

ΜΑΘΗΜΑ: “Πληροφορική με Εφαρμογές Στατιστικής”

ΕΡΓΑΣΤΗΡΙΟ 4: Βασικά Στατιστικά Μέτρα & Υπολογισμός με το CALC

Ανοίξτε το CALC και φορτώστε το αρχείο Lab4.ods (υπάρχει στη σελίδα του μαθήματος στο e-class, στα Έγγραφα). Κατεβάστε το στον υπολογιστή σας και στη συνέχεια ανοίξτε το με το LibreOffice CALC. Παρατηρήστε ότι έχει ένα μόνο φύλλο εργασίας, το οποίο ονομάζεται Dell_Stock και αφορά την τιμή της μετοχής της εταιρείας DELL από 12/7/2019 έως και 11/10/2019 (τα δεδομένα είναι πραγματικά και έχουν αντληθεί από το Yahoo Finance).

- Χρησιμοποιώντας τα δεδομένα της στήλης «Close» (είναι η τιμή κλεισίματος της μετοχής, σε \$) και τις κατάλληλες συναρτήσεις του CALC, να συμπληρώσετε τη στήλη «Τιμές» τον παρακάτω πίνακα (προσοχή, μη τον μεταφέρετε στο CALC)

ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ	Τιμές	Συνάρτηση
Αριθμητικός μέσος		
Γεωμετρικός μέσος		
Αρμονικός μέσος		
Ελάχιστο		
Μέγιστο		
1ο τεταρτημόριο		
3ο τεταρτημόριο		
Διάμεσος		
5% ποσοστιαίο σημείο		
10% ποσοστιαίο σημείο		
20% ποσοστιαίο σημείο		
80% ποσοστιαίο σημείο		
90% ποσοστιαίο σημείο		
95% ποσοστιαίο σημείο		
Τάξη (rank) της τιμής κλεισίματος στις 12 Αυγούστου 2019		

Οι συναρτήσεις που θα σας χρειαστούν (ή τουλάχιστον, κάποιες από αυτές) είναι οι AVERAGE, GEOMEAN, HARMEAN, QUARTILE, PERCENTILE. Για τον υπολογισμό του Rank (Τάξη) μιας παρατήρησης, χρησιμοποιήστε την εντολή RANK.AVG.

- Δοκιμάστε να υπολογίσετε τα τεταρτημόρια και τη μέγιστη/ελάχιστη τιμή (α) μόνο με χρήση της Quartile(array;...), (β) μόνο με χρήση της Percentile(array;...). Για παράδειγμα, η διάμεσος μπορεί να υπολογιστεί ως =median(array) ή =quartile(array;2) ή =percentile(array;0,5).
- Δοκιμάστε επίσης τις εντολές quartile(array;0) και quartile(array;4). Εντοπίστε τα αποτελέσματα αυτών των εντολών στον παραπάνω πίνακα.

Σημείωση: Όπου array είναι το πλέγμα (σύνολο) κελιών που βρίσκονται τα δεδομένα του ερωτήματος.

- Τι ποσοστό των ημερών εμφάνισαν τιμή κλεισίματος μετοχής μικρότερη από το μέσο όρο όλων των τιμών κλεισίματος; Να χρησιμοποιήσετε τη συνάρτηση COUNTIF. Η συνάρτηση αυτή συντάσσεται ως

=COUNTIF(range;criteria).

- Εφαρμογή της COUNTIF: 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 ← → “μέτρησε αν...”).

- Ένας τρόπος για να μη γράφετε συγκεκριμένες τιμές στο criteria είναι ο εξής: Αρχικά, υπολογίστε σε ένα κελί F3 το μέσο όρο των τιμών κλεισίματος (θα είναι =AVERAGE(E2:E66)). Στη συνέχεια, για να βρείτε το πλήθος των τιμών που είναι μικρότερο από το μέσο όρο, δώστε π.χ. στο F4 την εντολή =COUNTIF(E2:E66;"<"&F3).
 - Με τη συνάρτηση αυτή θα βρείτε τις περιπτώσεων (ημέρες) που αντιστοιχούν σε τιμές της Close μικρότερες από το μέσο όρο. Στη συνέχεια, να μετατρέψετε αυτό τον αριθμό σε ποσοστό (προσοχή, υπάρχει ένα ενδιάμεσο βήμα!).
3. Με τον ίδιο τρόπο, (i) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής \leq του Q_1 και (ii) Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής $>$ του Q_3 .
 4. Ποιος είναι ο αριθμός (και το αντίστοιχο ποσοστό) των ημερών με τιμή κλεισίματος μετοχής στο διάστημα $(Q_1 - 0.5 \cdot IQR, Q_3 + 0.5 \cdot IQR)$. Το $IQR = Q_3 - Q_1$. Εδώ, μπορείτε να χρησιμοποιήσετε μια παραλλαγή της COUNTIF, τη συνάρτηση 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 δουλέψετε και με την countif αλλά εκεί θα χρειαστεί λίγες περισσότερες πράξεις (**άσκηση HW!**).

Για να υπολογίσετε το ζητούμενο αριθμό ημερών με τιμή κλεισίματος μετοχής στο $(Q_1 - 0.5 \cdot IQR, Q_3 + 0.5 \cdot IQR)$, υπολογίστε αρχικά τα όρια του διαστήματος (χρησιμοποιήστε τις τιμές Q_1, Q_3 που βρήκατε στο (1)). Στη συνέχεια, δώστε την εντολή

=COUNTIFS(E2:E66;">"&cell1; E2:E66;"<"&cell2)

Στα cell1, cell2 είναι τα κελιά με τις τιμές του κάτω και του άνω ορίου του διαστήματος, αντίστοιχα.

5. 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 **Close** 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 (χρησιμοποιήστε τη συνάρτηση COUNT()), η μέση τιμή, η ελάχιστη και η μέγιστη τιμή, το 1ο, το 2ο και το 3ο τεταρτημόριο. Πότε παρατηρήθηκε η μεγαλύτερη ημερήσια μεταβολή και πότε η μικρότερη; Στο ερώτημα αυτό να χρησιμοποιήσετε μόνο τις παρακάτω συναρτήσεις του Calc: COUNT, SUM, QUARTILE.

6. Να δημιουργήσετε μια στήλη με όνομα DfM (Difference from Mean), όπου θα υπολογίσετε τη διαφορά της ημερήσιας τιμής κλεισίματος, από το μέσο όρο όλων των τιμών. Στη συνέχεια, να βρείτε το άθροισμα αυτών των τιμών. Τι παρατηρείτε;
7. Να φτιάξετε μια μεταβλητή με όνομα DfMabs και μια μεταβλητή με όνομα DfMsq. Στην 1^η να υπολογιστούν οι απόλυτες τιμές των διαφορών κάθε τιμής κλεισίματος από τα η μέση τιμή, στη 2^η τα τετράγωνα των διαφορών κάθε τιμής κλεισίματος από τα η μέση τιμή. Να βρείτε τα αθροίσματα των τιμών για τις 2 νέες μεταβλητές. Δώστε την εντολή =DEVSQ(E2:E66). Τι παρατηρείτε; Στο array E2:E66 είναι τα δεδομένα της "Close".

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.ods. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας πριν φύγετε).