

# ΚΕΦΑΛΑΙΟ 3

## ΜΗ ΠΑΡΑΜΕΤΡΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

### 3.1. Εισαγωγή

Ο Στατιστικός Έλεγχος Ποιότητας (*Statistical Quality Control*) είναι ένα χρήσιμο εργαλείο σε όλα τα στάδια εξέλιξης και παραγωγής ενός προϊόντος, όπως στον σχεδιασμό του, στον εργαστηριακό έλεγχο του σχεδιασμού που εφαρμόζεται, στη διάρκεια της παραγωγικής διαδικασίας καθώς και τον έλεγχο των εισερχόμενων υλικών που χρησιμοποιούνται. Ο αντικειμενικός σκοπός του Στατιστικού Ελέγχου Ποιότητας είναι η εξασφάλιση της υψηλής ποιότητας του παραγόμενου προϊόντος στα πλαίσια του βέλτιστου οικονομικού σχεδιασμού. Η ποιότητα ενός προϊόντος μπορεί να εκτιμηθεί και να ορισθεί με διάφορους τρόπους. Οι βασικές διαστάσεις της ποιότητας ενός αντικειμένου (*dimensions of Quality*) είναι η αξιοπιστία (*Reliability*), η απόδοση (*Performance*), η αντοχή (*Durability*), η εμφάνιση (*Aesthetics*), η δυνατότητα επισκευής (*Serviceability*) και η τήρηση των προδιαγραφών κατασκευής του (*Conformance to standards*). Λεπτομερής εξέταση των διαστάσεων της ποιότητας ενός αντικειμένου, δίνεται στην εργασία του Garvin (1987).

Κατά την παραγωγή ενός προϊόντος, η ικανοποίηση των αρχικών απαιτήσεων σχετικά με την ποιότητά του, μπορεί να εξασφαλισθεί μέσω μιας σταθερής και επαναλαμβανόμενης διαδικασίας, η οποία να λειτουργεί με μικρή μεταβλητότητα (*variability*) γύρω από τις προσδοκώμενες (ή ονομαστικές) τιμές της ποιότητας του αντικειμένου (βλ. Taguchi (1986)). Ο Στατιστικός Έλεγχος Διεργασιών (*Statistical Process Control*) εφαρμόζεται στην παρακολούθηση μιας παραγωγικής διαδικασίας για την επίτευξη της σταθερότητας και τη βελτίωση της ικανότητάς της, μέσω της μείωσης της μεταβλητότητας που παρουσιάζει. Τα βασικά εργαλεία που χρησιμοποιεί είναι το φύλλο ελέγχου (*Check sheet*), το διάγραμμα Pareto (*Pareto chart*), το διάγραμμα διασποράς (*Scatter diagram*), το Ιστόγραμμα (*Histogram*) και το διάγραμμα ελέγχου (*Control chart*). Για παράδειγμα, το διάγραμμα Pareto είναι ένα ραβδόγραμμα συχνοτήτων που χρησιμοποιείται για την ταξινόμηση των

ελαττωμάτων που προκύπτουν στην παραγωγή, ανάλογα με τη συχνότητα που αυτά εμφανίζονται (βλ. Albin & Friedman (1992) και Wilkinson (2006)). Για λεπτομερή περιγραφή του σχεδιασμού και της ερμηνείας των διαγραμμάτων *Pareto*, ο ενδιαφερόμενος αναγνώστης μπορεί να συμβουλευτεί το βιβλίο του Δαμιανού (1996).

Σε κάθε παραγωγική διαδικασία, ανεξάρτητα από το πόσο σωστά έχει σχεδιασθεί και πόσο προσεκτικά επιβλέπεται, υπάρχει μια έμφυτη ή φυσική μεταβλητότητα, η οποία οφείλεται στην επίδραση πολλών και (συνήθως) μικρών αναπόφευκτων αιτιών. Όταν η διεργασία λειτουργεί μόνο με την παρουσία της φυσικής μεταβλητότητας, τότε λέγεται πως βρίσκεται σε κατάσταση στατιστικού ελέγχου ή πιο απλά εντός ελέγχου (*in-control*). Διαφορετικά, όταν παρουσιάζονται μορφές ειδικής μεταβλητότητας που οφείλονται σε μηχανικές ρυθμίσεις, λειτουργικά λάθη ή ελαττωματικές πρώτες ύλες, τότε η διεργασία λέγεται πως βρίσκεται εκτός ελέγχου (*out-of-control*). Στις περιπτώσεις αυτές, ο Στατιστικός Έλεγχος Διεργασιών στοχεύει στη γρήγορη ανίχνευση των παραπάνω αιτιών, ώστε να διορθωθούν έγκαιρα μειώνοντας το οικονομικό κόστος που αποφέρει η παραγωγή ελαττωματικών προϊόντων. Αξίζει να σημειωθεί ότι σε πολλές περιπτώσεις ελέγχου, είναι εφικτό να εντοπισθούν μόνο ορισμένα αίτια που ευθύνονται για τη δυσλειτουργία της παραγωγής, ενώ ίσως η σημαντικότερη να παραμένει ανερμήνευτη. Για αναλυτικότερη περιγραφή και αντιμετώπιση του συγκεκριμένου φαινομένου, που ονομάζεται *masking problem*, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Guess *et al.* (1991) και Yaschin (2007).

Τα διαγράμματα ελέγχου, όπως αναφέρθηκε και παραπάνω, βοηθούν στον εντοπισμό ειδικών αιτιών μεταβλητότητας σε μια διεργασία, με στόχο την επίτευξη μιας εντός ελέγχου διαδικασίας. Διαισθητικά, σε περιπτώσεις ανεπιθύμητης μετατόπισης της διεργασίας, ένα διάγραμμα ελέγχου οφείλει να την ανιχνεύσει όσο το δυνατόν γρηγορότερα και να παράγει σήμα ότι η παραγωγική διαδικασία είναι εκτός ελέγχου. Είναι προφανές ότι η αποτελεσματικότητα ενός διαγράμματος ελέγχου καθορίζεται από το πόσο γρήγορα θα καταφέρει να δώσει έγκυρο σήμα ότι η διεργασία βρίσκεται εκτός ελέγχου. Επιπλέον, τα διαγράμματα ελέγχου χρησιμοποιούνται για την εκτίμηση των παραμέτρων μιας παραγωγικής διαδικασίας και γενικότερα για τον καθορισμό της ικανότητάς της, με σκοπό τη βελτίωση της ποιότητας του προϊόντος που παράγεται ή ισοδύναμα στη μείωση της μεταβλητότητας που παρουσιάζει.

Τα διαγράμματα Στατιστικού Ελέγχου Ποιότητας έχουν εισαχθεί στην εργασία του Shewhart (1926), ενώ στη συνέχεια έχουν προταθεί διάφορες παραλλαγές τους για την παρακολούθηση ενός συνεχούς χαρακτηριστικού. Για την περιγραφή του τρόπου κατασκευής και λειτουργίας ενός διαγράμματος ελέγχου, ως υποθέσουμε ότι σε μια παραγωγική διαδικασία ενδιαφερόμαστε να μελετήσουμε τη συμπεριφορά μιας κρίσιμης ποσότητας ενός (μετρήσιμου) χαρακτηριστικού  $X$  των προϊόντων που εξάγονται. Για το σκοπό αυτό λαμβάνουμε τυχαία δείγματα του προϊόντος σε διαφορετικές χρονικές στιγμές και χρησιμοποιώντας τις μετρήσεις αυτές, υπολογίζουμε την τιμή μιας κατάλληλα επιλεγμένης στατιστικής συνάρτησης που εκτιμά την υπό μελέτη κρίσιμη ποσότητα. Ένα τυπικό διάγραμμα ελέγχου είναι μια γραφική παράσταση του στατιστικής συνάρτησης που έχει μετρηθεί ή υπολογισθεί με τη βοήθεια ενός τυχαίου δείγματος, έναντι του αύξοντα αριθμού του δείγματος αυτού. Το διάγραμμα περιλαμβάνει δύο οριζόντιες γραμμές που ονομάζονται άνω και κάτω όριο ελέγχου (*upper and lower control limit*). Οι δύο ακραίες αυτές γραμμές έχουν κατάλληλα σχεδιασθεί ώστε, αν η διεργασία είναι εντός ελέγχου, τότε (σχεδόν) όλα τα σημεία να τοποθετούνται εντός των δύο γραμμών. Όσο τα δεδομένα εμφανίζονται εντός των ορίων ελέγχου και η συμπεριφορά τους δεν παρουσιάζει κάποια συστηματική συμπεριφορά, τότε υποθέτουμε ότι η παραγωγική διαδικασία βρίσκεται εντός ελέγχου και δεν προχωρούμε σε καμία διορθωτική παρέμβαση. Αντίθετα, αν κάποιο σημείο βρεθεί εκτός των ορίων ελέγχου, έχουμε ένδειξη ότι η διεργασία είναι εκτός ελέγχου και αντιμετωπίζουμε μια κατάσταση συναγερμού (*alarm*).

Στην εφαρμογή των διαγραμμάτων ελέγχου, δύο διαφορετικές φάσεις έχουν χρησιμοποιηθεί στη διεθνή βιβλιογραφία: Η Φάση I και η Φάση II (βλ. Woodall (2000)). Στη Φάση I, βασικός στόχος είναι ο έλεγχος παρελθοντικών δεδομένων, προκειμένου να διαπιστωθεί αν προέρχονται από εντός ελέγχου διεργασία ή όχι. Όταν η διαδικασία βρεθεί εντός ελέγχου, τότε μια συνήθης πρακτική είναι η εκτίμηση των παραμέτρων της με τη βοήθεια ενός δείγματος αναφοράς και χρήση αυτών για τον καθορισμό των ορίων ελέγχου που θα χρησιμοποιηθούν στη Φάση II. Γενικότερα, για το βέλτιστο οικονομικό σχεδιασμό των διαγραμμάτων ελέγχου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Bather (1963) και Lorenzen & Vance (1986).

Στη Φάση II, σκοπός είναι ο έλεγχος μελλοντικών δεδομένων ώστε να διαπιστωθεί αν η διεργασία συνεχίζει να παραμένει εντός ελέγχου ή αν έχει

μετατοπιστεί σε εκτός ελέγχου κατάσταση. Η πιο συνήθης μέθοδος για τη λήψη απόφασης είναι η εφαρμογή ενός διαγράμματος ελέγχου. Κάθε νέα παρατήρηση ή δείγμα παρατηρήσεων ελέγχεται και με τη βοήθεια συγκεκριμένων κανόνων διαπιστώνεται αν η κατάσταση της υπό παρακολούθηση διεργασίας έχει αλλάξει ή έχει μείνει σταθερή. Ο αριθμός των δειγμάτων ή υποσυνόλων που πρέπει να ληφθούν από την παραγωγική διαδικασία πριν το διάγραμμα ελέγχου δώσει το πρώτο σήμα για εκτός ελέγχου κατάσταση, είναι τυχαία μεταβλητή και ονομάζεται μήκος ροής (*run length*). Η κατανομή του μήκους ροής χρησιμοποιείται συχνά για την αξιολόγηση της απόδοσης ενός διαγράμματος ελέγχου, ενώ η αναμενόμενη τιμή αυτής της κατανομής, που ονομάζεται μέσο μήκος ροής (*average run length, ARL*), είναι ίσως το δημοφιλέστερο μέτρο σύγκρισης δύο διαγραμμάτων ελέγχου. Εξ ορισμού, το μήκος ροής είναι μια τυχαία μεταβλητή που παίρνει θετικές ακέραιες τιμές, συνεπώς το μέσο μήκος ροής (*ARL*) δεν είναι κατάλληλο μέτρο αξιολόγησης ενός διαγράμματος, όταν η κατανομή παρουσιάζει λοξότητα. Για το λόγο αυτό, σε πολλές περιπτώσεις χρησιμοποιούνται διαφορετικά μέτρα, όπως η διάμεσος ή άλλα ποσοστημόρια (βλ. Human & Graham (2008)). Το επιθυμητό είναι το μέσο μήκος ροής να είναι μεγάλο όταν η διεργασία βρίσκεται εντός ελέγχου και μικρό, όταν η διεργασία είναι εκτός ελέγχου. Η πιθανότητα λανθασμένου συναγερμού (*false alarm rate, FAR*) είναι η πιθανότητα να παράγει ένα διάγραμμα ελέγχου σήμα ότι η διεργασία μετατοπίστηκε σε εκτός ελέγχου κατάσταση, ενώ στην πραγματικότητα παραμένει εντός ελέγχου. Δύο διαγράμματα ελέγχου μπορούν να συγκριθούν με τη βοήθεια των εκτός ελέγχου μέσων μηκών ροής τους ( $ARL_{out}$ ), έχοντας όμως εξασφαλίσει ότι τα εντός ελέγχου μέσα μήκη ροής τους ( $ARL_{in}$ ) ή οι πιθανότητες λανθασμένου συναγερμού είναι (σχεδόν) ίσα. Με άλλα λόγια, αν δύο διαγράμματα ελέγχου σχεδιασθούν με τέτοιο τρόπο ώστε η απόδοση τους να είναι σχεδόν ίδια κατά τη διάρκεια της εντός ελέγχου λειτουργίας, τότε η συγκριτική τους αξιολόγηση μπορεί να επιτευχθεί με τη σύγκριση των  $ARL_{out}$  που παρουσιάζουν. Για περισσότερες λεπτομέρειες σχετικά με τον υπολογισμό του μέσου μήκους ροής ενός διαγράμματος ελέγχου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Fu *et al.* (2002) και Fu *et al.* (2003).

Στο παρόν Κεφάλαιο, θα παρουσιάσουμε σε συντομία γνωστά διαγράμματα ελέγχου για παρακολούθηση ενός ποιοτικού χαρακτηριστικού που έχουν προταθεί στη διεθνή βιβλιογραφία. Πιο συγκεκριμένα, στην Παράγραφο 3.2 θα αναφερθούμε

σε τεχνικές στατιστικού ελέγχου διεργασιών, στις οποίες βασική υπόθεση είναι ότι το υπό μελέτη χαρακτηριστικό ακολουθεί μια γνωστή κατανομή (συνήθως την κανονική κατανομή). Τα διαγράμματα αυτά είναι γνωστά ως παραμετρικά (*parametric control charts*) και μπορούν να εφαρμοσθούν μόνο αν η παραπάνω υπόθεση ικανοποιείται στην πράξη. Το αντικείμενο μελέτης της Παραγράφου 3.3, είναι διαγράμματα ελέγχου που μπορούν να εφαρμοσθούν ανεξάρτητα της κατανομής που ακολουθεί το υπό μελέτη χαρακτηριστικό και συνεπώς δεν προϋποθέτουν τη γνώση της κατανομής των παρατηρούμενων τιμών. Τα διαγράμματα αυτά είναι γνωστά ως μη παραμετρικά (*non-parametric*) ή ελεύθερα κατανομής (*distribution-free*). Τέλος, η Παράγραφος 3.4 περιλαμβάνει μια ξεχωριστή οικογένεια μη παραμετρικών διαγραμμάτων ελέγχου που βασίζονται, για τον καθορισμό των ορίων ελέγχου τους, στην εμπειρική κατανομή παρελθοντικών δεδομένων, γνωστή ως κατανομή αναφοράς (*reference distribution*).

### 3.2. Παραμετρικά διαγράμματα ελέγχου

Τα παραμετρικά διαγράμματα ελέγχου είναι τεχνικές που βασίζονται σε μια γνωστή κατανομή, δηλαδή η διεργασία θεωρείται πως ακολουθεί μια συγκεκριμένη κατανομή (συνήθως την κανονική). Αυτό σημαίνει ότι οι στατιστικές ιδιότητες των διαγραμμάτων αυτών είναι ακριβείς, μόνο αν ικανοποιείται η παραπάνω υπόθεση. Το πιο γνωστό παραμετρικό διάγραμμα ελέγχου είναι το διάγραμμα τύπου Shewhart. Αν  $W$  είναι η στατιστική συνάρτηση που εκτιμά την κρίσιμη ποσότητα του υπό μελέτη χαρακτηριστικού, τότε ένα τυπικό διάγραμμα *Shewhart* περιλαμβάνει μία κεντρική γραμμή (*center line, CL*) που αντιπροσωπεύει (συνήθως) τη μέση τιμή της  $W$  (όταν η διεργασία είναι εντός ελέγχου), και δύο ακόμη οριζόντιες γραμμές που ονομάζονται άνω και κάτω όριο ελέγχου (*upper and lower control limit, UCL and LCL*). Για την κατασκευή των παραπάνω διαγραμμάτων, χρησιμοποιείται συνήθως το μοντέλο ορίων  $L$  σίγμα που τοποθετεί τα όρια ελέγχου σε απόσταση  $L$  μονάδων τυπικής απόκλισης  $\sigma_w$  από την κεντρική γραμμή  $\mu_w$ , όπως φαίνεται παρακάτω

$$UCL = \mu_w + L\sigma_w$$

$$CL = \mu_w$$

$$LCL = \mu_w - L\sigma_w.$$

Αξίζει να σημειωθεί ότι δημοφιλέστερη επιλογή για την απόσταση  $L\sigma$  των ορίων από την κεντρική γραμμή, είναι τρεις μονάδες τυπικής απόκλισης, και το μοντέλο αυτό είναι γνωστό ως μοντέλο τριών σίγμα (*three sigma control chart*). Αν ενδιαφερόμαστε για τη μέση τιμή  $\mu$  του υπό μελέτη χαρακτηριστικού, τότε χρησιμοποιούμε το δειγματικό μέσο  $\bar{X}$  για εκτίμησή της και το διάγραμμα ελέγχου παίρνει τη μορφή

$$\begin{aligned} UCL &= \mu_{\bar{X}} + 3\sigma_{\bar{X}} \\ CL &= \mu_{\bar{X}} \\ LCL &= \mu_{\bar{X}} - 3\sigma_{\bar{X}}. \end{aligned} \tag{3.1}$$

Για το μοντέλο των τριών σίγμα που περιγράφεται στη (3.1), γνωρίζουμε ότι αν η κατανομή της στατιστικής συνάρτησης  $\bar{X}$  είναι κανονική και η διεργασία είναι εντός ελέγχου, τότε η πιθανότητα να πάρει η  $\bar{X}$  τιμή εκτός των ορίων ελέγχου ή ισοδύναμα το διάγραμμα ελέγχου να δώσει ένδειξη ότι η διεργασία είναι εκτός ελέγχου είναι ίση με 0.0027. Το γεγονός αυτό σημαίνει ότι το  $\bar{X}$ -διάγραμμα ( $\bar{X}$ -chart) παράγει λανθασμένο συναγερμό με πιθανότητα 0.27%. Στη συγκεκριμένη περίπτωση, η εκτός ελέγχου κατάσταση σημαίνει ότι ο μέσος της διεργασίας έχει μετατοπισθεί προς τα πάνω ή προς τα κάτω και αυτό μεταφράζεται σε αύξηση των παραγόμενων προϊόντων που δεν ικανοποιούν τις απαιτήσεις. Γενικότερα η ικανότητα ανίχνευσης τέτοιων μετατοπίσεων μειώνεται όταν λαμβάνονται από την παραγωγή μικρά μεγέθη δειγμάτων (βλ. Champ & Jones (2004) και Chakraborti (2006)).

Όταν η μέση τιμή και η διακύμανση της διεργασίας είναι γνωστές, τότε το μήκος ροής  $N$  ακολουθεί Γεωμετρική κατανομή με παράμετρο  $p$ , όπου  $p$  είναι η πιθανότητα να βρεθεί ένα σημείο του διαγράμματος εκτός των ορίων ελέγχου. Συνεπώς το μέσο μήκος ροής ενός  $\bar{X}$ -διαγράμματος είναι ίσο με

$$ARL = \frac{1}{p}.$$

Για μια διεργασία που βρίσκεται εντός ελέγχου και στην περίπτωση που χρησιμοποιούμε το μοντέλο των τριών σίγμα για την κατασκευή του  $\bar{X}$ -διαγράμματος, το εντός ελέγχου μέσο μήκος ροής, δεδομένου ότι η κατανομή του δειγματικού μέσου είναι κανονική, είναι

$$ARL_{in} = \frac{1}{0.0027} = 370.$$

Αξίζει να σημειωθεί ότι αν η μέση τιμή ή η διακύμανση της διεργασίας που μελετάμε, δεν είναι γνωστή, τότε πρέπει να την (ή τις) εκτιμήσουμε με τη βοήθεια των δεδομένων που διαθέτουμε. Κατά συνέπεια τα όρια ελέγχου του διαγράμματος που θα κατασκευασθεί θα στηρίζονται στις παραπάνω εκτιμήσεις. Για περισσότερες λεπτομέρειες σχετικά με τις μεθόδους εκτίμησης των ορίων ελέγχου ενός  $\bar{X}$ -διαγράμματος, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Nedumaran & Pignatiello (2001) και Tsai *et al.* (2005). Στις περιπτώσεις αυτές, η κατανομή του μήκους ροής δεν είναι η γεωμετρική, αφού τα ενδεχόμενα να δοθεί σήμα κινδύνου στο  $i$ -οστό δείγμα, όπου  $i = 1, 2, \dots$ , δεν είναι πλέον ανεξάρτητα (βλ. Chakraborti (2000)). Οι Ghosh *et al.* (1981) και Del Castillo (1996) μελέτησαν την περίπτωση που η διακύμανση είναι άγνωστη, ενώ οι Burroughs *et al.* (1993) παρουσίασαν αποτελέσματα σχετικά με τα διαγράμματα τύπου *Shewhart* σε περιπτώσεις που τόσο η μέση τιμή, όσο και η διακύμανση είναι άγνωστες.

Στα διαγράμματα τύπου *Shewhart*, η λήψη της απόφασης για το αν η διεργασία βρίσκεται εντός ή εκτός στατιστικού ελέγχου, στηρίζεται αποκλειστικά σε ένα μόνο δείγμα, αγνοώντας τις πληροφορίες που μπορούμε να πάρουμε από τα προηγούμενα δείγματα που εξέρχονται από τη διεργασία. Για το λόγο αυτό έχουν αναπτυχθεί διαγράμματα ελέγχου που αξιοποιούν περισσότερα από ένα δείγματα και ονομάζονται διαγράμματα ελέγχου με μνήμη (*control charts with memory*). Για παράδειγμα, το διάγραμμα ελέγχου κινούμενου μέσου (*moving average control chart*) χρησιμοποιεί τους μέσους των  $k$  πιο πρόσφατων δειγμάτων. Πιο συγκεκριμένα, η στατιστική συνάρτηση που απεικονίζεται στο διάγραμμα αυτό δίνεται από τον ακόλουθο τύπο

$$M_t = \frac{\bar{X}_t + \bar{X}_{t-1} + \dots + \bar{X}_{t-k+1}}{k}, \quad t \geq k,$$

όπου  $\bar{X}_i$  είναι ο μέσος των παρατηρήσεων του  $i$ -οστού δείγματος που λαμβάνεται από την παραγωγική διαδικασία. Για την παρουσίαση των ιδιοτήτων του διαγράμματος κινούμενου μέσου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο κείμενο των Derman & Ross (1997, σελ. 129–132). Πρόσθετα, αν τα παρελθοντικά δείγματα σταθμιστούν με εκθετικά βάρη, τότε προκύπτει το διάγραμμα ελέγχου

εκθετικά σταθμισμένου κινούμενου μέσου (*Exponentially Weighted Moving Average, EWMA*), στο οποίο απεικονίζεται η στατιστική συνάρτηση

$$W_t = \varepsilon \bar{X}_t + (1 - \varepsilon)W_{t-1}, \quad t \geq 1$$

όπου  $0 < \varepsilon < 1$ . Τα διαγράμματα *EWMA* προτάθηκαν αρχικά από τον Roberts (1959), ενώ οι ιδιότητες και γενικεύσεις τους μελετήθηκαν εκτενώς στη διεθνή βιβλιογραφία (βλ. Lucas & Saccucci (1990) ή Han & Tsung (2004)). Τα συγκεκριμένα διαγράμματα ελέγχου μπορούν να χρησιμοποιηθούν για την από κοινού παρακολούθηση της μέσης τιμής και της διακύμανσης μιας διεργασίας (βλ. Gan (1995)), ακόμη και σε περιπτώσεις που οι παράμετροι τους είναι άγνωστες και έχουν εκτιμηθεί (βλ. Jones *et al.* (2001)). Η απόδοση του παραπάνω διαγράμματος σε περιπτώσεις που η κανονικότητα του υπό μελέτη χαρακτηριστικού παραβιάζεται έχει εξετασθεί στην εργασία των Maravelakis *et al.* (2005). Πρόσθετα, για την ανίχνευση μετατόπισης του μέσου της διεργασίας, έχουν προταθεί στην εργασία του Page (1954), τα διαγράμματα συσσωρευμένων αθροισμάτων (*Cumulative Sum Control charts, CUSUM*), στα οποία απεικονίζεται η στατιστική συνάρτηση

$$Y_t = (\bar{X}_t - Y_0) + Y_{t-1}, \quad t \geq 1,$$

όπου  $Y_0$  είναι η τιμή στόχος. Η κατανομή του μήκους ροής ενός *CUSUM* διαγράμματος έχει μελετηθεί στις εργασίες των Brook & Evans (1972), Jones *et al.* (2004) και Luceno & Puig-Pey (2000), ενώ για το σχεδιασμό του, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Zantek (2006) και Rogerson (2006) (βλ. επίσης Αντζουλάκος (2006)).

### 3.3. Μη παραμετρικά διαγράμματα ελέγχου

Τα μη παραμετρικά ή ελεύθερα κατανομής διαγράμματα ελέγχου, όπως έχει ήδη αναφερθεί, έχουν ιδιότητες που δεν εξαρτώνται από την κανονικότητα ή κάποια άλλη συγκεκριμένη παραμετρική κατανομή. Αξίζει να σημειωθεί ότι ο όρος «μη παραμετρικά» δεν δηλώνει ότι στο σχεδιασμό των διαγραμμάτων αυτών δεν περιλαμβάνονται παράμετροι και κατά συνέπεια ο όρος «ελεύθερα κατανομής» φαίνεται να προσδιορίζει καλύτερα τη φύση των διαγραμμάτων αυτών. Για την ανάπτυξη και θεωρητική μελέτη των μη παραμετρικών διαγραμμάτων ελέγχου χρησιμοποιούνται κατά βάση οι κλασικές μη παραμετρικές τεχνικές συμπερασματολογίας που αντιμετωπίζουν αποτελεσματικά το πρόβλημα ελέγχου



υποθέσεων όταν η κατανομή δεν έχει καθορισθεί. Για περισσότερες λεπτομέρειες, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο κείμενο των Gibbons & Chakraborti (2003) ή τις εργασίες των Balakrishnan *et al.* (2006) και Aki (1987).

Το σημαντικότερο πλεονέκτημα των μη παραμετρικών διαγραμμάτων ελέγχου είναι ότι η εφαρμογή τους δεν απαιτεί την υπόθεση κάποιας συγκεκριμένης κατανομής για την υπό μελέτη διεργασία. Για το λόγο αυτό, το εντός ελέγχου μέσο μήκος ροής είναι το ίδιο για όλες τις συνεχείς κατανομές. Επιπλέον, η διακύμανση της διεργασίας δεν χρειάζεται να είναι γνωστή ή να εκτιμηθεί, ενώ επιπλέον τα διαγράμματα αυτά παρουσιάζουν αντοχή στην παρουσία έκτροπων παρατηρήσεων και σε περιπτώσεις που ο υπό μελέτη πληθυσμός ακολουθεί λοξή ή με βαριές ουρές κατανομή. Για μια λεπτομερή ανασκόπηση των μη παραμετρικών διαγραμμάτων ελέγχου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Chakraborti *et al.* (2001) και Chakraborti (2007).

Ας υποθέσουμε ότι παρακολουθούμε μια παραγωγική διαδικασία και έστω  $X_{i1}, X_{i2}, \dots, X_{in}$ ,  $i = 1, 2, \dots$ , δείγματα μεγέθους  $n$ , που λαμβάνουμε από τη διεργασία ( $X_{ij}$  είναι η  $j$ -οστή παρατήρηση του  $i$ -οστού δείγματος που λαμβάνεται από την παραγωγή) με σκοπό να ανιχνεύσουμε αν παραμένει εντός ελέγχου ή έχει μετατοπισθεί σε εκτός ελέγχου κατάσταση. Οι Amin *et al.* (1995) προκειμένου να κατασκευάσουν ένα μη παραμετρικό διάγραμμα ελέγχου τύπου *Shewhart* χρησιμοποίησαν τη προσημική στατιστική συνάρτηση (*sign test statistic*)

$$SN_i = \sum_{j=1}^n \text{sign}(X_{ij} - \theta_0),$$

όπου

$$\text{sign}(x) = \begin{cases} 1, & \alpha\nu x > 0 \\ 0, & \alpha\nu x = 0 \\ -1, & \alpha\nu x < 0, \end{cases}$$

και  $\theta_0$  είναι η εντός ελέγχου τιμή της διαμέσου της διεργασίας. Αν  $T_i$  είναι η συνήθης προσημική συνάρτηση (*usual sign statistic*) που μετρά το πλήθος των παρατηρήσεων των δειγμάτων που είναι μεγαλύτερες ή ίσες από τη διάμεσο  $\theta_0$ , τότε είναι γνωστό ότι

$$SN_i = 2T_i - n.$$

Δεδομένου ότι κατανομή της  $T_i$  είναι η διωνυμική με παράμετρος  $n$  και  $p = 0.5$ , έπεται ότι η εντός ελέγχου κατανομή της  $SN_i$  είναι συμμετρική γύρω από το μηδέν και συνεπώς τα όρια ελέγχου του προσημικού διαγράμματος (*sign control chart*) που προκύπτει θα είναι  $LCL = -c$  και  $UCL = c$  αντίστοιχα, όπου  $c$  είναι ένας θετικός ακέραιος αριθμός. Αξίζει να σημειωθεί ότι οι Amin *et al.* (1995) εφάρμοσαν για το παραπάνω διάγραμμα ελέγχου κανόνες ροών (*run rules*), παρουσιάζοντας αριθμητικά αποτελέσματα για το εντός ελέγχου μέσο μήκος ροής, ενώ χρησιμοποίησαν την προσημική συνάρτηση  $SN_i$  για την κατασκευή ενός *CUSUM* διαγράμματος (*sign CUSUM chart*) για την παρακολούθηση της διαμέσου μιας διεργασίας, αποδεικνύοντας μάλιστα ότι είναι αποτελεσματικότερο από το αντίστοιχο προσημικό διάγραμμα τύπου *Shewhart*. Για περισσότερες λεπτομέρειες σχετικά με τη χρήση της προσημικής συνάρτησης στην παρακολούθηση μιας διεργασίας, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην εργασία του Arnold (1985).

Ιδιαίτερα δημοφιλή είναι τα μη παραμετρικά διαγράμματα ελέγχου που χρησιμοποιούν προσημικές-βαθμολογικές συναρτήσεις (*signed-rank statistics*). Στην εργασία του Reynolds (1975), μελετώνται διαγράμματα που βασίζονται σε προσημικούς διαδοχικούς βαθμούς (*signed sequential ranks*) των παρατηρήσεων, ο Miller (1970) εισήγαγε έναν έλεγχο που χρησιμοποιεί την προσημική-βαθμολογική στατιστική συνάρτηση του *Wilcoxon*, ενώ οι McGilchrist & Woodyer (1975) πρότειναν ένα διάγραμμα *CUSUM* για τη μελέτη της διαμέσου της κατανομής βροχοπτώσεων που βασίζεται σε προσημικούς βαθμούς. Οι Bakir & Reynolds (1979) πρότειναν ένα διάγραμμα *CUSUM* για την παρακολούθηση του μέσου μιας διεργασίας, που βασίζεται στην ιδέα της βαθμολόγησης (*ranking*) των παρατηρήσεων εντός του δείγματος, στο οποίο ανήκουν. Η συγκεκριμένη ιδέα είχε αρχικά εφαρμοσθεί από τους *Wilcoxon et al.* (1963) και Van der Laan (1966) για την κατασκευή ενός ελέγχου για δύο διαδοχικά δείγματα.

Αν  $R_{ij}$  είναι ο βαθμός της απόλυτης τιμής  $|X_{ij}|$  της κάθε παρατήρησης μεταξύ όλων των απόλυτων τιμών των παρατηρήσεων του  $i$ -οστού δείγματος, για  $j = 1, 2, \dots, n$  και  $i = 1, 2, \dots$ , τότε η *Wilcoxon* προσημική-βαθμολογική στατιστική συνάρτηση που απεικονίζεται στο διάγραμμα των Bakir & Reynolds (1979) δίνεται στον ακόλουθο τύπο

$$SR_i = \sum_{j=1}^n \text{sign}(X_{ij})R_{ij} .$$

Η διαδικασία που προτείνεται, χρησιμοποιεί το συσσωρευμένο άθροισμα των  $SR_1, SR_2, \dots$ , τα οποία είναι μια ακολουθία ανεξάρτητων προσημικών–βαθμολογικών συναρτήσεων τύπου *Wilcoxon* και εκφράζουν τις διαφορές μεταξύ των αθροισμάτων των βαθμών των  $|X_{ij}|$  που αντιστοιχούν σε θετικές και αρνητικές τιμές αντίστοιχα. Το συγκεκριμένο διάγραμμα δίνει σήμα ότι η διεργασία είναι εκτός ελέγχου στο  $i$  – οστό δείγμα, αν ικανοποιείται μία εκ των δύο παρακάτω ανισώσεων

$$\sum_{i=1}^n (SR_i - k) - \min_{0 \leq m \leq n} \sum_{i=1}^m (SR_i - k) \geq h ,$$

ή

$$\max_{0 \leq m \leq n} \sum_{i=1}^n (SR_i + k) - \sum_{i=1}^m (SR_i + k) \geq h ,$$

όπου οι ποσότητες  $k, h$  είναι παράμετροι σχεδιασμού του διαγράμματος.

Για την παρακολούθηση της διαμέσου μιας διεργασίας, ας θεωρήσουμε ότι λαμβάνουμε διαδοχικά και ανεξάρτητα τυχαία δείγματα  $X_{i1}, X_{i2}, \dots, X_{in}$  και έστω ότι  $|X_{i1} - \theta_0|, |X_{i2} - \theta_0|, \dots, |X_{in} - \theta_0|$  είναι οι απόλυτες τιμές των αποκλίσεων των παρατηρήσεων από τη διάμεσο του  $i$  – οστού δείγματος και  $R_{ij}$  ο βαθμός της ποσότητας  $|X_{ij} - \theta_0|$ , για  $j = 1, 2, \dots, n$ , μεταξύ των  $n$  απόλυτων αποκλίσεων (από τη διάμεσο), όπου  $\theta_0$  είναι η εντός ελέγχου τιμή της διαμέσου. Ο Bakir (2004) χρησιμοποίησε την ακόλουθη μορφή προσημικής–βαθμολογικής συνάρτησης *Wilcoxon*

$$SR'_i = \sum_{j=1}^n \text{sign}(X_{ij} - \theta_0)R_{ij} , \quad i = 1, 2, \dots ,$$

για την κατασκευή ενός μη παραμετρικού διαγράμματος ελέγχου τύπου *Shewhart*, και υπολόγισε το εντός ελέγχου μέσο μήκος ροής ( $ARL_{in}$ ) και την πιθανότητα λανθασμένου συναγερμού ( $FAR$ ) με τη βοήθεια της κατανομής της συνάρτησης *Wilcoxon*  $W^+$  που εκφράζει το συνολικό άθροισμα των βαθμών μόνο για τις θετικές αποκλίσεις ( $X_{ij} - \theta_0$ ). Πιο συγκεκριμένα, αποδεικνύεται ότι ισχύει η σχέση

$$SR' = 2W^+ - \frac{n(n+1)}{2} ,$$

οπότε προκύπτει ότι η συνάρτηση  $SR'$  είναι συμμετρική γύρω από το μηδέν. Συνεπώς η εντός ελέγχου κατανομή της  $SR'$  υπολογίζεται από την κατανομή της  $W^+$  υπό τη μηδενική υπόθεση (*null distribution*) της που έχει μελετηθεί εκτενώς στη διεθνή βιβλιογραφία (βλ. Wilcoxon *et al.* (1972)).

Οι Chakraborti & Eryilmaz (2007) πρότειναν απλές παραλλαγές του διαγράμματος του Bakir (2004), εφαρμόζοντας σε αυτό κανόνες ροών τύπου 2-από-2. Πιο συγκεκριμένα, κατασκεύασαν τα διαγράμματα 2-από-2 KL και 2-από-2 DR, χρησιμοποιώντας τη στατιστική συνάρτηση  $SR'$ . Το πρώτο διάγραμμα δίνει σήμα ότι η διεργασία είναι εκτός ελέγχου αν οι τιμές της  $SR'$  των δύο πιο πρόσφατων δειγμάτων, βρίσκονται πάνω ή κάτω από τα όρια ελέγχου, ενώ στο δεύτερο διάγραμμα το σήμα παράγεται ακόμη και αν η μια τιμή βρίσκεται πάνω και η δεύτερη κάτω από τα όρια ελέγχου. Για μια γενική επισκόπηση κανόνων ροών που εφαρμόζονται σε διαγράμματα ελέγχου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Koutras *et al.* (2007), Bersimis *et al.* (2007) και Ρακιτζή (2008).

Οι Amin & Searcy (1991) πρότειναν ένα μη παραμετρικό διάγραμμα ελέγχου τύπου *EWMA* που βασίζεται στη βαθμολόγηση των παρατηρήσεων εντός του δείγματος. Συγκεκριμένα, η στατιστική συνάρτηση που απεικονίζεται στο διάγραμμα δίνεται στην ακόλουθη σχέση

$$Z_i = \lambda SR_i + (1 - \lambda)Z_{i-1}, \quad 0 < \lambda < 1,$$

όπου η αρχική τιμή  $Z_0$  είναι η τιμή στόχος που έχει καθορισθεί.

Στην εργασία του McDonald (1990), προτείνεται ένα διάγραμμα ελέγχου *CUSUM* που βασίζεται στους διαδοχικούς βαθμούς (*sequential ranks*) των παρατηρήσεων, οι οποίοι ορίστηκαν αρχικά στην εργασία του Parent (1965) μέσω του τύπου

$$R_i = 1 + \sum_{j=1}^{i-1} I(X_j < X_i),$$

όπου  $I(\cdot)$  είναι η συνήθης δείτρια συνάρτηση. Το διάγραμμα στηρίζεται στις συναρτήσεις  $U_i = R_i / (i + 1)$ ,  $i = 1, 2, \dots$ , οι οποίες είναι ανεξάρτητες τυχαίες μεταβλητές και ακολουθούν ομοιόμορφη κατανομή (σε εντός ελέγχου διεργασία) στο

$\left(\frac{1}{i+1}, \frac{2}{i+1}, \dots, \frac{i}{i+1}\right)$ . Το σήμα για εκτός ελέγχου κατάσταση δίνεται όταν

ικανοποιείται η ανίσωση  $T_i \geq h$ , όπου

$$T_i = \max(T_{i-1} + U_i - k, 0)$$

και  $h, k$  είναι οι παράμετροι σχεδιασμού του διαγράμματος.

Ολοκληρώνοντας τη σύντομη ανασκόπηση, αξίζει να αναφερθεί ότι οι Bhattacharya & Frierson (1981) πρότειναν ένα μη παραμετρικό διάγραμμα ελέγχου τύπου *CUSUM*, χρησιμοποιώντας σταθμισμένα αθροίσματα των διαδοχικών βαθμών  $R_i$  των παρατηρήσεων  $X_1, X_2, \dots, X_n$  που λαμβάνονται από την παραγωγική διαδικασία. Πιο συγκεκριμένα, αν εισάγουμε τις ποσότητες  $Z_i$

$$Z_i = \frac{R_i - (i+1)/2}{i},$$

τότε στο παραπάνω διάγραμμα απεικονίζεται μια κανονικοποιημένη μορφή των ποσοτήτων  $S_k = \sum_{i=1}^k Z_i$ ,  $1 \leq k \leq n$ , που δίνεται ως ακολούθως

$$S_n(t) = \sqrt{\frac{12}{n}} \{S_{[nt]} + (nt - [nt])Z_{[nt]+1}\}, \quad 0 \leq t \leq 1.$$

### 3.4. Μη παραμετρικά διαγράμματα ελέγχου με χρήση μιας εμπειρικής κατανομής αναφοράς

Μια ενδιαφέρουσα ιδέα στην κατασκευή και το σχεδιασμό ενός διαγράμματος ελέγχου είναι να υποθέσουμε ότι είναι διαθέσιμο ένα αρχικό δείγμα αναφοράς (*reference sample*) ή μια κατανομή αναφοράς (*reference distribution*), τα οποία αφορούν τη διεργασία που μας ενδιαφέρει όταν αυτή βρίσκεται εντός ελέγχου. Οι Park & Reynolds (1987) πρότειναν διαγράμματα ελέγχου τύπου *Shewhart* και *CUSUM* για την παρακολούθηση της παραμέτρου θέσης μιας παραγωγικής διαδικασίας, για περιπτώσεις που η εντός ελέγχου τιμή της παραμέτρου δεν είναι γνωστή. Τα συγκεκριμένα διαγράμματα στηρίζονται στις συναρτήσεις γραμμικής τοποθέτησης (*linear placement statistics*) που έχουν εισαχθεί στην εργασία των Orban & Wolfe (1982) για τη σύγκριση δειγμάτων που λαμβάνονται από τη διεργασία με ένα σταθερό δείγμα που έχει συλλεχθεί όταν η διεργασία βρισκόταν εντός ελέγχου. Οι Hackl & Ledolter (1991) χρησιμοποίησαν την έννοια των κανονικοποιημένων βαθμών (*standardized ranks*)  $R_i$  των παρατηρήσεων  $X_i$  που ορίζονται στην ακόλουθη σχέση

$$R_i = 2 \left( F_0(X_i) - \frac{1}{2} \right),$$

όπου  $F_0$  είναι η αθροιστική συνάρτηση κατανομής για την εντός ελέγχου διεργασία. Ωστόσο, αν η  $F_0$  δεν είναι γνωστή, τότε ο κανονικοποιημένος βαθμός παίρνει τη μορφή

$$R'_i = \frac{2(R_i^* - (n+1)/2)}{n},$$

όπου ένα δείγμα αναφοράς μεγέθους  $(n-1)$  είναι διαθέσιμο όταν η διεργασία είναι εντός ελέγχου. Επιπλέον, συμβολίζουμε με  $R_i^*$  το βαθμό της  $i$ -οστής παρατήρησης ως προς το δείγμα αναφοράς, έτσι ώστε

$$R_i^* = 1 + \sum_{j=1}^{n-1} I(X_i > X_j).$$

Η στατιστική συνάρτηση που απεικονίζεται στο προτεινόμενο διάγραμμα τύπου *EWMA* δίνεται από τον τύπο

$$T_i = (1 - \lambda)T_{i-1} + \lambda R_i,$$

όπου  $T_0$  είναι συνήθως το μηδέν και  $\lambda \in (0,1]$ .

Ο Yaschin (1992) μελέτησε την κατανομή του μήκους ροής ενός μη παραμετρικού διαγράμματος ελέγχου *CUSUM*, σε περιπτώσεις που η κατανομή της διεργασίας είναι άγνωστη. Συγκεκριμένα, για τη σημειακή εκτίμηση διαφόρων χαρακτηριστικών του μήκους ροής, πρότεινε την αντικατάσταση της πραγματικής κατανομής των δεδομένων, με μια εμπειρική κατανομή ενός δείγματος αναφοράς. Πρόσθετα, ο Bakir (2006) κατασκεύασε διαγράμματα ελέγχου για την παρακολούθηση του κέντρου μιας διεργασίας, βασιζόμενος σε προσημικές-βαθμολογικές συναρτήσεις, οι οποίες αντί να βαθμολογούν τις παρατηρήσεις του δείγματος που εξέρχεται από τη διεργασία, βαθμολογούν συναρτήσεις αυτών των παρατηρήσεων. Οι συγκεκριμένες προσημικές-βαθμολογικές συναρτήσεις ονομάζονται *signed rank-like statistics* και οφείλουν την ονομασία τους στον Moses (1963). Αν  $\delta$  είναι η διάμεσος του δείγματος αναφοράς που λαμβάνεται από την διεργασία όταν βρίσκεται εντός ελέγχου και  $R_{ij}$  είναι ο βαθμός της απόλυτης τιμής της απόκλισης  $|X_{ij} - \delta|$  ανάμεσα στις αποκλίσεις  $|X_{i1} - \delta|, |X_{i2} - \delta|, \dots, |X_{in} - \delta|$ , τότε το προτεινόμενο διάγραμμα του Bakir (2006) απεικονίζει την ακόλουθη στατιστική συνάρτηση

$$\varphi_i^* = \sum_{j=1}^n \text{sign}(X_{ij} - \delta) R_{ij}^*.$$

Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα μη παραμετρικά διαγράμματα ελέγχου που βασίζονται σε ένα δείγμα αναφοράς και χρησιμοποιούν διατεταγμένες παρατηρήσεις του δείγματος αυτού στον προσδιορισμό των ορίων ελέγχου αλλά και στον κανόνα απόφασης για το αν η διεργασία κρίνεται ότι βρίσκεται εντός ή εκτός στατιστικού ελέγχου. Οι Willemain & Runger (1996) πρότειναν το σχεδιασμό διαγραμμάτων ελέγχου με τη βοήθεια μιας εμπειρικής κατανομής αναφοράς (*empirical reference distribution*), η οποία αναφέρεται στα δεδομένα της διεργασίας όταν αυτή λειτουργεί κάτω από κανονικές συνθήκες, δηλαδή με την παρουσία μόνο της φυσικής μεταβλητότητας. Ας θεωρήσουμε ότι  $X_1, X_2, \dots, X_m$  είναι ένα δείγμα αναφοράς μεγέθους  $m$  που λαμβάνεται από την εντός ελέγχου διεργασία και έστω  $F$  η αθροιστική συνάρτηση κατανομής του. Τα όρια ελέγχου του προτεινόμενου διαγράμματος είναι δύο διατεταγμένες παρατηρήσεις  $X_{(k)}, X_{(b+k)}$  του παραπάνω δείγματος, όπου  $0 \leq k \leq m$  και  $1 \leq b \leq m - k + 1$  (σύμβαση:  $X_{(0)} = -\infty$  και  $X_{(m+1)} = +\infty$ ). Για την παρακολούθηση της παραγωγικής διαδικασίας, λαμβάνουμε μεμονωμένες παρατηρήσεις  $Y$  από αυτήν και τις συγκρίνουμε με τα όρια ελέγχου. Η πιθανότητα να βρεθεί μια παρατήρηση  $Y$  μεταξύ των ορίων ελέγχου δίνεται από την ακόλουθη σχέση

$$p = P(X_{(k)} \leq Y \leq X_{(b+k)}) = F(X_{(b+k)}) - F(X_{(k)}).$$

Δεδομένου ότι οι διατεταγμένες παρατηρήσεις είναι τυχαίες μεταβλητές, είναι προφανές ότι και η πιθανότητα  $p$  θα είναι μια τυχαία μεταβλητή. Πιο συγκεκριμένα η  $p$  θα ακολουθεί Βήτα κατανομή, με συνάρτηση πυκνότητας

$$g(p) = \frac{m!}{(b-1)!(m-b)!} p^{b-1} (1-p)^{m-b},$$

ενώ η αναμενόμενη τιμή και η διακύμανσή της, δίνονται αντίστοιχα ως ακολούθως

$$E(p) = \frac{b}{m+1}$$

και

$$Var(p) = \frac{b(m-b+1)}{(m+1)^2(m+2)}.$$

Για το προτεινόμενο διάγραμμα των Willemain & Runger (1996), δεδομένων των ορίων ελέγχου που έχουν επιλεγεί, το μήκος ροής  $N$  ακολουθεί γεωμετρική κατανομή με παράμετρο  $(1-p)$ . Συνεπώς, το αναμενόμενο μέσο μήκος ροής όταν η

διεργασία είναι εντός ελέγχου, δεδομένου ότι γνωρίζουμε την πιθανότητα  $p$ , δίνεται από την ακόλουθη σχέση

$$ARL_{in} = \frac{1}{1-p}.$$

Αξίζει να σημειωθεί ότι η ποσότητα  $ARL_{in}$  είναι τυχαία μεταβλητή με συνάρτηση πυκνότητας

$$h(ARL_{in}) = \frac{m!}{(b-1)!(m-b)!} \cdot \frac{(ARL_{in}-1)^{b-1}}{(ARL_{in})^{m+1}}, \quad ARL_{in} \geq 1$$

και μέση τιμή

$$E(ARL_{in}) = \frac{m}{m-b}. \quad (3.2)$$

Είναι προφανές από την τελευταία σχέση ότι, αν χρησιμοποιούμε δείγμα αναφοράς μεγέθους  $m$ , τότε η μέγιστη τιμή του αναμενόμενου εντός ελέγχου μέσου μήκους ροής του παραπάνω διαγράμματος με συμμετρικά όρια ελέγχου είναι ίση με  $m/4$ . Για το λόγο αυτό, προκειμένου να πετύχουμε στην πράξη την επιθυμητή τιμή για το  $ARL_{in}$  σε ένα συμμετρικό διάγραμμα ελέγχου, χρειαζόμαστε τουλάχιστον  $4E(ARL_{in})$  παρατηρήσεις (για την εύρεση του κατάλληλου (και ελάχιστου δυνατού) μεγέθους δείγματος για την ανάπτυξη ασφαλών στατιστικών συμπερασμάτων, βλ. επίσης τη μονογραφία του Κυριακούση (2008)). Με τη βοήθεια της (3.2), παρουσιάζονται στον Πίνακα 3.1 διάφοροι συνδυασμοί των παραμέτρων  $m, k, b$  του προτεινόμενου διαγράμματος που παράγουν  $ARL_{in} \cong 370.4$ .

Αξίζει να σημειωθεί ότι οι Ion & Klaassen (2005) πρότειναν ένα μη παραμετρικό διάγραμμα ελέγχου τύπου *Shewhart*, το οποίο αποτελεί ειδική περίπτωση του διαγράμματος των Willemain & Runger (1996) με την έννοια ότι κατασκευάζεται με τυχαία επιλογή μεταξύ των σχεδιασμών ( $k=0, b=n-n_a$ ) και ( $k=0, b=n-n_a+1$ ), όπου  $n_a = \lceil (n+1)a \rceil$  είναι το ακέραιο μέρος του αριθμού  $(n+1)a$ .

Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα μη παραμετρικά διαγράμματα ελέγχου που βασίζονται στις ελεύθερες κατανομής ποσότητες  $P_{(r)}$  που ονομάζονται στατιστικές συναρτήσεως προτεραιότητας (*precedence statistics*). Τα συγκεκριμένα διαγράμματα στηρίζονται σε ένα εντός ελέγχου δείγμα αναφοράς (*reference sample*) για την εκτίμηση των ορίων ελέγχου, ενώ από την παραγωγική διαδικασία εξετάζονται διαδοχικά και ανεξάρτητα τυχαία δείγματα (*test samples*), με τη βοήθεια



των οποίων λαμβάνεται η απόφαση για το αν η διεργασία παραμένει εντός ελέγχου ή όχι.

**Πίνακας 3.1.** Υπολογισμός του  $ARL_{in}$  για διάφορες τιμές των παραμέτρων σχεδιασμού του διαγράμματος των Willemain & Runger (1996)

$m$	$k$	$b$	$ARL_{in}$
741	3	736	370.5
1112	4	1105	370.7
1478	5	1473	370.5
3334	10	3315	370.4
5186	15	5157	370.4
7038	20	6999	370.4
8890	25	8841	370.4
10742	30	10683	370.4
14446	40	13667	370.4
18150	50	18051	370.4
21854	60	21735	370.4
29262	80	29103	370.4
36670	100	36471	370.4

Αξίζει να σημειωθεί ότι η λήψη τυχαίων δειγμάτων μεγέθους  $n$  αποτελεί μια βασική διαφορά των διαγραμμάτων αυτών σε σχέση με το διάγραμμα των Willemain & Runger (1996), στο οποίο ο κανόνας απόφασης στηρίζεται σε μεμονωμένες παρατηρήσεις.

Μία από τις πρώτες προσπάθειες διερεύνησης των ιδιοτήτων και της χρήσης των στατιστικών ελέγχων προτεραιότητας (*precedence tests*), ως μία γρήγορη και αποτελεσματική διαδικασία ελέγχου χρόνων ζωής, οφείλεται στον Epstein (1955). Συγκεκριμένα, ο Epstein εξέτασε την περίπτωση δύο πληθυσμών που προέρχονται από την κανονική κατανομή, και μελέτησε εμπειρικά έναν έλεγχο που βασίζεται στο πλήθος των παρατηρήσεων του ενός πληθυσμού που ξεπερνούν την τιμή της  $r$ -οστής διατεταγμένης παρατήρησης του δεύτερου. Ακολούθως, ο Nelson (1963) χρησιμοποίησε τη στατιστική συνάρτηση προτεραιότητας για την κατασκευή ενός ελεύθερου κατανομής ελέγχου χρόνων ζωής που βασίζεται στη διάταξη πρόωρων

αποτυχιών και επιτρέπει τη γρήγορη λήψη αποφάσεων ακόμη και σε περιπτώσεις (δεξιά) περικομμένων παρατηρήσεων.

Ας υποθέσουμε ότι  $X_1, X_2, \dots, X_m$  και  $Y_1, Y_2, \dots, Y_n$  είναι δύο ανεξάρτητα τυχαία δείγματα μεγέθους  $m$  και  $n$  που προέρχονται από τις κατανομές  $F_X$  και  $F_Y$  αντίστοιχα. Στόχος μας είναι ο έλεγχος της μηδενικής υπόθεσης ότι οι δύο κατανομές είναι ισόνομες, έναντι της εναλλακτικής ότι μία από τις δύο κατανομές είναι στοχαστικά μεγαλύτερη από την άλλη. Πιο συγκεκριμένα, η διατύπωση των δύο υποθέσεων δίνεται ακολούθως

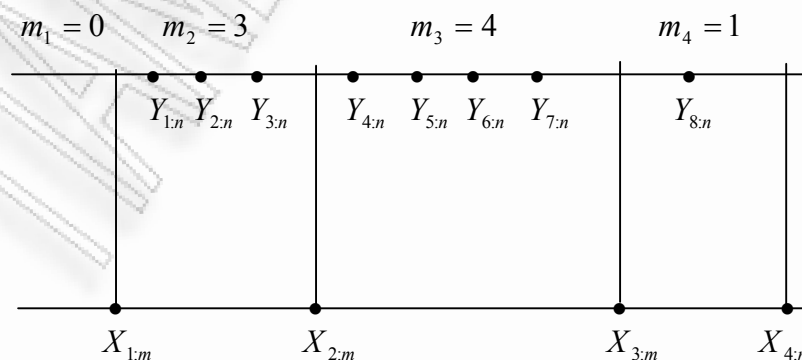
$$H_0 : F_X = F_Y \text{ έναντι } H_1 : F_X < F_Y. \quad (3.2)$$

Αν  $X_{1:m}, X_{2:m}, \dots, X_{m:m}$  και  $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$  είναι οι διατεταγμένες παρατηρήσεις των δύο παραπάνω τυχαίων δειγμάτων, τότε συμβολίζουμε με  $M_1$  το πλήθος των  $Y$ -παρατηρήσεων που είναι μικρότερες από την  $X_{1:m}$  και με  $M_i$  το πλήθος των  $Y$ -παρατηρήσεων που βρίσκονται μεταξύ των  $X_{i-1:m}$  και  $X_{i:m}$ ,  $i = 2, 3, \dots, m$ . Αξίζει να σημειωθεί ότι οι ποσότητες  $M_i$  σχετίζονται με τις γνωστές *exceedance* στατιστικές συναρτήσεις υπέρβασης (*exceedance statistics*), των οποίων οι ιδιότητες έχουν μελετηθεί στη διεθνή βιβλιογραφία (βλ. Fligner & Wolfe (1976) και Randles & Wolfe (1979)). Η στατιστική συνάρτηση προτεραιότητας, η οποία εκφράζει το πλήθος των  $Y$ -παρατηρήσεων που προηγούνται της  $r$ -οστής διατεταγμένης  $X$ -παρατήρησης, ορίζεται από τον ακόλουθο τύπο

$$P_{(r)} = \sum_{i=1}^r M_i. \quad (3.3)$$

Για παράδειγμα, όπως φαίνεται στο ακόλουθο διάγραμμα, για  $m = 4, n = 8, r = 4$

**Σχήμα 3.1.** Σχηματική παρουσίαση ενός στατιστικού ελέγχου προτεραιότητας



η στατιστική συνάρτηση  $P_{(r)}$  είναι ίση με

$$P_{(4)} = 0 + 3 + 4 + 1 = 8.$$

Είναι σαφές ότι μεγάλες τιμές της συνάρτησης  $P_{(r)}$  δίνουν ένδειξη για την απόρριψη της μηδενικής υπόθεσης  $H_0$  έναντι της εναλλακτικής  $H_1$ . Για καθορισμένο επίπεδο σημαντικότητας  $\alpha$ , η κρίσιμη περιοχή θα είναι της μορφής  $\{s, s+1, \dots, n\}$ , όπου

$$\alpha = P(P_{(r)} \geq s | H_0 : F_X = F_Y).$$

Υπό τη μηδενική υπόθεση  $H_0 : F_X = F_Y$ , τα δύο διαθέσιμα δείγματα  $X_1, X_2, \dots, X_m$  και  $Y_1, Y_2, \dots, Y_n$  είναι ισόνομα, οπότε το πλήθος των τρόπων επιλογής του  $Y$ -δείγματος μεγέθους  $n$  από ένα σύνολο  $(m+n)$  παρατηρήσεων, είναι ίσο με

$$\binom{m+n}{n}.$$

Παρατηρώντας ότι το συμβάν  $P_{(r)} = j$  εκφράζει το πλήθος των τρόπων

επιλογής  $j$  από τις πρώτες  $(j+r-1)$  διατεταγμένες παρατηρήσεις και στη συνέχεια  $(n-j)$  από τις υπόλοιπες  $(m+n-j-r)$  για τη δημιουργία του  $Y$ -δείγματος, μπορούμε να γράψουμε την ακόλουθη σχέση

$$P(P_{(r)} = j | H_0 : F_X = F_Y) = \frac{\binom{j+r-1}{j} \binom{m+n-j-r}{n-j}}{\binom{m+n}{m}}. \quad (3.4)$$

Εναλλακτικά, μπορούμε να καταλήξουμε στην κατανομή της στατιστικής συνάρτησης  $P_{(r)}$  (υπό τη μηδενική υπόθεση), αν, δεδομένης της τιμής  $X_{r:m} = x$ , γράψουμε την πιθανότητα ότι  $j$  παρατηρήσεις του δείγματος  $Y_1, Y_2, \dots, Y_n$  προηγούνται της  $X_{r:m}$  ως εξής

$$\begin{aligned} P(\text{πλήθος των } Y \text{- παρατηρήσεων πριν την } X_{r:m} = j | X_{r:m} = x) &= \\ &= \binom{n}{j} (F_Y(x))^j (1 - F_Y(x))^{n-j}. \end{aligned}$$

Στη συνέχεια, ολοκληρώνοντας ως προς  $x$  για να πάρουμε τη δέσμευση που τέθηκε για το  $X_{r:m}$ , λαμβάνουμε τα εξής

$$\begin{aligned} P(P_{(r)} = j) &= \\ &= \int_{-\infty}^{\infty} P(\text{πλήθος των } Y \text{- παρατηρήσεων πριν την } X_{r:m} = j | X_{r:m} = x) f_{r:m}(x) dx \end{aligned}$$

$$= \binom{n}{j} \int_{-\infty}^{\infty} (F_X(x))^j (1 - F_X(x))^{n-j} \times \frac{m!}{(r-1)!(m-r)!} (F_Y(x))^{r-1} (1 - F_Y(x))^{m-r} f_X(x) dx, \quad (3.5)$$

όπου η συνάρτηση  $f_{r:m}(x)$  ορίζεται στο πρώτο Κεφάλαιο.

Υπό τη μηδενική υπόθεση, μπορούμε να καταλήξουμε στην κατανομή της  $P_{(r)}$ , όπως αυτή δίνεται στην (3.4). Στον Πίνακα 3.2 παρουσιάζονται, για διαφορετικές τιμές των μεγεθών δειγμάτων  $m, n$  και  $r = 1, 2, 3, 4, 5$ , η κρίσιμη τιμή  $s$  και το ακριβές επίπεδο σημαντικότητας  $a$  που επιτυγχάνεται (το δυνατόν πλησιέστερα στην τιμή 0.05).

**Πίνακας 3.2.** Κρίσιμες τιμές  $s$  και ακριβή επίπεδα σημαντικότητας  $a$  για στατιστικούς ελέγχους προτεραιότητας

		$r = 1$		$r = 2$		$r = 3$		$r = 4$		$r = 5$	
$n$	$m$	$s$	$a$	$s$	$a$	$s$	$a$	$s$	$a$	$s$	$a$
10	10	4	0.043	6	0.029	7	0.035	8	0.035	9	0.029
10	15	3	0.052	4	0.064	5	0.062	6	0.053	7	0.042
15	15	4	0.049	6	0.040	7	0.054	8	0.064	10	0.033
15	20	3	0.070	5	0.040	6	0.046	7	0.047	8	0.045
20	20	4	0.053	6	0.046	7	0.064	9	0.041	10	0.048
20	25	4	0.033	5	0.052	6	0.064	8	0.034	9	0.036
25	25	4	0.055	6	0.049	8	0.037	9	0.048	10	0.057
25	30	4	0.037	5	0.061	7	0.038	8	0.045	9	0.049
30	30	4	0.056	6	0.051	8	0.040	9	0.052	10	0.063

Η επιλογή της συνάρτησης  $P_{(r)}$  ως στατιστική συνάρτηση που απεικονίζεται σε ένα διάγραμμα, οδηγεί στην κατασκευή ενός μη παραμετρικού διαγράμματος ελέγχου προτεραιότητας (*precedence control chart*). Πράγματι, η εντός ελέγχου κατανομή της  $P_{(r)}$ , όπως αυτή δίνεται στην (3.4), εξαρτάται μόνο από τις παραμέτρους σχεδιασμού

$m, n, j$  και όχι από την κατανομή της διεργασίας. Συνεπώς, ένας κανόνας απόφασης που βασίζεται στην  $P_{(r)}$ , αλλά και το εντός ελέγχου μήκος ροής, θα είναι ελεύθερα κατανομής, με την προϋπόθεση ότι η κατανομή της διεργασίας είναι συνεχής. Με άλλα λόγια, τα διαγράμματα ελέγχου που στηρίζονται σε στατιστικές συναρτήσεις προτεραιότητας, είναι ανεξάρτητα της κατανομής της διεργασίας, οπότε ανήκουν στην κατηγορία των μη παραμετρικών διαγραμμάτων.

Οι Eilbott & Nadles (1965) μελέτησαν τις ιδιότητες των στατιστικών ελέγχων προτεραιότητας, στις περιπτώσεις που οι κατανομές  $F_X, F_Y$  είναι εκθετικές, προσδιορίζοντας την ισχύ τους. Στην εργασία του Shorack (1967) τα αποτελέσματα των Eilbott & Nadles (1965) επεκτείνονται σε μία ευρύτερη κλάση εναλλακτικών υποθέσεων της μορφής  $H_A : F_Y(x) = (F_X(x))^\gamma$ , όπου  $\gamma$  είναι ένας θετικός ακέραιος αριθμός. Οι παραπάνω υποθέσεις, που έχουν εισαχθεί από τον Lehmann (1953), ονομάζονται εναλλακτικές τύπου *Lehmann* (*Lehmann alternatives*) και περιλαμβάνουν την κλάση των εκθετικών και των *Weibull* κατανομών. Είναι χρήσιμο να γίνει διάκριση των εναλλακτικών υποθέσεων *Lehmann* στις κλάσεις:

- (i)  $H_1 : F_Y(x) = (F_X(x))^\gamma$ , για  $\gamma > 1$  ( $\gamma < 1$ ), η οποία για μικρές τιμές της παραμέτρου  $\gamma$ , αντιστοιχεί σε μετατόπιση προς τα δεξιά (προς τα αριστερά) της κατανομής  $F_Y$  ως προς την  $F_X$
- (ii)  $H_2 : F_Y(x) = 1 - (1 - F_X(x))^\gamma$ , για  $\gamma > 1$  ( $\gamma < 1$ ), η οποία για μικρές τιμές της παραμέτρου  $\gamma$ , αντιστοιχεί σε μετατόπιση προς τα αριστερά (προς τα δεξιά) της κατανομής  $F_Y$  ως προς την  $F_X$ .

Για περισσότερες λεπτομέρειες σχετικά με τις εναλλακτικές υποθέσεις τύπου *Lehmann*, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Davies (1971) και Lehmann (1975). Οι Lin & Sukhatme (1992) μελέτησαν την ισχύ των στατιστικών ελέγχων προτεραιότητας έναντι της εναλλακτικής  $H_2$  για  $\gamma < 1$ , ενώ για τις υπόλοιπες κλάσεις εναλλακτικών υποθέσεων, αντίστοιχα αποτελέσματα παρουσιάζονται στην εργασία των Van der Laan & Chakraborti (2001). Ο Slud (1992) γενίκευσε τους ελέγχους προτεραιότητας, υλοποιώντας συγκρίσεις του  $r$ -οστού ποσοστημορίου Kaplan–Meier του ενός δείγματος, με το αντίστοιχο  $s$ -οστό ποσοστημόριο του δεύτερου δείγματος. Για αναλυτικότερη περιγραφή των ιδιοτήτων των στατιστικών ελέγχων προτεραιότητας, αλλά και ανασκόπηση της

σχετικής διεθνούς βιβλιογραφίας για πλήρη και περικομμένα δεδομένα, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Chakraborti & Van der Laan (1996, 1997) και τη μονογραφία των Balakrishnan & Ng (2006).

Αξίζει να σημειωθεί ότι η εναλλακτική υπόθεση  $H_1 : F_Y(x) = (F_X(x))^\gamma$  ανήκει στη γενικότερη κλάση εναλλακτικών  $H_A : F_Y(x) < F_X(x)$ , για  $\gamma > 1$ . Η ισχύς του ελέγχου είναι η πιθανότητα απόρριψης της μηδενικής υπόθεσης  $H_0$ , ενώ ισχύει η εναλλακτική υπόθεση. Υπό την  $H_1$ , η σχέση (3.5) γράφεται ως εξής

$$P(P_{(r)} = j | H_1 : F_Y(x) = (F_X(x))^\gamma) = \binom{n}{j} \int_{-\infty}^{\infty} (F_X(x))^j (1 - F_X(x))^{n-j} \frac{m!}{(r-1)!(m-r)!} \\ \times (F_X(x))^{\gamma(r-1)} \left\{ \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} (F_X(x))^k \right\} \\ \times \left\{ \gamma (F_X(x))^{\gamma-1} f_X(x) \right\} dx.$$

Υλοποιώντας κατάλληλους μετασχηματισμούς στο παραπάνω ολοκλήρωμα, συμπεραίνουμε ότι η ισχύς του στατιστικού ελέγχου προτεραιότητας δίνεται από την ακόλουθη σχέση

$$P(P_{(r)} \geq s | H_1 : F_Y(x) = (F_X(x))^\gamma) = \sum_{j=s}^n \frac{m!n!\gamma}{j!(r-1)!(m-r)!} \\ \times \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \frac{\Gamma(j + \gamma(r+k))}{\Gamma(n + \gamma(r+k) + 1)}.$$

Μια ιδιαίτερα δημοφιλής επιλογή στατιστικής συνάρτησης προτεραιότητας για τον έλεγχο ισονομίας δύο δειγμάτων και συνεπώς για την κατασκευή ενός ελεύθερου κατανομής διαγράμματος ελέγχου, είναι  $r = (n+1)/2$  με  $n$  περιττό. Στη συγκεκριμένη περίπτωση, η παρατήρηση  $Y_{(r)}$  του τυχαίου δείγματος που λαμβάνεται από τη διεργασία είναι η διάμεσος, η οποία συγκρίνεται με τα όρια ελέγχου που έχουν εκτιμηθεί από κατάλληλες διατεταγμένες παρατηρήσεις του δείγματος αναφοράς. Οι Janacek & Meikle (1997) πρότειναν την κατασκευή ενός μη παραμετρικού διαγράμματος ελέγχου τύπου *Shewhart* που βασίζεται στη διάμεσο με όρια που υπολογίζονται με τη βοήθεια ενός εντός ελέγχου δείγματος αναφοράς, ενώ οι Chakraborti *et al.* (2004) και Balakrishnan *et al.* (2008a, 2008c) μελέτησαν το

παραπάνω διάγραμμα, υπολογίζοντας το μέσο μήκος ροής του με τη βοήθεια μιας τεχνικής δέσμευσης (βλ. Chakraborti (2000)) και προτείνοντας ταυτόχρονα πρόσθετες γενικεύσεις του. Τα μη παραμετρικά διαγράμματα ελέγχου με χρήση της διαμέσου και ορισμένες νέες γενικεύσεις αυτών μελετώνται εκτενώς στο Κεφάλαιο 4 του παρόντος κειμένου.

Πρόσθετα, για τον έλεγχο υποθέσεων που διατυπώνονται στην (3.2) ή ισοδύναμα για την κατασκευή διαγραμμάτων ελέγχου προτεραιότητας, μπορεί να χρησιμοποιηθεί η στατιστική συνάρτηση *Wilcoxon rank-sum*, η οποία έχει εισαχθεί στην εργασία του Wilcoxon (1945). Ορισμένα νέα μη παραμετρικά διαγράμματα ελέγχου που κάνουν χρήση της παραπάνω βαθμολογικής συνάρτησης σε συνδυασμό με ένα εντός ελέγχου δείγμα αναφοράς για την εκτίμηση των ορίων ελέγχου, εξετάζονται εκτενώς στο Κεφάλαιο 5 του παρόντος κειμένου.

РАНЕКЪМЪО РЕПАА