4,8E-05	0,00057
2	2013	40	22	33	4	50	0,00027	0,00041	5,2E-05	0,00062
3	2013	41	26	38	5	53	0,00031	0,00046	5,6E-05	0,00064
4	2013	42	29	43	5	56	0,00035	0,00051	6,1E-05	0,00068
5	2013	43	33	48	6	60	0,0004	0,00058	6,8E-05	0,00072
6	2013	44	38	54	6	64	0,00045	0,00065	7,7E-05	0,00077
7	2013	45	44	65	8	71	0,00054	0,00078	9,4E-05	0,00086
8	2013	46	53	78	9	78	0,00066	0,00096	0,00012	0,00097
9	2013	47	62	88	10	83	0,00079	0,00113	0,00013	0,00107
0	2013	48	71	95	11	85	0,00094	0,00127	0,00014	0,00114
1	2013	49	79	99	10	85	0,0011	0,00137	0,00014	0,00117
2	2013	50	95	104	9	84	0,00132	0,00146	0,00012	0,00117
3	2013	51	115	113	8	84	0,0016	0,00157	0,00011	0,00118
4	2013	52	132	122	9	86	0,00183	0,00169	0,00012	0,0012
5	2013	53	146	131	10	89	0,00204	0,00183	0,00014	0,00125
6	2013	54	157	141	12	93	0,00223	0,002	0,00017	0,00133
7	2013	55	173	155	15	99	0,00252	0,00226	0,00022	0,00144
8	2013	56	196	171	19	104	0,00289	0,00253	0,00028	0,00154
9	2013	57	216	184	22	108	0,00319	0,00272	0,00033	0,00159
0	2013	58	235	194	24	110	0,00358	0,00295	0,00037	0,00168
1	2013	59	253	200	25	111	0,00403	0,00319	0,0004	0,00178
2	2013	60	278	207	26	111	0,0045	0,00335	0,00043	0,00179
3	2013	61	310	217	29	111	0,00501	0,00351	0,00046	0,0018
4	2013	62	337	228	31	114	0,00548	0,00371	0,00051	0,00187
5	2013	63	358	239	34	121	0,00592	0,00395	0,00057	0,002
6	2013	64	374	251	38	130	0,00627	0,00421	0,00063	0,00218
7	2013	65	403	265	42	141	0,00679	0,00447	0,00071	0,00239
8	2013	66	438	280	47	152	0,00744	0,00476	0,00081	0,00259
9	2013	67	460	295	52	162	0,00803	0,00515	0,00092	0,00283
0	2013	68	469	309	57	170	0,00875	0,00577	0,00107	0,00318
1	2013	69	465	323	62	177	0,00954	0,00664	0,00127	0,00364
2	2013	70	452	330	66	182	0,00993	0,00727	0,00146	0,004
3	2013	71	448	340	72	188	0,00999	0,00758	0,00161	0,0042
4	2013	72	458	363	80	200	0,00997	0,00791	0,00175	0,00437
5	2013	73	480	400	91	219	0,0102	0,00849	0,00193	0,00465
6	2013	74	517	451	103	243	0,01087	0,00948	0,00218	0,00513
7	2013	75	574	530	125	280	0,01223	0,0113	0,00267	0,00599
8	2013	76	627	616	151	320	0,01378	0,01354	0,00333	0,00705
9	2013	77	656	683	172	351	0,01502	0,01565	0,00397	0,00807
0	2013	78	660	731	190	373	0,01591	0,01762	0,0046	0,00903
1	2013	79	639	760	203	386	0,01632	0,01938	0,00521	0,00989
2	2013	80	619	801	223	404	0,01687	0,02175	0,00611	0,01104
3	2013	81	615	861	251	431	0,01796	0,02504	0,00737	0,01261
4	2013	82	611	917	276	456	0,01919	0,02864	0,00871	0,01434
5	2013	83	607	968	298	479	0,02095	0,03318	0,01033	0,01655
6	2013	84	604	1015	317	500	0,02393	0,0399	0,01264	0,01986
7										

Εικόνα 9: (qx) ανεξάρτητη πιθανότητα αποβίωσης

Στον παραπάνω πίνακα υπολογίσαμε την ανεξάρτητη μεταβλητή της πιθανότητας αποβίωσης, προκειμένου να τη χρησιμοποιήσουμε ώστε να βρούμε την ανεξάρτητη μεταβλητή της συνάρτησης επιβίωσης. Το υπολογίσαμε από τον τύπο:



$$q_x^j \cong 1 - [1 - (aq)_x^T]^{[(aq)_x^j / (aq)_x^T]}$$

Α	B	C	D	E	F	G	H	I	J	
1	ΕΤΟΣ	ΗΛΙΚΙΩΣ	ΚΑΡΚΙΝΟΣ	ΙΣΧ. ΚΑΡΔΙΟΠΑΘΕΙΑ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	px1	px2	px3	px4
2	2013	30	8	9	1	44	0,9999	0,99989	0,99998	0,99943
3	2013	31	11	11	2	47	0,99986	0,99987	0,99997	0,99942
4	2013	32	13	12	3	48	0,99984	0,99985	0,99997	0,99941
5	2013	33	14	14	3	48	0,99983	0,99983	0,99996	0,99942
6	2013	34	14	16	3	46	0,99983	0,99981	0,99996	0,99944
7	2013	35	14	18	3	43	0,99983	0,99978	0,99996	0,99947
8	2013	36	14	22	3	41	0,99983	0,99974	0,99996	0,9995
9	2013	37	15	25	3	41	0,99982	0,9997	0,99996	0,9995
10	2013	38	17	28	4	43	0,99979	0,99966	0,99996	0,99947
11	2013	39	19	30	4	47	0,99976	0,99963	0,99995	0,99943
12	2013	40	22	33	4	50	0,99973	0,99959	0,99995	0,99938
13	2013	41	26	38	5	53	0,99969	0,99954	0,99994	0,99936
14	2013	42	29	43	5	56	0,99965	0,99949	0,99994	0,99932
15	2013	43	33	48	6	60	0,9996	0,99942	0,99993	0,99928
16	2013	44	38	54	6	64	0,99955	0,99935	0,99992	0,99923
17	2013	45	44	65	8	71	0,99946	0,99922	0,99991	0,99914
18	2013	46	53	78	9	78	0,99934	0,99904	0,99988	0,99903
19	2013	47	62	88	10	83	0,99921	0,99887	0,99987	0,99893
20	2013	48	71	95	11	85	0,99906	0,99873	0,99986	0,99886
21	2013	49	79	99	10	85	0,99899	0,99863	0,99986	0,99883
22	2013	50	95	104	9	84	0,99868	0,99854	0,99988	0,99883
23	2013	51	115	113	8	84	0,9984	0,99843	0,99989	0,99882
24	2013	52	132	122	9	86	0,99817	0,99831	0,99988	0,9988
25	2013	53	146	131	10	89	0,99796	0,99817	0,99986	0,99875
26	2013	54	157	141	12	93	0,99777	0,998	0,99983	0,99867
27	2013	55	173	155	15	99	0,99748	0,99774	0,99978	0,99856
28	2013	56	196	171	19	104	0,99711	0,99747	0,99972	0,99846
29	2013	57	216	184	22	108	0,99681	0,99728	0,99967	0,99841
30	2013	58	235	194	24	110	0,99642	0,99705	0,99963	0,99832
31	2013	59	253	200	25	111	0,99597	0,99681	0,9996	0,99822
32	2013	60	278	207	26	111	0,9955	0,99665	0,99957	0,99821
33	2013	61	310	217	29	111	0,99499	0,99649	0,99954	0,9982
34	2013	62	337	228	31	114	0,99452	0,99629	0,99949	0,99813
35	2013	63	358	239	34	121	0,99408	0,99605	0,99943	0,998
36	2013	64	374	251	38	130	0,99373	0,99579	0,99937	0,99782
37	2013	65	403	265	42	141	0,99321	0,99553	0,99929	0,99761
38	2013	66	438	280	47	152	0,99256	0,99524	0,99919	0,99741
39	2013	67	460	295	52	162	0,99197	0,99485	0,99908	0,99717
40	2013	68	469	309	57	170	0,99125	0,99423	0,99893	0,99682
41	2013	69	465	323	62	177	0,99046	0,99336	0,99873	0,99636
42	2013	70	452	330	66	182	0,99007	0,99273	0,99854	0,996
43	2013	71	448	340	72	188	0,99001	0,99242	0,99839	0,9958
44	2013	72	458	363	80	200	0,99003	0,99209	0,99825	0,99563
45	2013	73	480	400	91	219	0,98998	0,99151	0,99807	0,99535
46	2013	74	517	451	103	243	0,98913	0,99052	0,99782	0,99487
47	2013	75	574	530	125	280	0,98777	0,9887	0,99733	0,99401
48	2013	76	627	616	151	320	0,98622	0,98646	0,99667	0,99295
49	2013	77	656	683	172	351	0,98498	0,98435	0,99603	0,99193
50	2013	78	660	731	190	373	0,98409	0,98238	0,9954	0,99097
51	2013	79	639	760	203	386	0,98368	0,98062	0,99479	0,99011
52	2013	80	619	801	223	404	0,98313	0,97825	0,99389	0,98896
53	2013	81	615	861	251	431	0,98204	0,97496	0,99263	0,98739
54	2013	82	611	917	276	456	0,98081	0,97136	0,99129	0,98566
55	2013	83	607	968	298	479	0,97905	0,96682	0,98967	0,98345
56	2013	84	604	1015	317	500	0,97607	0,9601	0,98736	0,98014
57										

Εικόνα 10: (px) ανεξάρτητη πιθανότητα επιβίωσης

Εδώ υπολογίσαμε την ανεξάρτητη πιθανότητα επιβίωσης για να καταλήξουμε στη συνάρτηση επιβίωσης και το βρήκαμε από τον παρακάτω τύπο:

$$p_x^j = 1 - q_x^j$$

	A	B	C	D	E	F	G	H	I	J	K
1	ΕΤΟΣ	ΗΛΙΚΙΑ	ΚΑΡΚΙΝΟΣ	ΙΣΧΑΙΜΙΚΗ ΚΑΡΔ	ΑΝΑΠΝΕΥΣΤΙΚΑ	ΆΛΛΕΣ ΑΙΤΙΕΣ	St(x1)	St(x2)	St(x3)	St(x4)	St(x0)
2	2013	30	8	9	1	44	1	1	1	1	1
3	2013	31	11	11	2	47	0,9999	0,99989	0,99998	0,99943	0,99931
4	2013	32	13	12	3	48	0,99976	0,99976	0,99996	0,99885	0,99856
5	2013	33	14	14	3	48	0,9996	0,99961	0,99992	0,99827	0,9978
6	2013	34	14	16	3	46	0,99943	0,99944	0,99989	0,99769	0,99702
7	2013	35	14	18	3	43	0,99926	0,99925	0,99985	0,99713	0,99624
8	2013	36	14	22	3	41	0,9991	0,99903	0,99982	0,99661	0,99546
9	2013	37	15	25	3	41	0,99893	0,99877	0,99978	0,99611	0,99467
0	2013	38	17	28	4	43	0,99875	0,99847	0,99974	0,99562	0,99383
1	2013	39	19	30	4	47	0,99854	0,99813	0,9997	0,99509	0,99293
2	2013	40	22	33	4	50	0,99831	0,99776	0,99965	0,99453	0,99195
3	2013	41	26	38	5	53	0,99804	0,99735	0,99959	0,99391	0,99088
4	2013	42	29	43	5	56	0,99773	0,9969	0,99954	0,99328	0,98974
5	2013	43	33	48	6	60	0,99737	0,99638	0,99948	0,9926	0,9885
6	2013	44	38	54	6	64	0,99698	0,99581	0,99941	0,99189	0,98715
7	2013	45	44	65	8	71	0,99652	0,99516	0,99933	0,99113	0,98567
8	2013	46	53	78	9	78	0,99599	0,99438	0,99924	0,99028	0,98396
9	2013	47	62	88	10	83	0,99533	0,99342	0,99912	0,98931	0,98194
0	2013	48	71	95	11	85	0,99454	0,9923	0,99899	0,98826	0,97966
1	2013	49	79	99	10	85	0,9936	0,99105	0,99885	0,98714	0,97717
2	2013	50	95	104	9	84	0,99252	0,98969	0,99871	0,98598	0,97456
3	2013	51	115	113	8	84	0,9912	0,98825	0,99859	0,98483	0,97188
4	2013	52	132	122	9	86	0,98962	0,9867	0,99847	0,98367	0,9691
5	2013	53	146	131	10	89	0,9878	0,98503	0,99835	0,98249	0,96619
6	2013	54	157	141	12	93	0,98579	0,98323	0,99822	0,98126	0,96308
7	2013	55	173	155	15	99	0,98359	0,98126	0,99805	0,97996	0,95972
8	2013	56	196	171	19	104	0,98111	0,97904	0,99783	0,97854	0,95595
9	2013	57	216	184	22	108	0,97827	0,97656	0,99755	0,97704	0,9518
0	2013	58	235	194	24	110	0,97515	0,97391	0,99722	0,97549	0,9474
1	2013	59	253	200	25	111	0,97166	0,97104	0,99685	0,97385	0,94267
2	2013	60	278	207	26	111	0,96775	0,96794	0,99645	0,97212	0,93762
3	2013	61	310	217	29	111	0,96339	0,96471	0,99603	0,97038	0,93241
4	2013	62	337	228	31	114	0,95857	0,96132	0,99556	0,96863	0,92704
5	2013	63	358	239	34	121	0,95331	0,95775	0,99506	0,96683	0,9214
6	2013	64	374	251	38	130	0,94767	0,95397	0,99449	0,96489	0,91541
7	2013	65	403	265	42	141	0,94173	0,94995	0,99387	0,96279	0,90899
8	2013	66	438	280	47	152	0,93534	0,94571	0,99316	0,96049	0,90213
9	2013	67	460	295	52	162	0,92838	0,94121	0,99236	0,95801	0,89479
0	2013	68	469	309	57	170	0,92092	0,93636	0,99144	0,9553	0,88685
1	2013	69	465	323	62	177	0,91286	0,93095	0,99038	0,95226	0,87799
2	2013	70	452	330	66	182	0,90415	0,92477	0,98912	0,94879	0,86787
3	2013	71	448	340	72	188	0,89517	0,91805	0,98768	0,945	0,85687
4	2013	72	458	363	80	200	0,88623	0,91109	0,98609	0,94103	0,84544
5	2013	73	480	400	91	219	0,87739	0,90388	0,98436	0,93691	0,83362
6	2013	74	517	451	103	243	0,86844	0,8962	0,98246	0,93255	0,8211
7	2013	75	574	530	125	280	0,859	0,8877	0,98032	0,92777	0,80737
8	2013	76	627	616	151	320	0,84849	0,87767	0,97769	0,92222	0,79135
9	2013	77	656	683	172	351	0,8368	0,86579	0,97444	0,91572	0,77255
0	2013	78	660	731	190	373	0,82423	0,85224	0,97057	0,90833	0,75133
1	2013	79	639	760	203	386	0,81112	0,83723	0,9661	0,90013	0,72806
2	2013	80	619	801	223	404	0,79788	0,821	0,96106	0,89123	0,7032
3	2013	81	615	861	251	431	0,78442	0,80314	0,95519	0,88139	0,67616
4	2013	82	611	917	276	456	0,77033	0,78303	0,94815	0,87027	0,64612
5	2013	83	607	968	298	479	0,75555	0,7606	0,9399	0,8578	0,61323
6	2013	84	604	1015	317	500	0,73972	0,73536	0,93019	0,8436	0,57705
7							0	0	0	0	0

Εικόνα 11: (St\_x) Συνάρτηση επιβίωσης (ανεξάρτητη)

Παρατηρούμε στον παραπάνω πίνακα την ανεξάρτητη συνάρτηση επιβίωσης των τεσσάρων αιτιών θανάτου, δηλαδή από τα κακοήθη νεοπλασμάτα, την ισχαιμική καρδιοπάθεια, τα αναπνευστικά προβλήματα και όλες τις υπόλοιπες αιτίες. Φαίνεται στον παραπάνω πίνακα ότι άτομα μεγαλύτερης ηλικίας έχουν μικρότερη πιθανότητα επιβίωσης. Όλα τα παραπάνω τα υπολογίσαμε με βάση την UDD μέθοδο (ομοιόμορφη κατανομή των απαυξημάτων στον πίνακα με πολλαπλά αίτια εξόδων).

## ΚΕΦΑΛΑΙΟ 6: Συσχέτιση των δεδομένων

Σε αυτό το κεφάλαιο, θα αναλύσουμε τα πιο διαδεδομένα μοντέλα που χρησιμοποιούνται για την μέτρηση της θνησιμότητας καθώς και τη συσχέτιση των Copulas με τις βασικές αιτίες θανάτων.

### Μοντέλο Hatzopoulos & Haberman (2009)

**GLM & Poisson κατανομή:** Ας υποθέσουμε την ύπαρξη μίας ομάδας ατόμων με την ίδια ηλικία. Σύμφωνα με τον Forfar et. al. (1988), εάν  $\Theta$  ορίζει τον αριθμό των θανάτων και  $R^C$  την κεντρική έκθεση στον κίνδυνο, με το  $\Theta$  (αλλά όχι το  $R^C$ ) να μοντελοποιείται σαν μια τυχαία μεταβλητή, τότε ο αριθμός των θανάτων ακολουθεί μια Poisson κατανομή με μέσο και διακύμανση  $R^C \mu$  αντίστοιχα, όπου το  $\mu$  ορίζει την ένταση θνησιμότητας. Ισχύει λοιπόν

$$\Theta \sim (R^C \mu)$$

Αυτό μπορεί να θεωρηθεί μια διαδικασία Poisson, κατά την οποία ο αριθμός των θανάτων ακολουθεί Poisson κατανομή με ένταση  $\mu$  (την ένταση θνησιμότητας). Οι τιμές  $\Theta$  και  $R^C$  είναι ελάχιστης στατιστικής ικανότητας για το  $\mu$ . Ως εκ τούτου είναι φυσικό να βασίζονται όλα τα στατιστικά συμπεράσματα σε αυτές τις δυο ποσότητες (Sverdrup 1965). Υποθέτουμε ότι η ένταση θνησιμότητας  $\mu$  είναι εν μέρει σταθερή εντός κάθε ηλικιακής κατηγορίας και περιόδου μελέτης έτσι ώστε ο ρυθμός  $\{\theta / R^C\}$  να αποτελεί τον εκτιμητή μέγιστης πιθανοφάνειας για το  $\mu$ .

Εκφράζοντάς το σαν ένα GLM που βασίζεται στις ανεξάρτητες τυχαίες μεταβλητές  $\{\Theta_X\}$ , παίρνουμε

$$\{\Theta_X\} = m_X = R_X^C \mu_X \quad \text{και} \quad \text{Var}(\Theta_X) = m_X$$

Με scale παράμετρο  $\phi = 1$ , αρχικά βάρη  $\omega_X = 1$  και συνάρτηση διακύμανσης

$$(m_X) = m_X$$

Για ευκολία, θα χρησιμοποιήσουμε το  $\mu_X$  ως σταθερή τιμή της έντασης θνησιμότητας σε όλο το ηλικιακό φάσμα παρά το  $\mu_{X+\frac{1}{2}}$ . Για αυτή τη συγκεκριμένη περίπτωση παίρνουμε την παρακάτω έκφραση του scale deviance,

$$S(c, f) = 2 \sum_x \left\{ y_x \log \left( \frac{y_x}{\hat{m}_x} \right) \right\} - (y_x - \hat{m}_x)$$

Όπου το  $y_x$  ορίζει τις παρατηρούμενες αποκρίσεις του  $\Theta_x$  και το  $\hat{m}_x$  ορίζει τις προσαρμοσμένες τιμές  $R_x^C \hat{\mu}_x$  για το συγκεκριμένο μοντέλο. Δηλαδή, η παραπάνω έκφραση για το scale deviance μπορεί να γραφτεί

$$S(c, f) = 2 \sum_x \left\{ \theta_x \log \left( \frac{\theta_x}{R_x^C \hat{\mu}_x} \right) \right\} - (\theta_x - R_x^C \hat{\mu}_x)$$

Ο Renshaw (1991), περιγράφει την εφαρμογή της  $\mu_x$ - εξομάλυνσης στο GLIM που βασίζεται στις υποθέσεις που σχετίζονται με τις κατανομές, σε συνδυασμό με την χρήση της φόρμουλας εκτιμητή log-link που έχει τη μορφή

$$\log(m_x) = \log(R_x^C) + \log(\mu_x) = \log(R_x^C) + \sum_j \beta_j x^j$$

Στην οποία το  $\log(R_x^C)$  συμπεριφέρεται σαν ένα offset. Σημειώνουμε ότι η φόρμουλα εξομάλυνσης

$$\log(\hat{m}_x(t)) = \sum_{j=1}^k \beta_{j-1}(t) L_{j-1}(x)$$

Δηλώνει ότι η ένταση θνησιμότητας  $\mu_x$  μοντελοποιείται σαν ένα εκθετικό πολυώνυμο στην ηλικία  $x$ .

Για κάθε έτος ανεξάρτητα, προσαρμόζω ένα ορθοκανονικό πολυώνυμο κ-βαθμού (Legendre). Εκτιμώ κ-GLM παραμέτρους για κάθε έτος μέγιστης πιθανοφάνειας, συνεχίζοντας με αυτό τον τρόπο για όλα τα έτη. Σχηματίζω έναν πίνακα με τις B-GLM παραμέτρους και σε αυτόν τον πίνακα εφαρμόζω την προσέγγιση σύμφωνα με το Hatzopoulos & Haberman (2011) στον οποίο κάνω principal component ανάλυση (SPCA) με σκοπό να εκτιμήσουμε τις B-GLM παραμέτρους. Έτσι, καταλήγω στο μοντέλο σύμφωνα με το Hatzopoulos & Haberman (2009) το οποίο είναι το εξής:

$$\log(\hat{m}_x(t)) = A(x) + \sum_{i=1}^p g_i(x) Y_i(t) + \varepsilon_x(t)$$

Όπου το  $(x)$  είναι ένα σταθερό σύνολο ηλικιών, που περιγράφει την θνησιμότητα από την ηλικία. Το  $g_i(x)$  περιγράφει την σπουδαιότητα των ηλικιών  $x$  σε απάντηση των τάσεων κατά τη μεταβολή του χρόνου. Τα  $g_i(x)$  και  $Y_i(t)$  εκφράζουν τις αλληλεπιδράσεις μεταξύ της ηλικίας και του χρόνου.

## Appendix

Παρακάτω παρατίθεται ο κώδικας που χρησιμοποιήθηκε στο matlab για την εφαρμογή των δεδομένων των πινάκων με συναρτήσεις Copulas.

```
clear all
clc

n=84;

data = xlsread('Dedomena'); %eisagw ta dedomena
x = data(:,2); %ilikies
dx_c = data(:,3); %aitia thanatou karkinos
dx2 = data(:,4); %aitia thanatou kardiagkeiakas pathiseis
dx3 = data(:,5); %aitia thanatou
dx4 = data(:,6); %oles oi ipoloipes aities thanatwn
ex = data(:,7); %kentriki ekthesi ston kindino

mx1 = dx_c./ex; %kentrikos rithmos thnisimotitas gia karkino
mx2 = dx2./ex; %kentrikos rithmos thnisimotitas gia
kardiagkeiakas pathiseis
mx3 = dx3./ex; %kentrikos rithmos
mx4 = dx4./ex; %kentrikos rithmos gia tis ipoloipes aities

logmx1 = log(mx1);
logmx2 = log(mx2);
logmx3 = log(mx3);
logmx4 = log(mx4);

px_1 = data(:,30); %pithanotites apoviwsis tvn eksartimenwn
metavlitwn
px_2 = data(:,31);
px_3 = data(:,32);
px_4 = data(:,33);

figure;
plot(x,px_1,'-',x,px_2,'-.',x,px_3,'*',x,px_4,'.')
ylabel('pithanotita apoviwsis')
xlabel('ages')

px_o = data(:,34);
figure;
plot(x,px_o,'-',x,px_1,'-.')
ylabel('st_x')
xlabel('ages')

st_x1 = data(:,43); %sinartiseis epiviwsis twn aneksartitwn
metavlitwn
st_x2 = data(:,44);
st_x3 = data(:,45);
```

```

st_x4 = data(:,46);

figure;
plot(x,st_x1,'-',x,st_x2,'-.',x,st_x3,'*',x,st_x4,'.')
ylabel('st_x')
xlabel('ages')

l_0 = 100000 %arithmos epizwntwn se akrivi ilikia gia ton
karkino

qx_c = data(:,12);
qx_overall = data(:,16);

% pairnoute to st_x1 kai to st_x2
w = ksdensity(st_x1,st_x1,'function','cdf')
z = ksdensity(st_x2,st_x2,'function','cdf')

figure;
scatterhist(w,z)
xlabel('w')
ylabel('z')

u = rand(1,6) %to U einai enas pinakas me 1 grammi kai oses
stiles theloume me times sto [0,1]
RHOHAT = copulafit('Gaussian',u) %sto rhohat epistrefei mia
ektimisi tou pinaka tis grammikis sixxetisis gia ena gaussian
copula

rng default
[Rho, nu] = copulafit('t',[w,z],'Method','ApproximateML');

[rhoha,nuhat,nuci] =
copulafit('t',[w,z],'Method','ApproximateML')

PARAMHAT = copulafit('Clayton',[w,z])
par = copulafit('Frank',[w,z])

%pairnoute to st_x1 me to st_x3
k = ksdensity(st_x1,st_x1,'function','cdf')
l = ksdensity(st_x3,st_x3,'function','cdf')

figure;
scatterhist(k,l)
xlabel('k')
ylabel('l')

u = rand(1,6) %to U einai enas pinakas me 1 grammi kai oses
stiles theloume me times sto [0,1]
RHOHAT = copulafit('Gaussian',u) %sto rhohat epistrefei mia
ektimisi tou pinaka tis grammikis sixxetisis gia ena gaussian
copula

```

```

rng default
[Rho, nu] = copulafit('t',[k,l],'Method','ApproximateML');

[rhoha,nuhat,nuci] =
copulafit('t',[k,l],'Method','ApproximateML')

PARAMHAT = copulafit('Clayton',[k,l])
par = copulafit('Frank',[k,l])

%pairnومه to st_x1 kai to st_x4
p = ksdensity(st_x1,st_x1,'function','cdf')
r = ksdensity(st_x4,st_x4,'function','cdf')

figure;
scatterhist(p,r)
xlabel('r')
ylabel('p')

u = rand(1,6) %to U einai enas pinakas me 1 grammi kai oses
stiles theloume me times sto [0,1]
RHOHAT = copulafit('Gaussian',u) %sto rhohat epistrefei mia
ektimisi tou pinaka tis grammikis sixxetisis gia ena gaussian
copula

rng default
[Rho, nu] = copulafit('t',[p,r],'Method','ApproximateML');

[rhoha,nuhat,nuci] =
copulafit('t',[p,r],'Method','ApproximateML')

PARAMHAT = copulafit('Clayton',[p,r])
par = copulafit('Frank',[p,r])

%gia thn apo koinou sinartisi epiviwsis
st_xo = data(:,47);

figure;
plot(x,st_x1,'-',x,st_xo,'.')
ylabel('sinartisi epiviosis')
xlabel('ages')

m = ksdensity(st_x1,st_x1,'function','cdf')
n = ksdensity(st_xo,st_xo,'function','cdf')

figure;
scatterhist(m,n)
xlabel('m')
ylabel('n')

u = rand(1,6) %to U einai enas pinakas me 1 grammi kai oses
stiles theloume me times sto [0,1]
RHOHAT = copulafit('Gaussian',u) %sto rhohat epistrefei mia
ektimisi tou pinaka tis grammikis sixxetisis gia ena gaussian

```

```
copula
```

```
rng default
```

```
[Rho, nu] = copulafit('t',[m,n],'Method','ApproximateML');
```

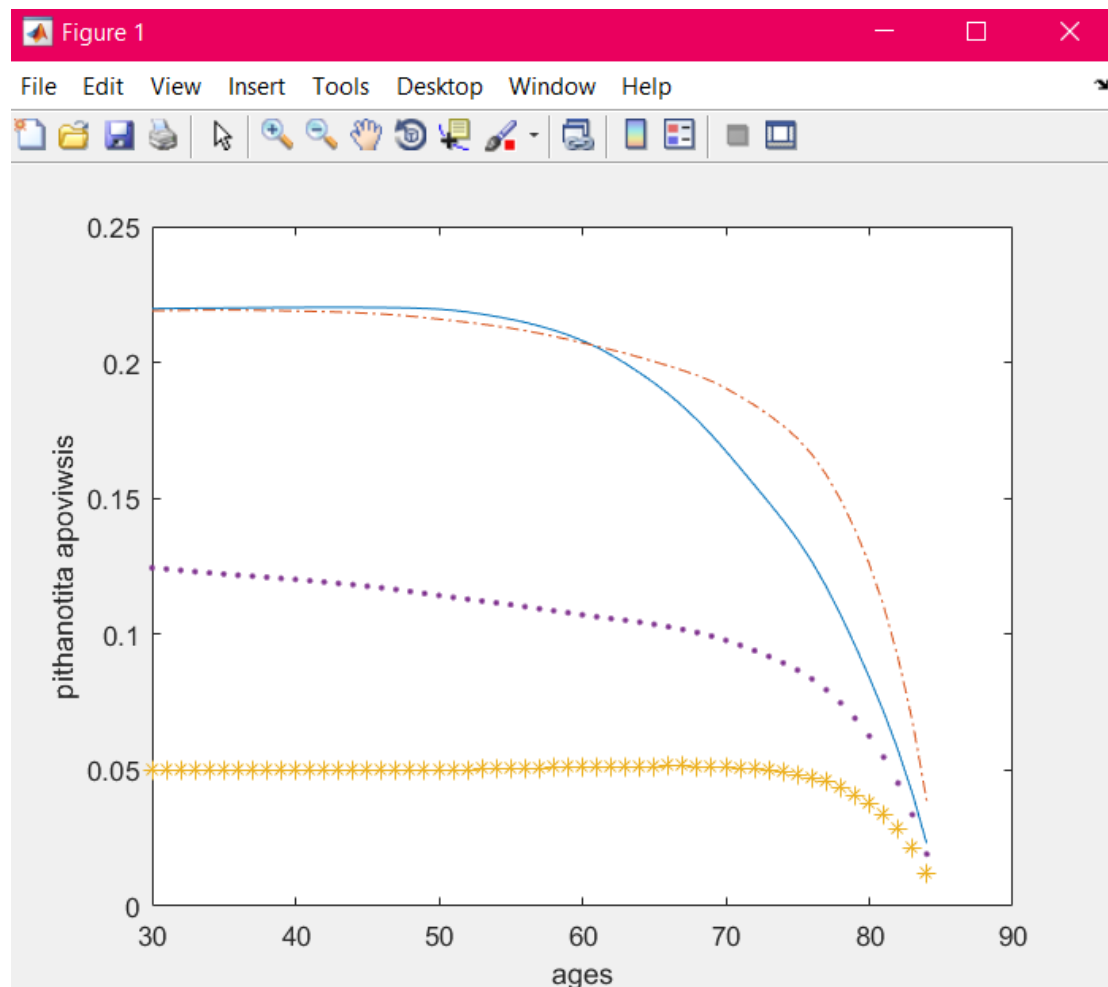
```
[rhoa,nuhat,nuci] =
```

```
copulafit('t',[m,n],'Method','ApproximateML')
```

```
PARAMHAT = copulafit('Clayton',[m,n])
```

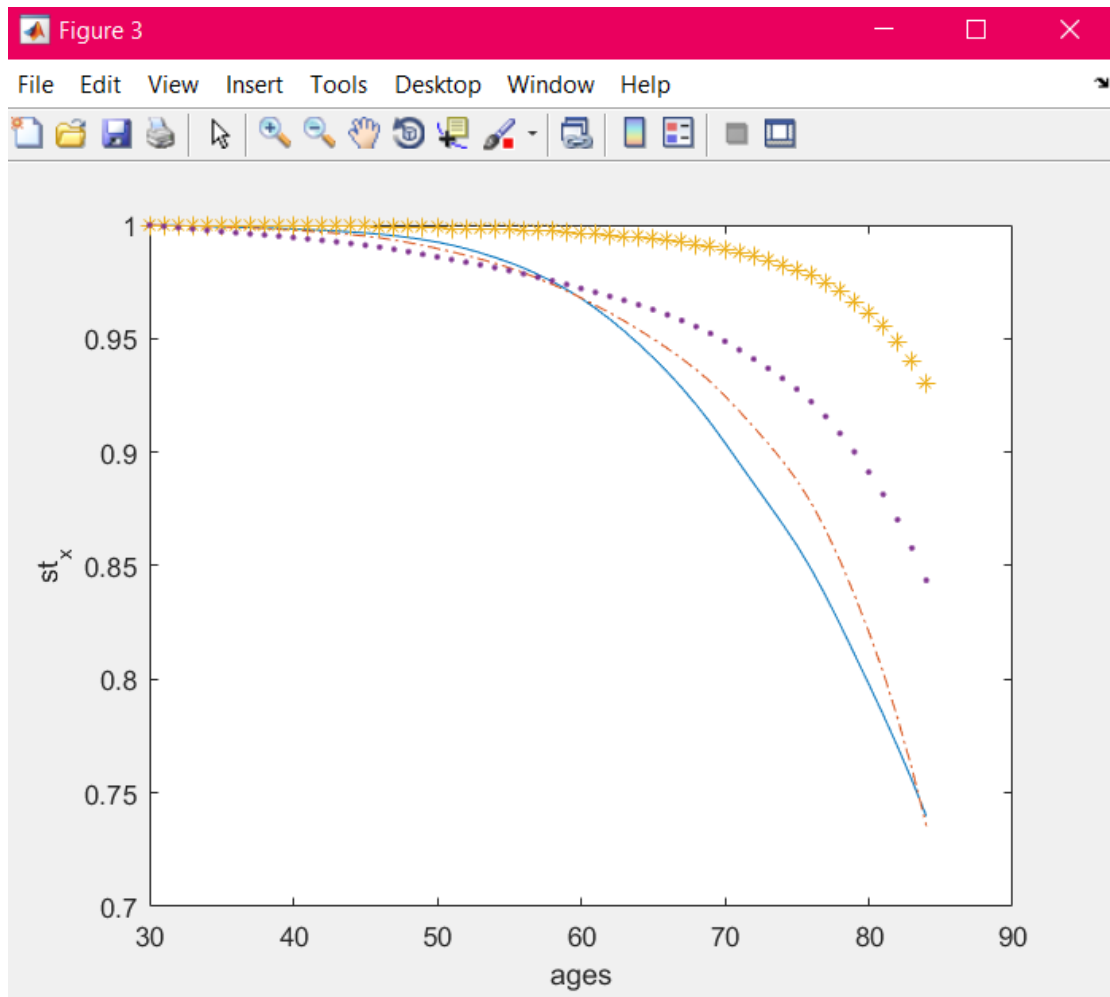
```
par = copulafit('Frank',[m,n])
```

Τα αποτελέσματα που προκύπτουν φαίνονται στα παρακάτω διαγράμματα:

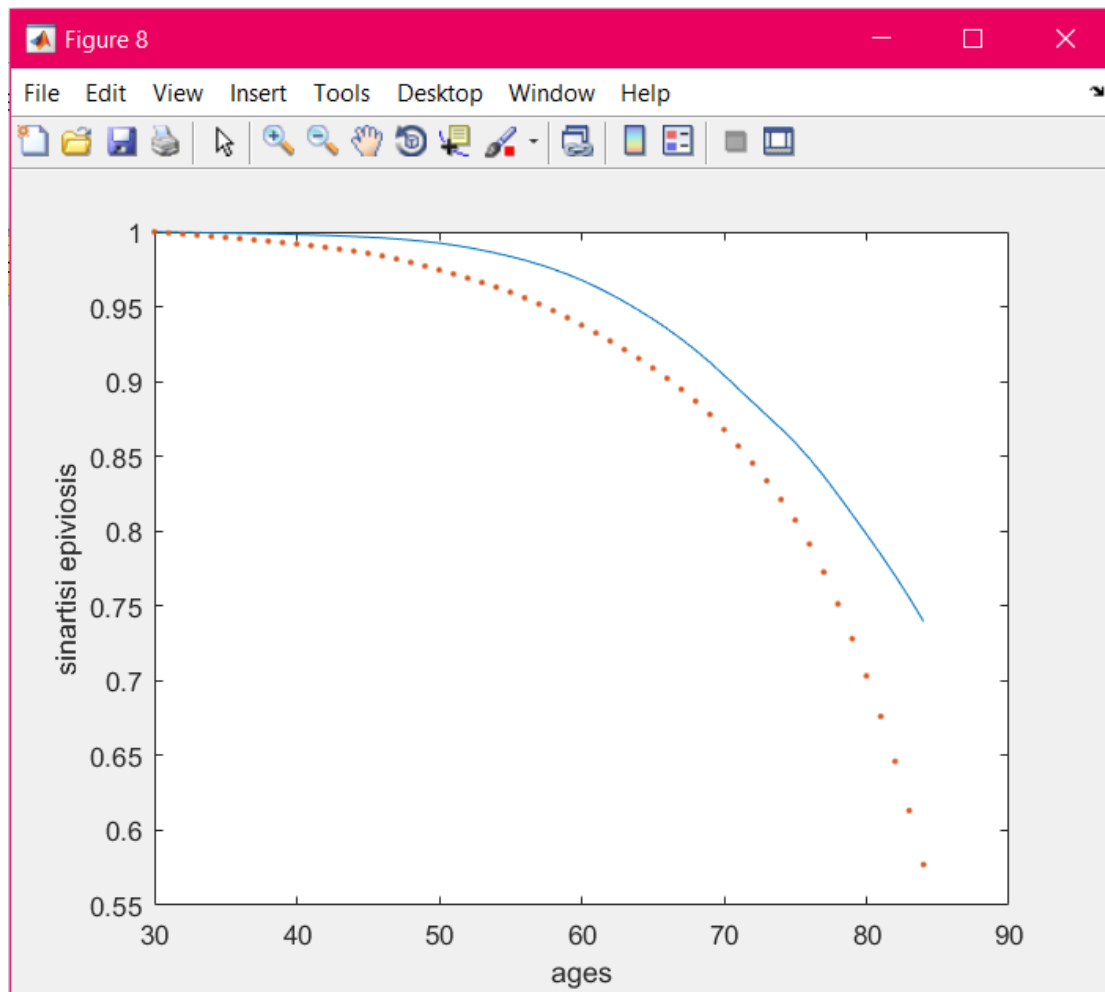




Στο οποίο βλέπουμε τις συναρτήσεις αποβίωσης των τεσσάρων αιτιών θανάτου.



Εδώ παρατηρούμε τις συναρτήσεις επιβίωσης των ανεξάρτητων μεταβλητών.



Τέλος βλέπουμε στο παραπάνω διάγραμμα τις συναρτήσεις επιβίωσης από κακοήθη νεοπλασμάτα και όλες τις υπόλοιπες αιτίες μαζί.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

1. Πέτρος Φ. Χατζόπουλος , «Μαθηματικά Ασφαλίσεων Ζωής»  
ΕΚΔΟΣΕΙΣ ΣΥΜΜΕΤΡΙΑ
2. Πτυχιακή εργασία φοιτήτριας Αθανασίας Ραγκούση «Μοντελοποίηση  
Ανιάτων Ασθενειών Στο Σύνολο Της Ελλάδας και Πρόβλεψη»
3. «Dependent Competing Risk: Cause elimination and its impact on  
survival» Dimitrina S. Dimitrova, Steven Haberman and Vladimir K.  
Kaishev November 2009
4. «Theory of Copulas and Their use to Generate Synthetic Data»  
Theodora K. Kokkinioti
5. «Copulas and their applications in risk management and finance»  
Ioannis P. Kokkinos
6. «Modelling the joint distribution of competing risks survival times  
using copula functions» Vladimir K. Kaishev, Dimitrina S. Dimitrova  
and Steven Haberman, Cass Business School, City University, London
7. [http:// www.iatronet.gr/ygeia/kardiologia/article/33103/isxaimiki-  
kardiopatheia-poy-ofeiletai-kai-pws-emfanizetai](http://www.iatronet.gr/ygeia/kardiologia/article/33103/isxaimiki-kardiopatheia-poy-ofeiletai-kai-pws-emfanizetai)
8. [https://el.wikipedia.org/wiki/ karkinos](https://el.wikipedia.org/wiki/karkinos) tou pneumonia
9. <http://www.iatropedia.gr/encyclopedia/karkinos-tou-pnevmona>