

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΜΗ ΠΑΡΑΜΕΤΡΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

Ιωάννης Γ. Κωνσταντίνου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Ιούνιος 2021

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

NON PARAMETRIC CONTROL
CHARTS AND APPLICATIONS

Ioannis G. Konstantinou

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfilment of the
requirements for the degree of Master of Science in Applied
Statistics

Piraeus, Greece

June 2021

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών.

Τα μέλη της Επιτροπής ήταν:

- Μάρκος Κούτρας (Επιβλέπων)
- Δημήτριος Αντζουλάκος
- Χαράλαμπος Ευαγγελάρας

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

Στην οικογένεια μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Μ. Κούτρα, για την πολύτιμη βοήθεια του και τις συνεχείς κατευθύνσεις που μου παρείχε σε όλη τη διάρκεια της συγγραφής της παρούσας διπλωματικής εργασίας. Είναι πολύ σημαντικό να επισημάνω ότι χωρίς τη συμβολή του, η εργασία αυτή δεν θα είχε την παρούσα μορφή.

Θα ήθελα επίσης να ευχαριστήσω, τα μέλη της τριμελούς επιτροπής, τον Αναπληρωτή Καθηγητή κ. Δ Αντζουλάκο και τον Επίκουρο Καθηγητή κ. Χ. Ευαγγελάρα, για τον χρόνο που αφιέρωσαν στη διόρθωση αυτής της εργασίας. Θα ήταν παράληψη να μην ευχαριστήσω τους καθηγητές του προγράμματος μεταπτυχιακών σπουδών για τις πολύτιμες γνώσεις που μου προσέφεραν σε όλο το διάστημα της φοίτησης μου σε αυτό.

Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου και τον αδερφό μου για τη συνεχή στήριξη και συμπαράσταση καθ' όλη τη διάρκεια των σπουδών μου μέχρι και την ολοκλήρωση της εργασίας αυτής.

Περίληψη

Ο Στατιστικός Έλεγχος Ποιότητας αποτελεί ένα πολύ σημαντικό μέρος της παραγωγής προϊόντων και υπηρεσιών. Κομμάτι αυτού είναι και ο Στατιστικός Έλεγχος Διεργασιών ο οποίος χρησιμοποιεί διάφορα εργαλεία για την παρακολούθηση μιας διεργασίας. Ένα από αυτά είναι τα Διαγράμματα Ελέγχου τα οποία θα αποτελέσουν το αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας. Πιο συγκεκριμένα, στην παρούσα διπλωματική εργασία θα παρουσιάσουμε τα Μη Παραμετρικά Διαγράμματα Ελέγχου.

Τα κλασικά δ.ε. αφορούν δεδομένα που ακολουθούν κάποια συγκεκριμένη κατανομή και είναι αρκετά αποδοτικά στον εντοπισμό μιας μεσαίας ή μεγάλης μετατόπισης σε κάποια παράμετρο της παραγωγικής διαδικασίας, τέτοια διαγράμματα είναι το διάγραμμα τύπου Shewhart, EWMA, CUSUM κ.α.. Εν τούτοις, όταν τα δεδομένα δεν ακολουθούν κάποια συγκεκριμένη κατανομή ή η κατανομή τους είναι άγνωστη, η μόνη λύση για την παρακολούθηση της διεργασίας είναι να κατασκευάσουμε μη παραμετρικά δ.ε.. Ένα χαρακτηριστικό των διαγραμμάτων αυτών είναι ότι έχουν ευαισθησία και μπορούν να εντοπίσουν ακόμη και μικρές μετατοπίσεις σε μια παραγωγική διαδικασία.

Στην παρούσα διπλωματική εργασία θα γίνει μια εισαγωγή στο Στατιστικό Έλεγχο Ποιότητας και στο Στατιστικό Έλεγχο Διεργασιών, σύντομη ιστορική αναδρομή στην ποιότητα, θα δοθεί ο ορισμός των μη παραμετρικών δ.ε. καθώς επίσης τα πλεονεκτήματα και τα μειονεκτήματα αυτών. Στη συνέχεια θα γίνει παρουσίαση μερικών από τις πιο σημαντικές μη παραμετρικές μεθοδολογίες που χρησιμοποιούνται σε μη παραμετρικά δ.ε. καθώς επίσης και σύγκριση μερικών εξ αυτών. Παράλληλα, θα γίνει μια σύντομη παρουσίαση και σύγκριση τεσσάρων νέων μεθοδολογιών μη παραμετρικών δ.ε. που έχουν εμφανιστεί στην πρόσφατη βιβλιογραφία με τη χρήση κώδικα στη γλώσσα προγραμματισμού R. Θα γίνει μια σύντομη βιβλιογραφική παρουσίαση σε πολυμεταβλητά δ.ε. καθώς και σε δύο πολυμεταβλητά δ.ε. για ταυτόχρονη παρακολούθηση του μέσου και της διασποράς. Τέλος, θα γίνει παρουσίαση περιοχών που εφαρμόζονται τα μη παραμετρικά δ.ε. όπως είναι η ιατρική, η συχνότητα διαδοχικών γεγονότων και η τρισδιάστατη εκτύπωση.

Abstract

Statistical Quality Control is a very important part of the production of products and services. Part of this is the Statistical Process Control which uses various tools to monitor a process, one of them and the subject of study of this dissertation is the Control Charts. More specifically, the Non-Parametric Control Charts.

Classical control charts (e.g. Shewhart type control chart, EWMA, CUSUM etc) can be applied to data that follow a specific distribution and are quite efficient in identifying a medium or large shift in some aspect of the production process. However, when the data do not follow a specific distribution or their distribution is unknown, the only solution for monitoring the process is to construct non-parametric control charts. One interesting feature of the non-parametric control charts is that they are quite sensitive and can detect even small shifts in a production process.

In this dissertation we shall first present an introduction to the Statistical Quality Control and Statistical Process Control, a brief historical flashback to quality, the definition of non-parametric control charts as well as the advantages and disadvantages. Next, we shall present some of the most important non-parametric methodologies of non-parametric control charts as well as a comparison between some of them. At the same time, we shall proceed to a brief presentation and comparison of four new non-parametric control charts methodologies that have appeared in recent literature using code in the programming language R. A brief bibliographical presentation will be made to multivariate control charts as well as to two multivariate control charts for simultaneous monitoring of the medium and dispersion. Finally, we shall discuss about areas where the non-parametric control charts can be applied such as medicine, frequency of sequential events and 3D printing.

Περιεχόμενα

Κατάλογος Πινάκων	xii
Κατάλογος Σχημάτων	xiv
Κατάλογος Συντομογραφιών	xvi
1. Εισαγωγή	1
1.1 Μια ιστορική αναδρομή στον έλεγχο ποιότητας	2
1.2 Τα 7 πιο χρήσιμα εργαλεία στο Στατιστικό Έλεγχο Διεργασιών	4
1.3 Μεταβλητότητα στην Παραγωγική Διαδικασία	5
1.4 Μη παραμετρικά δ.ε.	6
1.5 Φάση I και Φάση II μιας Παραγωγικής Διαδικασίας	10
1.6 Πληροφορίες για τα μη παραμετρικά δ.ε.	11
2. Μεθοδολογίες στα μη παραμετρικά διαγράμματα ελέγχου	13
2.1 Μέτρα απόδοσης για τη σύγκριση των δ.ε.	13
2.1.1 Δ.ε. βασισμένα στη συνάρτηση προτεραιότητας	14
2.1.2 Δ.ε. βασισμένα στη διάμεσο	15
2.1.3 Μονόπλευρα δ.ε.	16
2.2 Δ.ε. για τον Έλεγχο της μεταβλητότητας	18
2.2.1 Οι εκτιμητές S_n και Q_n	19
2.2.2 Σύγκριση των διαγραμμάτων S_n και Q_n με το διάγραμμα S	20
3. Μη παραμετρικά δ.ε. που βασίζονται στην Προσημική Συνάρτηση	22
3.1 Μη παραμετρικά δ.ε. τύπου Shewhart που βασίζονται στην SN	22
3.1.1 Δ.ε. για την παρακολούθηση της κεντρικής τιμής	23
3.1.2 Συμπεράσματα για τα δ.ε. που βασίζονται στη συνάρτηση SN	25
3.2 Δ.ε. που βασίζονται στην Προσημική-Βαθμολογική Συνάρτηση	26
3.2.1 Δ.ε. SR για τον έλεγχο του κέντρου της διεργασίας	26
3.2.2 Συμπεράσματα για τα δ.ε. που βασίζονται στη συνάρτηση SR	29

4.	Νέες μέθοδοι μη παραμετρικών διαγραμμάτων ελέγχου	30
4.1	Διάγραμμα <i>CUSUM</i> για το μέσο	30
4.2	Διάγραμμα <i>EWMA</i> που χρησιμοποιεί την προσημική συνάρτηση	31
4.3	Διάγραμμα <i>CUSUM</i> για τη διασπορά	33
4.4	Διάγραμμα <i>EWMA</i> για τη διασπορά	34
4.5	Σύγκριση νέων μεθοδολογιών	36
5.	Πολυμεταβλητά διαγράμματα ελέγχου	39
5.1	Διάφορα Πολυμεταβλητα δ.ε.	39
5.2	Πολυμεταβλητα Διαγράμματα ταυτόχρονου ελέγχου του μέσου και της διασποράς	40
6.	Εφαρμογές μη παραμετρικών διαγραμμάτων ελέγχου	42
6.1	Εφαρμογή σε δεδομένα του καρκίνου του μαστού	42
6.2	Εφαρμογή για το μέγεθος και τη συχνότητα διαδοχικών γεγονότων	43
6.3	Εφαρμογή σε δεδομένα τρισδιάστατης εκτύπωσης	44
Παράρτημα	Κώδικες υπολογισμού ARL_1 των διαγραμμάτων του Κεφαλαίου 4	47
Βιβλιογραφία		53

Κατάλογος Πινάκων

1.1	ARL_0 για διάφορες κατανομές διαγράμματος \bar{X} Shewhart.	11
2.1	Δ.ε. προτεραιότητας για καθορισμένο FAR .	17
2.2	ARL_0 για τα τρία δ. ε. για διάφορες κατανομές	20
2.3	Σύγκριση των ARL_1 για τα τρία δ.ε. και για διαφορετικές κατανομές.	21
3.1	$L^+(\mu)$ για διαγράμματα SN τύπου Shewhart όταν $n = 10$.	24
3.2	$L^+(\mu)$ σε συνάρτηση με το p για διαγράμματα SN όταν $n = 10$.	25
3.3	FAR και ARL_0^+ για $n = 4,5$ και 6 .	27
3.4	FAR και ARL_0^+ για $n = 8$ και 10 .	28
4.1	ARL_1 για τα διαγράμματα για τον μέσο με $n = 10$.	37
4.2	ARL_1 για τα διαγράμματα για τη διασπορά με $n = 10$.	37

Κατάλογος Σχημάτων

1	Βασικό δ.ε. τύπου Shewhart.	7
2	Διάγραμμα T_k .	40
3	Διάγραμμα $EWMA$ για δεδομένα καρκίνου του μαστού Coimbra.	42

Κατάλογος Συντομογραφιών

τ.μ.	τυχαία μεταβλητή
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας
σ.κ.	συνάρτηση κατανομής
δ.ε.	Διάγραμμα Ελέγχου ή Διαγράμματα Ελέγχου – Control Chart
<i>ΣΕΑ-SPC</i>	Statistical Process Control - Στατιστικός έλεγχος διεργασιών
<i>ΣΕΠ-SQC</i>	Statistical Quality Control - Στατιστικός έλεγχος ποιότητας
<i>ΚΟΘ</i>	Κεντρικό Οριακό Θεώρημα
<i>ARL</i>	Average Run Length – μέσο μήκος ροής
<i>ARL₀</i>	In-Control Average Runs Length – Εντός ελέγχου μέσο μήκος ροής
<i>ARL₁</i>	Out-of-Control Average Runs Length – Εκτός ελέγχου μέσο μήκος ροής
<i>FAR</i>	False Alarm Rate - Πιθανότητα λανθασμένου συναγερμού
<i>UCL</i>	Άνω όριο ενός δ.ε.
<i>CL</i>	Κεντρική γραμμή ή Τιμή-στόχος ενός δ.ε.
<i>LCL</i>	Κάτω όριο ενός δ.ε.
<i>CUSUM</i>	Cumulative Sum – συσσωρευμένο άθροισμα
<i>EWMA</i>	Exponentially Weighted Moving Average – Εκθετικά ζυγισμένος κινούμενος μέσος
<i>SN test</i>	Sign test – Προσημικός έλεγχος
<i>SR test</i>	Wilcoxon's Singed-Rank Test – προσημικός βαθμολογικός έλεγχος του Wilcoxon
<i>SN</i>	Sign Chart
Διάγραμμα	
<i>SR</i>	Singed-Rank Wilcoxon statistic Chart
Διάγραμμα	

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Η ποιότητα αποτελεί έναν από τους σημαντικότερους παράγοντες λήψης αποφάσεων για την επιλογή των ανταγωνιστικών προϊόντων και υπηρεσιών. Το φαινόμενο είναι ευρύτατο, ανεξάρτητα από το αν ο καταναλωτής είναι άτομο, βιομηχανικός οργανισμός, κατάστημα λιανικής πώλησης, τράπεζα, χρηματοπιστωτικό ίδρυμα ή στρατιωτικό αμυντικό πρόγραμμα. Κατά συνέπεια, η κατανόηση και η βελτίωση της ποιότητας είναι βασικοί παράγοντες που οδηγούν στην επιτυχία των επιχειρήσεων, την ανάπτυξη και την ενισχυμένη ανταγωνιστικότητα.

Η ποιότητα ενός προϊόντος μπορεί να περιγραφεί και να αξιολογηθεί με διάφορους τρόπους. Ωστόσο, ο παραδοσιακός ορισμός της ποιότητας βασίζεται στην άποψη ότι τα προϊόντα και οι υπηρεσίες πρέπει να πληρούν τις απαιτήσεις των καταναλωτών. Παράλληλα, οι στατιστικές μέθοδοι παίζουν πολύ σημαντικό ρόλο στη διαδικασία βελτίωσης της ποιότητας καθώς και στην κατασκευή και εξυπηρέτηση των βιομηχανιών.

Ο ποιοτικός έλεγχος είναι μια διαδικασία μέσω της οποίας η διοίκηση επιδιώκει να εξασφαλίσει ότι η ποιότητα του προϊόντος διατηρείται ή βελτιώνεται και τα σφάλματα κατασκευής μειώνονται ή εξαλείφονται. Αυτό επιτυγχάνεται με την επιθεώρηση των παραγόμενων προϊόντων. Οι δύο πιο δημοφιλείς μέθοδοι ελέγχου ποιότητας είναι ο Στατιστικός Έλεγχος Ποιότητας - ΣΕΠ (Statistical Quality Control - SQC) και ο Στατιστικός Έλεγχος Διεργασιών - ΣΕΔ (Statistical Process Control - SPC) ο οποίος είναι υποκατηγορία του ΣΕΠ.

Ο ΣΕΔ είναι μια τεχνική πρόληψης αστοχιών και αποτελείται από μεθόδους κατανόησης, παρακολούθησης, και βελτιστοποίησης της απόδοσης μιας βιομηχανίας. Ο ΣΕΠ είναι μια τεχνική που στοχεύει στην ανίχνευση και στην αφαίρεση σφαλμάτων αντικειμένων απόρριψης προκειμένου να ικανοποιηθούν οι προδιαγραφές ποιότητας όπως αυτές καθορίζονται.

Οι στόχοι του ΣΕΠ είναι η εξάλειψη των μη ποιοτικών προϊόντων, η εξάλειψη των επαναχρησιμοποιούμενων και των περιττών πηγών και η επίτευξη των στόχων αυτών με το χαμηλότερο δυνατό κόστος.

1.1 Μια ιστορική αναδρομή στον έλεγχο ποιότητας

Η ποιότητα αποτελεί αναπόσπαστο μέρος σχεδόν όλων των προϊόντων και υπηρεσιών. Ωστόσο, η συνειδητοποίηση της σπουδαιότητάς και η εισαγωγή των επίσημων μεθόδων για τον ποιοτικό έλεγχο και τη βελτίωση ήταν μια εξελικτική αξιοποίηση.

Ωστόσο, με την τυποποίηση των μεθόδων εργασίας ήρθε η έννοια των εργασιακών προτύπων σε πραγματικό χρόνο για την ολοκλήρωση της εργασίας και για ένα καθορισμένο αριθμό μονάδων που πρέπει να παράγονται ανά περίοδο. Ο Frank Gilbreth και άλλοι επέκτειναν αυτή την έννοια στη μελέτη του σχεδιασμού κίνησης και εργασίας. Πολλά από αυτά είχαν θετικό αντίκτυπο στην παραγωγικότητα, αλλά συχνά δεν τονίζουν επαρκώς την ποιότητα της εργασίας. Στη συνέχεια έχουμε μια ιστορική αναδρομή στον ποιοτικό έλεγχο (Montgomery, 2009).

- 1875: Ο Frederick W. Taylor εισάγει τις αρχές της «επιστημονικής διαχείρισης» για να διαιρέσει την εργασία σε μικρότερες και πιο εύκολα επιτυχημένες μονάδες.
- 1900-1930: Ο Henry Ford εισήγαγε τη γραμμή συναρμολόγησης με σκοπό την περαιτέρω βελτίωση των μεθόδων εργασίας για τη βελτίωση της παραγωγικότητας και της ποιότητας. Η Ford ανέπτυξε έννοιες συναρμολόγησης ανθεκτικές σε λάθη, αυτοέλεγχο και επιθεώρηση κατά τη διάρκεια της διαδικασίας.
- 1907-1908: Η AT & T ξεκινά συστηματική επιθεώρηση και δοκιμή προϊόντων και υλικών.
- 1908: Ο W.S. Gosset εισάγει την κατανομή t και αποτελέσματα από την εργασία του σχετικά με τον έλεγχο ποιότητας στο Guinness Brewery.
- 1919: Ιδρύεται στην Αγγλία ο Σύνδεσμος Τεχνικής Επιθεώρησης και αργότερα γίνεται το Ινστιτούτο Διασφάλισης Ποιότητας.
- 1920: Ο B. Dudding στη General Electric στην Αγγλία χρησιμοποιεί στατιστικές μεθόδους για τον έλεγχο της ποιότητας ηλεκτρικών λαμπτήρων.
- 1924 και 1931: Ο W.A. Shewhart εισάγει την ιδέα του Διαγράμματος Ελέγχου σε ένα τεχνικό μνημόνιο της Bell Laboratories και δημοσιεύει στατιστικές μεθόδους οικονομικού ελέγχου ποιότητας των παραγόμενων προϊόντων για χρήση σε μεθόδους παραγωγής και διαγραμμάτων ελέγχου.

- 1940: Το αμερικανικό τμήμα πωλήσεων δημοσιεύει έναν οδηγό για τη χρήση διαγραμμάτων ελέγχου για την ανάλυση δεδομένων διεργασίας.
- 1942: Στη Μεγάλη Βρετανία, δημιουργείται το Υπουργείο Παροχής Συμβουλευτικών Υπηρεσιών για τις στατιστικές μεθόδους και τον ποιοτικό έλεγχο.
- 1944: Δημοσιεύεται το βιβλίο «Industrial Quality Control».
- 1946: Ιδρύεται ο Διεθνής Οργανισμός Τυποποίησης (ISO) και διαμορφώνεται η Ιαπωνική Ένωση Επιστημόνων και Μηχανικών (JUSE).
- 1948: Ο G. Taguchi ξεκινά τη μελέτη και εφαρμογή του πειραματικού σχεδιασμού.
- 1950: Ο Deming ξεκινά την εκπαίδευση των ιαπωνικών βιομηχανικών διευθυντών, οι στατιστικές μέθοδοι ελέγχου ποιότητας αρχίζουν να διδάσκονται ευρέως στην Ιαπωνία και ο K. Ishikawa εισάγει το διάγραμμα αιτίου-αποτελέσματος.
- 1951: Ο A.B. Feigenbaum δημοσιεύει τη πρώτη έκδοση του βιβλίου «Total Quality Control».
- 1954: Ο E.S. Page εισάγει το διάγραμμα ελέγχου τύπου CUSUM.
- 1957: Οι J.M. Juran και F.M. Gryna δημοσιεύουν το βιβλίο «Quality Control Handbook» για πρώτη φορά.
- 1959: Ο S. Roberts εισάγει το διάγραμμα ελέγχου EWMA.
- 1960: Οι G.E.P. Box και J.S. Hunter γράφουν θεμελιώδη έγγραφα σχετικά με τους 2^{k-p} παραγοντικούς σχεδιασμούς.
- 1980-1990: Οι μέθοδοι πειραματικών σχεδιασμών εισάγονται και υιοθετούνται από μια ευρύτερη ομάδα οργανισμών.
- 1987: Το ISO δημοσιεύει το πρώτο πρότυπο συστημάτων ποιότητας και ξεκινά η καινοτομία των έξι-σίγμα (six-sigma) της Motorola.
- 1997: Η προσέγγιση έξι-σίγμα της Motorola αρχίζει να εξαπλώνεται σε άλλες βιομηχανίες.
- 2000: Εκδίδεται το πρότυπο ISO 9000:2000. Η διαχείριση της αλυσίδας εφοδιασμού και η ποιότητα των προμηθευτών καθίστανται ακόμα πιο κρίσιμοι παράγοντες στην επιτυχία μιας επιχείρησης. Οι δραστηριότητες βελτίωσης της ποιότητας επεκτείνονται πέραν του παραδοσιακού βιομηχανικού περιβάλλοντος σε πολλούς άλλους τομείς, συμπεριλαμβανομένων των χρηματοπιστωτικών υπηρεσιών, της υγειονομικής περίθαλψης, της ασφάλισης και των υπηρεσιών κοινής ωφελείας.

Οι στατιστικές μέθοδοι και η εφαρμογή τους στη βελτίωση της ποιότητας είχαν μακρά ιστορία. Μεγάλο μέρος της αύξησης του ενδιαφέροντος για τον ΣΕΠ οφείλεται στην επανάσταση στον τομέα της ποιότητας, η οποία προκλήθηκε από μια όλο και πιο ανταγωνιστική διεθνή αγορά. Μια άλλη αιτία της αύξησης της ερευνητικής

δραστηριότητας μπορεί να οφείλεται σε μεγάλο βαθμό στην αυξημένη ανάγκη για έρευνα.

Σήμερα υπάρχει μια παγκόσμια αναζωπύρωση του ενδιαφέροντος για την ποιότητα και τη διαχείριση της το οποίο προέρχεται σε μεγάλο βαθμό από την παγκόσμια επιτυχία των ιαπωνικών επιχειρήσεων. Στις Η.Π.Α. η ποιότητα έχει επικρατήσει ως μια ενοποιητική κεντρική ιδέα στις επιχειρηματικές στρατηγικές πολλών επιχειρήσεων τα τελευταία χρόνια. Εταιρείες όπως οι Xerox, Motorola, Ford, Federal Express, Corning Glass, Milliken κ.α. υιοθέτησαν από τα μέσα της δεκαετίας του 1980 εταιρικές στρατηγικές με την ποιότητα ως το καθοδηγητικό θέμα. Αυτές οι προσπάθειες, που συχνά χαρακτηρίζονται ως T.Q.M. ή "Διαχείριση Ολικής Ποιότητας", σε αντίθεση με τα προγενέστερα προγράμματα ποιότητας, επιδιώκουν να διατηρήσουν τις συνεχείς διαδικασίες όπως για παράδειγμα οι Deming (1986), Juran (1951, 1989), Ishikawa (1985), Godfrey and Kolesar (1988), Garvin (1987) και Imai (1986).

1.2 Τα 7 πιο χρήσιμα εργαλεία του Στατιστικού Ελέγχου Διεργασιών

Για να παραχθεί ένα προϊόν ικανοποιητικό προς τον καταναλωτή, η παραγωγική διαδικασία θα πρέπει να είναι σταθερή και επαναλαμβανόμενη. Για να καταστεί αυτό απαραίτητη προϋπόθεση είναι η διεργασία να λειτουργεί με όσο το δυνατόν μικρότερη μεταβολή στις τιμές στόχο που έχουν οριστεί για τα ποιοτικά χαρακτηριστικά του παραγόμενου προϊόντος. Για τη σταθερότητα μιας διεργασίας έχουν οριστεί διάφορα εργαλεία τα οποία εφαρμόζονται στον ΣΕΔ και σε κάθε παραγωγική διεργασία. Τα επτά πιο χρήσιμα εργαλεία είναι τα ακόλουθα:

1. Το Φύλλο Ελέγχου (Check Sheet)
2. Το Ιστόγραμμα και το Διάγραμμα Μίσχου-Φύλλων (Histogram και Stem-and-Leaf Plot αντίστοιχα)
3. Το Διάγραμμα Pareto (Pareto Chart)
4. Το Διάγραμμα Αιτίας-Αποτελέσματος (Cause-and-Effect Diagram)
5. Το Διάγραμμα Συγκέντρωσης Ελαττωμάτων (Defect Concentration Diagram)
6. Το Διάγραμμα Διασποράς/Διασκόρπισης (Scatter Plot)
7. Το Διάγραμμα Ελέγχου (Control Chart)

Τα παραπάνω επτά εργαλεία έχουν οριστεί και ως “The Magnificent Seven” (Αντζουλάκος, 2016). Ωστόσο, από τα επτά μόνο το Διάγραμμα Ελέγχου - Δ.Ε. θα αποτελέσει βασικό αντικείμενο μελέτης στα επόμενα κεφάλαια.

Η κατασκευή δ.ε. γίνεται με σκοπό την παρακολούθηση των χαρακτηριστικών της μέσης τιμής και της διακύμανσης λαμβάνοντας τυχαία δείγματα από τα παραγόμενα προϊόντα. Επίσης, αποτελεί μια τεχνική που χρησιμοποιείται ευρέως για την ανίχνευση ειδικών αιτιών μεταβλητότητας σε πραγματικό χρόνο σε μια διεργασία.

1.3 Μεταβλητότητα στην παραγωγική διαδικασία

Η κατανόηση της διακύμανσης των τιμών ενός χαρακτηριστικού ποιότητας έχει αδιαμφισβήτητο πρωταρχικό ρόλο στο ΣΕΔ. Μια συνηθισμένη αιτία μεταβλητότητας οφείλεται στη φύση της διεργασίας και δεν μπορεί να παραληφθεί χωρίς να αλλάξει η ίδια η διεργασία.

Ο Shewhart ήξερε ότι η μεταβλητότητα είναι εγγενής σε όλες τις φυσικές διεργασίες. Με μια συγκεκριμένη παραγωγική διαδικασία και ένα σταθερό σχέδιο, θα μπορούσαν να γίνουν λίγες ενέργειες ώστε να μειωθούν οι μεταβολές αυτές, εκτός από ένα θεμελιώδη ανασχεδιασμό ή ανασχεδιασμό προϊόντων ή διεργασιών ή και των δύο.

Οι αιτίες που μπορούν να αποδοθούν στη μεταβλητότητα είναι τα ασυνήθιστα σοκ ή άλλες διαταραχές στη διεργασία, οι οποίες μπορούν και πρέπει να αφαιρεθούν. Ένας σκοπός των δ.ε. είναι η διάκριση μεταξύ αυτών των δύο τύπων μεταβλητότητας, προκειμένου να αποφευχθεί η υπερβολική ή ανεπαρκής αντίδραση σε μια διεργασία. Η διάκριση μεταξύ των κοινών αιτιών και των αιτιών που μπορούν να αφαιρεθούν, εξαρτάται από το περιβάλλον της διεργασίας.

Μια κοινή αιτία σήμερα μπορεί να είναι μια σημαντική αιτία μεταβλητότητας στο μέλλον. Η προέλευση της αιτίας αυτής μπορεί να αλλάξει με μια αλλαγή στο δειγματοληπτικό σχέδιο. Στη συνέχεια πρέπει να προχωρήσουμε σε αφαίρεση της μόνο όταν έχει ικανό αντίκτυπο έτσι ώστε να έχει πρακτικές και οικονομικές επιπτώσεις στη βελτίωση της ποιότητας.

Οι δύο πηγές μεταβλητότητας διακρίνονται, σε εκείνες που δεν έχουν τη δυνατότητα να προσδιοριστούν με κάποιο τρόπο-τυχαίες (chance cause) και εκείνες που μπορούν να αποδοθούν σε κάποια αιτία (assignable cause). Όταν μια διεργασία λειτουργεί μόνο

υπό τυχαίες αιτίες τότε βρίσκεται σε κατάσταση στατιστικού ελέγχου. Το δ.ε. βοηθά τους ερευνητές να εντοπίζουν και να εξαλείφουν τα αίτια που οφείλονται κάπου, ώστε η διεργασία να διατηρείται εντός στατιστικού ελέγχου.

Στα πλαίσια του ελέγχου διεργασιών, το μοτίβο των τυχαίων αιτιών συχνά υποτίθεται ότι ακολουθεί κάποια παραμετρική κατανομή. Η συνηθέστερη υπόθεση είναι ότι η κατανομή πιθανότητας είναι η Κανονική. Οι στατιστικές ιδιότητες των συνήθως χρησιμοποιούμενων δ.ε. είναι ακριβείς μόνο εάν ισχύει αυτή η υπόθεση. Ωστόσο, η υποκείμενη διεργασία σε πολλές εφαρμογές δεν είναι Κανονική και αυτό έχει ως αποτέλεσμα, οι στατιστικές ιδιότητες των τυποποιημένων διαγραμμάτων να επηρεάζονται σε μεγάλο βαθμό από τέτοιες καταστάσεις.

1.4 Μη παραμετρικά δ.ε.

Όπως είδαμε στην προηγούμενη ενότητα, τα δ.ε. βοηθούν τους ερευνητές να εντοπίζουν και να εξαλείφουν τις μη τυχαίες αιτίες μεταβλητότητας. Σε περίπτωση αλλαγής της διεργασίας, ένα δ.ε. θα πρέπει να είναι σε θέση να το ανιχνεύσει το συντομότερο δυνατόν και να εκπέμψει ένα σήμα εκτός ελέγχου. Συνεπώς, όσο ταχύτερη είναι η ανίχνευση και το σήμα, τόσο πιο αποδοτικό είναι το διάγραμμα.

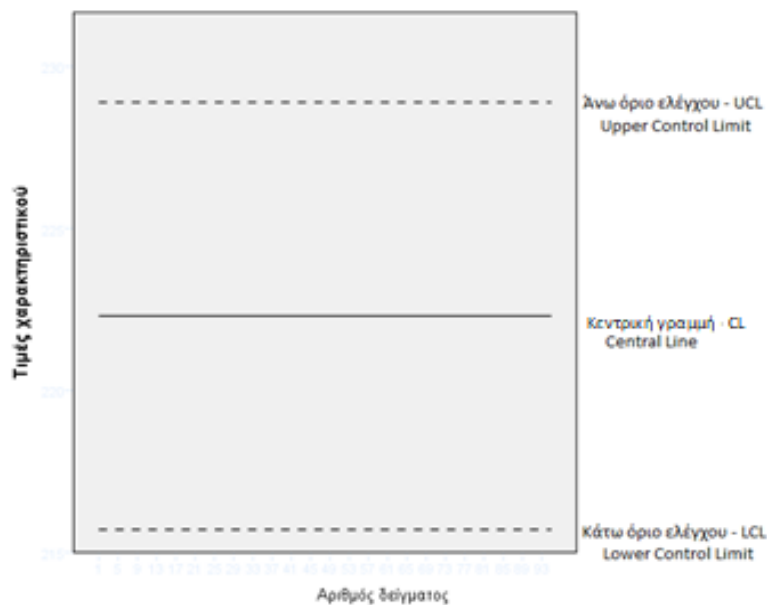
Όπως γνωρίζουμε, για τη χρήση των μη παραμετρικών δ.ε. δεν απαιτείται κάποια συγκεκριμένη κατανομή για τα χαρακτηριστικά τους.

Η βασική ιδέα ενός δ.ε. είναι απλή. Ας υποθέσουμε ότι παράγουμε αντικείμενα και υπάρχει κάποιο χαρακτηριστικό προς έλεγχο. Για την παρακολούθηση της διαδικασίας παραγωγής, λαμβάνονται N δείγματα από αυτή σε τακτά χρονικά διαστήματα και στη συνέχεια μετριέται το σημαντικό χαρακτηριστικό. Η παρουσία μιας πηγής μη τυχαίας μεταβλητότητας στο μετρούμενο χαρακτηριστικό υποθέτουμε ότι εμφανίζεται στα δείγματα ως μία μεταβολή στις μετρήσεις που είναι εκτός των ορίων μιας κανονικής παρτίδας. Αυτή η μεταβολή αποδίδεται σε μια μετατόπιση του μέσου ή της διακύμανσης και για την ανίχνευση τέτοιων αλλαγών χρησιμοποιούμε εκτιμώμενες τιμές τους από τα δείγματα μας. Επομένως μπορεί να γίνει μια σειρά από γραφικά τεστ για την απεικόνιση των παραπάνω μεταβλητών. Ένα σήμα συναγερμού υποδεικνύει ότι η διαδικασία βρίσκεται σε κατάσταση εκτός ελέγχου, πράγμα που μας δείχνει ότι απαιτείται η διερεύνηση της προέλευσης της πηγής μεταβλητότητας και εάν κριθεί αναγκαίο, λαμβάνονται μέτρα για την εξάλειψή της. Από την άλλη πλευρά, εάν δεν

υπάρχουν σημεία εκτός των ορίων ελέγχου, συνεχίζουμε να αντλούμε διαδοχικά δείγματα για να παρακολουθήσουμε την παραγωγική διαδικασία.

Όμως, εάν οι πληθυσμοί μας αποκλίνουν από την Κανονικότητα, τότε χρησιμοποιώντας στατιστικά τεστ για τη μέση τιμή, μπορεί να δημιουργηθούν διαγράμματα που είτε δεν θα ανιχνεύσουν πραγματικές αλλαγές στη διεργασία είτε θα δημιουργήσουν ψευδείς προειδοποιήσεις ενώ η διεργασία δεν έχει αλλάξει. Για να επιλύσουμε αυτό το πρόβλημα και να διατηρήσουμε την απαραίτητη δομή των κλασικών δ.ε., έχουν προταθεί διάφορες τροποποιήσεις που βασίζονται σε απλές μη παραμετρικές μεθόδους.

Ένα τυπικό δ.ε. τύπου Shewhart παρουσιάζεται στο Σχήμα 1.



Σχήμα 1: Βασικό δ.ε. τύπου Shewhart

Το δ.ε. κατασκευάζεται με γραφική αναπαράσταση των στατιστικών τιμών στον κατακόρυφο άξονα και τον αριθμό δείγματος στον οριζόντιο. Έχει συνήθως μια (οριζόντια) γραμμή στην κάθε πλευρά της κεντρικής γραμμής (CL). Η γραμμή πάνω από τη CL ονομάζεται ανώτερο όριο ελέγχου (UCL) ενώ η γραμμή κάτω από τη CL ονομάζεται χαμηλότερο όριο ελέγχου (LCL). Αυτές οι τρεις γραμμές τοποθετούνται στο διάγραμμα για να αποφανθούμε εάν η διεργασία βρίσκεται σε έλεγχο ή όχι. Η απόφαση αυτή βασίζεται κυρίως στο σχέδιο των σημείων που παρίστανται στο διάγραμμα σε σχέση με τα όρια ελέγχου.

Τα όρια ελέγχου αυτά καθορίζονται συνήθως σε ± 3 τυπικά σφάλματα από την κεντρική γραμμή, τα όρια ελέγχου αυτά αναφέρονται ως όρια «τριών-σίγμα» (Woodall and Montgomery, 1999).

Τα δ.ε. τύπου Shewhart χρησιμοποιούνται ως επί το πλείστον στην παρακολούθηση της διεργασίας μέσω του δειγματικού μέσου (διάγραμμα $\bar{X} - \bar{X}$ Chart), της μεταβλητότητας διεργασίας μέσω του δειγματικού εύρους (διάγραμμα $R - R$ Chart), της δειγματικής τυπικής απόκλισης (διάγραμμα $S - S$ Chart) ή του διαγράμματος διακύμανσης (S^2). Αυτά τα διαγράμματα είναι απλά, κατανοητά, εύκολο να εφαρμοστούν και χρησιμοποιούνται ευρέως στην παρακολούθηση των διεργασιών τη στιγμή της παραγωγικής διαδικασίας. Τα διαγράμματα τύπου Shewhart έχουν το μειονέκτημα ότι οι σ.σ. τους βασίζονται σε πληροφορίες μόνο για το τρέχον δείγμα, επομένως δεν είναι αποτελεσματικά για την ανίχνευση μικρών αλλαγών στη διεργασία (Zombade, 2019).

Η δυνατότητα ανίχνευσης μικρών αλλαγών στις παραμέτρους μπορεί να βελτιωθεί σημαντικά χρησιμοποιώντας δ.ε. από προηγούμενα δείγματα μαζί με το τρέχον δείγμα. Μια εναλλακτική προσέγγιση για την ανίχνευση μικρών μετατοπίσεων είναι τα διαγράμματα ελέγχου με μνήμη δηλαδή, το διάγραμμα συσσωρευτικού αθροίσματος – διάγραμμα *CUSUM* και το διάγραμμα εκθετικά σταθμισμένου κινητού μέσου – διάγραμμα *EWMA*. Αυτά τα δ.ε. έχουν σχεδιαστεί ώστε εξυπηρετούν αυτόν το σκοπό και είναι πολύ ευαίσθητα σε μικρές και μέτριες μεταβολές. Έτσι, όταν είναι επιθυμητή η γρήγορη ανίχνευση μικρών μετατοπίσεων, τα διαγράμματα αυτά αποτελούν καλύτερη εναλλακτική λύση αντί των δ.ε. τύπου Shewhart.

Σε περιπτώσεις όπου η υπόθεση κανονικότητας δεν ισχύει, τα συμβατικά δ.ε. δεν είναι αξιόπιστα στην παρακολούθηση της διεργασίας επειδή τα πραγματικά ποσοστά ψευδών συναγερμών (*FAR - false alarm rate*) τους θα μπορούσαν να είναι σημαντικά μεγαλύτερα ή μικρότερα από ένα υποτιθέμενο *FAR*. Σε περιπτώσεις όπου αυτό το πραγματικό ποσοστό είναι μεγαλύτερο από το υποτιθέμενο *FAR*, θα χανόταν πολύ εργατικό δυναμικό και πολλοί πόροι, επειδή η παραγωγική διαδικασία σταματάει πολύ συχνά. Από την άλλη πλευρά, αν το πραγματικό *FAR* είναι μικρότερο, τότε το διάγραμμα δεν θα εκπέμπει σήμα έγκαιρα για μια αλλαγή της διαδικασίας με συνεπεία να κατασκευάζονται πολλά ελαττωματικά προϊόντα χωρίς προειδοποίηση.

Τα μη παραμετρικά δ.ε. όμως μπορούν να χρησιμοποιηθούν όταν δεν είναι δεδομένο πως η κατανομή πιθανότητας είναι η Κανονική. Πρέπει να σημειωθεί ότι ο όρος μη

παραμετρικό δεν έχει σκοπό να υποδηλώσει ότι δεν υπάρχουν παράμετροι που να εμπλέκονται.

Τα κύρια πλεονεκτήματα των μη παραμετρικών δ.ε. είναι η ευελιξία τους, διότι δεν απαιτείται η υπόθεση οποιασδήποτε παραμετρικής κατανομής για τη διεργασία, τουλάχιστον όσον αφορά τη δημιουργία και την εφαρμογή τους. Στον τομέα του ΣΕΔ, αυτό είναι πολύ επωφέλης, ιδίως σε αρχικό στάδιο όπου δεν υπάρχουν πολλά διαθέσιμα δεδομένα για τη χρήση παραμετρικής διαδικασίας. Θα πρέπει να σημειωθεί ότι οι μη παραμετρικές μέθοδοι μπορεί να είναι λιγότερο αποτελεσματικές από τις αντίστοιχες παραμετρικές μεθόδους τους, υπό την προϋπόθεση ότι κάποιος έχει την πλήρη γνώση της υποκείμενης στοχαστικής διαδικασίας, στην πράξη οι πληροφορίες αυτές σπάνια, αν όχι ποτέ, είναι διαθέσιμες στον επαγγελματία ποιότητας.

Ένα ακόμη βασικό πλεονέκτημα των μη παραμετρικών δ.ε. είναι ότι η κατανομή του εντός ελέγχου μήκους ροής είναι η ίδια για όλες τις συνεχείς κατανομές. Τα μη παραμετρικά δ.ε. μοιράζονται τις ιδιότητες ανθεκτικότητας, των μη παραμετρικών ελέγχων και των διαστημάτων εμπιστοσύνης και ως εκ τούτου, είναι πολύ πιθανό να επηρεαστούν λιγότερο από ακραίες τιμές (outliers). Παράλληλα, τα διαγράμματα αυτά είναι συχνά πολύ πιο αποτελεσματικά από τις συνηθισμένες θεωρίες που ισοδυναμούν με την επίδραση των ασύμμετρων κατανομών και κατανομών με βαριές ουρές (heavy-tailed). Αυτά τα διαγράμματα μπορεί να είναι ιδιαίτερα χρήσιμα όταν μια διεργασία ξεκινά και είναι επιθυμητό να εφαρμοστεί ένα δ.ε. πριν υπάρξουν αρκετά δεδομένα για τη σωστή εκτίμηση της διακύμανσης.

Ο αριθμός των δειγμάτων που συλλέγονται πριν από το πρώτο σήμα εκτός ελέγχου ενός διαγράμματος είναι μια τ.μ. που ονομάζεται μήκος ροής. Εάν η κατανομή του εντός ελέγχου μήκους ροής είναι ίδια για κάθε κατανομή πιθανότητας, το διάγραμμα ονομάζεται μη παραμετρικό (Chakraborti et al., 2001).

Η ανάπτυξη και η εφαρμογή των μη παραμετρικών δ.ε. ήταν μάλλον αργή στον έλεγχο των βιομηχανικών διεργασιών. Υπάρχουν πολλοί λόγοι για αυτό. Οι επαγγελματίες μερικές φορές πιστεύουν ότι το Κ.Ο.Θ. (Κεντρικό Οριακό Θεώρημα) θα εξασφαλίσει με κάποιο τρόπο την αναμενόμενη απόδοση του διαγράμματος. Το πιο σημαντικό πρόβλημα είναι ότι τα δ.ε. πολλές φορές χρειάζεται να εφαρμοσθούν σε μεμονωμένες παρατηρήσεις (Montgomery, 1997) και δεν είναι δυνατή η εφαρμογή του Κ.Ο.Θ. (επειδή το μέγεθος του δείγματος είναι ένα). Έχει αποδειχθεί ότι στην περίπτωση αυτή η κατανομή δεν είναι ανθεκτική στα πιο συνήθη διαγράμματα.

Άλλοι λόγοι για την έλλειψη ενδιαφέροντος αποτελούν η μη διαθεσιμότητα επαρκών υπολογιστικών δομών και η αντίληψη ότι πρέπει να θυσιαστεί η αποτελεσματικότητα όταν χρησιμοποιούμε αυτές τις απλές τεχνικές που συχνά βασίζονται στην καταμέτρηση και την κατάταξη των παρατηρήσεων. Βέβαια η μειωμένη αποτελεσματικότητα μπορεί να αντισταθμιστεί με περισσότερες παρατηρήσεις. Ο πρώτος λόγος δεν αποτελεί πλέον πρόβλημα στην εποχή των υπολογιστών και ο τελευταίος δεν είναι απαραίτητα ορθός όπως έχει τεκμηριωθεί στη βιβλιογραφία στατιστικών ελέγχων και εκτιμήσεων. Αξίζει να σημειωθεί ότι, ακόμα και όταν η υποκείμενη κατανομή είναι στην πραγματικότητα Κανονική, η αποτελεσματικότητα ορισμένων μη παραμετρικών μεθόδων όπως ο προσημικός έλεγχος του Wilcoxon σε σχέση με τις αντίστοιχες θεωρητικές μεθόδους που αφορούν την Κανονική κατανομή, όπως το t-test είναι υψηλή περίπου 0.955 (Gibbons and Chakraborti, 1992). Τέλος, σημειώνεται ότι μεγάλο μέρος της ανάπτυξης της μη παραμετρικής μεθοδολογίας έχει πραγματοποιηθεί στα πλαίσια της στατιστικής εκτίμησης και ελέγχων υποθέσεων ενώ δεν έχει καταβληθεί αρκετή προσπάθεια για την κατανόηση του προβλήματος του πρακτικού ΣΕΠ.

1.5 Φάση I και II μιας Παραγωγικής Διαδικασίας

Για να ελεγχθεί η παραγωγική διαδικασία χρησιμοποιώντας τα δ.ε. υπάρχουν δύο φάσεις, η Φάση I και η Φάση II.

Φάση I (Phase I) : Στη Φάση I γίνεται χρήση των αναδρομικών δ.ε. με σκοπό να διαπιστωθεί αν η διεργασία βρίσκεται εκτός ελέγχου σύμφωνα με παλαιότερα δείγματα που έχουν καταγραφεί. Με τα διαγράμματα Φάσης I, αφού η διεργασία βρίσκεται εντός ελέγχου δίνεται η δυνατότητα στον διαχειριστή της παραγωγικής διαδικασίας να υπολογίσει τα όρια ελέγχου καθώς και την κεντρική τιμή. Αυτά τα όρια ελέγχου και η κεντρική τιμή χρησιμοποιούνται με σκοπό τη μελλοντική παρακολούθηση της διεργασίας.

Φάση II (Phase II) : Στη Φάση II γίνεται χρήση των δ.ε. που κατασκευάστηκαν στη Φάση I για τον συνεχή έλεγχο της παραγωγής έτσι ώστε η διεργασία να μην βρεθεί εκτός ελέγχου. Εδώ, η παραγωγική διαδικασία ελέγχεται συνεχώς από τον υπεύθυνο ποιότητας και μπορεί να ανιχνεύει γρήγορα μια μεταβολή στη διεργασία. Στην περίπτωση αυτή λαμβάνονται μέτρα για την επαναφορά της διεργασίας εντός ελέγχου.

Είναι πολύ σημαντικό να ξεχωρίσουμε τη χρήση ενός δ.ε. σε ένα σύνολο παλαιότερων δεδομένων για να προσδιοριστεί κατά πόσον μια διεργασία υπήρξε ή όχι

σε στατιστικό έλεγχο (Φάση I) και τη χρήση της Φάσης II όπου τα δείγματα λαμβάνονται διαδοχικά με την πάροδο του χρόνου για την ανίχνευση αλλαγών από τη Φάση I (Αντζουλάκος, 2016).

1.6 Πληροφορίες για τα μη παραμετρικά δ.ε.

Υπάρχει αρκετή αναφορά στη βιβλιογραφία που αφορά μη παραμετρικά δ.ε.. Ο Arnold (1985) πρότεινε ένα σύστημα για τη χρήση οικονομικών παραμέτρων για να επιλέγονται οι παράμετροι ενός δ.ε. βάσει του προσημικού τεστ. Οι Saad and Reynolds (1979) χρησιμοποίησαν το στατιστικό προσημικό έλεγχο για να ελέγξουν τη συμμετρία γύρω από ένα σταθερό επίπεδο, ενώ οι Hackl and Ledolter (1991) περιέγραψαν μια διεργασία βασισμένη σε εκθετικά εξομαλυμένες τάξεις.

Ο Chakraborti (2001) βασιζόμενος σε να διάγραμμα τύπου Shewhart που προτάθηκε από τους Janacek and Meikle (1997), κατασκεύασε ένα μη παραμετρικά δ.ε το οποίο βασίζεται στη διάμεσο και τα όρια του υπολογίζονται από ένα εντός ελέγχου δείγμα. Υπάρχουν πρακτικά πλεονεκτήματα στη χρήση δ.ε. τα οποία είναι βασισμένα στη διάμεσο. Σε γενικές γραμμές, αυτά τα δ.ε. είναι χρήσιμα επειδή οι διάμεσοι είναι πολύ λιγότερο ευαίσθητοι σε εσφαλμένες μετρήσεις ενώ μπορούν επίσης να προσδιοριστούν και σε περιπτώσεις όπου περιλαμβάνουν ποιοτικές μετρήσεις.

Παρόλο που τα συμβατικά δ.ε. με βάση το μέσο είναι οικεία και κατανοητά, για να λειτουργούν σωστά, θα πρέπει να χρησιμοποιούνται μεταβλητές που κατανέμονται Κανονικά. Έχει προταθεί μια εναλλακτική μη παραμετρική λύση, η οποία είναι πολύ αξιόπιστη όταν χρησιμοποιούνται σε μη Κανονικές κατανομές. Ο Wheeler (1995) επεσήμανε, ότι «οι υποθέσεις που χρησιμοποιούνται για τη μαθηματική επεξεργασία αποτελούν περιορισμούς οι οποίοι συνήθως δεν ισχύουν στην πράξη» ενώ παρόμοια θέση παίρνουν και οι Hoerl and Palm (1992).

Όσον αφορά την ανθεκτικότητα των παραδοσιακών δ.ε., οι Amin et al. (1995) διαπίστωσαν έντονη διαφορά στις τιμές του ARL_0 του διαγράμματος \bar{X} του Shewhart για διάφορες κατανομές. Υποθέτοντας μια γνωστή τυπική απόκλιση και ένα δείγμα μέγεθος 10, έχουν βρεθεί οι ακριβείς τιμές των ARL_0 του μονόπλευρου διαγράμματος \bar{X} του Shewhart με 3σ , οι τιμές αυτές παρουσιάζονται στο παρακάτω πίνακα.

Πίνακας 1.1: ARL_0 για διάφορες κατανομές διαγράμματος \bar{X} Shewhart.

Κατανομή	ARL_0
Ομοιόμορφη	1068.7
Κανονική	740.8
Διπλή Εκθετική	441.9
Cauchy	11.7

Αυτό μας δείχνει ο Πίνακας 1.1 είναι ότι για κατανομές με βαριές ουρές, εμφανίζονται εσφαλμένοι συναγερμοί πολύ πιο συχνά από το αναμενόμενο όταν η διεργασία λειτουργεί υπό έλεγχο. Παράλληλα, οι Yourstone and Zimmer (1992) εξετάσανε την επίδραση των παρεκκλίσεων από την Κανονικότητα και κατέληξαν στο συμπέρασμα ότι η επίδραση της λοξότητας και της κύρτωσης είναι σημαντική για το ARL_0 ενός διαγράμματος \bar{X} τύπου Shewhart, και ανέπτυξαν τα ασύμμετρα διάγραμμα τύπου Shewhart για τα κέντρα (centers) για μη Κανονικές κατανομές. Οι Schilling and Nelson (1976) μελέτησαν την επίδραση της μη Κανονικότητας στα όρια ελέγχου του διαγράμματος \bar{X} όταν η τυπική απόκλιση της διεργασίας είναι γνωστή. Τέλος, αξίζει να σημειωθεί ότι, ο Burr (1967) παρείχε τους πίνακες των προσαρμοσμένων τιμών των σταθερών που χρησιμοποιούνται στον υπολογισμό των ορίων ελέγχου του διαγράμματος \bar{X} για 28 διαφορετικές μη Κανονικές κατανομές.

ΚΕΦΑΛΑΙΟ 2

Μεθοδολογίες στα μη παραμετρικά διαγράμματα ελέγχου

2.1 Μέτρα απόδοσης για τη σύγκριση των δ.ε.

Η απόδοση ενός δ.ε. τυπικά μετράται με το μέσο μήκος ροής (ARL) και την τυπική απόκλιση του μήκους ροής ($SDRL$). Το ARL είναι ο μέσος αριθμός δειγμάτων που απεικονίζεται στο διάγραμμα πριν ανιχνευθεί το πρώτο σήμα εκτός ελέγχου, ενώ το $SDRL$ μετρά τη διασπορά της κατανομής του μήκους ροής όταν μια διεργασία είναι εκτός ελέγχου. Τα ARL και $SDRL$ είναι επιθυμητό να παίρνουν μικρές τιμές.

Το ARL διακρίνεται σε εντός ελέγχου μέσο μήκος ροής (ARL_0) και σε εκτός ελέγχου μέσο μήκος ροής (ARL_1).

Το ARL_0 , αντιπροσωπεύει τον μέσο αριθμό δειγμάτων που εμπίπτουν στα όρια ελέγχου πριν εμφανιστεί μια κατάσταση εκτός ελέγχου. Το ARL_1 , αντιπροσωπεύει τον μέσο αριθμό δειγμάτων που μετριέται μέχρι το πρώτο δείγμα να βρεθεί εκτός των ορίων ελέγχου όταν η διεργασία βγει εκτός ελέγχου. Για ένα δεδομένο ARL_0 το διάγραμμα με το μικρότερο ARL_1 καλείται το βέλτιστο διάγραμμα.

Για να συγκρίνουμε την αποτελεσματικότητα δύο δ.ε. για την ανίχνευση μετατοπίσεων σε μια παράμετρο της διεργασίας, προσαρμόζουμε τα όρια ελέγχου τους έτσι ώστε οι τιμές των ARL_0 να γίνουν ίδιες και στη συνέχεια να συγκριθούν με το ARL_1 σε διάφορες μετατοπίσεις στην παράμετρο της διεργασίας.

Για να μετρήσουμε την απόδοση ενός δ.ε. κατά τη Φάση I, μπορούμε να χρησιμοποιήσουμε την πιθανότητα να προκληθεί εκτός ελέγχου σήμα από μια διεργασία η οποία είναι στην πραγματικότητα εκτός ελέγχου. Το FAR είναι η πιθανότητα ένα δ.ε. να εκπέμψει εκτός ελέγχου σήμα για την διεργασία ενώ η διεργασία είναι εντός ελέγχου. Αυτή η προσέγγιση σχετίζεται με την «ανάλυση των μέσων» που έγινε από τον Wheeler (1995) και τον Ryan (2000). Το ARL όμως είναι η παράμετρος που χρησιμοποιείται συχνότερα, αν και η κατανομή του μήκους ροής είναι συχνά ασύμμετρη προς τα δεξιά. Ως εκ τούτου, μερικές φορές λαμβάνονται υπόψη και άλλα μέτρα όπως η διάμεσος ή άλλα τεταρτημόρια (Woodall, 2000).

Ένα ακόμη μέτρο για ανίχνευση της μεταβολής στη διεργασία είναι το ποσοστό σωστής ταξινόμησης για την ανίχνευση διαφορετικών μετατοπίσεων του μέσου και της διακύμανσης. Όσο μεγαλύτερο είναι το ποσοστό ορθών ταξινομήσεων για μια συγκεκριμένη μετατόπιση, τόσο πιο αποτελεσματικό είναι ένα διάγραμμα. Ωστόσο στην παρούσα εργασία δεν θα αναφερθούμε σε αυτό το μέτρο.

2.1.1 Δ.ε. βασισμένο στη συνάρτηση προτεραιότητας

Σε αυτή την ενότητα θα γίνει μια ανάλυση της μεθοδολογίας κατασκευής ενός μη παραμετρικού δ.ε. που βασίζεται στη συνάρτηση προτεραιότητας (precedence statistic) (βλ. Chakraborti (2004)).

Αρχικά, υποθέτουμε ότι έχουμε ένα τ.δ. $X = (X_1, X_2, \dots, X_m)$, από μια εντός ελέγχου διεργασία με άγνωστη συνεχή α.σ.κ. $F_X(X) = F(x - \theta)$ (θ είναι η παράμετρος θέσης). Τα X διατάσσονται σε αύξουσα σειρά ($X_{1:m} \leq X_{2:m} \leq \dots \leq X_{m:m}$) και υπολογίζονται δύο σ.σ., $X_{\alpha:m}$ και $X_{b:m}$ (για $1 \leq \alpha < b \leq m$). Στη συνέχεια, για να κατασκευάσουμε ένα δίπλευρο διάγραμμα ορίζουμε τα όρια ελέγχου με $LCL = X_{\alpha:m}$ και $UCL = X_{b:m}$. Αφού βρήκαμε τα όρια ελέγχου μπορούμε να τα χρησιμοποιήσουμε σε μελλοντικά δείγματα που λαμβάνονται ανεξάρτητα το ένα από το άλλο καθώς και από το αρχικό δείγμα ώστε ελέγξουμε εάν η διεργασία είναι εντός ελέγχου.

Για να γίνει αυτό πρέπει να υπολογίσουμε τη σ.σ. $Y_{j:n}$ από μια α.σ.κ. $G(x) = F(x - \theta_h)$, $h = 1, 2, \dots$, (θ_h η παράμετρος θέσης της h ομάδας από το δείγμα μεγέθους n που ελέγχεται). Μετά από τη συλλογή κάθε δείγματος, υπολογίζεται η $Y_{j:n}$, και συγκρίνεται με τα όρια ελέγχου. Αν η σ.σ. $Y_{j:n}$ βρεθεί εντός των LCL και UCL τότε η διεργασία βρίσκεται εντός ελέγχου, διαφορετικά η διεργασία βρίσκεται εκτός ελέγχου. Όταν μία διεργασία βρίσκεται εντός ελέγχου σημαίνει ότι η κατανομή του πληθυσμού που ελέγχεται είναι ίδια με αυτή του δείγματος αναφοράς.

Έστω τώρα, μια σ.σ. W_j που αντιπροσωπεύει τον αριθμό των παρατηρήσεων X που προηγούνται (δεν είναι μεγαλύτερες) της $Y_{j:n}$. Η σ.σ. αυτή ονομάζεται συνάρτηση προτεραιότητας.

Όταν η διεργασία βρίσκεται υπό έλεγχο, η ακριβής κατανομή πιθανότητας της σ.σ. προτεραιότητας μπορεί να βρεθεί με μαθηματικές-στατιστικές τεχνικές είτε συνδυαστικά. Η κατανομή πιθανότητας της σ.σ. προτεραιότητας δίνεται από το παρακάτω τύπο,

$$P(W_j = w) = \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{m}}, \quad w = 0, 1, \dots, m. \quad (2.1.1)$$

Η πιθανότητα το διάγραμμα προτεραιότητας να μην εκπέμψει σήμα ότι η διεργασία βρίσκεται εκτός ελέγχου εκφράζεται μέσω της σ.σ. W_j . Αυτό ισχύει διότι, οι τιμές των $Y_{j:n}$ βρίσκονται μεταξύ των $X_{a:m}$ και $X_{b:m}$ (για $1 \leq a < b \leq m$) και η σ.σ. W_j βρίσκεται μεταξύ a και $b - 1$. Έτσι, γενικά,

$$p = p(m, n, j; F, \theta, \theta_h) = P(X_{(a:m)} \leq Y_{j:n} \leq X_{(b:m)}) = P(a \leq W \leq b - 1) \quad (2.1.2)$$

Πρέπει να σημειωθεί ότι το συγκεκριμένο δ.ε. δεν εκπέμπει εκτός ελέγχου σήμα όταν τα $Y_{j:n}$ είναι είτε αυστηρά πάνω από το $X_{b:m}$ είτε είναι αυστηρά κάτω από το $X_{a:m}$.

Για την εφαρμογή των δ.ε. προτεραιότητας, έχουν αναφερθεί δύο προτάσεις. Η πρώτη είναι να χρησιμοποιηθεί μια καθορισμένη τιμή για το FAR . Οι Janacek and Meikle (1996) μεταξύ πολλών άλλων, το ανέπτυξαν, θέτοντας ως p_0 την εντός ελέγχου τιμή του p , έτσι ώστε το $1 - p_0$ να αντιπροσωπεύει το FAR . Έτσι,

$$p_0 = p(m, n, j; F, \theta, \theta) = P(a \leq W \leq b - 1).$$

Χρησιμοποιώντας τις εξισώσεις (2.1.1) και (2.1.2) μπορούν να βρεθούν τα a και b έτσι ώστε,

$$\sum_{w=a}^{b-1} \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{m}} \geq p_0 \quad (2.1.3)$$

Η ανισότητα (2.1.3) χρησιμοποιείται για να προβλέψει το διαχωρισμό του W_j . Τα όρια ελέγχου υπολογίζονται θέτοντας το p_0 ή το $1 - p_0$ αντιστοίχως σε κάποια επιθυμητή υψηλή ή χαμηλή τιμή. Βρίσκοντας το p_0 ή το $1 - p_0$, στη συνέχεια μπορούν να βρεθούν και τα $LCL = X_{a:m}$ και $UCL = X_{b:m}$.

Η δεύτερη πρόταση είναι να χρησιμοποιήσουμε ένα καθορισμένο ARL_0 , όπως το 370 ή το 500, προκειμένου να υπολογιστούν τα a και b .

2.1.2 Δ.ε. βασισμένο στη διάμεσο

Η χρήση του δ.ε. που βασίζεται στη διάμεσο έχει χρησιμοποιηθεί περισσότερο ως παράδειγμα και όχι τόσο στην πράξη, ωστόσο είναι ένα αρκετά σημαντικό δ.ε.. Μπορεί να χρησιμοποιηθεί και για άλλα δειγματικά τεταρτημόρια ή εκατοστημόρια πέρα της διαμέσου.

Για λόγους απλότητας το n είναι περιττός αριθμός, ας πούμε $2s + 1$ και η διάμεσος του δείγματος $Y_{j:n}$, ορίζεται με μοναδικό τρόπο με $j = s + 1$, με αποτέλεσμα η σ.σ. W_{s+1} να είναι η διάμεσος του Mathisen (1943).

Για περικό n η κατανομή του W_{s+1} δίνεται από την εξίσωση (2.1.1), είναι συμμετρική και μια λογική επιλογή για το b είναι το $m - a + 1$. Έτσι, χρησιμοποιώντας την ανισότητα (2.1.3) η σταθερά $a (\geq 1)$ μπορεί να προσδιοριστεί ως ο μεγαλύτερος ακέραιος έτσι ώστε,

$$\sum_{w=a}^{m-a} \frac{\binom{s+w}{w} \binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \geq p_0 \quad (2.1.4)$$

αφού βρεθεί η σταθερά a , αντικαθιστάμε το b με $m - a + 1$.

Χρησιμοποιώντας τη συμμετρία, η ανισότητα (2.1.4) μπορεί να ξαναγραφεί ως $P(0 \leq W_j \leq a - 1) \leq (1 - p_0)/2$. Εδώ πρέπει να σημειωθεί ότι, έχει γραφτεί ένα σημειωματάριο στο πρόγραμμα Mathematica (Wolfram, 1996) ώστε να λυθεί η εξίσωση (2.1.4) και χρησιμοποιώντας αυτό το πρόγραμμα στον Πίνακα 2.1 δίνονται οι τιμές των a και b για $m = 50, 100, 500, 1000$, $n = 5, 11, 25$ και $1 - p_0 = 0.01, 0.005, 0.0027$. Στο Πίνακα 2.1, δίνονται επίσης οι τιμές του ακριβούς FAR που επιτυγχάνεται σε κάθε ουρά, $P(0 \leq W_j \leq a - 1)$ και $P(m - a + 1 \leq W_j \leq m)$, μαζί με την πραγματική τιμή του ARL_0 . Για τον υπολογισμό του FAR απλά προσθέτουμε αυτές τις δύο πιθανότητες.

2.1.3 Μονόπλευρα δ.ε.

Σε ορισμένες εφαρμογές όπου μας ενδιαφέρουν μόνο υπερβολικά υψηλές ή χαμηλές τιμές, είναι πιο επιθυμητή η χρήση ενός μονόπλευρου δ.ε.. Για παράδειγμα, υποθέτουμε ότι υπάρχει μια κατάσταση όπου η μεταβλητή που μας ενδιαφέρει είναι η αντοχή στη δύναμη που ασκείται σε κάποιο υλικό. Μια μετατόπιση σε υψηλότερη δύναμη θα μπορούσε να προκαλέσει αστοχία υλικού, αλλά η μετατόπιση σε χαμηλότερη δύναμη μπορεί να μην αποτελεί πρόβλημα. Εάν για παράδειγμα, ανιχνευθούν υψηλότερες τιμές (ανεξάρτητα από το αν η παράμετρος έχει μετακινηθεί προς τα δεξιά), μπορούμε να χρησιμοποιήσουμε το $UCL = X_{b:m}$. Έτσι, για το διάγραμμα της διαμέσου και για συγκεκριμένο $FAR = 0.0027$, το b μπορεί να βρεθεί διπλασιάζοντας την πιθανότητα ουράς από 0.00135 έως 0.0027 και χρησιμοποιώντας τον Πίνακα 2.1.

Πίνακας 2.1: Δ.ε. προτεραιότητας για ένα καθορισμένο *FAR*. (Πηγή: Chakraborti, Van Der Laan και Van De Wiel (2004))

$1 - p_0$	n	j	Αποτελέσματα για συνεχείς τιμές για το m			
			50	100	500	1000
0.00100	5	3	(3,48)	(7,94)	(40,461)	(82,919)
			0.0036	0.00432	0.00477	0.00499
			635.7	213.8	114.5	103.5
	11	6	(7,44)	(15,86)	(83,418)	(167,834)
			0.00465	0.00428	0.00494	0.0497
			642.2	245.0	113.3	108.4
	25	13	(10,41)	(23,78)	(123,375)	(248,743)
			0.00305	0.00402	0.00473	0.00497
			109990.0	510.8	128.3	109.8
0.005	5	3	(2,49)	(5,96)	(31,470)	(64,937)
			0.0015	0.00176	0.00233	0.00246
			5671.0	678.4	242.3	215.1
	11	6	(5,46)	(13,88)	(72,429)	(146,855)
			0.00125	0.00225	0.0024	0.00244
			9503.0	574.5	240.9	219.8
	25	13	(9,42)	(21,80)	(118,383)	(239,762)
			0.00155	0.002	0.00243	0.00245
			44750.0	1488.0	261.0	227.5
0.0027	5	3	(1,50)	(4,97)	(24,476)	(51,950)
			0.0004	0.00102	0.00127	0.00129
			∞	1550.0	460.2	419.5
	11	6	(5,46)	(11,90)	(64,437)	(130,871)
			0.00125	0.00106	0.00131	0.00131
			9503.0	1630.0	456.1	409.8
	25	13	(8,43)	(19,82)	(110,391)	(224,777)
			0.00075	0.00092	0.00127	0.00133
			173700.0	5183.0	526.2	430.2

2.2 Διαγράμματα για τον Έλεγχο της μεταβλητότητας

Σε αυτή την ενότητα θα ασχοληθούμε με τα δ.ε. για τη παρακολούθηση της παραμέτρου της μεταβλητότητας.

Όπως αναφέρει η Das (2011) στην αντίστοιχη δημοσίευση της για τα δ.ε. για τον έλεγχο της μεταβλητότητας που αφορά μη Κανονικές διεργασίες, ο M.O.A. Abu-Shawiesh (2008) ανέπτυξε ένα ισχυρό δ.ε. για την παρακολούθηση της μεταβλητότητας με βάση το δείγμα *MAD* (Μέση απόλυτη απόκλιση από το διάμεσο). Το *MAD*, εισήχθη για πρώτη φορά από τον Hampel (1974) και είναι πολύ απλό και εύκολο να υπολογιστεί και έχει το καλύτερο δυνατό σημείο καμπής (50% διπλάσιο από το εύρος μεταξύ των τεταρτημορίων). Παρά όλα τα πλεονεκτήματα του, έχει επίσης κάποια μειονεκτήματα. Έτσι, πρώτον η αποτελεσματικότητά του σε Gaussian κατανομές είναι πολύ χαμηλή, ενώ η ασυμπτωτική απόδοση της διαμέσου του παραμένει 64%, το *MAD* είναι μόνο 37% αποδοτικό. Δεύτερον, αποτελεί μια συμμετρική προσέγγιση για τη διασπορά, επειδή κάποιος αρχικά εκτιμά τη διάμεσο και στη συνέχεια αποδίδει την ίδια σημασία στις θετικές και αρνητικές αποκλίσεις από αυτήν.

Το διάγραμμα *S* χρησιμοποιείται για τον έλεγχο της μεταβλητότητας και έτσι, η δειγματική τυπική απόκλιση *S* είναι το πιο συχνά χρησιμοποιούμενο μέτρο. Το κύριο πλεονέκτημα της είναι ότι είναι εύκολο να υπολογιστεί, είναι περιεκτική και λαμβάνει υπόψη όλες τις παρατηρήσεις στον υπολογισμό του, ενώ το κύριο μειονέκτημα είναι ότι δεν είναι ανθεκτική ακόμη και σε μικρές αποκλίσεις από την Κανονικότητα και μπορεί εύκολα να επηρεαστεί από ακραίες τιμές.

Ο εκτιμητής δίνεται από το \bar{S}/c_4 , ο \bar{S} αντιπροσωπεύει το μέσο των δειγματικών τυπικών αποκλίσεων από k δείγματα που λαμβάνονται από μια σταθερή διεργασία και οι τιμές του c_4 δίνονται στον Πίνακα 1 της Das στο αντίστοιχο άρθρο της. Τα όρια ελέγχου ενός διαγράμματος *S* είναι:

$$\begin{aligned}UCL &= \bar{S} + 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2}, \\CL &= \bar{S}, \\LCL &= \bar{S} - 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2}.\end{aligned}$$

Στόχος μας είναι να λάβουμε μια ανθεκτική εκτίμηση για το *S*.

Οι Rousseeuw and Croux (1993) πρότειναν δύο εναλλακτικούς εκτιμητές εκτός του *MAD*. Στη συνέχεια, στους δύο αυτούς εκτιμητικές προστέθηκαν οι συντελεστές

διόρθωσης c_n και d_n καθώς αυτοί οι συντελεστές κάνουν αυτούς τους εκτιμητές αμερόληπτους ακόμα και για μικρό μέγεθος δείγματος, οι συντελεστές αυτοί δόθηκαν από τους ίδιους το 1992. Οι δύο αμερόληπτοι ανθεκτικοί εκτιμητές για την τυπική απόκλιση είναι:

$$S_n = c_n 1.1926 \text{Med}_i\{\text{Med}_j|X_i - X_j\},$$

$$Q_n = d_n 2.2219 [Q(0.25)\text{της απόστασης } \{|X_i - X_j|, i < j\}].$$

όπου $\text{Med}_i\{\text{Med}_j|X_i - X_j\}$ σημαίνει ότι για κάθε i , η διάμεσος του $\{|X_i - X_j|; j = 1, \dots, n\}$ υπολογίζεται σύμφωνα με n αριθμούς, ο διάμεσος του οποίου δίνει την εκτίμηση του S_n . Οι τιμές 1.1926 και 2.2219 προτάθηκαν από τους συγγραφείς.

Σε αντίθεση με το MAD , ο S_n δεν χρειάζεται εκτίμηση θέσης στα δεδομένα. Εξετάζει την απόσταση μεταξύ των παρατηρήσεων η οποία ισχύει και για ασύμμετρες κατανομές. Παράλληλα, έδειξαν μέσω προσομοίωσης ότι ο S_n συμπεριφέρεται καλύτερα.

Εξέτασαν επίσης τις διαφορετικές ιδιότητες του εκτιμητή Q_n . Ο Q_n όπως και ο S_n , είναι επίσης μοναδικά καθορισμένος και κατάλληλος για ασύμμετρες κατανομές. Παράλληλα, χρησιμοποιώντας προσομοίωση απέδειξαν ότι ο Q_n είναι πολύ πιο αποτελεσματικός από τους άλλους δύο. Το αποτέλεσμα της προσομοίωσης επιβεβαίωσε την αμεροληψία των εκτιμητών για πεπερασμένα δείγματα.

Παράλληλα, υπογράμμισαν ότι ο S_n , είναι πιο αποτελεσματικός από τον Q_n για μικρά δείγματα. Ενώ, ένα ακόμη πλεονέκτημα τους είναι ότι δεν προϋποθέτουν συμμετρική κατανομή μοντέλου, αλλά μπορούν να θεωρηθούν ως μη παραμετρικά μέτρα εκτίμησης.

2.2.1 Οι εκτιμητές S_n και Q_n

Τα όρια ελέγχου του διαγράμματος με τον εκτιμητή S_n είναι :

$$UCL = c_4 \hat{\sigma} + 3 \hat{\sigma} \sqrt{1 - c_4^2} = c_4 c_n \bar{S}_n + 3 c_n \bar{S}_n \hat{\sigma} \sqrt{1 - c_4^2},$$

$$CL = c_4 \hat{\sigma} = c_4 c_n \bar{S}_n,$$

$$LCL = c_4 \hat{\sigma} - 3 \hat{\sigma} \sqrt{1 - c_4^2} = c_4 c_n \bar{S}_n - 3 c_n \bar{S}_n \hat{\sigma} \sqrt{1 - c_4^2},$$

όπου \bar{S}_n είναι ο μέσος του S_n .

Τα όρια ελέγχου του διαγράμματος με τον εκτιμητή Q_n είναι :

$$UCL = c_4\hat{\sigma} + 3\hat{\sigma}\sqrt{1 - c_4^2} = c_4d_n\overline{Q_n} + 3d_n\overline{Q_n}\hat{\sigma}\sqrt{1 - c_4^2},$$

$$CL = c_4\hat{\sigma} = c_4d_n\overline{Q_n},$$

$$LCL = c_4\hat{\sigma} - 3\hat{\sigma}\sqrt{1 - c_4^2} = c_4d_n\overline{Q_n} - 3d_n\overline{Q_n}\hat{\sigma}\sqrt{1 - c_4^2},$$

όπου $\overline{Q_n}$ είναι ο μέσος του Q_n .

Οι πίνακες με τις τιμές των c_n και d_n έχουν δοθεί από την Das (2011).

2.2.2 Σύγκριση των διαγραμμάτων S_n , Q_n και S

Οι Langenberg and Iglewicz (1986) τόνισαν το γεγονός ότι οι διεργασίες που παράγονται από κατανομές με βαρύτερες ουρές από την Κανονική παρουσιάζουν ιδιαίτερο ενδιαφέρον, τείνουν να έχουν περισσότερο από τον αναμενόμενο αριθμό σημείων εκτός των ορίων ελέγχου. Η απόδοση των παραπάνω διαγραμμάτων συγκρίνεται με εκείνη του διαγράμματος S τύπου Shewhart, για διαφορετικές κατανομές χρησιμοποιώντας ίσα μεγέθη δείγματος. Για τη σύγκριση της απόδοσης των τριών δ.ε. χρησιμοποιήθηκαν οι τιμές των ARL_0 και ARL_1 .

Πίνακας 2.2: ARL_0 για τα τρία δ.ε. για διάφορες κατανομές. (Πηγή: Das (2011))

	Κατανομή			
	Κανονική	Λογιστική	Διπλή Εκθετική	Cauchy
Διάγραμμα S τύπου Shewhart	370.8	373.5	377.2	365.9
Διάγραμμα S_n	370.6	370.3	370.4	370.5
Διάγραμμα Q_n	370.4	370.6	370.0	370.1

Πίνακας 2.3: ARL_1 για τα τρία δ.ε. για διάφορες κατανομές. (Πηγή: Das (2011))

		Κατανομή				
		Μετατόπιση λ	Κανονική	Λογιστική	Διπλή Εκθετική	Cauchy
Διάγραμμα S τύπου Shewhart	1		370.8	373.5	377.2	365.9
	1.5		13.5	10.6	8.2	36.6
	2		3.2	2	1.8	32.5
	2.5		1.5	1.2	1.1	1.8
Διάγραμμα S_n	1		370.6	370.3	370.4	370.5
	1.5		13.8	2.5	2.7	3.1
	2		3.1	1.5	1.8	2.2
	2.5		1.6	1	1.2	1.5
Διάγραμμα Q_n	1		370.4	370.6	370	370.1
	1.5		12.9	2.8	2.9	3.5
	2		3.3	1.3	1.5	2.9
	2.5		1.5	1.1	1.2	1.6

Από τον Πίνακα 2.3, φαίνεται πως η απόδοση των προτεινόμενων δ.ε. της ενότητας δεν είναι σε καμία περίπτωση χειρότερη από εκείνη του υπάρχοντος διαγράμματος S τύπου Shewhart σύμφωνα με το ARL_1 .

ΚΕΦΑΛΑΙΟ 3

Μη παραμετρικά διαγράμματα ελέγχου που βασίζονται στην προσημική συνάρτηση

Σε αυτό το κεφάλαιο, θα γίνει μια σύντομη παρουσίαση της μεθοδολογίας και σύγκριση των δ.ε. που χρησιμοποιούν τη σ.σ. προσημικού ελέγχου στη θέση του δειγματικού μέσου για τη ανίχνευση μιας μετατόπισης από το κέντρο της κατανομής των παρατηρήσεων.

Οι Bakir and Reynolds (1979) και ο Bakir (1977) ανέπτυξαν ένα μη παραμετρικό δ.ε. τύπου *CUSUM* χρησιμοποιώντας τη στατιστική προσημική-βαθμολογική συνάρτηση (συνάρτηση *SR*) του Wilcoxon (Wilcoxon signed-rank statistic). Η κατάταξη έγινε εντός ομάδων όπου τα δείγματα τους πάρθηκαν είτε σε κάθε σημείο είτε σε τεχνητά καθορισμένες ομάδες. Στη συνέχεια, ο Park (1985) στη μελέτη του, πρότεινε τη χρήση μη παραμετρικών δ.ε. όπως το διάγραμμα *SN* τύπου Shewhart και το προσημικό-βαθμολογικό διάγραμμα (διάγραμμα *SR*) για ασύμμετρες κατανομές.

Τα κύρια χαρακτηριστικά των δ.ε. που βασίζονται στη σ.σ. *SN* είναι τα ίδια με τα χαρακτηριστικά των μη παραμετρικών. Αυτά τα διαγράμματα βασίζονται σε πρόσημα που υπολογίζονται από τα δείγματα. Στην περίπτωση των δ.ε. που αφορούν τη θέση, υπολογίζεται η διάμεσος και δίνεται ένα πρόσημο (+) όταν μια παρατήρηση έχει τιμή που υπερβαίνει την εντός ελέγχου τιμή της διαμέσου. Στη περίπτωση των δ.ε. για τη μεταβλητότητα, στις παρατηρήσεις δίνεται ένα πρόσημο (+) όταν πέφτει κάτω από το πρώτο ή πάνω από το τρίτο τεταρτημόριο.

3.1 Μη Παραμετρικά δ.ε. τύπου Shewhart που βασίζονται στην συνάρτηση *SN*

Οι Amin et al. (2007) παρουσίασαν και ανέπτυξαν το διάγραμμα που βασίζεται στη σ.σ. *SN* (διάγραμμα *SN*), με σκοπό την ανίχνευση αυξήσεων της μεταβλητότητας της διεργασίας. Είναι εύκολο στην κατανόηση και τη χρήση του και αποφεύγει τα μειονεκτήματα των δ.ε. που βασίζονται στο δειγματικό μέσο.

3.1.1 Δ.ε. για την παρακολούθηση της κεντρικής τιμής

Τα διαγράμματα SN προϋποθέτουν τη σύγκριση κάθε παρατήρησης με την τιμή στόχο μ_0 και στη συνέχεια την καταγραφή για κάθε δείγμα όταν το αποτέλεσμα βρεθεί πάνω ή κάτω από το μ_0 (μ_0 θεωρείται ότι είναι η διάμεσος). Για ασύμμετρες κατανομές, το ο προσημικός έλεγχος είναι στην πραγματικότητα ένας έλεγχος αλλαγής στη διάμεσο των παρατηρήσεων και είναι πολύ εύκολο στην εφαρμογή.

Έστω,

$$SN_i = \sum_{j=1}^n \text{sign}(X_{ij} - \mu_0), \quad i = 1, 2, \dots \quad (3.1.1)$$

όπου

$$\text{sign}(t) = \begin{cases} 1, & \text{αν } t > 0 \\ 0, & \text{αν } t = 0 \\ -1, & \text{αν } t < 0. \end{cases}$$

Έτσι, η SN_i είναι ακριβώς η διαφορά μεταξύ του αριθμού των παρατηρήσεων πάνω και κάτω από το μ_0 στο i δείγμα. Χρησιμοποιώντας λοιπόν τη συνάρτηση SN_i , έχουμε το πλεονέκτημα να κρατήσουμε τα όρια ελέγχου συμμετρικά γύρω από το 0. Το ARL για το δίπλευρο δ.ε. είναι ίσο με $L(\mu) = 1/P(|SN_i| \geq a_2)$, $a_2 > 0$ είναι μια σταθερά, όπου η πιθανότητα μπορεί να υπολογισθεί για οποιαδήποτε κατανομή στην οποία μπορεί να εκτιμηθεί το $p = P\{X_{ij} > \mu_0\}$. Στη συνέχεια, χρησιμοποιώντας μόνο το ένα από τα όρια ελέγχου μπορεί να δημιουργηθεί ένα μονόπλευρο δ.ε.. Το ARL για την περίπτωση του μονόπλευρου διαγράμματος είναι ίσο με $L^+(\mu) = 1/P(SN_i \geq a_2)$ και το ARL_0 θα είναι το ίδιο για όλες τις κατανομές με διάμεσο μ_0 .

Δεδομένου ότι δεν χρησιμοποιούνται τα μεγέθη των παρατηρήσεων, η σ.σ. SN_i είναι σχετικά ανεπαρκής σε σύγκριση με το \bar{X}_i όταν οι παρατηρήσεις πλησιάζουν την Κανονική κατανομή. Από την άλλη πλευρά, αυτό το πλεονέκτημα της απόδοσης μπορεί να αντιστραφεί για κατανομές με πολύ βαριές ουρές.

Οι μεγαλύτερες δυνατές τιμές του ARL_0 συμμετρικών μονόπλευρων και δίπλευρων δ.ε. είναι 2^n και 2^{n-1} , αντίστοιχα όταν $p = 1/2$. Εάν το δείγμα που έχουμε είναι μέτριου μεγέθους, μπορεί να είναι δύσκολο να επιτευχθεί ακόμη και προσεγγιστικά με μια καθορισμένη τιμή $L^+(\mu_0)$. Στην πραγματικότητα, η μεγαλύτερη δυνατή τιμή του $L^+(\mu_0)$ όταν $n = 10$ είναι 1024, η οποία επιτυγχάνεται όταν $a_2 = 10$.

Η αποδοτικότητα των δ.ε. τύπου Shewhart για την ανίχνευση των μικρών μετατοπίσεων στην παράμετρο που μας ενδιαφέρει μπορεί συνήθως να βελτιωθεί με την εισαγωγή των προειδοποιητικών ορίων μέσα στα όρια ελέγχου. Για τα διαγράμματα \bar{X} τύπου Shewhart τα προειδοποιητικά όρια είναι $\mu_0 \pm w_1$ (w_1 είναι μια σταθερά που ισχύει $0 \leq w_1 < a_1$), όπου a_1 είναι μια σταθερά που συχνά είναι ίση με $3\sigma/\sqrt{n}$. Ένα σήμα εκπέμπεται εάν r διαδοχικά σημεία βρεθούν μεταξύ $\mu_0 + w_1$ και $\mu_0 + a_1$ ή αν r διαδοχικά σημεία βρεθούν μεταξύ $\mu_0 - a_1$ και $\mu_0 - w_1$ ή αν οποιοδήποτε σημείο δεν εμπίπτει στα όρια ελέγχου.

Το ARL ενός μονόπλευρου διαγράμματος μόνο με θετικά προειδοποιητικά όρια ελέγχου μπορούν να καθορισθεί σύμφωνα με τον Page (1962) από τον τύπο,

$$L^+(\mu) = \frac{1-p^r}{1-p_1-p_0(1-p_1^r)} \quad (3.1.2)$$

όπου $p_0 = P\{SN_i < w_2 | \mu\}$ και $p_1 = P\{w_2 \leq SN_i < a_2 | \mu\}$ (w_2 είναι μια σταθερά που ισχύει $0 \leq w_2 < a_2$). Το ARL του δίπλευρου διαγράμματος είναι ίσο με,

$$L(\mu) = \frac{L^-(\mu)L^+(\mu)}{L^-(\mu)+L^+(\mu)},$$

$L^-(\mu)$ είναι το ARL ενός διαγράμματος με αρνητικά προειδοποιητικά όρια ελέγχου μόνο. Το $L^-(\mu)$ μπορεί να υπολογισθεί με την ισότητα (3.1.2) αντικαθιστώντας $p_0 = P\{SN_i > -w_2 | \mu\}$ και $p_1 = P\{-a_2 < SN_i \leq -w_2 | \mu\}$.

Πίνακας 3.1: $L^+(\mu)$ για διαγράμματα SN τύπου Shewhart όταν $n = 10$. (Πηγή: Amin, Reynolds Jr. και Bakir (2007))

w_2	$a_2 = 8$				$a_2 = 10$			
	0	2	4	6	2	4	6	8
2	4.1	9.2	30.2	79.4	9.6	38.6	269.2	933.7
3	7.9	23.0	70.1	92.4	27.8	193.6	890.3	1023.0
4	13.5	43.6	88.4	93.1	73.0	593.7	1015.8	1023.0
5	21.2	66.9	92.3	93.1	175.4	911.2	1023.6	1024.0
6	31.0	81.5	93.0	93.1	364.4	1002.0	1024.0	1024.0
7	42.3	88.5	93.1	93.1	609.8	1020.3	1024.0	1024.0

Ο Πίνακας 3.1 μας δίνει τιμές για το $L^+(\mu)$ για διάφορες τιμές των a_2, w_2 και r . Οι τιμές του ARL για άλλες τιμές n μπορούν να υπολογισθούν με όμοιο τρόπο. Αν w_2 βρίσκεται κοντά στο a_2 και r είναι αρκετά μεγάλο, η εισαγωγή των προειδοποιητικών ορίων θα έχει μια μικρή επίδραση στο $L^+(\mu_0)$ αλλά μπορεί να μειωθεί σημαντικά για μικρές μετατοπίσεις του μ .

Πίνακας 3.2: $L^+(\mu)$ σε συνάρτηση με το p για διαγράμματα SN όταν $n = 10$. (Πηγή: Amin, Reynolds Jr. και Bakir (2007))

p	0.50	0.60	0.70	0.80	0.90	0.95
$a_2 = 10$	1024.0	165.4	35.4	9.3	2.9	1.7
$a_2 = 10, w_2 = 4, r = 6$	1002.8	127.8	19.5	5.9	2.7	1.7
$a_2 = 10, w_2 = 6, r = 4$	1015.8	151.1	25.9	6.3	2.5	1.6

Στον Πίνακα 3.2 συγκρίνονται τρεις διεργασίες, η μια χωρίς προειδοποιητικά όρια και δύο με προειδοποιητικά όρια. Η παρουσία των μηδενικών μεταβάλλει την κατανομή της SN_i , αλλά το αποτέλεσμα δεν θα πρέπει να είναι μεγάλο όσο η πιθανότητα είναι ίση με μηδέν ή είναι μικρή. Έτσι, μπορούν να χρησιμοποιηθούν τα ίδια όρια ελέγχου χωρίς να επηρεαστούν σημαντικά οι ιδιότητες της διεργασίας. Στην περίπτωση που εμφανιστούν πολλά μηδενικά, μπορούμε είτε να μετρήσουμε εναλλάξ ένα μηδέν ως (-) και στη συνέχεια το επόμενο ως (+).

3.1.2 Συμπεράσματα για τα δ.ε. που βασίζονται στη συνάρτηση SN

Από τα αποτελέσματα που αναφερθήκαν στην προηγούμενη ενότητα καθώς και από αποτελέσματα άλλων μελετών, είναι σαφές ότι εάν η κατανομή των δεδομένων είναι κοντά στην Κανονική και το μέγεθος του δείγματος δεν είναι πολύ μικρό, τότε η κατανομή του \bar{X} θα είναι κοντά στην Κανονική.

Στις περιπτώσεις που η κατανομή των δεδομένων έχει βαριές ουρές, οι μη παραμετρικές τεχνικές που βασίζονται στη συνάρτηση SN , έχουν το πλεονέκτημα να έχουν σταθερό ARL όταν η διεργασία βρίσκεται εντός ελέγχου και υψηλή αποτελεσματικότητα στην ανίχνευση μετατοπίσεων του μ . Αξίζει να σημειωθεί ότι, ο προσημικός έλεγχος μπορεί να εφαρμοστεί για τον έλεγχο του μέσου αντί της διάμεσου, αλλά τότε το ARL_0 δεν θα είναι ακριβώς η υπολογισμένη τιμή όπως το πραγματικό ARL_0 για το διάγραμμα \bar{X} και κατά συνέπεια δεν θα είναι ακριβώς ίση με την θεωρητική τιμή. Αυτά τα μη παραμετρικά δ.ε., φαίνεται να είναι πιο χρήσιμα όταν χρησιμοποιούνται για μεγάλα δείγματα ($n \geq 7$) για την ανίχνευση μικρών

μετατοπίσεων. Ο Duncan (1974) επισημαίνει ότι για την ανίχνευση μικρών μετατοπίσεων είναι επιθυμητά δείγματα μεγαλύτερα από τα συνηθισμένα (4 ή 5). Τα διαγράμματα *SN* απαιτούν μέτρια μεγέθη δείγματος ($n \geq 7$) προκειμένου να επιτευχθεί ένα αρκετά μεγάλο ARL_0 .

Σε αυτή την ενότητα παρουσιάσαμε έναν απλό μη παραμετρικό δ.ε. για τη διακύμανση, το διάγραμμα αυτό είναι πολύ χρήσιμο για την παρακολούθηση της μεταβλητότητας της διεργασίας όταν η κατανομή των παρατηρήσεων είναι μη κανονική ή ασύμμετρη, όπως μας αναφέρουν οι Amin, Reynolds Jr. και Bakir (1995) στην αντίστοιχη δημοσίευση τους για τα μη παραμετρικά δ.ε. που βασίζονται στην προσημική συνάρτηση.

3.2 Δ.ε. που βασίζονται στην Προσημική-Βαθμολογική συνάρτηση

Σε αυτή την ενότητα, προτείνεται ένα ανθεκτικό διάγραμμα τύπου Shewhart βασισμένο στη συνάρτηση *SR*. Το δ.ε. αυτό είναι ανθεκτικό σε κατανομές που δεν μοιάζουν με την Κανονική όπως και σε ακραίες τιμές, διότι οι ακραίες τιμές δεν επηρεάζουν τις βαθμολογίες και τα πρόσημα των παρατηρήσεων.

Ο προσημικός έλεγχος είναι πιθανώς η απλούστερη μέθοδος που μπορεί να χρησιμοποιηθεί αν και μπορεί να μην είναι πολύ αποτελεσματική. Οι Bakir and Reynolds (1979) και οι Amin and Searcy (1991) χρησιμοποίησαν τις εντός των ομάδων προσημικές-βαθμολογίες για την ανάπτυξη ενός μη παραμετρικού διαγράμματος *CUSUM* που βασίζεται στη συνάρτηση *SR* του Wilcoxon και ενός μη παραμετρικού διαγράμματος *EWMA* αντίστοιχα.

Παρακάτω, θα γίνει μια σύντομη παρουσίαση της μεθοδολογίας ενός μη παραμετρικού δ.ε. τύπου Shewhart που βασίζεται στη σ.σ. *SR* του Wilcoxon για τον έλεγχο ενός κέντρου διεργασιών.

3.2.1 Δ.ε. *SR* για τον έλεγχο του κέντρου της διεργασίας

Στο διάγραμμα *SR* τύπου Shewhart για τον έλεγχο του κέντρου της διεργασίας το χαρακτηριστικό ελέγχου είναι ο μέσος.

Η κατασκευή του διαγράμματος γίνεται, παίρνοντας ένα τ.δ. μεγέθους $n > 1$ από παρατηρήσεις που σχετίζονται με την έξοδο της διεργασίας σε κάθε δειγματοληψία i και δεδομένου ότι οι παρατηρήσεις αυτές είναι ανεξάρτητες και συνεχείς, για κάθε j

($j = 1, 2, \dots$), καθορίζεται η απόλυτη τάξη εντός της ομάδας των αποκλίσεων ($X_{ij} - \theta_0$) με,

$$R_{ij}^+ = 1 + \sum_{i=1}^n I(|X_{ij} - \theta_0| < |X_{ij} - \theta_0|)$$

όπου $I(u) = 1$ ή 0 αν u είναι αληθής ή ψευδής.

Όπως επισημαίνει ο Bakir (2004), το R_{ij}^+ δηλώνει την τάξη του $|X_{ij} - \theta_0|$ στην υποομάδα ($|X_{i1} - \theta_0|, |X_{i2} - \theta_0|, \dots, |X_{in} - \theta_0|$). Η σ.σ. του δ.ε. είναι,

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

όπου

$$\text{sign}(u) = \begin{cases} 1 & \text{αν } u > 0 \\ 0 & \text{αν } u = 0 \\ -1 & \text{αν } u < 0. \end{cases}$$

Για ένα δίπλευρο δ.ε. τα όρια ελέγχου είναι $UCL > 0$ και $LCL < 0$ τα οποία ικανοποιούν ένα συγκεκριμένο FAR ή ARL_0 . Επιλέγοντας $LCL = -UCL$ προκύπτει ένα συμμετρικό δίπλευρο διάγραμμα. Για το μονόπλευρο δ.ε., επιλέγεται μόνο ένα όριο ελέγχου. Οι τιμές των διαφόρων ορίων ελέγχου δίδονται στους Πίνακες 3.3 και 3.4.

Πίνακας 3.3: $FAR (p_0^+)$ και ARL_0^+ για $n = 4, 5$ και 6 . (Πηγή: Saad and Bakir (2004))

	$n = 4$		$n = 5$		$n = 6$	
UCL	ARL_0^+	p_0^+	ARL_0^+	p_0^+	ARL_0^+	p_0^+
10	16	0.06250000	6.40	0.1562500	4.57	0.2187500
11	∞	0.0	10.67	0.0937500	6.40	0.1562500
12			10.67	0.0937500	6.40	0.1562500
13			16	0.0625000	9.14	0.1093750
14			16	0.0625000	9.14	0.1093750
15			32	0.0312500	12.80	0.0781250
16			∞	0.0	12.80	0.0781250
17					21.33	0.0468750
18					21.33	0.0468750
19					32.00	0.0312500
20					32.00	0.0312500
21					64	0.0156000
22					∞	0.0

Αυτά τα δ.ε. εκπέμπουν ένα εκτός ελέγχου σήμα στην πρώτη φορά που λαμβάνεται δείγμα i για το οποίο ισχύει,

$$\begin{aligned}
 &SR_i > UCL, \text{ για ένα θετικό διάγραμμα,} \\
 &SR_i < LCL, \text{ για ένα αρνητικό διάγραμμα και} \\
 &SR_i \notin (UCL, LCL) \text{ για ένα δίπλευρο διάγραμμα.}
 \end{aligned}$$

Το ARL των διαγραμμάτων αυτών υπολογίζεται ως εξής:

$$ARL^- = E(L^-) = 1/p^- \text{ (όριο ελέγχου } LCL)$$

$$ARL^+ = E(L^+) = 1/p^+ \text{ (όριο ελέγχου } UCL)$$

$$ARL = E(L) = 1/p \text{ (δίπλευρο διάγραμμα } SR \text{ τύπου Shewhart).}$$

Πίνακας 3.4: $FAR(p_0^+)$ και ARL_0^+ για $n = 8$ και 10 . (Πηγή: Saad and Bakir (2004))

$n = 8$			$n = 10$		
UCL	ARL_0^+	p_0^+	UCL	ARL_0^+	p_0^+
18	8.00	0.12500000	25	8.61	0.116210937
20	10.24	0.09765625	27	10.34	0.096679687
22	13.47	0.07421875	29	12.49	0.080078125
24	18.29	0.05468750	31	15.28	0.065429687
26	25.60	0.00390625	33	18.96	0.052734375
28	36.57	0.02734375	35	23.81	0.041992187
30	51.20	0.01953125	37	31.03	0.032226562
32	85.33	0.01171875	39	40.96	0.024414062
34	128.00	0.00781250	41	53.89	0.018554687
36	256.00	0.00390625	43	73.14	0.013671875
> 36	∞	0.0	45	102.40	0.009765625
			47	146.29	0.006835037
			49	203.70	0.004882812
			51	341.33	0.002929687
			53	512.00	0.001953125
			55	1024.00	0.000976562
			> 55	∞	0.0

3.2.2 Συμπεράσματα για τα δ.ε. που βασίζονται στην Προσημική-Βαθμολογική συνάρτηση

Σε αυτή την ενότητα, έγινε μια σύντομη παρουσίαση ενός διαγράμματος SR τύπου Shewhart που βασίζεται σε ομαδοποιημένες παρατηρήσεις για την παρακολούθηση ενός γνωστού μ . Σε αντίθεση με το παραδοσιακό διάγραμμα \bar{X} του Shewhart, το προτεινόμενο διάγραμμα είναι μη παραμετρικό υπό την έννοια ότι το ARL_0 (ή το FAR) είναι το ίδιο για οποιαδήποτε κατανομή της διεργασίας που είναι συνεχής και συμμετρική. Δίνονται επίσης, πίνακες για την ακριβή τιμή του ARL_0 και του FAR μαζί με τα αντίστοιχα όρια ελέγχου. Το διάγραμμα αυτό συνιστάται ειδικά σε καταστάσεις όπου η κατανομή των δεδομένων είναι γνωστό ότι έχει βαρύτερες ουρές από τις ουρές της Κανονικής κατανομής ή έχει πολλές ακραίες τιμές. Σε τέτοιες περιπτώσεις, αυτό το διάγραμμα είναι πιο αποτελεσματικό από το παραδοσιακό διάγραμμα \bar{X} τύπου Shewhart.

ΚΕΦΑΛΑΙΟ 4

Νέες μέθοδοι μη παραμετρικών διαγραμμάτων ελέγχου

Στην ενότητα αυτή θα παρουσιαστούν αρχικά τέσσερα νέα μη παραμετρικά δ.ε.. Τα δύο από αυτά είναι τύπου *CUSUM* ενώ τα άλλα δύο θα είναι τύπου *EWMA*. Μετά την παρουσίαση των τεσσάρων διαγραμμάτων, ακολουθεί μια σύγκριση της απόδοσης μεταξύ των διαγραμμάτων που αφορούν τον μέσο και εκείνων που αφορούν την διασπορά.

4.1 Διάγραμμα CUSUM για τον μέσο

Στην ενότητα αυτή, θα παρουσιάσουμε ένα νέο δίπλευρο δ.ε. *CUSUM* για το μέσο όπως το πρότειναν οι Yang και Cheng (2010) στην αντίστοιχη δημοσίευσή τους.

Αρχικά υποθέτουμε ένα κρίσιμο χαρακτηριστικό, X με μέσο μ . Έστω $Y = X - \mu$ και $p = P(Y > 0)$, εάν η διεργασία ήταν εντός ελέγχου τότε $p = p_0$, αλλιώς αν ο μέσος έχει μετατοπισθεί τότε $p = p_1 \neq p_0$. Η παρακολούθηση των μικρών μετατοπίσεων του μ ισοδυναμεί με την παρακολούθηση μικρών μετατοπίσεων στο ποσοστό της διεργασίας. Εάν δεν μας δοθεί το p_0 , μπορεί να εκτιμηθεί με τη χρήση ενός προκαταρκτικού συνόλου δεδομένων. Για την παρακολούθηση του μέσου της διεργασίας, χρησιμοποιούμε ένα τ.δ. μεγέθους n από το X . Το Y ορίζεται ως εξής,

$$Y_j = X_j - \mu \text{ και } I_j = \begin{cases} 1, & \text{αν } Y_j > 0 \\ 0, & \text{διαφορετικά} \end{cases} \text{ για } j = 1, 2, 3, \dots, n.$$

Ο μέσος μ όταν οι παράμετροι της διεργασίας είναι άγνωστοι όταν η διεργασία βρίσκεται εντός ελέγχου υπολογίζεται ως εξής:

$$\hat{\mu} = \bar{\bar{X}} = \frac{\sum_{i=1}^m \bar{X}_i}{m}, \bar{X}_i = \frac{\sum_{j=1}^n X_{ij}}{m}.$$

Έστω M ο συνολικός αριθμός των $Y_j > 0$, τότε $M = \sum_{j=1}^n I_j \sim B(n, p_0)$ για μια εντός ελέγχου διεργασία.

Το εντός ελέγχου p υπολογίζεται ως εξής:

$$\hat{p} = \bar{p} = \frac{\sum_{i=1}^m M_i/n}{m}.$$

Έστω $\Delta = |p_0 - p_1|$, $\Delta > 0$ και μας ενδιαφέρει να εντοπίσουμε γρήγορα τη μετατόπιση. Οι συγγραφείς όρισαν δύο σ.σ. για την παρακολούθηση του διαγράμματος τη στιγμή i , Οι συναρτήσεις αυτές είναι:

$$C_i^+ = \max(0, C_{i-1}^+ + M_i - (np_0 + k_1)), \quad i = 1, 2, \dots$$

$$C_i^- = \min(0, C_{i-1}^- - (np_0 - k_1) + M_i), \quad i = 1, 2, \dots$$

όπου k_1 είναι μια τιμή αναφοράς με $k_1 = n\Delta/2$ και $C_0^+ = 0, C_0^- = 0$. Αν $C_i^+ \geq H_1$ ή $C_i^- \leq -H_1$, τότε η διεργασία βρίσκεται εκτός ελέγχου.

Οι δύο παράμετροι σχεδιασμού του διαγράμματος, k_1 και H_1 , επιλέγονται έτσι ώστε να ικανοποιούν συγκεκριμένα *ARL*.

Προκειμένου να υπολογιστεί το *ARL* του νέου διαγράμματος, πρέπει να εξεταστεί το ανώτερο και χαμηλότερο μισό του διαγράμματος ξεχωριστά. Το άνω μισό του διαγράμματος συμβολίζεται με το διάγραμμα $CUSUM^+$ θα χρησιμοποιηθεί για τον εντοπισμό ανοδικών μετατοπίσεων της διεργασίας, δηλαδή $p_1 = p_0 + \Delta$, ενώ το χαμηλότερο μισό, $CUSUM^-$ για να ελέγξει τις μετατοπίσεις προς τα κάτω, δηλαδή $p_1 = p_0 - \Delta$.

Το νέο δίπλευρο διάγραμμα *CUSUM* για τον μέσο, προτείνεται για την παρακολούθηση και τον εντοπισμό μικρών μετατοπίσεων του μέσου της διεργασίας. Τέλος, αποτελεί μια καλή εναλλακτική λύση στα διαγράμματα \bar{X} και *CUSUM* όταν η υποκείμενη κατανομή είναι μη Κανονική ή άγνωστη.

4.2 Διάγραμμα EWMA που χρησιμοποιεί την προσημική σ.σ.

Σε αυτήν την ενότητα, παρουσιάζουμε εν συντομία το μη παραμετρικό διάγραμμα *EWMA* που προτάθηκε από τους Yang et al. (2011) για την παρακολούθηση της απόκλισης από το στόχο της διεργασίας.

Ας υποθέσουμε ότι ένα χαρακτηριστικό X , έχει μια τιμή-στόχο T . Έστω ότι $Y = X - T$ και $p = P(Y > 0)$, εάν η διαδικασία είναι υπό έλεγχο τότε $p = 0.5$, αν η διαδικασία είναι εκτός ελέγχου, τότε $p = p_1 \neq 0.5$.

Για την παρακολούθηση της απόκλισης από το στόχο, ανά πάσα στιγμή λαμβάνεται ένα τυχαίο δείγμα μεγέθους n από το X και ορίζεται η εξής σ.σ.,

$$Y_j = X_j - T \text{ και } I_j = \begin{cases} 1, \text{ αν } Y_j > 0 \\ 0, \text{ διαφορετικά} \end{cases} \text{ για } j = 1, 2, \dots, n$$

Έστω M ο συνολικός αριθμός των $Y_j > 0$, τότε $M = \sum_{j=1}^n I_j \sim B(n, 0.5)$ για μια εντός ελέγχου διεργασία.

Αρχικά, ορίζουμε τις σ.σ. παρακολούθησης του διαγράμματος,

$$EWMA_{M_t} = \lambda M_t + (1 - \lambda)EWMA_{M_{t-1}}.$$

Η αρχική τιμή $EWMA_{M_0}$ είναι ο μέσος των M που είναι ίσος με $n/2$. Επίσης γνωρίζουμε ότι,

$$E(EWMA_{M_t}) = n/2 \text{ και } Var(EWMA_{M_t}) = \frac{\lambda[1-(1-\lambda)^{2t}]}{2-\lambda} (1/4n).$$

Τα όρια ελέγχου για το διάγραμμα $SN-EWMA$ βασίζονται συνήθως στην ασυμπτωτική τυπική απόκλιση της σ.σ. ελέγχου. Ως εκ τούτου, μπορούμε να κατασκευάσουμε το διάγραμμα $SN-EWMA$ με τα εξής όρια:

$$UCL_{EWMA_M} = n/2 + k \sqrt{\frac{\lambda}{2-\lambda} \cdot \frac{1}{4n}},$$

$$CL_{EWMA_M} = n/2,$$

$$LCL_{EWMA_M} = n/2 - k \sqrt{\frac{\lambda}{2-\lambda} \cdot \frac{1}{4n}}.$$

Αν $EWMA_M \geq UCL$ ή $EWMA_M \leq LCL$ τότε η διεργασία βρίσκεται εκτός ελέγχου. Οι δύο παράμετροι διαγράμματος, k και λ , επιλέγονται ώστε να επιτευχθεί ένα συγκεκριμένο ARL . Το ARL_0 του συγκεκριμένου διαγράμματος εξαρτάται από τα k , λ , και n .

Το νέο διάγραμμα $EWMA$ παρέχει μια εναλλακτική λύση όταν η κατανομή είναι άγνωστη ή μη Κανονική. Ενώ η απόδοση του είναι αρκετά καλή για διάφορες κατανομές.

4.3 Διάγραμμα CUSUM για τη διασπορά

Στην ενότητα αυτή, παρουσιάζεται ένα νέο διάγραμμα *CUSUM* με βάση την προσημική σ.σ. όπως αυτό προτάθηκε από τους Shirke and Barale (2018). Οι συγγραφείς ανέπτυξαν ένα διάγραμμα *SN* για τη μεταβλητότητα με βάση τα δεκατημόρια, το οποίο αποτελεί τροποποίηση ενός διαγράμματος *SN* με βάση τα τεταρτημόρια ελέγχου που δόθηκαν από τους Amin et al. (1995). Η διαδικασία κατασκευής του διαγράμματος που προτείνεται έχει ως εξής.

Αρχικά, υποθέτουμε ότι D_2 και D_8 είναι αντίστοιχα το 2^ο και το 8^ο δεκατημόριο της κατανομής της διεργασίας όταν βρίσκεται υπό έλεγχο. Υποθέτουμε στη συνέχεια ότι D_2 και D_8 είναι γνωστά από προηγούμενα δεδομένα, εάν δεν είναι γνωστά μπορούν ευκολά να υπολογιστούν.

Ορίζεται η εξής συνάρτηση,

$$U_{ij} = \begin{cases} 1, \text{αν } X_{ij} \leq D_2 \text{ ή } X_{ij} \geq D_8 \\ 0, \text{αν } X_{ij} > D_2 \text{ και } X_{ij} < D_8 \end{cases}$$

και $U_i = \sum_{j=1}^n U_{ij}$. Τα U_i ακολουθούν την Διωνυμική κατανομή με παραμέτρους n και p , με $p = (X_{ij} \leq D_2 \text{ ή } X_{ij} \geq D_8 | \sigma = \delta\sigma_0)$.

Οι Shirke et al. (2016) απέδειξαν ότι το παρόν διάγραμμα είναι καλύτερο από αυτό του Amin. Ανέλυσαν αυτήν την προσέγγιση και παρείχαν ένα μη παραμετρικό διάγραμμα *CUSUM* για την παρακολούθηση της διασποράς της διεργασίας σ^2 .

Οι μικρές μετατοπίσεις στη διασπορά μπορούν να ελεγχθούν με τη βοήθεια του ποσοστού των παρατηρήσεων που βρίσκεται στις ουρές. Όταν υπάρχει μια αλλαγή στη διασπορά, ισχύει $p = p_1$.

Ας υποθέσουμε $\Delta = |p_0 - p_1|$, $\psi > 0$ και ότι θέλουμε να εντοπίσουμε γρήγορα μια μετατόπιση p_1 . Ορίζεται μια σ.σ. παρακολούθησης του διαγράμματος *CUSUM* για το i δείγμα υποομάδας,

$$C_i^+ = \max(0, U_i - (np_0 + k_1) + C_{i-1}^+)$$

$$C_i^- = \max(0, (np_0 + k_1) - U_i + C_{i-1}^-)$$

όπου k_1 μια τιμή αναφοράς με $k_1 = \frac{n\Delta}{2}$. Οι αρχικές τιμές των C είναι ίσες με το μηδέν ($C_0^+ = 0$ και $C_0^- = 0$). Η διεργασία θεωρείται εκτός ελέγχου όταν $C_i^+ > H$ (για την ανίχνευση αύξησης της διασποράς) ή $C_i^- < H$ (για τη μείωση της), H είναι η

παράμετρος ενός μη παραμετρικού διαγράμματος *CUSUM*. Είναι σημαντικό να επιλεγούν οι κατάλληλες τιμές των k_1 και H ώστε να βελτιστοποιείται η απόδοση του μήκους ροής του διαγράμματος

Σε αυτή την ενότητα, παρουσιάσαμε ένα νέο μη παραμετρικό δ.ε. *CUSUM* βασισμένο στα δεκατημόρια ελέγχου για την ανίχνευση μικρών μετατοπίσεων στη διασπορά της διεργασίας. Δεδομένου ότι, όποια και αν είναι η κατανομή της διεργασίας το προτεινόμενο μη παραμετρικό διάγραμμα *CUSUM* διατηρεί το ίδιο ARL_0 . Επομένως, το προτεινόμενο διάγραμμα είναι μια καλύτερη εναλλακτική λύση για το διάγραμμα *CUSUM S²* όταν η κατανομή δεν είναι γνωστή εκ των προτέρων. Η χρήση του προτεινόμενου διαγράμματος *CUSUM* συνιστάται, λόγω της απλοποιημένης της διαδικασίας κατασκευής του.

4.4 Διάγραμμα EWMA για τη διασπορά

Σε αυτήν την ενότητα, εξετάζουμε εν συντομία το μη παραμετρικό διάγραμμα *EWMA* για την παρακολούθηση της μεταβλητότητας, που προτάθηκε από τους Yang and Arnold (2015) και στη συνέχεια τροποποιήθηκε από τον Haq (2017).

Αρχικά, υποθέτουμε πως οι παράμετροι της διεργασίας είναι όλες γνωστές εκ των προτέρων. Θεωρείται ότι τόσο ο μέσος μ_X όσο και η διακύμανση σ_X^2 της διεργασίας είναι γνωστές σταθερές.

Ωστόσο, όταν το μ_X και το σ_X^2 είναι άγνωστα, είναι σύνηθες να υπολογίζεται από παλαιότερα δεδομένα. Οι τύποι υπολογισμού τους είναι:

$$\bar{X} = \frac{\sum_{i=1}^m \bar{X}_i}{m} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij}$$

$$\overline{S_X^2} = \frac{1}{m} \sum_{i=1}^m S_{Xi}^2 = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

Στο συγκεκριμένο δ.ε. οι τιμές, εκτιμώνται από δείγματα διεργασιών από τη Φάση I. Στη Φάση II αυτές οι δύο εκτιμώμενες ποσότητες αντιμετωπίζονται ως σταθερές. Ξεκινώντας λοιπόν τη Φάση II υποθέτουμε ότι τα μ_X και σ_X^2 είναι γνωστά. Έτσι έχουμε την εξής σ.σ.,

$$X_{ij}^{**} = (X_{ij} - \mu_X)^2, j = 1, 2, \dots, n$$

Παρατηρείται ότι $E(X_{ij}^{**}) = \sigma_x^2$ για $j = 1, 2, \dots, n$. Χρησιμοποιώντας το X_{ij}^{**} μπορούμε να ορίσουμε τη μεταβλητή δείκτη I_{ij}^{**} ($1 \leq j \leq n$) ως εξής,

$$I_{ij}^{**} = \begin{cases} 1, & \text{αν } X_{ij}^{**} > \sigma_x^2 \\ 0, & \text{διαφορετικά} \end{cases}$$

Έστω τώρα, $A_t = \sum_{j=1}^n I_{ij}^{**} \sim B(n, p)$, ο συνολικός αριθμός των περιπτώσεων για τις οποίες ισχύει $X_{ij}^{**} > \sigma_x^2$ με $p = p_0 = P(X_{ij}^{**} > \sigma_x^2)$. Εδώ η τιμή του p εξαρτάται από την κατανομή των X_i .

Θα μπορούσε να κατασκευαστεί ένα μη παραμετρικό δ.ε. με το A_t , αλλά το πρόβλημα είναι η αντιστοίχιση του ARL_0 του σε ένα συγκεκριμένο επίπεδο. Για να διορθωθεί αυτό το πρόβλημα, οι συγγραφείς μετέτρεψαν το A_t με τη βοήθεια της συνάρτησης \sin^{-1} . Έτσι, πλέον έχουμε τη σ.σ. $B_t = \sin^{-1} \sqrt{A_t/n}$. Η κατανομή του B_t μπορεί στη συνέχεια να προσεγγιστεί με μια κανονική τ.μ. με μέσο $\sin^{-1} \sqrt{p}$ και διακύμανση $1/4n$. Η κατανομή του B_t δεν εξαρτάται από την κατανομή των X_t .

Χρησιμοποιώντας τα \bar{X} και $\overline{S_X^2}$ μπορούμε υπολογίσουμε το p_0 ως εξής:

$$\hat{p}_0 = \frac{1}{m} \sum_{i=1}^m \left(\frac{A_t}{n} \right) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^{**}$$

Έστω τώρα, $\{D_t\}$ μια ακολουθία $EWMA$ που βασίζεται σε μια IID ακολουθία $\{B_t\}$ ορίζεται ως εξής,

$$D_t = \lambda B_t + (1 - \lambda)D_{t-1}, \quad D_0 = \sin^{-1}(\sqrt{p_0}),$$

και $\lambda \in (0, 1]$ η σταθερά εξομάλυνσης.

Για μια εντός ελέγχου διεργασία ισχύει,

$$E(D_t) = E(B_t) = \sin^{-1}(\sqrt{p_0}) \quad \text{και} \quad Var(D_t) = \left(\frac{\lambda}{2-\lambda} \right) (1 - (1-\lambda)^{2t}) \left(\frac{1}{4n} \right).$$

Η ασυμπτωτική διακύμανση του D_t όταν $t \rightarrow \infty$ είναι ίση με $Var(D_t) = \left(\frac{\lambda}{2-\lambda} \right) \left(\frac{1}{4n} \right)$. Τα όρια ελέγχου του διαγράμματος $EWMA$ με βάση το D_t τη στιγμή t είναι:

$$UCL_t = \sin^{-1}(\sqrt{p_0}) + k \sqrt{\frac{\lambda}{2-\lambda} \cdot \frac{1}{4n}}$$

$$CL_t = \sin^{-1}(\sqrt{p_0}),$$

$$LCL_t = \sin^{-1}(\sqrt{p_0}) - k \sqrt{\frac{\lambda}{2-\lambda} \cdot \frac{1}{4n}}$$

όπου η τιμή του $k > 0$ επιλέγεται σύμφωνα με το επιθυμητό ARL_0 του διαγράμματος. Είναι σημαντικό να επιλεγούν οι κατάλληλες τιμές των λ και k ώστε να βελτιστοποιείται η απόδοση του μήκους ροής του διαγράμματος. Το προτεινόμενο διάγραμμα $EWMA$ εκπέμπει εκτός ελέγχου σήμα κάθε φορά που ισχύει $D_t < LCL_t$ ή $D_t > UCL_t$.

4.5 Σύγκριση νέων μεθοδολογιών

Σε αυτή την ενότητα θα γίνει σύγκριση της απόδοσης των νέων μεθοδολογιών που παρουσιάσαμε σε αυτό το Κεφάλαιο. Η σύγκρισή τους έγινε με τη χρήση προσομοίωσης και με 10000 επαναλήψεις, στη γλώσσα προγραμματισμού R και το μέτρο που χρησιμοποιήθηκε για την σύγκριση των διαγραμμάτων του παρόντος κεφαλαίου είναι το ARL_1 .

Τα δεδομένα αποτελούνται από 40 δείγματα 10 παρατηρήσεων τα οποία προέρχονται από τις εξής κατανομές:

- α) Κανονική, τα 8 πρώτα από την $N(0,1)$ και τα 32 τελευταία από την $N(2,1.6)$.
- β) Laplace, τα 8 πρώτα από τη $Laplace(0,0.5)$ και τα 32 τελευταία από τη $Laplace(1,3.2)$.
- γ) Βήτα, τα 8 πρώτα από τη $Beta(0.5,0.5)$ και τα 32 τελευταία από τη $Beta(0.4,0.15)$.
- δ) Ομοιόμορφη, τα 8 πρώτα από την $U(0,1)$ και τα 32 τελευταία από την $U(0.25,2.1)$.

Η διαφοροποίηση που έγινε στις κατανομές μετά το 8^ο δείγμα έχει σκοπό τον έλεγχο για τον έγκαιρο εντοπισμό αλλαγής στη διεργασία.

Οι σταθερές των διαγραμμάτων των παρακάτω πινάκων έχουν οριστεί ως εξής: $k_1 = 0.5$ για τα διαγράμματα $M-CUSUM$ και $D-CUSUM$, εφόσον θέλουμε να ανιχνεύσουμε μετατόπιση της τάξης του 0.1 και (προσεγγιστικά) $p_0 = 0.5$ και $p_0 = 0.4$ αντίστοιχα, $\lambda = 0.22$ και $k = 2.86$ για το διάγραμμα $D-EWMA$, ενώ για το διάγραμμα $SN-EWMA$ έχουν οριστεί $\lambda = 0.05$ και $k = 2.49$. Για το $M-CUSUM$ εφόσον έχουμε $n = 10$ και $k_1 = 0.5$ έχουμε όρια ελέγχου $H_1 = \pm 10.65$ (σύμφωνα με τον Πίνακα 1 των Yang και Cheng (2010)). Για το $D-CUSUM$ εφόσον έχουμε $n = 10$ και $k_1 = 0.5$ έχουμε όρια ελέγχου $H = 9.30$ (σύμφωνα με τον Πίνακα 2 των Shirke and Barale (2018)). Για τα

διαγράμματα *D-EWMA* και *SN-EWMA* οι τιμές των ορίων και του p_0 υπολογίζονται για το κάθε ένα ξεχωριστά και οι τιμές έχουν οριστεί κατάλληλα ώστε να ικανοποιούνται οι συγκεκριμένες τιμές του ARL_0 και σύμφωνα με τις προτάσεις των αρθρογράφων των δύο διαγραμμάτων αντίστοιχα.

Πίνακας 4.1: ARL_1 για τα διαγράμματα για τον μέσο με $n = 10$.

Διάγραμμα	Κατανομή			
	Κανονική	Laplace	Βήτα	Ομοιόμορφη
<i>M-CUSUM</i> (4.1)	16.28	17.54	14.27	11.78
<i>SN-EWMA</i> (4.2)	16.54	17.25	14.57	12.24

Πίνακας 4.2: ARL_1 για τα διαγράμματα για τη διασπορά με $n = 10$.

Διάγραμμα	Κατανομή			
	Κανονική	Laplace	Βήτα	Ομοιόμορφη
<i>D-CUSUM</i> (4.3)	10.66	10.36	12.59	12.00
<i>D-EWMA</i> (4.4)	9.9	9.58	12.65	11.69

Από τον Πίνακα 4.1 παρατηρούμε ότι, το διάγραμμα των Yang και Cheng (2010) (*M-CUSUM*) έχει ελάχιστα καλύτερη απόδοση από το διάγραμμα των Yang et al. (2011) (*SN-EWMA*), για κατανομές όπως η Κανονική, η Βήτα και η Ομοιόμορφη σύμφωνα με το ARL_1 . Το δεύτερο διάγραμμα έχει καλύτερη απόδοση από το πρώτο μόνο για την κατανομή Laplace σύμφωνα με το ARL_1 , καθώς το διάγραμμα *SN-EWMA* εντοπίζει την μετατόπιση στο μέσο πολύ κοντά στο 17^ο (17.25) δείγμα, ενώ το διάγραμμα *M-CUSUM* εντοπίζει την μετατόπιση ελάχιστα αργότερα (17.54). Τα δύο διαγράμματα γενικά, φαίνεται να έχουν καλύτερη απόδοση όταν η κατανομή της διεργασίας τείνει να μοιάζει με την Ομοιόμορφη.

Από τον Πίνακα 4.2 παρατηρούμε ότι, το διάγραμμα του Haq (2017) (*D-EWMA*) έχει καλύτερη απόδοση από το διάγραμμα των Shirke and Barale (2018) (*D-CUSUM*) για κατανομές όπως η Κανονική και η Laplace. Για κατανομές όπως η Βήτα, τα δύο διαγράμματα έχουν όμοια απόδοση και για κατανομές όπως η Ομοιόμορφη η απόδοση των δύο διαγραμμάτων έχει σχετικά μικρή διαφορά. Γενικά όμως, το διάγραμμα *D-EWMA* φαίνεται να έχει την καλύτερη απόδοση.

Τέλος, πρέπει να τονιστεί πως, η απόδοση των διαγραμμάτων μπορεί να αλλάξει για διαφορετικά δεδομένα και για διαφορετικό n συνεπώς δεν μπορούμε να αποφανθούμε με σιγουριά για κάποιο από τα διαγράμματα αυτά.

ΚΕΦΑΛΑΙΟ 5

Πολυμεταβλητα διαγράμματα ελέγχου

Σε αυτό το κεφάλαιο θα γίνει μια σύντομη βιβλιογραφική παρουσίαση πολυμεταβλητών μη παραμετρικά δ.ε., ενώ στο δεύτερο μέρος του θα γίνει μια σύντομη παρουσίαση δύο πολυμεταβλητών διαγραμμάτων τα οποία κάνουν ταυτόχρονο έλεγχο στον μέσο και στη διασπορά της διεργασίας.

5.1 Διάφορα Πολυμεταβλητα δ.ε.

Η Liu (1995) πρότεινε δύο πολυμεταβλητά δ.ε. βασισμένα στο βάθος των δεδομένων. Τα διαγράμματα αυτά είναι το διάγραμμα Q και το διάγραμμα r τα οποία αποτελούν γενικεύσεις των διαγραμμάτων \bar{X} και X . Για την κατασκευή των διαγραμμάτων χρησιμοποιήθηκε η κατανομή του δείγματος των παρατηρήσεων (F) και η κατανομή του δείγματος αναφοράς (G). Το διάγραμμα r έχει μόνο κάτω όριο ελέγχου και γνωρίζουμε πως η διεργασία τίθεται εκτός ελέγχου όταν η σ.σ. βρεθεί κάτω από το ποσοστό λανθασμένου συναγερμού (το οποίο είναι κοντά στο μηδέν). Για το διάγραμμα Q τώρα, υπάρχουν τρεις προτάσεις, εκείνη που είναι για μεγάλα μεγέθη δείγματος και εκείνο για μικρές τιμές του ποσοστό λανθασμένου συναγερμού.

Ένα αρκετά γνωστό μη παραμετρικό πολυμεταβλητό δ.ε., είναι αυτό του Boone (2010). Ο Boone βασισμένος στην προσέγγιση του Hettmansperger (2006) και στο διάγραμμα SN^2 της Das (2009), πρότεινε ένα πολυμεταβλητό SN δ.ε. (SN^2 Chart) το οποίο είναι μια γενίκευση του διαγράμματος SN και επέκταση του διαγράμματος της Das (2009). Για την παρακολούθηση της διαμέσου το διάγραμμα αυτό χρησιμοποιεί όπως καταλαβαίνουμε την προσημική συνάρτηση. Χρησιμοποιώντας και μετασχηματίζοντας κατάλληλα την προσημική συνάρτηση ο αρθρογράφος, κατασκεύασε την πολυμεταβλητή προσημική συνάρτηση και εν συνεχεία εντοπίζοντας την κατανομή της συνάρτησης SN^2 έδωσε τα όρια ελέγχου για το νέο διάγραμμα. Μελετώντας αυτό το διάγραμμα η Das (2009), σύγκρινε το διάγραμμα SN^2 με το διάγραμμα T^2 του Hotelling και σύμφωνα με το ARL_0 του παρατήρησε πως το διάγραμμα SN^2 έχει καλύτερη απόδοση σε σχέση με το διάγραμμα T^2 για μετατόπιση μικρότερη ή ίση του 1.5 ενώ για μετατόπιση μεγαλύτερη από την τιμή αυτή τα διαγράμματα δεν διαφέρουν πολύ.

Παράλληλα, ο Boone (2010) μαζί με το πολυμεταβλητό δ.ε. SN^2 , πρότεινε και το πολυμεταβλητό προσημικό-βαθμολογικό δ.ε. τύπου Shewhart (SR^2 Chart). Το διάγραμμα αυτό χρησιμοποιείται για την παρακολούθηση των διαμέσων και βασίζεται στην προσημική βαθμολογική συνάρτηση του Wilcoxon. Το διάγραμμα SR^2 αποτελεί και αυτό μια γενίκευση του διαγράμματος SR . Πρέπει να τονιστεί πως, για να υπολογιστεί η συνάρτηση SR^2 για κάθε δείγμα χρησιμοποιείται ο ασυμπτωτικός πίνακας συνδιακύμανσης.

5.2 Πολυμεταβλητα Διαγράμματα ταυτόχρονου ελέγχου του μέσου και της διασποράς **Διάγραμμα Max-MEWMA**

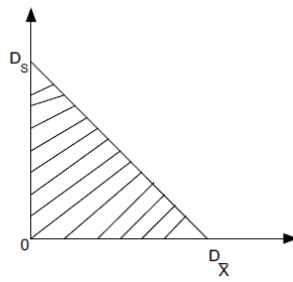
Οι Chen et al. (2005) ανέπτυξαν ένα πολυμεταβλητό δ.ε. $EWMA$ για τον ταυτόχρονο έλεγχο της μέσης τιμής και της διακύμανσης της διεργασίας. Η μέθοδος αυτή υπολογίζει αρχικά το διάνυσμα της μέσης τιμής χρησιμοποιώντας τις παρατηρήσεις X_i και τον πίνακα συνδιακύμανσης των μετρήσεων της πολυμεταβλητής κατανομής της διεργασίας. Στη συνέχεια, υπολογίζεται το διάνυσμα της διακύμανσης της διεργασίας. Γνωρίζοντας πως τα δύο διανύσματα είναι ανεξάρτητα και συνδυάζοντας τα, δημιούργησαν μια νέα σ.σ. για το νέο πολυμεταβλητό δ.ε. ($Max-MEWMA$) η οποία χρησιμοποιεί την μέγιστη τιμή ανάμεσα στην απόλυτη τιμή της μέσης τιμής και της διακύμανσης.

Μια μεγάλη τιμή της σ.σ. σημαίνει ότι ο μέσος της διεργασίας έχει απομακρυνθεί από το 0 ή/και η μεταβλητότητα της διαδικασίας έχει αλλάξει. Από την άλλη, μια μικρή τιμή της σ.σ. υποδηλώνει ότι ο μέσος της διεργασίας και η μεταβλητότητα έχουν παραμείνει κοντά στις επιθυμητές τιμές.

Το συγκεκριμένο δ.ε., μπορεί να είναι αρκετά καλό σε περίπτωση αλλαγής των μεγεθών δειγμάτων εύκολα και διατηρεί τις επιθυμητές ιδιότητες των διαγραμμάτων $EWMA$. Σε σύγκριση με το συνδυασμό του διαγράμματος X^2 και του διαγράμματος $|S|$, το διάγραμμα $Max-MEWMA$ έχει καλύτερη απόδοση στην ανίχνευση μικρών έως μέτριων αλλαγών σε μια διεργασία.

Διάγραμμα Semi – Circle

Οι Cheng and Mao (2008) προτείναν το πολυμεταβλητό διάγραμμα T_k . Η σ.σ. του συγκεκριμένου διαγράμματος χρησιμοποιεί τις μετρήσεις από την διεργασία και προσθέτει τις παραμέτρους D_x και D_s , για την θέση και την διασπορά αντίστοιχα. Αφού υπολογιστούν αυτές οι δύο συναρτήσεις, κατασκευάζεται ένα διάγραμμα σε σχήμα «τριγώνου» όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 2. Διάγραμμα T_k .

Πηγή: Cheng and Mao (2008)

Η σ.σ. που ελέγχεται συγκρίνεται με μια τιμή C η οποία εξαρτάται από την κατανομή της σ.σ. ($\chi_{nk,(1-a)}^2$). Αν η σ.σ. βρεθεί έξω από το «τρίγωνο» τότε η διεργασία βρίσκεται εκτός ελέγχου και πρέπει να ελεγχθεί η αιτία που την έφερε εκτός ελέγχου.

Το διάγραμμα T_k μπορεί να χρησιμοποιηθεί για τον ταυτόχρονο έλεγχο της μετατόπισης θέσης ή/και της αλλαγής διακύμανσης μιας πολυδιάστατης διεργασίας. Οι Cheng and Mao (2008) αναφέρουν στο άρθρο τους πως το διάγραμμα αποδίδει αρκετά καλά σε σχέση με το διάγραμμα $\chi^2 - |S|$ στις περισσότερες περιπτώσεις ενώ επίσης είναι πολύ απλό στην κατασκευή και την κατανόηση του. Πρέπει να τονιστεί ότι το συγκεκριμένο διάγραμμα μπορεί να μετατραπεί σε διάγραμμα τύπου $EWMA$ χρησιμοποιώντας μια χρονική ακολουθία.

ΚΕΦΑΛΑΙΟ 6

Εφαρμογές μη παραμετρικών διαγραμμάτων ελέγχου

Σε αυτό το κεφάλαιο θα γίνει μια σύντομη παρουσίαση τριών εφαρμογών των μη παραμετρικών δ.ε.. Τα διαγράμματα αυτά βρίσκουν εφαρμογές σε διάφορες επιστήμες όπως αυτή της ιατρικής, φυσικής, πληροφορικής κ.α..

6.1 Εφαρμογή σε δεδομένα του καρκίνου του μαστού

Ο καρκίνος του μαστού είναι ένας κακοήθης όγκος που προκύπτει από τα κύτταρα του μαστού. Οι επιπλοκές του μπορεί να επηρεάσουν σχεδόν κάθε μέρος του σώματος. Ο προσυμπτωματικός έλεγχος του είναι μια σημαντική στρατηγική για την έγκαιρη ανίχνευση και για να εξασφαλιστεί μεγαλύτερη πιθανότητα ο ασθενής να έχει ένα καλό αποτέλεσμα από τη θεραπεία. Τα ισχυρά μοντέλα πρόβλεψης που βασίζονται σε δεδομένα που μπορούν να συλλεχθούν και η ανάλυση αίματος επιδιώκεται να συμβάλουν σημαντικά, προσφέροντας έτσι περισσότερα εργαλεία προσυμπτωματικού ελέγχου και είναι σημαντικά για την ανίχνευση αλλαγών στον καρκίνο του μαστού.

Τα δ.ε. έχουν αυξανόμενη σημασία στις μελέτες βιοπαρακολούθησης. Για παράδειγμα, οι Rogerson and Yamada (2004) πρότειναν μια πολυμεταβλητή προσέγγιση του διαγράμματος *CUSUM* για την ανίχνευση αλλαγών σε χωρικά μοτίβα και την εφάρμοσαν σε δεδομένα καρκίνου του μαστού σε επίπεδο πολιτείας στις Βορειοανατολικές Ηνωμένες Πολιτείες. Τα αποτελέσματα σύγκρισης έδειξαν ότι το διάγραμμα είχε καλές επιδόσεις. Παρόμοια μη παραμετρικά δ.ε. έχουν χρησιμοποιηθεί και από άλλους ερευνητές στον κλάδο της ιατρικής.

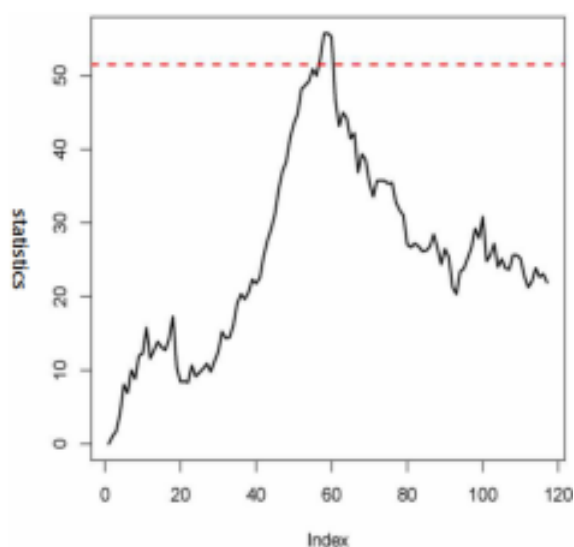
Οι Liu et al. (2015) εισήγαγαν ένα μη παραμετρικό πολυμεταβλητό διάγραμμα *EWMA* που βασίζεται στην κατάταξη ανεξάρτητων παρατηρήσεων X_i , πάνω στο οποίο βασίστηκαν οι Yue et al. (2020) για να κατασκευάσουν ένα διάγραμμα το οποίο τους βοήθησε στην πρόβλεψη και την ανίχνευση του καρκίνου του μαστού. Η σ.σ. του διαγράμματος που χρησιμοποίησαν είναι:

$$EWMA_i = \lambda Y_i + (1 + \lambda)EWMA_{i-1}, \quad 0 < \lambda \leq 1.$$

Στη συνέχεια, διεξήχθη μια μελέτη ασθενών-μαρτύρων με 116 συμμετέχοντες (64 ασθενείς με καρκίνο του μαστού και 52 υγιείς μάρτυρες) όπου για τον κάθε ένα παρατηρήθηκαν διάφορα κλινικά χαρακτηριστικά. Μελετήθηκαν διάφορα ποσοτικά χαρακτηριστικά όπως είναι η ηλικία (έτη), ο Δείκτης Μάζας Σώματος (kg/m²), η γλυκόζη (mg/dL), η ινσουλίνη (μU/mL) κ.α.. Τα χαρακτηριστικά είναι ανθρωπομετρικά δεδομένα και παράμετροι που μπορούν να συγκεντρωθούν σε μια ανάλυση αίματος ρουτίνας και μπορούν να χρησιμοποιηθούν ως βιοδείκτες καρκίνου του μαστού. Τα δεδομένα που χρησιμοποιήθηκαν από τους Yue et al. (2020) ονομάζονται “Breast Cancer Coimbra Data Set” από το UCI Machine Learning Repository καθώς είναι δημόσια διαθέσιμα. Σε αυτά τα δεδομένα παρατηρήθηκε ότι η υπόθεση της Κανονικότητας απορρίπτεται. Για το λόγο αυτό, για την απεικόνιση των δεδομένων χρησιμοποιήθηκε το μη παραμετρικό διάγραμμα των Liu et al..

Το διάγραμμα *EWMA* υπέδειξε στους συγγραφείς ότι υπάρχει αλλαγή μετά την 52^η παρατήρηση. Επομένως, χρησιμοποιήθηκαν τα εντός ελέγχου δεδομένα μέχρι την 50^η παρατήρηση για να βρεθούν τα όρια ελέγχου. Στη συνέχεια, χρησιμοποιήθηκαν τα όρια ελέγχου για την παρακολούθηση της υπόλοιπης διαδικασίας. Από τα κατάλοιπα της σ.σ. *EWMA* παρατηρήθηκε ότι η προτεινόμενη μέθοδος μπορεί να παρέχει σχετικά έγκαιρη ανίχνευση σε μια διαδικασία.

Τέλος οι συγγραφείς κατέληξαν στο συμπέρασμα ότι, τα δεδομένα της Coimbra για τον καρκίνο του μαστού δείχνουν ότι οι σ.σ. δεν εμπίπτουν στο όριο ελέγχου στην 53^η παρατήρηση. Αυτό σημαίνει ότι υπάρχει σημαντική ένδειξη η οποία δείχνει ότι κάποιος έχει υψηλό κίνδυνο να πάθει καρκίνο του μαστού.



Σχήμα 3: Διάγραμμα *EWMA* για δεδομένα καρκίνου του μαστού Coimbra.

Πηγή: (Liu et al. 2020)

6.2 Εφαρμογή για το μέγεθος και τη συχνότητα διαδοχικών γεγονότων

Τα περισσότερα δ.ε. που υπάρχουν στη βιβλιογραφία και αφορούν τους ενδιαμέσους χρόνους μεταξύ διαδοχικών γεγονότων επικεντρώνονται μόνο στην παρακολούθηση της συχνότητας χωρίς να λαμβάνουν υπόψη το μέγεθος ενός γεγονότος.

Οι Wu et al. (2009) επεσήμαναν ότι υπάρχουν δύο τύποι γεγονότων, τα αρνητικά όπως ένας σεισμός και τα θετικά όπως η εντολή αγοράς ενός προϊόντος. Παράλληλα, πρότειναν ένα δ.ε. $T&X$ (δηλαδή, ένα διάγραμμα T για την παρακολούθηση του χρονικού διαστήματος t μαζί με ένα διάγραμμα X). Προτείναν επίσης, ένα ενιαίο διάγραμμα για την παρακολούθηση της αναλογίας μεταξύ X και T , καθώς είναι πιο αποτελεσματικό από ένα μεμονωμένο διάγραμμα t , ένα μεμονωμένο διάγραμμα x .

Οι Huang et al. (2017) στο αντίστοιχο άρθρο τους παρουσίασαν ένα μη παραμετρικό διάγραμμα $EWMA$, το οποίο παρακολουθεί ταυτόχρονα τη συχνότητα και το μέγεθος των συμβάντων με βάση τη σ.σ. του Mathur, η οποία έχει αποδειχθεί αποτελεσματική στην ανίχνευση αλλαγών για διμεταβλητές διεργασίες.

Η σ.σ. του Marthur που χρησιμοποιήθηκε είναι:

$$M = \sum_{j=1}^n \sum_{i=1}^{nm} L_{ij}$$

όπου

$$L_{ij} \begin{cases} 1, \text{αν } Y_{1i}^2 > Y_{2i}^2 \\ 0, \text{διαφορετικά} \end{cases}$$

Η συνάρτηση του διαγράμματος Marthur- $EWMA$ ορίζεται ως εξής:

$$S_{M,j} = (1 - \lambda)S_{M,j-1} + \lambda M_j, j = 1, 2, 3, \dots$$

Για την εφαρμογή του χρησιμοποιήθηκε ένα σύνολο δεδομένων που περιέχει την ημερομηνία και τους θανάτους των καταστροφών ανθρακωρυχείων στην Αμερική. Τα δεδομένα ξεκινούν από το Μάιο του 1907 έως τον Απρίλιο του 2004 και αποτελούνται από 215 παρατηρήσεις σοβαρών καταστροφών ανθρακωρυχείων με περισσότερους από 10 θανάτους. Το αντίστοιχο διάγραμμα Mathur- $EWMA$ κατασκευάστηκε με $m = 50$, $n = 11$ και $\lambda = 0.1$ και τα όρια ελέγχου που βρέθηκαν είναι $(LCL, UCL) = (239.89, 310.10)$. Όπως παρατηρήθηκε, βρέθηκαν τρία σήματα εκτός ελέγχου στη 13^η, 14^η και 15^η θέση, γεγονός που επαλήθευσε την αποτελεσματικότητα του προτεινόμενου διαγράμματός τους.

6.3 Εφαρμογή σε δεδομένα τρισδιάστατης εκτύπωσης

Η τρισδιάστατη εκτύπωση είναι μια διαδικασία σύνθεσης ενός τρισδιάστατου προϊόντος που σχηματίζεται από διαδοχικά στρώματα υλικού υπό τον έλεγχο του υπολογιστή (Evans, 2012). Έχει χρησιμοποιηθεί για να δημιουργήσει μέρη αυτοκινήτων, εξαρτήματα smartphone, κ.α..

Για να μελετηθεί ένα προϊόν 3D εκτύπωσης, χρειάζεται ένα σαρωτή λέιζερ και τα δεδομένα της επάνω επιφάνειας ενός προϊόντος. Πριν από την εκτύπωση του προϊόντος, οι χρήστες πρέπει να καταχωρίσουν το σχεδιασμένο σχήμα του σε έναν υπολογιστή. Στη συνέχεια, ο εκτυπωτής εκτυπώνει το προϊόν σύμφωνα με αυτό το σχέδιο. Μετά την εκτύπωση ενός προϊόντος, ο σαρωτής λέιζερ καταγράφει τα δεδομένα της επιφάνειας του, με το προϊόν σταθερό σε περιστρεφόμενη θήκη τοποθετημένη πριν από το σαρωτή.

Κατά την παρακολούθηση της επιφάνειας του προϊόντος, είναι σημαντικό να ευθυγραμμισθούν γεωμετρικά οι θέσεις των δεδομένων των διαφόρων προϊόντων. Οι Zang and Qiu (2018) επεσήμαναν μερικά προβλήματα που αντιμετωπίζει ο έλεγχος ποιότητας στις 3D εκτυπώσεις όπως είναι ο όγκος των δεδομένων ο οποίος είναι μεγάλος. Αυτό απαιτεί υπολογιστικά αποτελεσματικές στατιστικές μεθόδους για την ανάλυση των δεδομένων και για την παρακολούθηση της διαδικασίας εκτύπωσης. Ως εκ τούτου, η επιφάνεια δεν μπορεί να περιγραφεί από ένα παραμετρικό μοντέλο και οι στατιστικές μέθοδοι θα πρέπει να είναι κατάλληλες για άνισους κατανομημένα σχεδιασμούς. Αυτά τα προβλήματα τους υπέδειξαν ότι η παρακολούθηση των προϊόντων εκτύπωσης 3D είναι δύσκολη. Όπως μας περιγράφουν οι συγγραφείς, οι μέθοδοι έγκαιρης παρακολούθησης βασίζονται σε γραμμική ή παραμετρική μοντελοποίηση. Νωρίτερα, οι ίδιοι στα πλαίσια της παρακολούθησης της Φάσης I παρουσίασαν τρόπους ανίχνευσης των ακραίων τιμών μεταξύ ενός συνόλου προϊόντων 3D εκτύπωσης χρησιμοποιώντας ένα διάγραμμα Shewhart. Στη συνέχεια, μελέτησαν και εφάρμοσαν ένα διάγραμμα *CUSUM* Φάσης II. Η μεθοδολογία τους επικεντρώθηκε κυρίως στην παρακολούθηση των κορυφών (των επιφανειών) των προϊόντων. Η προτεινόμενη διαδικασία παρακολούθησης Φάσης II έχει τρία κύρια στοιχεία.

- i. Καταχώριση επιφάνειας: όλες οι παρατηρούμενες επιφάνειες των προϊόντων εκτύπωσης 3D είναι γεωμετρικά ευθυγραμμισμένες.
- ii. Μη παραμετρική εκτίμηση επιφάνειας: Η επιφάνεια της κορυφής και άλλες ποσότητες εκτιμώνται από ένα σύνολο εντός ελέγχου δεδομένων, χρησιμοποιώντας μια μη παραμετρική τοπική διαδικασία εξομάλυνσης του πυρήνα.

- iii. Ηλεκτρονική παρακολούθηση της διαδικασίας: κατασκευάζεται ένα μη παραμετρικό διάγραμμα *CUSUM* Φάσης II για ηλεκτρονική παρακολούθηση προϊόντων 3D εκτύπωσης.

Έπειτα από ανάλυση της μεθοδολογίας τους πρότειναν την εξής σ.σ. για το διάγραμμα *CUSUM*:

$$Q_{ART,i}^+ = \max \left(0, Q_{ART,i-1}^+ + \frac{\Lambda_{ART,i} - E_0(\Lambda_{ART,i})}{\sqrt{Var_0(\Lambda_{ART,i})}} - k \right), \quad i = 1, 2, \dots$$

με $Q_{ART,0}^+ = 0$ ενώ ως $\Lambda_{ART,i}$ ορίζεται η συνολική διαφορά μεταξύ της i εκτιμώμενης ευθύγραμμης επιφάνειας από τη Φάση II και της κορυφής του προϊόντος της 3D εκτύπωσης.

Το συγκεκριμένο δ.ε. εκπέμπει εκτός ελέγχου σήμα για τη μετατόπιση της επιφάνειας όταν $Q_{ART,i}^+ > c_{ART}$.

Η εφαρμογή αφορούσε πραγματικά δεδομένα από τρισδιάστατο εκτυπωτή και σαρωτή λέιζερ. Το εντός ελέγχου προϊόν είχε ως χαρακτηριστικά το χώρο σχεδιασμού να είναι $[-30,30] \times [-30,30]$ (σε mm^2) και τις τιμές για το ύψος σε διαφορετικές (x,y) θέσεις $[10,30] mm$. Για κάθε προϊόν, ο αριθμός των παρατηρήσεων σε κάθε παρατηρούμενη επιφάνεια που παράγεται από το σαρωτή λέιζερ βρισκόταν μεταξύ 141.000 και 158.500. Στη συνέχεια, ο εκτυπωτής εκτύπωσε 15 προϊόντα και οι παρατηρούμενες επιφάνειες αυτών χρησιμοποιούνται ως δεδομένα ελέγχου. Για το διάγραμμα $Q_{ART,i}^+$ αυτό επιλέχθηκαν $k = 0.5$ και $ARL_0 = 20$. Τα όρια ελέγχου του διαγράμματος καθορίζονται από τη διαδικασία εκκίνησης με το μέγεθος δείγματος εκκίνησης $B = 10.000$.

Αφού υπολογίστηκε οι σ.σ., έδωσαν την εντολή στον υπολογιστή να εκτυπώσει τέσσερα προϊόντα εντός και πέντε εκτός των ορίων ελέγχου ως δεδομένα Φάσης II. Τα πέντε προϊόντα που προέκυψαν είχαν διαφορετικά σχήματα και μεγέθη. Και τα εννέα προϊόντα Φάσης II κατανέμονται με τρόπο που τα τέσσερα εντός ελέγχου προϊόντα τοποθετήθηκαν στις πρώτες τέσσερις θέσεις και τα πέντε προϊόντα εκτός ελέγχου τοποθετήθηκαν αργότερα με τα μεγέθη τους να αυξάνονται με την πάροδο του χρόνου.

Αυτό που παρατηρήθηκε για το διάγραμμα $Q_{ART,i}^+$ είναι ότι μπορεί να ανιχνεύσει καλά τη μετατόπιση της διαδικασίας στο πέμπτο χρονικό σημείο.

Πρέπει να σημειωθεί ότι οι ερευνητές κατασκεύασαν τέσσερα διαγράμματα, το ένα εκ των οποίων και αυτό με την καλύτερη απόδοση ήταν το $Q_{ART,i}^+$ και συνεπώς αυτός είναι ο λόγος που επιλέχθηκε.

Τέλος, όπως μας επισημαίνουν η μέθοδος που εφάρμοσαν έχει ορισμένους περιορισμούς αλλά απεδείχθη αρκετά αποτελεσματική. Για περισσότερες πληροφορίες μπορεί κάποιος να ανατρέξει στην αντίστοιχη βιβλιογραφία των Zang and Qiu (2018).

Παράρτημα

Κώδικες υπολογισμού ARL_1 για τα διαγράμματα του Κεφαλαίου 4

- **Δημιουργία τυχαίων δειγμάτων**

```
####Απαραίτητα πακέτα####  
#install.packages("jmuOutlier")  
#install.packages("reshape2")  
  
library(jmuOutlier)  
library(reshape2)  
  
#Πίνακες ARL1  
arl<-matrix(NA,10000,4);  
arl1<-matrix(NA,1,4);  
colnames(arl1)<-c("4.1","4.2","4.3","4.4");  
colnames(arl)<-c("4.1","4.2","4.3","4.4");  
  
####Προσομοίωση 10000 επαναλήψεων####  
for (z in 1:10000){  
  set.seed(z) #Σταθεροποιητής δειγμάτων  
  
  ###Δείγμα Κανονικής κατανομής###  
  mydata1<-matrix(rnorm(80,0,1),ncol=10,byrow = TRUE) #Τυπική Κανονική  
  κατανομή N(0,1)  
  mydata2<-matrix(rnorm(320,2,4),ncol=10,byrow = TRUE) #Κανονική κατανομή  
  N(2,4)  
  
  ###Δείγμα κατανομής Laplace###  
  #mydata1<-matrix(rlaplace(80,0,0.5),ncol=10,byrow = TRUE) #κατανομή  
  Laplace(0,0.5)  
  #mydata2<-matrix(rlaplace(320,1,3.2),ncol=10,byrow = TRUE) #κατανομή  
  Laplace(0,3.2)  
  
  ###Δείγμα κατανομής Βήτα###  
  #mydata1<-matrix(rbeta(8*10,shape1=0.5,shape2=0.5),ncol=10,byrow = TRUE)  
  #κατανομή Βήτα(0.5,0.5)  
  #mydata2<-matrix(rbeta(320,shape1=0.4,shape2=0.15),ncol=10,byrow =  
  TRUE) #κατανομή Βήτα(0.4,0.15)  
  
  ###Δείγμα Ομοιόμορφης κατανομής###  
  #mydata1<-matrix(runif(80,0,1),ncol=10,byrow = TRUE) # Ομοιόμορφη  
  κατανομή U(0,1)
```

```
#mydata2<-matrix(runif(320,0.25,2.1),ncol=10,byrow = TRUE)# #Ομοιόμορφη
κατανομή U(0.25,2.1)
```

```
mydata<-rbind(mydata1,mydata2) #Σύνολο δεδομένων
```

```
n<-length(mydata[1,])
m<-length(mydata[,1])
m1<-length(mydata1[,1]) #m δείγματος αναφοράς
mean(mydata) #μέση τιμή
sd(mydata)# τυπική απόκλιση
```

• Κώδικας για το διάγραμμα της ενότητας 4.1

```
###Υπολογισμός συναρτήσεων σύμφωνα με το δείγμα αναφοράς###
```

```
Xi<-matrix(NA,m1,1)
for (i in 1:m1) {
  Xi[i]=(sum(mydata[i,])/n)
}
mu2<-sum(Xi)/m1
```

```
###Υπολογισμός του Mt####
```

```
y<-matrix(NA,nrow=m,ncol=n)
M<-0
for (j in 1:n){
  for (i in 1:m) {
    y[i,j]<-mydata[i,j]-mu2
    if (y[i,j]>0) {
      y[i,j]<-1
    } else if (y[i,j]<=0) {
      y[i,j]<-0
    }
    M[[i]]<-sum(y[i,]==1)
  }
}
M
```

```
###Υπολογισμός pbar###
```

```
pbar=(sum(M[1:m1])/n)/m1;pbar
```

```
delta<-0.1
k1<-(n*delta)/2
```

```
###Υπολογισμός ορίων CL+ CL-###
```

```
delta<-0.1
k1<-(n*delta)/2
```

```

Cplus<-matrix(0,nrow=m)
Cmin<-matrix(0,nrow=m)
for (i in 1:m) {
  Cplus[[i]]<-max(0,ifelse(i==1,0,Cplus[i-1])+M[[i]]-(n*pbar+k1))
  Cmin[[i]]<-min(0,ifelse(i==1,0,Cmin[i-1])-(n*pbar-k1)+M[[i]])
}
####Υπολογισμός του ARL1####
H2<-10.65

for (i in 1:m) {
  arl[z,1]<-i
  if(Cplus[[i]]>=H2| Cmin[[i]]<=-H2) break}

```

- **Κώδικας για το διάγραμμα της ενότητας 4.2**

```

####Υπολογισμός της τιμής στόχου####
Tah<-matrix(NA,m1,1)
for (i in 1:m1) {
  Tah[i]=(sum(mydata[i,])/10)
}
mu5<-sum(Tah)/m1 #Τιμή στόχος

####Υπολογισμός του M####
Y<-matrix(0,nrow=m,ncol=n)
M<-matrix(0,nrow=m)

for (j in 1:n){
  for (i in 1:m) {
    Y[i,j]<-mydata[i,j]-mu5
    if (Y[i,j]>0) {
      Y[i,j]<-1
    } else if (Y[i,j]<=0) {
      Y[i,j]<-0
    }
    M[[i]]<-sum(Y[i,]==1)
  }
}
lamda<-0.05
k<-2.49

####Υπολογισμός της συνάρτησης EWMA####
EWMA<-matrix(0,nrow=m)
EWMA[1]<-n/2
for (i in 2:m){
  EWMA[i]<-(lamda*M[i])+((1-lamda)*(EWMA[i-1]))
}

```

```

}

####Όρια ελέγχου UCL,CL και LCL####
UCLewma<-5.63
CLewma<-n/2
LCLewma<-4.36

####Εναλλακτικός τρόπος υπολογισμού UCL,CL και LCL####
#UCLewma<-n/2+k*sqrt((lamda/(2-lamda))*(1/(4*n)))
#CLewma<-n/2
#LCLewma<-n/2-k*sqrt((lamda/(2-lamda))*(1/(4*n)))

####Υπολογισμός του ARL1####
for(i in 1:m){
  arl[z,2]<-i
  if(EWMA[i]>UCLewma | EWMA[i]==UCLewma | EWMA[i]<LCLewma |
EWMA[i]==LCLewma) break}

```

- **Κώδικας για το διάγραμμα της ενότητας 4.3**

```

####Υπολογισμός των D2 και D8####
d2 <- as.numeric(quantile(mydata1, prob = 0.20, type=5))#2nd Decile
d8 <- as.numeric(quantile(mydata1, prob = 0.80, type=5))#8th Decile

####Υπολογισμός Uij####
U<-matrix(0, nrow = m, ncol = n)
for (j in 1:n){
  for (i in 1:m){
    if (mydata[i,j] <= d2 | mydata[i,j] >= d8) {
      U[i,j] <- 1}
    else if (mydata[i,j] > d2 & mydata[i,j] < d8) {
      U[i,j] <- 0}
  }
}

####Αθροισμα των Ui####
Ui <- 0
for (j in 1:n){
  for (i in 1:m){
    Ui[[i]] <-sum(U[i,])
  }
}
psi <- 0.1; k1<-n*psi/2
p0<-0.4
####Υπολογισμός ορίων CL+ CL-####

```

```

Cplus<-matrix(0,nrow=m)
Cmin<-matrix(0,nrow=m)

for (i in 1:m) {
  Cplus[[i]]<-max(0,Ui[[i]]-(n*p0+k1)+ifelse(i==1,0,Cplus[i-1]))

  Cmin[[i]]<-max(0,(n*p0-k1)-Ui[[i]]+ifelse(i==1,0,Cmin[i-1]))
}
####Υπολογισμός του ARL1####
for (i in 1:m) {
  arl[z,3]<- i
  if(Cplus[[i]]>9.30 | Cmin[[i]]>9.30) break}

```

- **Κώδικας για το διάγραμμα της ενότητας 4.4**

```

####Υπολογισμός του Sx2####
Xbarbar<-sum(mydata1)/(m1*n)

Xibar<-matrix(NA,m1,1)
X2<-matrix(NA,m1,n)

for (i in 1:m1){
  for (j in 1:n){
    Xibar[i]<-mean(mydata1[i,])

    X2[i,j]<-(mydata1[i,j]-Xibar[i])^2
  }
}

Sx2<-(sum(X2))/(m1*(n-1));Sx2

####Υπολογισμός των X*, At και Bt####
Xstar2<-matrix(NA,m,n)

A<-matrix(NA,nrow=m)
B<-matrix(NA,nrow=m)

for (j in 1:n){
  for (i in 1:m){
    Xstar2[i,j]<-(mydata[i,j]-Xbarbar)^2
    if (Xstar2[i,j]>Sx2) {
      Xstar2[i,j]<-1
    } else
      Xstar2[i,j]<-0
  }
}

```

```

        A[[i]]<-sum(Xstar2[i,])
        B[[i]]<-asin(sqrt(A[[i]]/n))
    }
}

p0<-sum(A[1:m1]/n)/m1
p<-sum(A/n)/m

lamda<-0.22
k<-2.86

####Υπολογισμός του Dt####
D<-matrix(NA,nrow=m)
D[1]<-lamda*B[1]+(1-lamda)*asin(sqrt(p0))

for (i in 2:m){
    D[[i]]<-lamda*B[[i]]+(1-lamda)*D[i-1]
}

####Ορια ελέγχου UCL,CL και LCL####
UCLt<-(asin(sqrt(p0))+k*(sqrt((lamda/(2-lamda))*(1/(4*n))))))
CLt<-asin(sqrt(p0))
LCLt<-(asin(sqrt(p0))-k*(sqrt((lamda/(2-lamda))*(1/(4*n))))))

####Υπολογισμός του ARL1####
for(i in 1:m){
    arl[z,4]<-i
    if(D[[i]]>UCLt | D[[i]]<LCLt) break}

###Λήξη των 1000 επαναλήψεων###
z<-z+1}

###Στρογγυλοποίηση και εμφάνιση των τιμών ARL1###
arl1<-matrix(NA,1,4);
colnames(arl1)<-c("4.1","4.2","4.3","4.4");
arl1[1]<-round(mean(arl[,1]),2);arl1[2]<-round(mean(arl[,2]),2);
arl1[3]<-round(mean(arl[,3]),2);arl1[4]<-round(mean(arl[,4]),2);
arl1

#Ο κώδικας πρέπει να «τρέξει» όλος μαζί εκ νέου με τις κατάλληλες
τροποποιήσεις στην αρχή ώστε να λάβουμε τις τιμές του ARL1 για κάθε μια
από τις τέσσερις κατανομές.

```


ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Αντζουλάκος, Δ. (2016). *Στατιστικός Έλεγχος Ποιότητας, Σημειώσεις Παραδόσεων του μαθήματος Στατιστικός Έλεγχος Ποιότητας στα πλαίσια του ΠΜΣ στην «Εφαρμοσμένη Στατιστική»*, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.

Ξένα

Abu-Shawiesh, M., O., A. (2008). A simple robust control chart based on MAD, *Journal of Mathematics and Statistics*, **4**, 102-107.

Amin, R., W., Reynolds, M., R., & Bakir, S. (1995). Nonparametric Quality Control Charts Based on the Sign Statistic, *Communications in Statistics-Theory and Methods*, **24**, 1597-1623.

Amin, R., W., & Searcy, A., J. (1991). A Nonparametric Exponential Weighted Moving Average Control Scheme, *Communications in Statistics: Simulation and Computation*, **20**, 1049-1072.

Bakir, S., T. (1977). *Nonparametric Procedures for Process Control*, (Ph. D. Dissertation), Department of Statistics, Blacksburg, Virginia : Polytechnic Institute and State University.

Bakir, S., T. (2004). A Distribution-Free Shewhart Quality Control Chart Based on Signed-Ranks, *Quality Engineering*, **16**, 613-623.

Bakir, S., T., & Reynolds, M., R., Jr. (1979). A Nonparametric Procedure for Process Control Based on Within-Group Ranking, *Technometrics*, **21**, 175-183.

Boone, J., M. (2010). *Contributions to Multivariate Control Charting: Studies of the Z-Chart and four Nonparametric Charts*, Ph. D. Dissertation, University of Alabama.

Chakraborti, S., Van Der Laan, P., & Bakir, S., T. (2001). Nonparametric Control Charts: An Overview and Some Results, *Journal of Quality Technology*, **33**, 304-315.

Chakraborti, S., Van Der Laan, P., & Van De Wiel, M., A. (2004). A class of distribution-free control charts, *Applied Statistics*, **53**, 443-462.

Chen, G., Cheng, S., W., and Xie, H. (2005). A New Multivariate Control Chart for Monitoring Both Location and Dispersion, *Communications in Statistics – Simulation and Computation*, **34**, 203-217.

Cheng, S., W., & Mao, H. (2008). A Multivariate Semi-Circle Control Chart for Variabables Data, *Quality Technology and Quantitative Management*, **5**, 331–338.

- Das, N. (2009). A comparison study of three non-parametric control charts to detect shift in location parameters, *International Journal of Advanced Manufacturing Technology*, **41**, 799-807.
- Das, N. (2009). A New Multivariate Non-Parametric Control Chart Based on Sign Test, *Quality Technology and Quantitative Management*, **6**, 155-169.
- Das, N. (2011). Control Charts of Controlling Variability of Non-Normal Processes, *Economic Quality Control*, **26**, 121-131.
- Evans, B., (2012). *Practical 3D printers: The science and art of 3D printing*, New York, Apress Publisher.
- Duncan, A., J. (1974). *Quality Control and Industrial Statistics*, Richard D. Irwin: Homewood, Illinois, 4th Edition.
- Gibbons, J., D., & Chakraborti, S. (1992). *Nonparametric Statistical Inference*, 3rd ed., Marcel Dekker, New York.
- Hampel, F., R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383-393.
- Haq, A. (2017). A New Nonparametric EWMA Control Chart for Monitoring Process Variability, *Quality and Reliability Engineering International*, **33**.
- Huang, S., Yang, J., & Mukherjee, A. (2017). Nonparametric EWMA chart for simultaneous monitoring of event frequency and magnitude, *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore, 1878-1882.
- Janacek, G., J., & Meikle, S., E. (1997). Control Charts Based on Medians, *Journal of the Royal Statistical Society: Series D (The Statistician)*, **46**, 19-31.
- Langenberg, P., & Iglewicz, B. (1986). Trimmed mean X and R charts, *Journal of Quality Technology*, **18**, 152-161.
- Liu, R., Y. (1995). Control Charts for Multivariate Processes, *Journal of the American Statistical Association*, **20**, 1380-1387.
- Mathisen, H., C. (1943). A method of testing the hypothesis that two samples are from the same population, *Ann. Math. Statist.*, **14**, 188-194.
- Montgomery, D., C. (2009). *Introduction to Statistical Quality Control*, Wiley: New York, 6th Edition.
- Page, E., S. (1962). A Modified Control Chart with Warning Lines, *Biometrika*, **49**, 171-176.
- Park, C. (1985). Some Control Procedures Useful of One-Sided Asymmetrical Distributions, *Journal of the Korean Statistical Society*, **14**, 76-86.
- Rogerson, P., A., & Yamada, I. (2004). Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches, *Statistics in Medicine*, **23**, 2195-2214.
- Rousseeuw, P., J., & Croux, C. (1992). Time-efficient algorithms for two highly robust estimators of scale, *Computational Statistics*, **1**, 411-428.
- Rousseeuw, P., J., & Croux, C. (1993). Alternatives to the median absolute deviation, *Journal of the American Statistical Association*, **88**, 1273-1283.

- Shirke, D., T., & Barale, M., S. (2018). A nonparametric CUSUM chart for process dispersion, *Quality and reliability engineering international*, **34**, 858-866.
- Wolfram, S. (1996). *The Mathematica Book*, Cambridge: Cambridge University Press, 3rd Edition.
- Woodall, W., H., & Montgomery, D., C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology*, **31**, 376-386.
- Woodall, W., H. (2000). Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, **32**, 341-350.
- Wu, Z., Jiao, J., He, Z. (2009). A Control scheme for monitoring the frequency and magnitude of an event, *International Journal of Production Research*, **47**, 2887-2902.
- Yang, S., & Cheng, S. (2010). A New Non-parametric CUSUM Mean Chart, *Quality and Reliability Engineering International*, **27**, 867-875.
- Yang, S., Lin, J., & Cheng, S. (2011). A new nonparametric EWMA sign chart, *Expert Systems with Applications*, **38**, 6239-6243.
- Yue, J., Zhao, N., & Liu, L. (2020). Prediction and Monitoring Method for Breast Cancer: A Case Study for Data from the University Hospital Centre of Coimbra, *Cancer Management and Research*, **12**, 1887-1893.
- Zang, Y., & Qiu., P. (2018). Phase II monitoring of free-form surfaces: An application to 3D printing, *Journal of Quality Technology*, **50**, 379-390.
- Zombade, D., M., (2019). *Nonparametric Quality Control Techniques*, (Doctoral Dissertation), Solapur : Solapur University, India.