

# **ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>**

## **ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ**

### **Αριθμητικά Δεδομένα & Μέτρα Θέσης**

- Τα λογιστικά φύλλα χρησιμοποιούνται συνήθως ως αποθηκευτικά μέσα αριθμητικών δεδομένων. Έτσι, στην πράξη, όταν ανοίξει κανείς ένα φύλλο εργασίας του CALC, αντικρύζει μια εικόνα παρόμοια με την παρακάτω

The image shows a screenshot of a spreadsheet application window titled "01\_Excel\_2016-2017.xlsx - LibreOffice Calc". The window displays a grid of data with columns labeled A through J and rows numbered 1 through 37. The data consists of numerical values, likely representing financial or statistical information. The interface includes a menu bar (File, Edit, View, Insert, Format, Sheet, Data, Tools, Window, Help), a toolbar with various icons, and a status bar at the bottom showing "K12".

	A	B	C	D	E	F	G	H	I	J
1	5,14322	2,48785	4,66598	3,87729	4,31707	4,80484	5,03027	5,79127	2,62416	3,24554
2	5,56549	2,48110	1,28998	1,11941	4,70807	3,02708	3,85555	2,08592	2,00304	3,89938
3	5,87470	1,29584	2,50367	4,88542	5,75179	2,03753	5,56710	1,60485	1,91835	3,74670
4	5,72655	3,05756	2,17691	4,44176	1,28358	2,80051	5,74964	5,14469	2,61287	1,75218
5	4,84696	5,98888	4,97042	4,55341	2,89912	5,47571	5,74476	5,90852	2,51448	2,88814
6	4,25234	1,76498	5,45331	2,91454	2,65668	4,35428	5,88251	4,57555	2,72481	3,80347
7	2,06824	5,98769	2,90313	1,34264	3,30657	3,37855	3,17416	3,54935	1,03483	4,66561
8	5,88371	1,93734	3,67188	3,43564	2,63309	1,48746	4,90505	1,67955	5,94390	1,05644
9	1,98673	5,72534	3,85435	3,58684	1,75867	3,79442	1,62924	5,25964	4,95197	3,01472
10	3,14968	1,38024	3,94093	4,83470	4,41430	2,03162	2,78357	5,59966	3,37492	1,06424
11	1,70437	1,49396	2,01035	1,22942	5,96674	2,30770	2,90772	5,10833	2,71612	2,87810
12	2,70420	3,09745	4,37469	1,29419	5,44313	4,18244	3,67208	1,60721	5,08609	3,82416
13	1,60374	1,80617	2,31562	2,86185	1,62089	3,03861	1,72418	4,50421	4,59809	1,75114
14	1,18969	5,10368	4,32579	5,69516	4,72044	2,57352	2,08170	5,27800	2,25087	4,09713
15	5,61746	4,80389	4,28811	3,92248	2,34484	4,51697	4,38654	2,69521	3,79261	5,17151
16	1,43039	5,61229	3,10261	1,74610	2,86951	1,06878	2,17277	1,36161	2,98951	5,85184
17	4,16037	3,24840	1,79959	4,86723	3,56282	3,28472	5,86045	5,90953	3,95536	3,06095
18	5,69402	3,13685	5,45500	2,80333	3,15739	4,78781	5,63217	1,13880	3,03948	3,45524
19	1,81606	3,45463	2,39000	2,03955	1,09708	2,33528	1,21547	4,54342	4,97752	4,73836
20	3,67541	1,08480	2,68141	2,53189	5,02769	3,05949	1,16917	1,20145	2,13183	3,96266
21	1,72856	1,31566	1,55370	2,78397	2,42077	4,19508	2,18026	1,64019	4,34336	5,52149
22	3,42716	3,94866	3,65432	1,87921	2,31903	5,85951	5,04003	3,87602	2,98943	3,66854
23	3,52919	1,94515	4,15873	4,74090	4,47933	1,58113	4,21046	1,14134	1,66935	2,76501
24	5,61559	5,54553	2,96234	2,53902	4,61425	4,11589	5,00951	1,65470	3,70402	2,15117
25	5,79496	1,12162	1,22994	2,41082	4,70330	4,96991	4,87025	4,50385	1,84071	1,39658
26	1,30345	2,78583	4,89153	2,16462	4,73444	3,02288	3,27810	4,65730	3,56437	2,17772
27	3,04863	3,11638	2,48026	5,32644	3,16393	2,74562	2,95783	4,03109	5,27782	2,19013
28	5,81597	5,22120	1,79402	3,92232	5,23308	2,48203	1,28201	5,73098	4,59406	3,10141
29	4,09533	2,28692	5,25442	2,87427	4,58021	4,62526	1,10689	2,74918	1,92647	4,22394
30	5,55958	2,40506	2,15545	4,99370	4,55083	3,59530	1,37235	2,81254	4,46885	2,92066
31	5,08374	1,20725	4,74367	3,01804	2,38225	4,86408	4,56432	3,66996	4,57962	4,11821
32	2,14666	3,84792	5,26058	2,84018	1,39546	4,58880	4,85897	5,96120	1,57654	3,87958
33	3,91625	3,11813	1,34456	5,37524	5,13257	2,25752	4,89352	3,50345	4,69763	5,44640
34	2,47057	2,33753	1,26098	3,97879	1,84321	4,36801	3,25681	5,13232	5,99375	4,95152
35	3,24291	3,28265	3,66728	4,72545	3,50506	1,68737	2,43544	2,89653	2,65867	3,41643
36	1,67581	3,03922	1,27752	3,27990	3,84123	1,32781	3,90040	2,69777	3,62062	2,09404
37	2,60061	2,07422	4,67517	4,08538	2,43310	3,02234	3,96324	3,76005	3,89451	1,77613

Ένας σωρός (ή πλέγμα, *grid*) από δεδομένα (εδώ είναι μόλις 10 στήλες, δηλ. 10 μεταβλητές), τα οποία πρέπει να χειριστούμε κατάλληλα, προκειμένου να βγάλουμε συμπεράσματα ή/και να απαντήσουμε στα βασικά προβλήματα της έρευνάς μας, εργασίας μας κλπ

Ας ξεκινήσουμε με κάτι απλό:

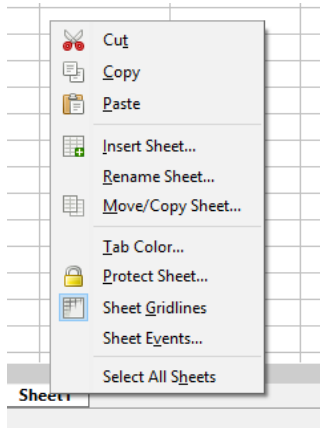
**Δραστηριότητα:** Στη διάθεσή μας, έχουμε τα παρακάτω αριθμητικά δεδομένα (αρχείο CH03\_ex01.ods):

15, 55, 70, 07, 48, 66, 69, 72, 77, 77, 40, 40, 52, 58, 58, 61, 60, 89, 39, 39, 60, 25.

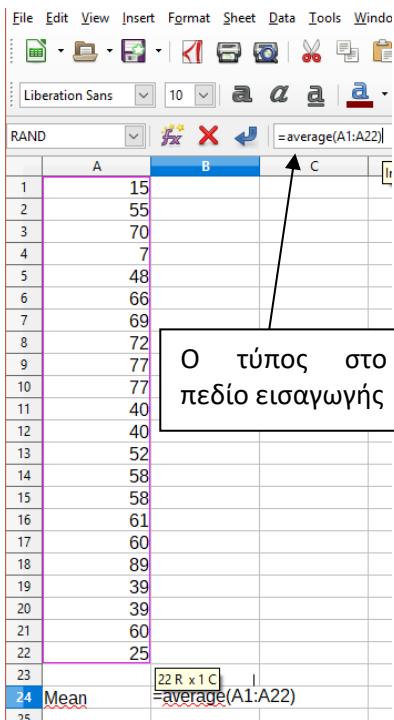
(α) Να εισάγετε τα δεδομένα αυτά στο Libre Office Calc.

(β) Να υπολογιστεί ο αριθμητικός μέσος των δεδομένων.

- Αρχικά, "ανοίγουμε" το Libre Office Calc, με έναν από τους 3 τρόπους που έχουμε ήδη αναφέρει, και θα πρέπει να βλέπουμε τη γνωστή εικόνα με τα κελιά.
- Παρατηρήστε ότι το βιβλίο εργασίας, έχει 1 φύλλο (Sheet1). Θα εισάγουμε τα δεδομένα στο 1ο φύλλο (Sheet1), το οποίο και θα μετονομάσουμε σε example1.



Για να το κάνουμε αυτό, πατάμε δεξί κλικ πάνω στο Sheet1, επιλέγουμε Μετονομασία (Rename), δίνουμε το όνομα example1 στο παράθυρο που εμφανίζεται και πατάμε Enter.



Στη συνέχεια, ξεκινάμε από το κελί A1 (πάνω αριστερά), πληκτρολογούμε τις τιμές τη μία μετά την άλλη, πατώντας το Enter (για να αλλάξουμε γραμμή) μετά από κάθε καταχώρηση. Έπειτα θα υπολογίσουμε τον αριθμητικό μέσο των τιμών που βρίσκονται στα κελιά A1 έως A22.

Χωρίς να είναι πάντοτε αυτό απαραίτητο, για λόγους καλύτερης παρουσίασης των ευρημάτων της ανάλυσής μας, εισάγουμε ως κείμενο στο κελί A24 τη λέξη Mean και πατάμε Enter.

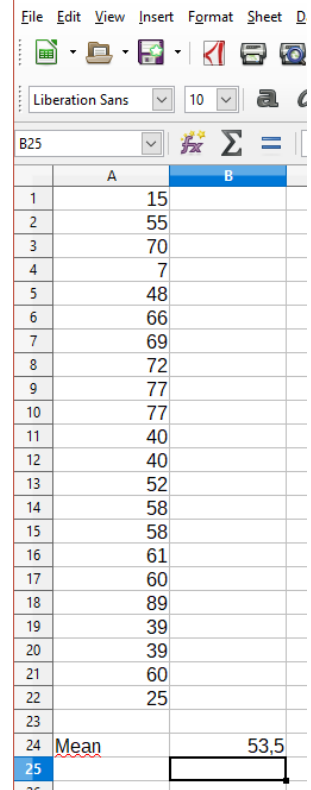
Στη συνέχεια, εισάγουμε στο κελί B24 τον παρακάτω τύπο

`=AVERAGE (A1 :A22 )`

Σύμφωνα με τη συνάρτηση που χρησιμοποιήσαμε, θέλουμε να βρούμε το μέσο όρο (*average*) των τιμών οι οποίες βρίσκονται στα κελιά A1 έως και A22 (A1:A22). Το Calc θα χρησιμοποιήσει τους αριθμούς που βρίσκονται στο "πλέγμα" (array) A1:A22 και θα βρει το μέσο όρο τους.

**Προσοχή:** Αν στο πλέγμα που δηλώσαμε, δεν είναι όλες οι τιμές αριθμητικές, τότε το Calc θα παρακάμψει αυτές τις τιμές και θα υπολογίσει τον αριθμητικό μέσο όρο, με αυτές που απομένουν.

Ο Μέσος όρος των 22 αυτών τιμών είναι 53.5. Αν οι τιμές αυτές αποτελούν τον πληθυσμό, τότε συμβολίζουμε  $\mu = 53.5$ . Αν οι τιμές αυτές είναι ένα δείγμα 22 ατόμων από τον πληθυσμό, τότε συμβολίζουμε  $\bar{x} = 53.5$ .



The screenshot shows a spreadsheet application window with the following data:

	A	B
1	15	
2	55	
3	70	
4	7	
5	48	
6	66	
7	69	
8	72	
9	77	
10	77	
11	40	
12	40	
13	52	
14	58	
15	58	
16	61	
17	60	
18	89	
19	39	
20	39	
21	60	
22	25	
23		
24	Mean	53.5
25		

- Ο τύπος υπολογισμού είναι  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (για τον αριθμητικό μέσο όρο του δείγματος, δηλ. το δειγματικό μέσο) και  $\mu = \frac{1}{N} \sum_{i=1}^N x_i$  (για τον αριθμητικό μέσο όρο του πληθυσμού, δηλ. τον πληθυσμιακό μέσο). **Το CALC δεν καταλαβαίνει τη διαφορά (εσείς όμως πρέπει να είστε σε θέση να την εντοπίζετε).**

- Στη συνέχεια, θα υπολογίσουμε το Γεωμετρικό Μέσο (Geometric Mean) και τον Αρμονικό Μέσο (Harmonic Mean). Στα κελιά A25, A26 εισάγουμε, αντίστοιχα, τις λέξεις geometric mean και harmonic mean. Στη συνέχεια, εισάγουμε στο κελί B25 τον παρακάτω τύπο

$$=GEOMEAN (A1 : A22)$$

ενώ εισάγουμε στο κελί B26 τον παρακάτω τύπο

$$=HARMEAN (A1 : A22)$$

Τα αποτελέσματα είναι 47,53685 (γεωμετρικός μέσος, *GM*) και 37,39651 (αρμονικός μέσος, *HM*).

$$GM = \sqrt[n]{x_1 x_2 \dots x_n}, \quad HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

## ΙΔΙΟΤΗΤΕΣ ΓΕΩΜΕΤΡΙΚΟΥ ΜΕΣΟΥ ΟΡΟΥ

Προφανώς, ο γεωμετρικός μέσος δεν ορίζεται αν στο δείγμα περιλαμβάνονται παρατηρήσεις με αρνητικές τιμές ή παρατηρήσεις που έχουν την τιμή μηδέν. Επίσης, αποδεικνύεται (μέσω της ανισότητας Cauchy-Schwarz) ότι  $GM \leq AM$ , όπου  $AM$  είναι ο (συνήθης) Αριθμητικός Μέσος (*Arithmetic Mean*).

**Παράδειγμα**<sup>1</sup> (για αυτοαξιολόγηση!): Να υπολογιστεί ο αριθμητικός και ο γεωμετρικός μέσος για τα παρακάτω σύνολα δεδομένων: Σύνολο  $A=\{6\ 8\ 10\ 10\ 10\ 12\ 16\}$  και Σύνολο  $B=\{6\ 8\ 10\ 10\ 10\ 12\ 48\}$ . Τι παρατηρείτε;

Ο γεωμετρικός μέσος επηρεάζεται λιγότερο από ακραίες τιμές στο δείγμα. Επιπλέον, είναι χρήσιμος όταν μεταβάλλονται μερικές μόνο από τις παρατηρήσεις μιας ακολουθίας παρατηρήσεων (επηρεάζεται λιγότερο, πιο ανθεκτικός από τον  $AM$ ). Για το λόγο αυτό

---

<sup>1</sup> Πανάρετος & Ξεκαλάκη, Εισαγωγή στη Στατιστική Σκέψη I



χρησιμοποιείται στον υπολογισμό του δείκτη *Financial Times Industrial Ordinary Share Index*<sup>2</sup>. Αξίζει πάντως να σημειωθεί ότι πλέον ο δείκτης αυτός δε χρησιμοποιείται ιδιαίτερα.

Ένας εναλλακτικός τύπος υπολογισμού είναι

$$\log GM = \frac{\sum_{i=1}^n \log X_i}{n},$$

και τελικά  $GM = 10^{\log GM}$ .

Επίσης, ο γεωμετρικός μέσος είναι χρήσιμος όταν τα δεδομένα αυξάνονται ή ελαττώνονται σύμφωνα με γεωμετρική πρόοδο.

Επίσης, μπορεί να χρησιμοποιηθεί για την εκτίμηση του συνολικού ποσού χρημάτων που επενδύεται με σύνθετο επιτόκιο (*compound interest*). Υπάρχει παράδειγμα στο τέλος του παρόντος Κεφαλαίου.

---

<sup>2</sup> [https://en.wikipedia.org/wiki/FT\\_30](https://en.wikipedia.org/wiki/FT_30)

## ΙΔΙΟΤΗΤΕΣ ΑΡΜΟΝΙΚΟΥ ΜΕΣΟΥ ΟΡΟΥ

- Αποδεικνύεται ότι<sup>3</sup>  $HM \leq GM \leq AM$ , όπου  $AM$  είναι ο (συνήθης) Αριθμητικός Μέσος (*Arithmetic Mean*),  $GM$  είναι ο Γεωμετρικός Μέσος και  $HM$  είναι ο Αρμονικός Μέσος. Συνήθως, ο Αρμονικός μέσος ( $HM$ ) χρησιμοποιείται ως ένας μέσος για ρυθμούς.
- **Παράδειγμα:** Έστω αυτοκίνητο το οποίο κινείται στη διαδρομή προς την πόλη Α με ταχύτητα 60km/h και στο ταξίδι της επιστροφής, κινείται με 40km/h. Τότε, δεν είναι σωστό να θεωρήσουμε ότι η μέση ταχύτητα κατά τη διάρκεια όλου του ταξιδιού ήταν 50km/h (δηλ. ο Αριθμητικός μέσος). Αυτή θα ήταν η μέση ταχύτητα αν το αυτοκίνητο είχε κινηθεί 1 ώρα με 60km/h και άλλη 1 ώρα με 40km/h.
- Προκειμένου να καθορίσουμε τη μέση ταχύτητα όταν έχουν χρησιμοποιηθεί διαφορετικές ταχύτητες πάνω στην ίδια απόσταση, χρησιμοποιείται ο αρμονικός μέσος. Έτσι,  $\frac{2}{\left(\frac{1}{60} + \frac{1}{40}\right)} = \frac{2}{\frac{100}{2400}} = 48\text{km/h}$ .

---

<sup>3</sup> [https://en.wikipedia.org/wiki/Geometric%E2%80%93harmonic\\_mean#Inequalities](https://en.wikipedia.org/wiki/Geometric%E2%80%93harmonic_mean#Inequalities)

- Ένα άλλο παράδειγμα είναι το εξής: Έστω ότι μια δακτυλογράφος μπορεί να δακτυλογραφήσει τιμολόγια με ρυθμό 30 την ώρα, καταστάσεις με ρυθμό 40 την ώρα και σημειώματα υπενθύμισης με ρυθμό 80 την ώρα. Έστω επίσης ότι θα πρέπει να δακτυλογραφηθούν ίσοι αριθμοί από κάθε κατηγορία των εγγράφων αυτών.
- Ο διεθυντής του γραφείου θέλει να καταγράψει τη μέση παραγωγικότητα ανά ώρα της δακτυλογράφου και υπολογίζει τον αριθμητικό μέσο, ο οποίος είναι  $(30+40+80)/3=50$  κείμενα την ώρα.
- Όμως, ο αριθμητικός μέσος, δε λαμβάνει υπόψη του ότι τα τιμολόγια χρειάζονται περισσότερο χρόνο να δακτυλογραφηθούν από ότι οι καταστάσεις και οι καταστάσεις χρειάζονται περισσότερο χρόνο απ' ότι τα σημειώματα υπενθύμισης. Έτσι, ο πραγματικός μέσος ρυθμός είναι  $\frac{3}{\left(\frac{1}{30}+\frac{1}{40}+\frac{1}{80}\right)} = \frac{3}{\frac{17}{240}} \approx 42$  κείμενα την ώρα.
- Αν  $n = 2$ , τότε  $H = G^2/A$ , όπου  $H$  είναι ο αρμονικός μέσος,  $G$  είναι ο γεωμετρικός μέσος και  $A$  είναι ο αριθμητικός μέσος (η απόδειξη αφήνεται για εξάσκηση!)

## ΆΛΛΑ ΜΕΤΡΑ ΘΕΣΗΣ

- **Υπολογισμός επικρατούσας τιμής (κορυφής):** Ως επικρατούσα τιμή  $M_0$  ορίζεται η τιμή που εμφανίζεται τις περισσότερες φορές στο δείγμα (ή αντίστοιχα, στον πληθυσμό).
- Στο παράδειγμα είναι  $M_0 = 39, 40, 58, 60$  και  $77$ , δηλαδή η κατανομή των δεδομένων είναι πολυκόρυφη (όλες αυτές οι παρατηρήσεις εμφανίζονται τον ίδιο αριθμό φορές στο δείγμα).
- Αν υπάρχει μοναδική τιμή για το  $M_0$ , τότε η κατανομή ονομάζεται μονοκόρυφη.
- Με το Calc: Στο κελί A27 εισάγουμε τη λέξη Mode και στο κελί B27 εισάγουμε τον τύπο `=MODE (A1 :A22)`.
- Το αποτέλεσμα με το Calc είναι 39. Αυτό συμβαίνει γιατί η συνάρτηση του Calc είναι έτσι φτιαγμένη ώστε να δίνει τη μικρότερη τιμή, όταν υπάρχουν περισσότερες από μια κορυφές στα δεδομένα.

**Υπολογισμός διαμέσου:** Ως διάμεσος  $\delta$  ορίζεται η μεσαία τιμή-παρατήρηση, όταν οι παρατηρήσεις είναι διατεταγμένες σε αύξουσα σειρά (δηλ. η μεσαία παρατήρηση στο διατεταγμένο δείγμα). Αν το μέγεθος του δείγματος είναι άρτιος (ζυγός) αριθμός, τότε η διάμεσος είναι η μέση τιμή των δύο μεσαίων παρατηρήσεων.

**Ορισμός:** Έστω δείγμα  $x_1, x_2, \dots, x_n$ . Τότε, αν  $n = 2k$ ,  $k \in \mathbb{Z}$ , η διάμεσος  $\delta = (x_{(k)} + x_{(k+1)})/2$  ενώ αν  $n = 2k + 1$ ,  $k \in \mathbb{Z}$ , η διάμεσος  $\delta = x_{(k+1)}$ .

**Στο παράδειγμα:** Το διατεταγμένο δείγμα είναι

7, 15, 25, 39, 39, 40, 40, 48, 52, 55, 58, 58, 60, 60, 61, 66, 69, 70, 72, 77, 77, 89

Προφανώς,  $n = 22$ , δηλ.  $k = 11$ , άρα αφού το πλήθος άρτιος αριθμός, τότε  $\delta = (x_{(11)} + x_{(12)})/2 = (58 + 58)/2 = 58$ . **Με το Calc:** Στο κελί A28 εισάγουμε τη λέξη Median και στο κελί B28 εισάγουμε τον τύπο =MEDIAN (A1 : A22) . Το αποτέλεσμα είναι 58.

Μπορούμε να δώσουμε και έναν "πιο προσωπικό τόνο" στο φύλλο εργασίας μας, μορφοποιώντας τα δεδομένα στο φύλλο εργασίας. Εδώ χρησιμοποιήθηκε έντονη γραμματοσειρά (**bold**) και ως φόντο μια απόχρωση του γκρι.

**Υπολογισμός τεταρτημορίων:** Το πρώτο τεταρτημόριο (συμβ.  $Q_1$ ) είναι η τιμή εκείνη για την οποία τουλάχιστον 25% των δεδομένων βρίσκονται πριν από αυτή και τουλάχιστον 75% των δεδομένων βρίσκονται μετά από αυτή.

- Το δεύτερο τεταρτημόριο (συμβ.  $Q_2$ ) είναι η τιμή εκείνη για την οποία τουλάχιστον 50% των δεδομένων βρίσκονται πριν από αυτή την τιμή και τουλάχιστον 50% των δεδομένων βρίσκονται μετά από αυτή. Το  $Q_2$  συμπίπτει με τη διάμεσο  $\delta$ .
- Το τρίτο τεταρτημόριο (συμβ.  $Q_3$ ) είναι η τιμή εκείνη για την οποία τουλάχιστον 75% των δεδομένων βρίσκονται πριν από αυτή και τουλάχιστον 25% των δεδομένων βρίσκονται μετά από αυτή.

## ΥΠΟΛΟΓΙΣΜΟΣ ΠΟΣΟΣΤΙΑΙΩΝ ΣΗΜΕΙΩΝ (ή ΠΟΣΟΣΤΗΜΟΡΙΩΝ)

**Ορισμός:** Έστω  $0 \leq \alpha \leq 1$ . Το  $100 \cdot \alpha$ -οστο ποσοστημόριο ( $100\alpha$  percentile) είναι η τιμή εκείνη για την οποία τουλάχιστον  $100 \cdot \alpha\%$  των δεδομένων βρίσκονται πριν από την τιμή αυτή και τουλάχιστον  $100 \cdot (1 - \alpha)\%$  των δεδομένων βρίσκονται μετά από αυτή.

- **Στο παράδειγμα:** Για το  $Q_2$ , που είναι το 0,50-ποσοστιαίο σημείο ή 50° εκατοστημόριο, έχουμε ήδη δει ότι είναι ίσο με τη διάμεσο, δηλ.  $Q_2=58$ . Για το  $Q_1$  (ή 0,25-ποσοστιαίο σημείο ή 25° εκατοστημόριο) θέλουμε τουλάχιστον 25% των παρατηρήσεων. Άρα,  $0.25 \cdot 22 = 5.5$ , δηλ. πρέπει να έχουμε τουλάχιστον 5.5 παρατηρήσεις μικρότερες από αυτή την τιμή  $Q_1$ . Επειδή ο αριθμός των παρατηρήσεων είναι ακέραιος, θα πάρουμε ως  $Q_1$  την 6η διατεταγμένη παρατήρηση. Οπότε, από το διατεταγμένο δείγμα, είναι  $Q_1=x_{(6)}=40$ .
- Όμοια, για το  $Q_3$  (ή 0,75-ποσοστιαίο σημείο ή 75° εκατοστημόριο), θέλουμε τουλάχιστον 75% των παρατηρήσεων. Άρα,  $0.75 \cdot 22 = 16.5$ , δηλ. πρέπει να έχουμε τουλάχιστον 16.5 παρατηρήσεις μικρότερες από αυτή την τιμή  $Q_3$ . Επειδή ο αριθμός

των παρατηρήσεων είναι ακέραιος, θα πάρουμε ως  $Q_3$  τη 17η διατεταγμένη παρατήρηση. Οπότε, από το διατεταγμένο δείγμα, είναι  $Q_3=x_{(17)}=69$ .

- Χρησιμοποιώντας το Calc: Αρχικά, γράφουμε τις λέξεις Q1, Q3 στα κελιά A29 και A30, αντίστοιχα. Στη συνέχεια, εισάγουμε στα κελιά B29 και B30 τους τύπους

=QUARTILE (A1 :A22 ; 1) και =QUARTILE (A1 :A22 ; 3)

( ή ισοδύναμα =PERCENTILE (A1 :A22 ; 0.25) και

=PERCENTILE (A1 :A22 ; 0.75) )

και πατάμε Enter.

- Το αποτέλεσμα είναι:  $Q_1=40$ ,  $Q_3=68.25$ . Τι έχει συμβεί και το Calc έδωσε διαφορετικά αποτελέσματα;
- **Επεξήγηση:** Ο αλγόριθμος υπολογισμού χρησιμοποιεί γραμμική παρεμβολή μεταξύ της  $x_{(k)}$  και της  $x_{(k+1)}$  διατεταγμένης παρατήρησης. Ειδικότερα<sup>4</sup>

---

<sup>4</sup> <https://www.ischool.utexas.edu/~wyllys/IRLISMaterials/quartiles.pdf>



- 1 Find the kth smallest member in the array of values, where:  
 $k = (\text{quart}/4) * (n-1) + 1$   
 If k is not an integer, truncate it but store the fractional portion (f) for use in step 3.  
 n = number of values in the array  
 quart = value between 0 and 4 depending on which quartile you want to find
- 2 Find the smallest data point in the array of values that is greater than the kth smallest -- the (k+1)th smallest member.
- 3 Interpolate between the kth smallest and the (k+1)th smallest values:  
 Output =  $a[k] + (f * (a[k+1] - a[k]))$   
 $a[k]$  = the kth smallest  
 $a[k+1]$  = the k+1th smallest

- Στο παράδειγμα, είναι  $\text{quart} = 3$ ,  $n = 22$ ,  $a[k] = x_{(k)}$  (k-οστη διατεταγμένη παρατήρηση),  $a[k+1] = x_{(k+1)}$  ((k+1)-οστη διατεταγμένη παρατήρηση),  $f$  = το δεκαδικό μέρος του αριθμού  $k = \left(\frac{\text{quart}}{4}\right) \cdot (n - 1) + 1$ .

**Σύνταξη της συνάρτησης QUARTILE:** =QUARTILE (array; k)

- Στο array, δίνουμε το "πλέγμα" (ή το σύνολο των αριθμών) για τους οποίους θέλουμε να βρούμε το **τεταρτημόριο**.
- Το  $k$  λαμβάνει τιμές στο σύνολο  $\{0, 1, 2, 3, 4\}$ , όπου 0 = ελάχιστη παρατήρηση, 1 = το 1ο τεταρτημόριο, 2 = το 2ο τεταρτημόριο (δηλ. η διάμεσος), 3 = το 3ο τεταρτημόριο και 4 = μέγιστη παρατήρηση.

**Σύνταξη της συνάρτησης PERCENTILE:** =PERCENTILE (array; k)

- Στο array, δίνουμε το "πλέγμα" (ή το σύνολο των αριθμών) για τους οποίους θέλουμε να βρούμε το **100·k-εκατοστημόριο ή ισοδύναμα το k-ποσοστιαίο σημείο**
- Το  $k \in [0,1]$ , για  $k=0$  θα πάρουμε την ελάχιστη παρατήρηση και για  $k=1$  θα πάρουμε τη μέγιστη παρατήρηση.

**Εφαρμογή:** Να υπολογιστούν τα (i) 5ο εκατοστημόριο (ή 0.05-ποσοστιαίο σημείο), (ii) 10ο εκατοστημόριο (ή 0.10-ποσοστιαίο σημείο), (iii) 20ο εκατοστημόριο (ή 0.20-ποσοστιαίο σημείο), (iv) 80ο εκατοστημόριο (ή 0.80-ποσοστιαίο σημείο), (v) 90ο εκατοστημόριο (ή 0.90-ποσοστιαίο σημείο) και (vi) 95ο εκατοστημόριο (ή 0.95-ποσοστιαίο σημείο).

**Λύση:** Γράφουμε την εντολή =PERCENTILE (A1:A22;k) σε διαφορετικά κελιά, όπου  $k = 0.05, 0.10, 0.20, 0.80, 0.90, 0.95$ . Η απάντηση φαίνεται στην παρακάτω εικόνα:

P5=	15,5
P10=	26,4
P20=	39,2
P80=	69,8
P90=	76,5
P95=	77

**Συμβολισμοί:** Για τα Τεταρτημόρια (*Quartiles*) είναι **Q1**: 1<sup>ο</sup> τεταρτημόριο, **Q2** (ή **δ**): 2<sup>ο</sup> τεταρτημόριο (διάμεσος), **Q3**: 3<sup>ο</sup> τεταρτημόριο. Για τα Ποσοστιαία σημεία (*Percentiles*): **Pa**,

$a \in \{1, 2, \dots, 99\}$ . Προφανώς,  $P_{25}=Q_1$ ,  $P_{50}=Q_2=\delta$ ,  $P_{75}=Q_3$ .

## ΥΠΟΛΟΓΙΣΜΟΣ ΤΑΞΗΣ (ή ΒΑΘΜΟΥ, rank)

**Ορισμός:** Ως τάξη (*rank*) μιας παρατήρησης, ορίζουμε τη σειρά που έχει η τιμή της παρατήρησης σε ένα σύνολο αριθμών. Δηλαδή, πόσες παρατηρήσεις είναι μικρότερες ή ίσες της τιμής της παρατήρησης αυτής.

**Παράδειγμα:** Έστω το σύνολο δεδομένων 0, 9, 4, 5, 2, 8, 12, 3. Τότε, οι τάξεις για κάθε μια από τις παρατηρήσεις είναι αντίστοιχα, 1, 7, 4, 5, 2, 6, 8, 3.

Αν υπάρχουν παρατηρήσεις οι οποίες εμφανίζονται περισσότερες από μια φορά στο σύνολο των δεδομένων (δηλ. υπάρχουν δεσμοί, *ties*), τότε ως τάξη σε κάθεμιά από αυτές δίνουμε το μέσο όρο των τάξεων (*midrank*) που θα δίναμε αν ήταν διαφορετικές.

**Παράδειγμα:** Έστω το σύνολο δεδομένων 1, 7, 6, 5, 2, 6, 4, 5,9 ,7, 3, 2, 7, 0, 4, 9, 11, 10, 3.

Αρχικά φτιάχνουμε το διατεταγμένο δείγμα

$i$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$x_{(i)}$	0	1	2	2	3	3	4	4	5	5	6	6	7	7	7	9	9	10	11
Rank	1	2	3.5	3.5	5.5	5.5	7.5	7.5	9.5	9.5	11.5	11.5	14	14	14	16.5	16.5	18	19

- Στο Calc, χρησιμοποιούμε τη συνάρτηση RANK. Για το παράδειγμα με τις 22 παρατηρήσεις, μπορούμε να κάνουμε το εξής ώστε να βρούμε την τάξη για κάθε μια από τις διαθέσιμες παρατηρήσεις.
- Γράφουμε στο κελί B1 τον τύπο =RANK(A1;\$A\$1:\$A\$22;1). Αν πατήσουμε Enter, το Calc δίνει την τιμή 2, το οποίο σημαίνει ότι στο σύνολο των δεδομένων που βρίσκεται στα κελιά A1:A22, η τιμή που βρίσκεται στο A1 (δηλ. το 15), έχει τάξη ίση με 2 (δηλ. στο δείγμα υπάρχουν 2 παρατηρήσεις μικρότερες ή ίσες από το 15).

- Αν είχαμε γράψει =RANK(A1;\$A\$1:\$A\$22;0), τότε η απάντηση θα ήταν 21. Αυτό συμβαίνει γιατί στην περίπτωση που το 3<sup>ο</sup> όρισμα είναι **ίσο με 1**, το δείγμα έχει διαταχθεί από τη μικρότερη προς τη μεγαλύτερη παρατήρηση (όπως το έχουμε ορίσει έως τώρα) ενώ αν είναι ίσο με 0, το δείγμα έχει διαταχθεί από τη μεγαλύτερη προς τη μικρότερη παρατήρηση (και σε αυτή την περίπτωση η τάξη αυτής της παρατήρησης είναι 21). Γενικά, εκτός αν αναφέρεται ξεκάθαρα, θα χρησιμοποιούμε τον ορισμό της τάξης **στο διατεταγμένο δείγμα κατά αύξουσα σειρά**.

**Σημαντικό:** Η συνάρτηση RANK όταν βρει δεσμούς (ties) δίνει και σε κάθε παρατήρηση το μικρότερο βαθμό και όχι το μέσο όρο. Αν θέλουμε να εργαστούμε με τον ορισμό που έχουμε δώσει για τους βαθμούς και να κάνουμε ακριβώς ότι έχουμε πει σε περίπτωση δεσμών, πρέπει να χρησιμοποιήσουμε τη συνάρτηση RANK.AVG. Η σύνταξή της είναι ακριβώς η ίδια με την (απλή) RANK.

- Όταν γράφουμε  $\$A\$1$ , σημαίνει πως η τιμή του κελιού A1 έχει "κλειδώσει". Όμοια, το  $\$A\$1:\$A\$22$  σημαίνει ότι έχει "κλειδώσει" το πλέγμα των τιμών μεταξύ αυτών των δύο κελιών.
- Αυτό διευκολύνει σε πολλές περιπτώσεις τον υπολογισμό των ζητούμενων ποσοτήτων. **Ο λόγος;** Μπορούμε να χρησιμοποιήσουμε την αυτόματη συμπλήρωση του Calc (αφού πρώτα έχουμε "κλειδώσει" την περιοχή των δεδομένων μας) και να υπολογίσουμε για κάθε μια από τις τιμές του πλέγματος, την τάξη του.
- **Πώς γίνεται η αυτόματη συμπλήρωση;** Επιλέγουμε ένα κελί (άρα γύρω του υπάρχει ένα μαύρο πλαίσιο) και τοποθετούμε τον κέρσορα στην κάτω δεξιά γωνία.
- Τότε, ο κέρσορας μετατρέπεται σε σταυρό + και μπορούμε να "σύρουμε" το πλαίσιο του κελιού, συμπληρώνοντας τα κελιά που επιθυμούμε (είτε προς τα κάτω, είτε προς τα δεξιά).

	A	B	C	D
1	15	7	2	21
2	55	15	10	
3	70	25	18	
4	7	39	1	
5	48	39	8	
6	66	40	16	
7	69	40	17	
8	72	48	19	
9	77	52	20	
10	77	55	20	
11	40	58	6	
12	40	58	6	
13	52	60	9	
14	58	60	11	
15	58	61	11	
16	61	66	15	
17	60	69	13	
18	89	70	22	
19	39	72	4	
20	39	77	4	
21	60	77	13	
22	25	89	3	

Αυτόματη συμπλήρωση  
με τη συνάρτηση Rank

**Τι πετυχαίνω;** Τα κελιά συμπληρώνονται με τη συνάρτηση (ή τον τύπο) και υπολογίζεται και η τιμή σύμφωνα με τη αυτή.

**Γιατί χρειάζεται το κλείδωμα;** Αν δεν το κάνουμε, καθώς θα προχωρεί προς τα κάτω η αυτόματη συμπλήρωση, εκτός από το πρώτο όρισμα (του οποίου η τιμή θέλουμε να αλλάζει σύμφωνα με τις τιμές στα κελιά από A1 έως και A22), θα αλλάζει και το "πλέγμα" για το σύνολο δεδομένων. Άρα, θα λαμβάνουμε λανθασμένα αποτελέσματα. Κλειδώνοντας την περιοχή των δεδομένων, το μόνο που αλλάζει είναι η τιμή της παρατήρησης (καθώς κατεβαίνουμε προς τα κάτω) και υπολογίζεται η τάξη της για το σύνολο δεδομένων μας.



## ΣΤΑΘΜΙΣΜΕΝΟΣ ΜΕΣΟΣ ΟΡΟΣ

**Ορισμός:** Ο σταθμισμένος μέσος όρος (*weighted mean* ή *weighted average*) των  $n$  παρατηρήσεων  $x_1, x_2, \dots, x_n$  δίνεται από τον τύπο

$$x_W = \sum_{i=1}^n w_i x_i,$$

όπου οι τιμές  $w_1, w_2, \dots, w_n$  ονομάζονται "βάρη" (*weights*) και ικανοποιούν τη σχέση  $\sum_{i=1}^n w_i = 1$ .

**Παράδειγμα:** Ας υποθέσουμε ότι ένα τεστ ξένης γλώσσας, αποτελείται από 5 επιμέρους διαδικασίες: (I) έκθεση, (II) κατανόηση κειμένου, (III) γραμματική, (IV) προφορικά και (V) κατανόηση ακουστικού κειμένου. Κάθε διαδικασία βαθμολογείται στην κλίμακα 0-100 και η βαρύτητά τους είναι αντίστοιχα 0.25, 0.20, 0.25, 0.15, 0.15. Να υπολογιστούν οι μέσες βαθμολογίες δύο εξεταζομένων (σταθμισμένες και αστάθμιστες), οι οποίοι έγραψαν:

*Εξεταζόμενος 1:* 70/100, 75/100, 82/100, 61/100, 65/100

*Εξεταζόμενος 2:* 77/100, 65/100, 92/100, 51/100, 68/100

**Λύση** (με το Calc, αρχείο CH03\_ex02) Σε ένα νέο φύλλο εργασίας, φτιάχνουμε τις παρακάτω στήλες:

- Weights, εκεί θα γράψουμε τα βάρη 0.25, 0.2, 0.25, 0.15, 0.15<sup>5</sup>
- Grades A, εκεί θα γράψουμε τους βαθμούς για τον 1ο εξεταζόμενο
- Prod A, εκεί θα υπολογίζουμε τα γινόμενα  $w_i x_i$ ,  $i = 1,2,3,4,5$ , για τον 1ο εξεταζόμενο
- Grades B, εκεί θα γράψουμε τους βαθμούς για το 2ο εξεταζόμενο
- Prod B, εκεί θα υπολογίζουμε τα γινόμενα  $w_i x_i$ ,  $i = 1,2,3,4,5$ , για το 2ο εξεταζόμενο

	A	B	C	D	E	
1	<u>Weights</u>	<u>Grades A</u>	<u>Prod A</u>	<u>Grades B</u>	<u>Prod B</u>	
2	0,25	70		77		
3	0,20	75		65		
4	0,25	82		92		
5	0,15	61		51		
6	0,15	65		68		
7						

- Για να βρούμε τις τιμές στη στήλη Prod A, θα πρέπει να βρούμε τα γινόμενα A2\*B2,

---

<sup>5</sup> Κατά την εισαγωγή των τιμών στο Calc, ως υποδιαστολή έχει χρησιμοποιηθεί το «,».

$A3*B3, \dots A6*B6$ . Αυτό μπορεί να γίνει γράφοντας στο κελί C2 τον τύπο  $=A2*B2$  και στη συνέχεια, χρησιμοποιώντας την αυτόματη συμπλήρωση, γράφουμε τον ίδιο τύπο στα κελιά C3 μέχρι και C6. Επειδή δεν έχουμε κλειδώσει κάποια από τις τιμές, καθώς θα κατεβαίνουν τα κελιά, οι όροι του γινομένου θα μεταβάλλονται σύμφωνα με τις τιμές στα κελια των στηλών A και B.

- Με τον ίδιο ακριβώς τρόπο, για να βρούμε τις τιμές στη στήλη Prod B, θα πρέπει να βρούμε τα γινόμενα  $A2*D2, A3*D3, \dots A6*D6$ . Αυτό μπορεί να γίνει γράφοντας στο κελί E2 τον τύπο  $=A2*D2$  και στη συνέχεια, χρησιμοποιώντας την αυτόματη συμπλήρωση, γράφουμε τον ίδιο τύπο στα κελιά E3 μέχρι και E6. Αφού το κάνουμε αυτό, η εικόνα που έχουμε είναι:

	A	B	C	D	E	
1	<u>Weights</u>	<u>Grades A</u>	<u>Prod A</u>	<u>Grades B</u>	<u>Prod B</u>	
2	0,25	70	17,5	77	19,25	
3	0,20	75	15	65	13	
4	0,25	82	20,5	92	23	
5	0,15	61	9,15	51	7,65	
6	0,15	65	9,75	68	10,2	
7						

Στη συνέχεια, θα βρούμε τους σταθμισμένους και αστάθμιστους μέσους όρους, για κάθε έναν από τους εξεταζόμενους.

Χρησιμοποιώντας τη συνάρτηση AVERAGE βρίσκουμε το μέσο όρο των τιμών στη στήλη B (για τον 1ο) και τη στήλη D (για το 2ο). Οι τιμές δίνονται στα κελιά B7 και D7, αντίστοιχα. Όμοια, αφού έχουμε βρει τα επιμέρους γινόμενα  $w_i x_i$ , αρκεί να τα αθροίσουμε. Αυτό μπορεί να γίνει με τη συνάρτηση SUM. Οι τιμές δίνονται στα κελιά C7 (για τον 1ο) και E7 (για το 2ο), αντίστοιχα. Οι τύποι είναι =SUM(B2 : B6) και =SUM(E2 : E6) .

	A	B	C	D	E
1	<b>Weights</b>	<b>Grades A</b>	<b>Prod A</b>	<b>Grades B</b>	<b>Prod B</b>
2	0,25	70	17,5	77	19,25
3	0,20	75	15	65	13
4	0,25	82	20,5	92	23
5	0,15	61	9,15	51	7,65
6	0,15	65	9,75	68	10,2
7		70,6	71,9	70,6	73,1

Παρατηρούμε ότι οι δύο εξεταζόμενοι έχουν γράψει κατά μέσο όρο την ίδια (αστάθμιστη) βαθμολογία (70,6) αλλά καλύτερη επίδοση έχει ο 2ος, με 73,1 έναντι 71,9 του 1ου. Αυτό

συμβαίνει λόγω της βαρύτητας κάθε διαδικασίας (η οποία πρέπει να ληφθεί υπόψη). Ο 2ος έχει γράψει αρκετά υψηλότερους βαθμούς στις 2 διαδικασίες με την υψηλότερη βαρύτητα και γι'αυτό καταφέρνει και συγκεντρώνει καλύτερη επίδοση.

**Εναλλακτικά:** Μπορούμε να χρησιμοποιήσουμε τη συνάρτηση SUMPRODUCT, η οποία υπολογίζει το άθροισμα των γινομένων των επιμέρους όρων για δύο σύνολα δεδομένων. Ειδικότερα, αν έχετε δύο διανύσματα  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , με τη βοήθεια της SUMPRODUCT μπορείτε να υπολογίσετε το εσωτερικό γινόμενο των  $\mathbf{x}$ ,  $\mathbf{y}$ , δηλ. το  $\sum_{i=1}^n x_i y_i$ .

Στο παραπάνω παράδειγμα, για το σταθμικό μέσο όρο του 1ου εξεταζόμενου, γράφουμε στο κελί C8 τον τύπο =SUMPRODUCT (A2:A6;B2:B6). Το Calc θα υπολογίσει τα γινόμενα μεταξύ των αντίστοιχων κελιών στις δύο στήλες και θα μας δώσει το άθροισμα αυτών. Με τον ίδιο τρόπο, για το σταθμικό μέσο όρο του 2ου εξεταζόμενου, γράφουμε στο κελί E8 τον τύπο =SUMPRODUCT (A2:A6;D2:D6).

## ΠΕΡΙΚΟΜΜΕΝΟΣ ΜΕΣΟΣ ΟΡΟΣ (TRIMMED MEAN)

**Ορισμός:** Ο  $k\%$  περικομμένος (ή ξακρισμένος) μέσος  $n$  παρατηρήσεων  $x_1, x_2, \dots, x_n$  υπολογίζεται ως ο μέσος όρος των παρατηρήσεων που απομένουν αν αφαιρεθούν οι  $k/2\%$  μικρότερες και οι  $k/2\%$  μεγαλύτερες παρατηρήσεις. Συνήθως  $k \in \{5, 10, 20\}$ .

Η χρήση του περικομμένου μέσου ενδείκνυται σε δεδομένα αγώνων ή βαθμολογιών, όπου αφαιρούνται οι μεγαλύτερες και οι μικρότερες βαθμολογίες (επιδόσεις), ώστε οι ακραίες αυτές τιμές να μη λαμβάνονται υπόψη στο μέσο όρο βαθμολογίας για τους διαγωνιζόμενους.

Στο Calc, μπορούμε να υπολογίσουμε τον 10% περικομμένο μέσο (δηλ.  $k = 10$ ) χρησιμοποιώντας την εντολή `=TRIMMEAN(A1:A22; 0, 10)`. Το αποτέλεσμα είναι 54,05.

**Παράδειγμα:** Έστω το παρακάτω σύνολο  $n=30$  αριθμητικών τιμών

2 5 11 6 4 7 3 3 5 8 9 10 14 19 15 15 13 6 21 18 17 33 22 25 16 39 47 2 12 14

Να υπολογιστεί με χρήση του CALC ο περικομμένος μέσος 5%, 10% και 20% των παραπάνω δεδομένων.

Αρκεί να εισάγουμε τις παραπάνω τιμές σε ένα πεδίο του CALC (π.χ. B2:B31) και στη συνέχεια με τις εντολές =TRIMMEAN(B2:B31;0,05), =TRIMMEAN(B2:B31;0,10), =TRIMMEAN(B2:B31;0,20) να υπολογίσουμε τα ζητούμενα μέτρα. Οι τιμές τους είναι 14.0333, 13.2857, 12.2917 ενώ ο 5% περικομμένος μέσος ταυτίζεται με το συνήθη αριθμητικό μέσο όρο.

Για να καταλάβουμε τι ακριβώς κάνει το CALC για τον υπολογισμό του περικομμένου μέσου, θα πρέπει αρχικά να τοποθετήσουμε τις τιμές σε αύξουσα τάξη μεγέθους,

2 2 3 3 4 5 5 6 6 7 8 9 10 11 12 13 14 14 15 15 16 17 18 19 21 22 25 33 39 47

Στη συνέχεια, υπολογίζουμε την τιμή  $\left\lfloor \left(\frac{k}{2}\right) \cdot n \right\rfloor$  όπου  $\lfloor x \rfloor$  είναι το ακέραιο μέρος ενός αριθμού  $x$  και το  $k$  είναι η αναλογία των τιμών που θέλουμε να εξαιρέσουμε. Δηλαδή, για

τον 5% περικομμένο μέσο, το  $k = 0.05$ . Άρα βρίσκουμε τις τιμές

$$5\% \text{ περικομμένος: } \left\lfloor \left( \frac{0.05}{2} \right) \cdot 30 \right\rfloor = \lfloor 0.75 \rfloor = 0$$

$$10\% \text{ περικομμένος: } \left\lfloor \left( \frac{0.10}{2} \right) \cdot 30 \right\rfloor = \lfloor 1.5 \rfloor = 1$$

$$20\% \text{ περικομμένος: } \left\lfloor \left( \frac{0.20}{2} \right) \cdot 30 \right\rfloor = \lfloor 3 \rfloor = 3$$

Στη συνέχεια, στο διατεταγμένο δείγμα, εξαιρούμε πλήθος τιμών ίσο με την τιμή που προέκυψε παραπάνω, από τα 2 άκρα του δείγματος. Δηλαδή

**5% περικομμένος μέσος:** Η τιμή προέκυψε 0, δεν εξαιρώ τιμές άρα ο 5% περικομμένος μέσος υπολογίζεται και από τις 30 διαθέσιμες μετρήσεις (γι' αυτό και ταυτίζεται με το συνήθη αριθμητικό μέσο).

2 2 3 3 4 5 5 6 6 7 8 9 10 11 12 13 14 14 15 15 16 17 18 19 21 22 25 33 39 47

**10% περικομμένος μέσος:** Η τιμή προέκυψε 1, εξαιρώ την 1<sup>η</sup> και την τελευταία τιμή στο διατεταγμένο δείγμα και άρα ο 10% περικομμένος μέσος υπολογίζεται από τις υπόλοιπες 28 διαθέσιμες μετρήσεις.



2 2 3 3 4 5 5 6 6 7 8 9 10 11 12 13 14 14 15 15 16 17 18 19 21 22 25 33 39 47

Υπολογίστε το μέσο όρο των 28 τιμών μέσα στο κόκκινο κουτί και επιβεβαιώστε την τιμή 13.2857. Είναι ο ζητούμενος 10% περικομμένος μέσος.

**20% περικομμένος μέσος:** Η τιμή προέκυψε 3, εξαιρώ τις 3 πρώτες και τις 3 τελευταίες τιμές στο διατεταγμένο δείγμα και άρα ο 20% περικομμένος μέσος υπολογίζεται από τις υπόλοιπες 24 διαθέσιμες μετρήσεις.

2 2 3 3 4 5 5 6 6 7 8 9 10 11 12 13 14 14 15 15 16 17 18 19 21 22 25 33 39 47

Υπολογίστε το μέσο όρο των 24 τιμών μέσα στο κόκκινο κουτί και επιβεβαιώστε την τιμή 12.1917. Είναι ο ζητούμενος 20% περικομμένος μέσος.

## ΕΦΑΡΜΟΓΗ (Γεωμετρικός Μέσος)

**Παράδειγμα<sup>6</sup>:** Έστω ότι επενδύουμε 1000€ σε ένα τραπεζικό προϊόν, το οποίο δίνει απόδοση 20% τον πρώτο χρόνο, 6% το δεύτερο χρόνο και 1% τον τρίτο χρόνο. Αν θέλουμε να δούμε πόση είναι η ετήσια απόδοση, είναι λάθος να υπολογίσουμε τον αριθμητικό μέσο  $(20+6+1)/3=9$ , δηλ. 9%. Σε αυτή την περίπτωση, η απόδοση στο τέλος του 3<sup>ου</sup> έτους θα ήταν  $1000*(1+0.09)^3 = 1295.03€$

**Υπενθύμιση:** Τύπος Ανατοκισμού

$$P_n = P(1 + \tau)^n$$

όπου  $P$ : το αρχικό κεφάλαιο (εδώ 1000€),  $n$ : το πλήθος των περιόδων (εδώ είναι  $n = 3$  έτη),  $\tau$ : το ετήσιο επιτόκιο και  $P_n$  η αξία του αρχικού κεφαλαίου στο τέλος της  $n$ -οστης περιόδου.

Χρησιμοποιώντας τον τύπο για την περίπτωση σύνθετου επιτοκίου (*compound interest*), η συνολική αξία του αρχικού κεφαλαίου στο τέλος του 3<sup>ου</sup> έτους είναι

---

<sup>6</sup> [https://www.financeformulas.net/Geometric\\_Mean\\_Return.html](https://www.financeformulas.net/Geometric_Mean_Return.html)

$$1.20*1.06*1.01*1000=1284.75,$$

το οποίο διαφέρει σημαντικά της περίπτωση που το συνολικό ποσό είχε υπολογιστεί με χρήση του αριθμητικού μέσου των αποδόσεων.

Για τον υπολογισμό της ετήσιας απόδοσης, χρησιμοποιούμε τον τύπο

$$GR = \sqrt[n]{(1 + r_1) \times (1 + r_2) \times \dots (1 + r_n)} - 1,$$

όπου  $r_i$  είναι η απόδοση την  $i$ -οστή περίοδο και  $n$  είναι το πλήθος των περιόδων. Έτσι, δεν είναι δύσκολο να επαληθεύσουμε ότι

$$GR = \sqrt[3]{(1 + 0.20) \times (1 + 0.06) \times \dots (1 + 0.01)} - 1 = 0.0871, \text{ δηλ.}$$

$$\text{και άρα } 1000*(1+0.0871)^3 = 1284.72.$$

Στη συνέχεια θα δούμε πώς μπορούμε να δουλέψουμε με το CALC. Αυτό που έχει ενδιαφέρον είναι να φτιάξουμε το φύλλο εργασίας ώστε να βλέπουμε τη μεταβολή των αποδόσεων στο τέλος του 3<sup>ου</sup> έτους, για διαφορετικές τιμές αρχικού κεφαλαίου (στο παράδειγμα είναι 1000€) αλλά και διαφορετικές αποδόσεις σε καθένα από τα 3 πρώτα έτη.

## ΛΥΣΗ ΜΕ ΤΟ CALC (στο εργαστήριο):

01\_example\_Geometric\_mean (compound interest).xlsx - LibreOffice Calc

File Edit View Insert Format Sheet Data Tools Window Help

	A	B	C	D	E	F	G
1	Αρχικό Κεφάλαιο (Επένδυση)	1.000,00 €		Αποδόσεις	Συνολικό Κεφάλαιο (Αξία Επένδυσης)		
2	Επιτόκιο 1ου χρόνου	20,00%	1ου έτους	200,00 €	1.200,00 €		
3	Επιτόκιο 2ου χρόνου	6,00%	2ου έτους	72,00 €	1.272,00 €		
4	Επιτόκιο 3ου χρόνου	1,00%	3ου έτους	12,72 €	1.284,7200 €		
5	Αριθμητική Μέση Τιμή επιτοκίου 3 ετ	9,00%	1ου έτους	90,00 €	1.090,00 €		
6			2ου έτους	98,10 €	1.188,10 €		
7			3ου έτους	106,93 €	1.295,03 €	Τύπος Ανατοκισμού	1.295,03 €
8	Υπολογισμός Ετήσιας Απόδοσης		1ου έτους	120,00%			
9			2ου έτους	106,00%			
10			3ου έτους	101,00%			
11			GR	8,71%	1.284,7200 €		
12			Geomean	108,71%			
13							

Αρχικά εισάγουμε στα κελιά A1 έως A4 τους τίτλους «Αρχικό Κεφάλαιο (Επένδυση)», «Επιτόκιο 1<sup>ου</sup> χρόνου», «Επιτόκιο 2<sup>ου</sup> χρόνου», «Επιτόκιο 3<sup>ου</sup> χρόνου». Στα B1 έως B4 εισάγουμε τις τιμές 1000 (σε ευρώ, το αρχικό κεφάλαιο), 0,2 (επιτόκιο 1<sup>ου</sup> χρόνου), 0,06 (2<sup>ου</sup> χρόνου), 0,01 (3<sup>ου</sup> χρόνου). Η απόδοση στο τέλος του 1<sup>ου</sup> έτους είναι  $1000 \cdot 0,20 = 200$

(γράφουμε στο D2: =B1\*B2) και η αξία της αρχικής επένδυσης στο τέλος του 1<sup>ου</sup> έτους είναι (υπολογισμός στο E2: =B2+D2).

Στη συνέχεια, η διαδικασία γίνεται με τον ίδιο τρόπο, μόνο που τώρα το «αρχικό» κεφάλαιο είναι η τιμή στο E2 (τέλος του 1<sup>ου</sup> έτους) και για να βρούμε την απόδοσή του γράφουμε στο D3: = E2\*B3. Έτσι, η απόδοση στο τέλος του 2<sup>ου</sup> έτους είναι  $1200 * 0,06 = 72$  και άρα στο τέλος του 2<sup>ου</sup> έτους η αξία της αρχικής επένδυσης είναι 1272 ευρώ (στο E3: =E2+D3). Με τον ίδιο τρόπο βρίσκουμε ότι στο τέλος του 3<sup>ου</sup> έτους η αξία της αρχικής επένδυσης είναι 1287,72 (στο D4: = E3\*B4 και στο E4: = E3+D4).

- Η τιμή στο D12 είναι η αξία του 1€ επένδυσης στην αρχή κάθε περιόδου (έτους). Άρα η απόδοση είναι  $D12 - 1 =$  η τιμή στο D11.
- Δεν είναι δύσκολο να διαπιστώσουμε ότι ο τύπος πρέπει να δώσουμε σε ένα κελί στο CALC είναι  $=((1+0,2)*(1+0,06)*(1+0,01))^{(1/3)}-1$
- Μεγαλύτερο όμως ενδιαφέρον έχει η κατασκευή ενός φύλλου εργασίας ώστε για διαφορετικές τιμές στο Αρχικό Κεφάλαιο καθώς επίσης και στα Επιτόκια του 1<sup>ου</sup>, του

2<sup>ου</sup> και του 3<sup>ου</sup>, να μπορούμε άμεσα να υπολογίζουμε τις Ετήσιες Αποδόσεις αλλά και το Συνολικό Κεφάλαιο στο τέλος κάθε έτους.

- Παρακάτω δίνεται ο τρόπος κατασκευής του φύλλου εργασίας (λεπτομέρειες στο Εργαστήριο).

	A	B	C	D	E	F	G
1	Αρχικό Κεφάλαιο (Επένδυση)	1.000,00 €		Αποδόσεις	Συνολικό Κεφάλαιο (Αξία Επένδυσης)		
2	Επιτόκιο 1ου χρόνου	20,00%	1ου έτους	=B1*B2	=B1+D2		
3	Επιτόκιο 2ου χρόνου	6,00%	2ου έτους	=E2*B3	=E2+D3		
4	Επιτόκιο 3ου χρόνου	1,00%	3ου έτους	=E3*B4	=E3+D4		
5	Αριθμητική Μέση Τιμή επιτοκίου 3 ετών	=AVERAGE(B2:B4)	1ου έτους	=B1*B5	=B1+D5		
6			2ου έτους	=E5*B5	=E5+D6		
7			3ου έτους	=E6*B5	=E6+D7	Τύπος Ανατοκισμού	=B1*(1+B5)^3
8	Υπολογισμός Ετήσιας Απόδοσης		1ου έτους	=(1+B2)			
9			2ου έτους	=(1+B3)			
10			3ου έτους	=(1+B4)			
11			GR	=GEOMEAN(D8:D10)-1	=B1*(1+D11)^3		
12			Geomean	=GEOMEAN(D8:D10)			
13							

**Παρατήρηση:** Ο (σωστός) υπολογισμός δίνεται στα πορτοκαλί κελιά ενώ στα πράσινα κελιά, χρησιμοποιείται το μέσο επιτόκιο των 3 ετών (και ο τύπος ανατοκισμού) για τον υπολογισμό του συνολικού κεφαλαίου.