



## Νέοι Αλγόριθμοι Στατιστικής Επεξεργασίας Δεδομένων μέσω του *StatsExcel*<sup>GR</sup> (Μυλωνάς Κ., 2012, 2018, 2022)

---

Σε συνέχεια του λογισμικού *StatsExcel*<sup>GR</sup> που παραλάβατε με το σύγγραμμα που αφορά στη "Στατιστική στις επιστήμες της συμπεριφοράς Ι", ετοίμασα-προγραμματίσα κατά τη διάρκεια της εκπαιδευτικής μου αδειάς (χειμερινό εξάμηνο ακαδ. έτους 2021-22) **δέκα (10) νέους αλγορίθμους στατιστικής ανάλυσης** και τους ενέταξα σε πέντε (5) αρχεία-ενότητες. Έχω αναρτήσει (μαζί με το παρόν αρχείο οδηγιών) τα αρχεία αυτά στα Έγγραφα του μαθήματος ΨΧ62 (Στατιστική Ι) αλλά και στα άλλα μαθήματά μου του Προγράμματος Σπουδών του Τμήματός μας (ΨΧ61, ΨΧ65 και, ΨΧ72).

### Οι Νέοι Αλγόριθμοι αφορούν στον υπολογισμό

- Στατιστικών Κριτηρίων ( $z$  &  $t$ , *Levene*, *Brown-Forsythe*, *Bartlett's*),
- Ορίων Εμπιστοσύνης για πέντε στατιστικούς δείκτες,
- Τακτικών Τιμών για περαιτέρω μη-παραμετρική ανάλυση, και
- Μη-παραμετρικών Δεικτών Συνάφειας *Tau-b* κατά *Kendall*, *Rho* κατά *Spearman* και τέλος, του πιθανοθεωρητικού δείκτη  $\gamma$ .

Μπορείτε, φυσικά, να "κατεβάσετε" τα αρχεία αυτά καθώς και το παρόν αρχείο οδηγιών και να τα αντιγράψετε στον υποφάκελλο του ηλεκτρονικού υπολογιστή σας, εκεί όπου έχετε αποθηκεύσει και τα υπόλοιπα αρχεία του οπτικού δίσκου.

**Σημαντική σημείωση:** Καθώς για πρακτικούς λόγους τα αρχεία δεν είναι κλειδωμένα, ελλοχεύει πάντα κίνδυνος να διαγραφούν κατά τη χρήση σημεία-φατνία των αρχείων που δεν πρέπει να διαγραφούν. Εάν συμβεί αυτό, οι αλγόριθμοι δεν θα λειτουργούν, δεν θα "τρέχουν". Για να αποφύγετε να "κατεβάζετε" εκ νέου τα αρχεία, συνιστάται να κρατήσετε πριν τη χρήση τους ένα αντίγραφο ασφαλείας σε επιμέρους υποφάκελλο του υπολογιστή σας.

**Τέλος, να σημειώσω ότι οι νέοι αλγόριθμοι είναι πιθανόν να περιέχουν "bugs",** καθώς είναι σε φάση beta και προφανώς επιδέχονται πολλών βελτιώσεων και αλλαγών. Εφόσον διαπιστωθεί κάποιο πρόβλημα στη λειτουργία τους, παρακαλώ να με ενημερώσετε.

Δείτε και αναλυτικές οδηγίες στις επόμενες σελίδες:

## Αλγόριθμος-αρχείο Tau-b.xls

- Αφορά στον υπολογισμό του μη-παραμετρικού δείκτη συνάφειας *Tau-b* κατά *Kendall*
- Ο υπολογισμός γίνεται μέσω μακροεντολών. Η βασική εφαρμογή θα ζητήσει να ενεργοποιήσει τις μακροεντολές (macros). Θα πρέπει να ενεργοποιηθούν **μόνον** αυτές του αλγορίθμου, όχι από άλλες πηγές (π.χ. από Foxit Corporation ή διαδικτυακές) καθώς οι μακροεντολές πολλές φορές είναι φορείς ιώσεων. Δείτε και τις οδηγίες εντός του αλγορίθμου και απαντήστε "enable macros" **μόνον** για ενεργοποίηση των μακροεντολών του Tau-b.xls όπως θα αναγράφεται κάτω από το *security warning* κατά την ενεργοποίηση του αρχείου. Δεν υφίσταται κίνδυνος εφόσον ακολουθήσει κανείς τα ανωτέρω
- Ακολουθώντας και τις οδηγίες εντός του αλγορίθμου, γυρίζουμε το πληκτρολόγιο στα ΑΓΓΛΙΚΑ και μεταφέρουμε με copy, paste special, values τα ερευνητικά στοιχεία των οποίων η συνάφεια μας ενδιαφέρει στις στήλες A και B ("X" και "Y" μετρήσεις) ΑΦΟΥ ΠΡΩΤΑ έχουμε διαγράψει πιθανώς υπαρκτά στοιχεία στις στήλες αυτές. Για να γίνει αυτό με ευκολία επιλέγουμε το φαντίο A3, κρατάμε πατημένα τα ctrl & shift μαζί και πιέζουμε πρώτα το → και μετά το ↓ ώστε να επιλέξουμε όλη την περιοχή με τα προηγούμενα δεδομένα. Ελέγχουμε ότι έχουμε επιλέξει όσα θέλουμε να διαγράψουμε και τίποτε περισσότερο και κατόπιν κρατώντας το αριστερό alt πλήκτρο πατημένο πιέζουμε e, a, και a
- Έχοντας μεταφέρει τα ερευνητικά στοιχεία στις στήλες A και B, κρατάμε πατημένο το **ctrl** και πιέζουμε **w** για να ενεργοποιηθεί η μακροεντολή
- Ο αλγόριθμος μετρά επιμέρους δείκτες που αφορούν στα δεδομένα και μετά το τέλος της καταμέτρησης επιστρέφει τον μη-παραμετρικό δείκτη *Tau-b* κατά *Kendall* (και σε στρογγυλοποιημένη μορφή)
- Για να μην χρειάζεται κάθε φορά που ενεργοποιείτε το αρχείο η διαγραφή των προηγούμενων στοιχείων, μπορεί κανείς να μην αποθηκεύσει το αρχείο μετά τον υπολογισμό και απλώς να καταγράψει τον στατιστικό δείκτη
- Ακολουθήστε και τις οδηγίες εντός του αλγορίθμου

## Αλγόριθμος-αρχείο Ranking X,Y plus.xls

- Αφορά στον υπολογισμό Τακτικών Τιμών για ζεύγος μετρήσεων X, Y για τις οποίες επίσης υπολογίζονται ο μη-παραμετρικός δείκτης συνάφειας *Tau-b* κατά *Kendall*, ο μη παραμετρικός δείκτης συνάφειας *Rho* κατά *Spearman* καθώς και, ο μη-παραμετρικός πιθανοθεωρητικός δείκτης συνάφειας  $\gamma$  (gamma). Ο υπολογισμός των Τακτικών Τιμών γίνεται χωριστά για X και χωριστά για Y
- Ο υπολογισμός γίνεται μέσω μακροεντολών. Η βασική εφαρμογή θα ζητήσει να ενεργοποιήσει τις μακροεντολές (macros). Θα πρέπει να ενεργοποιηθούν **μόνον** αυτές του αλγορίθμου, όχι από άλλες πηγές (π.χ. από Foxit Corporation ή διαδικτυακές) καθώς οι μακροεντολές πολλές φορές είναι φορείς ιώσεων. Δείτε και τις οδηγίες εντός του αλγορίθμου και απαντήστε "enable macros" **μόνον** για ενεργοποίηση των μακροεντολών του Ranking X,Y plus.xls όπως θα αναγράφεται κάτω από το *security warning* κατά την ενεργοποίηση του αρχείου. Δεν υφίσταται κίνδυνος εφόσον ακολουθήσει κανείς τα ανωτέρω
- Ακολουθώντας και τις οδηγίες εντός του αλγορίθμου, γυρίζουμε το πληκτρολόγιο στα ΑΓΓΛΙΚΑ και μεταφέρουμε με copy, paste special, values τα ερευνητικά στοιχεία των οποίων η συνάφεια μας ενδιαφέρει στις στήλες A και B ("X" και "Y" μετρήσεις) ΑΦΟΥ ΠΡΩΤΑ έχουμε διαγράψει πιθανώς υπαρκτά στοιχεία στις στήλες αυτές. Για να γίνει αυτό με ευκολία επιλέγουμε το φαντίο A3, κρατάμε πατημένα τα ctrl & shift μαζί και πιέζουμε πρώτα το → και μετά το ↓ ώστε να επιλέξουμε όλη την περιοχή με τα προηγούμενα δεδομένα. Ελέγχουμε ότι έχουμε επιλέξει όσα θέλουμε να διαγράψουμε και τίποτε περισσότερο και κατόπιν κρατώντας το αριστερό alt πλήκτρο πατημένο πιέζουμε e, a, και a
- Έχοντας μεταφέρει τα ερευνητικά στοιχεία στις στήλες A και B, κρατάμε πατημένο το **ctrl** και πιέζουμε **e** για να ενεργοποιηθεί η μακροεντολή
- Ο αλγόριθμος υπολογίζει και εμφανίζει σε δύο χωριστές στήλες τις Τακτικές Τιμές (ranks) για κάθε μέτρηση χωριστά (X και Y) οι οποίες μπορούν να αντιγραφούν και να χρησιμοποιηθούν περαιτέρω σε άλλους αλγορίθμους
- Ο αλγόριθμος επίσης μετρά επιμέρους δείκτες που αφορούν στα δεδομένα και μετά το τέλος της καταμέτρησης επιστρέφει τους τρεις μη-παραμετρικούς δείκτες. Σημ.: για τον *Spearman's Rho*, δεν γίνεται διόρθωση κατά *Siegel* (1956) για τους κόμβους, δηλαδή οι κόμβοι εξακολουθούν να επηρεάζουν το αποτέλεσμα σύμφωνα με τη θεωρία του *Spearman*
- Για να μην χρειάζεται κάθε φορά που ενεργοποιείτε το αρχείο η διαγραφή των προηγούμενων στοιχείων, μπορεί κανείς να μην αποθηκεύσει το αρχείο μετά τον υπολογισμό και απλώς να αντιγράψει όσα χρειάζεται πριν κλείσει το αρχείο χωρίς αποθήκευση
- Ακολουθήστε και τις οδηγίες εντός του αλγορίθμου

### Αλγόριθμος-αρχείο z&t tests.xls

- Αφορά στο στατιστικό έλεγχο σύγκρισης ενός δείγματος (μεγάλου, με  $N \geq 30$  ή μικρού, με  $N < 30$ ) με τον στατιστικό πληθυσμό
- Αναλόγως του μεθέθους του δείγματος υπολογίζεται το z-κριτήριο ή για μικρά δείγματα το t-κριτήριο κατά *Student*
- Ο υπολογισμός γίνεται επί των πρωτογενών X στοιχείων της υπό μελέτη μέτρησης (δεν χρησιμοποιείται μακροεντολή)
- Ακολουθήστε και τις οδηγίες εντός του αλγορίθμου

### **Αλγόριθμος-αρχείο HomogeneityTests.xls**

- Αφορά στο στατιστικό έλεγχο ως προς την πιθανή παραβίαση της συνθήκης ομοιογένειας της διασποράς υπό στατιστική ανάλυση σύγκρισης δύο ή περισσότερων ομάδων (έως 10 στη συγκεκριμένη εφαρμογή)
- Υπολογίζονται 3 στατιστικά κριτήρια ελέγχου της στατιστικής συνθήκης της ομοιογένειας και συγκεκριμένα (κατά σειρά αυστηρότητας) το κριτήριο *Levene*, το κριτήριο *Brown-Forsythe* και το κριτήριο *Bartlett's*
- Χρειάζονται τα πρωτογενή ερευνητικά στοιχεία για κάθε συγκρινόμενη ομάδα (δεν χρησιμοποιείται μακροεντολή)
- Ακολουθήστε και τις οδηγίες εντός του αλγορίθμου

## Αλγόριθμος-αρχείο CL\_z\_sd\_%\_r\_Δ.xls

- Αφορά στον υπολογισμό Ορίων Εμπιστοσύνης
- Περιλαμβάνει 5 επιμέρους αλγορίθμους και συγκεκριμένα Υπολογισμό Ορίων Εμπιστοσύνης i) Μέσου Όρου, ii) Δειγματικής Τυπικής Απόκλισης, iii) Ποσοστού, iv) Δείκτη Συνάφειας  $r$  κατά *Pearson* και, v) Διαφοράς δύο μέσων όρων ανεξαρτήτων ισοπληθών διαφορικών ομάδων
- Σε κάθε υπολογισμό, ο ερευνητής χρειάζεται να γνωρίζει στοιχεία που αφορούν στα ερευνητικά του δεδομένα, δηλαδή χρειάζεται να γνωρίζει κάποιους δείκτες καθώς οι υπολογισμοί δεν γίνονται επί των αρχικών  $X$  δεδομένων αλλά επί των στατιστικών δεικτών τους. Συγκεκριμένα, χρειάζεται να είναι γνωστά (κατά περίπτωση) τα ακόλουθα:
  - i) πλήθος συμμετεχόντων, μέσος όρος για τη συγκεκριμένη αριθμητική μέτρηση και, πληθυσμιακή τυπική απόκλιση  $s$ , ενώ ο ερευνητής θα πρέπει να έχει αποφασίσει εάν πρόκειται για μονοκατάληκτη ή δικατάληκτη ερευνητική στατιστική υπόθεση καθώς και το σε ποιο επίπεδο  $\alpha$  θα αφορά η εκτίμηση
  - ii) πλήθος συμμετεχόντων, δικατάληκτο επίπεδο στατιστικής σημαντικότητας  $\alpha$  για το οποίο θα γίνει η εκτίμηση των ορίων και, δειγματική διασπορά  $\hat{s}^2$
  - iii) πλήθος συμμετεχόντων, το παρατηρούμενο ποσοστό, δικατάληκτο επίπεδο στατιστικής σημαντικότητας  $\alpha$  για το οποίο θα γίνει η εκτίμηση των ορίων και, προαιρετικά το μέγεθος του στατιστικού πληθυσμού
  - iv) πλήθος συμμετεχόντων, ο παρατηρούμενος στατιστικός δείκτης  $r$  και, δικατάληκτο επίπεδο στατιστικής σημαντικότητας  $\alpha$  για το οποίο θα γίνει η εκτίμηση των ορίων
  - v) συνολικό πλήθος συμμετεχόντων (και στις δύο ισοπληθείς ομάδες), δειγματική διασπορά  $\hat{s}^2$  για την πρώτη ομάδα, δειγματική διασπορά  $\hat{s}^2$  για τη δεύτερη ομάδα, μέσος όρος της πρώτης ομάδας, μέσος όρος της δεύτερης ομάδας και, δικατάληκτο επίπεδο στατιστικής σημαντικότητας  $\alpha$  για το οποίο θα γίνει η εκτίμηση των ορίων
- Ακολουθήστε και τις οδηγίες εντός του αλγορίθμου και κάθε επιμέρους εφαρμογής του

*Καλή τύχη και καλά αποτελέσματα!*

*Κ. Μυλωνάς, Φεβρουάριος 2022*