



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ
ΚΑΙ ΔΙΟΙΚΗΣΗΣ

«Ανάπτυξη διαγραμμάτων ελέγχου
ποιότητας για δεδομένα ικανοποίησης
πελατών»

Διπλωματική εργασία

*Μαρία Κουτσούκου
ΑΜ:2016010011*

Επιβλέπων καθηγητής : Καθ. Ευάγγελος Γρηγορούδης

Περιεχόμενα

ΕΥΧΑΡΙΣΤΙΕΣ	5
ΠΕΡΙΛΗΨΗ.....	6
Κεφάλαιο 1 ^ο : Εισαγωγή.....	7
1.1 Γενικά	7
1.2 Έλεγχος δεδομένων ικανοποίησης πελατών	8
1.3 Δομή και στόχοι εργασίας.....	9
Κεφάλαιο 2 ^ο : Εισαγωγή στον στατιστικό έλεγχο ποιότητας	11
2.1 Γενικά	11
2.2 Αιτίες Μεταβλητότητας.....	12
2.3 Γενικές αρχές διαγραμμάτων ελέγχου	14
2.4 Βασική θεωρία διαγραμμάτων ελέγχου	16
2.5 Επιλογή ορίων ελέγχου	17
2.6 Μέσο μήκος ροής (Average Run Length).....	19
Κεφάλαιο 3 ^ο : Διαγράμματα ελέγχου για δεδομένα ικανοποίησης πελατών	22
3.1 Γενικά	22
3.2 Εναλλακτικά διαγράμματα ελέγχου.....	23
3.2.1 Τυπικά διαγράμματα ελέγχου Shewhart	23
3.2.2 Επέκταση του p διαγράμματος.....	27
3.2.3 Διάγραμμα ελέγχου X^2	28
3.3 Κατανομή Johnson.....	31
3.4 Μετασχηματισμός δεδομένων	36
3.4.1 Απλοί μετασχηματισμοί.....	36
3.4.2 Box-Cox μετασχηματισμός.....	36
Κεφάλαιο 4 ^ο : Εφαρμογή εναλλακτικών διαγραμμάτων ελέγχου.....	40
4.1 Σχεδιασμός προσομοίωσης	40
4.2 Δημιουργία δεδομένων	41
4.3 Παρουσίαση αποτελεσμάτων.....	42
4.4 Τυπικά διαγράμματα ελέγχου Shewhart	44
4.5 Διάγραμμα ελέγχου επέκτασης του X_p	46
4.5 Διάγραμμα ελέγχου X^2	48
4.6 Μετασχηματισμός δεδομένων	50
4.6.1 Απλοί μετασχηματισμοί.....	50
4.6.2 Μετασχηματισμός Box-Cox.....	58

Κεφάλαιο 5ο: Συμπεράσματα.....	62
5.1 Σύνοψη αποτελεσμάτων.....	62
5.2 Μελλοντικές επεκτάσεις	67
Βιβλιογραφία.....	68
Ηλεκτρονικές πηγές.....	69
Παράρτημα Α :Παράμετροι κατανομής Johnson	70
Παράρτημα Β :Πίνακας κατανομής χ^2	72

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της συγκεκριμένης διπλωματικής εργασίας θα ήθελα να ευχαριστήσω όσους ήταν δίπλα μου και με στήριξαν σε όλη την διάρκεια της, τόσο την οικογένεια μου αλλά και τους φίλους μου με τις χρήσιμες συμβουλές τους.

Επίσης θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Ευάγγελο Γρηγορούδη, Καθηγητή της Σχολής Μηχανικών Παραγωγής και Διοίκησης του Πολυτεχνείου Κρήτης για την συνεργασία που είχαμε και την καθοδήγηση του σε όλο το διάστημα εκπόνησης της διπλωματικής εργασίας.

Τέλος θα ήθελα να ευχαριστήσω τον διδακτορικό φοιτητή Ευστράτιο Καρτσωνάκη για την πολύτιμη βοήθεια του και την άψογη συνεργασία μας.

ΠΕΡΙΛΗΨΗ

Τα διαγράμματα στατιστικού ελέγχου διεργασιών αναπτύχθηκαν από τις αρχές του 1930, όπου η πρώτη εφαρμογή τους ήταν στον κατασκευαστικό/βιομηχανικό τομέα. Με την πάροδο των χρόνων οι εφαρμογές των διαγραμμάτων ελέγχου εξελίχθηκαν και σε άλλους τομείς πέρα από την παραγωγή. Σημαντικός παράγοντας ήταν πως μέσα από την εφαρμογή τους οι επιχειρήσεις μπορούν πλέον να προσδιορίσουν τι ικανοποιεί και τι δυσαρεστεί τους πελάτες και να εφαρμόσουν κατάλληλα μέτρα για βελτίωση προκειμένου να εξασφαλιστεί η ποιότητα.

Στην παρούσα διπλωματική εργασία δημιουργήθηκαν διαγράμματα ελέγχου για δεδομένα ικανοποίησης πελατών προκειμένου να ελεγχθεί ο έλεγχος ποιότητας ενός προϊόντος ή υπηρεσίας, αλλά και να εντοπιστούν τυχόν αλλαγές στο εσωτερικό ή εξωτερικό περιβάλλον. Σκοπός της διπλωματικής εργασίας είναι επίσης και ο έλεγχος των διαγραμμάτων αυτών για τα δεδομένα που προέρχονται από έρευνες ικανοποίησης. Τα δεδομένα που χρησιμοποιήθηκαν κατασκευάστηκαν τυχαία, ενώ πρέπει να σημειωθεί ότι τα συγκεκριμένα δεδομένα δεν ακολουθούν κανονική κατανομή. Για τον λόγο αυτό κατασκευάστηκαν τρία είδη διαγραμμάτων ελέγχου που μπορούν να εφαρμοστούν καλύτερα με δεδομένα ικανοποίησης πελατών. Αρχικά κατασκευάστηκε το Shewhart \bar{X} διάγραμμα, το \bar{X}_p ως μία επέκταση του \bar{X} Shewhart chart και το χ^2 διάγραμμα ελέγχου. Τέλος εφαρμόστηκε ο μετασχηματισμός δεδομένων με συναρτήσεις κατάλληλες για τα δεδομένα ικανοποίησης.

Τέλος με την χρήση του μέσου μήκους ροής (ARL) συγκρίνουμε τα διαγράμματα και μπορούμε να δούμε ποιο διάγραμμα εντόπισε την αλλαγή στην διεργασία αλλά και πόσο γρήγορα την εντόπισε. Έτσι θα καταλήξουμε στο διάγραμμα ελέγχου που μπορεί να αναλύσει καλύτερα τα δεδομένα μας και να δώσει τα καλύτερα αποτελέσματα. Μέσα από την κατάλληλη επιλογή διαγραμμάτων ελέγχου μπορούμε να έχουμε τόσο άμεσα και έμμεσα οφέλη για μία επιχείρηση που κατά επέκταση επηρεάζουν και την ικανοποίηση του πελάτη.

Κεφάλαιο 1^ο: Εισαγωγή

1.1 Γενικά

Τα τελευταία χρόνια ο όρος ποιότητα κυριαρχεί στις ζωές μας. Ζούμε σε μια εποχή όπου ο ανταγωνισμός μεταξύ των επιχειρήσεων είναι ιδιαίτερα αυξημένος και επηρεάζεται τόσο από την τιμή, αλλά και από την ποιότητα προϊόντος ή υπηρεσίας. Η ποιότητα ενός προϊόντος συνδέεται άμεσα με την ικανοποίηση ή όχι του πελάτη. Σύμφωνα με το βραβείο ποιότητας Malcom Baldrige, η ποιότητα του πελάτη ορίζεται ως βασική αξία. Μάλιστα από τις επτά κατηγορίες των βραβείων Baldrige του 1996, η εστίαση στον πελάτη και η ικανοποίηση των πελατών έλαβαν και συνεχίζουν να λαμβάνουν το μεγαλύτερο βάρος. Συνεπώς σκοπός μιας επιχείρησης είναι να εντοπίσει τι ικανοποιεί τον πελάτη και τι όχι. Η μέτρηση της ικανοποίησης του πελάτη μπορεί να βοηθήσει στην χάραξη μιας στρατηγικής και στην λήψη αποφάσεων μιας επιχείρησης. Μια επιχείρηση προκειμένου να ανταπεξέλθει στην νέα πραγματικότητα θα πρέπει να διεξάγει έρευνα ώστε να εντοπίσει τυχόν αλλαγές στο εξωτερικό περιβάλλον αλλά και στο εσωτερικό περιβάλλον της επιχείρησης. Μέσα από την έρευνα αυτή, μπορεί να λάβει νέα μέτρα και αποφάσεις που θα τη βοηθήσουν να αυξήσει την ανταγωνιστικότητα της απέναντι σε άλλες επιχειρήσεις και να οδηγήσουν σε αύξηση του κέρδους της και του μεριδίου αγοράς.

Συνεπώς οι κυριότεροι λόγοι μέτρησης ικανοποίησης του πελάτη είναι οι εξής:

- Η ικανοποίηση του πελάτη μας δίνει πληροφορίες για την τρέχουσα εικόνα της αγοράς και πως μπορεί να κινηθεί μελλοντικά η επιχείρηση.
- Η μέτρηση ικανοποίησης των πελατών μπορεί να αναδείξει ευκαιρίες στην συγκεκριμένη αγορά.
- Η διασφάλιση ποιότητας μιας επιχείρησης βασίζεται σε μια διαδικασία συνεχούς βελτίωσης, γεγονός το οποίο απαιτεί την ύπαρξη της διαδικασίας μέτρησης της ικανοποίησης. Έτσι η επιχείρηση είναι σε θέση να γνωρίζει τις ανάγκες των πελατών και να πραγματοποιήσει ενέργειες βελτίωσης.
- Η μέτρηση ικανοποίησης μπορεί να βοηθήσει την επιχείρηση να κατανοήσει καλύτερα την συμπεριφορά των πελατών ,να αναλύσουν τις επιθυμίες του και τις ανάγκες του .
- Το πρόβλημα της ύπαρξης διαφορετικής αντίληψης της ικανοποίησης ανάμεσα στον πελάτη και τη διοίκηση της εταιρείας μπορεί να προσδιοριστεί από την υλοποίηση ενός προγράμματος μέτρησης της ικανοποίησης. Με αυτόν τον τρόπο δίνεται η δυνατότητα να αμβλυνθούν αυτές οι διαφορές αντίληψης .

1.2 Έλεγχος δεδομένων ικανοποίησης πελατών

Μια έρευνα ικανοποίησης περιλαμβάνει τη συλλογή και την ανάκτηση όλων των δεδομένων από το περιβάλλον της έρευνας. Τα δεδομένα αυτά μπορούν να αναλυθούν μέσω του στατιστικού ελέγχου ποιότητας. Η πληρότητα των δεδομένων είναι μια απαίτηση η οποία δεν μπορεί πάντα να εξασφαλιστεί. Αυτό το γεγονός μπορεί να οφείλεται σε πλήθος παραγόντων σχεδιασμού, ανθρωπίνου δυναμικού, εξοπλισμού, λανθασμένης εκτέλεσης της έρευνας ή και εσκεμμένη απώλεια δεδομένων. Συνεπώς η απώλεια των δεδομένων μπορεί να οδηγήσει σε εσφαλμένα αποτελέσματα και να δημιουργηθεί αλλοίωση της πραγματικότητας της επιχείρησης που θέλουμε να μελετήσουμε. Επίσης σε περιπτώσεις όπου η έρευνα αφορά δεδομένα από την πραγματική ζωή, τα δεδομένα αυτά δεν είναι κανονικά κατανομημένα. Πολλές περιπτώσεις μάλιστα, δεν ακολουθούν κανονικές κατανομές περιπτώσεις όπως ο χρόνος κύκλου, οι κλήσεις πελατών, ο χρόνος αναμονής του πελάτη. Η εφαρμογή των στατιστικών διαγραμμάτων ελέγχου σε τέτοιες εφαρμογές μπορεί να οδηγήσει σε πιθανές δυσκολίες ή και λανθασμένα αποτελέσματα. Στις περιπτώσεις αυτές θα πρέπει να προσδιορίζεται η αιτία μη κανονικότητας και να λαμβάνονται τα κατάλληλα διορθωτικά μέτρα .

Συχνοί λόγοι για μη κανονικότητα των δεδομένων είναι :

- *Ακραίες τιμές:* Οι ακραίες τιμές μπορούν να παραμορφώσουν την κατανομή. Ο μέσος όρος της κατανομής είναι αρκετά ευαίσθητος σε ακραίες τιμές, οι οποίες μπορούν να οδηγήσουν σε μη κανονική κατανομή. Θα πρέπει σε μια έρευνα να προσδιοριστούν όλες οι ακραίες τιμές αρκετά μεγάλες ή αρκετά μικρές στο σύνολο των δεδομένων και να εξεταστούν ειδικές αιτίες στην διεργασία.
- *Υποομάδες/Επικάλυψη δύο ή περισσότερων διεργασιών:* Ένα σύνολο δεδομένων το οποίο είναι συνδυασμός δύο ή περισσότερων συνόλων δεδομένων που προέρχονται από δύο ή περισσότερες διεργασίες μπορούν επίσης να οδηγήσουν σε μια μη κανονική κατανομή δεδομένων. Αυτό μπορεί να οφείλεται στην πραγματικότητα σε δύο διαφορετικές διεργασίες, όπως δύο χειριστές ή δύο βάρδιες ή μια διαδικασία που αλλάζει συχνά.
- *Ανεπαρκής διάκριση δεδομένων:* Τα σφάλματα στρογγυλοποίησης ή οι συσκευές μέτρησης με χαμηλή ακρίβεια μπορούν να κάνουν τα πραγματικά συνεχή και κανονικά δεδομένα να φαίνονται διακριτά και μη κανονικά. Η χρήση ενός πιο ακριβούς συστήματος μέτρησης ή η συλλογή περισσότερων σημείων δεδομένων πρέπει να γίνεται για να ξεπεραστούν οι ανεπαρκείς διακρίσεις δεδομένων ή ο ανεπαρκής αριθμός διαφορετικών τιμών.
- *Μικρό μέγεθος δείγματος:* Αυτό μπορεί να οδηγήσει σε διάσπαρτη εμφάνιση κανονικά κατανομημένων δεδομένων.

- *Τιμές κοντά στα όρια ή στο μηδέν:* Εάν σε μια διεργασία οι τιμές βρίσκονται κοντά στο μηδέν ή κοντά στα όρια τότε η κατανομή που θα προκύψει θα είναι λοξά στα αριστερά.
- *Ταξινομημένα δεδομένα:* Σε περιπτώσεις όπου τα δεδομένα που συλλέγουμε από μία κανονικά κατανεμημένη διεργασία μπορούν να αντιστοιχούν και σε μία μη κανονικά κατανεμημένη εάν αντιπροσωπεύουν μόνο ένα δείγμα από την συνολική παραγωγική διαδικασία. Αυτό συμβαίνει όταν τα δεδομένα που συλλέγονται ταξινομούνται και στην συνέχεια αναλύονται.
- *Τα δεδομένα ακολουθούν διαφορετική κατανομή:* Εκτός από τους παραπάνω λόγους που αναφέρθηκαν όπου τα δεδομένα από μια διεργασία που κατανέμονται κανονικά μπορούν να εμφανιστούν ως μη κανονικά, υπάρχουν πολλοί τύποι δεδομένων που ακολουθούν μη κανονική κατανομή από την φύση τους. Σε τέτοιες περιπτώσεις τα δεδομένα αυτά θα πρέπει να αναλύονται με την χρήση διαφορετικών εργαλείων που δεν υποθέτουν κανονικότητα.

1.3 Δομή και στόχοι εργασίας

Η παρούσα διπλωματική εργασία αποσκοπεί στη διερεύνηση της χρήσης εναλλακτικών διαγραμμάτων ελέγχου για δεδομένα ικανοποίησης πελατών. Τα δεδομένα ικανοποίησης πελατών δεν ακολουθούν πάντα κανονική κατανομή. Συνεπώς τα διαγράμματα ελέγχου Shewhart δεν μπορούν να χρησιμοποιηθούν διότι μπορούν να οδηγήσουν σε εσφαλμένα αποτελέσματα σχετικά με τα επίπεδα ικανοποίησης και να μην εντοπίζουν τυχόν αλλαγές στο εσωτερικό αλλά και στο εξωτερικό περιβάλλον.

Στο πρώτο κεφάλαιο της διπλωματικής εργασίας γίνεται αναφορά σε βασικές έννοιες στον στατιστικό έλεγχο ποιότητας, ως η γνωστότερη μέθοδος ελέγχου διεργασιών και παρουσιάζονται οι σημαντικότερες τεχνικές που χρησιμοποιούνται. Η χρήση του στατιστικού ελέγχου ποιότητας έχει σκοπό τη βελτίωση ποιότητας ενός προϊόντος ή υπηρεσίας. Στο κεφάλαιο αυτό παρουσιάζονται και άλλες έννοιες που σχετίζονται με τον στατιστικό έλεγχο ποιότητας, όπως οι αιτίες μεταβλητότητας μιας διεργασίας και οι βασικές αρχές ενός διαγράμματος ελέγχου. Συγκεκριμένα παρουσιάζονται οι θετικές επιπτώσεις ενός διαγράμματος ελέγχου σε μια παραγωγική διαδικασία/διεργασία αλλά και τελικά στην επιχείρηση. Επίσης αναλύεται η βασική θεωρία των διαγραμμάτων, οι σχέσεις που χρησιμοποιούνται αλλά και πως υπολογίζονται τα όρια ελέγχου. Τέλος, γίνεται αναφορά στο μέσο μήκος ροής (ARL), πως υπολογίζεται αλλά και τι συμβολίζει για μια παραγωγική διαδικασία.

Στο δεύτερο κεφάλαιο της διπλωματικής εργασίας θα παρουσιαστούν τα διαγράμματα ελέγχου που θα χρησιμοποιηθούν για να αναπαραστήσουν τα δεδομένα ικανοποίησης πελατών. Αρχικά θα γίνει αναφορά γιατί η χρήση

συμβατικών διαγραμμάτων ελέγχου Shewhart οδηγεί σε εσφαλμένα αποτελέσματα και έτσι θα παρουσιαστούν τα εναλλακτικά διαγράμματα ελέγχου για δεδομένα ικανοποίησης πελατών. Θα παρουσιαστούν οι βασικές αρχές των διαγραμμάτων ελέγχου R , του X^2 και μια επέκταση του διαγράμματος ελέγχου p . Επίσης θα γίνει αναφορά και σε άλλες μεθοδολογίες που χρησιμοποιούνται για μετατροπή μη - κανονικών δεδομένων σε κανονικά όπως η κατανομή Johnson και μετασχηματισμός δεδομένων με τον μετασχηματισμό Box-Cox .

Στο επόμενο κεφάλαιο θα γίνει η εφαρμογή κάποιων εναλλακτικών διαγραμμάτων ελέγχου για τα δεδομένα ικανοποίησης πελατών, τα οποία παρουσιάζουν αρνητική ασυμμετρία. Τα δεδομένα αυτά έχουν δημιουργηθεί με τυχαίο τρόπο σε περιβάλλον Excel με χρήση της κατανομής Βήτα, αφού έχει γίνει καθορισμός των κατάλληλων παραμέτρων ώστε να μπορούμε να παρατηρήσουμε την μεταβολή των διαγραμμάτων ελέγχου .

Στο τελευταίο κεφάλαιο της διπλωματικής παρουσιάζονται τα συμπεράσματά που προέκυψαν από την σύγκριση των διαγραμμάτων ελέγχου. Γίνεται σύγκριση τόσο γραφικά αλλά και υπολογίζοντας τα σφάλματα πρώτου και δευτέρου είδους ώστε να υπολογιστεί το μέσο μήκος ροής (ARL). Μέσα από τον υπολογισμό του ARL θα γίνει σύγκριση των γραφημάτων και θα μελετηθούν χαρακτηριστικά όπως πόσο γρήγορα ένα διάγραμμα εντοπίζει τις αλλαγές στο εσωτερικό ή εξωτερικό περιβάλλον του. Τέλος γίνεται αναφορά σε μελλοντικές επεκτάσεις των διαγραμμάτων ελέγχου για την παρακολούθηση δεδομένων ικανοποίησης πελατών.

Κεφάλαιο 2ο: Εισαγωγή στον στατιστικό έλεγχο ποιότητας

2.1 Γενικά

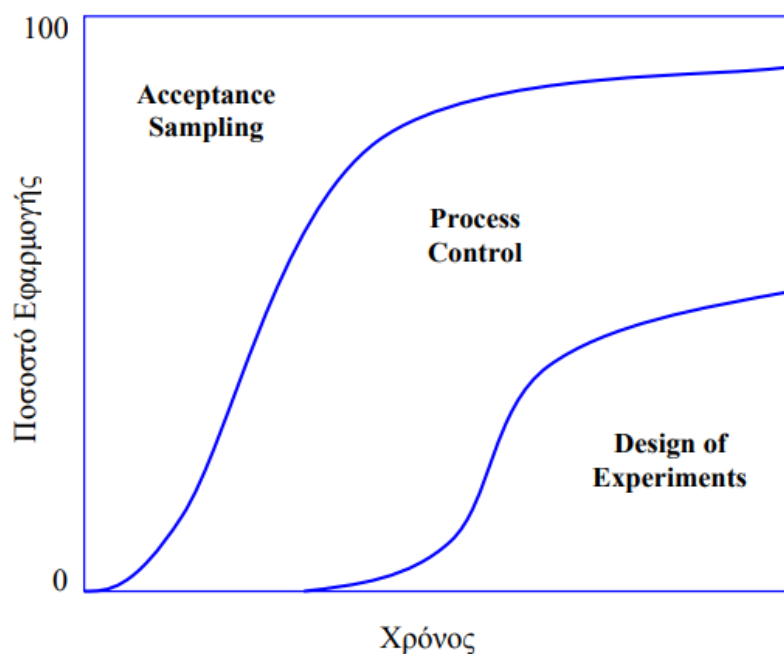
Ο στατιστικός έλεγχος ποιότητας είναι η γνωστότερη μέθοδος ελέγχου των διεργασιών με σκοπό τη βελτίωση της ποιότητας. Ο στατιστικός έλεγχος περιέχει ένα συγκεκριμένο αριθμό στατιστικών τεχνικών όπου οι σημαντικότερες είναι :

- Ο Στατιστικός Έλεγχος Διαδικασίας(Σ.Ε.Δ, Statistical Process Control)
- Η Δειγματοληψία Αποδοχής (Δ.Α., Acceptance Sampling)
- Ο Σχεδιασμός και η Ανάλυση Πειραμάτων (Design of experiments) .

Τα εργαλεία αυτά δίνουν τη δυνατότητα στο σύστημα ποιοτικού ελέγχου να διακρίνει πιο εύκολα και με μικρότερο κόστος τα προϊόντα εκείνα που δεν συμμορφώνονται με τις προδιαγραφές που έχουν οριστεί και με αυτόν τον τρόπο να πραγματοποιηθούν ενέργειες για διόρθωση. Επίσης μέσω του ποιοτικού ελέγχου γίνεται ο καθορισμός των ορίων των χαρακτηριστικών ποιότητας των προϊόντων και με αυτόν τον τρόπο μπορούμε να μετρήσουμε την απόδοση της παραγωγικής διαδικασίας. Για την επίτευξη της ποιότητας και της διασφάλισής της είναι σημαντικό η συνεχής έρευνα αλλά και η εκπαίδευση του προσωπικού στις κατάλληλες τεχνικές που ταιριάζουν καλύτερα με την παραγωγική διαδικασία.

Ο σχεδιασμός και η ανάλυση πειραμάτων περιέχει εκείνες τις στατιστικές τεχνικές οι οποίες μας βοηθούν στον εντοπισμό διαφόρων παραγόντων στις ποιοτικές παραμέτρους του τελικού προϊόντος. Η δειγματοληψία αποδοχής περιέχει στατιστικές τεχνικές που είναι απαραίτητες για να αποφασίσουμε αν θα απορριφθεί ή όχι μια παρτίδα. Ο στατιστικός έλεγχος διεργασιών είναι το σύνολο στατιστικών τεχνικών που χρησιμοποιούνται για τον έλεγχο της παραγωγικής διαδικασίας κατά τη διάρκειά της. Ο στατιστικός έλεγχος διεργασιών χρησιμοποιεί 7 κύρια εργαλεία τα οποία είναι: ιστόγραμμα, φύλλο ελέγχου, διάγραμμα Pareto, διάγραμμα αιτίας αποτελέσματος, διάγραμμα συγκέντρωσης ελαττωμάτων, διάγραμμα διασποράς. Ωστόσο το πιο σημαντικό είναι το διάγραμμα ελέγχου (Control Chart) .

Τα μοντέρνα συστήματα διασφάλισης της ποιότητας εστιάζονται σε στατιστικές τεχνικές όπως ο στατιστικός έλεγχος διεργασιών και ο σχεδιασμός πειραμάτων παρά στη δειγματοληψία αποδοχής, καθώς δεν υπάρχει ανατροφοδότηση στο σύστημα με κατάλληλες πληροφορίες ώστε να οδηγήσουν σε βελτίωση.



Σχήμα 2.1: Διάγραμμα φάσεων χρήσης στατιστικών μεθόδων.

2.2 Αιτίες Μεταβλητότητας

Σε κάθε διεργασία παραγωγής, ανεξάρτητα από το πόσο καλά σχεδιασμένη είναι ή πόσο καλά μελετημένη ή έχει συντηρηθεί, θα υπάρχουν μικρές διαφορές στα χαρακτηριστικά ποιότητας των παραγόμενων προϊόντων ή υπηρεσιών. Σύμφωνα με τον Shewhart υπάρχουν δύο είδη αιτιών που προκαλούν μεταβλητότητα:

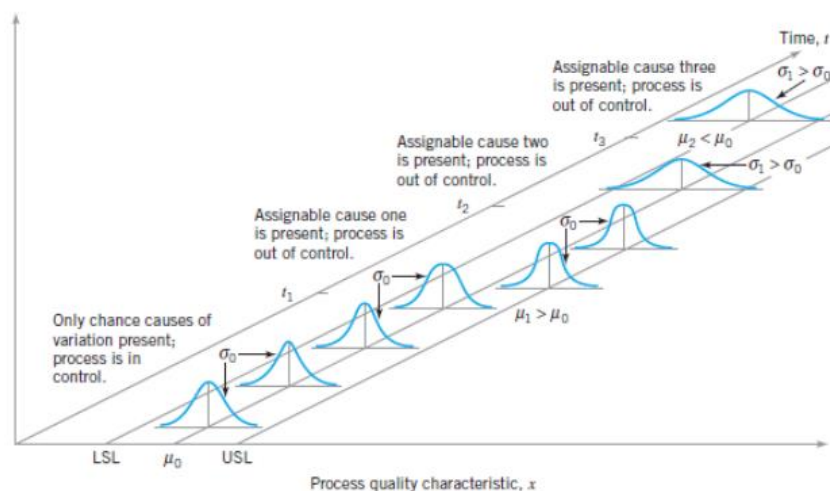
- i. Τα συνήθη αίτια (ή «τυχαία αίτια» ή «φυσιολογικές τυχαίες διακυμάνσεις» ή «τυχαίες διακυμάνσεις») είναι αναπόφευκτα αίτια σε μια παραγωγική διαδικασία. Τα αίτια αυτά οφείλονται σε ενδογενείς παράγοντες της παραγωγικής διαδικασίας, όπως είναι αλλαγές εξωτερικού περιβάλλοντος ή μικρές αποκλίσεις πρώτης ύλης και είναι αδύνατο να απαλειφθούν.
- ii. Τα ειδικά αίτια είναι αίτια τα οποία θα οδηγήσουν σε δημιουργία προβλήματος στην παραγωγική διαδικασία, καθώς προέρχονται από εξωτερικούς παράγοντες. Η παρουσία τέτοιων αιτιών μπορεί να οδηγήσει σε σημαντική απόκλιση και έτσι η παραγωγική διαδικασία να είναι «εκτός ελέγχου».

Για τον σχεδιασμό ενός προϊόντος είναι σημαντικό να καθοριστούν τα όρια ελέγχου για τα ποιοτικά χαρακτηριστικά του προϊόντος. Τα όρια αυτά είναι το άνω και το κάτω όριο ελέγχου ή αντίστοιχα UCL και LCL και ανάμεσα σε αυτά τα όρια πρέπει να βρίσκονται οι τιμές του προϊόντος ώστε το τελικό προϊόν να είναι αποδεχτό και υπό έλεγχο. Σε αυτήν την περίπτωση οι αιτίες μεταβλητότητας είναι τυχαίες. Σε αντίθετη

περίπτωση όταν δηλαδή τα σημεία βρίσκονται εκτός του άνω ή του κάτω ορίου ή παρουσιάζουν μια ανωμαλία η παραγωγική διαδικασία βρίσκεται εκτός ελέγχου. Σε αυτήν την περίπτωση οι αιτίες μεταβλητότητας οφείλονται σε ειδικά αίτια .

Στην περίπτωση ειδικής μεταβλητότητας μπορεί να υπάρξουν αλλαγές στις παραμέτρους της κατανομής που ακολουθούν οι τιμές του ποιοτικού χαρακτηριστικού του προϊόντος. Στην περίπτωση της κανονικής κατανομής η επίδραση μιας συστημικής αιτίας έχει ως αποτέλεσμα την μετατόπιση του μέσου της κατανομής ή της διακύμανσης της. Τα διαγράμματα ελέγχου του Shewhart περιέχουν όρια γύρω από την κανονική κατανομή των δεδομένων, η οποία είναι η επίδραση μόνο τυχαίων αιτιών στην παραγωγική διαδικασία. Κάθε παρατήρηση η οποία πέφτει εκτός ορίων οφείλεται σε κάποιο ειδικό αίτιο στην παραγωγική διαδικασία.

Στο Σχήμα 2.2 φαίνεται η επίδραση των ειδικών αιτιών μεταβλητότητας στην τιμή μ_0 και στην τυπική απόκλιση σ_0 σε διαφορετικές χρονικές στιγμές. Συγκεκριμένα την χρονική στιγμή t_1 η μέση τιμή μετατοπίζεται στην θέση $\mu_1 > \mu_0$, ενώ την χρονική στιγμή t_2 η μέση τιμή μετατοπίζεται στην θέση μ_0 ενώ η τυπική απόκλιση στην θέση $\sigma_1 > \sigma_0$ και την χρονική στιγμή t_3 ισχύει ότι $\mu_2 < \mu_0$.



Σχήμα 2.2: Επίδραση ειδικών αιτιών στην καμπύλη της κανονικής κατανομής.

Έτσι σκοπός του στατιστικού ελέγχου διεργασιών είναι η έγκαιρη ανίχνευση της ειδικής μεταβλητότητας σε μια παραγωγική διαδικασία ώστε να πραγματοποιηθούν οι απαραίτητες διορθωτικές κινήσεις έγκαιρα και να μην κατασκευαστούν αρκετά προϊόντα μη συμμορφούμενα με τις προδιαγραφές ποιότητας. Επίσης η ανεύρεση και η απομάκρυνση των ειδικών αυτών αιτιών από τη διεργασία οδηγεί σε μείωση της διασποράς στις τιμές των χαρακτηριστικών ποιότητας και κατά επέκταση

καλύτερη ποιότητα και οικονομία. Τα διαγράμματα ελέγχου είναι μια ευρέως διαδεδομένη τεχνική που χρησιμοποιείται για τον εντοπισμό των αιτιών μεταβλητότητας σε πραγματικό χρόνο σε μια παραγωγική διαδικασία (on-line process-monitoring), ώστε να αποφεύγεται η λειτουργία της διαδικασίας εκτός ελέγχου για μεγάλο χρονικό διάστημα, αλλά και η παραγωγή μεγάλου όγκου προϊόντων μη αποδεχτής ποιότητας. Ωστόσο η βελτίωση μιας παραγωγικής διαδικασίας είναι ευθύνη των ανθρώπων που την χειρίζονται και δεν μπορεί να προκύψει από τα διαγράμματα ελέγχου

2.3 Γενικές αρχές διαγραμμάτων ελέγχου

Όπως αναφέρθηκε και παραπάνω, ο στατιστικός έλεγχος διεργασιών είναι μια μεθοδολογία για παρακολούθηση της παραγωγικής διαδικασίας με σκοπό να εντοπιστούν οι ειδικές αιτίες μεταβλητότητας και να αντιμετωπιστούν με τις κατάλληλες διορθωτικές ενέργειες. Έτσι η διεργασία θα μπορέσει να τεθεί υπό στατιστικό έλεγχο και να απαλλαγεί από ειδικές αιτίες και τις επιδράσεις τους. Τα διαγράμματα ελέγχου είναι το σημαντικότερο εργαλείο στατιστικού ελέγχου. Τα διαγράμματα ελέγχου πέρα από τον εντοπισμό των ειδικών αιτιών χρησιμοποιούνται και στην ανάλυση των δυνατοτήτων της παραγωγικής διαδικασίας και στην επεξεργασία μετρήσεων για διαγνωστικούς σκοπούς. Η χρήση των διαγραμμάτων ελέγχου ξεκίνησε από βιομηχανίες παραγωγής και μεταποίησης προϊόντων και εξελίχθηκε και σε εταιρείες παροχής υπηρεσιών. Σε κάθε επιχείρηση το σημαντικότερο για την χρήση διαγραμμάτων ελέγχου είναι ο προσδιορισμός των κατάλληλων χαρακτηριστικών ποιότητας των προϊόντων ή υπηρεσιών. Τα διαγράμματα ελέγχου απέκτησαν ευρεία εφαρμογή καθώς αποφέρουν τόσο άμεσα και έμμεσα οφέλη σε μια επιχείρηση. Κάποια οφέλη είναι :

- Τα διαγράμματα ελέγχου είναι ένας αποδεδειγμένος τρόπος για τη βελτίωση της απόδοσης και της ποιότητας της παραγωγικής διαδικασίας. Αυτό επιτυγχάνεται μέσω της διάγνωσης προβλημάτων που οφείλεται σε κακό σχεδιασμό της διαδικασίας.
- Τα διαγράμματα ελέγχου μπορούν να εντοπίσουν πιθανά ελαττώματα και να προβούν σε κατάλληλες διορθωτικές ενέργειες. Είναι εργαλεία που βοηθούν να διατηρηθεί ο στατιστικός έλεγχος.
- Τα διαγράμματα ελέγχου βοηθούν στο να αποφευχθούν οι υπερβολικά συχνές και μη απαραίτητες επεμβάσεις στην παραγωγική διεργασία. Αυτό συμβαίνει καθώς μπορούν να διακρίνουν μεταξύ τυχαίων και ειδικών αιτιών.
- Τα διαγράμματα ελέγχου μπορούν να ελαττώσουν το κόστος εσωτερικών αστοχιών καθώς και στην αύξηση της παραγωγικής διαδικασίας. Το γεγονός αυτό οφείλεται στον έγκαιρο εντοπισμό των προβλημάτων της παραγωγικής

διαδικασίας και σε οποιαδήποτε αλλαγή με αποτέλεσμα την αποφυγή ελαττωματικών προϊόντων.

- Τα διαγράμματα ελέγχου δίνουν χρήσιμες πληροφορίες που μπορούν να οδηγήσουν σε συγκεκριμένα συμπεράσματα. Τα συμπεράσματα αυτά μπορούν να προκύψουν από παρατηρήσεις σε μοτίβα σημείων. Έτσι, από συστηματική παρακολούθηση διαδικασιών, καταγραφή μετρήσεων και δημιουργία αρχείου από πληροφορίες αξιολόγησης διαδικασιών, μπορούν να προκύψουν χρήσιμα συμπεράσματα που θα βοηθήσουν στην βελτίωση της απόδοσης της διεργασίας.
- Τα διαγράμματα ελέγχου παρέχουν αξιόπιστη ανάλυση των δυνατοτήτων της παραγωγικής διαδικασίας. Μέσω των διαγραμμάτων μπορούμε να πάρουμε τιμές σημαντικών παραμέτρων και να αναλύσουμε την σχέση τους με τον χρόνο. Έτσι μπορούν να προκύψουν δυνατότητες της διεργασίας, το οποίο θα συμβάλει στην βελτίωση της.

Τα διαγράμματα ελέγχου χρησιμοποιούνται κυρίως για:

- Τη δημιουργία ενός εργαλείου στατιστικού ελέγχου της παραγωγικής διαδικασίας.
- Την παρακολούθηση της διεργασίας και την προειδοποίηση όταν βρεθεί εκτός στατιστικού ελέγχου.
- Τον καθορισμό των δυνατοτήτων της διεργασίας.

Παρακάτω παρουσιάζονται τα βήματα για την ανάπτυξη διαγραμμάτων ελέγχου:

1. Προετοιμασία

- Επιλογή του χαρακτηριστικού ποιότητας είτε μεταβλητή (variable) είτε ιδιότητα (attribute).
- Καθορισμός του μεγέθους και της συχνότητας της δειγματοληψίας.
- Επιλογή κατάλληλου διαγράμματος

2. Συλλογή δεδομένων

- Καταγραφή δεδομένων.
- Υπολογισμός στατιστικών μεγεθών : μέση τιμή, εύρος τιμών.
- Σύστημα αξόνων.

3. Καθορισμός ορίων ελέγχου

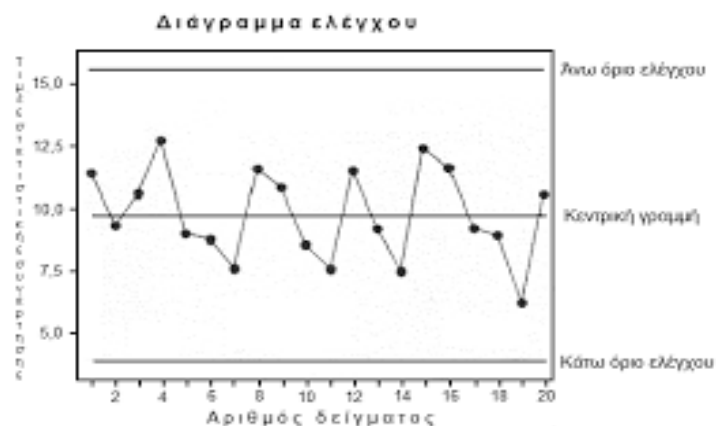
- Σχεδιασμός κεντρικής γραμμής του διαγράμματος ελέγχου.
- Υπολογισμός άνω και κάτω ορίου ελέγχου.

4. Ανάλυση και ερμηνεία

- Εντοπισμός σημείων που βρίσκονται εκτός ορίων ελέγχου.
 - Περιορισμός σημείων που βρίσκονται εκτός ελέγχου.
5. Χρήση ενός εργαλείου -τεχνική επίλυση προβλημάτων
- Αναγνώριση περιπτώσεων εκτός στατιστικού ελέγχου και λήψη κατάλληλων διορθωτικών ενεργειών.
6. Προσδιορισμός δυνατοτήτων της διεργασίας από τα δεδομένα του διαγράμματος ελέγχου .

2.4 Βασική θεωρία διαγραμμάτων ελέγχου

Βασικός σκοπός των διαγραμμάτων ελέγχου είναι ο εντοπισμός συστηματικών μεταβολών των παραμέτρων της παραγωγικής διαδικασίας οι οποίες προκαλούν απόκλιση των χαρακτηριστικών ποιότητας από τις προδιαγραφές. Το διάγραμμα ελέγχου είναι μια γραφική παρουσίαση ενός χαρακτηριστικού ποιότητας σε συνάρτηση ή με τον χρόνο ή με τον αριθμό δείγματος. Οι τιμές του χαρακτηριστικού ποιότητας έχουν προκύψει από τυχαία δείγματα που λαμβάνονται κατά την παραγωγική διαδικασία. Αν το χαρακτηριστικό ποιότητας είναι χαρακτηριστικό διαλογής τότε η εκτιμήτρια αφορά ποσοστό ελαττωματικών ή αριθμό ελαττωμάτων ή αριθμό ελαττωμάτων, ενώ αν είναι χαρακτηριστικό μέτρησης αφορά τη μέση τιμή και τη διασπορά της κατανομής.



Σχήμα 2.3: Ένα τυπικό διάγραμμα ελέγχου.

Στο Σχήμα 2.3 απεικονίζεται ένα τυπικό διάγραμμα ελέγχου ή αλλιώς απλό διάγραμμα ελέγχου ή διάγραμμα ελέγχου Shewhart . Όπως παρατηρούμε στο σχήμα υπάρχει μια κεντρική γραμμή (central line ,CL) η οποία αντιστοιχεί στην τιμή της παραμέτρου όταν η διαδικασία βρίσκεται υπό στατιστικό έλεγχο. Στην περίπτωση

διαγράμματος ποσοστού ελαττωματικών πρόκειται για την μέση τιμή, η οποία υπολογίζεται όταν η παραγωγική διαδικασία κυλάει ομαλά και είναι υπό την επίδραση μόνο τυχαίων αιτιών και όχι ειδικών αιτιών. Οι άλλες δύο οριζόντιες γραμμές αντιστοιχούν στο άνω όριο ελέγχου (upper control limit ,UCL) και στο κάτω όριο ελέγχου (lower control limit ,LCL). Στην περίπτωση όπου τα σημεία του διαγράμματος βρίσκονται εντός των ορίων ελέγχου, τότε η διεργασία βρίσκεται «υπό έλεγχο». Στην περίπτωση όπου έστω και ένα σημείο βρεθεί εκτός ορίων ελέγχου, η διεργασία είναι «εκτός ελέγχου». Η σχετικά μεγάλη απόσταση από την κεντρική γραμμή αποτελεί ένδειξη μεταβολής του χαρακτηριστικού ποιότητας και θα πρέπει να διερευνηθεί η συστηματική αυτή αιτία και να ληφθούν τα κατάλληλα διορθωτικά μέτρα. Υπάρχει και η περίπτωση όπου τα σημεία πέφτουν στην μία περιοχή που ορίζουν τα δύο όρια ελέγχου, δηλαδή τα σημεία παρουσιάζουν μια μη τυχαία συμμετρία ή περιοδικότητα και αυτό αποτελεί μια ένδειξη ότι η διεργασία είναι «εκτός ελέγχου». Το διάγραμμα ελέγχου είναι μια γραφική παράσταση της πορείας μιας εκτιμήτριας σε συνάρτηση με τον χρόνο ή με τον αριθμό του δείγματος. Η εκτιμήτρια αυτή στην ορολογία του στατιστικού ελέγχου ονομάζεται στατιστική δείγματος (sample statistic). Η στατιστική δείγματος είναι μία από τα παρακάτω :

- Σε περιπτώσεις σε χαρακτηριστικά διαλογής είναι: το ποσοστό ή ο αριθμός ελαττωματικών ,ο αριθμός ελαττωμάτων στο δείγμα ή ο αριθμός ελαττωμάτων ανά μονάδα δείγματος.
- Σε περιπτώσεις σε χαρακτηριστικά μέτρησης είναι: η μέση τιμή δείγματος ,η τυπική απόκλιση δείγματος ή το εύρος τιμών δείγματος .

Οι γενικές σχέσεις που περιγράφουν ένα διάγραμμα ελέγχου Shewhart για οποιαδήποτε στατιστική ελέγχου w είναι:

$$UCL = \mu_w + k\sigma_w$$

$$CL = \mu_w$$

$$LCL = \mu_w - k\sigma_w$$

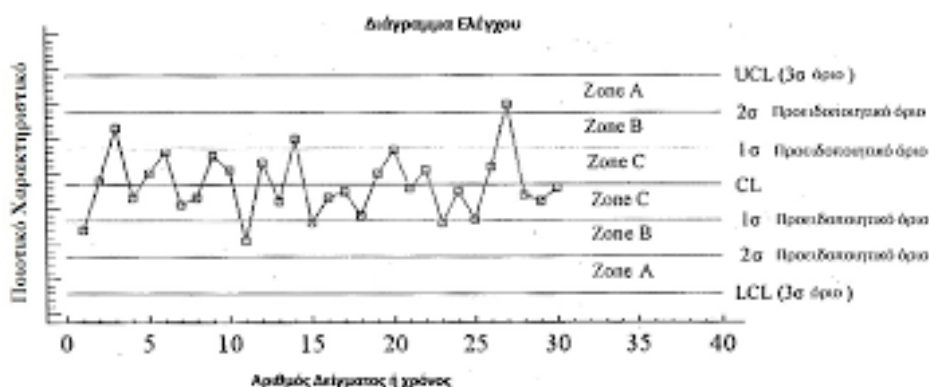
Όπου μ_w είναι η μέση τιμή και σ_w η τυπική απόκλιση της w σε κατάσταση στατιστικού ελέγχου και k είναι η παράμετρος θέσης των ορίων ελέγχου από την κεντρική γραμμή. Όταν το $k = 3$ έχουμε κατασκευή ορίων ελέγχου τριών τυπικών αποκλίσεων (three sigma control limits, 3σ) και η κατανομή που υποθέτουμε να περιγράψει την στατιστική ελέγχου είναι συνήθως κανονική κατανομή.

2.5 Επιλογή ορίων ελέγχου

Η επιλογή των ορίων ελέγχου είναι ιδιαίτερα σημαντική για τη δημιουργία ενός διαγράμματος ελέγχου, καθώς παίζει σημαντικό ρόλο στη διακριτική ικανότητα του διαγράμματος. Όσο πιο στενά είναι τα όρια, δηλαδή όσο πιο κοντά στην κεντρική

γραμμή βρίσκονται, τόσο πιο ευαίσθητο θα είναι το διάγραμμα, καθώς όλο και περισσότερα σημεία θα είναι εκτός ορίων. Συνεπώς τα μικρά όρια (μικρή τιμή του k) μπορούν να εντοπίζουν πιο εύκολα τη μεταβολή των σημείων, αλλά έχουν την τάση να πέφτουν σε σφάλματα πρώτου είδους. Από την άλλη πλευρά, ένα διάγραμμα με μεγαλύτερα όρια ελέγχου (μεγάλη τιμή του k) εντοπίζει πιο δύσκολα μια ειδική αιτία που θα εμφανιστεί, αλλά θα αποφύγει την τάση να πέσει σε σφάλματα πρώτου είδους. Η πιθανότητα σφάλματος δευτέρου είδους μπορεί να ελαττωθεί με την αύξηση του δείγματος, καθώς όσο περισσότερες πληροφορίες περιέχονται σε μεγαλύτερο δείγμα βοηθούν στη διαχωριστική ικανότητά του. Τα όρια ελέγχου με παράμετρο θέσης k ονομάζονται όρια τυπικών αποκλίσεων k . Συνήθως τα περισσότερα διαγράμματα ελέγχου χρησιμοποιούν όρια 3 τυπικών αποκλίσεων τα οποία αντιστοιχούν σε $\alpha=0,0027$ ή σε όρια 3,09 τυπικών αποκλίσεων που αντιστοιχούν σε $\alpha=0,002$ όταν η κατανομή είναι κανονική. Ωστόσο η επιλογή της θέσης των ορίων θα πρέπει να καθορίζεται και από παράγοντες όπως είναι το κόστος. Συγκεκριμένα, αν το κόστος διακοπής μιας παραγωγικής διαδικασίας Α για να εντοπιστεί μια ανύπαρκτη ειδική αιτία εξαιτίας σφάλματος πρώτου είδους είναι αρκετά μεγαλύτερο από μια διεργασία Β και επίσης οι συνέπειες λειτουργίας εκτός ελέγχου για την διεργασία Β είναι πολύ δυσμενείς, τότε συμπεραίνουμε πως τα όρια λειτουργίας της διεργασίας Β θα πρέπει να είναι πιο στενά από την Α.

Συχνό φαινόμενο για να κάνουμε πιο ευαίσθητο ένα διάγραμμα ελέγχου είναι να χρησιμοποιούμε προειδοποιητικά όρια ελέγχου. Τα προειδοποιητικά όρια ελέγχου απέχουν συνήθως ± 2 τυπικές αποκλίσεις της στατιστικής δείγματος από την κεντρική οριακή γραμμή. Στο Σχήμα 2.4 απεικονίζεται ένα διάγραμμα ελέγχου με όρια ελέγχου 3 τυπικών αποκλίσεων και προειδοποιητικά όρια 2 τυπικών αποκλίσεων. Σε περίπτωση όπου ένα ή περισσότερα σημεία βρίσκονται μεταξύ προειδοποιητικών ορίων και ορίων ελέγχου, τότε αποτελεί μια ένδειξη μετατόπισης της στατιστικής δείγματος.



Σχήμα 2.4: Διάγραμμα ελέγχου με προειδοποιητικά όρια.

Τα όρια αυτά συνήθως ορίζονται από τον μηχανισμό ποιοτικού ελέγχου όταν χρειάζεται κατά την παραγωγική διαδικασία. Η χρήση τους αυξάνει την ευαισθησία του διαγράμματος ελέγχου, αλλά και την πιθανότητα σφάλματος πρώτου είδους.

Σύμφωνα με το Σχήμα 2.4 υπάρχουν κάποια υποδείγματα που υποδεικνύουν οι παρακάτω κανόνες που δείχνουν πως η διεργασία είναι εκτός στατιστικού ελέγχου. Κάποιοι από αυτούς είναι:

- Ένα ή περισσότερα σημεία εκτός ορίων ελέγχου.
- 2 ή 3 συνεχόμενα σημεία στην ζώνη A (σε μια από τις δύο ζώνες A).
- 4 ή 5 σημεία πέραν της ζώνης C (σε μία από τις δύο περιοχές).
- 8 συνεχόμενα σημεία από την ίδια πλευρά της κεντρικής γραμμής.
- 6 συνεχόμενα σημεία σε αύξουσα ή φθίνουσα διάταξη.
- 15 συνεχόμενα σημεία εντός της ολικής ζώνης C .
- 14 συνεχόμενα σημεία σε εναλλασσόμενη μορφή.
- 8 συνεχόμενα σημεία πέρα της ολικής ζώνης C.
- Οποιαδήποτε ασυνήθιστη ή μη τυχαία ακολουθία σημείων.
- Ένα ή περισσότερα σημεία κοντά στα όρια ελέγχου ή τα προειδοποιητικά όρια.

Οι κανόνες αυτοί έχουν δημιουργηθεί με τέτοιο τρόπο ώστε η πιθανότητα εσφαλμένης ένδειξης όταν η διαδικασία είναι εντός στατιστικού ελέγχου να είναι ίδιας τάξης μεγέθους με την αντίστοιχη πιθανότητα του απλού κανόνα για διάγραμμα 3 τυπικών αποκλίσεων. Ο συνδυασμός ενός ή περισσότερων κανόνων από αυτούς μαζί με την υπέρβαση των ορίων ελέγχου οδηγούν στην αύξηση της συνολικής πιθανότητας σφάλματος πρώτου είδους. Επίσης οι σειρές των σημείων που περιεγράφηκαν παραπάνω είναι αρκετά σημαντικές καθώς μπορεί να φανερώσουν μια μη τυχαία συμπεριφορά, αλλά υπάρχουν και μοτίβα σημείων τα οποία μπορεί να δείχνουν μια κατάσταση εκτός στατιστικού ελέγχου.

2.6 Μέσο μήκος ροής (Average Run Length)

Το μέσο μήκος ροής (ARL) είναι ένα σημαντικό εργαλείο για τα διαγράμματα ελέγχου, καθώς εκφράζει τον μέσο αριθμό δειγμάτων από την στιγμή της μεταβολής στην παραγωγική διαδικασία μέχρι να εμφανιστεί το πρώτο δείγμα το οποίο θα δώσει την ένδειξη της μεταβολής. Εξαιτίας της παρουσίας σφάλματος δεύτερου είδους δεν είναι σίγουρο ότι η μεταβολή αυτή στην παραγωγική διαδικασία θα εντοπιστεί αμέσως.

Το μέσο μήκος ροής υπολογίζεται από τον παρακάτω τύπο:

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

Όπου το p εκφράζει την πιθανότητα ένα σημείο του διαγράμματος να βρεθεί εκτός ορίων ελέγχου. Όπως προκύπτει το p :

$$p = 1 - P(LCL \leq W \leq UCL) = 1 - F_0(\mu + K\sigma) + F_0(\mu - K\sigma)$$

Ισχύει για μια ποσότητα W η οποία έχει εντός ελέγχου μέση τιμή μ και τυπική απόκλιση σ και συνάρτηση κατανομής F_0 .

Σε κάθε δείγμα μετά την μεταβολή η πιθανότητα ένδειξης ισούται με την ισχύ του διαγράμματος $1 - \beta$. Ο αριθμός των δοκιμών μέχρι την πρώτη ένδειξη να βρεθεί ένα σημείο εκτός ελέγχου είναι τυχαία μεταβλητή που ακολουθεί γεωμετρική κατανομή.

Όταν η διαδικασία είναι υπό στατιστικό έλεγχο, τότε το ARL εκφράζει τον μέσο αριθμό δειγμάτων μέχρι την πρώτη εσφαλμένη ένδειξη εκτός ελέγχου. Η πιθανότητα του σφάλματος είναι α και άρα σε κατάσταση στατιστικού ελέγχου έχουμε $ARL=1/\alpha$. Στην περίπτωση όπου χρησιμοποιούμε διάγραμμα με 3σ όρια ελέγχου και κανονική κατανομή, γνωρίζουμε ότι το $\alpha=0,0027$ δηλαδή η πιθανότητα να βρεθεί ένα σημείο εκτός ελέγχου τότε το εντός ελέγχου ARL_0 είναι:

$$\begin{aligned} p &= 1 - P(LCL \leq W \leq UCL) | W \sim N(\mu, \sigma^2) \\ &= 1 - P(\mu - K\sigma \leq W \leq \mu + K\sigma) | W \sim N(\mu, \sigma^2) \\ &= 1 - \Phi(k) + \Phi(-k) = 2\Phi(-k) \end{aligned}$$

οπότε

$$ARL_0 = \frac{1}{0,0027} = 370$$

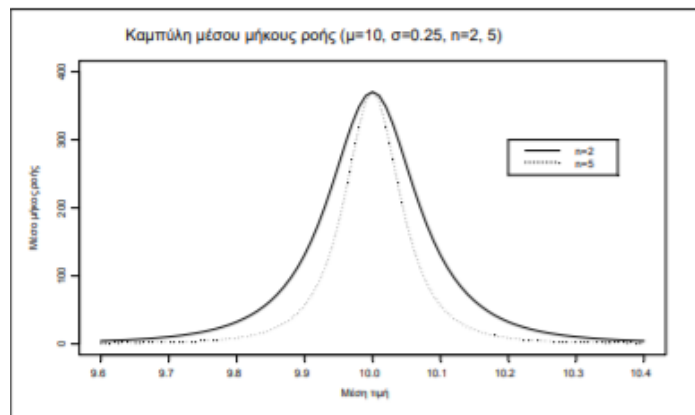
Δηλαδή κάθε 370 δείγματα κατά μέσο όρο θα λαμβάνουμε εσφαλμένη ένδειξη ότι η διαδικασία είναι εκτός ελέγχου. Στην περίπτωση όπου η διεργασία βρίσκεται εκτός ελέγχου τότε το εκτός ελέγχου ARL_Δ είναι ίσο με:

$$ARL_\Delta = \frac{1}{1 - \beta}$$

Όπου β ισούται με την πιθανότητα να βρεθεί ένα σημείο εντός των ορίων ελέγχου σε αυτήν την μετατόπιση.

Γενικά στην εντός ελέγχου διεργασία προτιμάμε να έχουμε μεγάλο ARL_0 , καθώς δεν θέλουμε να διακόπτουμε μια παραγωγική διαδικασία η οποία λειτουργεί ομαλά. Αντίθετα σε μια εκτός ελέγχου διεργασία θέλουμε να έχουμε όσο γίνεται μικρότερο ARL_Δ ώστε να εντοπίζεται πιο γρήγορα η μετατόπιση της διεργασίας.

Στο Σχήμα 2.5 φαίνεται η περίπτωση που έχουμε 3σ όρια ελέγχου και για $X \sim N(10, 0.25^2)$. Από το σχήμα προκύπτει πως για μικρές μετατοπίσεις του μέσου επιπέδου της διεργασίας το ARL_Δ δεν είναι ικανοποιητικό.



Σχήμα 2.5: Καμπύλη μέσου μήκους ροής (ARL)

Πολλές φορές η χρήση του ARL ως μέτρο απόδοσης της παραγωγικής διαδικασίας δεν είναι τόσο ικανοποιητική, καθώς υπάρχει απόκλιση ανάμεσα στο θεωρητικό ARL και στο πρακτικό. Αυτό οφείλεται στο γεγονός ότι το ARL ακολουθεί γεωμετρική κατανομή, η οποία είναι ασύμμετρη για μικρές τιμές του p και η τυπική απόκλιση λαμβάνει μεγάλες τιμές έτσι η μέση τιμή δεν μπορεί να θεωρηθεί αντιπροσωπευτικό μέτρο κεντρικής τάσης της κατανομής.

Πέρα από το μέσο μήκος ροής, εάν μια συστηματική αιτία επηρεάζει την παραγωγική διαδικασία από την έναρξη, ορίζουμε το ATS όπου είναι ο μέσος χρόνος μέχρι κάποιο δείγμα να εμφανίσει την παρουσία της αιτίας αυτής. Το ATS ορίζεται από τον τύπο :

$$ATS = ARL \cdot h$$

όπου το h συμβολίζει τον χρόνο μεταξύ διαδοχικών δειγμάτων.

Επίσης στην περίπτωση όπου η διαδικασία λειτουργεί σε στατιστικό έλεγχο αρχικά και στην συνέχεια υφίσταται επίδραση μιας ειδικής αιτίας που μεταβάλλει την κατανομή του χαρακτηριστικού ποιότητας, τότε χρησιμοποιούμε το AATS όπου είναι ο μέσος χρόνος από την στιγμή της μεταβολής μέχρι την εμφάνιση στο διάγραμμα ελέγχου. Η ειδική αιτία συμβαίνει στο διάστημα h μεταξύ δύο διαδοχικών δειγμάτων. Το AATS δίνεται από την σχέση :

$$AATS = (h - \tau) + h(ARL - 1) = \frac{h}{1 - \beta} - \tau = \frac{h}{1 - \beta} - \frac{h}{2}$$

όπου τ είναι ο μέσος χρόνος από την αρχή του διαστήματος h μέχρι την στιγμή που επιδρά η ειδική αιτία και το $h - \tau$ είναι ο μέσος χρόνος από την στιγμή της μεταβολής μέχρι το επόμενο δείγμα.

Η τελευταία ισότητα στον τύπο προκύπτει εάν θεωρηθεί ότι η συστηματική αιτία συμβαίνει περίπου στο μέσο του διαστήματος μεταξύ των δύο δειγμάτων.

Κεφάλαιο 3^ο: Διαγράμματα ελέγχου για δεδομένα ικανοποίησης πελατών

3.1 Γενικά

Σύμφωνα με τα προηγούμενα, σκοπός των διαγραμμάτων ελέγχου και των στατιστικών εργαλείων είναι ο εντοπισμός συστηματικών μεταβολών των παραμέτρων της παραγωγικής διαδικασίας οι οποίες προκαλούν απόκλιση των χαρακτηριστικών ποιότητας από τις προδιαγραφές, τόσο για την επιχείρηση αλλά και για το πελάτη.

Η πιο διαδεδομένη μέθοδος για μέτρηση ικανοποίησης είναι οι έρευνες πελατών. Οι εταιρείες διεξάγουν συχνές έρευνες για να μετρήσουν την ικανοποίηση των πελατών και αναλύουν τα δεδομένα που λαμβάνουν με στατιστικά εργαλεία προκειμένου να αντλήσουν τα κατάλληλα αποτελέσματα. Η χρήση των στατιστικών εργαλείων, όπως είναι τα διαγράμματα ελέγχου ξεκίνησε από την εφαρμογή στον κατασκευαστικό τομέα και έπειτα αναπτύχθηκε και σε διαφορετικούς τομείς, λόγω των θετικών επιπτώσεων όπως η βελτίωση ποιότητας. Η πρώτη εφαρμογή ξεκίνησε όταν ο Shewhart χρησιμοποίησε τα στατιστικά εργαλεία στην Western Electric. Σκοπός της εφαρμογής αυτής ήταν να περιορίσουν την απόκλιση στα τηλεφωνικά παράπονα όπου λάμβαναν.

Ωστόσο, η εφαρμογή μείωσης της απόκλισης οδηγούσε σε λανθασμένα αποτελέσματα. Το πρόβλημα εντοπίστηκε στο ότι η μείωση της απόκλισης στην Western Electric έπρεπε να χωριστεί σε 2 υποομάδες. Η πρώτη ομάδα αφορούσε αντιμετώπιση σφαλμάτων που προκλήθηκαν από κοινές αιτίες ως ειδικές αιτίες και η δεύτερη ομάδα αφορά αντιμετώπιση σφαλμάτων που προκλήθηκαν από ειδικές αιτίες ως κοινές αιτίες. Οι κοινές αιτίες είναι ένα κομμάτι της διαδικασίας που δύσκολα παραλείπεται. Για την αντιμετώπιση κοινών αιτιών απαιτείται μεγάλη αλλαγή, όπως αλλαγή στην καθοδήγηση της παραγωγικής διαδικασίας ή στο προϊόν. Αντίθετα, για την αντιμετώπιση των ειδικών αιτιών, η μείωση της μεταβολής μπορεί να αποφευχθεί. Συνεπώς η αντιμετώπιση της κάθε αιτίας είναι διαφορετική (Deming, 1986).

Το διάγραμμα ελέγχου που θα δημιουργηθεί για να εντοπίσει την ειδική και την κοινή αιτία είναι διαφορετικό και πρέπει να μελετηθεί τόσο από την πλευρά του επιθυμητού αποτελέσματος, αλλά και από την αιτία που προκάλεσε το επιθυμητό αποτέλεσμα. Όταν μια διαδικασία βρίσκεται σε στατιστικό έλεγχο, θα παράγει το επιθυμητό αποτέλεσμα μέσα στα προβλεπόμενα όρια με τυχαία συμπεριφορά και τα αποτελέσματα θα είναι προβλέψιμα (Deming, 1986). Μια διαδικασία η οποία είναι σε στατιστικό έλεγχο παράγει αποτελέσματα που είναι ανεξάρτητα και κανονικά κατανομημένα (Montgomery, 1996). Μια διαδικασία βρίσκεται εκτός ελέγχου όταν συμβαίνουν και άλλοι τύποι μεταβολών στο σύστημα. Πιθανές καταστάσεις εκτός ελέγχου είναι όταν ένα σημείο βρίσκεται εκτός ορίων ή τέτοιες καταστάσεις

περιγράφονται και από τους κανόνες του Montgomery (1996). Όσο περισσότεροι κανόνες για τον εντοπισμό πιθανών σημείων εκτός ορίων ελέγχου, προκύπτουν περισσότερα σφάλματα πρώτου είδους ή εσφαλμένα αποτελέσματα. Αντίθετα τα σφάλματα δεύτερου είδους φαίνεται να μειώνονται. Συνεπώς η χρήση των διαγραμμάτων ελέγχου μπορεί να χρησιμοποιηθεί για να εντοπίσει τη μεταβολή αλλά όχι για να λύσει το πρόβλημα.

Η εφαρμογή των στατιστικών διαγραμμάτων ελέγχου σε τέτοιες εφαρμογές μπορεί να επηρεάσει τα αποτελέσματα λόγω των δεδομένων που λαμβάνουμε. Τα δεδομένα που λαμβάνουμε από μια έρευνα ικανοποίησης δεν είναι πάντα κανονικά, λόγω της φύσης των δεδομένων ή των διαδικασιών καθαυτών, γι' αυτό υπάρχουν διάφοροι τρόποι για την ανάλυση των μη κανονικών δεδομένων. Τα διαγράμματα ελέγχου ανάλογα το είδος των μεταβλητών μπορούν να χωριστούν σε διαγράμματα πολυμεταβλητής (multivariate) και σε μονομεταβλητής (univariate) ανάλυσης. Τα διαγράμματα ελέγχου πολυμεταβλητών χρησιμοποιούνται όταν έχουμε δύο ή περισσότερα χαρακτηριστικά ποιότητας που πρέπει να παρακολουθήσουμε.

Ένα ευρέως γνωστό εργαλείο στατιστικού ελέγχου είναι τα διαγράμματα ελέγχου Shewhart. Η χρήση των διαγραμμάτων ελέγχου Shewhart προϋποθέτει πως τα δεδομένα είναι κανονικά κατανεμημένα, όμως αυτό δεν συμβαίνει πάντα σε δεδομένα ικανοποίησης πελατών. Γι' αυτό έχουν αναπτυχθεί εναλλακτικά διαγράμματα ελέγχου που θα αναλυθούν στη συνέχεια. Επιπλέον θα γίνει χρήση της κατανομής Johnson, όπου τα δεδομένα μπορούν να μετατραπούν σε κανονικά χρησιμοποιώντας συναρτήσεις της κατανομής αυτής. Τέλος θα χρησιμοποιηθεί ως εργαλείο για την ανάλυση δεδομένων ο μετασχηματισμός δεδομένων με χρήση κατάλληλων συναρτήσεων, έτσι ώστε τα δεδομένα να ταιριάζουν στην κανονική κατανομή.

3.2 Εναλλακτικά διαγράμματα ελέγχου

3.2.1 Τυπικά διαγράμματα ελέγχου Shewhart

Τα διαγράμματα ελέγχου Shewhart αποτελούν μια διαδεδομένη επιλογή για ανάλυση δεδομένων ικανοποίησης με την πάροδο του χρόνου, λόγω της ικανότητάς τους να εντοπίζουν γρήγορα αλλαγές στη διεργασία. Έτσι σύμφωνα με τον Hayes (1992), ο οποίος πρότεινε μια εφαρμογή των διαγραμμάτων ελέγχου Shewhart με το p διάγραμμα ελέγχου (ή np chart) για ανάλυση ικανοποίησης του αριθμού των πελατών. Η χρήση του γραφήματος αυτού προϋποθέτει ότι η έρευνα λαμβάνει μόνο 2 πιθανές απαντήσεις από τους πελάτες δηλαδή ικανοποιημένος ή δυσαρεστημένος, ή περισσότερες επιλογές οι οποίες όμως ομαδοποιούνται σε 2 κατηγορίες. Γι' αυτό τον λόγο το γράφημα μπορεί να χρησιμοποιήσει την διωνυμική κατανομή για να περιγράψει τις απαντήσεις που έλαβε από την έρευνα.

Επίσης ο Hayes (1992) πρότεινε τη χρήση των \bar{x} και s chart (όπου s η εκτιμώμενη τυπική απόκλιση) όταν παρουσιάζονται πολλές απαντήσεις στον πελάτη. Συνήθως αυτές οι απαντήσεις κυμαίνονται από το 1 έως το 5, όπου το 1 αντιπροσωπεύει τους πιο δυσαρεστημένους και το 5 τους πιο ικανοποιημένους πελάτες. Για το γράφημα αυτό, ο Hayes (1992) προτείνει το σχεδιασμό του υπολογίζοντας τον μέσο όρο όλων των απαντήσεων και σχεδιάζοντας τον μέσο όρο της υποομάδας για το \bar{x} γράφημα και την τυπική απόκλιση της υποομάδας για το s γράφημα. Οι προϋποθέσεις για την χρήση των παραπάνω γραφημάτων δεν πληρούνται για τα δεδομένα μιας έρευνας. Αυτό συμβαίνει καθώς στις έρευνες το μέγεθος δείγματος είναι μεγάλο (δηλαδή τα μεγέθη υποομάδων στο διάγραμμα ελέγχου) και έτσι το κεντρικό οριακό θεώρημα μπορεί να χρησιμοποιηθεί αλλά όχι σε όλες τις περιπτώσεις. Η εκτίμηση της τυπικής απόκλισης για τον υπολογισμό των ορίων ελέγχου προϋποθέτει παρατηρήσεις πελατών οι οποίες είναι κανονικά κατανεμημένες.

Σε περιπτώσεις όπου έχουμε μεγάλο μέγεθος δείγματος μπορεί να εξαλειφθεί η επίδραση της εσφαλμένης υπόθεσης, αλλά η φύση των δεδομένων της έρευνας καθιστά πιο δύσκολη την εφαρμογή ενός διαγράμματος \bar{x} χρησιμοποιώντας τυπικές αποκλίσεις της υποομάδας. Η χρήση διαγραμμάτων ελέγχου, συγκεκριμένα για τα διαγράμματα \bar{x} και s , προϋποθέτει μικρό δείγμα μέγεθος υποομάδων. Για τα δεδομένα που προέρχονται από έρευνες ικανοποίησης, το ποσοστό απόκρισης από μία έρευνα μεταβάλλεται και ποικίλει συνήθως από περίοδο σε περίοδο, μεταβάλλοντας και το μέγεθος κάθε υποομάδας. Η μεταβολή του μεγέθους υποομάδας συνεπάγεται με μεταβολή και των ορίων ελέγχου από παρατήρηση σε παρατήρηση. Συνεπώς σύμφωνα με τον Hayes (1992) τα παραπάνω γραφήματα δεν μπορούν να χρησιμοποιηθούν για τον χειρισμό μεταβλητών μεγεθών που προέρχονται από δεδομένα έρευνας. Σε περιπτώσεις όπου το μέγεθος του δείγματος είναι μεγάλο, τα εύρη υποτιμούν την τυπική απόκλιση και έτσι προκύπτουν όρια τα οποία είναι στενά.

Στον στατιστικό ποιοτικό έλεγχο το γράφημα \bar{x} , το γράφημα R και το γράφημα s είναι τύποι γραφημάτων ελέγχου που χρησιμοποιούνται για την συλλογή συνεχών δεδομένων ή μεταβλητών. Ξεκινώντας από το διάγραμμα \bar{x} , παρουσιάζονται γραφικά και συνολικά οι στατιστικοί έλεγχοι των υποθέσεων.

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Όπου μ είναι η μέση τιμή της κατανομής του X κατά την στιγμή της δειγματοληψίας και μ_0 η ονομαστική μέση τιμή που χαρακτηρίζει την κατάσταση στατιστικού ελέγχου της δειγματοληψίας. Η τιμή μ_0 συχνά ταυτίζεται με την επιθυμητή τιμή του χαρακτηριστικού ποιότητας.

Τα όρια ελέγχου υπολογίζονται στατιστικά από τους παρακάτω τύπους για μια σταθερή διαδικασία. Η μέση τιμή κάθε τυχαίου δείγματος μεγέθους n υπολογίζεται από την σχέση :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Όπου $x_i = 1, 2, 3, \dots, n$ είναι οι συγκεκριμένες τιμές του δείγματος.

$$UCL = \bar{x} + k \frac{\sigma}{n}$$

$$CL = \bar{x}$$

$$LCL = \bar{x} - k \frac{\sigma}{n}$$

Μια ευρέως γνωστή μέθοδος για τον εντοπισμό κοινών και ειδικών αιτιών που προκαλούν μεταβολή σε μια παραγωγική διαδικασία είναι το διάγραμμα ελέγχου R . Σκοπός του διαγράμματος R είναι η παρακολούθηση της μεταβλητότητας της διαδικασίας. Συγκεκριμένα τυχαία δείγματα μεγέθους n λαμβάνονται από την παραγωγική διαδικασία ανά h χρονικές μονάδες. Το εύρος κάθε δείγματος υπολογίζεται και καταγράφεται στο διάγραμμα ελέγχου. Η τιμή εύρους εκτός ορίων δείχνει πως η τυπική απόκλιση του χαρακτηριστικού ποιότητας έχει αυξηθεί (δηλαδή σημείο πάνω από το UCL) ή ότι υπάρχει πρόβλημα στην διαδικασία μέτρησης και καταγραφής τιμών (σημείο κάτω από το LCL). Στην περίπτωση όπου έχει προσδιοριστεί η τιμή της τυπικής απόκλισης του X , η κεντρική γραμμή και τα όρια ελέγχου k τυπικών αποκλίσεων (σ_R) του διαγράμματος R τότε ο υπολογισμός των ορίων ελέγχου γίνεται με τους παρακάτω τύπους :

$$UCL = d_2\sigma + Kd_3\sigma$$

$$CL = d_2\sigma$$

$$LCL = d_2\sigma - Kd_3\sigma$$

Όταν τα δείγματα που λαμβάνουμε ακολουθούν την κανονική κατανομή με μέση τιμή μ και τυπική απόκλιση σ , η μέση τιμή και τυπική απόκλιση του R αντίστοιχα είναι $\mu_R = d_2\sigma$ και $\sigma_R = d_3\sigma$.

Στην περίπτωση όπου η ονομαστική τυπική απόκλιση δεν έχει προσδιοριστεί με ακρίβεια, η κατασκευή του διαγράμματος R βασίζεται στην εκτίμηση της σ από την ανάλυση δυνατοτήτων της παραγωγικής διαδικασίας. Αν η εκτίμηση αυτή έχει προσδιοριστεί από το μέσο εύρος των δειγμάτων που χρησιμοποιήθηκαν, τότε αυτό το μέσο εύρος χαρακτηρίζει και την κεντρική οριακή γραμμή. Η εκτιμήτρια της τυπικής απόκλισης του εύρους δείγματος σε αυτήν την περίπτωση είναι:

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2}$$

Ο υπολογισμός των ορίων ελέγχου του R γίνεται σύμφωνα με τους παρακάτω τύπους:

$$UCL = \bar{R} + k d_3 \frac{\bar{R}}{d_2}$$

$$CL = \bar{R}$$

$$LCL = \bar{R} - k d_3 \frac{\bar{R}}{d_2}$$

Για $k=3$, τα όρια ελέγχου υπολογίζονται από τις παρακάτω σχέσεις:

$$UCL = D_4 \times \bar{R}$$

$$CL = \bar{R}$$

$$LCL = D_3 \times \bar{R}$$

όπου

$$D_3 = 1 - 3 \frac{d_3}{d_2} = \frac{D_1}{d_2}$$

$$D_4 = 1 + 3 \frac{d_3}{d_2} = \frac{D_2}{d_2}$$

Όπου οι τιμές D_3, D_4, d_2, d_3 λαμβάνονται από πίνακα (Wadsworth et al, Stephens and Godfrey, 1986) και εξαρτώνται από το μέγεθος του δείγματος. Η τιμή του \bar{R} υπολογίζεται ως ο μέσος όρος του εύρους του δείγματος n , ενώ η τιμή R υπολογίζεται ως η διαφορά μεταξύ της μεγαλύτερης και μικρότερης τιμής του δείγματος:

$$R = x_{max} - x_{min}.$$

Το διάγραμμα R έχει σκοπό τον εντοπισμό συστημικών αιτιών που έχουν μεταβάλει την φυσική μεταβλητότητα της διαδικασίας, αλλά επίσης υπόκειται και στον κίνδυνο σφαλμάτων πρώτου και δεύτερου είδους. Για τον κατάλληλο εντοπισμό των σφαλμάτων απαιτείται η κατανομή του εύρους δείγματος R . Η κατανομή του εύρους R δεν είναι κανονική ακόμα και στην περίπτωση όπου η κατανομή του χαρακτηριστικού ποιότητας είναι κανονική. Η κατανομή του R εξαρτάται από το μέγεθος του δείγματος n .

Το πρόβλημα που προκύπτει με την εφαρμογή του διαγράμματος R σε δεδομένα που προέρχονται από έρευνα ικανοποίησης πελατών είναι πως η κατασκευή του προϋποθέτει πως η κατανομή του δείγματος είναι κανονικά κατανεμημένη και τα όρια χρησιμοποιούν το εύρος της υποομάδας προκειμένου να υπολογίσουν την τυπική απόκλιση.

Σύμφωνα με όσα αναφέρθηκαν παραπάνω σχετικά με την καταλληλότητα των διαγραμμάτων για δεδομένα έρευνας παρακάτω θα παρουσιαστούν κάποιες εναλλακτικές διαγραμμάτων ελέγχου. Το πρώτο διάγραμμα ελέγχου είναι μια επέκταση του p chart αλλά πιο κατάλληλη για δεδομένα που προέρχονται από έρευνες ικανοποίησης και έχουμε περισσότερα από δύο πιθανά αποτελέσματα και

το δεύτερο διάγραμμα βασίζεται στην χ^2 στατιστική ,όπου χρησιμοποιείται για έλεγχο υποθέσεων .

Πίνακας 3.1 Πίνακας τιμών παραμέτρων κατασκευής για τα διαγράμματα ελέγχου διασποράς και μέσης τιμής

n	A ₂	D ₃	D ₄	A ₃	B ₃	B ₄
2	1.88	0	3.27	2.66	0	3.27
3	1.02	0	2.57	1.95	0	2.57
4	0.73	0	2.28	1.63	0	2.27
5	0.58	0	2.11	1.43	0	2.09
6	0.48	0	2.00	1.29	0.03	1.97
7	0.42	0.08	1.92	1.18	0.12	1.88
8	0.37	0.14	1.86	1.10	0.19	1.81
9	0.34	0.18	1.82	1.03	0.24	1.76
10	0.31	0.22	1.78	0.98	0.28	1.72
11	0.29	0.26	1.74	0.93	0.32	1.68
12	0.27	0.28	1.72	0.89	0.35	1.65
13	0.25	0.31	1.69	0.85	0.38	1.62
14	0.24	0.33	1.67	0.82	0.41	1.59
15	0.22	0.35	1.65	0.79	0.43	1.57
16	0.21	0.36	1.64	0.76	0.45	1.55
17	0.20	0.38	1.62	0.74	0.47	1.53
18	0.19	0.39	1.61	0.72	0.48	1.52
19	0.19	0.40	1.60	0.70	0.50	1.50
20	0.18	0.41	1.59	0.68	0.51	1.49

3.2.2 Επέκταση του p διαγράμματος

Σε αντίθεση με το παραπάνω η επέκταση του p γραφήματος ή \bar{x}_p σε περίπτωση όπου έχουμε μόνο 2 κατηγορίες, η οποία είναι παρόμοια με τα παραδοσιακά γραφήματα, είναι η πιο κατάλληλη μέθοδος για δεδομένα ικανοποίησης πελατών. Αυτό οφείλεται στη διαφορά κατασκευής των ορίων ελέγχου.

Συγκεκριμένα για τον υπολογισμό των ορίων ελέγχου η διασπορά που υπολογίζεται δε βασίζεται στην κανονική κατανομή αλλά στην κατανομή που βρίσκεται κάτω από το άθροισμα των x_s , όπου ως x_s ορίζουμε το σύνολο των απαντήσεων x . Το γράφημα αυτό βασίζεται στο μεγάλο μέγεθος δείγματος έτσι ώστε η κατανομή των \bar{x} να είναι περίπου κανονική. Τα όρια ελέγχου υπολογίζονται χρησιμοποιώντας την πραγματική διακύμανση των δεδομένων ικανοποίησης του πελάτη και χρησιμοποιούμε ± 3 τυπικές αποκλίσεις.

Υποθέτουμε ότι έχουμε πιθανές απαντήσεις των πελατών που κυμαίνονται από 1 έως k , όπου k η υψηλότερη τιμή και 1 η χαμηλότερη. Έτσι το X μπορεί να πάρει τιμές από

το 1 έως το k και η πιθανότητα να δώσει κάποιος πελάτης απάντηση x είναι διακριτή τυχαία μεταβλητή με συνάρτηση μάζας πιθανότητας $f(x) = Pr(X = x) = p_x$.

Η αναμενόμενη μέση τιμή και η τυπική απόκλιση είναι :

$$\mu_x = \sum_{1}^k x p_x$$

$$\sigma_x^2 = \sum_{1}^k x^2 p_x - \left(\sum_{1}^k x p_x \right)^2$$

Η αναμενόμενη τιμή του δείγματος υπολογίζεται από το δείγμα χρησιμοποιώντας

$$\mu_{\bar{x}} = \mu_x \text{ και } \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Συνεπώς τα όρια ελέγχου που προκύπτουν είναι :

$$\sum_{1}^k x p_x \pm \frac{3}{\sqrt{n}} \sqrt{\sum_{1}^k x^2 p_x - \left(\sum_{1}^k x p_x \right)^2}$$

Από τα όρια ελέγχου συμπεραίνουμε πως η τιμή τους μεταβάλλεται ανάλογα με το μέγεθος δείγματος n. Γι' αυτό τον λόγο, σε περίπτωση σταθερών ορίων ελέγχου, υπολογίζουμε τα όρια χρησιμοποιώντας την μέση τιμή του μεγέθους δείγματος n. Η περίπτωση αυτή είναι ιδιαίτερα χρήσιμη σε περιπτώσεις μεταβλητού μεγέθους δείγματος (Gitlow et al. 1989).

Για τον υπολογισμό του p_x όταν η τιμή του δεν είναι γνωστή μπορεί να υπολογιστεί από τα δεδομένα που λαμβάνουμε από την έρευνα ικανοποίησης.

$$\widehat{p}_x = \frac{\sum_{j=1}^m Y_{xj}}{\sum_{j=1}^m n_j}$$

Συλλέγουμε m δεδομένα από το συνολικό δείγμα n_j και Y_{xj} ισούται με τον πραγματικό αριθμό των πελατών που δίνουν απάντηση x στο δείγμα j.

Η παραπάνω σχέση \widehat{p}_x αναφέρεται στο ποσοστό των πελατών όπου δίνουν απάντηση x συνδυάζοντας όλα τα δείγματα. Υποθέτει επίσης πως το m είναι αρκετά μεγάλο ώστε τυχόν σημεία εκτός ορίων ελέγχου εξαλείφονται.

3.2.3 Διάγραμμα ελέγχου χ^2

Ένα διαφορετικό γράφημα για δεδομένα ικανοποίησης πελατών είναι μια διαδικασία η οποία χρησιμοποιεί μετρήσεις καλής προσαρμογής (χ^2 goodness of fit) ώστε να συγκρίνει την πραγματική με την θεωρητική τιμή της κατανομής. Οι εφαρμογές του γραφήματος μελετήθηκαν από τον Duncan (1950) και τον Marcucci (1985) για

δεδομένα από τον κατασκευαστικό τομέα. Ωστόσο η μεθοδολογία αυτή αποτελεί μια κατάλληλη επιλογή για δεδομένα έρευνας, καθώς και για έλεγχο υποθέσεων για κατηγορηματικά δεδομένα.

Υποθέτουμε πως έχουμε k κατηγορίες δηλαδή k πιθανές απαντήσεις που δίνει ο πελάτης. Ο πελάτης δίνει μια απάντηση x με πιθανότητα p_x και ο πραγματικός αριθμός πελατών που δίνουν απάντηση x στο δείγμα j ισούται με Y_{xj} . Εάν είναι γνωστός ο πραγματικός αριθμός πελατών υπολογίζεται από τον τύπο :

$$\widehat{p}_x = \frac{\sum_{j=1}^m Y_{xj}}{\sum_{j=1}^m n_j}$$

Στην συνέχεια ο πραγματικός αριθμός πελατών συγκρίνεται με τον θεωρητικό/ αναμενόμενο αριθμό πελατών που δίνουν απάντηση x χρησιμοποιώντας την μέτρηση καλής προσαρμογής .

$$\chi_j^2 = \frac{\sum_{x=1}^k (Y_{xj} - E_{xj})^2}{E_{xj}}$$

Όπου $E_{xj} = n_j \times p_x$, δηλαδή ο αναμενόμενος αριθμός πελατών που δίνουν απάντηση x στο δείγμα j και n_j είναι ο συνολικός αριθμός πελατών στο δείγμα j .

Η παραπάνω σχέση αποτελεί την κανονική μέτρηση καλής προσαρμογής και προσεγγίζει την κατανομή χ^2 (Chi- squared του Pearson) η οποία χρησιμοποιείται σε οποιαδήποτε μεταβλητή που μπορεί να υπολογιστεί η συνάρτηση αθροιστικής κατανομής. Η κατανομή χ^2 χρησιμοποιείται στις κοινές δοκιμές χ^2 για την καλή προσαρμογή μιας παρατηρούμενης κατανομής σε μια θεωρητική, την ανεξαρτησία δύο κριτηρίων ταξινόμησης ποιοτικών δεδομένων και την εκτίμηση του διαστήματος εμπιστοσύνης για την τυπική απόκλιση πληθυσμού που ακολουθεί κανονική κατανομή από την τυπική απόκλιση δείγματος.

Το τεστ καλής προσαρμογής χ^2 μπορεί να εφαρμοστεί σε δεσμευμένα δεδομένα, δηλαδή δεδομένα τα οποία χωρίζονται σε κατηγορίες, καθώς η τιμή της στατιστικής δοκιμής εξαρτάται από τον τρόπο δέσμευσης των δεδομένων. Ένα μειονέκτημα της μεθόδου αυτής είναι πως απαιτεί μεγάλο μέγεθος δείγματος ώστε η προσέγγιση χ^2 να είναι έγκυρη. Το χ^2 τεστ αποτελεί ένα μέτρο της καλής προσαρμογής που είναι το άθροισμα των διαφορών μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων αποτελέσματος (δηλαδή πλήθος παρατηρήσεων), το καθένα στο τετράγωνο και διαιρούμενο με τις αναμενόμενες παρατηρήσεις. Η τιμή που προκύπτει μπορεί να συγκριθεί με μία κατανομή χ^2 ώστε να ελεγχθεί ο έλεγχος καλής προσαρμογής. Η κατανομή χ^2 έχει $k-1$ βαθμούς ελευθερίας και υποθέτουμε ότι το E_{xj} είναι τουλάχιστον 5. Σε περιπτώσεις όπου ο αναμενόμενος αριθμός πελατών είναι μικρότερος από 5, οι κατηγορίες πρέπει να συγχωνευτούν, έτσι ώστε να ικανοποιείται ο συγκεκριμένος περιορισμός. Σε αυτήν την περίπτωση οι βαθμοί ελευθερίας μειώνονται.

Μια σημαντική διαφορά του γραφήματος χ^2 είναι πως έχει μόνο άνω όριο ελέγχου. Σε περίπτωση δηλαδή όπου ξεπεραστεί το άνω όριο ελέγχου, μια ειδική αιτία έχει επηρεάσει την διεργασία. Μια τέτοια ειδική αιτία μπορεί να οφείλεται είτε σε αλλαγές στην επιχείρηση, είτε σε αλλαγές στο περιβάλλον που έχουν επηρεάσει την ικανοποίηση του πελάτη σημαντικά.

Για να είναι ο μέσος ρυθμός ψευδούς συναγερμού του διαγράμματος ελέγχου ίδιος με αυτά των γνωστών διαγραμμάτων ελέγχου, το UCL υπολογίζεται έτσι ώστε η πιθανότητα να βρίσκεται εκτός ορίων ελέγχου όταν οι αναλογίες δεν έχουν αλλάξει να είναι 0.0027.

Παρακάτω παρουσιάζεται ο Πίνακας 3.2 που δίνει τις τιμές στα άνω όρια ελέγχου(UCL) ανάλογα με τις τιμές k και την κατανομή χ^2 . Το k μπορεί να πάρει τιμές από το 2 έως το 10, καθώς τιμές μεγαλύτερες του 10 για δεδομένα ικανοποίησης πελατών δεν χρησιμοποιούνται.

Πίνακας 3.2 : Πίνακας τιμών UCL για διάφορες τιμές κατηγοριών (k).

k	UCL
2	9.000
3	11.829
4	14.156
5	16.251
6	18.205
7	20.062
8	21.846
9	23.574
10	25.257

Το διάγραμμα χ^2 θεωρείται από τα πιο κατάλληλα γραφήματα για δεδομένα που προέρχονται από έρευνα. Αυτό συμβαίνει καθώς για τον υπολογισμό των ορίων ελέγχου δεν στηρίζεται στην υπόθεση ότι τα δεδομένα είναι κανονικά, αλλά εξαρτάται από τον αριθμό των κατηγοριών k . Λόγω της εξάρτησης από τον αριθμό των κατηγοριών k της έρευνας, το άνω όριο ελέγχου (UCL) δεν επηρεάζεται από το μέγεθος του δείγματος και από τη μεταβλητότητα του. Ωστόσο πέρα από τα θετικά του γραφήματος χ^2 , είναι δύσκολη η εφαρμογή του με το χέρι, καθώς και η ερμηνεία του λόγω των στατιστικών αθροισμάτων που περιλαμβάνει ο τύπος

3.3 Κατανομή Johnson

Τα χαρακτηριστικά ποιότητας που προέρχονται από έρευνες αναλύονται με στατιστικά εργαλεία και συχνά απαιτείται να είναι κανονικά κατανεμημένα. Σε περιπτώσεις όπου ένα χαρακτηριστικό ποιότητας δεν είναι κανονικά κατανεμημένο τα εργαλεία που υποθέτουν κανονική κατανομή μπορεί να οδηγήσουν σε εσφαλμένα αποτελέσματα. Μια εναλλακτική προσέγγιση για την αντιμετώπιση του προβλήματος αυτού είναι ο μετασχηματισμός των μη κανονικών δεδομένων σε κανονικά με την χρήση της κατανομής Johnson. Για την κατάλληλη επιλογή του μετασχηματισμού θα χρησιμοποιηθεί το τεστ των Shapiro and Wilk. Το τεστ Shapiro–Wilk είναι ένα τεστ κανονικότητας σε στατιστικές συχνοτήτων και δημιουργήθηκε το 1965 από τους Samuel Sanford Shapiro και Martin Wilk .

Το 1949 ο Johnson ανέπτυξε μια οικογένεια κατανομών μιας μεταβλητής X όπου μετασχηματίζεται σε μια τυπική κανονική κατανομή. Η κατανομή Johnson περιλαμβάνει κάθε συνδυασμό μέσης τιμής, τυπικής απόκλισης, ασυμμετρίας και κύρτωσης. Η οικογένεια των κατανομών Johnson μπορεί να χρησιμοποιηθεί επίσης για να μοντελοποιήσει ένα δείγμα μονομεταβλητών δεδομένων (univariate data), για να υπολογίσει κρίσιμες τιμές, αλλά και ως γεννήτρια τυχαίων αριθμών για προσομοίωση Monte Carlo.

Υπάρχουν 3 διαφορετικοί τύποι της οικογένειας Johnson :

- SB: Φραγμένη κατανομή (Bounded distribution)
- SU: Μη φραγμένη κατανομή (Unbounded distribution)
- SL: Λογαριθμική κατανομή (Lognormal distribution)

Ο Πίνακας 3.3 παρουσιάζει τις συναρτήσεις που χρησιμοποιούνται για τον μετασχηματισμό του X σε μια τυπική κανονική μεταβλητή Z κατά το σύστημα Johnson.

Πίνακας 3.3: Πίνακας μετασχηματισμών για κάθε οικογένεια Johnson.

Johnson family	Transformation	Parameter Conditions	X Condition
S_B	$Z = \gamma + \eta \ln \left(\frac{X - \varepsilon}{\lambda + \varepsilon - X} \right)$	$\eta, \lambda > 0,$ $-\infty < \gamma,$ $\varepsilon < \infty$	$\varepsilon < X$ $< \varepsilon + \lambda$
S_L	$Z = \gamma + \eta \ln(X - \varepsilon)$	$\eta > 0,$ $-\infty < \gamma,$ $\varepsilon < \infty$	$X > \varepsilon$
S_U	$Z = \gamma + \eta \sinh^{-1} \left(\frac{X - \varepsilon}{\lambda} \right)$	$\eta, \lambda > 0,$ $-\infty < \gamma,$ $\varepsilon < \infty$	$-\infty < X < \infty$

Προκειμένου να γίνει ο μετασχηματισμός των μη κανονικών δεδομένων σε κανονικά, πρέπει να γίνει κατάλληλη επιλογή της κατανομής Johnson και να υπολογιστούν οι παράμετροι της κάθε οικογένειας. Έτσι εξετάζονται μέθοδοι που χρησιμοποιούν ένα δείγμα εκατοστημόριων για να επιλέξουν την κατάλληλη κατανομή Johnson και τις άγνωστες παραμέτρους της. Ο Johnson (1949) πρότεινε μια μέθοδο η οποία χρησιμοποιεί 4 εκατοστημόρια. Η διαδικασία αυτή αργότερα χρησιμοποιήθηκε σε μηχανολογικά δεδομένα. Η επιλογή αυτών των εκατοστημορίων μπορεί να είναι και τυχαία αλλά πολλοί συγγραφείς (π.χ. Bukac (1972), Slifker and Shapiro (1980), Bowman and Shenton (1988,1989)) κατέληξαν στην επιλογή 4 συμμετρικών τυπικών κανονικών αποκλίσεων $-sz, -z, z, sz$, όπου s και z είναι τυχαίες, θετικές σταθερές και ισχύει $s > 1$.

Θεωρούμε X μια τυχαία μεταβλητή η οποία ακολουθεί μια κατανομή η οποία ανήκει στην οικογένεια Johnson και δίνεται από τον Πίνακα 3.3. Θεωρούμε επίσης q_1, q_2, q_3, q_4 ως τις περιοχές μιας τυπικής κανονικής καμπύλης κάτω από $-sz, -z, z, sz$ αντίστοιχα. Σύμφωνα με τα παραπάνω θα ισχύει $q_1 = \Phi(-sz), q_2 = \Phi(-z), q_3 = \Phi(z), q_4 = \Phi(sz)$, όπου Φ η αθροιστική συνάρτηση μιας τυπικής κανονικής μεταβλητής. Θεωρούμε x_i ως το q_i^{th} τεταρτημόριο της κατανομής για $x=1, 2, 3, 4$. Έτσι ορίζουμε την ποσοστιαία αναλογία (Quantile Ratio) :

$$QR = \frac{(x_4 - x_3)(x_2 - x_1)}{(x_3 - x_2)^2}$$

Για την επιλογή της SU κατανομής : $QR < \frac{(s-1)^2}{4}$, όπου $1 < s \leq 3$

Για την επιλογή της SB κατανομής : $QR > \frac{(s-1)^2}{4}$, όπου $s \geq 3$.

Με τον παραπάνω τρόπο υπολογίζονται και οι παράμετροι των κατανομών για δείγμα X_1, X_2, \dots, X_n . Ένα πλεονέκτημα της μεθοδολογίας είναι πως μπορούμε να επιλέξουμε τιμή του s ώστε να μην απαιτείται παρεμβολή δεδομένων για τον υπολογισμό των εκατοστημορίων. Για οποιαδήποτε τιμή $s > 1$ δεν υπάρχει τρόπος να ξεχωριστούν οι κατανομές της οικογένειας Johnson, εκτός από την τιμή $s=3$. Γι' αυτό τον λόγο επιλέγουμε να χρησιμοποιήσουμε $s=3$ και με το κριτήριο του Silker and Shapiro ανάλογα την τιμή του QR καθορίζονται και οι κατανομές της οικογένειας Johnson.

- Εάν το X έχει κατανομή SB τότε $QR < 1$
- Εάν το X έχει κατανομή SL τότε $QR = 1$
- Εάν το X έχει κατανομή SU τότε $QR > 1$

Οι Chou et al. (1994) παρουσιάζουν κατάλληλους τύπους για τον υπολογισμό του QR σε περιπτώσεις όπου η μεταβλητή ακολουθεί διάφορες κατανομές, όπως

ομοιόμορφη κατανομή, κατανομή βήτα ή γάμμα, λογαριθμική κατανομή, κανονική ή t κατανομή. Κάθε κατανομή ταξινομείται ως μέλος της οικογένειας Johnson ανάλογα με το εύρος του, τα μέτρα λοξότητας του και την κυρτότητα του και συμβολίζεται με β_1, β_2 . Οι παράμετροι αυτοί θεωρούνται άγνωστοι, εφόσον η κατανομή του X είναι άγνωστη. Σε αυτές τις περιπτώσεις όπου έχουμε δεδομένα που προέρχονται από άγνωστες κατανομές, η πραγματική τιμή του QR είναι άγνωστη και υπολογίζεται από τον παρακάτω τύπο :

$$\widehat{QR} = \frac{(\widehat{x}_4 - \widehat{x}_3)(\widehat{x}_2 - \widehat{x}_1)}{(\widehat{x}_3 - \widehat{x}_2)^2}$$

Το \widehat{x}_i ορίζεται ως το q_i^{th} τεταρτημόριο του δείγματος με μέγεθος n . Το \widehat{x}_i είναι η j^{th} παρατήρηση όπου ως $j = [nq_i] = 1$ ή $j = [(n+1)q_i]$. Σύμφωνα με τους Shilker and Shapiro προτείνεται $j = nq_i + 0,5$ και παρεμβολή σε περιπτώσεις όπου το j δεν είναι ακέραιο. Για μέτρια και μεγάλα δείγματα, θα υπάρχει μικρή διαφορά για την χρήση αυτών των τύπων.

Εφόσον το \widehat{QR} αποτελεί μια εκτίμηση αναλογίας από στατιστικές συναρτήσεις, για τον υπολογισμό του αποτελεί μια εκτίμηση καθώς εξαρτάται σε μεγάλο βαθμό από τον πίνακα συνδιακύμανσης των 4 στατιστικών τάξεων. Επίσης, εφόσον το \widehat{x}_i εξαρτάται από το z τότε και η τιμή του \widehat{QR} θα εξαρτάται από το z , το n και τα δείγματα δεδομένων. Σύμφωνα με την εξίσωση επιλογής οικογένειας Johnson αν αντικαταστήσουμε το QR με την εκτίμηση \widehat{QR} τότε η τιμή z που επιλέγουμε μπορεί να καθορίσει την οικογένεια Johnson. Σε περιπτώσεις όπου η κατανομή του X είναι γνωστή τότε το κριτήριο μπορεί να ισχύει για όλες τις περιπτώσεις του z . Όσο και αν μεταβάλλεται η μεταβλητότητα του ποσοστού δείγματος σημαίνει ότι το κριτήριο δεν θα κάνει διακρίσεις. Αυτό υποδηλώνει ότι για να εξασφαλιστεί η καλή εφαρμογή για ένα δείγμα, είναι καλύτερο να εξεταστεί η εφαρμογή της κατανομής Johnson για πολλές τιμές του z , επιλέγοντας εκείνη που μετατρέπει καλύτερα τα δεδομένα σε κανονικότητα.

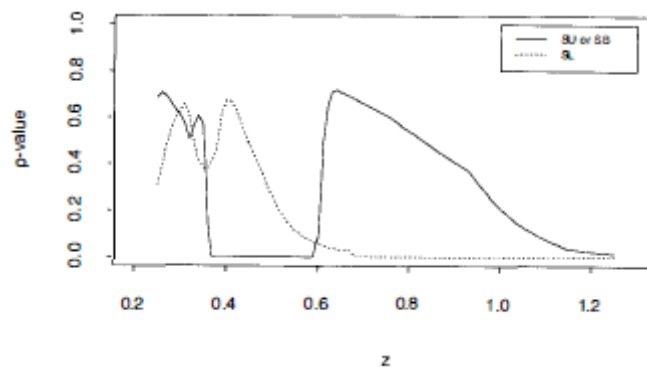
Στην συνέχεια, για να εντοπίσουμε την καλύτερη κατανομή Johnson για ένα σύνολο δεδομένων θα χρησιμοποιήσουμε το τεστ των Shapiro-Wilk. Το στατιστικό αυτό τεστ W προϋποθέτει μια τιμή α η οποία βρίσκεται μεταξύ 0 και 1 και χρησιμοποιείται για τον έλεγχο της κανονικότητας. Οι μικρές τιμές W υποδεικνύουν μη κανονικότητα.

Έτσι η υπόθεση κανονικότητας απορρίπτεται σε συγκεκριμένο επίπεδο α , εάν δηλαδή η τιμή του W είναι μικρότερη από μία τιμή w_α , όπου w_α είναι το α^{th} τεταρτημόριο της στατιστικής δοκιμής (Conover, 1980). Για ένα σύνολο δεδομένων, η τιμή p υποδηλώνει την πιθανότητα να λάβουμε μια τιμή W μικρότερη από την πραγματική τιμή που υπολογίστηκε για το δεδομένα.

Συνεπώς οι μικρές τιμές p υποδηλώνουν μη κανονικότητα, ενώ οι μεγάλες τιμές p υποδεικνύουν δεδομένα με κανονική κατανομή. Όπως αναφέρθηκε και παραπάνω, η επιλογή μιας κατανομής Johnson εξαρτάται από την κατάλληλη επιλογή του z , έτσι

για το βέλτιστο z που θα ταιριάζει στα προσαρμοσμένα δεδομένα Johnson θα ορίζουμε και την μέγιστη τιμή W . Στην περίπτωση αυτή η κατανομή Johnson θα θεωρείται η καταλληλότερη για τα δεδομένα που έχουμε.

Σημαντικό κομμάτι της διαδικασίας είναι η εύρεση της βέλτιστης τιμής z . Προκειμένου να επιτευχθεί χρειάζεται καθορισμός του εύρους της βέλτιστης τιμής. Σύμφωνα με τους Owen and Li (1988), ορίζεται ένα εύρος του z από 0,05 έως 5,00 με προσαυξήσεις 0,05 ώστε να χωρέσουν τα δεδομένα της κατανομής Johnson. Οι συγγραφείς θεωρούν πως όλες οι μέγιστες τιμές του W υπερβαίνουν την τιμή 0.90 αλλά δεν αναφέρουν τις τιμές του p . Οι τιμές του $W > 0.90$ δεν προϋποθέτουν πάντα καλή εφαρμογή και έτσι σε πολλές περιπτώσεις όπου η τιμή του W μεγαλύτερη του 0.90 αντιστοιχεί σε τιμή του p 0.05 ή και μικρότερη. Από μελέτες που πραγματοποιήθηκαν σε δείγματα διαφόρων μη κανονικών κατανομών, προέκυψε πως δεν υπάρχει κάποια συσχέτιση μεταξύ του z και του p (βλ. Σχήμα 3.1).



Σχήμα 3.1: Γράφημα μεταξύ της τιμής p και z .

Σύμφωνα με μια μελέτη προσομοίωσης από τους Mandraccia et al. (1996) με δεδομένα από τη Sony Semiconductor Company των ΗΠΑ, προκύπτει ότι οι τιμές p είναι γενικά μεγαλύτερες από 0,4 όταν χρησιμοποιούμε ένα πλέγμα z με εύρος από 0,25 έως 1,25 με αύξηση της τάξης του 0,01 για ένα ευρύ φάσμα κατανομών. Έτσι, προκύπτει ότι το δυνατό εύρος για το βέλτιστο z είναι $S = \{z : z = 0,25, 0,26, \dots, 1.25\}$, το οποίο αποτελείται από 101 τιμές z . Συνεπώς η χρήση τιμών z εκτός εύρους καθώς και οι τιμές του p μικρότερες από 0,05 υποδεικνύουν κακή εφαρμογή. Η χρήση ακραίων τιμών z οδηγεί σε ακραία ποσοστά που έχουν μεγάλο τυπικό σφάλμα και είναι ευαίσθητα σε ακραίες τιμές.

Για ένα σύνολο δεδομένων X_1, \dots, X_n που προέρχεται από μια άγνωστη κατανομή θα παρουσιαστούν κάποια βήματα προκειμένου να εντοπιστεί η βέλτιστη εφαρμογή για τα δεδομένα.

- 1) Υπολογισμός του W και του αντίστοιχου p ώστε να υποθέσουμε κανονικότητα δεδομένων. Ο υπολογισμός του W πραγματοποιείται από συναρτήσεις σε

κατάλληλο πρόγραμμα ενώ η τιμή του p δίνεται από τον Royston (1982, 1983). Σε περίπτωση μη κανονικότητας η διαδικασία δεν μπορεί να συνεχιστεί.

- 2) Για κάθε z σε $S = \{z : z = 0,25, 0,26, \dots, 1,25\}$, υπολογίζουμε το \widehat{QR} . Σε περίπτωση όπου $\widehat{QR} < 1$ τα δεδομένα ταιριάζουν σε SB, SL διαφορετικά ταιριάζουν σε SU, SL. Για το μετασχηματισμό οι τύποι δίνονται από τον Πίνακα 3.3 και οι παράμετροι υπολογίζονται από τους τύπους στο Παράρτημα Α.
- 3) Γίνεται έλεγχος των δεδομένων που ταιριάζουν σε κατανομές SB, SL και εάν τα δεδομένα βρίσκονται εκτός του εύρους του X . Σε περίπτωση όπου κάποια δεδομένα βρεθούν εκτός του εύρους του X , τότε η κατανομή δεν ταιριάζει με τα δεδομένα και δεν χρησιμοποιούμε την κατανομή.
- 4) Για κάθε z στο S και την αντίστοιχη προσαρμοσμένη κατανομή (ή μετασχηματισμό), γίνεται υπολογισμός του W και της σχετική p τιμή για το μετασχηματισμένο σύνολο δεδομένων.
- 5) Η εκτιμώμενη καλύτερη εφαρμογή του Johnson συνδέεται με τον μετασχηματισμό έχοντας τη μεγαλύτερη τιμή του W , η οποία θα αντιστοιχεί και στη μεγαλύτερη τιμή του p , εφόσον υπάρχει αντιστοιχία των δύο αυτών τιμών. Έτσι, εφόσον τα μετασχηματισμένα δεδομένα είναι σχεδόν κανονικά, η διαδικασία στατιστικού ελέγχου μπορεί να εφαρμοστεί.

Για την κατάλληλη διερεύνηση των δεδομένων και την προσαρμογή σε μια κατανομή Johnson είναι ιδιαίτερα χρήσιμη η στατιστική W . Αυτό συμβαίνει καθώς η επιλογή μετασχηματισμού μόνο με βάση το \widehat{QR} μπορεί να οδηγήσει σε εσφαλμένα αποτελέσματα. Εάν τα δεδομένα ταιριάζουν με μια SB κατανομή, τότε το σύνολο δεδομένων μπορεί να ανήκει σε ένα εύρος $(\varepsilon, \varepsilon + \lambda)$. Αντίστοιχα εάν τα δεδομένα ταιριάζουν με μια κατανομή SL, θα ανήκουν στο εύρος (ε, ∞) . Έτσι είναι απαραίτητος ο έλεγχος του εύρους των δεδομένων.

Εφόσον οι μετασχηματισμοί βασίζονται σε εκτιμώμενες παραμέτρους, μπορεί να προκύψουν μη ικανοποιητικές προσαρμογές. Αυτό το φαινόμενο αντιστοιχεί σε μια χαμηλή μέγιστη τιμή p . Ωστόσο, στην προσαρμογή μη κανονικών δεδομένων από ένα ευρύ φάσμα κατανομών, φαίνεται πως το σύστημα Johnson είναι γενικά ευέλικτο ώστε να βρεθεί ένας αποδεκτός μετασχηματισμός δεδομένων.

Στα γραφήματα μεταβλητών συχνά υποθέτουμε πως το χαρακτηριστικό ποιότητας είναι κανονικά κατανομημένο και υπολογίζουμε την εκτίμηση τυπικής απόκλισης με την υπόθεση αυτή. Από την παραπάνω μεθοδολογία φαίνεται ότι οι αποκλίσεις από την κανονικότητα μπορούν να οδηγήσουν σε αύξηση του σφάλματος μέσου τετραγώνου (mean-square error) της εκτίμησης της τυπικής απόκλισης πληθυσμού. Ωστόσο το πρόβλημα αυτό μπορεί να διορθωθεί με τη βέλτιστη εφαρμογή του μετασχηματισμού Johnson που ταιριάζει καλύτερα στα δεδομένα. Η μεθοδολογία αυτή είναι ιδιαίτερα χρήσιμη και αποτελεσματική για μικρά σύνολα δεδομένων (περίπου 10). Η μεθοδολογία αυτή δίνει τη δυνατότητα στους χρήστες να επιλέξουν

τους κατάλληλους μετασχηματισμούς στην κανονικότητα και έτσι διασφαλίζει την εγκυρότητα της υπόθεσης κανονικότητας για τις επιλεγμένες τεχνικές στατιστικού ελέγχου.

3.4 Μετασχηματισμός δεδομένων

3.4.1 Απλοί μετασχηματισμοί

Ένας άλλος τρόπος αντιμετώπισης της μη κανονικότητας των δεδομένων είναι ο μετασχηματισμός δεδομένων με τη χρήση κατάλληλων συναρτήσεων. Με αυτόν τον τρόπο το σύνολο των δεδομένων που έχουμε θα προσεγγίζει περίπου την κανονική κατανομή και έτσι θα έχουμε την δυνατότητα εφαρμογής των στατιστικών εργαλείων που βασίζονται στην υπόθεση κανονικότητας.

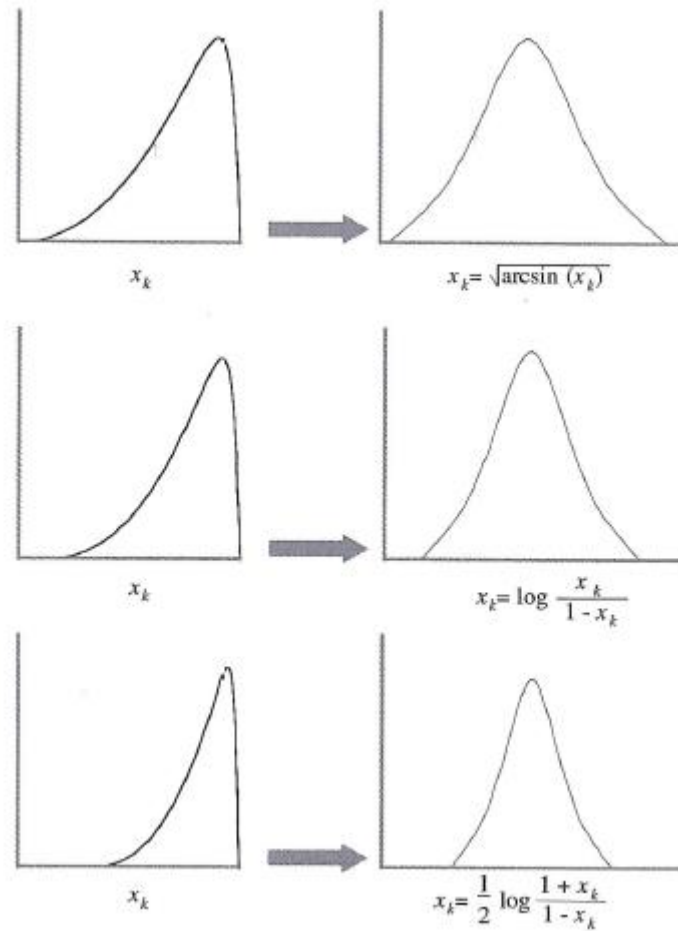
Στο Σχήμα 3.2 παρουσιάζονται κάποιες βασικές περιπτώσεις όπου λαμβάνουμε από έρευνες ικανοποίησης, όπου συναντάμε προβλήματα κατανομής. Η εφαρμογή στατιστικών τεχνικών που υποθέτουν κανονική κατανομή θα οδηγήσει σε εσφαλμένα αποτελέσματα λόγω έλλειψης διακύμανσης τους. Οι τρεις περιπτώσεις που θα αναλυθούν έχουν αυξημένη λοξότητα. Για κάθε περίπτωση μπορεί να χρησιμοποιηθεί ένας διαφορετικός μετασχηματισμός ώστε να αποκτήσει κανονική μορφή. Οι μετασχηματισμοί αυτοί είναι μονοτονικοί δηλαδή διατηρούν την σειρά των παρατηρήσεων στις κατανομές. Συγκεκριμένα αν μια παρατήρηση έχει μεγαλύτερη τιμή από μια άλλη αυτή η διαφορά θα διατηρηθεί και μετά τον μετασχηματισμό όπως ισχύει και για την μέγιστη και την ελάχιστη τιμή.

Οι μετασχηματισμοί όπως φαίνεται προσπαθούν να ελαττώσουν την λοξότητα των δεδομένων. Οι μετασχηματισμοί αυτοί δεν μπορούν να δημιουργήσουν διακύμανση.

3.4.2 Box-Cox μετασχηματισμός

Ο Box-Cox είναι ένας ιδιαίτερα χρήσιμος μετασχηματισμός ο οποίος μετατρέπει μη κανονικά δεδομένα σε κανονικά. Ο μετασχηματισμός αυτός πήρε το όνομα του από τους στατιστικούς George Box και Sir David Roxbee Cox το 1964. Ο μετασχηματισμός αυτός έχει σκοπό τον εντοπισμό ενός κατάλληλου λ το οποίο θα χρησιμοποιηθεί για την μετατροπή των δεδομένων σε κανονικό σχήμα. Το λ αυτό λαμβάνει τιμές από -5 έως 5, μέχρι να βρεθεί η καλύτερη τιμή που προσεγγίζει την κανονική. Ο μετασχηματισμός έχει την μορφή :

$$y(\lambda) = \begin{cases} \frac{y^\lambda}{\lambda}, & \text{αν } \lambda \neq 0 \\ \log(y), & \text{αν } \lambda = 0 \end{cases}$$



Σχήμα 3.2: Μετασχηματισμοί δεδομένων

Η παραπάνω σχέση χρησιμοποιείται μόνο για θετικά δεδομένα και το y είναι η παράμετρος μετασχηματισμού.

Σε περίπτωση αρνητικών τιμών χρησιμοποιείται ο παρακάτω τύπος, όπου λ_2 μια σταθερά ώστε τα δεδομένα να γίνουν θετικά πριν μετασχηματιστούν.

$$y(\lambda) = \begin{cases} \frac{(y + \lambda_3)^{\lambda_1} - 1}{\lambda_1}, & \text{αν } \lambda_1 \neq 0 \\ \log(y + \lambda_2), & \text{αν } \lambda_2 = 0 \end{cases}$$

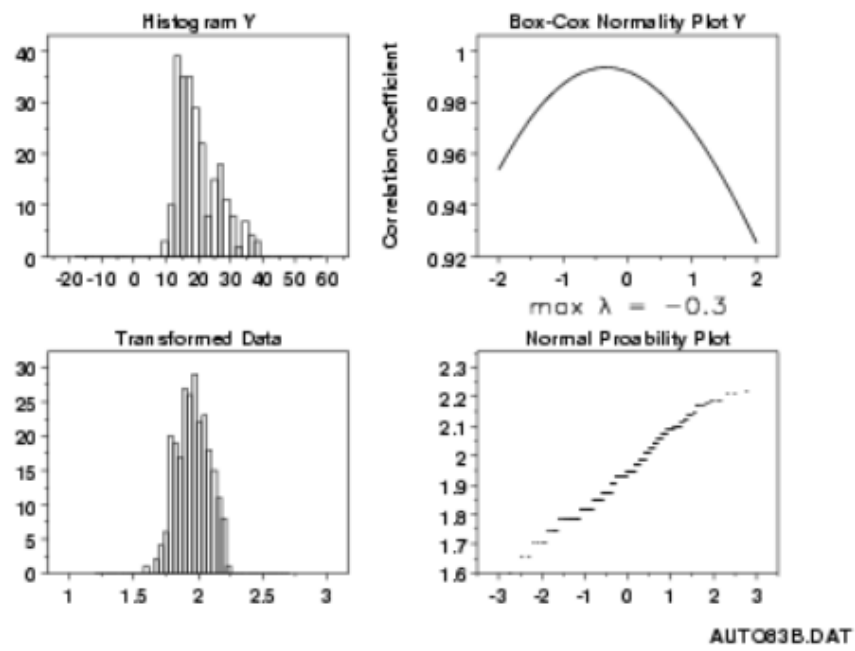
Παρακάτω παρουσιάζεται ένας πίνακας με τους κοινούς μετασχηματισμούς Box-Cox, όπου Y' ο μετασχηματισμός των αρχικών δεδομένων Y . Στην περίπτωση που $\lambda=0$, ο μετασχηματισμός δεν είναι Y (γιατί αυτό θα ισούταν με 1 για κάθε τιμή), αλλά είναι ίσος με τον λογάριθμο του Y .

Πίνακας 3.4: Μετασχηματισμοί Box-Cox.

Lambda value (λ)	Transformed data (Y)
-3	$Y^{-3} = \frac{1}{Y^3}$
-2	$Y^{-2} = \frac{1}{Y^2}$
-1	$Y^{-1} = \frac{1}{Y^1}$
-0.5	$Y^{-0.5} = \frac{1}{Y^{0.5}}$
0	$\log(Y)$
0.5	$Y^{0.5} = Y^{0.5}$
1	$Y^1 = Y$
2	Y^2
3	Y^3

Ο μετασχηματισμός Box-Cox δεν αποτελεί εγγύηση για την κανονικότητα των δεδομένων. Ο μετασχηματισμός προσπαθεί να βρει τη βέλτιστη επιλογή του λ, ώστε τα δεδομένα να προσεγγίζουν την κανονική κατανομή, αλλά για καλύτερα αποτελέσματα απαιτείται ένα γράφημα πιθανοτήτων. Ένα μέτρο κανονικότητας του μετασχηματισμού είναι ο υπολογισμός του συντελεστή συσχέτισης μιας κανονικής γραφικής παράστασης πιθανότητας. Η συσχέτιση μπορεί να υπολογιστεί μεταξύ των μεταβλητών του κάθετου άξονα και του οριζόντιου του σχεδίου πιθανότητας. Ο οριζόντιος άξονας αντιστοιχεί στην τιμή λ και ο κατακόρυφος άξονας αντιστοιχεί στον συντελεστή συσχέτισης από την κανονική γραφική παράσταση πιθανότητας μετά την εφαρμογή του μετασχηματισμού Box-Cox. Το διάγραμμα Box-Cox αποτελεί μια γραφική παράσταση του υπολογισμού του συντελεστή συσχέτισης .

Παρακάτω παρουσιάζεται ένα παράδειγμα γραφικής απεικόνισης Box-Cox. Συγκεκριμένα στο πάνω αριστερά ιστόγραμμα παρουσιάζεται ένα σύνολο δεδομένων το οποίο δεν ακολουθεί την κανονική κατανομή καθώς παρουσιάζει μια δεξιά λοξότητα. Το διάγραμμα κανονικότητας του Box-Cox δείχνει ότι για $\lambda = -0,3$ έχουμε την μέγιστη τιμή του συντελεστή συσχέτισης. Μετά την χρήση για $\lambda = -0,3$ προκύπτει το νέο ιστόγραμμα δεδομένων το οποίο φαίνεται να έχει την μορφή κανονικής κατανομής. Η κανονικότητα των δεδομένων επαληθεύεται από το γράφημα πιθανοτήτων.



Σχήμα 3.3: Παράδειγμα γραφικής απεικόνισης Box-Cox.

Κεφάλαιο 4ο: Εφαρμογή εναλλακτικών διαγραμμάτων ελέγχου

4.1 Σχεδιασμός προσομοίωσης

Σκοπός της διπλωματικής εργασίας είναι η ανάλυση δεδομένων που προέρχονται από έρευνες ικανοποίησης πελατών. Όπως αναφέρθηκε και παραπάνω τα δεδομένα αυτά δεν ακολουθούν την κανονική κατανομή λόγω της φύσης των δεδομένων ή των διαδικασιών καθαυτών, γι' αυτό υπάρχουν διάφοροι τρόποι για την ανάλυση των μη κανονικών δεδομένων. Ένας τρόπος ανάλυσης των δεδομένων είναι η δημιουργία κατάλληλων διαγραμμάτων ελέγχου.

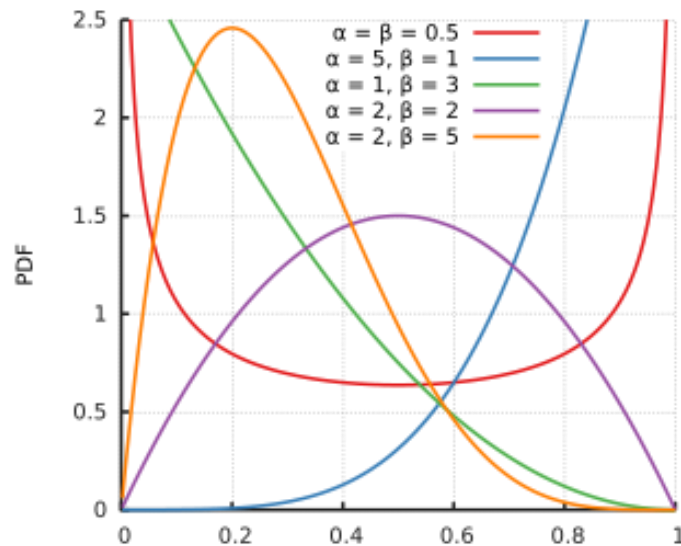
Ξεκινώντας θα δημιουργηθούν δεδομένα ανάλογα με δεδομένα που λαμβάνουμε από μια έρευνα ικανοποίησης. Τα δεδομένα αυτά στην πραγματικότητα ανάλογα την ποιότητα του προσφερόμενου προϊόντος ή υπηρεσίας τείνουν να ακολουθούν μια λοξή κατανομή γερμένη προς τα δεξιά. Συνεπώς η κατανομή αυτή δεν μοιάζει με κανονική. Για να δημιουργήσουμε τα κατάλληλα δεδομένα θα χρησιμοποιηθεί περιβάλλον Excel, τόσο για τη δημιουργία δεδομένων αλλά και για την δημιουργία διαγραμμάτων ελέγχου. Για την προσομοίωση των δεδομένων ικανοποίησης χρησιμοποιήθηκε η κατανομή Βήτα. Η κατανομή Βήτα είναι μια οικογένεια συνεχών κατανομών πιθανοτήτων που ορίζονται στο διάστημα $[0, 1]$ παραμετροποιούνται από δύο θετικές παραμέτρους σχήματος, που συμβολίζονται με άλφα (α) και βήτα (β), που εμφανίζονται ως εκθέτες στην τυχαία μεταβλητή και ελέγχουν το σχήμα της κατανομής. Η κατανομή Βήτα έχει εφαρμοστεί για να μοντελοποιήσει τη συμπεριφορά τυχαίων μεταβλητών που περιορίζονται σε διαστήματα πεπερασμένου μήκους σε μια μεγάλη ποικιλία επιστημονικών κλάδων.

Στο θεώρημα Bayes, η κατανομή Βήτα είναι η συζευγμένη προηγούμενη κατανομή πιθανότητας για τις κατανομές Bernoulli, διωνυμικές, αρνητικές διωνυμικές και γεωμετρικές κατανομές. Η κατανομή βήτα είναι ένα κατάλληλο μοντέλο για την τυχαία συμπεριφορά ποσοστών και αναλογιών. Γενικά η κατανομή Βήτα αποτελεί μια ευέλικτη κατανομή καθώς μεταβάλλοντας τις παραμέτρους α και β η γραφική της μπορεί να πάρει διάφορες μορφές που πολλές από αυτές προσεγγίζουν την κανονική κατανομή.

Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Βήτα:

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} I_{(0,1)}(x)$$

με $\alpha, \beta > 0$



Σχήμα 4.1: Γραφική απεικόνιση συνάρτησης πυκνότητας πιθανότητας της κατανομής Βήτα

4.2 Δημιουργία δεδομένων

Σε περιβάλλον Excel για την καλύτερη κατανόηση των στατιστικών διαδικασιών χρησιμοποιήθηκε η κατανομή Βήτα με διαφορετικούς συνδυασμούς παραμέτρων α και β . Στην πρώτη περίπτωση χρησιμοποιήθηκε $\alpha=15$ και $\beta=4$, στην δεύτερη περίπτωση χρησιμοποιήθηκε $\alpha=\beta=15$ και στην τρίτη περίπτωση χρησιμοποιήθηκε $\alpha=3$ και $\beta=2$. Σκοπός των διαφορετικών παραμέτρων είναι ο εντοπισμός της καλύτερης προσομοίωσης της κατανομής Βήτα στην μορφή της κανονικής κατανομής όπου εμείς επιθυμούμε. Τα δεδομένα που θέλουμε να δημιουργήσουμε αντιστοιχούν σε μια περίοδο 40 εβδομάδων. Συγκεκριμένα κάθε εβδομάδα δημιουργούμε 100 απαντήσεις με κλίμακα το 1 έως το 5. Το 1 αντιστοιχεί στην περίπτωση όπου ο πελάτης είναι δυσαρεστημένος, ενώ το 5 στην περίπτωση όπου ο πελάτης είναι ικανοποιημένος.

Παρακάτω θα παρουσιαστούν τα βήματα που ακολουθήθηκαν για τη δημιουργία των δεδομένων και έπειτα θα γίνει η παρουσίαση των διαγραμμάτων ελέγχου.

1. Ξεκινώντας την δημιουργία δεδομένων, δημιουργούμε 100 τυχαίους αριθμούς μέσα από την συνάρτηση RAND() του Excel και σε διάστημα $[0,1]$ για κάθε εβδομάδα.
2. Με χρήση της συνάρτησης BETA.INV του Excel επιστρέφουμε την αντίστροφη της αθροιστικής συνάρτησης πυκνότητας πιθανότητας της κατανομής βήτα. Τα ορίσματα της συνάρτησης είναι BETA.INV(probability, alpha, beta, [lower bound], [upper bound]). Στην περίπτωση μας ως πιθανότητα χρησιμοποιούμε τον τυχαίο αριθμό που δημιουργήσαμε προηγουμένως, για τα α και β αντικαθιστούμε τις τιμές ανάλογα την περίπτωση και ως άνω και κάτω όριο

ορίσαμε το 5 και το 1, αντίστοιχα, λόγω της κλίμακας της έρευνας ικανοποίησης.

3. Με την εντολή ROUND() έγινε στρογγυλοποίηση στην κοντινότερη ακέραια τιμή της παραπάνω τιμής.
4. Η διαδικασία αυτή ακολουθήθηκε για σύνολο 40 εβδομάδες , όπου κάθε εβδομάδα δημιουργούνται 100 απαντήσεις με κλίμακα από το 1 έως το 5. Το 1 αντιστοιχεί στην περίπτωση όπου ο πελάτης είναι δυσαρεστημένος ενώ το 5 αντιστοιχεί στην περίπτωση όπου ο πελάτης είναι ικανοποιημένος.
5. Για κάθε εβδομάδα έγινε υπολογισμός των απαντήσεων για κάθε κατηγορία ξεχωριστά. Με την εντολή COUNTIF() έγινε μέτρηση των κελιών κάθε εβδομάδας όπου πληρούν ένα κριτήριο. (COUNTIF(B2:B501),"1"), για την καταμέτρηση των κελιών όπου η απάντηση ισούται με 1).
6. Για κάθε τιμή που υπολογίστηκε παραπάνω υπολογίζουμε την πιθανότητα μια ομάδα ανθρώπων να έχουν δώσει απάντηση x από το σύνολο των απαντήσεων δηλαδή p_x , όπου $x= 1, 2, 3, 4$.
7. Την ίδια διαδικασία ακολουθώ για κάθε εβδομάδα ώστε να υπολογίσω το p_x .
8. Τέλος θα γίνει εφαρμογή των εναλλακτικών διαγραμμάτων ελέγχου και των αντίστοιχων τύπων .

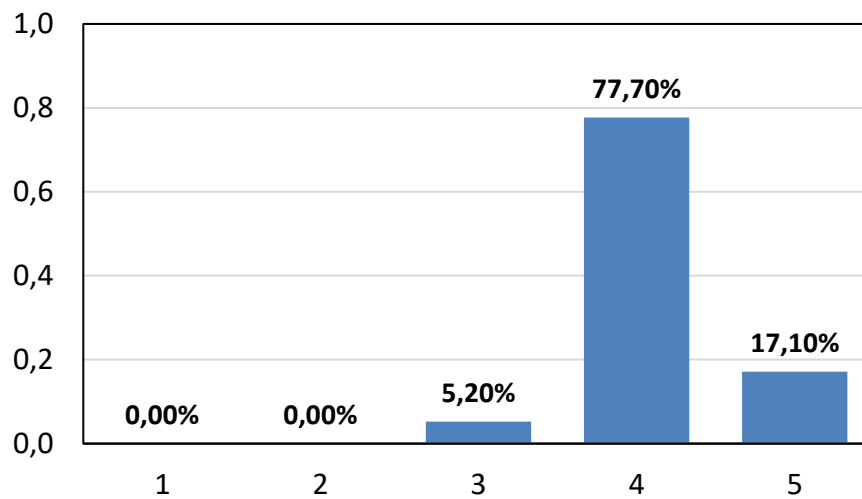
4.3 Παρουσίαση αποτελεσμάτων

Τα δεδομένα που δημιουργήθηκαν από την παραπάνω διαδικασία παρουσιάζονται στα παρακάτω γραφήματα για κάθε περίπτωση των παραμέτρων α και β .

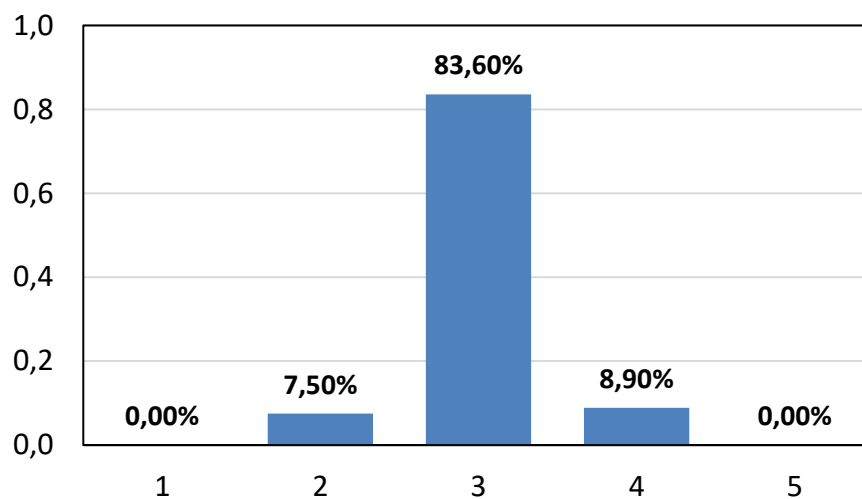
Για την πρώτη περίπτωση(1^ο σετ δεδομένων) όπου $\alpha=15$ και $\beta=4$ η κατανομή των δεδομένων ικανοποίησης σε όλη την περίοδο 40 εβδομάδων κυμαίνεται παρόμοια με την κανονική αλλά είναι γερμένη προς τα δεξιά.

Για την δεύτερη περίπτωση(2^ο σετ δεδομένων) όπου $\alpha=\beta=15$ η κατανομή των δεδομένων μοιάζει περισσότερο με την κανονική κατανομή, καθώς τα δεδομένα συγκεντρώνονται κυρίως στο μέσο της κλίμακας.

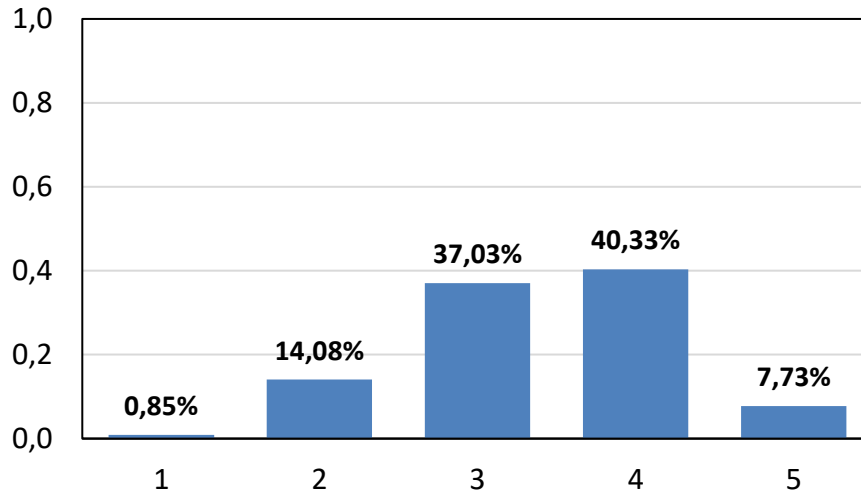
Για την περίπτωση όπου $\alpha=3$ και $\beta=2$ (3^ο σετ δεδομένων) η κατανομή των δεδομένων ικανοποίησης σε όλη την περίοδο 40 εβδομάδων κυμαίνεται παρόμοια με την κανονική αλλά είναι γερμένη προς τα δεξιά.



Σχήμα 4.1: Κατανομή του 1^{ου} σετ δεδομένων



Σχήμα 4.2: Κατανομή του 2^{ου} σετ δεδομένων



Σχήμα 4.3: Κατανομή του 3^{ου} σετ δεδομένων

4.4 Τυπικά διαγράμματα ελέγχου Shewhart

Τα τυπικά διαγράμματα ελέγχου υποθέτουν πως τα δεδομένα που συλλέγονται είναι κανονικά. Όμως τα δεδομένα μας είναι βαθμωτά και έτσι οι παρατηρήσεις που προκύπτουν δεν ακολουθούν την κανονική κατανομή αλλά την προσεγγίζουν.

Η χρήση του διαγράμματος \bar{x} προϋποθέτει πως το δείγμα που χρησιμοποιείται είναι σχετικά μικρό και παραμένει σταθερό, όμως κάτι τέτοιο δε συμβαίνει με τα δεδομένα της έρευνας (100 απαντήσεις ανά βδομάδα).

Για την εφαρμογή του διαγράμματος \bar{x} υπολογίστηκε η τιμή του μέσου όρου:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$UCL = \bar{x} + k \frac{\sigma}{n}$$

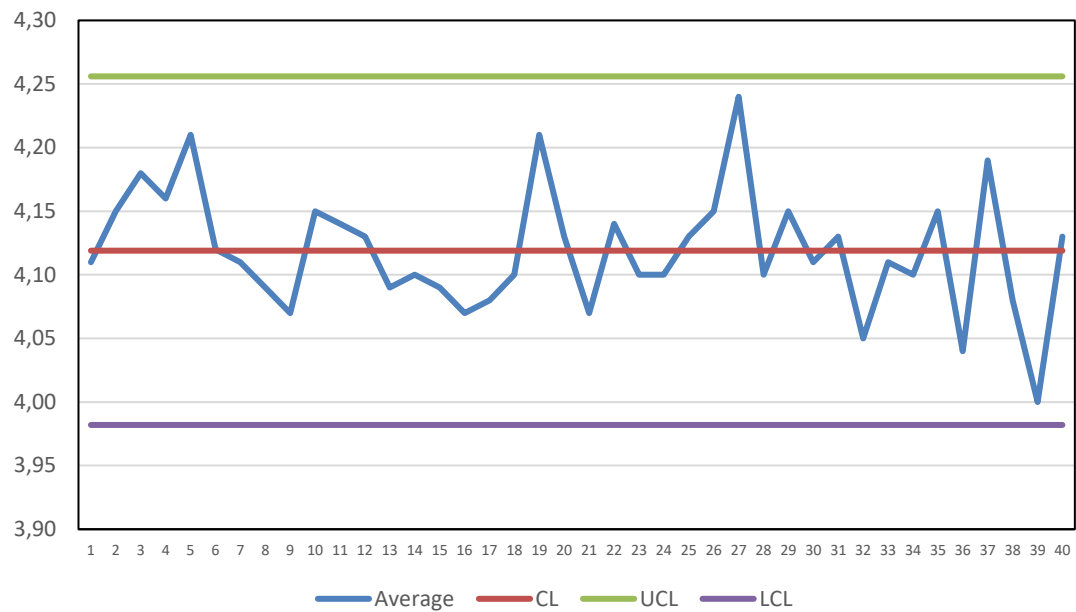
$$CL = \bar{x}$$

$$LCL = \bar{x} - k \frac{\sigma}{n}$$

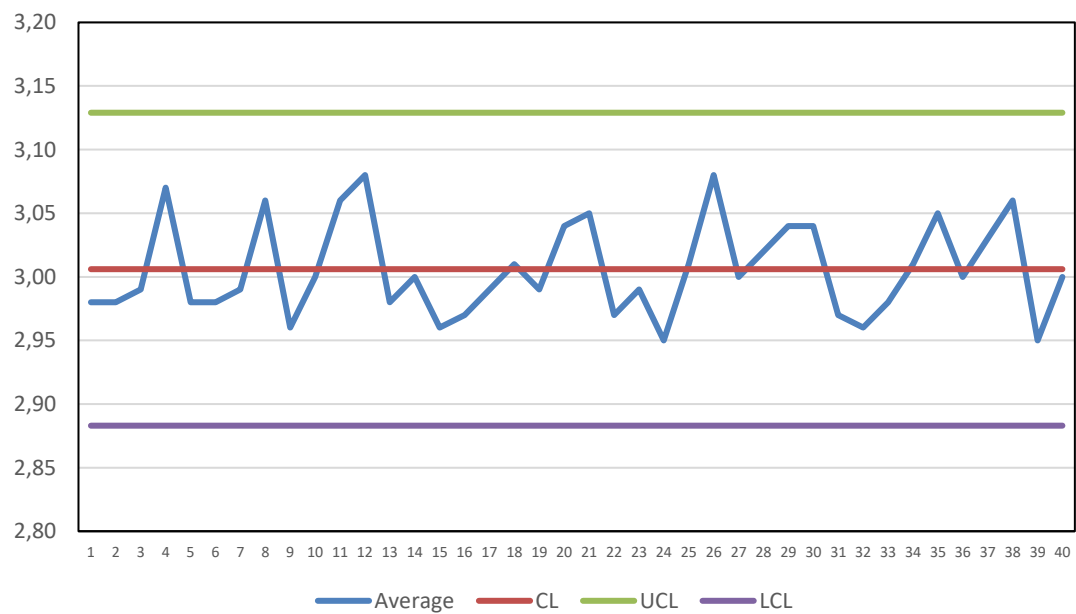
Για τον υπολογισμό της τυπικής απόκλισης σ χρησιμοποιήθηκε ο τύπος:

$$\sigma = \frac{\bar{s}}{c_4}$$

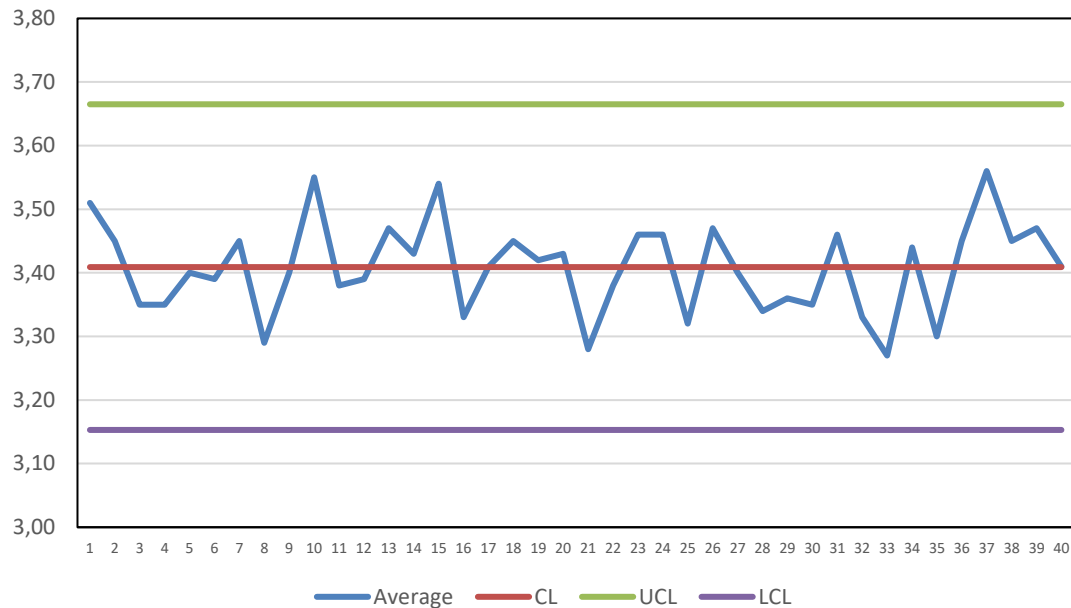
$$c_4 = \sqrt{\frac{2}{n-1} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}-\frac{1}{2})}}$$



Σχήμα 4.4: X-bar για το 1^ο σετ δεδομένων ($\alpha=15$, $\beta=4$)



Σχήμα 4.5: X-bar για το 2^ο σετ δεδομένων ($\alpha=\beta=15$)



Σχήμα 4.6: X-bar για το 3^ο σετ δεδομένων (α=3, β=2)

4.5 Διάγραμμα ελέγχου επέκτασης του Χρ

Για την κατασκευή του διαγράμματος ελέγχου της επέκτασης του Χρ χρησιμοποιείται η τυπική απόκλιση η οποία δεν βασίζεται στην κανονική κατανομή αλλά στην κατανομή των δεδομένων. Συγκεκριμένα η κατασκευή των ορίων βασίζεται στην κατανομή του αθροίσματος των x_s . Το γράφημα αυτό βασίζεται στο μέγεθος της υποομάδας να είναι μεγάλο αρκετά ώστε να είναι περίπου κανονικά κατανεμημένο και να χρησιμοποιούνται ± 3 τυπικές αποκλίσεις. Έχουμε πιθανές απαντήσεις των πελατών που κυμαίνονται από 1 έως 5, όπου 5 η υψηλότερη τιμή και 1 η χαμηλότερη. Έτσι το Χ μπορεί να πάρει τιμές από το 1 έως το 5 και η πιθανότητα να δώσει κάποιος πελάτης απάντηση x είναι διακριτή τυχαία μεταβλητή με συνάρτηση μάζας πιθανότητας $f(x)=\Pr(X=x)=p_x$.

Για κάθε εβδομάδα υπολογίζω τη μέση τιμή και την τυπική απόκλιση χρησιμοποιώντας τους παρακάτω τύπους .

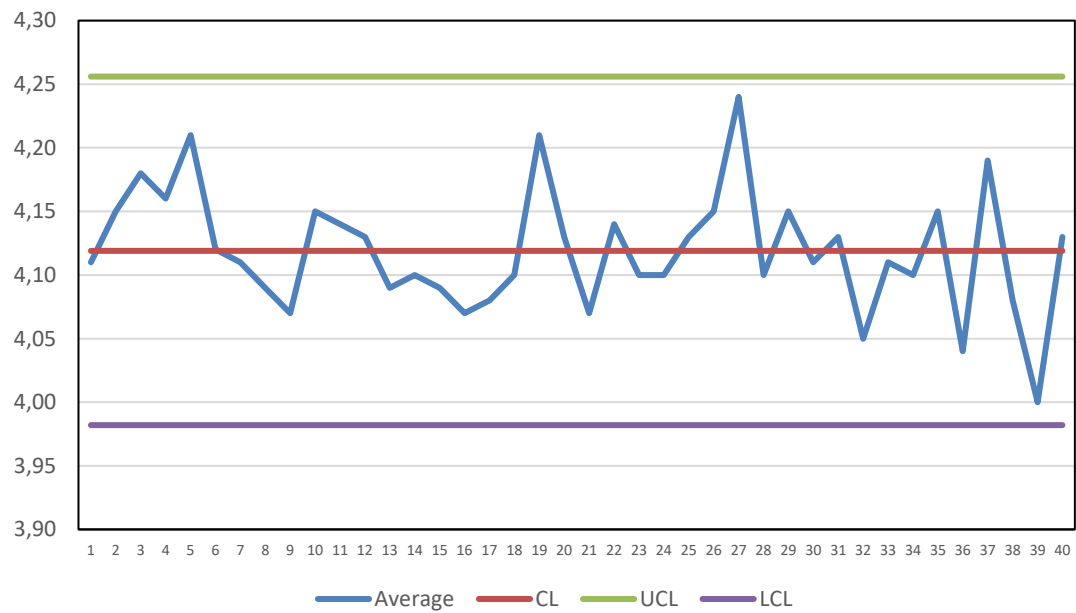
$$\mu_x = \sum_{1}^k x p_x$$

$$\sigma_x^2 = \sum_{1}^k x^2 p_x - \left(\sum_{1}^k x p_x \right)^2$$

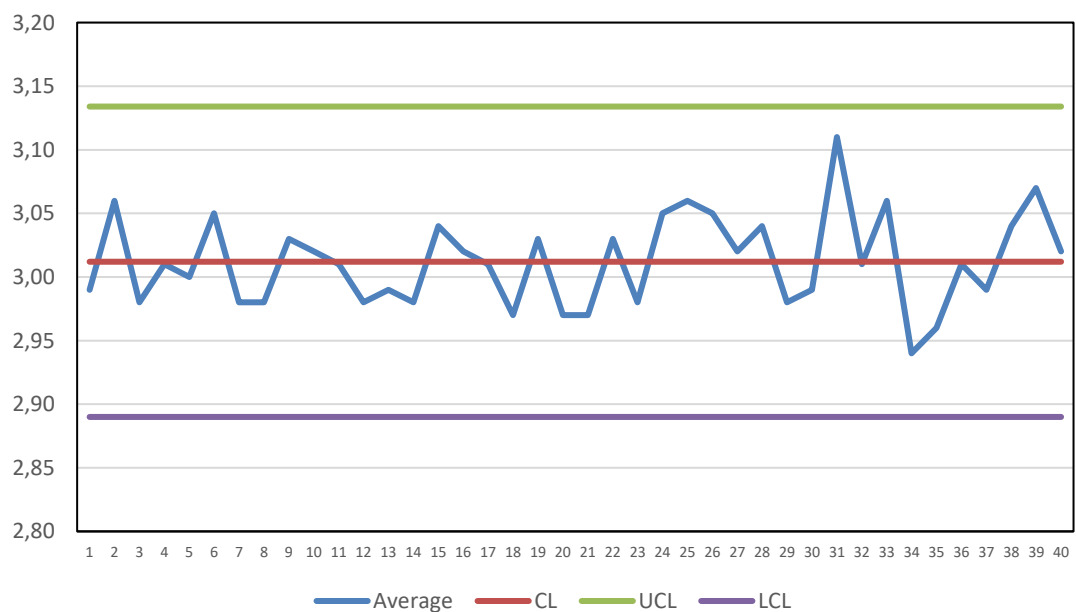
Στην συνέχεια υπολογίζω τα όρια ελέγχου σύμφωνα με τον παρακάτω τύπο και χρησιμοποιώντας ως μέση τιμή τον μέσο όρο όλων των δειγμάτων για σταθερά όρια ελέγχου.

$$\sum_{1}^k xp_x \pm \frac{3}{\sqrt{n}} \sqrt{\sum_{1}^k x^2 p_x - \left(\sum_{1}^k xp_x\right)^2}$$

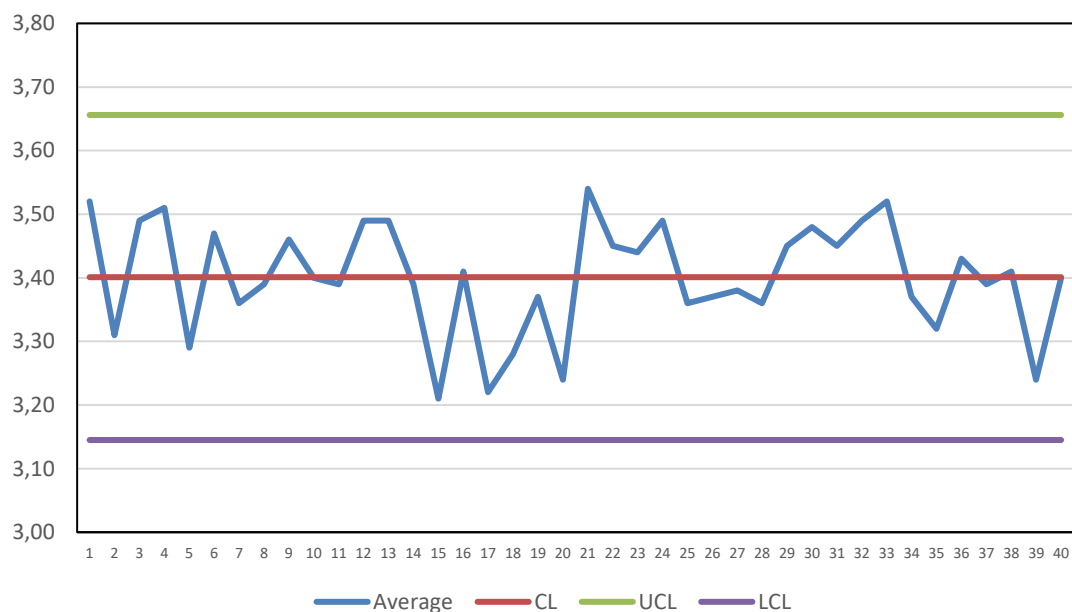
Την παραπάνω διαδικασία την εφαρμόζω για κάθε περίπτωση παραμέτρων και υπολογίζω τα όρια ελέγχου και τα διαγράμματα ελέγχου.



Σχήμα 4.7: Διάγραμμα ελέγχου \bar{X} για 1^ο σετ δεδομένων (α=15, β=4)



Σχήμα 4.8: Διάγραμμα ελέγχου \bar{X} για 2^ο σετ δεδομένων (α=β=15)



Σχήμα 4.9: Διάγραμμα ελέγχου \bar{X} για 3^ο σκετ δεδομένων ($\alpha=3, \beta=2$)

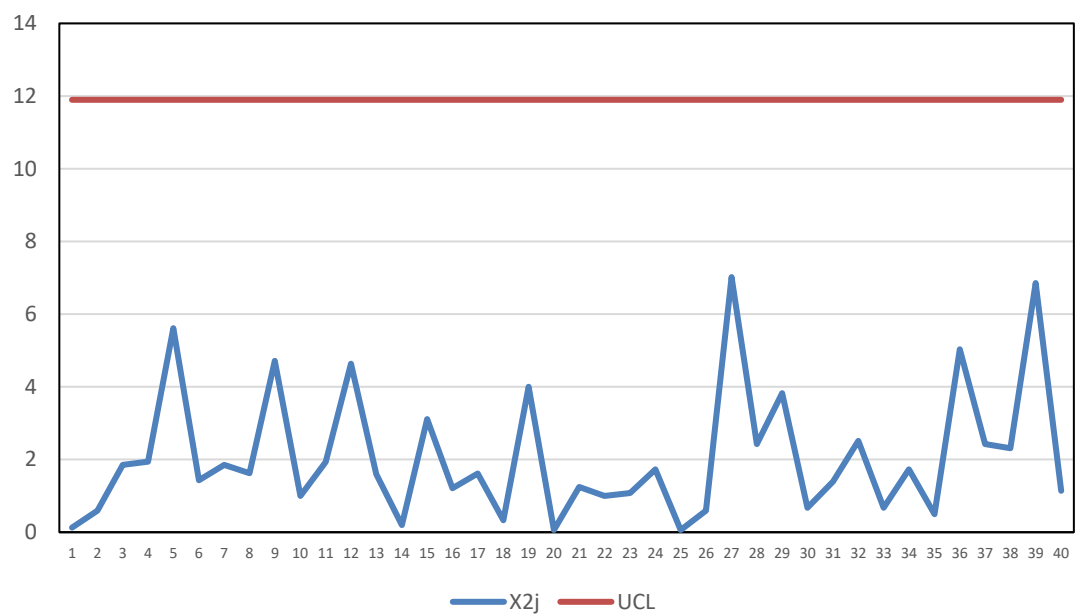
4.5 Διάγραμμα ελέγχου χ^2

Το διάγραμμα ελέγχου χ^2 χρησιμοποιεί μετρήσεις καλής προσαρμογής (χ^2 goodness of fit) ώστε να συγκρίνει την πραγματική τιμή της κατανομής με την θεωρητική τιμή. Έχουμε $k=5$ κατηγορίες και ορίζουμε ως p_x το ποσοστό των πελατών οι οποίοι δίνουν απάντηση x . Για τη δημιουργία του διαγράμματος ελέγχου συγκρίνουμε τον πραγματικό αριθμό πελατών που έδωσαν απάντηση x με τον θεωρητικό ή αναμενόμενο αριθμό πελατών που έδωσαν απάντηση x με την χρήση του τύπου:

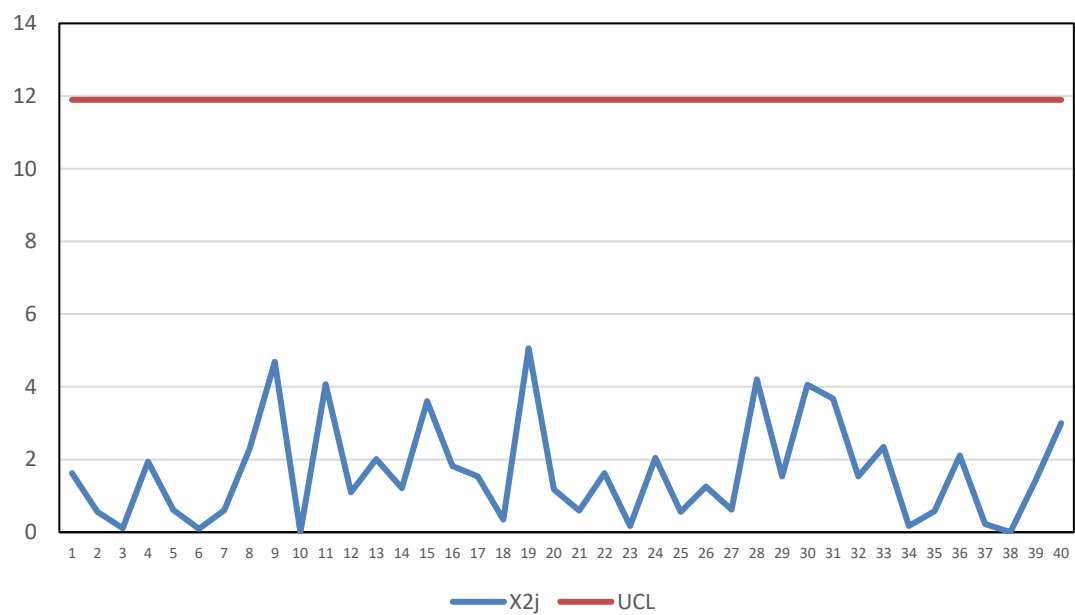
$$\chi_j^2 = \frac{\sum_{x=1}^k (Yxj - Exj)^2}{Exj}$$

Για τον υπολογισμό της θεωρητικής τιμής ή αναμενόμενης τιμής Exj υπολογίζουμε τον μέσο όρο των απαντήσεων όπου λαμβάνουμε κάθε εβδομάδα για κάθε κατηγορία x . Λόγω ότι η παραπάνω διαδικασία είναι σχεδόν χ^2 κατανεμημένη έχει $k-1$ βαθμούς ελευθερίας και υποθέτουμε ότι το Exj είναι τουλάχιστον 5. Στις περιπτώσεις όπου δεν ισχύει αυτή η υπόθεση, συνδυάζουμε τις κατηγορίες ώστε οι απαντήσεις να γίνουν τουλάχιστον 5 και έτσι οι βαθμοί ελευθερίας ελαττώνονται.

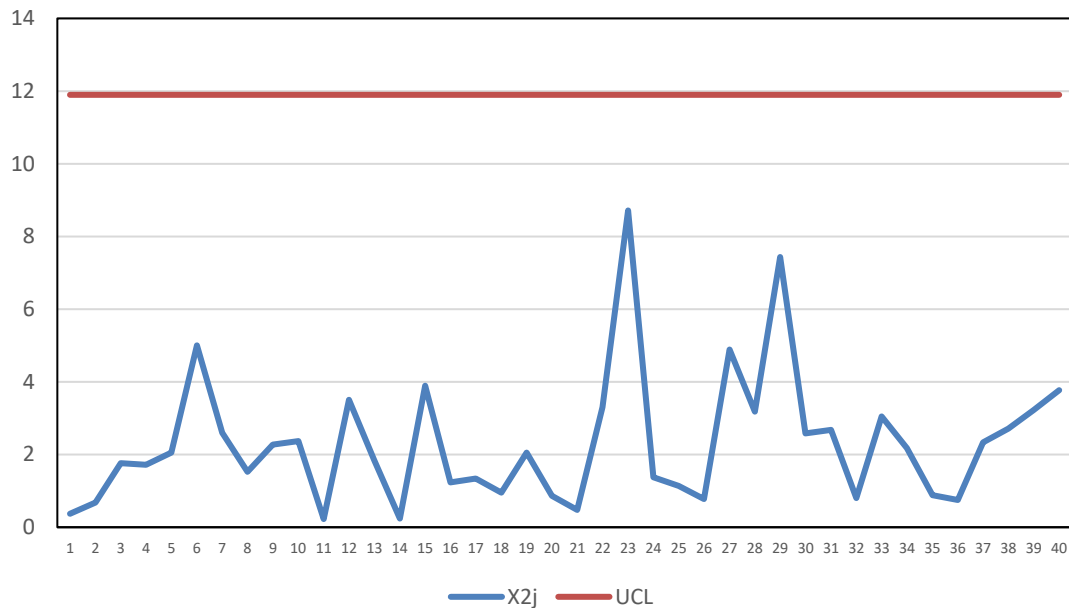
Για κάθε εβδομάδα ακολουθώ την παραπάνω διαδικασία και υπολογίζω το χ_j^2 . Το διάγραμμα ελέγχου χ^2 έχει μόνο άνω όριο ελέγχου (UCL) και ανάλογα την τιμή των κατηγοριών k ορίζουμε την τιμή του. Συγκεκριμένα για τα δεδομένα μας το k ισούται με 3 λόγω της συγχώνευσης των κατηγοριών άρα το UCL ισούται με 11,829, από τον Πίνακα 3.2.



Σχήμα 4.10: Διάγραμμα ελέγχου X^2 για 1^ο σετ δεδομένων ($\alpha=15, \beta=4$)



Σχήμα 4.11: Διάγραμμα ελέγχου X^2 για 2^ο σετ δεδομένων ($\alpha=\beta=15$)



Σχήμα 4.12: Διάγραμμα ελέγχου χ^2 για 3^ο σετ δεδομένων ($\alpha=3$, $\beta=2$)

4.6 Μετασχηματισμός δεδομένων

4.6.1 Απλοί μετασχηματισμοί

Ένας τρόπος αντιμετώπισης των μη κανονικών δεδομένων είναι ο μετασχηματισμός δεδομένων σε κανονικά με την χρήση κατάλληλων συναρτήσεων. Η διαδικασία θα πραγματοποιηθεί επίσης σε περιβάλλον Excel.

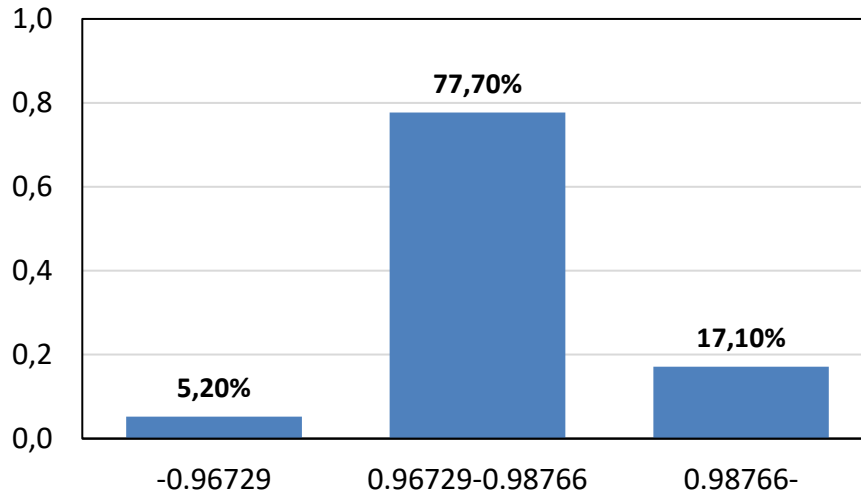
Συγκεκριμένα για την υλοποίηση της παραπάνω διαδικασίας θα εφαρμόσουμε τους παρακάτω τύπους μετασχηματισμού στα ήδη τυχαία δεδομένα που έχουμε δημιουργήσει με την χρήση της κατανομής Βήτα.

Περίπτωση 1^η

Για να δημιουργήσουμε την κατάλληλη κατανομή δεδομένων που ταιριάζει σε δεδομένα τα οποία προέρχονται από έρευνα ικανοποίησης και προσεγγίζουν την κανονική κατανομή θα χρησιμοποιήσουμε την αντίστροφη λογαριθμική .

$$\log \frac{x}{x-1} \rightarrow \frac{e^x}{1+e^x}$$

Τα δεδομένα που προέρχονται από το 1^ο σετ δεδομένων έπειτα από την χρήση του λογαριθμικού μετασχηματισμού έχουν κατανομή η οποία προσεγγίζει την κανονική και είναι γερμένη προς τα δεξιά.



Σχήμα 4.13: Κατανομή δεδομένων για το 1ο σετ δεδομένων (λογαριθμικός μετασχηματισμός)

Στην συνέχεια γίνεται εφαρμογή των κατάλληλων στατιστικών εργαλείων. Συγκεκριμένα θα εφαρμόσουμε τα διάγραμμα ελέγχου \bar{X} . Για την εφαρμογή του διαγράμματος ελέγχου θα χρησιμοποιηθούν οι τύποι :

$$UCL = \bar{x} + k \frac{\sigma}{n}$$

$$CL = \bar{x}$$

$$LCL = \bar{x} - k \frac{\sigma}{n}$$

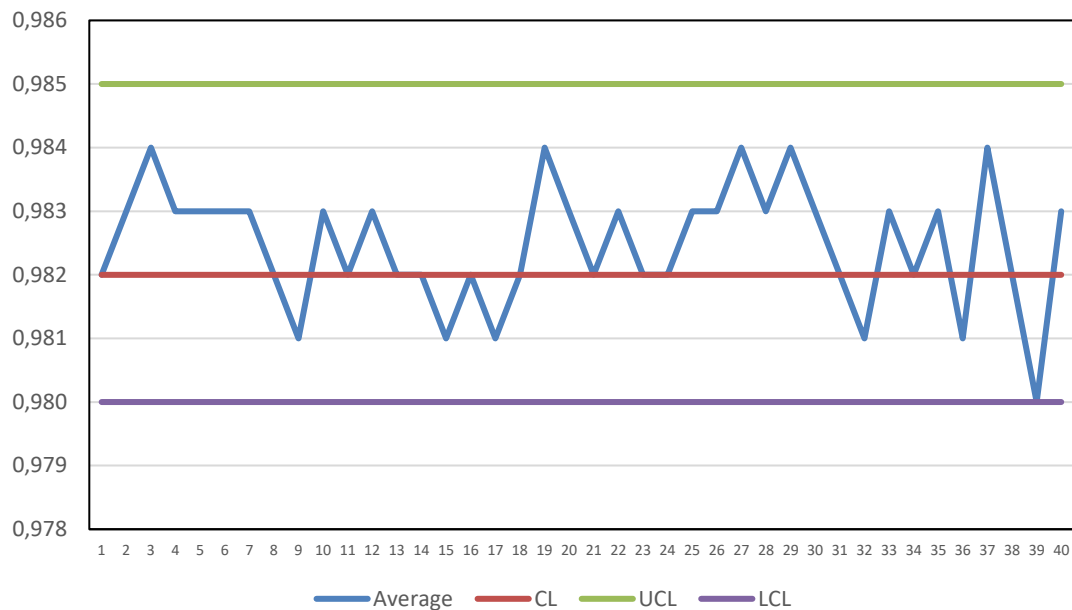
Όπου ως n είναι τα 100 δείγματα όπου λαμβάνουμε κάθε εβδομάδα και $k=3$. Η τιμή του σ υπολογίζεται από τον τύπο:

$$\sigma = \frac{\bar{s}}{c_4}$$

$$c_4 = \sqrt{\frac{2}{n-1} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}-\frac{1}{2})}}$$

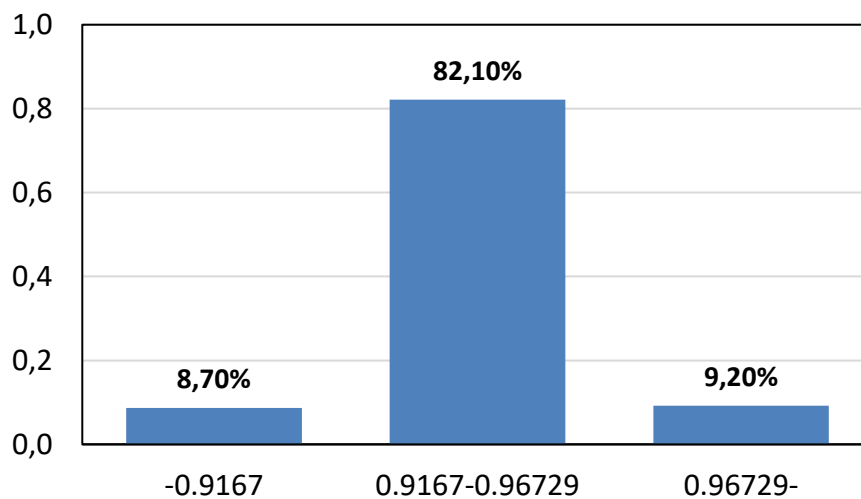
Η τιμή του μέσου όρου του εύρους \bar{x} υπολογίζεται χρησιμοποιώντας τον μέσο όρο των δειγμάτων που λαμβάνουμε.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$



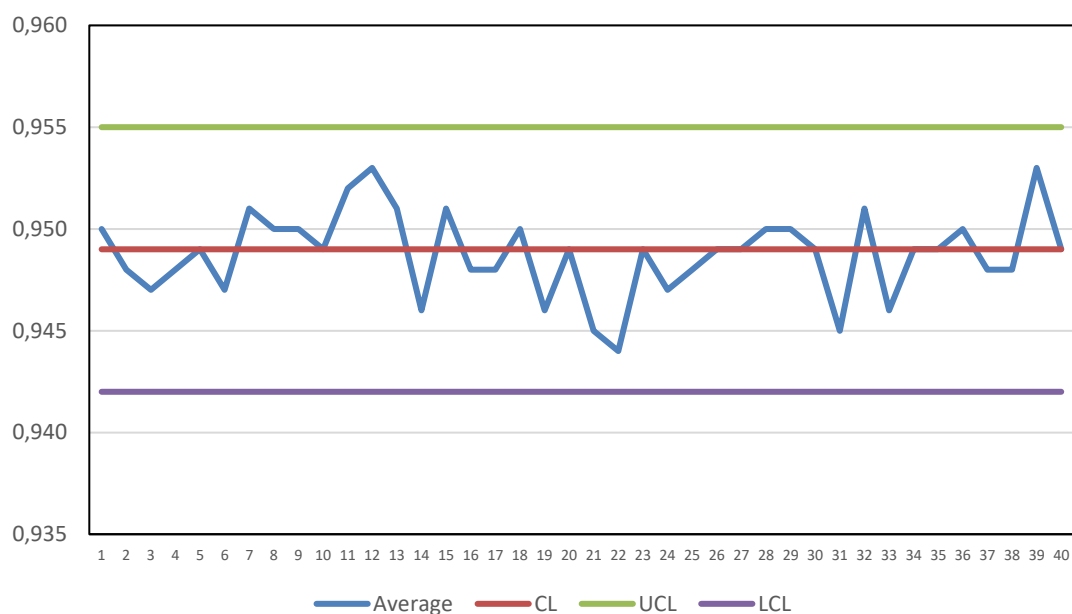
Σχήμα 4.14: Διάγραμμα ελέγχου X-bar για την εφαρμογή του λογαριθμικού μετασχηματισμού για το 1ο σετ δεδομένων

Τα δεδομένα που προέρχονται από το 2^ο σετ δεδομένων έπειτα από την χρήση του λογαριθμικού μετασχηματισμού έχουν κατανομή η οποία προσεγγίζει την κανονική αλλά δεν είναι γερμένη προς τα θετικά.



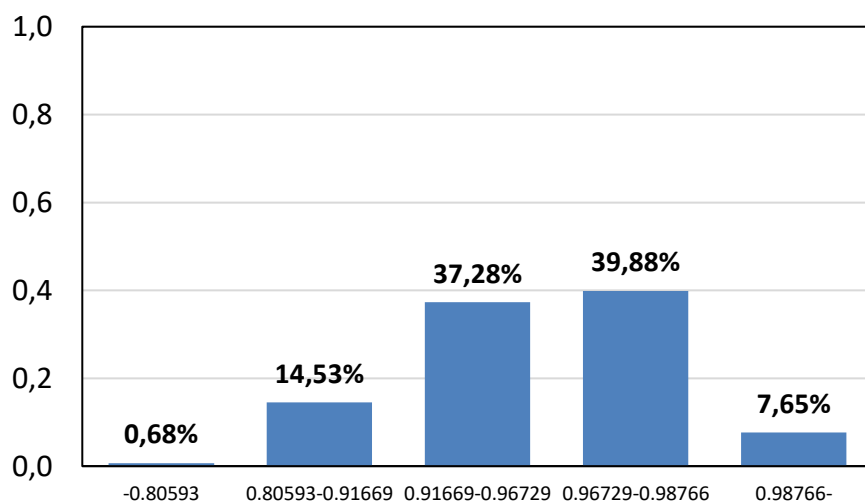
Σχήμα 4.15: Κατανομή δεδομένων για το 2ο σετ δεδομένων (λογαριθμικός μετασχηματισμός)

Στην συνέχεια γίνεται εφαρμογή των κατάλληλων στατιστικών εργαλείων. Συγκεκριμένα θα εφαρμόσουμε τα διάγραμμα ελέγχου X, με την χρήση των τύπων που εφαρμόστηκαν και παραπάνω .



Σχήμα 4.16: Διάγραμμα ελέγχου X-bar για την εφαρμογή του λογαριθμικού μετασχηματισμού για το 2ο σετ δεδομένων

Τα δεδομένα που προέρχονται από το 3^ο σετ δεδομένων έπειτα από την χρήση του λογαριθμικού μετασχηματισμού έχουν κατανομή η οποία δεν προσεγγίζει την κανονική κατανομή. Συνεπώς δεν μπορούν να εφαρμοστούν τα στατιστικά εργαλεία που χρησιμοποιήθηκαν παραπάνω.



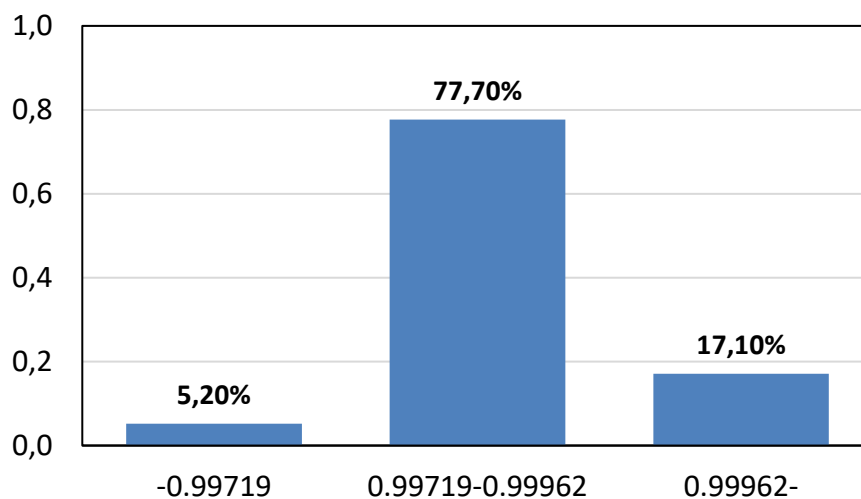
Σχήμα 4.17: Κατανομή δεδομένων για το 3ο σετ δεδομένων (λογαριθμικός μετασχηματισμός)

Περίπτωση 2^η

Ένας άλλος μετασχηματισμός δεδομένων για να δημιουργήσουμε την κατάλληλη κατανομή δεδομένων που ταιριάζει σε δεδομένα τα οποία προέρχονται από έρευνα ικανοποίησης παρουσιάζεται παρακάτω.

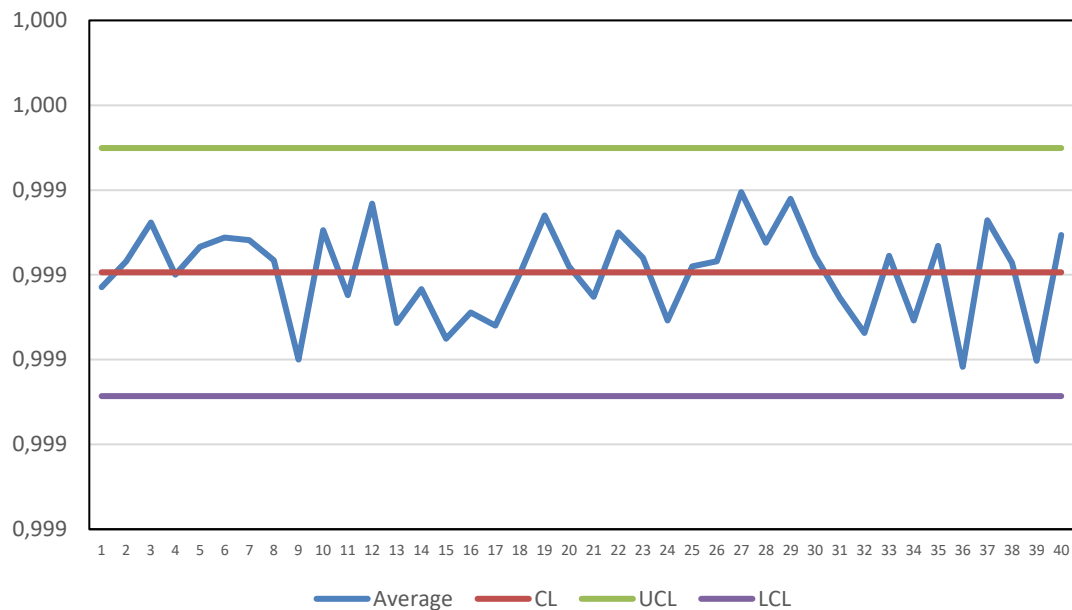
$$\frac{1}{2} \log \frac{1+x}{1-x} \rightarrow \frac{e^{2x} - 1}{e^{2x} + 1}$$

Τα δεδομένα που θα προκύψουν από τον παραπάνω μετασχηματισμό αντιστοιχούν στο 1^ο σετ δεδομένων και φαίνεται να έχουν κατανομή η οποία προσεγγίζει την κανονική.



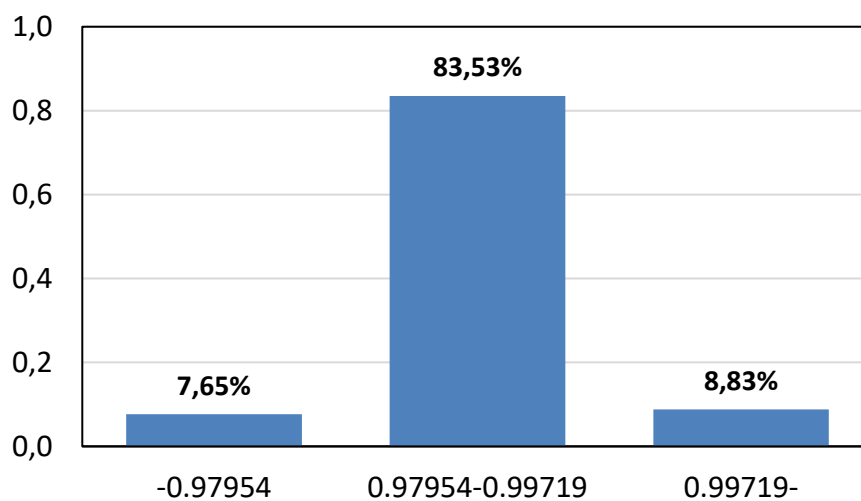
Σχήμα 4.18: Κατανομή δεδομένων για το 1ο σετ δεδομένων (εκθετικός μετασχηματισμός)

Το αντίστοιχο διάγραμμα ελέγχου Χ, με την χρήση των τύπων που εφαρμόστηκαν και στις προηγούμενες περιπτώσεις παρουσιάζεται στο Σχήμα 4.19.

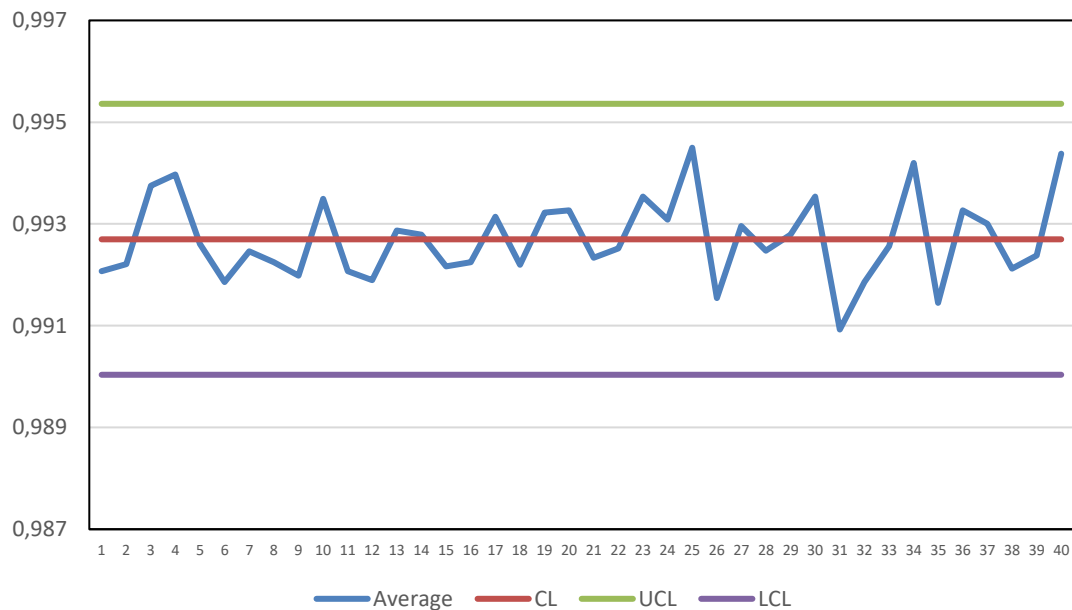


Σχήμα 4.19: Διάγραμμα ελέγχου X-bar για την εφαρμογή του εκθετικού μετασχηματισμού για το 1ο σετ δεδομένων

Τα δεδομένα που προέρχονται από το 2^ο σετ δεδομένων επίσης φαίνεται να προσεγγίζουν την κανονική κατανομή όπως φαίνεται από το παρακάτω γράφημα.

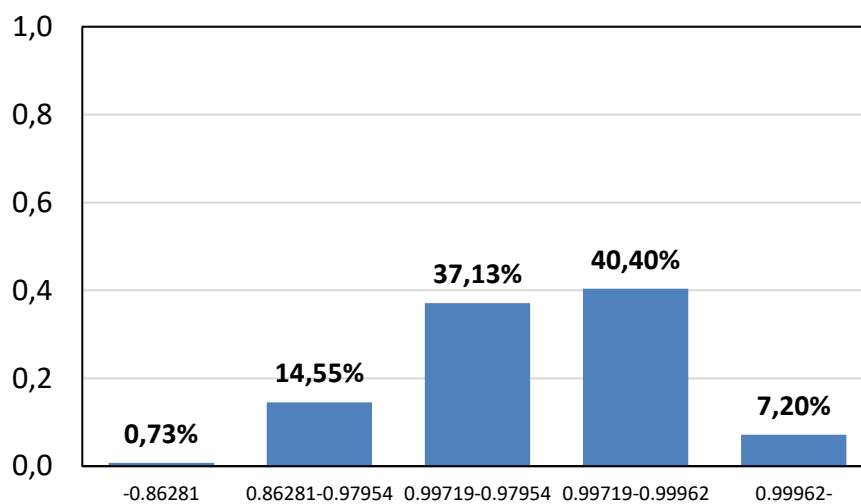


Σχήμα 4.20: Κατανομή δεδομένων για το 2ο σετ δεδομένων (εκθετικός μετασχηματισμός)



Σχήμα 4.21: Διάγραμμα ελέγχου X-bar για την εφαρμογή του εκθετικού μετασχηματισμού για το 2ο σετ δεδομένων

Τα δεδομένα που προέρχονται από το 3^ο σετ δεδομένων δεν φαίνεται να προσεγγίζουν την κανονική κατανομή όπως φαίνεται από το παρακάτω γράφημα



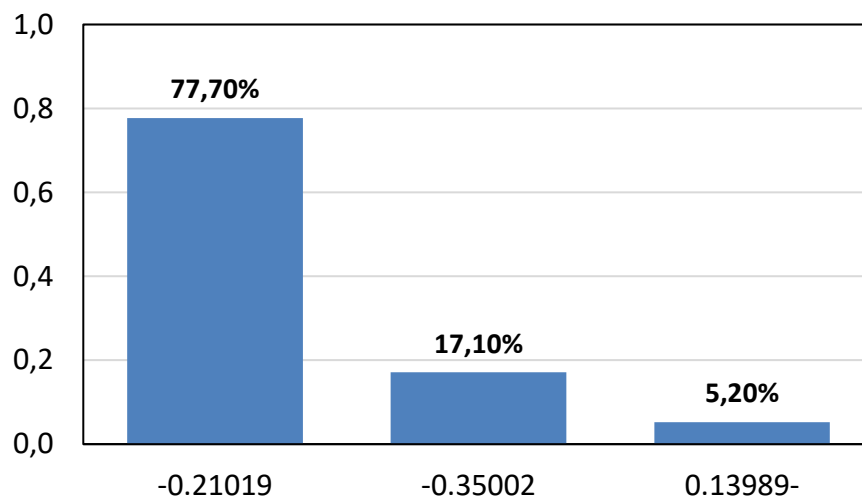
Σχήμα 4.22: Κατανομή δεδομένων για το 3ο σετ δεδομένων (εκθετικός μετασχηματισμός)

Περίπτωση 3^η

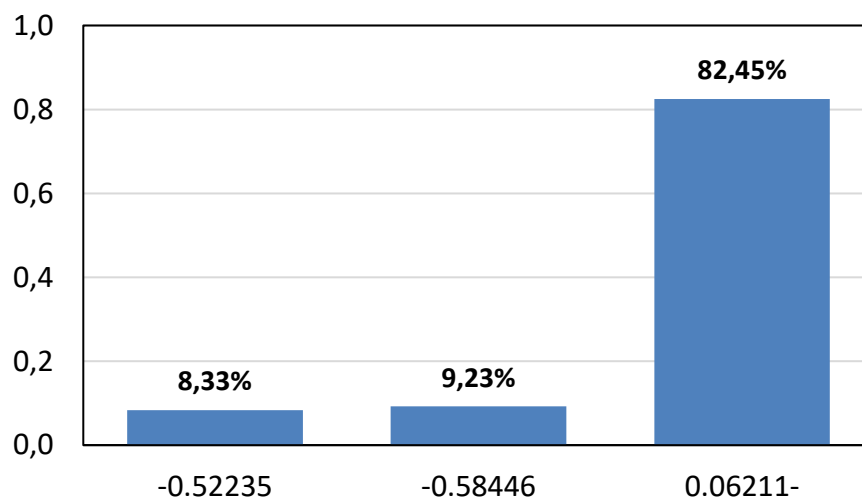
Ένας άλλος μετασχηματισμός δεδομένων για να δημιουργήσουμε την κατάλληλη κατανομή δεδομένων που ταιριάζει σε δεδομένα τα οποία προέρχονται από έρευνα ικανοποίησης παρουσιάζεται παρακάτω.

$$\sin(x^2)$$

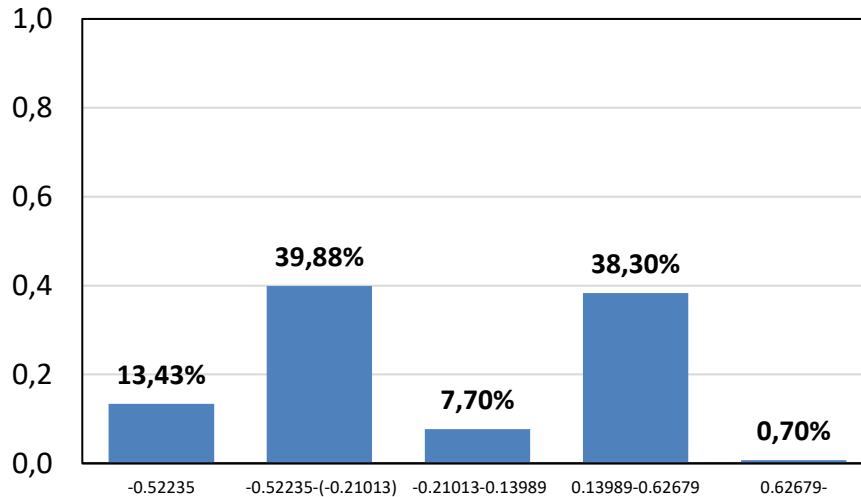
Τα δεδομένα από όλα τα σετ δεδομένων δεν φαίνεται να προσεγγίζουν την κανονική κατανομή γι' αυτό ο συγκεκριμένος μετασχηματισμός δεν θα χρησιμοποιηθεί.



Σχήμα 4.23: Κατανομή δεδομένων για το 1ο σετ δεδομένων (ημιτονοειδής μετασχηματισμός)



Σχήμα 4.24: Κατανομή δεδομένων για το 2ο σετ δεδομένων (ημιτονοειδής μετασχηματισμός)



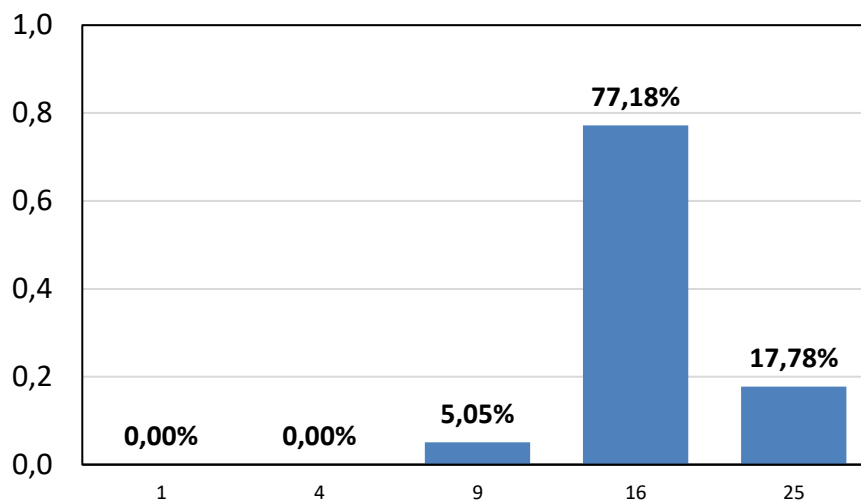
Σχήμα 4.25: Κατανομή δεδομένων για το 3ο σετ δεδομένων (ημιτονοειδής μετασχηματισμός)

4.6.2 Μετασχηματισμός Box-Cox

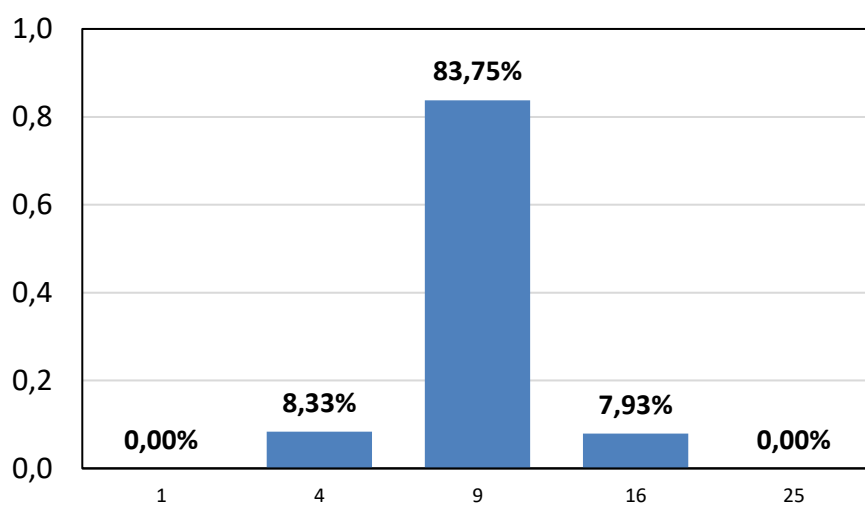
Ο Box-Cox είναι ένας μετασχηματισμός ο οποίος μετατρέπει μη κανονικά δεδομένα σε κανονικά. Ο μετασχηματισμός αυτός έχει σκοπό τον εντοπισμό ενός κατάλληλου λ το οποίο θα χρησιμοποιηθεί για την μετατροπή των δεδομένων σε κανονικό σχήμα. Το λ αυτό λαμβάνει τιμές από -5 έως 5, μέχρι να βρεθεί η καλύτερη τιμή που προσεγγίζει την κανονική κατανομή όπου εμείς επιθυμούμε .

Για τον μετασχηματισμό των δεδομένων θα χρησιμοποιηθεί ο Πίνακας 3.4 που παρουσιάστηκε σε προηγούμενο κεφάλαιο.

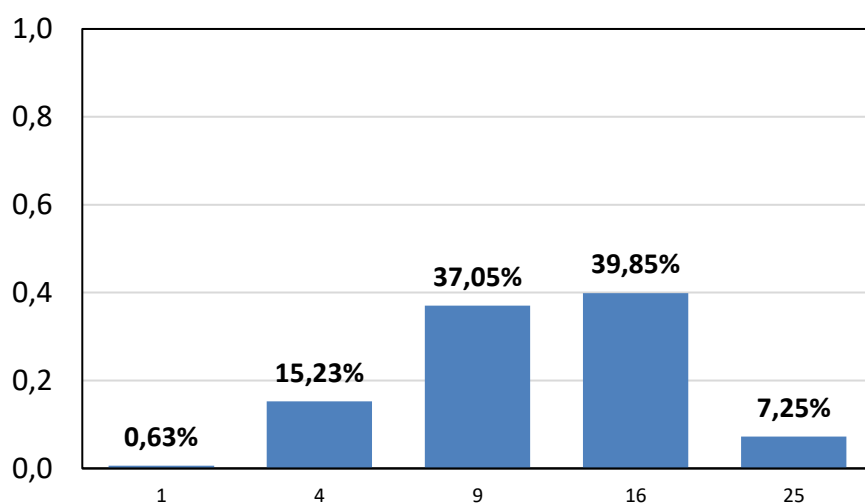
Για $\lambda=2$ παρουσιάζεται η απεικόνιση των δεδομένων για κάθε σετ δεδομένων.



Σχήμα 4.26: Κατανομή δεδομένων για το 1ο σετ δεδομένων (Box-Cox για $\lambda=2$)

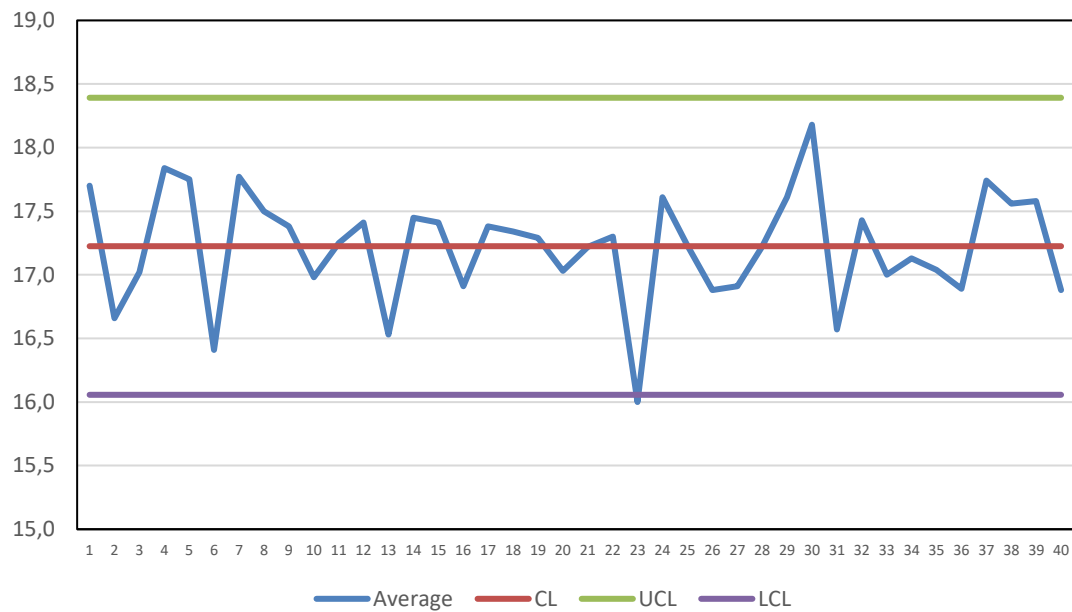


Σχήμα 4.27: Κατανομή δεδομένων για το 2ο σετ δεδομένων (Box-Cox για $\lambda=2$)

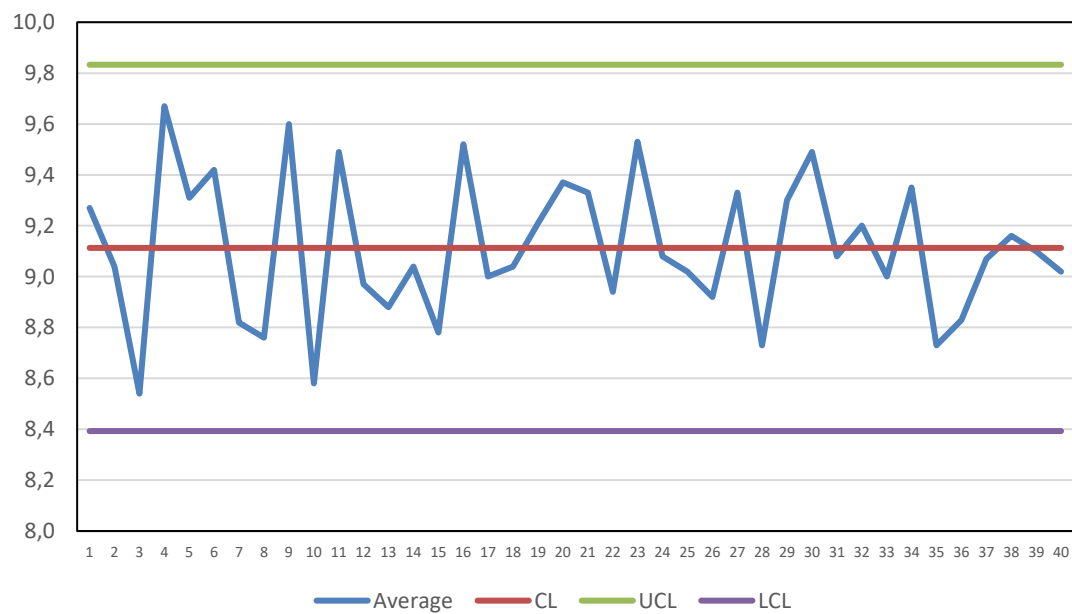


Σχήμα 4.28: Κατανομή δεδομένων για το 3ο σετ δεδομένων (Box-Cox για $\lambda=2$)

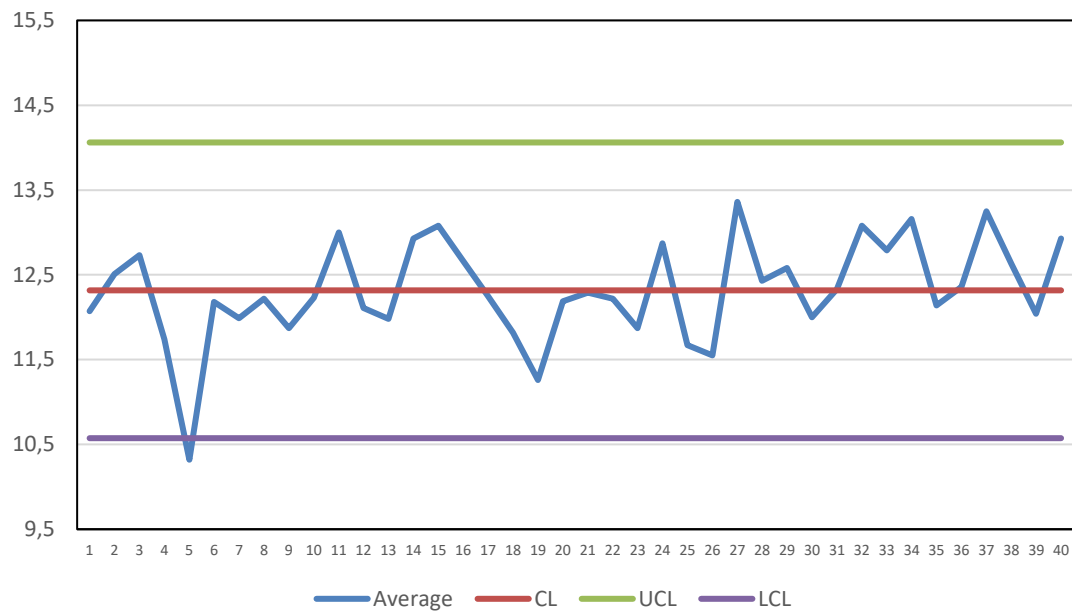
Στην συνέχεια γίνεται εφαρμογή των στατιστικών εργαλείων εφόσον τα δεδομένα μας προσεγγίζουν την κανονική κατανομή γερμένη προς τα δεξιά. Συγκεκριμένα θα γίνει εφαρμογή του διαγράμματος X-bar.



Σχήμα 4.29: Διάγραμμα ελέγχου X-bar για την εφαρμογή του μετασχηματισμού Box-Cox για το 1ο σετ δεδομένων



Σχήμα 4.30: Διάγραμμα ελέγχου X-bar για την εφαρμογή του μετασχηματισμού Box-Cox για το 2ο σετ δεδομένων



Σχήμα 4.31: Διάγραμμα ελέγχου X-bar για την εφαρμογή του μετασχηματισμού Box-Cox για το 3ο σετ δεδομένων

Κεφάλαιο 5ο: Συμπεράσματα

5.1 Σύνοψη αποτελεσμάτων

Στην προηγούμενη ενότητα παρουσιάστηκαν τα εναλλακτικά διαγράμματα ελέγχου για δεδομένα ικανοποίησης πελατών, τα οποία δεν ακολουθούν την κανονική κατανομή αλλά μια κατανομή η οποία είναι γερμένη προς τα δεξιά. Αρχικά συγκρίνοντας τα διαγράμματα με μια πρώτη ματιά παρατηρούμε πως όλα τα διαγράμματα ελέγχου φαίνονται να βρίσκονται εντός ορίων ελέγχου για όλα τα σετ δεδομένων που δημιουργήθηκαν. Σε κάποια διαγράμματα παρατηρούνται κάποια σημεία εκτός ελέγχου. Στην περίπτωση όπου η διαδικασία βρίσκεται εκτός στατιστικού ελέγχου, δείχνει πως η διαδικασία επηρεάζεται από συστηματικές αιτίες (ή ειδικές αιτίες). Συστηματικές αιτίες μπορεί να θεωρούνται είτε μηχανικές ή ηλεκτρονικές βλάβες ή αλλαγές στην ρύθμιση ή ανθρώπινα σφάλματα, γενικά αιτίες που προέρχονται από εξωτερικούς παράγοντες. Σημαντικό για τις συστηματικές αιτίες είναι ότι μπορούν να απαλειφθούν μετά από κατάλληλη διορθωτική επέμβαση σε αντίθεση με τις τυχαίες αιτίες. Οι συστηματικές αιτίες έχουν μεγαλύτερη επίδραση στην διασπορά των τιμών σε σχέση με τις τυχαίες αιτίες και κατά κανόνα οδηγεί σε μη αποδεκτή στάθμη ποιότητας.

Ένας απλός κανόνας ένδειξης εκτός ελέγχου είναι όταν η διαδικασία είναι εκτός ορίων ελέγχου δηλαδή μόνο όταν οι συστηματικές αιτίες προκαλούν σχετικά μεγάλες μεταβολές στην διαδικασία. Σε περίπτωση όπου η μεταβολή είναι μικρή, ανάλογα μικρή θα είναι και η ισχύς του διαγράμματος ελέγχου, με αποτέλεσμα η καθυστέρηση στον εντοπισμό της συστηματικής αιτίας. Επίσης ορισμένες συστηματικές αιτίες δεν επιφέρουν μόνιμες αλλαγές αλλά μπορεί να εμφανίσουν σταδιακές μεταβολές στην διαδικασία. Γι' αυτό σημαντικό είναι να παρατηρηθούν τα διαγράμματα ελέγχου τόσο και ως προς την μορφή των χρονοσειρών. Γενικά οι χρονοσειρές χωρίζονται σε δύο κατηγορίες ανάλογα με τον τύπο της μεταβολής που δημιουργεί.

- Χρονοσειρές που οφείλονται σε στιγμιαία μεταβολή της τιμής της παραμέτρου, η οποία στην συνέχεια διατηρείται σταθερή.
- Χρονοσειρές που οφείλονται σε συνεχή μεταβολή μιας παραμέτρου .

Ξεκινώντας για το διάγραμμα ελέγχου \bar{x} παρατηρούμε πως αν μια τιμή βρίσκεται εκτός ορίων ελέγχου ή παρατηρείται σειρά πολλών μέσων τιμών από την μία μόνο πλευρά της κεντρικής γραμμής τότε υπάρχει περίπτωση να έχει υπάρξει μεταβολή της μέσης τιμής του \bar{x} σε σχέση με την ονομαστική τιμή μ_0 , πιθανό να οφείλεται σε μια συστηματική αιτία. Σημαντικό ρόλο παίζει η κατασκευή των ορίων ελέγχου για το διάγραμμα ελέγχου \bar{x} . Στα όρια ελέγχου παρουσιάζεται η τυπική απόκλιση σ και πρέπει να γίνει αναφορά στις δύο περιπτώσεις της. Η πρώτη περίπτωση είναι όταν η τυπική απόκλιση μπορεί να θεωρηθεί ως δεδομένη η ονομαστική τιμή δηλαδή σ_0 και

η δεύτερη περίπτωση είναι όταν έχουμε μόνο μια εκτιμήτρια της τυπικής απόκλισης σ . Σε αυτήν την περίπτωση στις σχέσεις των ορίων ελέγχου χρησιμοποιείται η εκτιμήτρια της σ δηλαδή $\frac{\bar{s}}{c_4}$, όπως και στην περίπτωση μας.

Χρησιμοποιώντας το γράφημα \bar{X} αλλάζουν τα όρια ελέγχου. Τα όρια ελέγχου υπολογίζονται χρησιμοποιώντας την πραγματική απόκλιση (variance) των δεδομένων ικανοποίησης. Έτσι χρησιμοποιώντας πιο κατάλληλα γραφήματα με πραγματική απόκλιση μόνο τυχαίες αιτίες μπορούν να επηρεάσουν τα δεδομένα ικανοποίησης και να οδηγήσουν σε εκτός ορίου δεδομένα.

Ξεκινώντας από το 1^ο σετ δεδομένων (δηλαδή για $\alpha=15$ και $\beta=4$) παρατηρούμε πώς όλα τα διαγράμματα που χρησιμοποιήθηκαν δηλαδή το διάγραμμα ελέγχου \bar{x} , το διάγραμμα ελέγχου p αλλά και το X^2 μπορούν να εντοπίσουν τις μεταβολές που υπάρχουν στο σύστημα. Συγκεκριμένα φαίνεται πως και τα 3 διαφορετικά ως προς τον τρόπο κατασκευής διαγράμματα παρουσιάζουν μεταβλητότητα. Όσο αναφορά το διάγραμμα ελέγχου p παρατηρούμε ότι μοιάζει με το διάγραμμα X -bar. Ωστόσο για τον υπολογισμό των ορίων ελέγχου του p διαγράμματος η διασπορά που υπολογίζεται δε βασίζεται στην κανονική κατανομή αλλά στην κατανομή που βρίσκεται κάτω από το άθροισμα των x_s , όπου ως x_s ορίζουμε το σύνολο των απαντήσεων x . Το γράφημα αυτό βασίζεται στο μεγάλο μέγεθος δείγματος έτσι ώστε η κατανομή των \bar{x} να είναι περίπου κανονική. Τα όρια ελέγχου υπολογίζονται χρησιμοποιώντας την πραγματική διακύμανση των δεδομένων ικανοποίησης του πελάτη και χρησιμοποιούμε ± 3 τυπικές αποκλίσεις. Αντίθετα το διάγραμμα ελέγχου X^2 παρουσιάζει κάποια άλματα στις τιμές του. Αυτό οφείλεται στο γεγονός πως για τον υπολογισμό των ορίων ελέγχου του X^2 δεν βασίζεται στην υπόθεση ότι η κατανομή των δεδομένων είναι κανονική και επίσης για δεδομένο αριθμό κατηγοριών το άνω όριο ελέγχου δεν αλλάζει αλλά παραμένει σταθερό.

Στο 2^ο σετ δεδομένων ($\alpha=\beta=15$) ξεκινώντας από το διάγραμμα ελέγχου \bar{x} παρατηρούμε πως και τα 3 διαφορετικά ως προς τον τρόπο κατασκευής διαγράμματα παρουσιάζουν μεταβλητότητα και κανένα σημείο δεν φαίνεται να είναι εκτός ορίων. Στην περίπτωση του X^2 διαγράμματος ελέγχου δεν υπάρχει κάποια άνοδο στις τιμές ικανοποίησης. Υπάρχουν κάποιες τιμές οι οποίες φαίνεται να είναι κοντά στα όρια, ωστόσο κάτι τέτοιο δεν φαίνεται στο X^2 το οποίο σημαίνει πως δεν υπάρχει κάποια ειδική αιτία να μεταβάλλει την κατανομή των δεδομένων αλλά πρόκειται για κάποια τυχαία.

Στο 3^ο σετ δεδομένων ($\alpha=3$, $\beta=2$) το διάγραμμα ελέγχου \bar{x} δεν φαίνεται να έχει εντοπίσει κάποια μεταβολή των δεδομένων εκτός ορίων ελέγχου. Στο σετ αυτό για το διάγραμμα p μεταξύ των εβδομάδων 14-19 παρατηρούμε μια τάση των δεδομένων να βρίσκονται κάτω από την κεντρική γραμμή. Αντίστοιχα στο X^2 δεν παρατηρούμε κάποια απόκλιση αλλά κάποια άλματα στα δεδομένα.

Πέρα από την ανάλυση της γραφικής αναπαράστασης των σημείων στην περίοδο των 40 εβδομάδων, ένας άλλος τρόπος μέτρησης των διαγραμμάτων ελέγχου είναι ο υπολογισμός του ARL δηλαδή του μέσου αριθμού δειγμάτων από την στιγμή της μεταβολής στην διαδικασία μέχρι το πρώτο σημείο που θα βρεθεί εκτός ορίων ελέγχου. Η χρήση του ARL(μέσο μήκος ροής) προκύπτει και για να εντοπίσουμε πιο γρήγορα την αλλαγή στα δεδομένα μας αλλά και πιο διάγραμμα είναι πιο ακριβές για δεδομένα που προέρχονται από δεδομένα έρευνας ικανοποίησης. Τα δεδομένα που λαμβάνουμε ,υποθέτουν πως το μέγεθος του δείγματος είναι σταθερό, γι' αυτό για τον υπολογισμό του ARL μπορούμε να εφαρμόσουμε και την κανονική κατανομή.

Σε κατάσταση στατιστικού ελέγχου έχουμε:

$$ARL0 = \frac{1}{\alpha}$$

Το α είναι το σφάλμα πρώτου είδους το οποίο ισούται με την πιθανότητα το δείγμα να είναι εκτός ορίων ενώ διαδικασία να είναι υπό στατιστικό έλεγχο.

Για το διάγραμμα ελέγχου \bar{x} η πιθανότητα σφάλματος πρώτου είδους εξαρτάται από τα όρια ελέγχου και την κατανομή της στατιστικής του δείγματος \bar{x} .

Η μέση τιμή της \bar{x} είναι μ και η τυπική απόκλιση της \bar{x} είναι $\frac{\sigma}{\sqrt{n}}$, όπου μ και σ είναι η μέση τιμή και η τυπική απόκλιση αντίστοιχα του δείγματος. Όταν η κατανομή του δείγματος είναι κανονική τότε και η κατανομή της μέσης τιμής του δείγματος είναι επίσης κανονική και τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή $N(0,1)$. Επίσης σύμφωνα με το κεντρικό οριακό θεώρημα η κατανομή της Z προσεγγίζεται από την κανονική κατανομή $N(0,1)$ και σε περιπτώσεις όπου η κατανομή του X δεν θεωρείται κανονική. Τέτοιες περιπτώσεις είναι όταν το μέγεθος δείγματος είναι πολύ μεγάλο είτε η κατανομή του X προσεγγίζει αρκετά καλά την κανονική.

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z$$

Συγκεκριμένα ο υπολογισμός του σφάλματος πρώτου είδους γίνεται ως εξής:

$$\alpha = P(\bar{x} > \mu_0 + k \frac{\sigma}{\sqrt{n}} | \mu = \mu_0) + P(\bar{x} < \mu_0 - k \frac{\sigma}{\sqrt{n}} | \mu = \mu_0) = P\left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > k\right) + P\left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -k\right) = P(Z > k) + P(Z < -k) = 2\Phi(-k)$$

Όπου $\Phi(z)$ η συνάρτηση αθροιστικής πιθανότητας της μεταβλητής Z .

Για το διάγραμμα ελέγχου \bar{X} και X_p έχουμε $\alpha=0,0027$, συνεπώς

$$ARL0 = \frac{1}{\alpha} = 370,37$$

για όλα τα σετ δεδομένων.

Συγκεκριμένα, η τιμή 370 σημαίνει πως απαιτούνται τόσα δείγματα ώστε να εντοπιστεί η μεταβολή στην παραγωγική διαδικασία.

Και στα 3 σετ δεδομένων δεν παρατηρείται μεταβολή ως προς τον υπολογισμό του ARL καθώς εξαρτάται μόνο από το k το οποίο είναι σταθερό.

Για το διάγραμμα X^2 :

$$\alpha = P(x > UCL) = P(x > 11,829) = 1 - P(x < 11,829) = 1 - P(x^2 < 11,829) = 1 - 0,994 = 0,006.$$

$$ARL_0 = \frac{1}{\alpha} = \frac{1}{0,006} = 166,66.$$

Για τον υπολογισμό του σφάλματος πρώτου είδους χρησιμοποιήθηκε η κατανομή του X^2 , όπου ακολουθούν τα δεδομένα σε αυτήν την περίπτωση. Η τιμή του α προέκυψε σύμφωνα με τον πίνακα της κατανομής όπως φαίνεται και στο Παράρτημα Β μετά από γραμμική παρεμβολή. Η τιμή του ARL_0 φαίνεται πως δεν επηρεάζεται από τις διαφορετικές περιπτώσεις των παραμέτρων α και β , δηλαδή από τα διαφορετικά σετ δεδομένων.

Συγκρίνοντας επίσης τα παραπάνω διαγράμματα ελέγχου με τα διαγράμματα ελέγχου που προέκυψαν από τον απλό μετασχηματισμό των δεδομένων αλλά και από τον μετασχηματισμό Box-Cox παρατηρούμε πόσο καλά προσεγγίζουν την κανονική κατανομή. Τα δεδομένα με τον μετασχηματισμό προσομοιάζουν αρκετά καλά την μορφή των δεδομένων που λαμβάνουμε από μία έρευνα ικανοποίησης, δηλαδή δεδομένα τα οποία προσεγγίζουν την κανονική κατανομή αλλά λοξά κατανεμημένα. Επίσης παρατηρούμε πως με την χρήση του μετασχηματισμού Box-Cox παρόλο που η κατανομή δεδομένων είναι ικανοποιητική, τα διαγράμματα που δημιουργήθηκαν δεν αντιστοιχούν σε διαγράμματα που βρίσκονται σε στατιστικό έλεγχο. Γεγονός το οποίο παρατηρείται από τα μοτίβα σημείων αλλά και από τον μεγάλο αριθμό σημείων κοντά στα όρια ελέγχου. Χρησιμοποιώντας ως εργαλείο το ARL μπορούμε να δούμε ποιο διάγραμμα ελέγχου εντόπισε πιο γρήγορα την μεταβολή στο διάγραμμα.

Για το διάγραμμα ελέγχου X έχουμε $\alpha = 0,0027$, οπότε όμοια με πριν:

$$ARL_0 = \frac{1}{\alpha} = 370,37$$

Η τιμή του ARL_0 είναι σταθερή και ίδια για όλα τα σετ δεδομένων καθώς η τιμή του σφάλματος πρώτου είδους εξαρτάται μόνο από το k το οποίο μένει σταθερό. Επίσης ο υπολογισμός του ARL_0 παραμένει ίδιος και για απλό μετασχηματισμό αλλά και για τον Box-Cox μετασχηματισμό.

Η επιλογή ενός κατάλληλου διαγράμματος ελέγχου είναι σημαντική καθώς θα βοηθήσει στην μείωση του σφάλματος πρώτου και δεύτερου είδους. Επίσης μέσα από την κατάλληλη επιλογή ενός διαγράμματος ελέγχου μπορούμε να παρατηρήσουμε και πόσο γρήγορα θα εντοπιστεί μια αιτία που θα προκαλέσει

μεταβολή. Η χρήση των απλών διαγραμμάτων Shewhart δεν μπορεί να εντοπίσει την μεταβλητότητα στο δείγμα που έχουμε λάβει και γενικά σε μεγάλα δείγματα δεδομένων. Γι' αυτό επιλέγουμε να χρησιμοποιήσουμε εναλλακτικά διαγράμματα ελέγχου, μεταβάλλοντας και τις παραμέτρους των δεδομένων ώστε να έχουμε καλύτερα αποτελέσματα για την αποτελεσματικότητά τους.

Το χ^2 παρατηρούμε πως μπορεί να χειριστεί πιο εύκολα δεδομένα τα οποία έχουν μεταβαλλόμενο μέγεθος υποομάδων. Ωστόσο το χ^2 είναι δύσκολο στον υπολογισμό λόγω των στατιστικών εργαλείων που χρησιμοποιεί. Παρατηρούμε επίσης πως το διάγραμμα ελέγχου χ^2 έχει μικρότερη τιμή ARL0 σε σχέση με τα υπόλοιπα διαγράμματα ελέγχου. Η τιμή του ARL0 δείχνει πόσες τιμές θέλουμε να λάβουμε ώστε να εντοπιστεί μεταβολή στην διεργασία. Έτσι βλέπουμε πως το διάγραμμα αυτό θέλει λιγότερα δείγματα σε σχέση με τα υπόλοιπα διαγράμματα ελέγχου. Σε περιπτώσεις εντός ελέγχου διεργασία προτιμάμε να έχουμε μεγάλο ARL0 καθώς δεν θέλουμε να διακόπτουμε μια παραγωγική διαδικασία η οποία λειτουργεί ομαλά, αντίθετα σε μια εκτός ελέγχου διεργασία θέλουμε να έχουμε όσο γίνεται μικρότερο ARL_Δ ώστε να εντοπίζεται πιο γρήγορα η μετατόπιση της διεργασίας. Γενικά μια κατάσταση στατιστικού ελέγχου σημαίνει ότι η κατανομή του χαρακτηριστικού ποιότητας παραμένει σταθερή και με την μικρότερη φυσική μεταβλητότητα. Ωστόσο αυτό δεν σημαίνει ότι αυτό το επίπεδο ποιότητας είναι το αποδεκτό. Ο στατιστικός έλεγχος ορίζεται με βάση την μεταβλητότητα της παραγωγικής διαδικασίας ανεξάρτητα από εξωτερικούς παράγοντες (ειδικές αιτίες). Μόνο με μικρή επίδραση των τυχαίων αιτιών μπορούμε να θεωρήσουμε ότι η έρευνα είναι ικανοποιητική.

Συνοψίζοντας όσο αναφορά από τον υπολογισμό των ARL, το χ^2 φαίνεται να μην έχει τα καλύτερα αποτελέσματα σε σχέση με τα υπόλοιπα διαγράμματα. Καλύτερα αποτελέσματα φαίνεται να έχει το διάγραμμα ελέγχου χ_p , καθώς παρουσιάζει καλύτερο ARL0. Τέλος φαίνεται πως και η χρήση μετασχηματισμένων δεδομένων τόσο με την χρήση λογαριθμικών συναρτήσεων αλλά και η χρήση του μετασχηματισμού Box-Cox προσεγγίζουν αρκετά ικανοποιητικά την κανονική κατανομή για όλα τα σετ δεδομένων που δημιουργήθηκαν. Ο μετασχηματισμός Box-Cox δεν αποτελεί εγγύηση για την κανονικότητα των δεδομένων. Ο μετασχηματισμός προσπαθεί να βρει τη βέλτιστη επιλογή του λ , ώστε τα δεδομένα να προσεγγίζουν την κανονική κατανομή, αλλά για καλύτερα αποτελέσματα απαιτείται ένα γράφημα πιθανοτήτων.

5.2 Μελλοντικές επεκτάσεις

Η διπλωματική αυτή εργασία βασίστηκε σε άρθρα που ήδη υπάρχουν σχετικά με την ανάλυση δεδομένων ικανοποίησης και έχει ως σκοπό να μελετήσει διάφορες περιπτώσεις διαγραμμάτων ελέγχου κατά πόσο μπορούν να εφαρμοστούν σε δεδομένα ικανοποίησης πελατών. Η κατάλληλη επιλογή διαγράμματος ελέγχου είναι σημαντική για την μείωση σφαλμάτων πρώτου είδους και δευτέρου είδους. Σε περιπτώσεις όπου τα δεδομένα είναι κανονικά, τα διαγράμματα ελέγχου Shewhart είναι ικανοποιητικά. Σε περιπτώσεις όπου τα δεδομένα δεν ακολουθούν κανονική κατανομή, τα συμβατικά διαγράμματα ελέγχου μπορούν να οδηγήσουν σε εσφαλμένα αποτελέσματα. Η επιλογή των διαγραμμάτων ελέγχου λοιπόν θα πρέπει να γίνεται με βάση τα χαρακτηριστικά των δεδομένων και τα διαγράμματα αυτά θα πρέπει να δίνουν ένα επιθυμητό μέσο μήκος ροής (ARL) για καταστάσεις εντός και εκτός στατιστικού ελέγχου. Ανάλογα με τα χαρακτηριστικά των δεδομένων μπορούμε να χρησιμοποιήσουμε διαγράμματα τα οποία μπορούν να χειριστούν μεταβαλλόμενα μεγέθη υποομάδων.

Με την πάροδο των χρόνων έχουν γίνει αρκετές έρευνες πάνω στο κομμάτι της εφαρμογής μη κανονικών δεδομένων σε διαγράμματα στατιστικού ελέγχου. Έχουν πραγματοποιηθεί αρκετές εφαρμογές σε διαφορετικούς τομείς. Έτσι έχουν προκύψει διάφορες τεχνικές ανάλογα με τον τύπο του προβλήματος, ποιο διάγραμμα ελέγχου προτείνεται για να δώσει τα κατάλληλα αποτελέσματα. Η εφαρμογή των διαγραμμάτων ελέγχου μέχρι σήμερα έχει καταφέρει να αποδείξει την σημαντικότητα τους αλλά και την εξέλιξη τους από την αρχική εφαρμογή τους στον κατασκευαστικό τομέα. Ωστόσο υπάρχουν ακόμα θέματα τα οποία χρειάζονται έρευνα. Ένα ζήτημα το οποίο επηρεάζει την έρευνα είναι η χρήση των διαγραμμάτων ελέγχου για την παρακολούθηση δεδομένων ικανοποίησης πελατών. Μερικά από αυτά περιλαμβάνουν το διάστημα που μεσολαβεί από τον χρόνο που χορηγείται η έρευνα μέχρι τον χρόνο που λαμβάνουν τα αποτελέσματα γεγονός το οποίο επηρεάζει τον εντοπισμό των ειδικών αιτιών. Σημαντικό να αναφερθεί είναι πως η χρήση των διαγραμμάτων ελέγχου μπορεί να χρησιμοποιηθεί για εντοπίσει την μεταβολή σε μια διαδικασία αλλά όχι για να λύσει το πρόβλημα. Οι μάνατζερ της επιχείρησης είναι εκείνοι που θα πρέπει να κρίνουν τα αποτελέσματα που θα λάβουν και να λάβουν τις κατάλληλες αποφάσεις.

Βασική μελλοντική επέκταση είναι η ανάπτυξη μιας ολοκληρωμένης διαδικασίας προσομοίωσης με δημιουργία μεγάλου αριθμού δεδομένων με διαφορετικές παραμέτρους.

Βιβλιογραφία

NIST. 1990. Malcolm Baldrige National Quality Award— 1989. Videocassette. Gaithersburg, Md.: National Institute of Standards and Technology.

Chanseok Park and Min Wang, January 2022 , Factors for Constructing Control Limits.

Wood,M. (1994), “Statistical methods for monitoring service process”, International Journal of service Industry Management .

Bart MacCarthy, Thananya Wasusri (2002) , A review of non-standard applications of statistical rocess control (SPC) charts , International Journal of Quality & Reliability Management.

John J. Spitzep (2007), “Hypothesis testing in models using the box-cox transformation”, Journal of Statistical Computation and Simulation.

Wardell, D.G and Candia, M.R (1996), “Statistical process monitoring of customer satisfaction survey data” , Quality Management Journal , Vol.3 No.4 pp36-50.

Douglas C. Montgomery (2005). Introduction to Statistical Quality Control, Fifth Edition, John Wiley & Sons, Inc.

Derek R. Allen, T. R. N. Rao (2000), Analysis of Customer Satisfaction Data, ”Preparing Customer satisfaction data”, Asq Pr

Steven Durtka, Department of Mathematical Sciences Montana State University (2018), Simulation Study of Average Run Lengths in Statistical Process Control and Effect of Non-Normally Distributed Data on ARL.

YM Chou, AM Polansky, RL Mason - Journal of Quality Technology, 1998 - Taylor & Francis, ”Transforming Non-Normal Data to Normality in Statistical Process Control “.

Ταγαράς Γιώργος (2001). Στατιστικός Έλεγχος Ποιότητας, Εκδόσεις Ζήτη.

Ποσειδών Εμμ. Ζαίρης (2010) , Στατιστική Μεθοδολογία -Αναθεωρημένη έκδοση , Εκδόσεις Κριτική ΑΕ.

Γεώργιος Δ. Τσιότρας (2016) , Διοίκηση Ολικής Ποιότητας , Εκδόσεις Π.Χ Πασχαλίδης και Broken Hill Publisers Ltd.

Ηλεκτρονικές πηγές

<https://www.isixsigma.com/tools-templates/normality/tips-recognizing-and-transforming-non-normal-data/>.

<https://www.isixsigma.com/tools-templates/control-charts/non-normal-data-needs-alternate-control-chart-approach/>.

<https://www.isixsigma.com/tools-templates/normality/dealing-non-normal-data-strategies-and-tools/>.

https://en.wikipedia.org/wiki/Beta_distribution.

https://en.wikipedia.org/wiki/Goodness_of_fit.

<https://www.tandfonline.com/doi/full/10.1080/00949658208810579>.

[https://www.researchgate.net/publication/235302234_A_review_of_non-standard_applications_of_statistical rocess control SPC charts](https://www.researchgate.net/publication/235302234_A_review_of_non-standard_applications_of_statistical_process_control_SPC_charts).

Παράρτημα Α :Παράμετροι κατανομής Johnson

Έστω x_1, x_2, x_3, x_4 τα ποσοστά δείγματος και ο QR είναι η ποσοστιαία αναλογία.

Ορίζουμε επίσης $\hat{x}_L = \hat{x}_2 - \hat{x}_1$, $\hat{x}_M = \hat{x}_3 - \hat{x}_2$, $\hat{x}_U = \hat{x}_4 - \hat{x}_3$.

Για την S_B κατανομή :

$$\begin{aligned}\hat{\eta} &= z / \cosh^{-1}(1/2[(1 + \frac{\hat{x}_M}{\hat{x}_U})(1 + \frac{\hat{x}_M}{\hat{x}_L})]^{1/2}) \\ \hat{\gamma} &= \hat{\eta} \sinh^{-1} \left[\frac{\left(\frac{\hat{x}_M}{\hat{x}_L} - \frac{\hat{x}_M}{\hat{x}_U} \right) \left\{ \left(1 + \frac{\hat{x}_M}{\hat{x}_U} \right) \left(1 + \frac{\hat{x}_M}{\hat{x}_L} \right) - 4 \right\}^{1/2}}{2(\hat{x}_M^2/(\hat{x}_L \hat{x}_U) - 1)} \right] \\ \hat{\lambda} &= \frac{\hat{x}_M \left[\left\{ \left(1 + \frac{\hat{x}_M}{\hat{x}_U} \right) \left(1 + \frac{\hat{x}_M}{\hat{x}_L} \right) - 2 \right\}^2 - 4 \right]^{1/2}}{(\hat{x}_M^2/(\hat{x}_L \hat{x}_U) - 1)} \\ \hat{\varepsilon} &= \frac{1}{2}(\hat{x}_3 + \hat{x}_2 - \hat{\lambda} + \frac{\hat{x}_M(\frac{\hat{x}_M}{\hat{x}_L} - \frac{\hat{x}_M}{\hat{x}_U})}{(\hat{x}_M^2/(\hat{x}_L \hat{x}_U) - 1)})\end{aligned}$$

Για την S_L κατανομή:

$$\begin{aligned}\hat{\eta} &= \frac{2z}{\ln \left(\frac{\hat{x}_U}{\hat{x}_M} \right)} \\ \hat{\gamma} &= \hat{\eta} \ln \left[\frac{\frac{\hat{x}_U}{\hat{x}_M} - 1}{(\frac{\hat{x}_U}{\hat{x}_M})^{1/2}} \right] \\ \hat{\varepsilon} &= \frac{1}{2} \left[\hat{x}_3 + \hat{x}_2 - \hat{x}_M \left(\frac{\frac{\hat{x}_U}{\hat{x}_M} + 1}{\frac{\hat{x}_U}{\hat{x}_M} - 1} \right) \right]\end{aligned}$$

Για την S_u κατανομή:

$$\begin{aligned}\hat{\eta} &= \frac{2z}{\cosh^{-1} \left[\frac{1}{2} \left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} \right) \right]} \\ \hat{\gamma} &= \hat{\eta} \sinh^{-1} \left[\frac{\frac{\hat{x}_L}{\hat{x}_M} - \frac{\hat{x}_U}{\hat{x}_M}}{2 \left(\frac{\hat{x}_L \hat{x}_U}{\hat{x}_M^2} - 1 \right)^{1/2}} \right]\end{aligned}$$

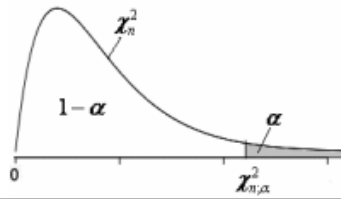
$$\hat{\lambda} = \frac{2 \hat{x}_M \left(\frac{\hat{x}_L \hat{x}_U}{\hat{x}_M^2} - 1 \right)^{1/2}}{\left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} - 2 \right) \left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} + 2 \right)^{1/2}}$$

$$\hat{\varepsilon} = \frac{1}{2} [\hat{x}_3 + \hat{x}_2 + \frac{\hat{x}_M (\frac{\hat{x}_L}{\hat{x}_M} - \frac{\hat{x}_U}{\hat{x}_M})}{\left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} - 2 \right)}]$$

Όπου $\widehat{QR} = \frac{(\widehat{\chi^U} \widehat{\chi^L})}{(\widehat{\chi^M})^2}$

- Εάν $\widehat{QR} \leq 1$ τότε η κατανομή SU υποθέτει ότι τα γ και λ δεν υπάρχουν.
- Εάν $\widehat{QR} \geq 1$ τότε η κατανομή SB υποθέτει ότι το λ δεν υπάρχει.
- Εάν $\widehat{QR} = 1$ τότε η κατανομή SB υποθέτει ότι το γ και ε δεν υπάρχει.
- Εάν $\frac{\widehat{\chi^U}}{\widehat{\chi^M}} \leq 1$ τότε η κατανομή SL υποθέτει ότι το λ δεν υπάρχει.
- Εάν $\frac{\widehat{\chi^U}}{\widehat{\chi^M}} = 1$ τότε η κατανομή SL υποθέτει ότι το η και το ε δεν υπάρχει.

Παράρτημα Β :Πίνακας κατανομής χ^2



n	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.414	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.335
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.706	22.164	24.4331	26.509	55.756	59.342	63.691	66.766
50	27.991	29.708	32.3574	34.764	67.505	71.420	76.154	79.490
60	35.535	37.485	40.4817	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.7576	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.1532	60.392	101.879	106.629	112.329	116.321
90	59.196	61.754	65.6466	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.2219	77.930	124.342	129.561	135.807	140.169