Λορέντζος Χαζάπης • Γιάννης Ζάραγκας

Διοίκηση Λειτουργιών

τα τετράδια μιας Οδύσσειας

τετράδιο 30

Χρήσιμα αλλά και κρίσιμα θέματα Στατιστικής

Αθήνα 2022

τετράδιο 30
Χρήσιμα αλλά και κρίσιμα θέματα Στατιστικής

ΠΕΡΙΛΗΨΗ

Ο Γιάννης και ο Λορέντζος συζητούν για την αξιοποίηση των χρήσιμων τεχνικών της Στατιστικής που δίνουν άμεση βοήθεια στη διοίκηση των Λειτουργιών. Θεωρούν ότι ασχολείται με την πρόβλεψη της συμπεριφοράς ενός χαρακτηριστικού μεγέθους κάποιου **πληθυσμού**, όταν η πληροφόρηση που έχουμε είναι οι τιμές ενός τυχαίου **δείγματος** (ή περισσοτέρων) από τον συγκεκριμένο πληθυσμό. Το δείγμα μάς δίνει μιαν εκτίμηση του χαρακτηριστικού μεγέθους του πληθυσμού και είναι προφανές πως όσο μεγαλύτερο είναι το δείγμα, τόσο πιο ακριβής είναι και η εκτίμηση. Η Στατιστική στηρίζεται στον λογισμό των πιθανοτήτων και χρησιμοποιεί μαθηματική σκέψη.

Η κλίμακα μέτρησης των πιθανοτήτων σχεδιάζεται από το p=0 - αδύνατο να συμβεί το γεγονός, έως p=1 - το γεγονός είναι βέβαιο. Μετά εξετάζουν μαζί δύο κλασικά εργαλεία της θεωρίας των πιθανοτήτων, τους κανόνες του αθροίσματος (αφορά αμοιβαία αποκλειόμενα γεγονότα) και του γινομένου των πιθανοτήτων (όταν τα γεγονότα να είναι ανεξάρτητα μεταξύ τους).

Στη συζήτηση για τις κατανομές των πιθανοτήτων, συμφωνούν ότι το εμβαδόν που περικλείεται από την καμπύλη και τον άξονα των τιμών αντιστοιχεί στη βεβαιότητα ή στην πιθανότητα 100% όλες οι τιμές του πληθυσμού περικλείονται από αυτή την καμπύλη. Ερευνούν τους ορισμούς των τεσσάρων βασικών τους παραμέτρων, που είναι η μέση, η κεντρική και η πιθανότερη τιμή, καθώς και η τυπική απόκλιση, που είναι το μέτρο της διασποράς της κατανομής.

Εξετάζονται οι ιδιότητες και ο τρόπος αξιοποίησης των βασικών κατανομών, Σταθερής, Κανονικής, καθώς και των κατοπτρικών Poison και Εκθετικής.

Στο πρόβλημα πώς μπορούμε από ένα δείγμα να κάνουμε εκτίμηση των παραμέτρων της κατανομής ενός πληθυσμού, ασχολούνται με τη Σημειακή και την εκτίμηση Εύρους ή Διαστήματος. Ξαναθυμούνται το Κεντρικό Οριακό Θεώρημα και τη σχέση του περιθωρίου εμπιστοσύνης (ακρίβεια εκτίμησης της παραμέτρου) με το ζητούμενο επίπεδο εμπιστοσύνης.

Τέλος, ασχολούνται και με τα διαγράμματα ελέγχου των διακυμάνσεων κάποιας μεταβλητής.

Τα τετράδια της Διοίκησης των Λειτουργιών δημοσιεύονται με ελεύθερη πρόσβαση στον ιστότοπο www.operations.gr
Παρατηρήσεις και σχόλια ευπρόσδεκτα στο Facebook (www.facebook.com/operations.gr)

Επιμέλεια κειμένου: Γιάννης Πλαχούρης

30η Συζήτηση

Χρήσιμα αλλά και κρίσιμα θέματα Στατιστικής

Πληροφόρηση και πρακτική χρήση της Στατιστικής

*Λορέντζο, στη διάρκεια των συζητήσεών μας, θίξαμε διάφορα θέματα της Στατιστικής, αλλά θέλω να επικεντρωθούμε στη σημερινή συζήτηση στις βασικές έννοιες και στις απλές τεχνικές της. Πάλι πλακώθηκα με το Λογιστήριο - όταν τους είπα ότι περιμένω πάνω από ένα μήνα τα αποτελέσματα του προηγούμενου, επειδή ισχυρίσθηκαν πως η καθυστέρηση των 15 ημερών για την έκδοση των μηνιαίων αποτελεσμάτων, είναι κάτι που συμβαίνει κατά μέσο όρο για χρόνια. Όταν ρώτησα «με ποια τυπική απόκλιση;», πέταξαν τη μπάλα στην εξέδρα, λέγοντάς μου με περισσή αθωότητα: «μα είπαμε περίπου κάθε 15 του μηνός έχεις τα αποτελέσματα, τι γκρινιάζεις;». Ε! όπως κατάλαβες έφυγα χωρίς να μιλήσω, γιατί όπως έλεγε η θεία μου η Πουλχερία: «**Μὴ δῶτε τὸ ἅγιον τοῖς κυσίν**». Άλλωστε ούτε αυτό θα το ερμήνευαν!*

Βλέπω με χαρά ότι επιδεικνύεις ηγετική συμπεριφορά αυθόρμητα και όχι παρορμητικά!

Δηλαδή;

Γιάννη, ο παρορμητικός αντιδρά, βγάζει συναισθήματα χωρίς επιλογή. Δεν επιδεικνύει ευελιξία και δεν υπηρετεί με μια πρόσφορη συμπεριφορά τον στόχο του. Αντίθετα, ο αυθόρμητος έχει επιλέξει τις συμπεριφορές του με βάση τις αξίες του. Έχει προσπαθήσει ενσυνείδητα να τις εφαρμόσει, οπότε του γίνονται συνήθειες και, καθώς είναι προϊόντα της επιλογής του, εκδηλώνονται αυθόρμητα, χωρίς δισταγμό, σύμφωνες με τις αξίες του, ουσιαστικές για τον στόχο του.

Υπονοείς ότι από παρόρμηση θα είχα γίνει έξαλλος. Όμως νομίζω ότι αρκεί να είσαι ακέραιος και να πιστεύεις στην πρόδραση, ώστε να μπορείς να φέρεσαι, όπως λες, ηγετικά, χωρίς να «σηκώνεις ατμό». Άλλωστε τι θα κέρδιζα; Τέλος πάντων. Και ποια είναι η βασική αξία που υπηρέτησα στην περίπτωση του Λογιστηρίου;

«Μην ανακατεύεσαι με τα πίτουρα γιατί θα σε φάνε οι κότες»!

Λορέντζο, πάντα βρίσκεις ευκαιρία να τονίζεις την αστεία πλευρά της διοίκησης. Η αλήθεια είναι ότι ξεκινήσαμε για ένα πολύ σοβαρό θέμα, τη Στατιστική, αλλά δεν μπορώ να μη διασκεδάζω με την κακοποίησή της από «σοφούς συμβούλους». Για παράδειγμα, η εταιρία που ερευνά τη θέση μας στον ανταγωνισμό αναφέρει διάφορα στατιστικά αποτελέσματα με ακρίβεια ερευνητικού εργαστηρίου βιοχημείας, όπως ότι: «το ξενοδοχείο μας θα το προτιμούσαν το 27,395% των κατοίκων της Ηπείρου». Όταν ρώτησα για τα περιθώρια εμπιστοσύνης αυτού του αποτελέσματος με κοίταζαν, όχι μόνον οι άλλοι αλλά και τα δικά μας στελέχη, με ύφος τέτοιο που έδειχνε ότι είμαι UFO.

Τέτοια διαμάντια ακούς από πολλούς «μορφωμένους». Απλό παράδειγμα στις ειδήσεις: δεν καταλαβαίνουν ότι η πτώση του ποσοστού από 10% στο 5% είναι εξαιρετικά μεγάλη και πολύ πιο σημαντική από την πτώση του ποσοστού από το 60% στο 50%. Γιάννη, η σοφή σου θεία Πουλχερία θα έλεγε επίσης ότι: «Υπάρχουν τα ψέματα, τα τρομακτικά ψέματα και η στατιστική».

Λορέντζο, εγώ πιστεύω στη μεγάλη βοήθεια που μπορούν να μου δώσουν οι μέθοδοι της Στατιστικής, όποτε χρησιμοποιούνται σωστά, όπως στο τι μπορεί να συμβεί σχετικά με τις τιμές του μεγέθους που παρακολουθούμε. Βέβαια ο δάσκαλος μάς είπε και το απίθανο: «Η στατιστική είναι σαν το μπικίνι. Δείχνει πάρα πολλά, αλλά όχι, κατ' ανάγκη, και τα κρίσιμα. Εξαρτάται από εσένα να τα ανακαλύψεις!». Αλήθεια, οι τεχνικές της πρόβλεψης της ζήτησης[1], που έχουμε δει, είναι τεχνικές της Στατιστικής;

[1] Βλ. *Η πρόβλεψη της ζήτησης*, Τετράδιο 8.

Γιάννη, από τη στιγμή που μας δίνουν μιαν αξιόπιστη πρόβλεψη, θεωρώ ότι αποτελούν απλές, αλλά έξυπνες στατιστικές τεχνικές. Άλλωστε, η Στατιστική ορίζεται ως μια συλλογή τεχνικών που χρησιμοποιούνται για τη συλλογή δεδομένων και ακολουθεί η ανάλυση και παρουσίαση των δεδομένων. Ασχολείται ουσιαστικά με την πρόβλεψη της συμπεριφοράς ενός χαρακτηριστικού μεγέθους κάποιου **πληθυσμού**, όταν η πληροφόρηση που έχουμε είναι οι τιμές ενός τυχαίου **δείγματος** (ή περισσοτέρων) από τον συγκεκριμένο πληθυσμό. Το δείγμα μάς δίνει μιαν εκτίμηση του χαρακτηριστικού μεγέθους του πληθυσμού και είναι προφανές πως όσο μεγαλύτερο είναι το δείγμα, τόσο πιο ακριβής είναι και η εκτίμηση. Η Αναλυτική Στατιστική ασχολείται με την εκτίμηση διαφορών και σχέσεων των δεδομένων, στηρίζεται στον λογισμό των πιθανοτήτων και χρησιμοποιεί μαθηματική σκέψη.

Πιθανότητες

*Ας ξεκινήσουμε με την έννοια της πιθανότητας και ας προσπαθήσουμε να αποφύγουμε τα πολλά μαθηματικά. Στις σημειώσεις μου από το Ίδρυμα, διαβάζω ότι ορίζουμε ως πιθανότητα τη σχετική συχνότητα εμφάνισης ενός γεγονότος. Μάλιστα αν ένα γεγονός **E** εμφανίζεται **f** φορές σε **n** επαναλήψεις (μέγεθος δείγματος) τότε η πιθανότητα το E να συμβεί σε κάθε επόμενη επανάληψη είναι*

$$Pr(E) = f/n$$

Πολύ σωστά. Η κλίμακα μέτρησης των πιθανοτήτων σχεδιάζεται από το p=0 που σημαίνει ότι είναι **αδύνατο** να συμβεί το γεγονός, για παράδειγμα, ότι μπορώ να διασχίσω πεζή την Ανταρκτική μόνος μου. Η άλλη άκρη της κλίμακας είναι το p=1 που σημαίνει ότι το γεγονός είναι **βέβαιο**, όπως ότι: Αν ξεχάσω το ψωμί στη φρυγανιέρα θα γίνει κάρβουνο.

*Λορέντζο, στο ερώτημα ποια είναι η πιθανότητα ένα ζάρι να φέρει 5 σε μία ρίψη, όλοι απαντάμε 1/6. Η υπόθεση που βρίσκεται στη βάση της απάντησης είναι ότι το ζάρι δεν είναι πειραγμένο και γίνονται πάρα πολλές ρίψεις. Όμως στον κόσμο της παραγωγής, σχεδόν σε όλα τα εξεταζόμενα γεγονότα, έχουμε μεγάλο αριθμό επαναλήψεων και επεμβαίνουν τυχαίοι ή/και άγνωστοι ενδογενείς παράγοντες. Αποτέλεσμα, η σχετική πιθανότητα στις μετρήσεις ενός χαρακτηριστικού μεγέθους είναι διαφορετική από τη βεβαιότητα του 1/6. Προφανώς είναι μια εκτίμηση της πιθανότητας, η οποία τείνει προς τη βεβαιότητα, όσο αυξάνει ο αριθμός των τιμών του δείγματος **n**.*

Ακριβώς. Μήπως στις σημειώσεις σου έχεις τους κανόνες του αθροίσματος και του γινομένου των πιθανοτήτων; Θυμάμαι ότι υπάρχουν, αλλά δεν θυμάμαι πώς τους χρησιμοποιείς.

Λογικό είναι. Όταν τελείωσες το Ίδρυμα, δεν είχε ανακαλυφθεί ακόμη η πιθανοθεωρία!

Γιάννη, άσε τις κοροϊδίες και ψάξε σε παρακαλώ.

*Λοιπόν, το βρήκα: Ο κανόνας του **αθροίσματος των πιθανοτήτων** αφορά **αμοιβαία αποκλειόμενα** γεγονότα. Να ένα απλό παράδειγμα: Έστω ότι παρακολουθώ μια ραπτομηχανή και καταγράφω τον αριθμό των σπασιμάτων της κλωστής κάθε 10 λεπτά. Με επιμονή κάνω τις παρατηρήσεις μου για μια εβδομάδα επί 7 ώρες ανά ημέρα επί 6 δεκάλεπτα ανά ώρα, δηλαδή 203 μετρήσεις (έχω χάσει 7 δεκάλεπτα, λόγω εξωτερικών επεμβάσεων). Έχω καταγράψει τα εξής:*

Κατανομή σπασιμάτων ανά10 λεπτο σε ραπτομηχανή		
Αριθμός σπασιμάτων ανά 10λεπτο	Αριθμός 10λέπτων (συχνότητα)	Πιθανότητα (σχετική συχνότητα)
0	70	0,34
1	93	0,46
2	26	0,13
3	8	0,04
4	4	0,02
5	2	0,01
6 (ή περισσότερα)	0	0,00
	203	1,00

Εάν ζητήσω την πιθανότητα να έχω λιγότερα από 2 σπασίματα στο 10λεπτο είναι προφανές ότι ζητώ την πιθανότητα να έχω 0 σπασίματα και 1 σπάσιμο. Τα γεγονότα αυτά δεν μπορούν να συμβούν ταυτόχρονα, συμβαίνει ή το ένα ή το άλλο. Είναι επομένως ***αμοιβαία αποκλειόμενα*** *(mutually exclusive). Σε αυτή την περίπτωση η πιθανότητα που ζητώ είναι το άθροισμα των 2 πιθανοτήτων. Θα είναι δηλαδή ίση με 0,34 + 0,46 = 0,80. Άρα στο 80% των μετρήσεων, δηλαδή στο 80% των δεκαλέπτων θα έχω λιγότερα από 2 σπασίματα ανά δεκάλεπτο.*

Και ο κανόνας του **γινομένου των πιθανοτήτων;**

Θα πρέπει τα γεγονότα να είναι ***ανεξάρτητα μεταξύ τους*** *και να προσπαθώ να προσδιορίσω την πιθανότητα κάποιο ή όλα να επηρεάσουν μια κατάσταση. Δες ένα παράδειγμα που βρήκα στις σημειώσεις μου: Έστω ότι ένα μηχάνημα προστατεύεται από 3 ηλεκτρικές ασφάλειες, Α, Β και Γ. Λειτουργεί μόνο αν και οι τρεις είναι λειτουργικές. Οι μετρήσεις μας έδειξαν ότι, σε έναν παραγωγικό κύκλο, οι πιθανότητα να καούν είναι 1%, 2% και 3% αντίστοιχα. Ποια είναι η πιθανότητα να σταματήσει η μηχανή σε έναν κύκλο από αστοχία των ασφαλειών;*

Γιάννη, πρέπει να ήσουν μεγάλος σπασίκλας, όπως λέγαμε τότε πριν 50 χρόνια κάποιους «υπερβολικά συνεπείς» συμμαθητές!

Έλα τώρα, σταμάτα τις κοροϊδίες.

Ωραία και πώς προχωρούμε; Η πιθανότητα να σταματήσει η μηχανή εξαρτάται από πολλά γεγονότα (να καεί η Α, η Β, η Γ, οι Α+Β μαζί, οι Α+Γ…).

Λορέντζο, είναι επομένως προτιμότερο να ασχοληθούμε με την πιθανότητα να μην καεί καμιά από τις ασφάλειες, ώστε να μη σταματήσει η μηχανή σε κάποιον παραγωγικό κύκλο. Σ' αυτήν την περίπτωση, οι αντίστοιχες πιθανότητες να μην καούν οι ασφάλειες Α, Β και Γ είναι 99%, 98% και 97%. Στην περίπτωση αυτή που τα γεγονότα είναι ***ανεξάρτητα μεταξύ τους*** *(λ.χ. το κάψιμο μιας ασφάλειας δεν επηρεάζει, ούτε επηρεάζεται από την κατάσταση των άλλων), ισχύει ο κανόνας του πολλαπλασιασμού. Άρα το μηχάνημα λειτουργεί με πιθανότητα σε κάθε κύκλο: 0.99 Χ 0,98 Χ 0,97 = 0,94*

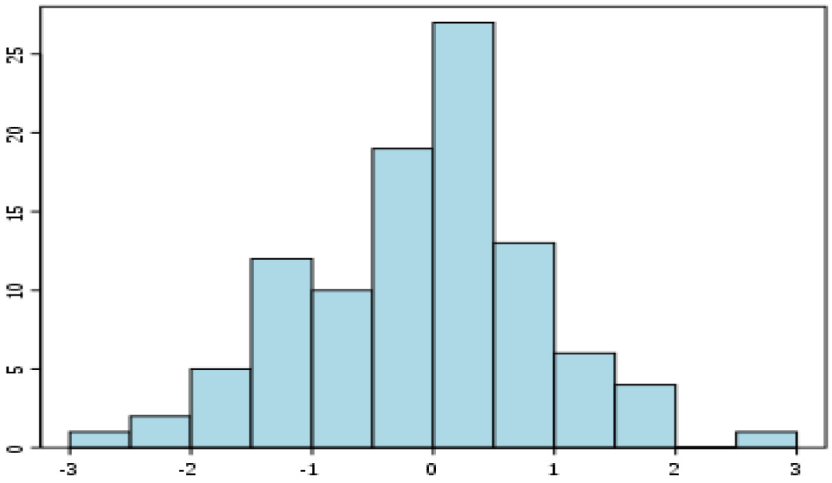
Μου λες δηλαδή, ότι η πιθανότητα να σταματήσει η μηχανή σε κάποιον κύκλο είναι 1 – 0,94 = 0,06 ή 6%

Ακριβώς!

Ιστόγραμμα. Κατανομή συχνότητας

Γιάννη, ας προχωρήσουμε. Στη στατιστική εξετάζουμε τον τρόπο που συμπεριφέρεται κάποιο σημαντικό χαρακτηριστικό από ένα γεγονός που μας ενδιαφέρει, όπως τα τεμάχια που παράγει ανά ώρα ένας εργαζόμενος. Το πλήθος των τιμών που μπορεί να λάβει το χαρακτηριστικό (εδώ ο **πληθυσμός**) θεωρείται άπειρο. Όμως υπάρχουν χαρακτηριστικά που λαμβάνουν **συνεχείς** τιμές, όπως για παράδειγμα τα ύψη των μαθητών ενός σχολείου, αλλά και **διακριτά** μεγέθη τιμών, όπως για παράδειγμα οι τιμές που μπορεί να βρούμε, σε πολλές ρίψεις ενός ζαριού. Οι **συχνότητες παρουσίασης** κάθε τιμής σε ένα δείγμα (σε μεταβλητές με διακριτές τιμές), ή να εμφανισθεί κάθε τιμή (σε συνεχείς μεταβλητές) σε κάποια διαστήματα τιμών που έχω ορίσει, καταγράφονται και μπορούν να παρουσιασθούν σ' ένα διάγραμμα που καλείται **ιστόγραμμα**. Είναι ένα διάγραμμα με τις μετρήσεις ενός χαρακτηριστικού (π.χ. το ύψος των μαθητών) στον οριζόντιο άξονα και τη συχνότητα εμφάνισης των διαφόρων μεγεθών στον κάθετο.

Θυμάμαι από την περίοδο που συζητούσαμε τα εργαλεία ποιότητας, ότι όταν κατασκευάζουμε[2] *το ιστόγραμμα των **n** τιμών του δείγματος που έχουμε πάρει, οπτικοποιούμε την κατανομή του μεγέθους και αντιλαμβανόμαστε καλύτερα τη μεταβλητότητά του. Από τη μορφή του συνάγουμε και τη μορφή που πρέπει να έχει και η κατανομή των τιμών του πληθυσμού. Οι μπάρες (συνήθως) εφάπτονται μεταξύ τους με πλάτος κάποια περιοχή μέτρησης (π.χ. ύψος μαθητών πάνω από 1,70 έως 1,75 περιλαμβανομένου) και ύψος τη συχνότητα που οι μετρήσεις έπεσαν σ' αυτή την περιοχή. Μια τυπική μορφή είναι η ακόλουθη:*



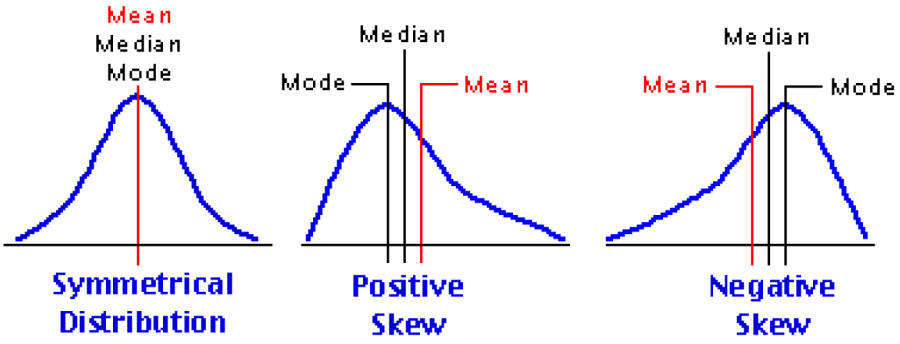
Σε πόσες περιοχές **n** (άρα και μπάρες) να χωρίσω τα δεδομένα; Άκου έναν πρακτικό κανόνα για να σε εντυπωσιάσω: Αριθμός μπαρών = log(**n**) / log(2).

Πράγματι με εντυπωσίασες. Καλά! Μπορείς να θυμάσαι τέτοιες λεπτομέρειες; Και να μιλάς για λογαρίθμους προφανώς ελάχιστες φορές από το Πολυτεχνείο. Εμείς ούτε που ασχοληθήκαμε. Τους υπολογισμούς πια τους κάνουν μηχανές και δεν χρειάζονται ως πρόσθετα εργαλεία οι άνθρωποι! Ας μιλήσουμε τώρα λιγάκι για την κατανομή των πιθανοτήτων των τιμών ενός πληθυσμού.

2 Βλ. *Εργαλεία Ποιότητας,* Τετράδιο 18α.

Βασικές παράμετροι των Κατανομών των Πιθανοτήτων

Γιάννη, σε έναν πληθυσμό με άπειρο αριθμό τιμών, αντί για τη **συχνότητα** των εμφανιζόμενων τιμών του δείγματος του χαρακτηριστικού που εξετάζουμε, έχουμε την **πιθανότητα** να εμφανισθεί η συγκεκριμένη τιμή. Το σύνολο των πιθανοτήτων εμφανίζεται πλέον σε μια κατανομή των πιθανοτήτων. Βασική λογική στις κατανομές των πιθανοτήτων είναι ότι το **εμβαδόν** που περικλείεται από την καμπύλη και τον άξονα των τιμών αντιστοιχεί στη βεβαιότητα ή στην πιθανότητα 100% όλες οι τιμές του πληθυσμού που μπορεί να εμφανισθούν περικλείονται από αυτή την καμπύλη. Όπως έχουμε ήδη πει η μορφή της κατανομής των πιθανοτήτων κατά συνθήκη μπορεί να είναι μια από τις παρακάτω τρεις.



Και ποιες είναι οι βασικές παράμετροι (measures of central tendency – μέτρα θέσεως – κεντρικές τιμές) που τις χαρακτηρίζουν;

Έλα τώρα. Όχι μόνο τις θυμάσαι, αλλά και τις χρησιμοποιείς. Βασικό μέτρο κάθε κατανομής είναι η **μέση τιμή** (mean - $\boldsymbol{\mu}$):

$$\boldsymbol{\mu} = \frac{\sum_1^N x_i}{N} \quad \text{για τον πληθυσμό,}$$

Όταν ασχολούμαστε με **δείγμα** του πληθυσμού, τότε αναφερόμαστε στον **μέσο όρο** (average - $\overline{\boldsymbol{x}}$) των μετρήσεων:

$$\overline{\boldsymbol{x}} = \frac{\sum_1^N x_i}{N} \quad \text{για το δείγμα}$$

Ονομάζεται και **αριθμητικός μέσος**. Υπάρχουν και ο γεωμετρικός και ο αρμονικός μέσος, αλλά σπάνια χρησιμοποιούνται.

*Λορέντζο, η **κεντρική τιμή** ή διάμεσος ή μέσος (Median), πως ορίζεται;*

Είναι η τιμή που ευρίσκεται στο μέσον της κατανομής συχνότητας κάποιων μετρήσεων και μας δηλώνει ότι υπάρχουν ίσες πιθανότητες να εμφανισθεί μια τιμή μεγαλύτερη ή μικρότερη από αυτήν. Οι επιφάνειες που περικλείονται από την καμπύλη της κατανομής, αριστερά και δεξιά από την κεντρική τιμή, είναι ίσες. Έχουμε δηλαδή την ίδια πιθανότητα 50% να είναι οι τιμές μεγαλύτερες ή μικρότερες από αυτή. Σε αντίθεση με τη μέση τιμή, η κεντρική δεν επηρεάζεται από τις ακραίες τιμές μιας κατανομής. Να γιατί στα θύματα του κορωνοϊού, αναφέρουν τη διάμεσο της ηλικίας: γιατί ένας υπεραιωνόβιος, με τη χρήση της μέσης τιμής, θα άλλαζε την εικόνα της κατανομής των ηλικιών.

*Και η **πιθανότερη τιμή** (Mode);*

Γιάννη, όπως λέει και το όνομά της, είναι η τιμή του μεγέθους ή της μεταβλητής που εμφανίζεται πιο συχνά. Και άκου και κάποιες πρακτικές που προκύπτουν από τη μελέτη των κατανομών:

> Αν χ = απόσταση του μέσου όρου από την κεντρική τιμή, τότε 3χ η απόσταση του μέσου όρου από την πιθανότερη τιμή.
>
> Ο νόμος του 5: Με 93,75% πιθανότητα, η κεντρική τιμή ενός πληθυσμού ευρίσκεται μεταξύ της μικρότερης και της μεγαλύτερης τιμής από ένα δείγμα 5 μετρήσεων του πληθυσμού.

*Λορέντζο, στις σημειώσεις μου από το Πολυτεχνείο βρίσκω ότι η **διασπορά των τιμών** είναι η άλλη βασική παράμετρος για τον προσδιορισμό μιας κατανομής. Ένα απλό μέτρο διασποράς είναι το **εύρος τιμών** (Range), δηλαδή η διαφορά μεταξύ της μικρότερης και μεγαλύτερης τιμής του δείγματος. Είναι ένα ατελές μέτρο γιατί δεν δείχνει πως διασπείρονται οι υπόλοιπες τιμές.*

Έχεις απόλυτο δίκαιο. Το μέτρο της διασποράς των τιμών του πληθυσμού που χρησιμοποιείται κυρίως είναι η **μεταβλητότητα**. Μας δείχνει τη διασπορά των τιμών ενός χαρακτηριστικού (της μεταβλητής) του πληθυσμού ή του δείγματος και συμβολίζεται με το **σ^2**. Με το σίγμα **σ**, συμβολίζουμε την **τυπική απόκλιση** (standard deviation) της κατανομής και με το **s** το μέτρο της διασποράς μιας μετρούμενης ποσότητας (ενός δείγματος).

*Θυμάμαι ότι είναι το μέτρο που ουσιαστικά προσδιορίζει και τη μορφή της κάθε κατανομής. Όσο μεγαλύτερη είναι η τυπική απόκλιση, τόσο πιο πλατιά είναι η καμπύλη της κατανομής, δηλαδή η διασπορά των τιμών δεξιά και αριστερά από τη μέση τιμή είναι μεγαλύτερη. Ο βασικός τύπος (μη με ρωτάς γιατί δεν διαιρώ με **n**, έτσι συνηθίζεται!) είναι ο ακόλουθος:*

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Στη βιομηχανία πολλές φορές χρησιμοποιείται και ο **συντελεστής μεταβλητότητας** (CV – coefficient of variation), που μετρά τη σχετική μεταβλητότητα ως ποσοστό του μέσου όρου. Είναι δηλαδή:

$$CV = \frac{100s}{\bar{x}}$$

Λορέντζο, θυμάμαι ακόμη ότι έχουμε ασχοληθεί με τη μεταβλητότητα του μεγέθους της εκροής κάποιας διεργασίας, γιατί μας αποκαλύπτει τη μεταβλητότητα των εισροών και, το πιθανότερο, των ελεγχουσών εισροών.

Πολύ σωστά. Η μεταβλητότητα είναι βασικό εργαλείο στην παραγωγή.

Προτείνω τώρα, να διατρέξουμε λίγο τις συνηθισμένες κατανομές που συναντάμε στις Λειτουργίες.

Οι συνηθισμένες κατανομές

1. Σταθερή Κατανομή

Γιάννη, την ονομάζουμε καταχρηστικά κατανομή και αφορά τις περιπτώσεις που οι μετρήσεις μας δίδουν **ακριβώς** το ίδιο αποτέλεσμα. Στη θεωρία αναμονής (παρίσταται με το D - deterministic) έχου-

με σταθερή κατανομή αν το χρονικό διάστημα μεταξύ αφίξεων είναι σταθερό. Η κατανομή έχει τυπική απόκλιση σ=0, χωρίς δηλαδή καμμιά μεταβλητότητα ($σ^2$).

Κάτι τέτοιο προφανώς συμβαίνει στη ροϊκή παραγωγή, όπου κάθε μονάδα προϊόντος φτάνει με μια αλυσίδα σταθερής ταχύτητας, σε κάθε σταθμό επεξεργασίας ή συναρμολόγησης. Αλλά και ο χρόνος εξυπηρέτησης μπορεί να υπακούει σε σταθερή κατανομή αν χρησιμοποιούμε ρομπότ. Και η κανονική κατανομή;

2. Κανονική Κατανομή

Αν, για παράδειγμα, καταγράψουμε με ένα όργανο ακρίβειας το βάρος του υγρού προϊόντος σε κάθε φιάλη μιας παραγωγικής διαδικασίας, δεν θα πάρουμε μόνο μια τιμή. Θα πάρουμε μια σειρά από τιμές, με τις οποίες, αν είναι από μεγάλο δείγμα, θα μπορούμε να κατασκευάσουμε το ιστόγραμμα. Αν δεν υπάρχει κάποιο τεχνικό πρόβλημα στη διεργασία και σχεδιάσουμε την περιβάλλουσα του ιστογράμματος, θα πάρουμε μια καμπύλη, η οποία **προσεγγίζει** σε πολύ καλό βαθμό την κανονική κατανομή. Ονομάζεται και καμπύλη του **Gauss** και οι βασικές της παράμετροι, μέση, κεντρική και πιθανότερη τιμή **συμπίπτουν**.

Ξέρω ότι είναι η πλέον διαδεδομένη κατανομή και συνήθως προσεγγίζει τα μετρούμενα μεγέθη (βάρος, διαστάσεις ή κάποιο ποιοτικό χαρακτηριστικό) σ' ένα παραγόμενο προϊόν. Δείχνει κατά πόσο η κατανομή των μετρήσεων ακολουθεί μια κανονικότητα, όπου ο μέσος όρος συμπίπτει ή είναι πολύ κοντά με τη συχνότερα εμφανιζόμενη τιμή και τη διάμεσο. Πολλά φυσικά μεγέθη ακολουθούν αυτή την τάση της κανονικότητας και αναμένουμε να υπάρχει η κανονική μορφή και στις μετρήσεις που κάνουμε στα αντίστοιχα δείγματα.

Γιάννη, μας προειδοποιεί επομένως, για ανωμαλίες ή δυσάρεστες εκπλήξεις όταν, για παράδειγμα, αναμένουμε κανονικότητα και οι μετρήσεις μας δίνουν διάγραμμα με συσσωρεύσεις μακριά από το κέντρο. Εξ ίσου σημαντικό είναι ότι μας βοηθά να αντιληφθούμε το μέγεθος της μεταβλητότητας (όσο πιο πλατιά είναι η καμπύλη, τόσο πιο μεγάλη είναι η μεταβλητότητα) του μεγέθους που μετράμε.

*Λορέντζο, πράγματι η κανονική κατανομή ορίζεται μόνο από τη **μέση τιμή** και την **τυπική απόκλιση** του πληθυσμού ή του δείγματος.*

Ναι, αυτά είναι τα δύο μεγέθη που ορίζουν την μορφή κάθε κανονικής κατανομής. Στη συζήτηση για τα $6σ$[3] θυμάσαι που αναφερθήκαμε στην πιθανότητα οι τιμές της μεταβλητής να βρίσκονται σε περιοχές της καμπύλης ανάλογα με την απόστασή τους από τη μέση τιμή, εκφρασμένη σε αριθμό τυπικών αποκλίσεων (σ). Ανεξάρτητα από το πλάτος της καμπύλης ή πόσο μεγάλη είναι η τυπική απόκλιση, όσο απομακρυνόμαστε από τη μέση τιμή, έχουμε όλο και μικρότερη πιθανότητα να βρούμε μια μέτρηση.

Είναι λογικό γιατί το εμβαδόν της καμπύλης προς τα δύο άκρα - ουρές συνεχώς μικραίνει, όσο τα πλησιάζουμε.

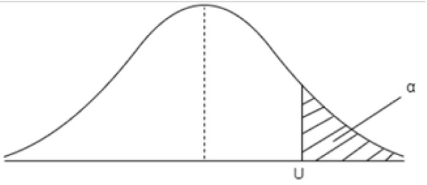
Γιάννη, με τη λογική της απόστασης από τη μέση τιμή, εκφρασμένη σε αριθμό τυπικών αποκλίσεων, χειριζόμαστε τα θέματα που αφορούν την κανονική κατανομή ανεξάρτητα από τις τιμές των δύο βασικών παραμέτρων της κατανομής (είτε του δείγματος, είτε του πληθυσμού). Στους Πίνακες της κανονικής κατανομής που σου δίνω, το **u** εκφράζει την απόσταση σε αριθμό τυπικών αποκλίσεων από τη μέση τιμή ($\bar{x}$) μιας μέτρησης (**x**) και το **α** το ποσοστό του εμβαδού ή της πιθανότητας που είναι στη μια ουρά από την τιμή x μέχρι το τέλος της καμπύλης (που θεωρητικά είναι το άπειρο). Σε πολλούς πίνακες

---

3 Βλ. *Η μεθοδολογία 6σ (Six Sigma),* Τετράδιο 26.

που θα βρεις σε διάφορα βιβλία, μπορεί το **u** να εκφράζεται με το γράμμα **z** και το **α** να αντιστοιχεί στο εμβαδόν (πιθανότητα) και των δύο άκρων – ουρών της κατανομής. Να ένας κλασικός πίνακας Στατιστικής για την κανονική κατανομή:

Η περιοχή στο άκρο - ουρά της Κανονικής Κατανομής. Οι εγγραφές στον πίνακα είναι τιμές:



a = Pr (u>U) για δεδομένες τιμές του U

U	α	U	α	U	α	U	α	U	α
0	0,5	0,6	0,2743	1,2	0,1151	1,8	0,0359	2,4	0,0082
0,02	0,492	0,62	0,2676	1,22	0,1112	1,82	0,0344	2,42	0,0078
0,04	0,484	0,64	0,2611	1,24	0,1075	1,84	0,0329	2,44	0,0073
0,06	0,4761	0,66	0,2546	1,26	0,1038	1,86	0,0314	2,46	0,007
0,08	0,4681	0,68	0,2483	1,28	0,1003	1,88	0,0301	2,48	0,0066
0,1	0,4602	0,7	0,242	1,3	0,0968	1,9	0,0287	2,5	0,0062
0,12	0,4522	0,72	0,2358	1,32	0,0934	1,92	0,0274		
0,14	0,4443	0,74	0,2296	1,34	0,0901	1,94	0,0262		
0,16	0,4364	0,76	0,2236	1,36	0,0869	1,96	0,025		
0,18	0,4286	0,78	0,2177	1,38	0,0838	1,98	0,0239		
0,2	0,4207	0,8	0,2119	1,4	0,0808	2	0,0228	2,6	0,0047
0,22	0,4129	0,82	0,2061	1,42	0,0778	2,02	0,0217		
0,24	0,4052	0,84	0,2005	1,44	0,0749	2,04	0,0207		
0,26	0,3974	0,86	0,1949	1,46	0,0721	2,06	0,0197		
0,28	0,3897	0,88	0,1894	1,48	0,0694	2,08	0,0188		
0,3	0,3821	0,9	0,1841	1,5	0,0668	2,1	0,0179	2,7	0,0034
0,32	0,3745	0,92	0,1788	1,52	0,0643	2,12	0,017		
0,34	0,3669	0,94	0,1736	1,54	0,0618	2,14	0,0162		
0,36	0,3594	0,96	0,1685	1,56	0,0594	2,16	0,0154		
0,38	0,352	0,98	0,1635	1,58	0,0571	2,18	0,0146		
0,4	0,3446	1	0,1587	1,6	0,0548	2,2	0,0139	2,8	0,0026
0,42	0,3372	1,02	0,1539	1,62	0,0526	2,22	0,0132		
0,44	0,33	1,04	0,1492	1,64	0,0505	2,24	0,0126		
0,46	0,3228	1,06	0,1446	1,66	0,0485	2,26	0,0119		
0,48	0,3156	1,08	0,1401	1,68	0,0465	2,28	0,0113		
0,5	0,3085	1,1	0,1357	1,7	0,0446	2,3	0,0107	2,9	0,0019
0,52	0,3015	1,12	0,1314	1,72	0,0427	2,32	0,0102		
0,54	0,2946	1,14	0,1271	1,74	0,0409	2,34	0,0096		
0,56	0,2877	1,16	0,123	1,76	0,0392	2,36	0,0091		
0,58	0,281	1,18	0,119	1,78	0,0375	2,38	0,0086		
0,6	0,2743	1,2	0,1151	1,8	0,0359	2,4	0,0082	3	0,0014

Λορέντζο, ας κάνω μια δοκιμή με τους πίνακες της στατιστικής. Για u = 1 (μία τυπική απόκλιση αριστερά και μία δεξιά από τη μέση τιμή), βρίσκουμε ότι το εμβαδόν στο διάστημα μεταξύ της γραμμής της μιας τυπικής απόκλισης και του τέλους της καμπύλης (στη μία ουρά της κατανομής), είναι το α = 15,87% του

συνόλου. Άρα μια τιμή έχει πιθανότητα 31,74% (= 2 x 15,87) να είναι εκτός του κεντρικού διαστήματος, στις δύο ουρές της κατανομής, δεξιά και αριστερά.

Η πιθανότητα δηλαδή μια τιμή του πληθυσμού να βρεθεί μέσα στο διάστημα των -1σ έως +1σ είναι:

$$1 - 0,3174 = 0,683 \text{ ή } 68,3\%$$

Στην περίπτωση που το διάστημα μεγαλώσει σε δύο τυπικές αποκλίσεις, από την τιμή -2σ αριστερά μέχρι +2σ δεξιά από τη μέση τιμή, το εμβαδόν περιορίζεται στο α = 2,28% ή στο 4,56% στις δύο ουρές.

Επομένως, η πιθανότητα μια τιμή του πληθυσμού να βρεθεί μέσα στο διάστημα των -2σ έως +2σ είναι:

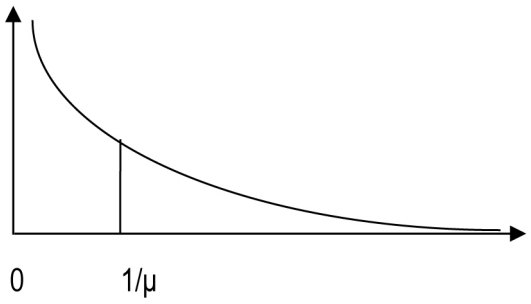
$$1 - 0,0456 = 0,954 \text{ ή } 95,4\%$$

Γιάννη, πρόσεξε και μια άλλη δυνατότητα. Στην περίπτωση που η στατιστική μέθοδος που χρησιμοποιούμε, απαιτεί τα δεδομένα να ακολουθούν την κανονική κατανομή και κάτι τέτοιο δεν συμβαίνει, μπορούμε να προχωρήσουμε σε μετατροπή τους. Μια μέθοδος είναι να υπολογίσουμε τους λογαρίθμους τους, οπότε λαμβάνουμε κανονική κατανομή (lognormal), κάνουμε τους υπολογισμούς μας όπως περιγράψαμε και επιστρέφουμε στα αρχικά δεδομένα.

Κατανομές Poisson και Εκθετική

Λορέντζο, έχω την εντύπωση ότι έχεις κάποιο κόλλημα με τους λογαρίθμους! Ας είναι. Είδαμε ότι η λογική και τα συμπεράσματα της Εκθετικής και της κατανομής Poisson αξιοποιούνται κατά κανόνα στα προβλήματα αναμονής. Η συνηθέστερη η περίπτωση αφίξεων συμβαίνει ***κατά τυχαίο τρόπο****, οπότε έχουμε την* ***κατανομή Poisson*** *με μέση τιμή* ***λ*** *αφίξεις/χρονικό διάστημα. Αντίστοιχα, το χρονικό διάστημα μεταξύ των αφίξεων ακολουθεί τη λεγόμενη* ***Εκθετική Κατανομή****, με μέση τιμή* ***1/λ****. Αυτές οι δύο «****κατοπτρικές****» κατανομές* ***προσεγγίζουν*** *με μεγάλη ακρίβεια πολλά πραγματικά προβλήματα και μας δίνουν τη δυνατότητα να τα μελετήσουμε καταλήγοντας σε πρακτικές λύσεις.*

Βασικό γνώρισμα της εκθετικής κατανομής είναι πως υπάρχει μεγάλη μεταβλητότητα στους χρόνους και πως – κατά διαστήματα – οι χρόνοι είναι πολύ μεγάλοι. Άρα είναι μεγάλη η **τυπική απόκλιση** και **ισούται με τη μέση τιμή 1/λ**. Οπότε η κατανομές αυτές ορίζονται από μόνο την μέση τιμή (λ ή 1/λ). Την ίδια λογική της τυχαιότητας ακολουθεί και η κατανομή των χρόνων εξυπηρέτησης με μέση τιμή και τυπική απόκλιση **1/μ**, όταν η μέση τιμή είναι μ εξυπηρετήσεις στη μονάδα του χρόνου.



Από τις σημειώσεις προηγούμενης συζήτησης[4] *βλέπω ότι στην εκθετική κατανομή, είναι πολύ αυξημένη η πιθανότητα για τον χρόνο εξυπηρέτησης, να είναι μικρότερος από τη μέση τιμή, σε σχέση με την πιθανότητα ο χρόνος εξυπηρέτησης να είναι μεγαλύτερος. Μάλιστα η πιθανότητα ο χρόνος να είναι*

[4] Βλ. *Εξυπηρέτηση και συστήματα αναμονής*, Τετράδιο 16.

μικρότερος από το μισό της μέσης τιμής, είναι μεγαλύτερη από το να κείται στο διάστημα από το 50 έως το 150 % της μέσης τιμής!

Γιάννη, να ένα παράδειγμα: Σε κατανομή με μ = 1 εξυπηρέτηση/ώρα δηλαδή 1/μ = 1 ώρα, η πιθανότητα να είναι ο χρόνος εξυπηρέτησης μικρότερος από τη μια ώρα, προκύπτει από τους τύπους ότι είναι 63,21%. Δηλαδή, είναι αυστηρά πτωτική η πιθανότητα να παρουσιασθεί ένα **συγκεκριμένο** χρονικό διάστημα μεταξύ αφίξεων (ή του χρόνου εξυπηρετήσεων), όσο αυτό το χρονικό διάστημα μεγαλώνει, ή αλλιώς είναι μεγαλύτερη η πιθανότητα να διαρκέσει η επόμενη εξυπηρέτηση χρόνο t από το να διαρκέσει 2t.

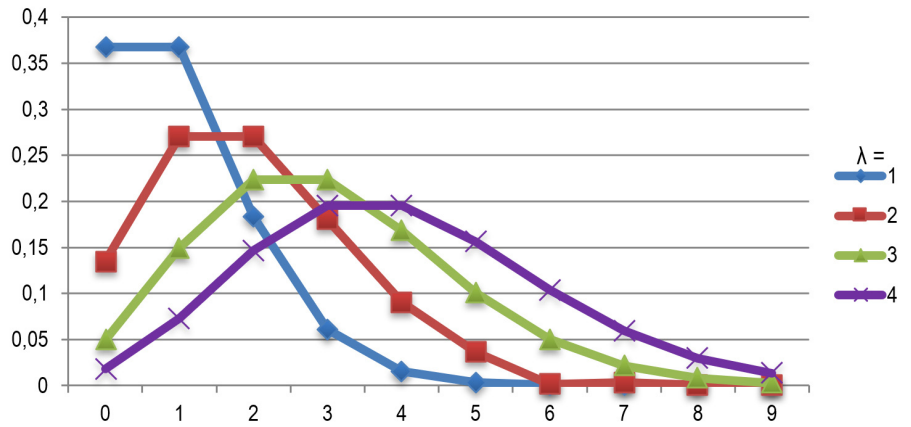
*Μπορούμε, με επιστημονικούς όρους, να καταλήξουμε ότι, στην απεικόνιση της πυκνότητας πιθανότητας, το εμβαδόν αριστερά από τη μέση τιμή είναι πολύ πιο μεγάλο από το εμβαδόν δεξιά. Την κατανομή τη λέμε επίσημα, «**αρνητική**» **εκθετική κατανομή**, επειδή πάει προς τα κάτω.*

Κύρια ιδιότητα της Poisson και της Εκθετικής, που τις κάνουν να ξεχωρίζουν, είναι ότι οι αντίστοιχες κατανομές δεν έχουν μνήμη. Η άφιξη ενός νέου πελάτη δεν επηρεάζεται από το πότε αφίχθηκε ο προηγούμενος. Ένας πελάτης μπορεί να εμφανισθεί αμέσως μετά τον προηγούμενο ή και μετά, ύστερα από μεγάλο χρονικό διάστημα. Αντίστοιχα στην εξυπηρέτηση, έστω και αν έχεις ήδη ασχοληθεί με έναν πελάτη πολύ, δεν σημαίνει ότι «όπου να 'ναι τελειώνει!». Όμως, αυτή η ιδιότητα της «αμνησίας» - όσο μεγαλώνει η παράμετρος «ρυθμός αφίξεων λ» ή «ρυθμός εξυπηρετήσεων μ» - χάνει σιγά-σιγά τη δύναμή της, επειδή η κατανομή Poisson **μεταπίπτει** στην Κανονική κατανομή, πάντα βέβαια με πολύ μεγάλη τυπική απόκλιση.

Μια άλλη σημαντική ιδιότητα, που συζητήσαμε στην περίπτωση των ουρών αναμονής, είναι ότι η αύξηση της μεταβλητότητας του χρόνου εξυπηρέτησης αυξάνει «εκθετικά» το μέγεθος των ουρών.

Η μορφή της Poisson

Ακριβώς, Γιάννη. Ας δούμε τώρα πώς μπορεί να χρησιμοποιήσει κάποιος την κατανομή Poisson. Όπως είπαμε, για να προσδιορίσει κανείς την κατανομή, πρέπει να γίνουν τυχαίες παρατηρήσεις, ώστε να ευρεθεί η μέση τιμή 1/λ του χρόνου μεταξύ αφίξεων, άρα και η τυπική απόκλιση (1/λ). Αντίστοιχα θα λογαριασθούν οι αφίξεις λ στη μονάδα του χρόνου. Από πίνακες βρίσκει κανείς την πιθανότητα συγκεκριμένου αριθμού αφίξεων στη μονάδα του χρόνου για κάθε λ. Στο κλασικό διάγραμμα της κατανομής Poisson, εμφανίζεται στον κάθετο άξονα η πιθανότητα να έχω τον αριθμό των αφίξεων στη μονάδα του χρόνου που καταγράφω στον οριζόντιο άξονα. Έτσι για μέσο όρο λ = 3 αφίξεις στη μονάδα του χρόνου, η πιθανότητα να έχω 3 αφίξεις στον αντίστοιχο χρόνο είναι περίπου 22% και για να έχω 4 αφίξεις είναι περίπου 17%.



*Επομένως, βλέποντας στο διάγραμμα την πράσινη καμπύλη (λ = 3), μπορώ να υπολογίσω την πιθανότητα να έχω μέχρι και 4 αφίξεις με βάση τον κανόνα του **αθροίσματος των πιθανοτήτων**, που αφορά **αμοιβαία αποκλειόμενα** γεγονότα. Στο άθροισμα των πιθανοτήτων έχω καμιά και μία και δύο και τρεις και τέσσερις αφίξεις, δηλαδή:*

5% + 15% + 22% + 22% + 17% = 81%

Αντίστοιχα η πιθανότητα να έχω περισσότερες από 5 αφίξεις είναι: 100% - 81% = 19%

Θυμίζω και μια άλλη ιδιότητα της εκθετικής κατανομής, την ιδιότητα του αθροίσματος: Ας υποθέσουμε ότι διάφοροι τύποι πελατών έρχονται στο σύστημα αναμονής, καθένας με τη δική τους εκθετική κατανομή. Τότε, η κατανομή των χρόνων μεταξύ των διαδοχικών αφίξεων του συνόλου των πελατών είναι πάλι εκθετική, με παράμετρο το άθροισμα των επί μέρους παραμέτρων λ. Δεν ισχύει το ίδιο και για τη διαφορά των δύο μεταβλητών. Αντίθετα, η διαφορά δύο μεταβλητών, που η κάθε μία ακολουθεί κανονική κατανομή, ακολουθεί και αυτή την κανονική κατανομή.

Λορέντζο, θυμάμαι ότι αυτή η ιδιότητα έχει και άλλη μορφή: Το άθροισμα από πολλές ανεξάρτητες εκθετικές τυχαίες μεταβλητές με την ίδια παράμετρο έχει εκθετική κατανομή με το άθροισμα των παραμέτρων. Αν έχω 5 σταθμούς ελέγχου διαβατηρίων σε ένα αεροδρόμιο, οι οποίοι διεκπεραιώνουν το έργο τους με εκθετική κατανομή με το ίδιο μ, τότε ο χρόνος, μέχρι το επόμενο πέρας της εξυπηρέτησης σε κάποιο σταθμό, έχει εκθετική κατανομή με παράμετρο το γινόμενο 5μ. Δηλαδή το σύστημα εργάζεται σαν ουρά αναμονής με έναν σταθμό ελέγχου με εκθετική κατανομή και μέση τιμή 1/5μ.

Εκτίμηση των βασικών παραμέτρων μιας κατανομής

Και πώς μπορούμε από ένα δείγμα να προβλέψουμε τις παραμέτρους της κατανομής ενός πληθυσμού;

Γιάννη, αυτό το λέμε «**εκτίμηση των παραμέτρων της κατανομής**». Χρησιμοποιούμε δύο μορφές εκτίμησης μιας παραμέτρου, όπως για παράδειγμα της μέσης τιμής μιας μεταβλητής. Τη Σημειακή και την εκτίμηση Εύρους ή Διαστήματος.

Σημειακή Εκτίμηση είναι κατ' αρχήν η εκτίμηση που προκύπτει από ένα δείγμα. Για παράδειγμα η μέση τιμή ενός δείγματος είναι $\bar{x} = \frac{\Sigma x}{n}$ και είναι η σημειακή εκτίμηση της μέσης τιμής **μ** του πληθυσμού. Το $\bar{x}$ ονομάζεται ο εκτιμητής – estimator του **μ**.

*Άρα, αντίστοιχα και η μεταβλητότητα του δείγματος $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ είναι η σημειακή εκτίμηση της μεταβλητότητας **σ²** του πληθυσμού;*

Η ακρίβεια της εκτιμηθείσας τιμής εξαρτάται από το μέγεθος της τυπικής απόκλισης **της κατανομής του εκτιμητή**, γιατί αυτή καθορίζει πόσο κοντά στη ζητούμενη τιμή του πληθυσμού θα πέσουν οι τιμές της παραμέτρου σε επαναλαμβανόμενα δείγματα. Καταλαβαίνω από το ύφος σου, ότι δεν κατάλαβες τίποτα!

Δεν έχεις άδικο…

Λοιπόν, με κάποιο μαγικό τρόπο (που δεν θυμάμαι πως αποδεικνύεται), ξέρουμε ότι η κατανομή του $\bar{x}$ έχει μέση τιμή το **μ** και η τυπική απόκλιση (αυτό που ονομάζουμε το τυπικό λάθος του $\bar{x}$ ισούται με το $s/\sqrt{n}$. Μάλιστα αυτή η κατανομή τείνει να είναι **κανονική**.

Λορέντζο, μου λες επομένως, ότι τα μεγάλα δείγματα μειώνουν την τυπική απόκλιση της κατανομής του εκτιμητή $\bar{x}$ και δίνουν ακριβέστερες εκτιμήσεις για τη μέση τιμή μ του πληθυσμού. Η ακρίβεια

όμως αυξάνει κατά την τετραγωνική ρίζα του μεγέθους του δείγματος. Επομένως, για να διπλασιασθεί η ακρίβεια, πρέπει να τετραπλασιασθεί το μέγεθος του δείγματος!

Ακριβώς. Αντίθετα, η **Εκτίμηση του Εύρους**, προτείνει από την ανάλυση του δείγματος ένα εύλογο – πιθανό εύρος τιμών το **Περιθώριο Εμπιστοσύνης** (μεταξύ δύο **ορίων εμπιστοσύνης**) για την παράμετρο του πληθυσμού που υπολογίζουμε. Το περιθώριο συσχετίζεται με το **ζητούμενο Επίπεδο Εμπιστοσύνης** ότι η πραγματική τιμή της παραμέτρου ευρίσκεται μέσα στο προτεινόμενο εύρος τιμών. Το εύρος του περιθωρίου ορίζεται από το ζητούμενο επίπεδο εμπιστοσύνης, το μέγεθος του δείγματος και τη μεταβλητότητα του δείγματος. Όσο μεγαλύτερο είναι το δείγμα, τόσο ακριβέστερη και η εκτίμηση της κάθε παραμέτρου του πληθυσμού που εξετάζεται.

Αντιλαμβάνομαι όμως ότι είναι πολύ σημαντικό το μέγεθος του ζητούμενου επιπέδου εμπιστοσύνης. Στις σημειώσεις μου από το Πολυτεχνείο βρίσκω μια πολύ χρήσιμη αναφορά:

> *Όσο μεγαλύτερο είναι το ζητούμενο επίπεδο εμπιστοσύνης, τόσο μεγαλύτερο καταλήγει να είναι και το περιθώριο εμπιστοσύνης, άρα και λιγότερο ακριβής η εκτίμηση της παραμέτρου.*

Για παράδειγμα, με επίπεδο εμπιστοσύνης 90%, το περιθώριο εμπιστοσύνης της πρόβλεψης (που έχει γίνει με τη συγκεκριμένη μέθοδο και συγκεκριμένο μέγεθος δείγματος) του ποσοστού ψήφων για το πρώτο κόμμα μπορεί να είναι ± 3%, ενώ, για επίπεδο εμπιστοσύνης 95%, το περιθώριο εμπιστοσύνης μπορεί να ανεβαίνει στο ± 10%. Συνήθως ζητείται το επίπεδο εμπιστοσύνης του 95%.

Γιάννη, πρόσεξε τώρα: Αν αξιοποιήσουμε τη λογική που περιγράψαμε για τη χρήση των πινάκων στις περιπτώσεις της κανονικής κατανομής, λαμβάνοντας υπ' όψη όσα είπαμε πριν για την κατανομή του $\bar{x}$ και την τυπική της απόκλιση, τότε τα Όρια Εμπιστοσύνης του **μ** για δεδομένο Επίπεδο Εμπιστοσύνης ίσο με 1 - 2α είναι

$$\frac{\bar{x}-Ua\cdot\sigma}{\sqrt{n}} \quad \text{έως} \quad \frac{\bar{x}+Ua\cdot\sigma}{\sqrt{n}}$$

Για το επίπεδο εμπιστοσύνης 90% το *α* = 0,05 και το *Ua* είναι από τους πίνακες 1,6449. Με λίγα λόγια:

> Το επίπεδο Εμπιστοσύνης αντιπροσωπεύει την συχνότητα με την οποία η αληθινή τιμή της άγνωστης παραμέτρου του πληθυσμού θα βρεθεί εντός του περιθωρίου Εμπιστοσύνης, σε συνεχείς δειγματοληψίες.

Κεντρικό Οριακό Θεώρημα

Λορέντζο, νομίζω ότι τα βασικά για το περιθώριο εμπιστοσύνης τα είπαμε. Με βάση αυτό που είπαμε για την μέση τιμή της κατανομής των $\bar{x}$ θυμήθηκα ακόμη και κάτι που ήταν σαν γρίφος, σχετικό με τη μέση τιμή των μέσων όρων με ένα όνομα σιδηρόδρομο!

Γιάννη, είσαι απίθανος. Θυμήθηκες μια εξαιρετική έννοια της στατιστικής που έχει και σχέση με όσα συζητήσαμε πριν. Είναι το Κεντρικό Οριακό Θεώρημα, σύμφωνα με το οποίο:

> Η μέση τιμή, των μέσων όρων μεγάλου αριθμού ανεξάρτητων παρατηρήσεων, ακολουθεί κατά προσέγγιση κανονική κατανομή, **ανεξάρτητα** από το ποια κατανομή ακολουθούν οι παρατηρήσεις.

Μάλιστα, όταν το μέγεθος του δείγματος είναι μεγαλύτερο του 20, να είσαι βέβαιος ότι ο μέσος όρος των δειγμάτων ακολουθεί κανονική κατανομή. Αυτό συμβαίνει γιατί σε πολλά φαινόμενα και πειράματα, οι τιμές των διαφόρων μεταβλητών είναι αποτέλεσμα αθροιστικής επίδρασης πολλών ανεξάρτητων αιτίων-παραγόντων, κανένα από τα οποία δεν υπερισχύει των άλλων. Για παράδειγμα: Ο χρόνος αναμονής σε μια ουρά είναι αποτέλεσμα πολλών παραγόντων, όπως η ημέρα της εβδομάδας, η ώρα της ημέρας, η αποτελεσματικότητα του υπαλλήλου, το είδος της συναλλαγής που διεκπεραιώνεται, κ.ά. Καθένας από τους παράγοντες αυτούς επιφέρει ένα θετικό ή αρνητικό αποτέλεσμα και όλοι μαζί αθροιστικά συντελούν στη διαμόρφωση του τελικού αποτελέσματος. Τέτοια χαρακτηριστικά-μεταβλητές, εμφανίζονται σε πολλά φαινόμενα και πειράματα.

Επομένως, το Κεντρικό Οριακό Θεώρημα συνδέει την κανονική κατανομή με οποιαδήποτε άλλη κατανομή (αφού δεν προϋποθέτει να ακολουθούν οι παρατηρήσεις την κανονική κατανομή), γεγονός το οποίο απαντάει, επίσης, στο ερώτημα, γιατί η κανονική κατανομή βρίσκει εφαρμογή σε μεγάλο πλήθος φαινομένων και πειραμάτων.

Μια άλλη εφαρμογή είναι ότι ο υπολογισμός του περιθωρίου εμπιστοσύνης της μέσης τιμής, που έχουμε κάνει προηγουμένως, ισχύει και όταν τα δεδομένα δεν ακολουθούν την κανονική κατανομή.

Διάγραμμα ελέγχου (control chart – run chart)

Λορέντζο, τα διαγράμματα Παρέτο[5] είναι στατιστικά εργαλεία;

Κατά κάποιον τρόπο ναι. Ουσιαστικά αναλύουν στοιχεία του παρελθόντος που υποψιαζόμαστε ότι προκαλούν προβλήματα ποιότητας σε μια παραγωγική διαδικασία, με στόχο τη διόρθωση και όχι την πρόληψη. Για την πρόληψη χρησιμοποιούμε τα **διαγράμματα ελέγχου** μελετώντας στοιχεία της τρέχουσας παραγωγής και αξιοποιώντας απλές έννοιες της στατιστικής.

Θυμάμαι ότι έχουμε αναφερθεί σ' αυτά[6]. Τα αξιοποιούμε για να παρακολουθούμε τη διακύμανση των τιμών της μεταβλητής που μας ενδιαφέρει, κατά πόσον ευρίσκεται μεταξύ των ορίων ελέγχου. Είχαμε μάλιστα τονίσει ότι, όταν η διεργασία κινείται εντός των ορίων, «δείχνει απλώς ότι αυτή υλοποιείται κατά τρόπο συνεπή με τον σχεδιασμό της και όχι κατ' ανάγκη ότι η υπηρεσία ή το προϊόν ευρίσκεται μέσα στα όρια των αποδεκτών προδιαγραφών από τον πελάτη».

Γιάννη, βλέπω ότι έχεις και εξαιρετική μνήμη. Ας δούμε λίγο καλύτερα τα πολύ χρήσιμα αυτά εργαλεία πρόληψης της κακής ποιότητας. Ελέγχουμε σε ορισμένα χρονικά διαστήματα έναν αριθμό **η** δειγμάτων **κάθε φορά**. Με βάση αυτά τα στοιχεία, υπολογίζουμε τη μέση γραμμή και τα όρια ελέγχου του διαγράμματος. Ο καθορισμός των ορίων είναι θέμα απλής στατιστικής και μπορεί ο καθένας να βρει τους **τύπους καθορισμού των ορίων**, όπως και τα στατιστικά εργαλεία για τον καθορισμό της **δυνατότητας της διεργασίας**, σε όλα τα βιβλία ποιοτικού ελέγχου. Τα όρια **συνήθως** υπολογίζονται σε απόσταση 3σ από τη μέση τιμή της μεταβλητής.

Νομίζω ότι είναι σημαντικό να μπορεί κανείς να «διαβάζει» ένα διάγραμμα ελέγχου και να επεμβαίνει όποτε υπάρχουν ενδείξεις «εκτός ελέγχου». Για το σκοπό αυτό καλό είναι να χωρίζουμε σε τρείς ζώνες ίσου πλάτους την περιοχή μεταξύ μέσου όρου και άνω ορίου και σε άλλες τρεις την περιοχή προς το κάτω όριο, με την ένδειξη Α στις εξωτερικές ζώνες και C της εφαπτόμενες στην ευθεία του μέσου όρου.

---

5 Βλ. *Εργαλεία εστίασης στον στόχο μας,* Τετράδιο 7.

6 Βλ. *Εργαλεία για την επίτευξη της ποιότητας,* Τετράδιο 18α.

Οι κλασικές ενδείξεις για καταστάσεις «εκτός ελέγχου» είναι:

1. Ένα ή περισσότερα σημεία εκτός των ορίων.
2. Δύο σημεία από τρία συνεχόμενα στη ζώνη Α (στην ίδια πλευρά).
3. Τέσσερις μετρήσεις από πέντε συνεχόμενες στη ζώνη Β ή και πέρα από αυτήν (στην ίδια πλευρά).
4. Εννέα συνεχόμενα σημεία στη μία πλευρά από την ευθεία του μέσου όρου.
5. Έξι συνεχόμενα σημεία αυξανόμενα ή μειούμενα, ανεξάρτητα θέσης στο διάγραμμα.
6. Δεκατέσσερα συνεχόμενα σημεία που εναλλάσσονται, ανεξάρτητα θέσης στο διάγραμμα.
7. Δεκαπέντε συνεχόμενα σημεία που περιορίζονται μέσα στα όρια των δύο κεντρικών ζωνών C (ανεξάρτητα πλευράς).

Γιάννη, πρόσεξε: Σε περιπτώσεις που σημεία/τιμές της μεταβλητής πέφτουν εκτός των ορίων (ή ακολουθούν τις προαναφερθείσες - στο run chart - ακολουθίες των 9 ή 6 σημείων), τότε λέμε ότι η διεργασία είναι εκτός ελέγχου. Κάτι τέτοιο συμβαίνει κατά συνθήκη από λάθη των εργαζομένων και απρόβλεπτα γεγονότα, όπως αλλαγή στην πρώτη ύλη, πτώση τάσης του ηλεκτρικού ρεύματος, ή στην περίπτωση των υπηρεσιών, από βιασύνη και παραβίαση της μεθόδου της εξυπηρέτησης. Πρέπει να άρουμε αυτές τις ανωμαλίες για να επιστρέψει η διεργασία εντός ορίων.

Επίσης, βλέπω από τις σημειώσεις μου ότι η ονομασία των διαφόρων τύπων διαγραμμάτων ελέγχου προέρχεται από το όνομα της μεταβλητής που παρακολουθούμε. Έτσι έχουμε:

Διαγράμματα Χ: παρακολουθούμε τη διακύμανση της μέσης τιμής του ποιοτικού χαρακτηριστικού που μας ενδιαφέρει από τα **n** δείγματα που επιθεωρούμε κάθε φορά.

Διαγράμματα R: η μεταβλητή που παρακολουθούμε είναι η διαφορά μεταξύ μέγιστης και ελάχιστης τιμής της μεταβολής του ποιοτικού χαρακτηριστικού.

Διαγράμματα P: η μεταβλητή είναι ο λόγος των ελαττωματικών προς το σύνολο των ελεγχθέντων του κάθε δείγματος.

Διαγράμματα C: η μεταβλητή είναι το ποσοστό των ελαττωμάτων ανά μονάδα υλικού (λ.χ. ένα αυτοκίνητο), όταν κάθε μονάδα ελέγχεται για περισσότερα από ένα χαρακτηριστικά.

Νομίζω ότι είπαμε πολλά, αλλά και πάρα πολύ λίγα για ένα θέμα που είναι τεράστιο, εξαιρετικά χρήσιμο, πολύ δύσκολο που παιρεταίρω ανάλυση ξεφεύγει από τα όρια αυτού του τετραδίου. Πραγματικά θαυμάζω τους ανθρώπους που ξέρουν Στατιστική και ακόμη περισσότερο αυτούς που τη χρησιμοποιούν σωστά στα καθημερινά προβλήματα των λειτουργιών. Καληνύχτα, Γιάννη.

Εύχομαι και εσύ να έχεις μια ήρεμη νύχτα χωρίς …στατιστικές διακυμάνσεις!

Έννοιες

Πιθανότητα είναι η σχετική συχνότητα εμφάνισης ενός γεγονότος.

Ο κανόνας του αθροίσματος των πιθανοτήτων αφορά αμοιβαία αποκλειόμενα γεγονότα και ο κανόνας του γινομένου των πιθανοτήτων αφορά γεγονότα που είναι ανεξάρτητα μεταξύ τους.

Βασικές παράμετροι (measures of central tendency – μέτρα θέσεως – κεντρικές τιμές) των κατανομών:

- Της κατανομής του πληθυσμού η μέση τιμή μ (Mean) και του δείγματος ο μέσος όρος $\bar{x}$ (average) των μετρήσεων. Επίσης η κεντρική τιμή ή διάμεσος ή μέσος (Median) και η πιθανότερη τιμή (Mode);
- Της κατανομής του πληθυσμού η τυπική απόκλιση σ (standard deviation) και με το s το μέτρο της διασποράς ενός δείγματος.

Σε τυχαίες αφίξεις με μέση τιμή λ αφίξεις/χρονικό διάστημα έχουμε κατανομή Poisson και το χρονικό διάστημα μεταξύ των αφίξεων ακολουθεί την Εκθετική κατανομή, με μέση τιμή 1/λ (κατοπτρικές κατανομές). Οι κατανομές αυτές δεν έχουν μνήμη.

Η μέση τιμή δείγματος $\bar{x} = \frac{\Sigma x}{n}$ είναι η σημειακή εκτίμηση της μέσης τιμής μ του πληθυσμού. Η κατανομή του $\bar{x}$ έχει μέση τιμή το μ και τυπική απόκλιση (ή τυπικό λάθος του $\bar{x}$) ίση με $s/\sqrt{n}$ (η κατανομή τείνει στην κανονική).

Εμπειρίες

Στην κανονική κατανομή, μέση, κεντρική και πιθανότερη τιμή συμπίπτουν.

Στην εκθετική κατανομή, είναι πολύ αυξημένη η πιθανότητα, ο χρόνος εξυπηρέτησης να είναι μικρότερος από τη μέση τιμή, σε σχέση με την πιθανότητα ο χρόνος εξυπηρέτησης να είναι μεγαλύτερος.

Μάλιστα, η πιθανότητα ο χρόνος να είναι μικρότερος από το μισό της μέσης τιμής είναι μεγαλύτερη από το να κείται στο διάστημα από το 50 έως το 150 % της μέσης τιμής!

Η πιθανότητα να διαρκέσει η επόμενη εξυπηρέτηση χρόνο t είναι μεγαλύτερη από το να διαρκέσει 2t.

Στην περίπτωση των ουρών αναμονής, είναι ότι η αύξηση της μεταβλητότητας του χρόνου εξυπηρέτησης αυξάνει «εκθετικά» το μέγεθος των ουρών.

Όσο μεγαλύτερο είναι το ζητούμενο επίπεδο εμπιστοσύνης, τόσο μεγαλύτερο καταλήγει να είναι και το περιθώριο εμπιστοσύνης, άρα και λιγότερο ακριβής η εκτίμηση της παραμέτρου.

Το επίπεδο Εμπιστοσύνης αντιπροσωπεύει τη συχνότητα στην οποία η αληθινή τιμή της άγνωστης παραμέτρου του πληθυσμού θα βρεθεί σε συνεχείς δειγματοληψίες εντός του περιθωρίου Εμπιστοσύνης.

Προτεινόμενα βιβλία και κείμενα

Yamane, Taro, «Statistics: An Introductory Analysis», 3rd ed., New York (N.Y.): Harper and Row, 1973.

Κίτσος Χρήστος, Παρ. «Υπολογιστική Στατιστική», Αθήνα, Εκδόσεις Νέων Τεχνολογιών.

Κιόχος Πέτρος Απ., «Στοιχεία Στατιστικής», Ίδρυμα Ευγενίδου, 1985.

Montgomery D., Runger G. “Applied Statistics and Probability for Engineers”, 7th ed. 2018, John Wiley & Sons, Inc.

Soong T.T., “Fundamentals of probability and statistics for engineers”, 2004, John Wiley & Sons Ltd.

Montgomery Douglas, “Introduction to Statistical Quality Control”, 7th Ed., 2013, John Wiley & Sons Inc.

Leaf G.A.V., “Practical Statistics for The Textile Industry: Part I”, Quality Control and Assessment Series, The Textile Institute,1984.