

ΜΑΘΗΜΑ: "Πληροφορική με Εφαρμογές Στατιστικής"

ΕΡΓΑΣΤΗΡΙΟ 5: Βασικά Στατιστικά Μέτρα & Υπολογισμός με το CALC II

Άσκηση 1^η: Ανοίξτε το CALC και φορτώστε το αρχείο Lab5.ods (υπάρχει στη σελίδα του μαθήματος στο e-class, στα Έγγραφα). Το dataset είναι από το βιβλίο των Berk and Carey (2000) *Data Analysis with Microsoft Excel*, Duxbury (ελαφρώς τροποποιημένο). Κατεβάστε το στον υπολογιστή σας και στη συνέχεια ανοίξτε το με το LibreOffice CALC. Παρατηρήστε ότι έχει ένα μόνο φύλλο εργασίας, στο οποίο δίνονται τιμές των παρακάτω μεταβλητών

Price: Τιμή πώλησης του σπιτιού
Square feet: Το μέγεθος του σπιτιού σε τετραγωνικά πόδια (μονάδα μέτρησης εμβαδού, 1 square feet = 0.09290304 square meters)
Age: Ηλικία του σπιτιού (σε έτη)
Features: Πλήθος αντικειμένων οικιακού εξοπλισμού (π.χ. πλυντήριο πιάτων, ψυγείο, φούρνος μικροκυμάτων, το πολύ 11)
NE Sector: Βρίσκεται στη νοτιοανατολική πλευρά της πόλης (Yes/No)
Corner Lot: Το σπίτι βρίσκεται στη γωνία του δρόμου (Yes/No)
Offer Pending: Εκκρεμεί προσφορά για το σπίτι (Yes/No)
Annual Tax: Εκτιμώμενο ποσό ετήσιου φόρου ακίνητης περιουσίας

Χρησιμοποιώντας τα δεδομένα της στήλης «Price» και τις κατάλληλες συναρτήσεις του CALC, να απαντήσετε στις παρακάτω ερωτήσεις:

- Υπολογίστε τα Q_1 , Q_3 , $IQR = Q_3 - Q_1$. Στη συνέχεια, υπολογίστε τις τιμές $KOT = \max\{X_{(1)}, Q_1 - 1.5 \cdot IQR\}$, $AOT = \min\{X_{(n)}, Q_3 + 1.5 \cdot IQR\}$. Να βρεθεί το ποσοστό των σπιτιών με τιμή πώλησης μικρότερη της KOT, το αντίστοιχο ποσοστό των σπιτιών με τιμή πώλησης μεγαλύτερη της AOT και το συνολικό ποσοστό σπιτιών που ικανοποιούν τη συνθήκη $\{Τιμή Πώλησης < KOT \text{ ή } Τιμή Πώλησης > AOT\}$
- Να υπολογιστεί και να ερμηνευτεί η τιμή του συντελεστή μεταβλητότητας CV (ως προς την ομοιογένεια του δείγματος). Δίνεται ότι $CV = \frac{s}{\bar{x}}$.
- Να υπολογιστούν οι τιμές των μέτρων $\frac{1}{n} \sum |x_i - A|$, όπου το $A = \bar{X}$ (άρα θέλετε τη μέση απόλυτη απόκλιση, MAD, και μπορείτε να χρησιμοποιήσετε τη συνάρτηση AVEDEV του CALC) καθώς και η διάμεση απόλυτη απόκλιση (δηλ. η διάμεσος των αποκλίσεων $|x_i - \delta|$, $i = 1, 2, \dots, n$, όπου δ η δειγματική διάμεσος).
- Να υπολογιστεί ο 10% και ο 20% περικομμένος μέσος (χρησιμοποιήστε τη συνάρτηση =TRIMMEAN() του CALC).
- Να υπολογιστεί ο συντελεστής GINI και η μέση διαφορά GINI για τις τιμές της μεταβλητής Price.
- Να κατασκευαστεί η καμπύλη Lorenz για τις τιμές της μεταβλητής Price. Δε χρειάζεται να κάνετε τη γραφική παράσταση παρά μόνο να βρείτε τα ζεύγη τιμών που πρέπει να απεικονιστούν στο επίπεδο.
- Να βρεθεί η μέση τιμή πώλησης των σπιτιών με λιγότερα από 1,600 square feet και ηλικία το πολύ 7 έτη. Να χρησιμοποιήσετε τη συνάρτηση

=AVERAGEIFS(Average range ; Range 1; Criteria 1 Range 2; Criteria 2; ...),

όπου στο Average range δώστε το πλέγμα των κελιών από τα οποία θα προκύψει ο μέσος όρος, στο Range 1 είναι το εύρος τιμών στις οποίες θα εφαρμοστεί το Criteria 1 (δηλ. Το 1^ο κριτήριο, το οποίο είναι Square_Feet > 1600) και αντίστοιχα βάζετε και τα υπόλοιπα κριτήρια. Αν δεν υπάρχουν άλλα κριτήρια, προφανώς δε βάζετε κάτι.

Σημειώστε ότι στην περίπτωση που έχετε μόνο ένα κριτήριο (και όχι τουλάχιστον 2 που πρέπει να ικανοποιούνται), προτείνεται η χρήση της συνάρτησης =AVERAGEIF(...) (χρησιμοποιήστε το HELP του CALC για να δείτε πληροφορίες για αυτή τη συνάρτηση). Εδώ όμως έχετε και 2^ο κριτήριο, Age <=7.

8. Να βρεθεί το μοναδικό πλήθος τιμών της AGE. Η διαδικασία είναι η εξής:

- βήμα 1. Επιλέξτε το array με τις τιμές της AGE
- βήμα 2. Από το Menu, επιλέξτε Data / More Filters / Standard Filter.
- βήμα 3. Στο παράθυρο διαλόγου, στην 1^η γραμμή σχετικά Φίλτρα Κριτηρίων (Filter Criteria) δώστε “None” στο Όνομα Πεδίου (Field Name).
- βήμα 4. Πατήστε πάνω στο Options (Επιλογές) για να εμφανιστούν περισσότερες επιλογές.
- βήμα 5. Επιλέξτε το «No duplications» (όχι διπλοεγγραφές, μοναδικές τιμές)
- βήμα 6. Επιλέξτε το “Copy results to:” και δώστε τον πλήρη προορισμό κελιού στο πεδίο.
- βήμα 7. Αποεπιλέξτε το “Keep filter criteria”.
- βήμα 8. Αν το array που επιλέξατε στο 1 δεν περιλαμβάνει τίτλο, αποεπιλέξτε το “Range contains column labels”.
- βήμα 9. Αν θέλετε να υπάρχει διάκριση μεταξύ Κεφαλαίων ή όχι (εδώ δε θέλετε κάτι τέτοιο γιατί αναζητάτε μοναδικές εγγραφές σε αριθμητικά δεδομένα) επιλέξτε «Case sensitive”.
- βήμα 10. Πατάμε OK.

Σημείωση: Για όσους χρησιμοποιούν EXCEL ή Google Sheets, δοκιμάστε τη συνάρτηση =COUNTUNIQUE(...)

- 9. Να υπολογιστούν και να ερμηνευτούν οι συντελεστές ασυμμετρίας και κύρτωσης. Χρησιμοποιήστε τις αντίστοιχες συναρτήσεις =SKEW() και =KURT() του CALC.
- 10. Να υπολογιστεί ο συντελεστής ασυμμετρίας με τύπο

$$\frac{Q_3 - \delta - (\delta - Q_1)}{Q_3 - Q_1}$$

- 11. Να υπολογιστεί ο συντελεστής κύρτωσης με τύπο

$$\frac{IQR/2}{P_{90} - P_{10}}$$

όπου P_{90} , P_{10} είναι αντίστοιχα το 0.90- και 0.10-ποσοστιαίο σημείο.

Άσκηση 2^η: Έστω ότι επενδύουμε 1000€ σε ένα τραπεζικό προϊόν, το οποίο δίνει απόδοση 15% τον πρώτο χρόνο, 10% τον δεύτερο και τον τρίτο χρόνο, 4% τον τέταρτο χρόνο και 2% τον πέμπτο χρόνο. Να φτιάξετε ένα φύλλο εργασίας στο CALC ώστε να παρακολουθήσετε το πώς μεταβάλλονται οι αποδόσεις μεταξύ των 5 ετών και ποια είναι η αξία του αρχικού κεφαλαίου στο τέλος του 5^{ου} έτους. Να συγκρίνετε τα αποτελέσματά σας με την περίπτωση κάποιος (εσφαλμένα) είχε θεωρήσει ως ετήσιο επιτόκιο, το μέσο επιτόκιο των 5 ετών. Θεωρήστε ότι το παρακάτω επενδυτικό σχέδιο προτείνεται από την τράπεζα Α και συγκρίνετε με τα αντίστοιχα επενδυτικά σχέδια των τραπεζών Β, Γ και Δ. Τα ετήσια επιτόκια δίνονται στον παρακάτω πίνακα:

Έτος	Τράπεζα Α	Τράπεζα Β	Τράπεζα Γ	Τράπεζα Δ
1	15%	20%	22%	21%
2	10%	15%	12%	16%
3	10%	6%	5%	7%
4	4%	6%	5%	4%
5	2%	1%	1%	2%

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 5ο εργαστήριο. Να δώσετε το όνομα LAB5.ods. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας πριν φύγετε).