

ΜΑΘΗΜΑ: “Πληροφορική με Εφαρμογές Στατιστικής”

ΕΡΓΑΣΤΗΡΙΟ 4: Περιγραφική Στατιστική με το CALC

ΜΕΡΟΣ Α

Στο προηγούμενο εργαστήριο υπολογίσατε βασικά περιγραφικά στατιστικά μέτρα για τα δεδομένα στο αρχείο Lab3.ods. Το αρχείο με τα αποτελέσματα βρίσκεται στο eclass (Εγγραφα / Υλικό Εργαστηρίων (Calc), αρχείο Lab4.ods). Να το κατεβάσετε και να το αποθηκεύσετε στον υπολογιστή σας. Στη συνέχεια, ανοίξτε το με το Libre Office Calc και συμπληρώστε τις τιμές των μέτρων που λείπουν χρησιμοποιώντας τις αντίστοιχες συναρτήσεις του CALC.

1. Χρησιμοποιώντας τα αποτελέσματα του πίνακα στο αρχείο που μόλις ανοίξατε, να απαντήσετε στα παρακάτω ερωτήματα:

(i) Σε ποιο διάστημα βρίσκεται το 50% των παρατηρήσεων του δείγματος (δεδομένα της στήλης Closing); Προσοχή, να μην περιέχονται ακραίες τιμές σε αυτό.

(ii) Σε ποιο διάστημα βρίσκεται το 90% των παρατηρήσεων του δείγματος (δεδομένα της στήλης Closing); Προσοχή, να μην περιέχονται ακραίες τιμές σε αυτό.

(iii) Τι ποσοστό των τιμών κλεισίματος της μετοχής είναι το πολύ 9,35;

(iv) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής \leq του Q1.

(v) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής $>$ του Q3.

(vi) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής $< \bar{x}$.

(vii) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής στο διάστημα (8,3743 , 10,4119).

(viii) Να ερμηνεύσετε την τιμή της μέσης απόλυτης απόκλισης.

(ix) Είναι η κατανομή των δεδομένων συμμετρική; Τι μπορείτε να πείτε για την κύρτωση; Δώστε ερμηνεία των 2 αυτών μέτρων.

(x) Να υπολογίσετε την τιμή του συντελεστή μεταβλητότητας. Είναι το δείγμα ομοιογενές;

(xi) Ποια είναι τα όρια των διαστημάτων με απόσταση 1s, 2s και 3s (δηλ. 1, 2 και 3 δειγματικών τυπικών αποκλίσεων) από το δειγματικό μέσο των τιμών;

- Κάτω Όριο Άνω Όριο
- Κάτω Όριο Άνω Όριο
- Κάτω Όριο Άνω Όριο

(xii) Να βρείτε τα ποσοστά των παρατηρήσεων του δείγματος (τιμές της στήλης Closing) που βρίσκονται στα διαστήματα $\bar{x} \pm ks$, $k = 1,2,3$.

Υπόδειξη για το (xii): Να χρησιμοποιήσετε τη συνάρτηση **COUNTIF**. Η συνάρτηση αυτή συντάσσεται ως

=COUNTIF(range;criteria).

Στο range δίνουμε το σύνολο των αριθμητικών δεδομένων και στο criteria δίνουμε τα κριτήρια που θέλουμε να ικανοποιούνται. Για παράδειγμα, αν θέλουμε τις τιμές που είναι μεγαλύτερες από 50, θα πρέπει να το δώσουμε ως “>50”. Η COUNTIF κάνει καταμέτρηση των περιπτώσεων (cases) οι οποίες ικανοποιούν τα κριτήρια που βάλαμε (COUNT-IF \leftarrow \rightarrow “μέτρησε αν...”).

Μια παραλλαγή της COUNTIF είναι η **COUNTIFS**. Η συνάρτηση αυτή συντάσσεται ως

=COUNTIFS(range1;criteria1;range2;criteria2...),

και μας δίνει τη δυνατότητα να μετρήσουμε τις περιπτώσεις που ικανοποιούν περισσότερα από ένα κριτήρια. Μπορείτε να τη χρησιμοποιήσετε στο (δ).

Σημειώστε ότι με την COUNTIFS είναι (σχεδόν) άμεση η απάντηση αλλά απαιτείται περισσότερη δουλειά στη συγγραφή του τύπου.

ΜΕΡΟΣ Β

1. Να δημιουργήσετε μια στήλη με όνομα **DRD** (*daily relative difference*), όπου θα υπολογίσετε την ημερήσια σχετική μεταβολή στην τιμή κλεισίματος της μετοχής. Οι τιμές αυτές είναι στη στήλη Closing και οι νέες τιμές να δημιουργηθούν με βάση το πηλίκο

(Τιμή κλεισίματος την ημέρα i – τιμή κλεισίματος την ημέρα $(i-1)$)/(τιμή κλεισίματος την ημέρα $(i-1)$)

Να δοθούν: Το μέγεθος του συνόλου των τιμών της μεταβλητής DRD, η μέση τιμή, η τυπική απόκλιση, η ελάχιστη και η μέγιστη τιμή, το 1ο, το 2ο και το 3ο. Πότε παρατηρήθηκε η μεγαλύτερη ημερήσια μεταβολή και πότε η μικρότερη; Στο ερώτημα αυτό να χρησιμοποιήσετε μόνο τις παρακάτω συναρτήσεις του Calc: COUNT, SUM, SUMSQ, QUARTILE.

Πριν φύγετε, να αποθηκεύσετε (σε δικό σας φάκελο, είτε στο σκληρό δίσκο του Η/Υ στο εργαστήριο, είτε σε “φλασάκι”) το βιβλίο εργασίας που δημιουργήσαμε στο 4ο εργαστήριο. Να δώσετε το όνομα LAB4.xlsx. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας πριν φύγετε).