

Μέτρηση Διοικητικών Βαρών: Συνολική μεθοδολογία δειγματοληψίας και υπολογισμών

Εισαγωγή

Για τη μέτρηση του διοικητικού βάρους επιλέχθηκε ως βάση το Τυποποιημένο Μοντέλο Κόστους (ΤΜΚ) (Standard Cost Model - SCM), καθώς είναι η πιο διαδεδομένη μεθοδολογία για τον εντοπισμό, την αξιολόγηση και τον υπολογισμό του διοικητικού κόστους που προκύπτει όταν οι πολίτες και οι επιχειρήσεις συμμορφώνονται με τις διοικητικές απαιτήσεις. Ωστόσο, η μέθοδος αυτή καθορίζει πρωταρχικώς τα στοιχεία εκείνα που πρέπει να μετρηθούν καθώς και τον τρόπο με τον οποίο γίνεται η αναγωγή από μεμονωμένα δεδομένων σε όλη την οικονομία. Η προσαρμοσμένη μεθοδολογία, που έχει επελέγη, βασίζεται στο Τυποποιημένο Μοντέλο Κόστους έχει ως στόχο:

1. Την καλύτερη περιγραφή της πραγματικής κατάστασης
2. Τον ελάχιστο δυνατό συνολικό αριθμό ξεχωριστών σημείων έρευνας (ερωτηματολόγια ή άλλα)

Θεωρητικά, για να υπολογίσουμε το μέσο κόστος που επιβαρύνει έναν πολίτη (ή μια επιχείρηση) για μια δεδομένη διαδικασία, πρέπει πρώτα να εντοπίσουμε και να μετρήσουμε μια σειρά από συγκεκριμένες περιπτώσεις. Οι συγκεκριμένες περιπτώσεις πρέπει να θεωρούνται επαρκώς αντιπροσωπευτικές, ώστε οι μετρήσεις να είναι κατά μέσο όρο και να επεκταθούν σε όλη την Ελλάδα. Αυτός ο τρόπος λειτουργίας διασφαλίζει τη σκοπιμότητα και διασφαλίζει την κατανομή πολύτιμων πόρων. Στην πράξη, ο υπολογισμός του κόστους περιλαμβάνει τον προσδιορισμό ενός ορίου για τον ελάχιστο έγκυρο αριθμό συγκεκριμένων περιπτώσεων με βάση την αξιολόγηση των «σημείων μέτρησης» και μια μέθοδο επιλογής των τοποθεσιών όπου θα πραγματοποιηθούν οι μετρήσεις.

Μια σημαντική πρόκληση είναι ότι οι συνήθεις στατιστικοί τύποι που χρησιμοποιούνται για τον προσδιορισμό του μεγέθους του δείγματος για μια αντιπροσωπευτική έρευνα (οι οποίοι βασίζονται στη διάσταση του στατιστικού πληθυσμού και στο επίπεδο εμπιστοσύνης που θέλουμε να επιτύχουμε) μπορεί να είναι δύσκολο να εφαρμοστούν.

Οι λόγοι για αυτό είναι η σημαντική ποικιλομορφία σε όλη την επικράτεια και οι περιορισμένοι πόροι, που καθιστούν την έρευνα ενός μεγαλύτερου δείγματος ανέφικτη.

Για να διασφαλιστεί η ακρίβεια, είναι υψίστης σημασίας η λήψη και η χρήση δεδομένων για έναν αριθμό βασικών παραμέτρων. Οι παράμετροι αυτές αφορούν το σημείο επαφής των θεσμών με τον πολίτη.

Η προτεινόμενη λύση βασίζεται στην ακόλουθη προσέγγιση:

- **Βήμα 1:** ομαδοποίηση σε συστάδες (clustering) των τοποθεσιών όπου λαμβάνει χώρα η διαδικασία με βάση την ομοιότητα – χρησιμοποιώντας χαρακτηριστικά που σχετίζονται με την παράμετρο που θέλουμε να μετρήσουμε/εκτιμήσουμε.
- **Βήμα 2:** έρευνα μόνο της πιο αντιπροσωπευτικής τοποθεσίας για κάθε ομάδα και πραγματοποίηση ενός περιορισμένου αριθμού συνεντεύξεων/μετρήσεων σε κάθε ομάδα -

ο αριθμός αυτός να προσδιορίζεται από τυπικούς στατιστικούς τύπους και αποδοχή της «ατέλειας» στο επίπεδο ακρίβειας.

- Βήμα 3: υπολογισμοί με βάση τις μετρήσεις σε επιλεγμένες τοποθεσίες.

Οι Προκλήσεις

Για να μετρήσετε με ακρίβεια μια διαδικασία, π.χ. το κόστος έναρξης ατομικής επιχείρησης ενός ελεύθερου επαγγελματία, εστιάζουμε σε ορισμένες βασικές παραμέτρους που πρέπει αρχικά να μετρηθούν επιτόπου και στη συνέχεια να συνδυαστούν στους υπολογισμούς του βάρους.

Για παράδειγμα, εάν ο στόχος είναι να υπολογιστεί το κόστος της αναμονής στο γκισέ, ο χρόνος αναμονής που προκύπτει πρέπει να πολλαπλασιαστεί με το αντίστοιχο ποσό των ωριαίων αποδοχών του αιτούντος.

Με την τήρηση αυτού του τύπου, λειτουργούμε με την παραδοχή ότι ο χρόνος που χάνεται στο ταμείο αφαιρείται από τον χρόνο που ο επιχειρηματίας θα μπορούσε να έχει περάσει στην εργασία του. Στην πραγματικότητα, όμως, έχουμε μετρήσει τον χρόνο αναμονής και τις ωριαίες απολαβές ενός συγκεκριμένου επιχειρηματία. Τέτοιες παράμετροι αποτελούν συγκεκριμένες τιμές και δεν έχουν στατιστική αξία σε μεγαλύτερη κλίμακα. Προκειμένου να πραγματοποιηθεί ένας υπολογισμός που έχει στατιστική αξία σε μεγαλύτερο πληθυσμό, είναι απαραίτητο να αντιμετωπιστεί το πρόβλημα της στατιστικά σημαντικής δειγματοληψίας. Η στατιστικά σημαντική δειγματοληψία δεν αναφέρεται μόνο στο μέγεθος του δείγματος αλλά και στην κατανομή του μέσου όρου του δείγματος έτσι ώστε η στατιστική τιμή να είναι ανέπαφη.

Επιστρέφοντας στο παράδειγμα της αναμονής στο γκισέ, το ερώτημα παραμένει: πώς μπορεί να υπολογιστεί ένα στατιστικά σημαντικό μέτρο του χρόνου αναμονής;

Προβληματισμοί

Μια πρόταση για την απάντηση αυτής της ερώτησης θα ήταν να υπολογίσετε τον χρόνο αναμονής στο γκισέ κάθε γραφείου που εμπλέκεται στη διαδικασία. Από τα δεδομένα που συλλέχθηκαν θα προκύψει ένας μέσος όρος σταθμισμένος στον αριθμό των υποθέσεων που υποβλήθηκαν σε επεξεργασία. Αυτός ο αριθμός θα αντιπροσωπεύει τους χρόνους αναμονής που έχετε ως «βάρους» κάθε γραφείο σε αναλογία με τον αριθμό των επιχειρηματιών που είχαν πρόσβαση στο γραφείο. Αυτό ουσιαστικά σημαίνει ότι ένα γραφείο με παρατεταμένους χρόνους αναμονής θα ασκήσει πολύ μικρή επίδραση στον υπολογισμό του μέσου χρόνου αναμονής, εάν οι επιχειρηματίες που επιλέγουν αυτό το γραφείο είναι λίγοι.

Τα στοιχεία για τον υπολογισμό θα είναι:

- Χρόνοι αναμονής στο γκισέ για κάθε γραφείο
- Ο αριθμός των υποθέσεων που διεκπεραιώνονται από κάθε γραφείο

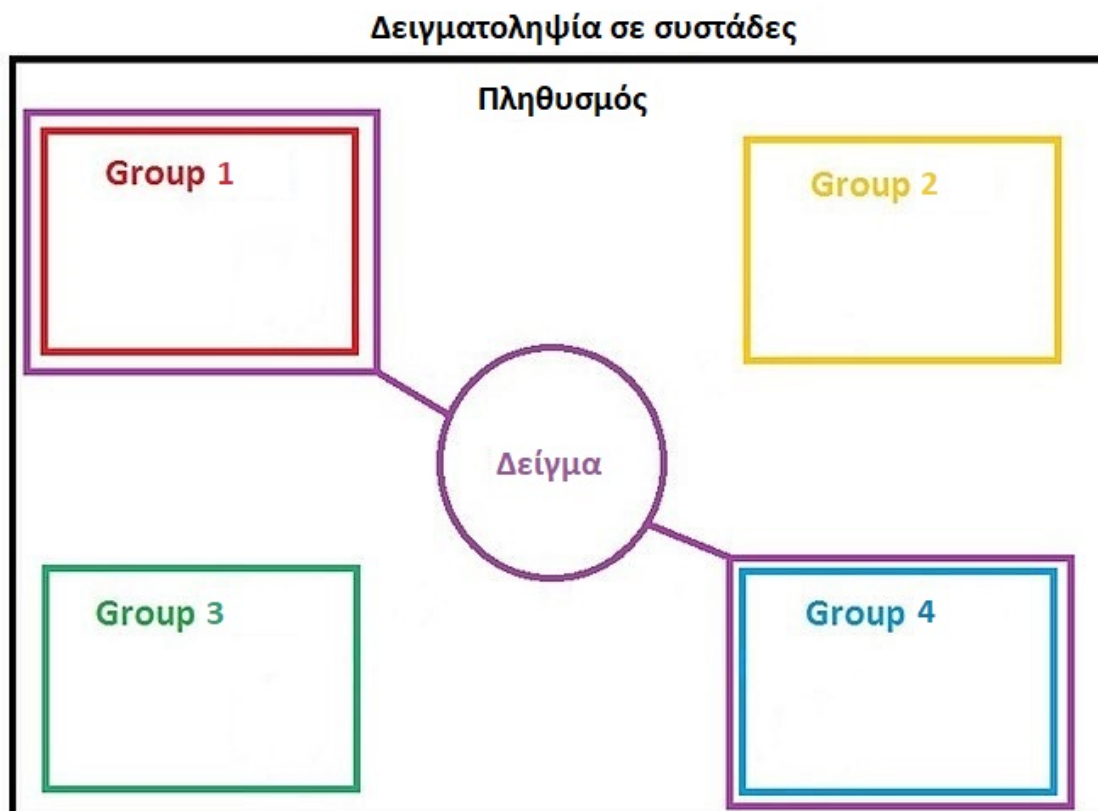
Αν και αυτή η προσέγγιση είναι αποτελεσματική στη θεωρία, είναι ελαττωματική στην πράξη. Η αξιολόγηση του μέσου χρόνου αναμονής για κάθε μεμονωμένο μετρητή (για παράδειγμα μέσω συνεντεύξεων) δεν μπορεί να συνδυαστεί με διαθέσιμους (ή ακόμη και δυνητικά διαθέσιμους)

πόρους. Δεδομένης της αδυναμίας λήψης συνεντεύξεων από όλους τους διαθέσιμους κλάδους, ποιοι καθοριστικοί παράγοντες περιορίζουν τους κλάδους στους οποίους πρέπει να γίνει συνέντευξη; Και πώς μπορούν να εφαρμοστούν τα αποτελέσματα της έρευνας; Αυτά είναι τα ζητήματα που επιδιώκει να επιλύσει η προτεινόμενη μεθοδολογία.

Βήμα 1: Ομαδοποίηση σε συστάδες (Clustering) και επιλογή «αντιπροσωπευτικών τοποθεσιών»

Το πρώτο βήμα, είναι η ομαδοποίηση των γραφείων (τοποθεσίες που σχετίζονται με τη διαδικασία που μετράται) που έχουν παρόμοια χαρακτηριστικά, λαμβάνοντας υπόψη μόνο τα χαρακτηριστικά που θεωρούνται ότι ενδέχεται να επηρεάσουν τη διαδικασία που μετράται (π.χ. χρόνος αναμονής για εγγραφή ως ελεύθερος επαγγελματίας). Για παράδειγμα, ο αριθμός των γκισέ σε κάθε γραφείο ή ο αριθμός των υπαλλήλων που είναι υπεύθυνοι για τη συγκεκριμένη διαδικασία και ο αριθμός των διαδικασιών που έχουν ολοκληρωθεί. Τέτοιοι καθοριστικοί παράγοντες αποτελούν τις βασικές παραμέτρους που πρέπει να αξιολογηθούν. Ωστόσο, τα δεδομένα για αυτά μπορεί να μην είναι πάντα διαθέσιμα, οπότε πρέπει να αρκεστούμε στους «δεύτερους καλύτερους» δείκτες που μπορούν να είναι σχετικοί με τον προσδιορισμό των χαρακτηριστικών των γραφείων.

Στις στατιστικές, αυτή η τεχνική είναι γνωστή ως «ανάλυση συστάδων». Η ομαδοποίηση χρησιμοποιείται για την ταξινόμηση αντικειμένων σε ομάδες (συστάδες). Αυτή η ταξινόμηση των ομάδων βασίζεται σε παρόμοια χαρακτηριστικά. Για παράδειγμα καλής πρακτικής, βλέπε 1.1



1.1 Παράδειγμα clustering

Ας υποθέσουμε ότι είστε ο επικεφαλής ενός καταστήματος ενοικίασης και θέλετε να κατανοήσετε τις προτιμήσεις των πελατών σας για να αναβαθμίσετε την επιχείρησή σας. Μπορείτε να δείτε λεπτομέρειες για κάθε πελάτη και να σχεδιάσετε μια μοναδική επιχειρηματική στρατηγική για κάθε έναν από αυτούς; Σίγουρα όχι. Αυτό όμως που μπορείτε να κάνετε είναι να συγκεντρώσετε όλους τους πελάτες σας σε 10 ομάδες, ας πούμε, με βάση τις αγοραστικές τους συνήθειες και να χρησιμοποιήσετε μια ξεχωριστή στρατηγική για τους πελάτες σε καθμία από αυτές τις 10 ομάδες. Και αυτό είναι που λέμε ομαδοποίηση σε συστάδες.

Πηγή: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>

Για τη διεξαγωγή δειγματοληψίας σε ομάδες πρέπει να ολοκληρωθούν τα ακόλουθα βήματα:

- 1) Προσδιορίστε τον πληθυσμό-στόχο και το μέγεθος του δείγματος
- 2) Δημιουργήστε ένα πλαίσιο δειγματοληψίας (δημιουργήστε μια πλήρη λίστα με όλες τις μονάδες ενδιαφέροντος)
- 3) Προσδιορίστε τις ομάδες (οι ομάδες πρέπει να αποκλείονται αμοιβαία)
- 4) Επιλέξτε συστάδες (τυχαία)

Για την εφαρμογή αυτών των τεσσάρων βημάτων, βλέπε 1.2.

1.2 Εφαρμογή των απαιτούμενων βημάτων για δειγματοληψία συστάδων-Παράδειγμα

Η Lulu θέλει να κάνει κάποια έρευνα μάρκετινγκ για την καμπάνια της Donna. Θέλει συγκεκριμένα πληροφορίες από τις διάφορες ομάδες ενδιαφερόντων του σχολείου. Στο σχολείο, διαπίστωσε ότι το 30% των μαθητών ασχολείται με τον αθλητισμό, το 25% των μαθητών ασχολείται με μια ακαδημαϊκή λέσχη, το 20% των μαθητών ασχολείται με μια λέσχη τέχνης ή θεάτρου και το 25% ασχολείται με ένα μουσικό κλαμπ. Κανένας από τους μαθητές δεν συμμετέχει σε περισσότερες από μία λέσχες και όλοι οι μαθητές συμμετέχουν σε μία. Η Lulu γνωρίζει ότι η Donna είναι μέλος αθλητικής λέσχης και οι μαθητές στίβου υποστηρίζουν την καμπάνια της Donna.

Κατά τη χρήση της δειγματοληψίας σε συστάδες, υπάρχουν δύο βασικές θεωρήσεις:

- Τα τυχαία δείγματα συμπλέγματος δεν μπορούν να διασταυρώνονται.
- Τα τυχαία δείγματα συμπλέγματος πρέπει να περιλαμβάνουν όλα τα μέλη ενός πληθυσμού.

Με άλλα λόγια, οι συστάδες πρέπει να είναι αμοιβαίως αποκλειόμενες. Σε αυτήν την περίπτωση, η Lulu δεν μπορεί να χωρίσει τον σχολικό πληθυσμό σε ομάδες όπου οι μαθητές συμμετέχουν σε περισσότερες από μία λέσχες ή είναι μέλη σε περισσότερες από μία συστάδες. Η Lulu αποφάσισε να διεξάγει έρευνα μόνο με τους μαθητές που

ασχολούνται με τις τέχνες, το θέατρο και τα μουσικά κλαμπ. Εφόσον η Lulu γνωρίζει ήδη ότι κανένας μαθητής δεν συμμετέχει σε περισσότερες από μία από τις κατηγορίες συλλόγων, γνωρίζει ότι το δείγμα της ομάδας δεν έχει αλληλεπικαλύψεις.

Γνωρίζουμε, επίσης, ότι τα δείγματα συστάδων πρέπει να περιλαμβάνουν όλα τα μέλη ενός πληθυσμού, που σημαίνει ότι όλοι οι μαθητές στο σχολείο πρέπει να είναι μέλη μιας από τις τέσσερις λέσχες. Εάν υπάρχουν μαθητές που δεν είναι μέλη μιας από τις λέσχες, τότε το δείγμα συστάδας δεν λειτουργεί.

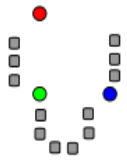
Εάν η Lulu δεν έχει χρόνο να δώσει μια έρευνα σε όλους τους μαθητές της λέσχης, μπορεί να χρησιμοποιήσει αμφίδρομη δειγματοληψία ομαδοποίησης σε συστάδες, η οποία είναι μια μέθοδος δειγματοληψίας που περιλαμβάνει τον διαχωρισμό του πληθυσμού σε συστάδες και, στη συνέχεια, την επιλογή τυχαίων δειγμάτων από καθεμία από αυτές. Η Lulu μπορεί να χρησιμοποιήσει απλή τυχαία δειγματοληψία για να επιλέξει μέλη κάθε συστάδας ή λέσχης, για να δώσει μια έρευνα.

Πηγή:

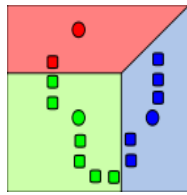
<https://study.com/academy/lesson/cluster-random-samples-definition-selection-examples.html>

Από τεχνική άποψη, για την πραγματοποίηση της ταξινόμησης γραφείων (ή άλλων τοποθεσιών/θεσμών για έρευνα) σε ομάδες, χρησιμοποιείται μια τεχνική που ονομάζεται K-Means. Αυτή η τεχνική είναι ένας εποπτευόμενος αλγόριθμος εκμάθησης που χρησιμοποιείται για την αναγνώριση και ταξινόμηση προτύπων και θα εκτελέσει αυτόματα αυτές τις ομαδοποιήσεις.

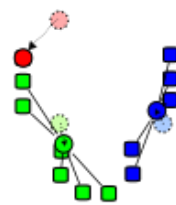
Επίδειξη του τυπικού αλγόριθμου ομαδοποίησης σε συστάδες



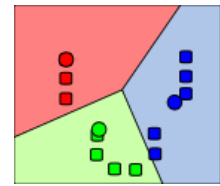
Σχήμα 1. k αρχικοί "μέσοι" (σε αυτή την περίπτωση $k=3$) δημιουργούνται τυχαία εντός του τομέα των δεδομένων (χρωματισμένοι)



Σχήμα 2. Οι k συστάδες δημιουργούνται συσχετίζοντας κάθε παρατήρηση με τον πλησιέστερο μέσο



Σχήμα 3. Το κεντροειδές των k συστάδων γίνεται ο νέος μέσος



Σχήμα 4. Τα βήματα 2 και 3 επαναλαμβάνονται μέχρι να επιτευχθεί σύγκλιση

Η εφαρμογή αυτού του αλγόριθμου επιτρέπει να:

- ορίσετε τη βέλτιστη ομαδοποίηση σε συστάδες (το k), δηλαδή τον αριθμό των συστάδων που αποτυπώνουν με ακρίβεια την κατάσταση, δεδομένης της κατανομής των μεταβλητών σε επιλεγμένες παραμέτρους (βλ. παρακάτω)
- ορίσετε το μαθηματικό κέντρο (μέσο όρο) κάθε σημείου στη συστάδα (το «κεντροειδές»).
- χρησιμοποιήσετε το «κεντροειδές» για να επιλέξετε την πιο αντιπροσωπευτική τοποθεσία για κάθε συστάδα (την πλησιέστερη σε αυτό από μαθηματικούς όρους, δηλαδή τη θέση της οποίας τα χαρακτηριστικά είναι πιο κοντά σε εκείνα του «κεντροειδούς»)¹.

Για ομαδοποίηση, χρειάζονται δεδομένα για πολλές παραμέτρους (στις οποίες θα λειτουργήσει ο αλγόριθμος). Είναι επομένως σημαντικό να καθοριστεί για ποιες παραμέτρους μπορούμε να λάβουμε ταυτόχρονα δεδομένα και ποιες είναι σχετικές.

Αυτές οι παράμετροι συχνά περιλαμβάνουν ένα ή περισσότερα από τα ακόλουθα:

1. Ιστορικά στοιχεία για τη συγκεκριμένη διαδικασία (αριθμός αιτήσεων ανά έτος, και, κατά περίπτωση, αριθμός αποφάσεων, προσφυγών κ.λπ.)
2. Δεδομένα για τη συγκεκριμένη τοποθεσία (διοικητικό γραφείο, επίπεδα προσωπικού, πιθανώς δείκτες του συνολικού φόρτου εργασίας, π.χ. δεδομένα για άλλες διαδικασίες που πρέπει να χειριστεί το γραφείο)
3. Πληθυσμός της περιοχής που καλύπτεται/εξυπηρετείται από την τοποθεσία (διοικητικό γραφείο)
4. Μέσο εισόδημα ή κατά κεφαλή ΑΕΠ της περιοχής που καλύπτεται/εξυπηρετείται από την τοποθεσία (διοικητικό γραφείο).

¹ Λάβετε υπόψη ότι δεν πρόκειται για ομαδοποίηση που βασίζεται στη γεωγραφική εγγύτητα αλλά στην ομοιότητα σε επιλεγμένες παραμέτρους (δηλαδή οι τοποθεσίες μέσα σε μία συστάδα μπορεί στην πραγματικότητα να βρίσκονται σε διαφορετικές περιοχές).

Για περισσότερα σχετικά με τον αλγόριθμο *k*-means, δείτε 1.3.

1.3 Αλγόριθμος K-means

Για να κατανοήσετε καλύτερα τον αλγόριθμο, είναι χρήσιμο να ανατρέξετε στα σχήματα 1-4.

Ο αλγόριθμος *k-means* ομαδοποιεί δεδομένα προσπαθώντας να διαχωρίσει δείγματα σε *n* ομάδες ίσης διακύμανσης, ελαχιστοποιώντας ένα κριτήριο γνωστό ως αδράνεια² ή άθροισμα τετραγώνων εντός συστάδας. Αυτός ο αλγόριθμος απαιτεί να καθοριστεί ο αριθμός των συστάδων. Δουλεύει καλά σε μεγάλο αριθμό δειγμάτων και έχει χρησιμοποιηθεί σε μεγάλο εύρος τομέων εφαρμογής.

Με βασικούς όρους, ο αλγόριθμος έχει τρία βήματα.

Το πρώτο βήμα επιλέγει τα αρχικά κεντροειδή, με τη πιο βασική μέθοδο να είναι αυτή όπου επιλέγετε *k* δείγματα από το σύνολο δεδομένων *X*. Μετά την αρχικοποίηση, το *k-means* γίνεται από την επανάληψη των δύο άλλων βημάτων.

Το πρώτο βήμα εκχωρεί κάθε δείγμα στο πλησιέστερο κεντροειδές.

Το δεύτερο βήμα δημιουργεί νέα κεντροειδή, λαμβάνοντας τη μέση τιμή όλων των δειγμάτων που έχουν εκχωρηθεί σε κάθε προηγούμενο κεντροειδές. Η διαφορά μεταξύ του παλιού και του νέου κεντροειδούς υπολογίζεται, και ο αλγόριθμος επαναλαμβάνει αυτά τα δύο τελευταία βήματα έως ότου αυτή η τιμή είναι μικρότερη από ένα κατώφλι. Με άλλα λόγια, επαναλαμβάνεται μέχρι να μην μετακινηθούν σημαντικά τα κεντροειδή.

Πηγή: <https://scikit-learn.org/stable/modules/clustering.html#k-means>

Αφού τα γραφεία (ή άλλες τοποθεσίες κ.λπ.) κατηγοριοποιηθούν σε ομάδες, για καθεμία από αυτές η ομάδα μέτρησης:

- επιλέγει το γραφείο που εκπροσωπεί την ομάδα,
- διεξάγει συνέντευξη(-εις) στο συγκεκριμένο γραφείο.

Τα αποτελέσματα που συλλέγονται από αυτό το γραφείο μπορούν επομένως να είναι ενδεικτικά των υπόλοιπων γραφείων της ίδιας ομάδας δεδομένου ότι έχουν κατηγοριοποιηθεί ως «παρόμοια» όσον αφορά τη διαδικασία που ερευνάται και τα καθοριστικά χαρακτηριστικά, π.χ. ο χρόνος αναμονής στο γκισέ.

Βήμα 2: Καθορισμός του μεγέθους δείγματος

Αφού προσδιορίσουμε τον αριθμό των συστάδων που απαιτούνται και ορίσουμε το μαθηματικό κέντρο κάθε σημείου της συστάδας στο βήμα 1, πρέπει να ορίσουμε το μέγεθος του δείγματος,

² Η αδράνεια μπορεί να αναγνωριστεί ως ένα μέτρο του πόσο εσωτερικά συνεκτικές είναι οι συστάδες: [2.3. Clustering — scikit-learn 0.23.2 documentation \(scikit-learn.org\)](https://scikit-learn.org/stable/modules/clustering.html#k-means)

δηλαδή τον πραγματικό αριθμό μετρήσεων για κάθε τοποθεσία (διοικητικό γραφείο ή άλλη τοποθεσία/υπηρεσία).

Οι μετρήσεις μπορεί να είναι συνεντεύξεις ή άμεσες παρατηρήσεις των αιτούντων κ.λπ. – αυτό το βήμα χρησιμεύει για τον καθορισμό του αριθμού των «δεδομένων μέτρησης».

Για περισσότερες πληροφορίες σχετικά με τον προσδιορισμό του μεγέθους του δείγματος, βλ. 2.1.

Το μέγεθος του δείγματος μπορεί να είναι είτε:

- «προκαθορισμένο» με βάση τους πόρους (στατιστικοί τύποι για το μέγεθος του δείγματος θα μας επιτρέψουν να προσδιορίσουμε το επίπεδο εμπιστοσύνης που θα παράγει ένα τέτοιο μέγεθος δείγματος), ή
- προσδιορίζεται χρησιμοποιώντας αυτούς τους ίδιους τύπους και ορίζοντας ένα επιθυμητό επίπεδο εμπιστοσύνης και αξιοπιστίας.

2.1 Καθορισμός μεγέθους δείγματος

Τα μεγέθη δειγμάτων μπορούν να επιλεγούν με διάφορους τρόπους:

– χρήση της εμπειρίας – μικρά δείγματα, αν και μερικές φορές αναπόφευκτα, μπορούν να οδηγήσουν σε μεγάλα διαστήματα εμπιστοσύνης και κίνδυνο λαθών στον έλεγχο στατιστικών υποθέσεων.

– χρησιμοποιώντας μια διακύμανση-στόχο για μια εκτίμηση που θα προκύψει από το δείγμα που τελικά λήφθηκε, δηλ. εάν απαιτείται υψηλή ακρίβεια (στενό διάστημα εμπιστοσύνης) αυτό μεταφράζεται σε χαμηλή διακύμανση στόχου του εκτιμητή.

– χρήση στόχου για την ισχύ μιας στατιστικής δοκιμής που θα εφαρμοστεί μόλις συλλεχθεί το δείγμα.

– χρησιμοποιώντας ένα επίπεδο εμπιστοσύνης, δηλαδή όσο μεγαλύτερη είναι η απαιτούμενη εμπιστοσύνη, τόσο μεγαλύτερο είναι το μέγεθος του δείγματος (δεδομένης μιας σταθερής απαίτησης ακρίβειας).

Το επαρκές μέγεθος δείγματος για ένα δεδομένο επίπεδο εμπιστοσύνης επηρεάζεται από τον τύπο κατανομής. Συνήθως, θα γνωρίζουμε την πραγματική κατανομή μόνο όταν ολοκληρωθεί η μέτρηση, αλλά γενικά μπορούμε να υποθέσουμε ότι οι περισσότερες πτυχές της μέτρησης θα ακολουθήσουν μια λεγόμενη «Κανονική» (Gaussian) κατανομή (η οποία απαιτεί σχετικά χαμηλό μέγεθος δείγματος σε σύγκριση με πιο «διασπαρμένες» διανομές). Σε ορισμένες περιπτώσεις, μπορεί επίσης να έχουμε πληροφορίες για τα άνω και/ή τα κάτω όρια ορισμένων στοιχείων προς μέτρηση, τα οποία επιτρέπουν και πάλι την ίδια εμπιστοσύνη με ένα σχετικά μικρότερο δείγμα.³

³ Για περισσότερες πληροφορίες σχετικά με τον τύπο προσδιορισμού μεγέθους: [Sample Size \(Definition, Formula\) | Calculate Sample Size \(wallstreetmojo.com\)](#)

Βήμα 3: Μετρήσεις και Υπολογισμοί

Συνολικά, οι κλασικοί τύποι για τον προσδιορισμό του μεγέθους του δείγματος ορίζουν τον αριθμό των σημείων δεδομένων που πρέπει να περιλαμβάνονται στο στατιστικό δείγμα, αλλά τέτοιοι τύποι υποθέτουν ότι το δείγμα δεν είναι προκατειλημμένο. Για να ξεπεράσουμε αυτόν τον περιορισμό, όπως αναφέρθηκε παραπάνω, εφαρμόζουμε:

- 1) ένας σταθμισμένος μέσος όρος για τα σημεία δεδομένων των γραφείων (για παράδειγμα ο χρόνος αναμονής),
- 2) μια μέθοδος ομαδοποίησης σε συστάδες για τη μείωση του αριθμού των γραφείων που πρέπει να ερευνηθούν για τη συλλογή αυτών των σημείων δεδομένων (βήμα 1).

Το επόμενο βήμα είναι να μετρήσετε/εκτιμήσετε πραγματικά το κόστος της διαδικασίας (όσον αφορά τις αμοιβές, τον χρόνο ταξιδιού, τον χρόνο αναμονής, το χαμένο εισόδημα κ.λπ.). Ενώ ορισμένα κόστη κωδικοποιούνται (επίσημα τέλη), άλλα πρέπει είτε να μετρηθούν «επιτόπου», είτε μέσω συνεντεύξεων, είτε να εκτιμηθούν με βάση στατιστικά δεδομένα (π.χ. χρόνος και κόστος μεταφοράς).

Οι μετρήσεις «επί τόπου» θα γίνονται σε βάση τυχαίας επιλογής. Αυτό μπορεί να είναι πιο δύσκολο στην περίπτωση διαδικασιών που δεν μπορούν να μετρηθούν άμεσα «επί τόπου» αλλά απαιτούν εξ αποστάσεως συνεντεύξεις (ενδέχεται να αντιμετωπίσουμε ζητήματα εμπιστευτικότητας των δεδομένων επικοινωνίας κ.λπ.), επομένως αξίζει περαιτέρω συζήτηση και εξέταση σε επιπλέον επαναλήψεις αυτής της σημείωσης.

Όσον αφορά τα δεδομένα που πρέπει να εκτιμηθούν με βάση τα διαθέσιμα στατιστικά δεδομένα (π.χ. μέσοι μισθοί ή απόσταση και χρόνος ταξιδιού, κ.λπ.), μια προκαταρκτική πρόταση θα ήταν η χρήση μιας λεγόμενης «στοχαστικής προσέγγισης» (“*Monte Carlo*”) να υπάρχει επαρκής αναπαράσταση των χαρακτηριστικών του υπό εξέταση πληθυσμού, της κατανομής διαφορετικών παραμέτρων κ.λπ.

3.1 Στοχαστική Προσέγγιση

Η στοχαστική προσέγγιση βασίζεται στην ιδιότητα να περιγράφεται καλά από μια τυχαία κατανομή πιθανοτήτων. Αυτή η προσέγγιση αξιολογεί την πιθανότητα διαφόρων αποτελεσμάτων, χρησιμοποιώντας τυχαίες μεταβλητές. Για την καλύτερη κατανόηση αυτής της προσέγγισης, είναι χρήσιμο να κατανοήσουμε τον ορισμό της ακριβώς αντίθετης προσέγγισης, δηλαδή της ντετερμινιστικής προσέγγισης.

Η ντετερμινιστική προσέγγιση παράγει τα ίδια αποτελέσματα για ένα συγκεκριμένο σύνολο εισροών, ανεξάρτητα από το πόσες φορές θα επανυπολογίσετε το μοντέλο. Αυτό συμβαίνει επειδή οι μαθηματικές ιδιότητες είναι γνωστές και δεν υπάρχει χρήση τυχαιοποίησης. Όταν χρησιμοποιείται μια ντετερμινιστική προσέγγιση, τυχόν αβέβαιοι παράγοντες είναι εξωτερικοί στην εφαρμογή του μοντέλου.

Από την άλλη πλευρά, η στοχαστική προσέγγιση βασίζεται στην τυχαιοποίηση. Είναι μια εγγενώς τυχαία προσέγγιση όπου οι αβέβαιοι παράγοντες ενσωματώνονται στο μοντέλο. Επομένως, αυτό το μοντέλο παράγει ένα πλήθος απαντήσεων, εκτιμήσεων και

αποτελεσμάτων για να δούμε τη διαφορετική επίδρασή τους στη λύση. Η ίδια διαδικασία συνήθως επαναλαμβάνεται πολλές φορές κάτω από διαφορετικά σενάρια.⁴

Πηγή: [Stochastic Modeling Definition \(investopedia.com\)](https://www.investopedia.com/terms/s/stochastic-modeling-definition/)

Η προσομοίωση Monte Carlo (βλ. 3.2) χρησιμοποιεί στατιστικές πληροφορίες για την προσομοίωση ενός δείγματος με χαρακτηριστικά συνεπή με τις πληροφορίες που συλλέγονται. Για παράδειγμα, ένα δείγμα 20.000 νέων επιχειρήσεων μετά την κατανομή του 2019 θα μπορούσε να δημιουργηθεί ως προς την κατοικία, το Επιμελητήριο αναφοράς, το γραφείο ΕΦΚΑ, τη ΔΟΥ κ.λπ. Σύμφωνα με το Τυποποιημένο Μοντέλο Κόστους, είναι δυνατός ο υπολογισμός όλων των δαπανών-επιβαρύνσεων (μεταφορές, αναμονή, αμοιβές κ.λπ.), σαν να αποτελούσαν υποθέσεις από συνέντευξη. Αυτό το δείγμα, αν και προσομοιωμένο, σέβεται όλα τα πραγματικά στατιστικά δεδομένα που έχουν συλλεχθεί και επιτρέπει έναν υπολογισμό ικανό να λαμβάνει υπόψη την αλληλεπίδραση μεταξύ των παραμέτρων.

Εξετάστε, για παράδειγμα, το κόστος αναμονής (στο γκισέ ή για μεταφορά). Από τη μία πλευρά, αυτό εξαρτάται από την αποτελεσματικότητα του γραφείου και από την απόσταση του γραφείου από την κατοικία. Από την άλλη, καθορίζεται και από τις ωριαίες απολαβές του επιχειρηματία. Προφανώς, αυτές οι ποσότητες δεν είναι ανεξάρτητες. Στα προάστια λιγότεροι επιχειρηματίες θα μεταφραζόταν σε πιο αποτελεσματικά γραφεία, ίσως σε χαμηλότερες αποδοχές. Στο κέντρο της Αθήνας η κατάσταση θα μπορούσε να είναι το αντίθετο.

3.2 Η προσομοίωση Monte Carlo

Η μέθοδος Monte Carlo βασίζεται σε επαναλαμβανόμενη τυχαία δειγματοληψία για τη λήψη αριθμητικών αποτελεσμάτων. Το βασικό στοιχείο αυτής της μεθόδου είναι ότι χρησιμοποιεί την τυχειότητα για την επίλυση προβλημάτων που είναι κατ' αρχήν ντετερμινιστικά. Στην ουσία, οι μέθοδοι Monte Carlo αποτελούν πειράματα. Ο πειραματισμός είναι η χρήση τυχαίων αριθμών για την εκτίμηση ορισμένων συναρτήσεων μιας κατανομής πιθανότητας. Ένα πρόβλημα που δεν έχει μια στοχαστική συνιστώσα μερικές φορές μπορεί επίσης να τεθεί ως πρόβλημα με μια συνιστώσα που μπορεί να προσδιοριστεί με την προσδοκία κάποιας συνάρτησης μιας τυχαίας μεταβλητής. Το πρόβλημα στη συνέχεια λύνεται με την εκτίμηση της αναμενόμενης τιμής με τη χρήση ενός προσομοιωμένου δείγματος από την κατανομή της τυχαίας μεταβλητής.

Οι μέθοδοι Monte Carlo χρησιμοποιούν τυχαίους αριθμούς, επομένως για να εφαρμοστεί μια μέθοδος Monte Carlo, είναι απαραίτητο να υπάρχει μια πηγή τυχαίων αριθμών. Οι μέθοδοι Monte Carlo χρησιμοποιούνται με διάφορους τρόπους στις στατιστικές. Χρησιμοποιούνται ευρέως στην ανάπτυξη στατιστικών μεθόδων.

Παράδειγμα: Πιθανότητες για Μπλάκτζακ

Ένα Μπλάκτζακ σε φύλλα αποτελείται από έναν Άσο και ένα φύλλο δέκα πόντων. Ας υποθέσουμε ότι θέλετε να βρείτε την πιθανότητα να πάρετε ένα blackjack (ένα «21» σε

⁴ Για περισσότερες πληροφορίες δείτε: [Stochastic Model / Process: Definition and Examples - Statistics How To](#)

φύλλα). Οι άσοι αξίζουν 11 πόντους και οι παρακάτω κάρτες αξίζουν 10: Βαλές, Ντάμα, Ρήγας. Θα μπορούσατε να γράψετε όλες τις πιθανότητες:

- Δέκα Σπαθιά / Άσος Σπαθί
- Βαλές Σπαθί / Άσος Σπαθί
- Ντάμα Σπαθί / Άσος Σπαθί
- Ρήγας Σπαθί / Άσος Σπαθί

Εάν σημειώνατε όλους τους πιθανούς συνδυασμούς φύλλων (συμπεριλαμβανομένων όλων εκείνων των συνδυασμών δύο φύλλων που δεν είναι άθροισμα 21, θα βρείτε ότι η πιθανότητα να πάρετε ένα Μπλάκτζακ είναι περίπου 1:21. Με μικρούς αριθμούς, όπως μια τράπουλα, ο υπολογισμός του δειγματικού χώρου (δηλαδή όλων των πιθανών αποτελεσμάτων) είναι αρκετά απλός και μικρός. Αλλά αν έχετε μεγαλύτερο αριθμό εισροών — ας πούμε χιλιάδες κάρτες, στη συνέχεια, ο εντοπισμός ενός δειγματικού χώρου χρησιμοποιώντας μια πιθανολογική μέθοδο όπως αυτή γίνεται δύσκολος.

Ένας άλλος τρόπος για να υπολογίσετε την πιθανότητα να αποκτήσετε ένα Μπλάκτζακ είναι να επιλέξετε δύο φύλλα έναν καθορισμένο αριθμό φορές (ας πούμε, εκατό φορές) και να καταγράψετε τα αποτελέσματα. Όσες περισσότερες φορές παίρνετε δείγμα δύο φύλλων, τόσο πιο κοντά θα πλησιάζετε στον «πραγματικό» αριθμό 1:21. Για παράδειγμα, αν επιλέξετε δύο φύλλα χίλιες φορές, πιθανότατα θα πλησιάσετε πολύ το 1:21. Εάν διαλέξετε δύο φύλλα δεκάδες φορές, πιθανότατα δεν θα πλησιάσετε καθόλου — μπορεί να έχετε «τύχη» ή μπορεί να μην πάρετε καθόλου «21». Αυτός είναι ουσιαστικά ο τρόπος με τον οποίο λειτουργούν οι προσομοιώσεις Monte Carlo.

Αντί να γράψει το δειγματικό χώρο (που είναι αυτό που κάναμε στο πρώτο μέρος αυτού του παραδείγματος), η μέθοδος Monte Carlo λαμβάνει δείγματα και εντοπίζει το πιο πιθανό αποτέλεσμα, δημιουργώντας ένα στοχαστικό μοντέλο. Για το λόγο ότι αυτή η προσομοίωση χρησιμοποιεί μια πολύ απλή κλήρωση (σε αυτό το παράδειγμα, δύο φύλλα) και την επαναλαμβάνει ξανά και ξανά, την έκανε να ονομάζεται επίσης «Μέθοδος Στατιστικών Δοκιμών».

Πηγές: [Monte Carlo Simulation / Method - Statistics How To](#) and [Monte Carlo Method - an overview | ScienceDirect Topics](#)