

# ΚΕΦΑΛΑΙΟ 3

## ΠΙΝΑΚΕΣ ΕΠΙΒΙΩΣΗΣ (LIFE TABLES)

Τα μέτρα θνησιμότητας που παρουσιάστηκαν στο προηγούμενο κεφάλαιο ικανοποιούν πολλές ανάγκες του ερευνητή στην προσπάθεια του να περιγράψει τη θνησιμότητα ενός πληθυσμού ή να συγκρίνει επίπεδα θνησιμότητας μεταξύ πληθυσμών. Υπάρχουν όμως πληθώρα άλλων ερωτημάτων που ο ερευνητής καλείται να απαντήσει στη μελέτη της θνησιμότητας και για τα οποία ο υπολογισμός κάποιου συντελεστή είναι ανεπαρκής.

Ένας εναλλακτικός τρόπος μέτρησης της θνησιμότητας ενός πληθυσμού είναι η κατασκευή ενός πίνακα επιβίωσης. Σκοπός ενός τέτοιου πίνακα είναι η αναλυτική παρουσίαση της ειδικής κατά ηλικία θνησιμότητας ενός πληθυσμού. Η χρήση των πινάκων επιβίωσης είναι ευρύτατη και αναγκαία στη δημογραφική - δημομετρική ανάλυση, όχι μόνο για την μελέτη της θνησιμότητας αλλά και για τις προβολές πληθυσμού, τις μελέτες γεννητικότητας κ.α. στη βιοστατιστική έρευνα, στον ασφαλιστικό τομέα (ιδιωτική και κοινωνική ασφάλιση), στην περίθαλψη και στην κοινωνική έρευνα ευρύτερα.

### 3.1 Είδη Πινάκων Επιβίωσης

#### A. Πίνακες επιβίωσης γενεάς (Cohort life tables)

Ας θεωρήσουμε μια ομάδα αποτελούμενη από έναν αριθμό  $l_0$  ατόμων. Αυτά τα άτομα έχουν γεννηθεί το ίδιο ημερολογιακό έτος, αποτελούν δηλαδή τον πληθυσμό μιας γενεάς (cohort, generation). Ας θεωρήσουμε ακόμα ότι τα άτομα αυτά αποτελούν μια κλειστή ομάδα, ότι δηλαδή δεν υπάρχει μετανάστευση από ή προς αυτή την ομάδα. Με την πάροδο του χρόνου η γενεά αυτή σταδιακά συρρικνώνεται εφ' όσον τα μέλη της σταδιακά εκλείπουν. Η κατά ηλικία θνησιμότητα των μελών μιας τέτοιας “ιδανικής” γενεάς μπορεί να περιγραφεί από ένα πίνακα επιβίωσης γενεάς.

**Πρόβλημα:** Για την ολοκλήρωση ενός τέτοιου πίνακα θα πρέπει να μελετάται αυτό το σύνολο ατόμων από τη γέννηση μέχρι την εξαφάνισή του. Θα απαιτηθούν, κατά συνέπεια περίπου 100-110 χρόνια.

#### B. Πίνακας επιβίωσης περιόδου (Period life tables)

Ένας πίνακας επιβίωσης περιόδου αποτελεί την αποτύπωση της κατά ηλικία θνησιμότητας ενός πληθυσμού κάποιας γεωγραφικής περιοχής, κάποιας χρονικής περιόδου.

Ο τύπος αυτός είναι προσέγγιση ενός πίνακα επιβίωσης γενεάς. Τα μέλη του πληθυσμού στον οποίο αναφέρεται ανήκουν βέβαια σε διαφορετικές γενεές. Είναι αντιμετώπιση “cross - sectional”. Δεν αναφέρεται σε μια γενεά αλλά κατασκευάζεται από μια “συνθετική γενεά” και μελετάται η εξέλιξη της βάσει των συνθηκών θνησιμότητας της περιόδου αναφοράς.

**Πρόβλημα:** Στερείται ομοιογένειας. Περιλαμβάνει άτομα που ανήκουν σε 100-110 διαφορετικές γενεές και φέρουν τα χαρακτηριστικά (βιολογικά και κοινωνικά) της γενεάς

τους. Είναι όμως η μόνη πρακτική λύση. Ένας τέτοιος πίνακας αποτελεί στιγμιαία (φωτογραφική) αποτύπωση της θνησιμότητας ενός ετερόκλιτου πληθυσμού.

Ακόμα, οι πίνακες επιβίωσης χωρίζονται σε δύο κατηγορίες όσον αφορά τη μορφή τους.

### A. Πλήρεις πίνακες επιβίωσης (Complete life tables)

Παρουσιάζουν τη θνησιμότητα κάθε ηλικίας:  $x=0, 1, 2, \dots, w$ .

### B. Συνεπτυγμένοι πίνακες επιβίωσης (Abridged life tables)

Παρουσιάζουν τη θνησιμότητα ομάδων ηλικιών, συνήθως εύρους 5 ή 10 ετών, με εξαίρεση συνήθως το πρώτο έτος ζωής, (την ηλικία μηδέν) που λόγω της ιδιαιτερότητας της εμφανίζεται ξεχωριστά.

Οι πίνακες της πρώτης κατηγορίας είναι αναλυτικότεροι, ακριβέστεροι και για κάποιες χρήσεις (όπως στη βιοστατιστική, και στην αναλογιστική επιστήμη) αναγκαίοι. Αντίθετα οι πίνακες της δεύτερης κατηγορίας είναι περιληπτικότεροι, βολικότεροι (περιλαμβάνουν σειρές 10-16 αριθμών αντί 100 - 110), ενώ για πολλούς πληθυσμούς είναι οι μόνοι υπάρχοντες λόγω ελλείψεων και σφαλμάτων των δημογραφικών δεδομένων (π.χ. Ελλάδα).

## 3.2. Δομή Πινάκων Επιβίωσης

Ας θεωρήσουμε μια γενεά (cohort) αποτελούμενη από  $l_0$  αριθμό ατόμων (συνήθως  $l_0=100.000$ ). Το  $l_0$  είναι ένας συμβατικός αριθμός, αναφέρεται σε γεννήσεις ζώντων και αποτελεί το βασικό πλήθος το οποίο αφορά ο πίνακας επιβίωσης και πάνω στον οποίο οικοδομείται. Αποκαλείται **ρίζα** του πίνακα (radix).

Θα μελετηθεί πώς αυτός ο αριθμός ατόμων  $l_0$  μειώνεται διαχρονικά λόγω της θνησιμότητας. Η κατά ηλικία θνησιμότητα μιας τέτοιας γενεάς, όπως αναφέρθηκε προηγούμενα, περιγράφεται από ένα πίνακα επιβίωσης γενεάς.

Στην πράξη η κατά ηλικία θνησιμότητα ενός πληθυσμού μπορεί να περιγραφεί από ένα πίνακα επιβίωσης περιόδου κάτω από τη θεώρηση μιας υποθετικής γενεάς, που έχει τα ίδια κατά ηλικία επίπεδα θνησιμότητας όπως ο πληθυσμός αυτός.

Στη συνέχεια θα θεωρήσουμε ένα πίνακα επιβίωσης περιόδου σαν πίνακα επιβίωσης μιας υποθετικής γενεάς και θα ορίσουμε τις βασικές συναρτήσεις που περιλαμβάνει ένας τέτοιος πίνακας<sup>1</sup>.

- $l_0$

Αρχικά ορίζεται ο αριθμός γεννήσεων των ατόμων υπό παρατήρηση της υποθετικής γενεάς  $l_0$ , κατά σύμβαση ίσος με 100.000.

- $l_x$

Ο αριθμός των ατόμων που φθάνουν την ακριβή ηλικία  $x$ , συμβολίζεται  $l_x$ .

---

<sup>1</sup> Οι συναρτήσεις του πίνακα επιβίωσης θα αναφέρονται στο διάστημα ηλικίας εύρους  $n$  ετών:  $[x, x+n)$

Το κλειδί του πίνακα επιβίωσης αποτελεί η στήλη των πιθανοτήτων θανάτου (probabilities of dying),

- ${}_n q_x$

είναι η πιθανότητα κάποιος ακριβούς ηλικίας  $x$  να πεθάνει στο διάστημα ηλικίας  $[x, x+n)$ .

Η πιθανότητα θανάτου είναι μια δεσμευμένη πιθανότητα με τη δέσμευση ότι κάποιος έχει επιβιώσει μέχρι την ακριβή ηλικία  $x$  ενώ

- ${}_n p_x$

είναι η πιθανότητα κάποιος ακριβούς ηλικίας  $x$  να επιβιώσει όλο το διάστημα ηλικίας  $[x, x+n)$ .

Μεταξύ των δύο πιθανοτήτων ισχύει η σχέση

$${}_n p_x = 1 - {}_n q_x$$

- $S(x)$

Ο λόγος του  $\ell_x$  προς το  $\ell_0$  αποτελεί την πιθανότητα επιβίωσης (survival probability),  $S(x)$ , την πιθανότητα δηλαδή κάποιος να φθάσει την ακριβή ηλικία  $x$ , να επιβιώσει στο διάστημα ηλικίας  $[0, x]$ .

$$S(x) = \frac{\ell_x}{\ell_0}$$

Αυτή η πιθανότητα δεν περιλαμβάνεται συνήθως στον πίνακα επιβίωσης, είναι όμως κεντρικής σημασίας στη δημομετρική και βιοστατιστική ανάλυση.

Η πιθανότητα επιβίωσης θα μπορούσε επίσης να εκφραστεί μέσω των πιθανοτήτων  $p_i$  και  $q_i$  ως εξής:

$$S(x) = \prod_{i=0}^{x-1} p_i = \prod_{i=0}^{x-1} (1 - q_i)$$

- ${}_n d_x$

είναι ο συνολικός αριθμός θανάτων στο διάστημα ηλικίας  $[x, x+n)$ .

Αυτός ο αριθμός δεν μπορεί από μόνος του να εκφράσει την θνησιμότητα αφού αυτή εξαρτάται και από το μέγεθος του πληθυσμού στον οποίο αναφέρεται.

Ισχύει

$${}_n d_x = \ell_x - \ell_{x+n}$$

ή εναλλακτικά:

$${}_n d_x = \ell_x \cdot {}_n q_x$$

- ${}_nL_x$

είναι ο συνολικός αριθμός ετών ζωής που βιώνονται από τα άτομα του πληθυσμού στο διάστημα ηλικίας  $[x, x+n)$ .

Κάθε μέλος του πληθυσμού που βιώνει όλο το διάστημα ηλικίας  $[x, x+n)$ , συμβάλλει στον υπολογισμό του  ${}_nL_x$  κατά  $n$  έτη ζωής, ενώ κάθε ένας που πεθαίνει κάποτε μέσα στο ίδιο διάστημα, κάτω από την υπόθεση της ομοιόμορφης κατανομής θανάτων μέσα στο διάστημα ηλικίας, συμβάλλει στο  ${}_nL_x$  κατά μέσο όρο με  $n/2$  αριθμό ετών ζωής. Έτσι το  ${}_nL_x$  μπορεί να εκτιμηθεί από τον προσεγγιστικό τύπο:

$${}_nL_x = n \cdot \ell_{x+n} + \frac{n}{2} \cdot d_x$$

αλλά  $d_x = \ell_x - \ell_{x+n}$

Άρα 
$${}_nL_x = \frac{n}{2} (\ell_x + \ell_{x+n})$$

Αν θεωρήσουμε την ηλικία  $x$  συνεχή μεταβλητή τότε:

$${}_nL_x = \int_x^{x+n} \ell(t) \cdot dt = \int_0^n \ell(x+t) dt \approx n \left( \frac{\ell_x + \ell_{x+n}}{2} \right)$$

- $T_x$

είναι ο συνολικός αριθμός ετών ζωής που πρόκειται να βιώσουν τα  $\ell_0$  άτομα του υποθετικού πληθυσμού στο διάστημα ηλικίας  $[x, w)$  όπου  $w$  η ηλικία που δεν φτάνει κανείς ( $w-1$ : η ανώτατη ηλικία στον πληθυσμό).

$$T_x = \sum_{i \geq x} {}_nL_i$$

Ακόμα: 
$$T_{x+1} = T_x - {}_nL_x \Rightarrow {}_nL_x = T_x - T_{x+n}$$

Θεωρώντας την ηλικία  $x$  συνεχή μεταβλητή έχουμε:

$$T_{x+1} = \int_0^{w-x} \ell(x+t) \cdot dt$$

Τέλος, η περισσότερο ενδιαφέρουσα στήλη του πίνακα επιβίωσης είναι η στήλη της προσδοκώμενης ζωής,  $e_x^0$  (expectation of life, life expectancy).

Η στήλη αυτή εκφράζει την αναμενόμενη (μέση) υπολειπόμενη ζωή των ατόμων ηλικίας  $x$  του πληθυσμού στον οποίο αναφέρεται ο πίνακας επιβίωσης.

$$e_x^0 = \frac{T_x}{l_x}$$

Έτσι η αναμενόμενη ηλικία θανάτου κάποιου ατόμου ηλικίας  $x$  ισούται με:

$$x + e_x^0$$

Αν η  $x$  θεωρηθεί συνεχής τότε:

$$e_x^0 = \frac{1}{l_x} \int_0^{w-x} l(x+t) \cdot dt$$

### 3.3. Ο Πίνακας Επιβίωσης σαν ένας Στάσιμος Πληθυσμός

Μια εναλλακτική εξήγηση των μεγεθών που περιλαμβάνει ο πίνακας επιβίωσης συνδέεται με την θεώρηση ενός στάσιμου πληθυσμού. Στάσιμος πληθυσμός είναι εκείνος του οποίου το συνολικό μέγεθος και η κατά ηλικία σύνθεση παραμένουν σταθερά διαχρονικά. Σε ένα τέτοιο υποθετικό πληθυσμό, ο αριθμός γεννήσεων κάθε χρόνο είναι σταθερός και ίσος με τον αριθμό θανάτων και η ειδική κατά ηλικία θνησιμότητα παραμένει σταθερή διαχρονικά.

Αν θεωρήσουμε ότι ο πίνακας επιβίωσης αναφέρεται σε ένα στάσιμο πληθυσμό στον οποίο συμβαίνουν  $l_0$  (=100.000) γεννήσεις κάθε χρόνο (και βέβαια  $l_0$  θάνατοι) και ότι στο στάσιμο αυτό πληθυσμό αντιστοιχούν τα επίπεδα της κατά - ηλικία θνησιμότητας όπως αυτά δίνονται από την στήλη  ${}_nq_x$  του πίνακα επιβίωσης, τότε και οι υπόλοιπες συναρτήσεις του πίνακα θα αναφέρονται όχι σε μια υποθετική γενεά αλλά σε αυτόν το στάσιμο πληθυσμό.

Έτσι έχουμε:

$l_x$  : ο αριθμός ατόμων που φθάνουν την ηλικία  $x$  κάθε χρόνο στο στάσιμο πληθυσμό.

${}_n d_x$  : ο αριθμός θανάτων σε ηλικίες του διαστήματος  $[x, x+n)$  που συμβαίνουν κάθε χρόνο στο στάσιμο πληθυσμό.

${}_n L_x$  : ο μέσος αριθμός ατόμων ηλικιών  $[x, x+n)$  στο στάσιμο πληθυσμό.

${}_n T_x$  : ο μέσος αριθμός ατόμων ηλικιών  $[x, w)$  στο στάσιμο πληθυσμό.

Κάθε μια από τις δύο διαφορετικές θεωρήσεις του πίνακα επιβίωσης είναι κατάλληλη για διαφορετικές εφαρμογές. Η πρώτη θεώρηση (της υποθετικής γενεάς) εφαρμόζεται σε μελέτες δημόσιας υγείας, στην ανάλυση θνησιμότητας, στον υπολογισμό συντελεστών επιβίωσης για να εκτιμηθεί ο πληθυσμός, η γεννητικότητα, η αναπαραγωγή. Η δεύτερη θεώρηση (του στάσιμου πληθυσμού) εφαρμόζεται σε συγκρίσεις θνησιμότητας και σε μελέτες της σύνθεσης του πληθυσμού.

Το σκεπτικό κατασκευής ενός πίνακα επιβίωσης εφαρμόζεται και για την περιγραφή άλλων δημογραφικών γεγονότων εκτός της θνησιμότητας π.χ. οι πίνακες οικονομικής δραστηριότητας του πληθυσμού όπου αντί να θεωρήσουμε την ηλικία των μελών του πληθυσμού, θεωρούμε τα χρόνια άσκησης κάποιας οικονομικής δραστηριότητας, ή οι πίνακες γαμηλιότητας που με εντελώς αντίστοιχο τρόπο μελετούν αντί της θνησιμότητας την κατά ηλικία γαμηλιότητα του πληθυσμού.

Έχει ενδιαφέρον να παρατηρήσουμε ότι οι συναρτήσεις  ${}_nq_x, {}_np_x, {}_nd_x$  και  ${}_nL_x$  αναφέρονται στο διάστημα ηλικίας  $[x, x+n)$ , ενώ οι  $l_x$  και  $e_x^0$  αναφέρονται στην ακριβή ηλικία  $x$ . Ακόμη, ενώ οι τιμές των  $l_x, {}_nd_x, {}_nL_x$  και  $T_x$  εξαρτώνται από την τιμή του  $l_0$ , που αποτελεί κάποιο υποθετικό πλήθος το οποίο ορίζεται αυθαίρετα, οι τιμές των  ${}_nq_x, {}_np_x$  και  $e_x^0$  είναι ανεξάρτητες από την τιμή του  $l_0$ , και αφορούν τον πραγματικό πληθυσμό στον οποίο αναφέρεται ο πίνακας επιβίωσης. Ακόμα, εφ' όσον:

$$l_0 = \sum_{x \geq 0} {}_nd_x, \quad e_x^0 = \frac{T_0}{l_0} \quad \text{και} \quad T_0 = \sum_{x \geq 0} {}_nL_x$$

τότε:

$$e_x^0 = \frac{\sum_{x \geq 0} {}_nL_x}{\sum_{x \geq 0} {}_nd_x} = \frac{\bar{P}}{D} = \frac{1}{A\Sigma\Theta}$$

### 3.4 Προσεγγιστικές Σχέσεις μεταξύ ${}_nm_x$ και ${}_nq_x$

Έστω  ${}_n\alpha_x$  ο αναμενόμενος αριθμός ετών που βιώνει στο διάστημα ηλικίας  $[x, x+n)$  ένα τυχαίο άτομο ηλικίας  $x$  που πεθαίνει μέσα στο ίδιο διάστημα ηλικίας ( $0 < {}_n\alpha_x < n$ ).

Έστω ακόμα  ${}_nf_x = \frac{{}_n\alpha_x}{n}$ ,  $0 < {}_nf_x < 1$

Το  ${}_nf_x$  είναι ο αναμενόμενος σχετικός αριθμός ετών που βιώνεται στο διάστημα ηλικίας  $[x, x+n)$  από ένα τυχαίο άτομο του πληθυσμού, ηλικίας  $x$  που πεθαίνει μέσα σε αυτό το διάστημα ηλικίας, (ένα ποσοστό του διαστήματος ηλικιών  $[x, x+n)$ ).

Έτσι το  ${}_nL_x$  ο συνολικός αριθμός ετών ζωής που θα βιωθούν από τα  $l_x$  άτομα στο διάστημα ηλικιών  $[x, x+n)$  αποτελείται από:

1.  $n$  αριθμό ετών για κάθε άτομο που βίωσε όλο το διάστημα  $[x, x+n)$ , ήτοι για  $l_{x+n}$  άτομα (συνολικά αυτοί συνέβαλαν κατά  $n \cdot l_{x+n}$  έτη ζωής).

2.  ${}_n\alpha_x$  αριθμό ετών για κάθε έναν από αυτούς που πεθαίνουν κάποτε μέσα στο ίδιο διάστημα ηλικιών, ήτοι για  ${}_nd_x$  άτομα (συνολικά αυτοί συνέβαλαν κατά  ${}_n\alpha_x \cdot {}_nd_x$  έτη ζωής).

Άρα:

$${}_nL_x = n \cdot l_{x+n} + {}_n\alpha_x \cdot {}_nd_x \Rightarrow$$

$$\Rightarrow {}_nL_x = n (l_x - {}_nd_x) + n \cdot {}_nf_x \cdot {}_nd_x \Rightarrow$$

$$\Rightarrow {}_nL_x = n \left[ l_x - {}_nd_x (1 - {}_nf_x) \right]$$

Όμως ο ειδικός συντελεστής θνησιμότητας για το διάστημα ηλικίας  $[x, x+n)$ , βάσει των συναρτήσεων του πίνακα επιβίωσης ισούται με:

$${}_n m_x = \frac{{}_n d_x}{{}_n L_x} \Rightarrow$$

$${}_n m_x = \frac{{}_n d_x}{n \left[ \ell_x - {}_n d_x (1 - {}_n f_x) \right]}$$

Αλλά:  ${}_n d_x = \ell_x \cdot {}_n q_x$

Άρα

$${}_n m_x = \frac{\ell_x \cdot {}_n q_x}{n \cdot \ell_x \left[ 1 - {}_n q_x (1 - {}_n f_x) \right]} \Rightarrow$$

$$\Rightarrow {}_n m_x = \frac{{}_n q_x}{n \left[ 1 - {}_n q_x (1 - {}_n f_x) \right]}$$

και 
$${}_n q_x = \frac{{}_n m_x}{1 + n(1 - {}_n f_x) \cdot {}_n m_x} \quad (1)$$

### Ειδικές περιπτώσεις:

Αν υποθέσουμε ομοιόμορφη κατανομή θανάτων μέσα στο διάστημα ηλικίας  $[x, x+n)$ .

Τότε:  ${}_n a_x = \frac{n}{2}$  και κατά συνέπεια  ${}_n f_x = \frac{1}{2}$

Έτσι η σχέση (1) γίνεται:

$${}_n m_x = \frac{{}_n q_x}{n \left( 1 - \frac{1}{2} {}_n q_x \right)} \quad (2)$$

Λύνοντας την (2) ως προς  ${}_n q_x$  έχουμε:

$${}_n q_x = \frac{{}_n m_x}{\frac{1}{2} \left( 1 + \frac{n}{2} {}_n m_x \right)} \quad (3)$$

Έτσι για  $n=1$ , βάσει των σχέσεων (2) και (3), τα κατά ηλικία μεγέθη  $m_x$  και  $q_x$  γίνονται:

$$m_x = \frac{q_x}{1 - \frac{1}{2}q_x} \quad \text{και} \quad q_x = \frac{m_x}{1 + \frac{1}{2}m_x}$$

### 3.5 Ένταση Θνησιμότητας $\mu(x)$

(Force of mortality, Intensity of mortality, Hazard function)

Αν η ηλικία  $x$  θεωρηθεί μια συνεχής μεταβλητή, στο διάστημα  $[x, x+\Delta x)$ , όπου  $\Delta x \rightarrow 0$ , συμβαίνουν  $\ell(x) - \ell(x+\Delta x)$  θάνατοι.

Ορίζεται σαν **ένταση θνησιμότητας**  $\mu(x)$  η οριακή τιμή του ειδικού κατά ηλικία συντελεστή θνησιμότητας του διαστήματος  $[x, x+\Delta x)$ , όταν  $\Delta x \rightarrow 0$ , (το εύρος του διαστήματος ηλικίας είναι απειροελάχιστο).

$$\mu(x) = \lim_{\Delta x \rightarrow 0} \Delta x \cdot m_x \Rightarrow$$

$$\mu(x) = \lim_{\Delta x \rightarrow 0} \frac{D_x d_x}{D_x L_x} \Rightarrow \quad (4)$$

$$\mu(x) = \lim_{\Delta x \rightarrow 0} \frac{\ell(x) - \ell(x + \Delta x)}{\int_x^{x+\Delta x} \ell(t) \cdot dt}$$

Όμως είναι γνωστό ότι η παράγωγος μιας συνάρτησης  $f(x)$  ορίζεται σαν:

$$\frac{df(x)}{dx} = \lim_{t \rightarrow 0} \frac{f(x+t) - f(x)}{t}$$

Άρα η (4) γίνεται:

$$\mu(x) = -\frac{d\ell(x)}{dx} \cdot \frac{1}{\ell(x)} \Rightarrow$$

$$\mu(x) = -\frac{d \ln \ell(x)}{dx} \Rightarrow$$

$$-\int_0^x m(t) dt = \ln \ell(x) + k \Rightarrow$$

$$\Rightarrow \exp\left(-\int_0^x \mu(t) dt\right) = \ell(x) \cdot e^k.$$

Θέτοντας  $\frac{1}{e^k}$  έχουμε:



$$\ell(x) = c \cdot \exp\left(-\int_0^x m(t) dt\right) \quad (5)$$

Όμως για  $x=0$  η (1) γίνεται:

$$\ell(0) = c$$

Άρα

$$\ell(x) = \ell(0) \cdot \exp\left(-\int_0^x m(t) dt\right) \Rightarrow \quad (6)$$

$$S(x) = \ell(x) \cdot \exp\left(\int_0^x m(t) dt\right) \quad (7)$$

όπου,  $S(x) = \frac{\ell(x)}{\ell(0)}$  είναι η πιθανότητα επιβίωσης στο διάστημα ηλικίας  $[0, x)$ .

Η τελευταία αυτή σχέση που εκφράζει την πιθανότητα επιβίωσης (survival probability),  $S(x)$ , σαν συνάρτηση της έντασης θνησιμότητας,  $\mu(x)$ , αποτελεί το μαθηματικό ορισμό της πιθανότητας επιβίωσης και είναι κεντρικής σημασίας στη Δημομετρική ανάλυση και στη Βιοστατιστική.

Ακόμα, εφ' όσον:  ${}_n p_x = \frac{\ell(x+n)}{\ell(x)}$

από την σχέση (6) έχουμε:

$${}_n p_x = \exp\left(-\int_x^{x+n} \mu(t) \cdot dt\right) \Rightarrow \quad (8)$$

$$1 - {}_n p_x = 1 - \exp\left(-\int_x^{x+n} \mu(t) \cdot dt\right)$$

Υποθέτοντας σταθερή ένταση θνησιμότητας στο διάστημα ηλικίας  $[x, x+1)$  έχουμε:

$$p_x = \exp(-\mu_x) \Rightarrow \mu_x = -\ln(p_x)$$

Αυτή η τελευταία σχέση είναι φυσικά προσεγγιστική. Βάσει αυτής της έκφρασης εκτιμάται η ένταση θνησιμότητας για την κάθε ηλικία βάσει των πιθανοτήτων  ${}_1 q_x$ , του πίνακα επιβίωσης.

### 3.6 Δημομετρική Ανάλυση των Συναρτήσεων του Πίνακα Επιβίωσης

Υποθέτουμε ότι κάποια χρονική στιγμή (την ονομάζουμε χρονική στιγμή 0) έχουμε  $N$  αριθμό ατόμων που όλα έχουν την ίδια ηλικία  $x$  (η  $x$  θεωρείται συνεχής μεταβλητή). Θα παρακολουθήσουμε την εξέλιξη αυτής της ομάδας  $N$  ατόμων μέχρι την εξαφάνιση της. Έστω  $T_j$  η χρονική στιγμή που το άτομο πεθαίνει ( $j = 1, 2, \dots, N$ ). Η απόσταση του από την χρονική στιγμή 0 (το μέγεθος του δηλαδή) είναι ίση με την προσδοκώμενη ζωή του ατόμου  $j$  μετά την ηλικία  $x$ .

Η ηλικία θανάτου θα είναι:  $x+T_j$

Άρα οι διάφορες  $T_j$ ,  $j = 1, \dots, N$  εκλαμβάνονται σαν τυχαίες μεταβλητές ανεξάρτητες μεταξύ τους και ισόνομες.

Συμβολίζουμε τη συνάρτηση κατανομής των  $T_j$  με:

$$F_T(t) = P(T_j \leq t)$$

κατά συνέπεια  $F_T(t)$  είναι η πιθανότητα κάποιος που έκλεισε την ηλικία  $x$  να πεθάνει πριν κλείσει την ηλικία  $x+t$  (να πεθάνει στο διάστημα  $[x, x+t)$ ), το οποίο ταυτίζεται με την πιθανότητα θανάτου  ${}_t q_x$  του πίνακα επιβίωσης.

Άρα :

$$F_T(t) \equiv {}_t q_x$$

και έτσι η πιθανότητα  ${}_t p_x$ , κάποιος που έκλεισε την ηλικία  $x$  να επιβιώσει στο διάστημα ηλικίας  $[x, x+t)$  μπορεί να εκφραστεί σαν:

$${}_t p_x = 1 - {}_t q_x = 1 - F_T(t) = P(T_j > t)$$

Έστω  $Y$  ο αριθμός ατόμων που έφθασαν την ηλικία  $x+t$  από τα  $N$  άτομα της ομάδας που είχαν ηλικία  $x$  ακριβώς τη χρονική στιγμή 0.

Η  $Y$  μπορεί να θεωρηθεί μια διακριτή τυχαία μεταβλητή που ακολουθεί διωνυμική κατανομή με παραμέτρους,  $N$  και  ${}_t p_x$  ( $Y \sim \text{Bin}(N, {}_t p_x)$ ) και κατανομή πιθανοτήτων:

$$P(Y = y) = \binom{N}{y} {}_t p_x^y (1 - {}_t p_x)^{N-y}, \quad y = 0, 1, \dots, N$$

Κατά συνέπεια η αναμενόμενη τιμή της  $Y$ ,  $E(Y)$  θα ισούται με:

$$E(Y) = N \cdot {}_t p_x$$

Η τιμή  $E(Y)$  είναι αναμενόμενος αριθμός ατόμων που επιβιώνουν την ηλικία  $x+t$  στο διάστημα  $[x, x+t)$ .

(Σύγκρινε  $l_{x+t} = l_x \cdot {}_t p_x$  για  $N = l_x$ ).

Η διακύμανση της  $Y$ ,  $V(Y)$ , είναι:

$$V(Y) = N \cdot {}_t p_x \cdot {}_t q_x$$

Ο αριθμός ατόμων που πεθαίνουν στο διάστημα ηλικίας  $[x, x+t)$  ισούται με:

$$U = N - Y$$

Η  $U$  είναι και αυτή μια διακριτή τυχαία μεταβλητή που ακολουθεί διωνυμική κατανομή με παραμέτρους  $N$  και  ${}_t q_x = 1 - {}_t p_x$  ( $U \sim \text{Bin}(N, {}_t q_x)$ ) με αναμενόμενη τιμή

$$E(U) = N \cdot {}_t q_x$$

(Σύγκρινε:  ${}_t d_x = \ell_x \cdot {}_t q_x$ ,  ${}_t d_x = E(U)$ ,  $\ell_x = N$ ) και διακύμανση ίση με τη διακύμανση της Y:

$$V(U) = N \cdot {}_t q_x \cdot {}_t p_x$$

Ας υποθέσουμε τώρα ότι  $x=0$ . Ήτοι, τη χρονική στιγμή 0 έχουμε N γεννήσεις ζώντων. Ο αναμενόμενος αριθμός αυτών που θα φθάσουν την ηλικία  $x$  θα είναι ίσος με:

$$N \cdot {}_x p_0 = \ell_x$$

(Σύγκρινε:  $\ell_x = \ell_0 \cdot {}_x p_0$  όπου  $\ell_0$  ο αριθμός γεννήσεων ζώντων τη χρονική στιγμή 0, η ρίζα του πίνακα επιβίωσης).

Ας θεωρήσουμε τώρα  $T_j$  τη χρονική στιγμή θανάτου ενός τυχαίου μέλους  $j$  του συνόλου N γεννήσεων. Το  $T_j$  λοιπόν είναι το διάστημα ζωής αυτού του ατόμου.

$$P(T_j > x) = {}_x p_0 = S(x)$$

Έστω  $V_x$  η υπολειπόμενη ζωή ενός τυχαίου ατόμου ηλικίας  $x$ .

Το συνολικό διάστημα ζωής του θα είναι:

$$T_j = x + V_x$$

$$E(V_x) = e_x^0 \Rightarrow$$

$$E(T_j) = x + E(V_x) \Rightarrow$$

$$x + e_x^0 = E(T_j)$$

Ακόμα:

$$\begin{aligned} {}_t p_x &= P(V_x \geq t) = \\ &= P(T_j \geq x+t | T_j \geq x) = \\ &= \frac{P(T_j \geq x+t \cap T_j \geq x)}{P(T_j \geq x)} = \\ &= \frac{P(T_j \geq x+t)}{P(T_j \geq x)} = \\ &= \frac{x+t p_0}{p} = \frac{S(x+t)}{S(x)} = \frac{\ell_{x+t}}{\ell} \end{aligned}$$

$$\text{Έτσι } {}_t p_x = \frac{\ell_{x+t}}{\ell_x}$$

και κατ' αντιστοιχία εφ' όσον  ${}_t q_x = F_{T'}(t)$

$${}_t q_x = F_{T'}(t) = P(T_j \leq x+t | T_j \geq x) = \frac{P(x < T_j < x+t)}{P(T_j > x)} =$$

$${}_t q_x = F_{T'}(t) = P(T_j \leq x+t | T_j \geq x) =$$

$$= \frac{P(x < T_j < x+t)}{P(T_j > x)} =$$

$$= \frac{P(T_j > x) - P(T_j > x+t)}{P(T_j > x)} =$$

$$= \frac{{}_x p_0 - {}_{x+t} p_0}{{}_x p_0} = 1 - \frac{{}_{x+t} p_0}{{}_x p_0} = 1 - \frac{\ell_{x+t}}{\ell_x} = 1 - {}_t p_x$$

Ακόμα εφ' όσον  $T_j$  είναι αναμενόμενο διάστημα ζωής ενός τυχαίου ατόμου ηλικίας μηδέν, η ένταση θνησιμότητας  $\mu(t)$  ορίζεται από την σχέση:

$$\mu(t) = \frac{f_T(t) dt}{1 - F_T(t)}$$

όπου,  $f_T(t)$  η συνάρτηση πυκνότητας του  $T_j$  και  $F_T(t)$  η συνάρτηση κατανομής του  $T_j$ .

Αλλά

$$F_T(t) = P(T_j \leq t) = {}_t q_0$$

και

$$f_{T_j}(t) dt = P(t \leq T_j \leq t+dt) dt \rightarrow 0$$

Έτσι έχουμε:

$$\mu(t) dt = \frac{P(t \leq T_j \leq t+dt)}{P(T_j \geq t)} \Rightarrow$$

$$\mu(t) dt = P(t \leq T_j \leq t+dt | T_j \geq t)$$

Η ένταση θνησιμότητας λοιπόν είναι η πιθανότητα κάποιος ηλικίας του διαστήματος  $[t, t+dt)$  να πεθάνει πριν φθάσει την ηλικία  $t+dt$  (μέσα στο διάστημα ηλικίας) όταν  $dt \rightarrow 0$ .

$${}_tq_x = P(T_j < x+t | T_j \geq x) \Rightarrow$$

Αλλά:

$${}_tq_x = \frac{P(x \leq T_j \leq x+t)}{P(T_j \geq x)} \Rightarrow$$

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

Έτσι:

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{{}_tq_x}{t} &= \lim_{t \rightarrow 0} \frac{F_T(x+t) - F_T(x)}{1 - F_T(x)} \cdot \frac{1}{t} = \\ &= \frac{f_T(x)}{1 - F_T(x)} = \mu_x \end{aligned}$$

### 3.7. Μέθοδοι περιγραφής της κατά ηλικία θνησιμότητας

Όπως ήδη έχει αναφερθεί στο Κεφάλαιο 2, η μορφή της κατά ηλικία θνησιμότητας εμφανίζει πολλές ομοιότητες στους διάφορους ανθρώπινους πληθυσμούς. Τα κοινά αυτά χαρακτηριστικά αποτέλεσαν αφορμή για την ανάπτυξη γενικών μεθόδων περιγραφής της.

Οι μέθοδοι αυτοί χωρίζονται σε δύο βασικές κατηγορίες:

1. Πρότυποι πίνακες επιβίωσης,
2. Παραμετρικά υποδείγματα τα οποία παρουσιάζουν την θνησιμότητα ενός πληθυσμού σαν συνάρτηση της ηλικίας.

#### 3.7.1. Πρότυποι (Τυπικοί) πίνακες επιβίωσης (Model life tables)

Η πρώτη επίσημη συλλογή προτύπων πινάκων επιβίωσης κατασκευάστηκε από το Τμήμα Πληθυσμού του Οργανισμού Ηνωμένων Εθνών το 1955. Αυτοί οι πίνακες βασίζονται σε ανάλυση 158 εμπειρικών συνεπτυγμένων πινάκων επιβίωσης για κάθε φύλο, διάφορων χωρών και αφορούν χρονικές περιόδους από την αρχή του αιώνα μας ως το 1950. Η κάθε μια από τις σειρές των πιθανοτήτων θανάτου  ${}_nq_x$  για κάθε  $x > 0$ , εκφράστηκε σαν ένα πολώνυμο δευτέρου βαθμού της αμέσως προηγούμενης (της  ${}_nq_x$  που αφορούσε την αμέσως προηγούμενη ηλικία στους εμπειρικούς πίνακες επιβίωσης). Στη συνέχεια, αρχίζοντας από κάποια συγκεκριμένη τιμή του  $q_0$  και με την διαδοχική χρήση των πολωνύμων των οποίων οι παράμετροι εκτιμήθηκαν προηγουμένα με τη χρήση των εμπειρικών δεδομένων, δημιουργήθηκε μια σειρά προτύπων πινάκων επιβίωσης. Το κύριο μειονέκτημα αυτής της συλλογής προτύπων πινάκων είναι ότι στερούνται ευελιξίας εφόσον για κάθε δεδομένο επίπεδο της βρεφικής θνησιμότητας,  $q_0$  αντιστοιχεί μόνο ένας πίνακας επιβίωσης.

Οι Ledermann και Breas<sup>2</sup> το 1959 εφάρμοσαν παραγοντική ανάλυση στα δεδομένα των εμπειρικών πινάκων επιβίωσης στους οποίους στηρίχθηκαν οι πρότυποι πίνακες της συλλογής του ΟΗΕ. Τα αποτελέσματα της ανάλυσης τους έδειξαν ότι απαιτούνται περισσότερες της μίας διαστάσεις για μια επιτυχή προσέγγιση της μορφής εξέλιξης της κατά ηλικία θνησιμότητας στους διάφορους πληθυσμούς. Η παραγοντική ανάλυση έδειξε ότι οι διαφοροποιήσεις αυτής της εξέλιξης μπορούν επαρκώς να περιγραφούν με τη χρήση τριών παραγόντων. Ο πρώτος που εκφράζει το 77% της συνολικής μεταβλητότητας αυτών των μορφών της κατά ηλικία εξέλιξης της θνησιμότητας, συνδέονται με την θνησιμότητα ενηλίκων (ηλικιών μεγαλύτερων των 40) και με τα χαρακτηριστικά των ακραίων ηλικιών (κάτω των 5 ετών και άνω των 70 ετών) αντίστοιχα, Ο Bourgois-Pichat<sup>3</sup> το 1962 συνέχισαν την ανάλυση των Ledermann και Breas και έδειξαν ότι συμπληρωματικές διαστάσεις είναι αναγκαίες για μια πλήρη περιγραφή της κατά ηλικίας θνησιμότητας στο φάσμα των διαφόρων εμπειρικών πληθυσμών.

Μια άλλη συλλογή προτύπων πινάκων επιβίωσης αναπτύχθηκε το 1966 από τους Coale και Demeny<sup>4</sup> στο Πανεπιστήμιο του Princeton. Οι πίνακες αυτοί στηρίχθηκαν σε ανάλυση 326 εμπειρικών συνεπτυγμένων πινάκων επιβίωσης διαφόρων χωρών και διαφορετικών περιόδων. Η αποτύπωση αυτών των σειρών  ${}_nq_x$  σε διαγράμματα ως προς την ηλικία  $x$ , οδήγησε στην κατάταξή τους σε τέσσερις “οικογένειες” οι οποίες ήταν ομοιογενείς ως προς τη μορφή της κατά ηλικία εξέλιξής τους. Στην συνέχεια, έγινε εφαρμογή απλών μοντέλων γραμμικής παλινδρόμησης που εξέφραζαν τα  ${}_nq_x$ , και τους λογάριθμους τους,  $\log(10000 \cdot {}_nq_x)$  σαν συνάρτηση της προσδοκώμενης ζωής στην ηλικία 10,  $e_{10}^0$ , στους πίνακες κάθε οικογένειας.

Έτσι κατασκευάστηκαν τέσσερις σειρές πρότυπων πινάκων. Κάθε μια από αυτές τις σειρές αναφέρεται στην βιβλιογραφία σαν Δυτική, Ανατολική, Βόρεια και Νότια περιφερειακή οικογένεια θνησιμότητας.

Το 1979, μια σειρά προτύπων πινάκων επιβίωσης παρουσιάστηκε στην βιβλιογραφία από τους Petrioli και Berti<sup>5</sup>. Αυτοί οι πίνακες βασίζονται σε μια συνάρτηση της ηλικίας  $r(x)$  την οποία αποκαλούν “ανθεκτικότητα” και η οποία ορίζεται σαν τον λόγο του μέσου ετήσιου αριθμού θανάτων μετά την ηλικία  $x$  προς το μέσο αριθμό θανάτων πριν την ηλικία  $x$ :

$$r(x) = \frac{\ell_x / w - x}{(1 - \ell_x \cdot x)}$$

όπου  $\ell_x$  είναι η πιθανότητα επιβίωσης μέχρι την ηλικία  $x$  και  $w$  το ανώτατο όριο ζωής. Η συνάρτηση  $r(x)$  εξομαλύνεται με την χρήση του παραμετρικού μοντέλου:

$$r(x) = k x^2 \cdot (w - x)^b$$

<sup>2</sup> Ledermann, S., Breas, R. (1959): Les dimensions de la mortalité. *Population*, 14, 4, 637-682.

<sup>3</sup> Bourgois-Pichat (1951): La mesure de la mortalité infantile. *Population* 2, 233-248.

<sup>4</sup> Coale, A., J., Demeny, P. (1966): *Regional Model Life Tables and Stable Populations*, Princeton, New Jersey

<sup>5</sup> Petrioli, L, Bereti M. (1979): *Modelli di Mortalità*. Franco Angeli Editore, Milan.

Έτσι σειρές προτύπων επιβίωσης κατασκευάζονται με τη χρήση συνδυασμών των παραμέτρων  $k, a, b$ .

Οι πιο σύγχρονες σειρές προτύπων πινάκων παρουσιάστηκαν το 1982 από τον OECD. Αυτοί οι πίνακες κατασκευάστηκαν με την εφαρμογή παρόμοιας μεθόδου σαν αυτή των Ledermann-Breas.

Οι διάφοροι πίνακες επιβίωσης, είναι σαφές ότι δεν αντικατοπτρίζουν κάποια πραγματικότητα εφ' όσον προκύπτουν από επεξεργασίες που βασίζονται σε εμπειρικούς πίνακες. Τέτοιοι πίνακες όμως εξυπηρετούν τις ανάγκες της δημομετρικής ανάλυσης πολλαπλά. Χρησιμοποιούνται κύρια για να εκπροσωπήσουν (σαν υποκατάστατα βέβαια) την κατά ηλικία θνησιμότητα πληθυσμών των οποίων τα εμπειρικά δεδομένα θνησιμότητας είναι χαμηλής αξιοπιστίας ή/και ελλιπή. Ένα ακόμα ευρύ πεδίο εφαρμογής τους είναι οι προβολές πληθυσμού όπως θα δούμε στο Κεφάλαιο 8. Το κύριο μειονέκτημα των προτύπων πινάκων επιβίωσης είναι ότι παρουσιάζονται αριθμητικά και αυτό περιορίζει την ευελιξία τους και την ευκολία στην χρήση τους.

### 3.7.2 Παραμετρικά υποδείγματα θνησιμότητας (Laws of mortality)

Μια εναλλακτική μέθοδος περιγραφής της κατά ηλικία θνησιμότητας είναι η εφαρμογή παραμετρικών υποδειγμάτων που εμφανίζουν τη θνησιμότητα σαν συνάρτηση της ηλικίας. Τέτοια υποδείγματα αναφέρονται στη δημογραφική βιβλιογραφία σαν “Νόμοι Θνησιμότητας” (Laws of Mortality). Αυτά τα υποδείγματα χρησιμοποιούνται κύρια στην Δημομετρία για την εξομάλυνση των εμπειρικών δεδομένων, την ανάλυση τάσεων και την κατασκευή πληθυσμιακών προβολών. Τα υποδείγματα αυτά αποτελούν ένα σαφώς περισσότερο ευέλικτο τρόπο παρουσίασης της κατά - ηλικία θνησιμότητας από αυτό των προτύπων πινάκων επιβίωσης. Η διαδικασία ανάπτυξης τέτοιων υποδειγμάτων έχει αρχίσει πολύ πριν τις αρχές του αιώνα μας.

Η πρώτη προσπάθεια ανάπτυξης ενός τέτοιου υποδείγματος είναι ο τύπος de Moivre ο οποίος παρουσιάστηκε στη βιβλιογραφία το 1725. Ο τύπος αυτός εμφανίζει την ένταση θνησιμότητας  $\mu(x)$  σαν συνάρτηση της ηλικίας:

$$\mu(x) = 1/(w - x)$$

ή στην εναλλακτική μορφή όπου εμφανίζεται η πιθανότητα επιβίωσης  $S(x)$  σαν συνάρτηση της ηλικίας:

$$S(x) = 1 - \frac{x}{w}$$

όπου  $w$  αντιπροσωπεύει την ανώτατη δυνατή ηλικία.

Ο τύπος αυτός περιγράφει ικανοποιητικά την κατά ηλικία θνησιμότητα των μεγάλων ηλικιών αλλά αποτυγχάνει στις μικρότερες ηλικίες, κάτι που σχολιάστηκε από τον ίδιο το δημιουργό του, ο οποίος τον εφάρμοσε για τις ανάγκες υπολογισμών.

Ο πιο διαδεδομένος ίσως νόμος θνησιμότητας έχει αναπτυχθεί από τον Gompertz<sup>6</sup> το 1825. Βασίζεται στην θεώρηση ότι η ένταση θνησιμότητας εξελίσσεται εκθετικά με τη ηλικία:

---

<sup>6</sup> Gompertz, B. (1925): On the Nature of the function Expressive of the Law of Human Mortality in Mathematical Demography. Ed. by Dand Smith and Nathan Keyfitz, Springer Verlag, 1977, 279-282.

$$\mu(x) = bc^x$$

όπου  $b$  και  $c$  θετικές παράμετροι, ή με βάση την πιθανότητα επιβίωσης  $S(x)$ , στη μορφή:

$$S(x) = \exp\left[-\frac{b}{\log c}(c^x - 1)\right]$$

Και αυτό το υπόδειγμα δεν επιτυγχάνει να περιγράψει επιτυχώς τη θνησιμότητα των μικρών ηλικιών ενώ περιγράφει ικανοποιητικά την θνησιμότητα των ηλικιών μετά τα σαράντα.

Ο Makeham<sup>7</sup> το 1860 ανέπτυξε ένα υπόδειγμα της μορφής:

$$\mu(x) = \alpha + b \cdot c^x$$

ή εναλλακτικά

$$S(x) = \exp\left(-\alpha x - \frac{b}{\log c}(c^x - 1)\right)$$

προσέθεσε δηλαδή μια σταθερά στην έκφραση Gompertz. Όμως, και αυτή η έκφραση δεν κατορθώνει να περιγράψει ικανοποιητικά την θνησιμότητα του συνολικού διαστήματος ηλικιών. Εμφανίζει καλή εφαρμογή στις μεσαίες ηλικίες, ενώ υπερεκτιμά τη θνησιμότητα των μεγάλων ηλικιών και αποτυγχάνει στην περιγραφή της θνησιμότητας των μικρών ηλικιών. Εν τούτοις, έχει ευρύτητα χρησιμοποιηθεί από αναλογιστές.

Ο Perks<sup>8</sup> το 1932 πρότεινε την εφαρμογή ενός γραμμικού υποδείγματος, της απλής μορφής  $\alpha + \beta x$ , στη συνάρτηση  $\mu(x)/(1 - \mu(x))$ , η οποία ελάχιστα διαφέρει από την  $\mu(x)$  στις νεαρές ηλικίες ενώ φθίνει ταχύτερα από την τελευταία στις μεγάλες ηλικίες.

Ο Thiele<sup>9</sup> το 1872 πρότεινε ένα μοντέλο για την περιγραφή της θνησιμότητας του συνολικού διαστήματος ζωής. Η προσπάθεια του βασίστηκε στην υπόθεση ότι οι διάφορες αιτίες θανάτων μπορούν να χωριστούν σε τρεις διακριτές ομάδες, οι οποίες επιδρούν κύρια στην παιδική ηλικία, στους ενήλικους και στους υπερήλικους αντίστοιχα.

---

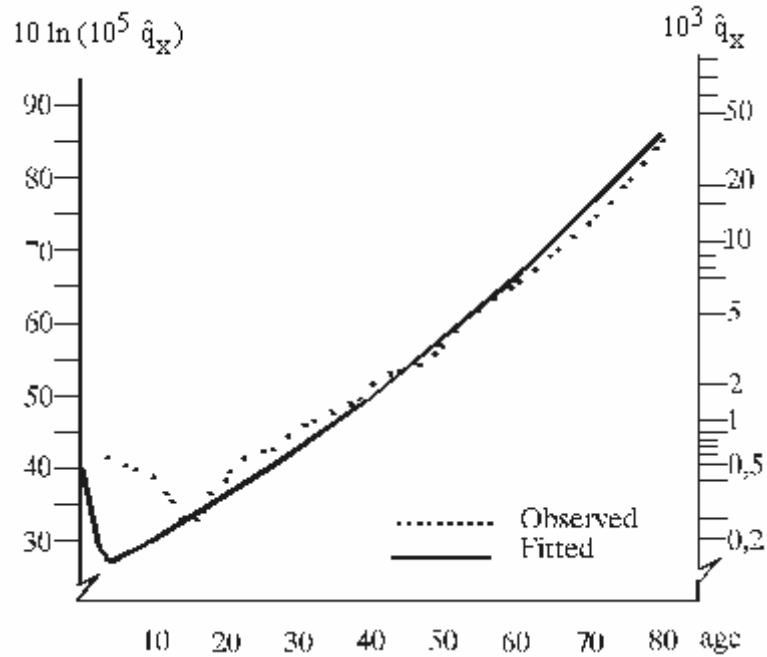
<sup>7</sup> Makaham, W. (1987), (1989): On the Law of Mortality in Mathematical Demography ed. by David Smith and Nathan Keyfitz, Springer Verlag 1977, 283-288.

<sup>8</sup> Perks W. (1932): On some experiments in the graduation of mortality statistics.

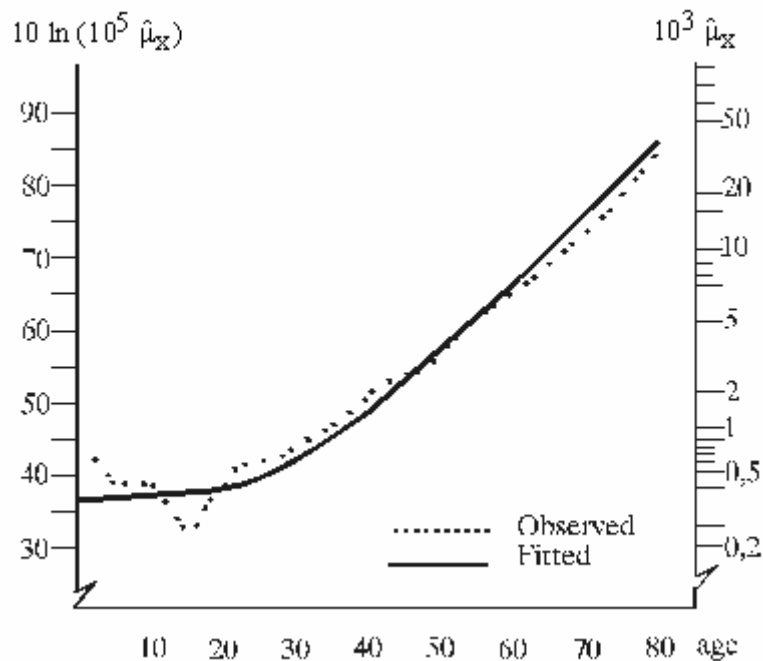
<sup>9</sup> Thiele, PN, (1972): On a Mathematical Formula to express the rate of mortality throughout the whole life, Journal of the Institute of Actuaries, 29,1, 97-107.



**Διάγραμμα 3.1** Εμπειρικές τιμές  $q_x$  του Σουηδικού γυναικείου πληθυσμού της περιόδου 1961-70 και εκτιμήσεις του με τη χρήση του μοντέλου Wittstein.



**Διάγραμμα 3.2** Τιμές  $\mu_x$  του Σουηδικού γυναικείου πληθυσμού της περιόδου 1961-70 και εκτιμήσεις τους με τη χρήση του μοντέλου Thiele.



Βάσει αυτής της έκφρασης η ένταση θνησιμότητας διαχωρίζεται σε τρεις προσθετικούς όρους:

$$\mu(x) = \mu_1(x) + \mu_2(x) + \mu_3(x)$$

όπου

$$\mu_1(x) = \alpha_1 \exp(-\beta_1 x)$$

$$\mu_2(x) = \alpha_2 \exp(-\beta_2^2 (x-c)^2 / 2)$$

$$\mu_3(x) = \alpha_3 \exp(\beta_3 x)$$

Η κύρια συμβολή στο  $\mu(x)$  προέρχεται από το  $\mu_1(x)$ ,  $\mu_2(x)$ ,  $\mu_3(x)$  στην παιδική ηλικία, την ενήλικη ζωή και γήρανση αντίστοιχα ενώ η συμβολή των υπολοίπων δύο όρων είναι ασήμαντη στα τρία αυτά διαστήματα.

Αυτός ο τύπος δεν έδειξε καλή εφαρμογή σε εμπειρικά δεδομένα ενώ η εφαρμογή του είναι ιδιαίτερα ανεπιτυχής στην αρχή του διαστήματος ηλικιών (στην παιδική ηλικία). Εν τούτοις η σκέψη στην οποία στηρίχθηκε υπήρξε θεμελιώδης για την κατασκευή επιτυχών υποδειγμάτων που παρουσιάστηκαν στην βιβλιογραφία περισσότερο από εκατό χρόνια αργότερα.

Ο Wittstein<sup>10</sup> το 1883 πρότεινε το υπόδειγμα:

$$q_x = \frac{1}{m} \alpha^{-(mx)^n} + \alpha^{-(w-x)^n}$$

όπου:

- $q_x$  η πιθανότητα θανάτου και
- $w$  η ανώτατη δυνατή ηλικία ενώ
- $\alpha, m$  παράμετροι που πρέπει να εκτιμηθούν.

Η υπόθεση στην οποία στηρίζεται αυτός ο τύπος είναι ότι οι αιτίες θανάτου μπορούν να χωριστούν σε δύο ομάδες, εκείνες που επιδρούν στην παιδική ηλικία και εκείνες που επιδρούν στους ενήλικους και ηλικιωμένους.

Τα διαγράμματα 3.1 και 3.2 παρουσιάζουν εφαρμογές των υποδειγμάτων Wittstein και Thiele αντίστοιχα σε εμπειρικά δεδομένα θνησιμότητας.

Το 1889 ο Makeham πρότεινε μια δεύτερη έκφραση η οποία περιέχει ένα αθροιστικό όρο ακόμα από την αρχική του:

$$\mu(x) = \alpha + hx = bc^x$$

ή εναλλακτική σε όρους της  $S(x)$ :

$$S(x) = \exp\left(-\alpha x + \frac{h}{2} x^2 - \frac{b}{\log c} (c^x - 1)\right)$$

Η εμπειρία έδειξε ότι και αυτό το μοντέλο είναι επιτυχές όταν εφαρμόζεται σε ενήλικους μόνο, στο ηλικιακό διάστημα δηλαδή που η ένταση θνησιμότητας αυξάνει με την ηλικία.

Ο Oppermann το 1870, πρότεινε ένα νόμο θνησιμότητας κατάλληλο μόνο για την περιγραφή της βρεφικής και παιδικής θνησιμότητας:

<sup>10</sup> Wittstein, T. (1883): The Mathematical Law of Mortality. Journal of the Institute of Actuaries, 24, 153-173.

$$\mu(x) = \frac{\alpha}{\sqrt{x}} + b + c\sqrt{x}$$

ή σε όρους της πιθανότητας επιβίωσης  $S(x)$ :

$$S(x) = \exp\left(-2\alpha\sqrt{x} - bx - \frac{2}{3}cx\sqrt{x}\right)$$

Οι Βρετανοί αναλογιστές το 1932 έχοντας συνεχές ενδιαφέρον για την εξομάλυνση της κατά ηλικία θνησιμότητας μέσω παραμετρικών υποδειγμάτων, ανέπτυξαν τον τύπο του Perks<sup>11</sup>:

$$\frac{q_x}{1-q_x} = \alpha + b + c^x$$

Λίγα χρόνια αργότερα Βρετανοί βιολόγοι ανέπτυξαν μια σχέση ανάλογη εκείνης του Perk, όπου  $\alpha=0$ , συνέδεσαν δηλαδή το  $\ln(q_x/1-q_x)$  με μια ευθεία γραμμή:

$$\ln\left(\frac{q_x}{1-q_x}\right) = b + c \cdot x$$

Αυτός ο τύπος χρησιμοποιήθηκε από το Institute and Faculty of Actuaries στη Μ. Βρετανία για την εξομάλυνση της εμπειρικής κατά ηλικία θνησιμότητας των συνταξιούχων και των ασφαλισμένων της περιόδου 1967-70 της χώρας.

Ένα παραμετρικό υπόδειγμα για την περιγραφή και εξομάλυνση της κατά-ηλικία επιβίωσης παρουσιάστηκε, σχετικά το 1982 από τους Mode και Busby<sup>12</sup>. Το υπόδειγμα αυτό αναφέρεται στην πιθανότητα επιβίωσης  $S(x)$  την οποία παρουσιάζει σύμφωνα με την έκφραση:

$$S(x) = \begin{cases} S_0(x), & 0 \leq x \leq \delta_0 \\ S_0(\delta_0) S_1(x - \delta_0) S_2(x - \delta_1), & \delta_0 < x < \delta_1 \\ S_0(\delta_0) S_1(\delta_1 - \delta_0), & x \geq \delta_1 \end{cases}$$

$$S_0(x) = \exp[a_0(\exp(-\beta_0 x) - 1)]$$

$$S_1(x) = \exp\left[\frac{\beta_1 \gamma_1^3}{3} - a_1 x + \frac{\beta_1}{3}(x - \gamma_1)^3\right]$$

$$S_2(x) = \exp[-a_2 x - \beta_2(\exp(\gamma_2 x) - 1)]$$

<sup>11</sup> Perks, W. (1932): On some experiments in the graduation of mortality statistics. Journal of the Institute of Actuaries, 63, 12-57.

<sup>12</sup> Mode, C., Busby, R. (1982): An eight parameter model of human mortality-the single decrement case. Bulletin of Mathematical Biology, 44, 647-659.

όπου:  $S_0(x)$  είναι η πιθανότητα κάποιος να φτάσει την ηλικία  $(x \leq \delta_0)$ , ενώ η  $\delta_0$  είναι μια θετική παράμετρος.

Στη συνέχεια είναι η δεσμευμένη πιθανότητα κάποιος που έφτασε την ηλικία  $\delta_0$  να επιβιώσει μέχρι την ηλικία  $x$ , όπου  $\delta_0 \leq x \leq \delta_1$  και η  $\delta_1$  είναι μια θετική παράμετρος. Τέλος  $S_2(x - \delta_1)$  είναι η δεσμευμένη πιθανότητα, κάποιος που έφτασε την ηλικία να φτάσει την ηλικία  $(x \geq \delta_1)$ .

Τα  $\delta_0, \delta_1$  είναι θετικοί παράμετροι, των οποίων οι τιμές θα επιλεγούν. Αυτό το σύστημα αλληλένδετων εξισώσεων, περιγράφει την κατά ηλικία επιβίωση του συνολικού διαστήματος ηλικιών πολύ ικανοποιητικά. Τα μειονεκτήματα αυτού του συστήματος είναι πρώτον η πολυπλοκότητα της και δεύτερον το γεγονός ότι εμπεριέχει αυθαιρεσία εφ' όσον το  $\delta_0$  και  $\delta_1$  επιλέγονται.

Ένας πολύ διαφορετικός τύπος υποδείγματος που έχει τελευταία αποκτήσει ιδιαίτερο ενδιαφέρον στην Δημογραφία είναι ένα σύστημα που παρουσιάστηκε το 1971 από τον W. Brass<sup>13</sup>.

Σύμφωνα με αυτό το σύστημα δύο διαφορετικές σειρές πιθανοτήτων επιβίωσης συνδέονται με τη σχέση:

$$\log\left(\frac{1-S(x)}{S(x)}\right) = \alpha + \beta \log\left(\frac{1-S'(x)}{S'(x)}\right)$$

όπου:  $S(x)$  και  $S'(x)$  η πιθανότητες επιβίωσης που θέλουμε να εξομαλύνουμε και  $S'(x)$  οι αντίστοιχες ενός πρότυπου πίνακα.

Το 1980 οι Heligman και Pollard<sup>14</sup> παρουσίασαν στο Journal of the Institute of Actuaries ένα αναλυτικό υπόδειγμα για την περιγραφή της θνησιμότητας του συνολικού διαστήματος ζωής.

Το υπόδειγμα αυτό συνδέει το λόγο συμπληρωματικών πιθανοτήτων  $q_x / p_x$  με την ηλικία μέσω της έκφρασης:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D \cdot \exp\left(-E\left(\log \frac{x}{F}\right)^2\right) + G \cdot H$$

όπου  $q_x$ : η πιθανότητα θανάτου στην ηλικία  $x$  και  $p_x = 1 - q_x$ .

Τα  $A, B, C, \dots, H$  είναι παράμετροι.

Ειδικά για  $x=0$ :

$$\frac{q_0}{p_0} = A^{B^C} + G$$

<sup>13</sup> Brass, W. (1974): Mortality models and their use in Demography. Transactions of the Faculty of Actuaries, 33, 2, 239.

<sup>14</sup> Heligman, L. Pollard, J. M. (1980): The Age Pattern of Mortality. Journal of the Institute of Actuaries, 107, 49-80.

Η θνησιμότητα της βρεφικής και παιδικής ηλικίας εκφράζεται κύρια από τον πρώτο αθροιστικό όρο του υποδείγματος. Ο πρώτος αυτός όρος είναι φθίνουσας εκθετικής μορφής με έντονα φθίνοντα ρυθμό. Περιλαμβάνει τρεις παραμέτρους. Η παράμετρος  $A$  παίρνει τιμές στο διάστημα  $(0,1)$  και είναι περίπου ίση με το  $q_1$ , συνδέεται επίσης στενά με το ρυθμό πτώσης των επιπέδων θνησιμότητας στο πρώτο αυτό διάστημα ζωής.

Η παράμετρος  $B$  είναι ενδεικτική για την ένταση της θνησιμότητας του πρώτου χρόνου ζωής,  $q_0$ , η οποία βάσει του υποδείγματος παίρνει τιμές στο διάστημα  $(q_1, 0.5)$ .

Η συμβολή της για τις επόμενες ηλικίες ( $x \geq 1$ ) είναι ασήμαντη. Αν  $B=0$  τότε  $q_0 = 0.5$  ανεξάρτητα από τις τιμές των άλλων παραμέτρων. Όσο ψηλότερες είναι οι τιμές του  $B$  τόσο πληρέστερη είναι η τιμή του  $q_0$  στο  $q_1$ , όταν το  $A$  και  $C$  διατηρηθούν σταθερά.

Οι τρεις αθροιστικοί όροι του υποδείγματος με τη σειρά που εμφανίζονται εκφράζουν κύρια την θνησιμότητα της βρεφικής - παιδικής ηλικίας, των νεαρών ενηλίκων και των μεσήλικων-υπερήλικων αντίστοιχα.

Η παράμετρος  $A$  συνδέεται κύρια με το επίπεδο της παιδικής θνησιμότητας και παίρνει τιμές στο διάστημα  $(0,1)$ .

Η παράμετρος  $B$  συνδέεται με την θέση της  $q_0$  στο διάστημα  $(q_1, 0.5)$ . Για τις τιμές  $x \geq 1$  η συμβολή της είναι ασήμαντη.

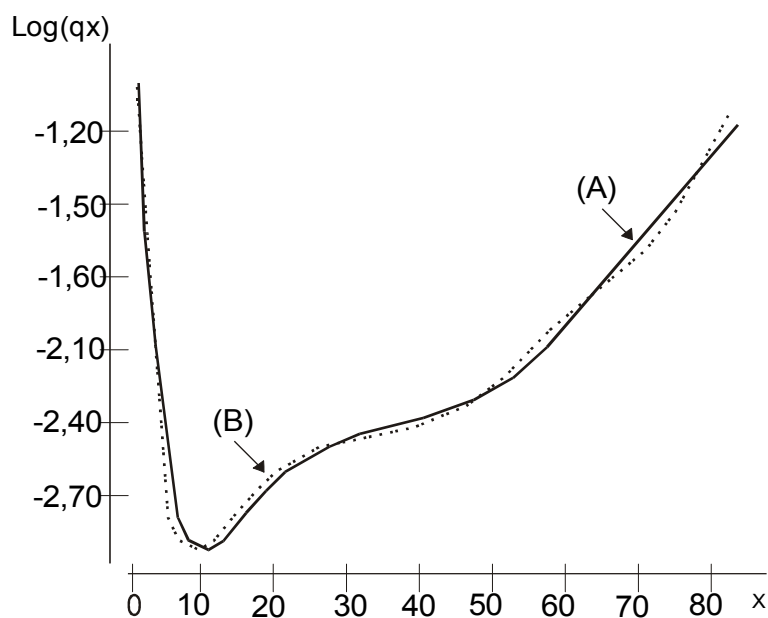
Η παράμετρος  $C$  είναι στενά συνδεδεμένη με τον ρυθμό μείωσης της θνησιμότητας της παιδικής ηλικίας, το ρυθμό δηλαδή που ο άνθρωπος προσαρμόζεται βιολογικά στο περιβάλλον στην αρχή της ζωής του. Η παράμετρος  $C$  παίρνει επίσης τιμές στο διάστημα  $(0, 1)$ . Ο δεύτερος αθροιστικός όρος του υποδείγματος Heligman - Pollard (στη συνέχεια θα αναφέρεται με την σύντμηση H&P8) περιγράφει την θνησιμότητα των ηλικιών 12 έως 35 περίπου. Αντανακλά την θνησιμότητα ατυχημάτων (accident hump) που επιδρά σε αυτές τις ηλικίες. Ο όρος αυτός μπορεί να θεωρηθεί σαν επιπρόσθετος στην βασική εξέλιξη της κατά ηλικία θνησιμότητας όπως αυτή εκφράζεται από τον πρώτο και τρίτο αθροιστικό όρο. Η μαθηματική έκφραση αυτού του όρου είναι παρόμοια της συνάρτησης πυκνότητας της λογαριθμοκανονικής κατανομής και περιλαμβάνει τρεις παραμέτρους:

Η  $F$  συνδέεται με τη θέση της καμπύλης ατυχημάτων στο άξονα της ηλικίας, η  $E$  συνδέεται αντίστροφα με το εύρος αυτής της καμπύλης (μεγάλες τιμές του  $E$  συνδέονται με μια ευρεία καμπύλη ατυχημάτων και αντίστροφα), η παράμετρος  $D$  συνδέεται με την ένταση αυτής της καμπύλης. Τέλος, ο τρίτος όρος του υποδείγματος H&P8 περιγράφει την θνησιμότητα των ηλικιών μετά τα 35 περίπου. Είναι εκθετικής μορφής όπως το υπόδειγμα Gompertz και αντανακλά την αυξητική πορεία της κατά ηλικία θνησιμότητας αυτών των ηλικιών. Ο όρος αυτός περιλαμβάνει 2 παραμέτρους, την  $G$  η οποία συνδέεται με το επίπεδο θνησιμότητας και η  $H$  που αντανακλά το ρυθμό αύξησης της θνησιμότητας αυτών των ηλικιών.

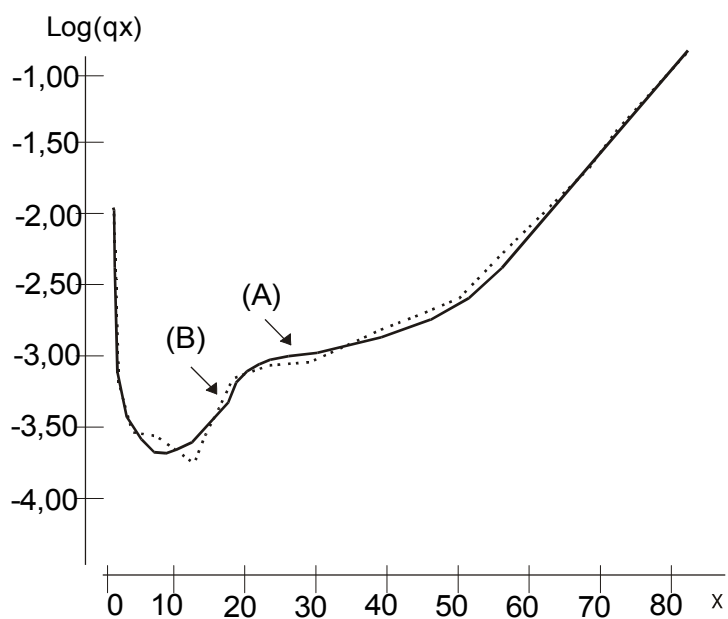
Οι διάφορες εφαρμογές του υποδείγματος των Heligman και Pollard σε ένα ευρύτατο φάσμα εμπειρικών δεδομένων διαφόρων χωρών έδειξαν ότι το υπόδειγμα παρέχει μια πολύ ικανοποιητική περιγραφή της κατά ηλικία θνησιμότητας στους διάφορους πληθυσμούς. Εν τούτοις ενέχει κάποια πηγή συστηματικού σφάλματος, το οποίο προέρχεται από το δεύτερο αθροιστικό όρο του μοντέλου. Ο όρος αυτός ο οποίος περιγράφει την καμπύλη ατυχημάτων (accident hump) δίνει συστηματικά στο σύνολο των εφαρμογών σε εμπειρικά δεδομένα, μια μετατόπιση της καμπύλης αυτής προς τα δεξιά. Η καμπύλη δηλαδή επικεντρώνεται σε μεγαλύτερη ηλικία από εκείνη που εκφράζουν τα εμπειρικά δεδομένα.

(Βλέπε διαγράμματα 3.3 και 3.4)

**Διάγραμμα 3.3.** Εμπειρικές τιμές του Σουηδικού γυναικείου πληθυσμού της περιόδου 1926-30 και εκτιμήσεις τους με τη χρήση των μοντέλων H&P8 (γραμμή A) και H&P9 (γραμμή B).



**Διάγραμμα 3.4** Εμπειρικές τιμές του Γερμανικού ανδρικού πληθυσμού της περιόδου 1988-89 και εκτιμήσεις τους με χρήση των μοντέλων H&P8 (γραμμή A) και H&P9 (γραμμή B).



Η μετατόπιση αυτή αν και συστηματική, δεν είναι το ίδιο έντονη στους διάφορους πληθυσμούς. Συνέπεια αυτής της μετατόπισης της καμπύλης ατυχημάτων (όπου αυτή εμφανίζεται έντονη) είναι κάποια υποεκτίμηση της θνησιμότητας των πολύ μεγάλων ηλικιών (μετά τα 70).

Αυτό ίσως αποτελεί κάποιο πρόβλημα, ιδιαίτερα σε εφαρμογές όπου μας ενδιαφέρει ή όσο το δυνατόν καλύτερη αποτύπωση της κατά-ηλικία θνησιμότητας (πχ. αναλογιστικές εφαρμογές).

Για την εξάλειψη αυτού του προβλήματος, μια απλή τροποποίηση του μεσαίου όρου του μοντέλου H&P8 προστέθηκε στην βιβλιογραφία το 1992 από την Κωστάκη<sup>15</sup> με την εισαγωγή μιας ακόμα παραμέτρου:

Το μοντέλο των εννέα παραμέτρων (H&P9) είναι της μορφής:

$$\frac{q_x}{p_x} = \begin{cases} A^{(x+B)^C} + D \exp \left[ -E_1 \left( \log \frac{x}{F} \right)^2 \right] + GH^x & \text{για } x \leq F \\ A^{(x+B)^C} + D \exp \left[ -E_2 \left( \log \frac{x}{F} \right)^2 \right] + GH^x & \text{για } x \geq F \end{cases}$$

Ο πρώτος και ο τρίτος αθροιστικός όρος της H&P9 είναι ίδιοι με τους αντίστοιχους H&P8, ενώ στο δεύτερο όρο η παράμετρος  $E$  έχει αντικατασταθεί από δύο παραμέτρους και οι οποίες συνδέονται με τη διασπορά της καμπύλης ατυχημάτων αριστερά και δεξιά από την κορυφή αντίστοιχα.

Τα διαγράμματα 3.3 και 3.4 παρουσιάζουν εφαρμογές των δύο διαφορετικών εκδόσεων του υποδείγματος Heligman - Pollard (H&P8 και H&P9) σε εμπειρικά δεδομένα.

### 3.8. Επιλογή Πινάκων Επιβίωσης

Ο πίνακας επιβίωσης αποτελεί βασικό εργαλείο και αναγκαίο βοήθημα του δημογράφου για την ανάλυση της θνησιμότητας του πληθυσμού αλλά και άλλες προσεγγίσεις όπως οι προβολές πληθυσμού, η μελέτη άλλων δημογραφικών φαινομένων όπως της γαμηλιότητας του πληθυσμού. Για να καταρτιστεί ένας πίνακας επιβίωσης απαιτούνται αξιόπιστα στοιχεία κατά ηλικία ή ομάδες ηλικιών τα οποία αφορούν τους θανάτους και τα αντίστοιχα πληθυσμιακά μεγέθη της χρονικής περιόδου στην οποία αναφέρεται ο πίνακας. Σε πολλούς πληθυσμούς όμως λόγω έλλειψης τέτοιων στοιχείων για το σύνολο του πληθυσμού ή λόγω γνωστών προβλημάτων στην ποιότητα αυτών των στοιχείων, αντί να κατασκευαστεί ένας πίνακας που θα βασίζεται στα ελλιπή ή προβληματικά δεδομένα, εναλλακτικά επιλέγεται ένας πίνακας κάποιας πρότυπης συλλογής πινάκων επιβίωσης σαν κατάλληλη προσέγγιση της κατά ηλικία θνησιμότητας του συγκεκριμένου πληθυσμού. Στη διαδικασία επιλογής ενός κατάλληλου πίνακα επιβίωσης ο οποίος θα λειτουργήσει σαν προσέγγιση του πραγματικού αλλά ελλιπή πίνακα, διεξάγονται στατιστικοί έλεγχοι οι οποίοι ελέγχουν την καταλληλότητα του με βάση τη μερική πληροφόρηση η οποία παρέχεται από τα ελλιπή ή ασταθή στοιχεία του πληθυσμού αναφοράς.

Οι στατιστικοί έλεγχοι που παρουσιάζονται στην συνέχεια εξετάζουν αν κάποιος συγκεκριμένος πίνακας που επιλέγεται από κάποια τυπική συλλογή είναι κατάλληλος να χρησιμοποιηθεί σαν ικανοποιητική προσέγγιση του συγκεκριμένου πληθυσμού αναφοράς.

<sup>15</sup> Kostaki, A. (1992): "A nine-parameter version of the Heligman-Pollard formula" *Mathematical Population Studies*, Vol 3 (4), 277-288.

### 3.9 Έλεγχοι υποθέσεων

Έστω  $\theta_x$  ο παρατηρούμενος αριθμός θανάτων ατόμων ηλικίας  $x$  και  $E_x$  ο παρατηρούμενος εκτεθειμένος στον κίνδυνο του θανάτου αριθμός ατόμων ηλικίας  $x$  τότε το

$q_x = \frac{\theta_x}{E_x}$  αποτελεί την παρατηρούμενη συχνότητα θανάτου στην ηλικία  $x$ .

Έστω ακόμα  $q_x$  ότι είναι οι αντίστοιχες πιθανότητες θανάτου όπως αυτές παρουσιάζονται σε κάποιο τυπικό πίνακα επιβίωσης.

Σύμφωνα με την υπόθεση μας,  $H_0$  τα εμπειρικά δεδομένα όπως αυτά εκφράζονται από τα  $q_x$ , προέρχονται από πληθυσμό με κατά ηλικία πιθανότητες θανάτου  $q_x$ , αυτές του τυπικού πίνακα επιβίωσης. Αντίθετα σύμφωνα με την εναλλακτική υπόθεση  $H_1$  ο τυπικός πίνακας θεωρείται ακατάλληλος να εκφράσει την κατά ηλικία θνησιμότητα του πληθυσμού αναφοράς.

Αν η  $H_0$  είναι αληθής τότε το  $\theta_x$  μπορεί να θεωρηθεί τυχαία διακριτή μεταβλητή η οποία ακολουθεί διωνυμική κατανομή με παραμέτρους  $E_x$  και  $q_x$ :

$$\theta_x \sim \text{Bin}(E_x, q_x)$$

όπου

$$E(\theta_x) = E_x q_x \text{ και}$$

$$\text{Var}(\theta_x) = E_x q_x p_x \text{ με } p_x = 1 - q_x$$

Έτσι:

$$\theta_x \sim N(E_x q_x, \sqrt{E_x q_x p_x})$$

επομένως

$$\frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}} \sim N(0, 1)$$

και κατά συνέπεια:

$$\left[ \frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}} \right]^2 \sim \chi^2(1)$$

Άρα

$$\sum_x \left[ \frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}} \right]^2 \sim \chi^2(n)$$

όπου  $n$  ο αριθμός των ηλικιών ή των ομάδων ηλικιών που συμμετέχουν.

Η συνάρτηση ελέγχου λοιπόν είναι της μορφής:

$$\sum_x \left[ \frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}} \right]^2$$



Αν η τιμή της συνάρτησης ελέγχου είναι μεγαλύτερη της αντίστοιχης τιμής του πίνακα της κατανομής  $\chi^2$  για επιλεγμένο επίπεδο σημαντικότητας  $\alpha$  και βαθμούς ελευθερίας  $n$  τότε η  $H_0$  απορρίπτεται.

Το παράδειγμα που ακολουθεί αναφέρεται σε δεδομένα πληθυσμού (στήλη (2) και (3) του πίνακα 1) ηλικιών 35 έως 70 ετών αναλογιστικών πινάκων και στα αντίστοιχα μεγέθη του τυπικού πίνακα ELT 10 (της σειράς English Life, Tables Number 10, Males).

Σύμφωνα με τους υπολογισμούς του Πίνακα 1, η τιμή της συνάρτησης ελέγχου είναι 34,48. Η τιμή είναι μικρότερη από την αντίστοιχη τιμή του πίνακα της κατανομής  $\chi^2$  για  $\alpha=0.05$  και 36 βαθμούς ελευθερίας, (ίση με 50). Άρα βάσει των αποτελεσμάτων αυτού του ελέγχου, για  $\alpha=0.05$  ή  $H_0$  δεν απορρίπτεται.

**Πίνακας 1** Παράδειγμα ελέγχου  $\chi^2$

E.L.T. 10 Males

x	$E_x$	$\theta_x$	$q_x$	$E_x q_x$	$E_x p_x q_x$	$(\hat{e}_x - E_x \cdot q_x) / E_x \cdot q_x \cdot p_x$	(7) <sup>2</sup>
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
35	1051	1	.00421	4.42	4.41	-1.63	2.66
36	940	6	.00447	4.20	4.18	0.88	0.77
37	1048	2	.00474	4.97	4.94	-1.33	1.78
38	716	3	.00502	3.59	3.57	-0.31	0.10
39	719	2	.00531	3.82	3.80	-0.93	0.87
40	1051	4	.00562	5.91	5.87	-0.79	0.62
41	1042	4	.00598	6.23	6.19	-0.90	0.80
42	1804	12	.00639	11.53	11.45	0.14	0.02
43	1468	7	.00687	10.09	10.02	-0.97	0.95
44	1576	16	.00741	11.68	11.59	1.27	1.61
45	1647	18	.00799	13.16	13.05	1.34	1.79
46	1861	16	.00861	16.02	15.89	-0.01	0.00
47	1669	13	.00925	15.44	15.30	-0.62	0.39
48	1624	14	.00990	16.08	15.92	-0.52	0.27
49	1157	11	.01057	12.23	12.10	-0.35	0.12
50	2193	19	.0128	24.74	24.46	-1.16	1.35
51	1803	20	.01206	21.74	21.48	-0.38	0.14
52	2402	31	.01295	31.11	30.70	-0.02	0.00
53	2120	27	.01393	29.53	29.12	-0.47	0.22
54	2406	38	.01499	36.07	35.52	0.32	0.11
55	1975	37	.01614	31.88	31.36	0.91	0.84
56	2564	38	.01744	44.72	43.94	-1.01	1.03
57	1798	36	.01890	33.98	33.34	0.35	0.12
58	2536	51	.02050	51.99	50.92	-0.14	0.02
59	2511	71	.02224	55.84	54.60	2.05	4.21
60	1858	32	.02415	44.87	43.79	-1.95	3.78
61	1835	54	.02630	48.26	46.99	0.84	0.70
62	1393	47	.02875	40.05	38.90	1.11	1.24
63	1462	40	.03150	46.05	44.60	-0.91	0.82
64	1245	34	.03455	43.01	41.53	-1.40	1.96
65	1064	46	.03791	40.34	38.81	0.91	0.83
66	1502	74	.04162	62.51	59.91	1.48	2.20
67	875	36	.04568	39.97	38.14	-0.64	0.41
68	927	38	.05014	46.48	44.15	-1.28	1.63
69	497	29	.05502	27.34	25.84	0.33	0.11
70	983	60	.06033	59.32	55.74	0.09	0.01
Total	-	-	-	-	-	-	34.48

## Περιορισμός του ελέγχου $\chi^2$

Ακόμα και αν η  $H_0$  δεν απορριφθεί με τον έλεγχο  $\chi^2$  τα δεδομένα χρειάζονται βαθύτερη διερεύνηση. Απαιτείται δηλαδή η διεξαγωγή συμπληρωματικών ελέγχων για την διερεύνηση συστηματικών αποκλίσεων μεταξύ των εμπειρικών  $q_x$  και εκείνων του τυπικού πίνακα,  $q_x$ . Ο γενικός έλεγχος  $\chi^2$  αποτυγχάνει πολλές φορές να ανακαλύψει προβλήματα όπως:

(α) την ύπαρξη ενός αριθμού υψηλών αποκλίσεων για κάποια  $x$  που ισοβαθμίζονται από έναν αριθμό πολύ χαμηλών αποκλίσεων για άλλες ηλικίες.

(β) υψηλές αθροιστικές αποκλίσεις που αφορούν μόνο ένα μέρος του διαστήματος ηλικιών.

(γ) υψηλό αριθμό θετικών ή αρνητικών αποκλίσεων για ένα μέρος του διαστήματος ηλικιών ή για όλες.

(δ) πολλές αρνητικές ή θετικές αποκλίσεις στη σειρά.

Ασυνέπειες του τύπου (α) μπορούν να εντοπιστούν συγκρίνοντας την τιμή του

$$\left[ \frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}} \right]^2$$

για κάποιες συγκεκριμένες ηλικίες,  $x_1$  με την τιμή του πίνακα της κατανομής  $\chi^2$  δια προκαθορισμένο επίπεδο σημαντικότητας  $\alpha$  και 1 βαθμό ελευθερίας.

Ασυνέπειες των τύπων (β), (γ) ή (δ) δείχνουν ότι η σειρά των κατά ηλικία πιθανοτήτων θανάτου που αφορούν τον πληθυσμό αναφοράς βρίσκεται σε σημαντικά υψηλότερα ή χαμηλότερα επίπεδα από την αντίστοιχη σειρά των πιθανοτήτων θανάτου του τυπικού πίνακα, με σοβαρές οικονομικές συνέπειες για τον ασφαλιστικό φορέα.

Έτσι καταλήγουμε ότι ένας γενικός έλεγχος  $\chi^2$  δεν αρκεί και ότι συμπληρωματικοί έλεγχοι θα ήταν αναγκαίοι για τον έλεγχο των επί μέρους πιθανών προβλημάτων.

Στη συνέχεια θα παρουσιαστούν οι εξής έλεγχοι:

- (i) Μεμονωμένων τυποποιημένων αποκλίσεων
- (ii) Μεμονωμένων απόλυτων τυποποιημένων αποκλίσεων
- (iii) Αθροιστικών αποκλίσεων για μέρη του Πίνακα
- (iv) Προσήμων
- (v) Ομάδων Προσήμων
- (vi) Αλλαγών Προσήμων

Οι έλεγχοι (i), (ii) και (iii) εξετάζουν αν οι τυποποιημένες αποκλίσεις ακολουθούν κανονική κατανομή, ενώ οι έλεγχοι (iv), (v) και (vi) εξετάζουν αν οι αποκλίσεις αυτές είναι τυχαία κατανεμημένες ως προς την ηλικία.

Για το σύνολο των ελέγχων η υπόθεση  $H_0$  είναι πάντοτε η ίδια δηλαδή ότι ο τυπικός πίνακας είναι κατάλληλος να χρησιμοποιηθεί σαν προσέγγιση της θνησιμότητας του πληθυσμού αναφοράς, ενώ σύμφωνα με την εναλλακτική υπόθεση  $H_1$  τα μεγέθη του τυπικού πίνακα εμφανίζουν στατιστικά σημαντικές αποκλίσεις από τα αντίστοιχα του πληθυσμού αναφοράς.

(i) Έλεγχος Μειονομένων Τυποποιημένων Αποκλίσεων

**Χρήση:** Διερεύνηση ύπαρξης σφαλμάτων τύπου (α).

Αν η  $H_0$  είναι αληθής τότε ο παρατηρούμενος αριθμός θανάτων στην ηλικία  $x$ ,  $\theta_x$  μπορεί να θεωρηθεί ότι ακολουθεί κατά προσέγγιση κανονική κατανομή με αναμενόμενη τιμή  $E_x q_x$  και διακύμανση  $E_x q_x p_x$ . Ακόμα αν θεωρήσουμε ότι οι θάνατοι στις διάφορες διαδοχικές ηλικίες επέρχονται ανεξάρτητα, τότε οι τυποποιημένες αποκλίσεις

$$\frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}}$$

μπορεί να θεωρηθούν ότι αποτελούν ανεξάρτητες παρατηρήσεις της τυποποιημένης κανονικής κατανομής.

Κατά συνέπεια θα πρέπει (αν η  $H_0$  είναι αληθής) όχι περισσότερο από 5% του συνόλου των τυποποιημένων αποκλίσεων να ξεπερνούν κατ' απόλυτη τιμή την τιμή

$$z_{1-\alpha/2} (=1,96 \text{ για } \alpha=0,05)$$

Σύμφωνα με την τυποποιημένη κανονική κατανομή τα ποσοστά των τυποποιημένων αποκλίσεων που βρίσκονται χονδρικά εντός των ορίων  $(-\infty, -3)$ ,  $(-3, -2)$ ,  $(-2, -1)$ ,  $(-1, 0)$ ,  $(0, 1)$ ,  $(1, 2)$ ,  $(2, 3)$ ,  $(3, +\infty)$  είναι 0%, 2%, 14%, 34%, 34%, 14%, 2% και 0% αντίστοιχα.

Έτσι οι αριθμοί των τυποποιημένων αποκλίσεων που βρίσκονται εντός αυτών των ορίων πρέπει να συγκριθούν με τους αντίστοιχους αναμενόμενους αριθμούς που θα υπολογιστούν με βάση τον αριθμό παρατηρήσεων και τα προαναφερθέντα ποσοστά.

Η σύγκριση αυτή θα γίνει με τον έλεγχο  $\chi^2$  και συνάρτηση ελέγχου:

$$\sum_{i=1}^k \frac{(f_i - E(f_i))^2}{E(f_i)}$$

όπου  $k$  ο αριθμός των διαστημάτων.

Αν δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των εμπειρικών και των αναμενόμενων συχνοτήτων αν δηλαδή η  $H_0$  είναι αληθής τότε η συνάρτηση ελέγχου θα ακολουθεί κατανομή  $\chi^2$  με  $(k - 1)$  βαθμούς ελευθερίας.

Στον πίνακα 2 (στήλη 1) δίνονται οι συχνότητες των τυποποιημένων αποκλίσεων που βρίσκονται μέσα στα όρια των διαφόρων διαστημάτων όπως αυτές προκύπτουν σύμφωνα με τη στήλη (7) του Πίνακα 1. Στη στήλη (2) του Πίνακα 2 εμφανίζονται οι αντίστοιχες αναμενόμενες συχνότητες.

**Πίνακας 2** Παρατηρούμενος και αναμενόμενος συχνότητες τυποποιημένων αποκλίσεων

Όρια διαστημάτων	(1) $f_i$	(2) $E(f_i)$
$(-\infty, -3)$	0	$36 \cdot 0,00 = 0,00$
$(-3, -2)$	0	$36 \cdot 0,02 = 0,72$
$(-3, -1)$	7	$36 \cdot 0,14 = 5,04$
$(-1, 0)$	15	$36 \cdot 0,34 = 12,24$
$(0, 1)$	9	$36 \cdot 0,34 = 12,24$
$(1, 2)$	4	$36 \cdot 0,14 = 5,04$
$(2, 3)$	1	$36 \cdot 0,02 = 0,72$
$(3, +\infty)$	0	$36 \cdot 0,00 = 0,00$

Λαμβάνοντας υπόψιν τον προορισμό του ελέγχου  $\chi^2$  σύμφωνα με τον οποίο η αναμενόμενη συχνότητα κάθε διαστήματος πρέπει να είναι μεγαλύτερη του 5, οι πρώτες τρεις και οι τελευταίες τρεις τάξεις ενοποιούνται και έτσι έχουμε

$$\begin{aligned} \chi^2 &= \sum_{i=1}^4 \frac{(f_i - E(f_i))^2}{E(f_i)} \\ &= \frac{(7 - 5,76)^2}{5,76} + \frac{(15 - 12,24)^2}{12,24} + \frac{(9 - 12,24)^2}{12,24} + \frac{(5 - 5,76)^2}{5,76} \\ &= 1,85 \end{aligned}$$

Η τιμή 1.85 όμως είναι μικρότερη της αντίστοιχης τιμής του πίνακα της κατανομής  $\chi^2$  για  $\alpha=0.05$  και  $(k-1)=3$  βαθμούς ελευθερίας ( $\chi_{0,05}^2 = 7.815$ ). Έτσι η υπόθεση  $H_0$  ότι δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ εμπειρικών και αναμενόμενων αποκλίσεων δεν απορρίπτεται και κατά συνέπεια αποδεχόμαστε ότι ο τυπικός πίνακας εκφράζει τα εμπειρικά δεδομένα.

**(ii) Έλεγχος μεμονομένων απολύτων τυποποιημένων αποκλίσεων**

**Χρήση:** Διερεύνηση σφαλμάτων τύπου (α).

Αν η  $H_0$  (όπως προηγούμενα έχει οριστεί) είναι αληθής τότε οι τυποποιημένες αποκλίσεις

$$\frac{\theta_x - E_x q_x}{\sqrt{E_x q_x p_x}}$$

μπορεί να θεωρηθούν ότι αποτελούν ανεξάρτητες παρατηρήσεις της τυποποιημένης κανονικής κατανομής.

Γνωρίζουμε ότι

$$P\left(z < \frac{2}{3}\right) = 0.50$$

Άρα οι μισές από τις τιμές των τυποποιημένων αποκλίσεων θα πρέπει κατ' απόλυτη τιμή να ξεπερνούν τα 2/3.

Αν έχουμε  $n$  ηλικίες και συμβολίζουμε με  $N$  τον αριθμό των απόλυτων τυποποιημένων αποκλίσεων των  $n$  αυτών ηλικιών που είναι μεγαλύτερες του 2/3 τότε:

Αν η  $H_0$  είναι αληθής το  $N$  μπορεί να θεωρηθεί διακριτή μεταβλητή που ακολουθεί διωνυμική κατανομή με παράμετρος  $n$  και 1/2.

Άρα

$$E(N) = \frac{n}{2} \text{ και } \text{Var}(N) = \frac{n}{4} \Rightarrow \sigma_n = \frac{\sqrt{n}}{2}$$

Έτσι το  $N$  θα αποτελέσει τη συνάρτηση ελέγχου διωνυμικού μονοκατάληκτου ελέγχου (ή απορρίπτεται αν η τιμή του  $N$  βρίσκεται στο 100% ανώτατο όριο της διωνυμικής κατανομής με παραμέτρους  $n$  και 1/2). (βλέπε Πίνακα 3).

Αν ο αριθμός των ηλικιών που θεωρούνται είναι επαρκώς μεγάλος ( $n > 20$ ), τότε η κατανομή της  $N$  μπορεί να θεωρηθεί προσεγγιστικά κανονική με παραμέτρους  $n/2$  και  $\frac{\sqrt{n}}{2}$ .

Έτσι η συνάρτηση ελέγχου μετατρέπεται σε:

$$z = \frac{N - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = \frac{2N - n}{\sqrt{n}}$$

η οποία αν η  $H_0$  είναι αληθής, ακολουθεί τυποποιημένη κανονική κατανομή. Έτσι η  $H_0$  απορρίπτεται αν η  $Z > Z_{1-\alpha}$ . Στο παράδειγμα του Πίνακα 1, οι 21 από τις 36 τυποποιημένες αποκλίσεις της στήλης (7) είναι μεγαλύτερες του 2/3, έτσι η συνάρτηση ελέγχου είναι ίση με:

$$\frac{2 \cdot 21 - 36}{\sqrt{36}} = 1.00 < 1.645$$

Άρα με  $\alpha = 0.05$  η  $H_0$  δεν απορρίπτεται.

### Πίνακας 3

Το άνω 100 α% όριο της δυνωμικής κατανομής με παραμέτρους  $n$  και  $1/2$ . (Οι τιμές του πίνακα περιλαμβάνονται στην περιοχή απόρριψης).

Το κάτω 100α% όριο βρίσκεται αφαιρώντας των τιμή του Πίνακα από το αντίστοιχο  $n$ .

n	a=0,005	0,01	0,025	0,05
5	-	-	-	5
6	-	-	6	6
7	-	7	7	7
8	8	8	8	7
9	9	9	8	8
10	10	10	9	9
11	11	10	10	9
12	11	11	10	10
13	12	12	11	10
14	13	12	12	11
15	13	13	12	12
16	14	14	13	12
17	15	14	13	13
18	15	15	14	13
19	16	15	15	14
20	17	16	15	15

#### (iii) Έλεγχος αθροιστικών αποκλίσεων

**Χρήση:** Διερεύνηση ύπαρξης σφαλμάτων τύπου (β).

Αν η  $H_0$  είναι αληθής τότε  $\theta_x \sim N(E_x q_x, \sqrt{E_x q_x p_x})$  και κατά συνέπεια αν τα  $\theta_x$  είναι ανεξάρτητα μεταξύ τους τότε:

$$\sum_{x=x_1}^{x_k} (\theta_x - E_x q_x) \sim N\left(0, \sqrt{\sum_{x=x_1}^{x_k} E_x q_x p_x}\right) \Rightarrow$$

$$\frac{\sum_{x=x_1}^{x_k} (\theta_x - E_x q_x)}{\sqrt{\sum_{x=x_1}^{x_k} E_x q_x p_x}} \sim N(0,1)$$

Έτσι η  $H_0$  απορρίπτεται αν η απόλυτη τιμή αυτής της συνάρτησης ξεπερνά το  $a/2$  ανώτατο όριο της τυποποιημένης κατανομής.

Ο έλεγχος αυτός μπορεί να εφαρμοστεί στο συνολικό διάστημα ηλικίας που έχει αρχικά θεωρηθεί ή σε μέρη του. Βασική προϋπόθεση όμως είναι η επιλογή των διαστημάτων ηλικιών να μην συνδέεται με τις τιμές των αποκλίσεων (πχ. επιλογή διαστήματος ηλικιών στις οποίες αντιστοιχούν αποκλίσεις με ίδια πρόσημα). Η επιλογή των διαστημάτων πρέπει να γίνεται τυχαία. Ακόμα δεν πρέπει να εφαρμόζεται σε επικαλυπτόμενα διαστήματα ηλικίας γιατί σε τέτοια περίπτωση το κοινό επίπεδο σημαντικότητας πάυει να είναι το

προκαθορισμένο  $\alpha$  (το αναμενόμενο ποσοστό των ελέγχων οι οποίοι καταλήγουν σε απόρριψη της  $H_0$  όταν η  $H_0$  είναι αληθής παύει να είναι ίσο με  $\alpha$ ).

Στο παράδειγμα του Πίνακα 1, εφαρμόζοντας τον έλεγχο αυτό στο συνολικό διάστημα ηλικίας (βλέπε Πίνακα 4) έχουμε: τιμή της συνάρτησης ελέγχου ίση με:

$$\frac{-12.15}{\sqrt{972.12}} = 0.39$$

**Πίνακας 4** Αθροιστικές αποκλίσεις του παραδείγματος του Πίνακα 1

Ηλικία $x$	Αθροιστικές Αποκλίσεις	Διακύμανση	Ηλικία $x$	Αθροιστικές Αποκλίσεις	Διακύμανση
35	-3,42	4,41	53	-20,47	244,04
36	-1,62	8,59	54	-18,54	279,56
37	-4,59	13,53	55	-13,42	310,92
38	-5,18	17,10	56	-20,14	354,86
39	-7,00	20,90	57	-18,12	388,20
40	-8,91	26,77	58	-19,11	439,12
41	-11,14	32,96	59	-3,95	493,72
42	-10,67	44,41	60	-16,82	537,51
43	-13,76	54,43	61	-11,08	584,50
44	-9,44	66,02	62	-4,13	623,40
45	-4,60	79,07	63	-10,18	668,00
46	-4,62	94,96	64	-19,19	709,53
47	-7,06	110,26	65	-13,53	748,34
48	-9,14	126,18	66	-2,04	808,25
49	-10,37	138,28	67	-6,01	846,39
50	-16,09	162,74	68	-14,49	890,54
51	-17,83	184,22	69	-12,83	916,38
52	-17,94	214,92	70	-12,15	972,12

Έτσι για  $\alpha=0.05$  η  $H_0$  δεν απορρίπτεται εφ' όσον  $|-0.39| < 1.96$ . Ένας τρόπος αναλυτικότερης διερεύνησης με τη χρήση αυτού του ελέγχου είναι να χωρίσουμε τα στοιχεία σε τέσσερις ισοπληθείς ομάδες (οκτώ ηλικιών η κάθε μία) και να διεξάγουμε τον έλεγχο σε κάθε μία από αυτές ξεχωριστά. Οι συναρτήσεις ελέγχου θα είναι:

$$\text{Διάστημα [35, 43]: } \frac{-13.76}{\sqrt{54.34}} = 1.86$$

$$\text{Διάστημα [44,52]: } \frac{[-17.94 - (-13.76)]}{\sqrt{214.92 - 54.34 - 0.33}} = -0.33$$

$$\text{Διάστημα [53, 61]: } \frac{[-11.8 - (-17.94)]}{\sqrt{584.50 - 214.92}} = 0.36$$

$$\text{Διάστημα [62, 70]: } \frac{[-12.15 - (-11.08)]}{\sqrt{972.12 - 584.50}} = -0.05$$

και στις τέσσερις περιπτώσεις για  $\alpha=0.05$  η  $H_0$  δεν απορρίπτεται.

#### (iv) Έλεγχος προσήμων

**Χρήση:** Διερεύνηση ύπαρξης σφαλμάτων τύπου ( $\gamma$ )

Αν η  $H_0$  είναι αληθής τότε οι αποκλίσεις των  $\theta_x$  από τις αναμενόμενες τιμές τους ( $E_x q_x$ ) είναι ανεξάρτητες τυχαίες μεταβλητές. Αν εμφανιστεί μεγάλος αριθμός θετικών ή αρνητικών προσήμων θα είναι ένδειξη συστηματικής απόκλισης των παρατηρούμενων  $\theta_x$  από τα αντίστοιχα μεγέθη του τυπικού πίνακα  $E_x q_x$ . Άρα ο τυπικός πίνακας επιβίωσης θα υποεκτιμά ή θα υπερεκτιμά συστηματικά την εμπειρική πραγματικότητα.

Αν έχουμε  $n$  ηλικίες, και συμβολίζουμε με  $N$  τον αριθμό των προσήμων των αποκλίσεων ( $\theta_x - E_x q_x$ ) με την μεγαλύτερη συχνότητα τότε:

Αν  $H_0$  είναι αληθής το  $N \sim \text{Bin}(n, 1/2)$

Έτσι για προκαθορισμένο επίπεδο σημαντικότητας  $\alpha$ , η  $H_0$  απορρίπτεται αν η τιμή του  $N$  βρίσκεται στο άνω 5% όριο της δυνωμικής κατανομής.

Αν το  $n$  είναι μεγάλο τότε:

$$N \sim N\left(\frac{n}{2}, \frac{\sqrt{n}}{2}\right)$$

και έτσι η συνάρτηση ελέγχου μετατρέπεται σε:

$$T = \frac{N - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \quad \text{ή} \quad \frac{2N - n}{\sqrt{n}}$$

η οποία αν  $H_0$  είναι αληθής, ακολουθεί τυποποιημένη κανονική κατανομή.

Έτσι η  $H_0$  απορρίπτεται αν  $T > Z_{1-\alpha}$

Στο παράδειγμα του πίνακα 1 έχουμε:  $N=22$ ,  $n=36$

$$T = \frac{2 \cdot 22 - 36}{\sqrt{36}} = 1.33 < 1.645$$

Έτσι για  $\alpha=0.05$  η  $H_0$  δεν απορρίπτεται.

#### (v) STEVEN'S test για ομάδες προσήμων

**Χρήση:** Διερεύνηση ύπαρξης σφαλμάτων τύπου ( $\delta$ ).

Αν  $H_0$  αληθής, τα πρόσημα των αποκλίσεων είναι ανεξάρτητα μεταξύ τους και το ίδιο πιθανό να είναι θετικά ή αρνητικά.

Ακόμη και αν ο αριθμός θετικών και αριθμός αρνητικών προσήμων είναι αποδεκτός (βάσει του ελέγχου προσήμων) μπορεί εν τούτοις ο αριθμός των ομάδων θετικών προσήμων (και κατά συνέπεια και ο αριθμός ομάδων αρνητικών προσήμων) να είναι σημαντικά μικρός (αντίστοιχα σημαντικά μεγάλος ή αντίστροφα).

Έστω  $n_1$  θετικά πρόσημα.



Για να χωριστούν αυτά σε  $t$  ομάδες θα πρέπει να θέσουμε όρια μεταξύ των προσήμων σε  $(t-1)$  από τις  $(n_1 - 1)$  πιθανές θέσεις.

Για παράδειγμα αν έχουμε πέντε θετικά πρόσημα ( $n_1=5$ ) και θέλουμε να τα χωρίσουμε σε τρεις ομάδες ( $t=3$ ) τότε θα πρέπει να θέσουμε όρια σε  $(t-1)=2$  από τις  $(n_1 - 1)=4$  πιθανές θέσεις

$$\begin{array}{cccccc} + & + & + & + & + & + \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \end{array}$$

Ο αριθμός πιθανών τρόπων για να γίνει αυτό θα είναι:

$$\binom{n_1 - 1}{t - 1}$$

Αν τώρα  $n_2$  ο αριθμός των αρνητικών προσήμων τότε θα έχουμε  $n_2+1$  θέσεις να τοποθετήσουμε τις  $t$  ομάδες θετικών προσήμων και ο αριθμός των πιθανών μεταθέσεων θα είναι:

$$\binom{n_2 + 1}{t}$$

Για παράδειγμα αν  $n_2=5$  τότε τις τρεις ομάδες θετικών προσήμων ( $t=3$ ) μπορούμε να τις τοποθετήσουμε σε  $n_2+1=6$  πιθανές θέσεις, δια- γραμματικά:

$$\begin{array}{cccccc} - & - & - & - & - & n_2 = 5 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \end{array} \quad n_2 + 1 = 6 \text{ θέσεις}$$

Τέλος ο συνολικός αριθμός μεταθέσεων  $n_1$  θετικών και  $n_2$  αρνητικών είναι

$$\binom{n_1 + n_2}{n_1}$$

Έτσι η πιθανότητα  $p$  να έχουμε  $t$  ομάδες θετικών προσήμων αν έχουμε  $n_1$  θετικά και  $n_2$  αρνητικά πρόσημα θα είναι:

$$p = \frac{\binom{n_1 - 1}{t - 1} \binom{n_2 + 1}{t}}{\binom{n_1 + n_2}{n_1}} = \frac{\binom{n_1 - 1}{t - 1} \binom{n_2 + 1}{n_2 + 1 - t}}{\binom{n_1 + n_2}{n_1}}$$

Αν έχουμε  $g$  ομάδες θετικών προσήμων, η πιθανότητα να έχουμε το ανώτατο  $g$  ομάδες θετικών προσήμων ( $g$  ή λιγότερες) υπολογίζεται από το άθροισμα των  $p$  για όλα τα  $t \leq g$  ( $t \geq 1$ ). Αν αυτή η συνολική πιθανότητα είναι ( $\alpha=5$ ) τότε η  $H_0$  απορρίπτεται.

Μπορεί να αποδειχθεί ότι:

$$E(g) = \frac{n_1(n_1 + 1)}{n_1 + n_2}$$

$$\text{Var}(g) = \frac{(n_1 n_2)^2}{(n_1 + n_2)^3}$$

Έτσι αν  $H_0$  αληθής τότε :  $G = \frac{g - E(g)}{\sqrt{\text{Var}(g)}}$

Η  $H_0$  δεν γίνεται αποδεκτή αν η τιμή της  $G$  βρίσκεται στο 5% κατώτατο όριο της κανονικής κατανομής ήτοι  $G < -1.645$  (μονοκατάληκτος έλεγχος)

**Σημείωση:** Απορρίπτουμε την  $H_0$  αν το  $g$  είναι σημαντικά μικρότερο του  $E(g)$  δηλαδή αν το  $G$  παίρνει αρνητική τιμή και είναι μικρότερη κάποιου αρνητικού όριου (για  $\alpha=0.05$  το όριο είναι  $-1.645$ ).

Στα δεδομένα του Πίνακα 1 έχουμε 9 ομάδες θετικών προσήμων,  $g=9$ .

Ακόμα  $E(g)=8.94$   $\text{Var}(g)=2.03$

Άρα

$$G = \frac{g - E(g)}{\sqrt{\text{Var}(g)}} > 0$$

αφού  $g > E(g)$

Άρα η  $H_0$  δεν απορρίπτεται.

#### (vi) Διωνυμικό test για τις αλλαγές προσήμων

**Χρήση:** Διερεύνηση ύπαρξης σφαλμάτων τύπου ( $\delta$ )

Αν  $H_0$  είναι αληθής τότε τα πρόσημα δύο διαδοχικών αποκλίσεων είναι το ίδιο πιθανό να είναι ίδια ή αντίθετα (τυχαιότητα)

Έστω  $N$  ο αριθμός αλλαγών προσήμων μίας σειράς  $n$  προσήμων.

Έτσι η  $H_0$  απορρίπτεται αν το  $N$  βρίσκεται στο κατώτατο  $\alpha 100\%$  κάτω όριο της τυποποιημένης κανονικής κατανομής.

Στο παράδειγμα του Πίνακα 1

$N=17$ ,  $n=36$  και

$$S = \frac{2 \cdot 17 - 36 + 1}{\sqrt{31 - 1}} = -0.17 > -16.45$$

άρα η  $H_0$  δεν απορρίπτεται.

Εφ' όσον η  $H_0$  δεν απορριφτεί από το σύνολο αυτών των ελέγχων μπορούμε να θεωρήσουμε ότι ο τυπικός πίνακας επιβίωσης εκφράζει την πραγματικότητα όπως αυτή προσδιορίζεται από τα εμπειρικά στοιχεία. Αν ο πρώτος γενικός έλεγχος καταλήξει σε απόρριψη της  $H_0$  τότε είναι άσκοπο να συνεχίσουμε με τους υπόλοιπους ελέγχους. Απόρριψη της  $H_0$  στο γενικό έλεγχο είναι ένδειξη ότι ο τυπικός πίνακας είναι ακατάλληλος και θα πρέπει να προχωρήσουμε σε επιλογή ενός άλλου πίνακα.