

ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΙΑ ΡΙΖΟΣΠΑΣΤΙΚΗ ΑΝΑΘΕΩΡΗΣΗ ΚΑΘΙΕΡΩΜΕΝΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΟΡΩΝ

Αλκιβιάδης Αλεξ. Γούσης

1. Περίληψη

Η συμβολή αυτή έχει ως σκοπό την παρουσίαση προτάσεων για μία ριζοσπαστική αναθεώρηση συγκεκριμένων καθιερωμένων στατιστικών όρων. Στηρίζεται τόσο σε μία εμπειριστατωμένη ετυμολογική διερεύνηση, όσον και σε μία ιστορική υποστήριξη μέσω των πρωτογενών εργασιών των γεννητόρων αυτών των όρων. Το δεύτερο αυτό σκέλος απεδείχθη σε ωρισμένες περιπτώσεις εξαιρετικά χρήσιμο για την σαφή κατανόηση της έννοιας του αντίστοιχου στατιστικού όρου. Εκτός όμως των ανωτέρω δύο προσεγγίσεων, ελήφθησαν υπ' όψιν και οι αποδόσεις των όρων αυτών σε άλλες γλώσσες.

2 Summary

The aim of this contribution is the presentation of propositions for a radical revision of specific established statistical terms. It is based on a thorough etymological investigation and on an historical literature review, based on the original papers of the parents of these terms. This second approach has been proved to be very useful for a clear understanding of the meaning of the corresponding statistical terms. Apart from these two approaches, the way of translation of these terms in other languages has been taken into consideration.

3 Προτάσεις

- **Regression : Συμμεταβολή**

Μεταξύ των πολυπληθών μελετών του περί την κληρονομικότητα, ο sir Francis Galton ασχολήθηκε και με το πως αυτή επιδρά στο ύψος των ατόμων [1]. Για τον λόγο αυτόν ανέθεσε στον φίλο του Karl Pearson, γνωστή και σημαντική μορφή της στατιστικής, την διενέργεια δειγματοληψίας με σκοπό την λήψη ενός αντιπροσωπευτικού δείγματος από το σύνολο των Αγγλικών οικογενειών και του εζήτησε να καταγράψει, για κάθε οικογένεια, το ύψος του πατέρα και το ύψος του υιού.

Ο Pearson παρέδωσε στον Galton 1708 ζεύγη υψών, τα οποία ο δεύτερος μετέφερε σε ένα καρτεσιανό διάγραμμα, τοποθετώντας στον άξονα των X τα ύψη των πατέρων (μέσοι όροι ανά κλάση ύψους) και στον άξονα των Y τους αντίστοιχους μέσους όρους των υιών οι οποίοι ανήκαν στην αυτή κλάση των πατέρων. Εξετάζοντας την τελική μορφή αυτού του σημειοδιαγράμματος, έκρινε ότι η προκύψασα σχέση μεταξύ των X και των Y, θα ήταν δυνατόν να ερμηνευθεί μέσω ενός κλασικού γραμμικού μοντέλου.

Η χαραχθείσα ευθεία γραμμή, η οποία είχε την αριθμητική έκφραση $Y = 33.73 + 0.516X$, επέτρεψε στον Galton να διαπιστώσει ότι :

- αν και υψηλοί πατέρες είχαν την τάση να έχουν επίσης υψηλούς υιούς, εντούτοις ο μέσος όρος των υιών οι οποίοι ανήκαν σε μία δεδομένη κλάση υψηλών πατέρων, ήταν μικρότερος εκείνου των πατέρων τους.
- αν και χαμηλοί πατέρες είχαν την τάση να έχουν επίσης χαμηλούς υιούς, εντούτοις ο μέσος όρος των υιών οι οποίοι ανήκαν σε μία δεδομένη κλάση χαμηλών πατέρων, ήταν μεγαλύτερος εκείνου των πατέρων τους.

Το διαπιστωθέν αυτό φαινόμενο ο Galton το ονόμασε «οπισθοδρόμηση των υψών των υιών προς τον μέσον όρο των υψών όλων των πατέρων», ενώ την αντιστοιχούσα ευθεία ως «ευθεία οπισθοδρόμησης». Εις το σημείο αυτό υπενθυμίζουμε ότι ο αρχικά επιλεγείς από τον Galton όρος δεν ήταν “*regression*” αλλά “*reversion*”.

Είναι βέβαια γεγονός ότι ο όρος “*regression*” καθιερώθηκε στην Αγγλική γλώσσα και κανείς δεν αποπειράται να τον αντικαταστήσει ή έστω να τον βελτιώσει. Το αυτό παρατηρείται και στις περισσότερες Ευρωπαϊκές χώρες οι οποίες διαθέτουν την ίδια λέξη, αν και ολοένα και περισσότερο εκφράζονται απόψεις για την ανάγκη αντικατάστασής του από άλλον πιά επιτυχημένο όρο.

Η ανωτέρω αναπτυχθείσα ιστορική αναφορά καταδεικνύει ότι η χρησιμοποίηση στη χώρα μας του όρου «παλινδρόμηση», ο οποίος αποτελεί απλά μία από τις αποδόσεις στην ελληνική γλώσσα του όρου «*regression*», δεν έχει καμία απολύτως σχέση με την πραγματικότητα η οποία συνατάται στον χώρο της μοντελοποίησης (γραμμικής ή μη), καθόσον είναι σε όλους γνωστό ότι τα κατά περίπτωση χρησιμοποιούμενα για την περιγραφή κάποιας συσχέτισης μαθηματικά μοντέλα, ιδιαίτερα στον χώρο της έρευνας, ουδεμία σχέση έχουν με «παλινδρόμηση», «οπισθοδρόμηση» ή ακόμη «αναστροφή».

Με βάση την ανωτέρω αναπτυχθείσα τεκμηρίωση, προτείνεται η υιοθέτηση του όρου «συμμεταβολή», ο οποίος πιστεύουμε αποδίδει πιστότερα αυτό το οποίο όλοι εννοούμε στην πραγματικότητα ως “*regression*” .

- **Variance : Παραλλακτικότητα**

Ο πρώτος όρος ο οποίος αποδόθηκε στο γνωστό στατιστικό μέγεθος ήταν “*fluctuation*” από τον Francis Ysidro Edgeworth, το 1888. Τριάντα ακριβώς χρόνια αργότερα ο R.A. Fisher [2], πρότεινε τον καθιερωθέντα όρο “*variance*”.

Εάν ο Fisher δεν είχε αλλάξει τον όρο του Edgeworth, τότε πράγματι το στατιστικό αυτό μέγεθος, θα έπρεπε να αποδοθεί ως «*διακύμανση*», δεδομένου ότι ετυμολογικώς ο όρος “*fluctuation*” προέρχεται από την Λατινική λέξη “*fluctus : κύμα*”. Όμως ο καθιερωθείς όρος του Fisher, ο οποίος αποτελεί έναν εντελώς διαφορετικό όρο από εκείνον του Edgeworth, μας υποχρεώνει να προτείνουμε και να υιοθετήσουμε έναν κατάλληλότερο όρο.

Οι ορισμοί των δύο συγγενών όρων “*variability*” και “*variance*”, οι οποίοι δίνονται σε έγκριτα λεξικά [3] και συμπίπτουν μεταξύ τους, έχουν ως ακολούθως :

Variability : Η κατάσταση ή η ιδιότητα του μεταβλητού (variable)

Variance : Η κατάσταση ή η ιδιότητα του παραλάσσοντος (variant)

Με βάση τα ανωτέρω αναπτυχθέντα, προτείνεται ο όρος «παραλλακτικότητα» για τον όρο “*variance*”, και ο όρος «μεταβλητότητα» για τον όρο “*variability*”.

- **Standard deviation : Πρότυπη απόκλιση**

Ο όρος αυτός προτάθηκε από τον Karl Pearson [4] το 1894, ενώ στην ελληνική γλώσσα αποδόθηκε ως «*τυπική απόκλιση*». Λαμβάνοντας υπ’ όψιν την απόδοση του όρου αυτού στην γερμανική και στην γαλλική γλώσσα, έχουμε την ακόλουθη αντιστοιχία :

Standard deviation – Standard Abweichung – Ecart-type.

Θεωρούμε ότι προφανώς η ελληνική απόδοση βασίστηκε στην αντίστοιχη γαλλική. Όμως, ο γαλλικός όρος δεν διατυπώνεται ως “*écart typique*”, δηλαδή «*τυπική απόκλιση*». Για όσους γνωρίζουν την γαλλική, η παράθεση, μετά από ένα ουσιαστικό, της λέξης “*type*”, διαχωριζομένης από το ουσιαστικό με μία παύλα, αποδίδει στο ουσιαστικό την έννοια του «*υπόδειγμα*», «*πρότυπο*». Κατ’ αυτόν τον τρόπο, όταν π.χ. χαρακτηρίζουμε έναν καθηγητή ως “*professeur-type*”, αυτό σημαίνει ότι τον χαρακτηρίζουμε ως «*καθηγητή-υπόδειγμα*» ή ως «*καθηγητή-πρότυπο*».

Ένα επίσης σημαντικό στοιχείο το οποίο συνηγορεί στην υιοθέτηση του όρου «*πρότυπη απόκλιση*», ελήφθη δε υπ’ όψιν και από τον Pearson, προέρχεται από την ίδια την θεωρία

της, κατά Gauss, κανονικής κατανομής. Αποδεικνύεται μαθηματικώς ότι η ανηγμένη καμπύλη της κανονικής κατανομής διαθέτει, εκατέρωθεν του κέντρου αυτής, δύο συμμετρικά σημεία καμπής. Το πρώτο εξ' αυτών αντιστοιχεί ακριβώς στο 1σ , ενώ το δεύτερο ακριβώς στο 2σ . Στο σημείο αυτό πρέπει να τονίσουμε ότι, μεταξύ των διαφόρων μέτρων διασποράς τα οποία χρησιμοποιούνται στη Στατιστική, μόνον το συγκεκριμένο μέτρο διαθέτει αυτή την σημαντική ιδιότητα.

Με βάση τα ανωτέρω αναπτυχθέντα, προτείνεται η υιοθέτηση του όρου «**πρότυπη απόκλιση**» ως αρμόζουσα απόδοση, η οποία ευρίσκεται και σε πλήρη συμφωνία με τις προαναφερθείσες στις τρεις γλώσσες, διατυπώσεις.

- **Accuracy : Ακρίβεια**
- **Trueness : Ορθότητα**
- **Precision : Πιστότητα**
- **Bias : Στρεβλότητα**

Η απόδοση στην ελληνική γλώσσα των ανωτέρω τεσσάρων βασικώτατων όρων της στατιστικής, εξακολουθεί, ακόμη και σήμερα να ταλανίζει ένα μεγάλο αριθμό ατόμων οι οποίοι τους χρησιμοποιούν σε καθημερινή βάση. Η παρατηρούμενη πανσπερμία των αποδόσεων ενός εκάστου εξ' αυτών των όρων, πιστεύουμε ακράδαντα ότι οφείλεται κυρίως στη σύγχυση η οποία προκαλείται λόγω του, φαινομενικώς, ταυτοσήμου των εννοιών των όρων “accuracy” και “precision”, στην αγγλική γλώσσα.

Οι προτεινόμενες αποδόσεις των τεσσάρων όρων βασίζονται σε μία διεξοδική ετυμολογική διερεύνηση, στους ορισμούς οι οποίοι διατυπώνονται στο [5] καθώς και στις αντίστοιχες αποδόσεις τους σε άλλες γλώσσες.

Για να καταστεί πιά εύκολη η κατανόηση της ακολουθηθείσας διαδικασίας, θα πρέπει ευθύς εξ' αρχής να τονισθεί ότι, ενώ οι δύο πρώτοι όροι καθώς και ο τέταρτος όρος είναι «μέτρα θέσεως», ο τρίτος όρος είναι «μέτρο διασποράς». Με βάση αυτή την, σημαντική πιστεύουμε, διευκρίνιση, δικαιολογείται και ο χαρακτηρισμός «φαινομενικώς», τον οποίο αναφέραμε στην πρώτη παράγραφο.

Σύμφωνα με τους αντίστοιχους ορισμούς οι οποίοι δίνονται στο [5],

- ❖ «Ακρίβεια είναι η εγγύτητα μεταξύ ενός αποτελέσματος δοκιμής και της αποδεκτής τιμής αναφοράς», ενώ στην ακολουθούσα σημείωση διευκρινίζεται ότι «ο όρος αυτός, εφαρμοζόμενος σε ένα σύνολο πειραματικών αποτελεσμάτων, περιλαμβάνει έναν συνδυασμό τυχαίων και συστηματικών σφαλμάτων».
- ❖ «ορθότητα είναι η εγγύτητα μεταξύ του αριθμητικού μέσου όρου ενός μεγάλου αριθμού αποτελεσμάτων δοκιμής και της αποδεκτής τιμής αναφοράς»
- ❖ «πιστότητα είναι η εγγύτητα μεταξύ ανεξαρτήτων αποτελεσμάτων δοκιμής, ληφθέντων υπό ρητώς διατυπωθείσες συνθήκες »
- ❖ «στρεβλότητα είναι διαφορά μεταξύ της μαθηματικής προσδοκίας των αποτελεσμάτων δοκιμής και της αποδεκτής τιμής αναφοράς».

Λαμβάνοντας υπ' όψιν την προαναφερθείσα σημείωση την ακολουθούσα τον ορισμό του όρου «ακρίβεια», ο ISO τροποποίησε τον τίτλο της νέας έκδοσης των προτύπων της σειράς ISO 5725 – Parts 1 to 6, ως ακολούθως : «Ακρίβεια (ορθότητα και πιστότητα) των μετρητικών μεθόδων και αποτελεσμάτων»[6].

Ο όρος «ακρίβεια» είναι ένας γενικός όρος και αναφέρεται τόσο στην «ορθότητα» όσο και στην «πιστότητα». Μέχρι πρότινος ο όρος «ακρίβεια» χρησιμοποιείτο για να καλύψει μόνον την συνιστώσα η οποία σήμερα ονομάζεται «ορθότητα», πλην όμως κατέστη σαφές και αναγκαίο ότι θα έπρεπε να υποδηλώνει την συνολική μετατόπιση ενός αποτελέσματος δοκιμής από μία αποδεκτή τιμή αναφοράς, μετατόπιση οφειλόμενη τόσο στις τυχαίες όσο και στις συστηματικές επιδράσεις.

Είς ότι αφορά τον όρο «*bias*» ο οποίος έχει αποδοθεί στην ελληνική είτε ως «μεροληψία» είτε ως «βιασμός», θεωρούμε ότι ουδείς εξ' αυτών έχει κάποια σχέση με την πραγματικότητα, τόσο από ετυμολογικής, ιδιαίτερα όμως, όσο και από στατιστικής άποψης.

Ετυμολογικώς, ο όρος “bias” προέρχεται από τον γαλλικό όρο “biais”, ο οποίος σημαίνει «λοξός, στρεβλός, μη ορθός».

Από στατιστικής άποψης, π.χ., στην περίπτωση της αξιολόγησης μίας αναλυτικής μεθόδου , μέσω της διοργάνωσης Συγκριτικών Διεργαστηριακών Δοκιμών, η διαπίστωση ότι η μέθοδος παρουσιάζει “bias”, δεν αποδεικνύεται λόγω κάποιας «μεροληψίας» ή κάποιου «βιασμού»,

αλλά μετά την συστηματική και εις βάθος αξιολόγηση των αποτελεσμάτων, δια της εφαρμογής συγκεκριμένων στατιστικών τύπων και κριτηρίων.

Με βάση τα ανωτέρω, όταν αποδειχθεί στατιστικώς ότι μία μέθοδος παρουσιάζει “*bias*”, αυτό σημαίνει ότι το προκύψαν τελικό αποτέλεσμα, το οποίο εκφράζεται από τον συνολικό μέσον όρο των αποτελεσμάτων των εις το τέλος της στατιστικής αξιολόγησης παραμεινάντων εργαστηρίων, συγκρινόμενο με την αποδεκτή τιμή αναφοράς, δεν είναι «ορθό», δηλαδή «στρεβλό».

Εις ό,τι αφορά τον όρο “*precision*”, όπως αναφέραμε και προηγουμένως, είναι ένα μέτρο διασποράς και ως εκ τούτου αντανακλά την ικανότητα μίας μεθόδου ή μίας συσκευής να παρέχει, ει δυνατόν, τα αυτά αποτελέσματα, δηλαδή να παρουσιάζει μηδενική διασπορά, άρα απόλυτη «*πιστότητα*».

Ο ανωτέρω προτεινόμενος όρος, βασίζεται στον γαλλικό όρο “*fidélité*”, ο οποίος αποτελεί την γαλλική απόδοση του όρου “*precision*”. Θεωρούμε ότι είναι επιτυχής και διαθέτει την ιδιότητα να θέτει ένα οριστικό τέλος στην, για πολλούς, ασαφή σχέση του με τον όρο “*accuracy*”.

ΣΥΝΟΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΠΡΟΤΕΙΝΟΜΕΝΩΝ ΟΡΩΝ

Αγγλική	Γαλλική	Γερμανική	Ελληνική
regression	regression	Regression	συμμεταβολή
variance	variance	Varianz	παραλλακτικότητα
variability	variabilité	Variabilität	μεταβλητότητα
standard deviation	écart-type	Standardabweichung	πρότυπη απόκλιση
accuracy	exactitude	Genauigkeit	ακρίβεια
trueness	justesse	Richtigkeit	ορθότητα
precision	fidélité	Präzision	πιστότητα
bias	biais	Systematische Abweichung	στρεβλότητα

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Galton F., “Regression towards mediocrity in hereditary stature”, Journal of the Anthropological Institute, pp. 246-263, (1886)
- [2] Fisher R.A., “The correlation between relations on the supposition of Mendelian inheritance”, Transactions of the Royal Society of Edinburgh, Vol. 52, pp. 399-433, (1918)
- [3] Oxford English Dictionary, Third Edition, (1997)
Merriam-Webster Collegiate Dictionary, (1998)
The Wordsmyth English Dictionary-Thesaurus, ARTFL Project, (1998)
Webster’s New Twentieth Century Dictionary of the English Language, 2nd Edition, (1962)
- [4] Pearson K., “Contributions to the mathematical theory of evolution”, Philosophical Transactions of the Royal Society of London, Series A, vol. 185, pp. 71-110, (1894)
- [5] ISO 3534-1:1993, Statistics – Vocabulary and symbols – Part 1 : Probability and general statistical terms.
- [6] ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results – Part 1 : General principles and definitions

Αλκιβιάδης Αλεξ. Γούσης

Υπεύθυνος Διαπίστευσης Φορέων Πιστοποίησης
Εθνικό Σύστημα Διαπίστευσης

Σισίνη 8, 115 28 Αθήνα
Αριθμός τηλεφώνου : 720 4603
Αριθμός τηλεμοιτύπου : 720 4500
e-mail : gousisa@ypan.gr