

Μενεξές, Γ. , Οικονόμου, Α. (2002). Σφάλματα και Παρανοήσεις στους Στατιστικούς Ελέγχους Υποθέσεων. Υπέρβαση μέσω της Ανάλυσης Δεδομένων. *Τετράδια Ανάλυσης Δεδομένων-Data Analysis Bulletin*, 2, 52-64.

ΣΦΑΛΜΑΤΑ ΚΑΙ ΠΑΡΑΝΟΗΣΕΙΣ ΣΤΟΥΣ ΣΤΑΤΙΣΤΙΚΟΥΣ ΕΛΕΓΧΟΥΣ ΥΠΟΘΕΣΕΩΝ. ΥΠΕΡΒΑΣΗ ΜΕΣΩ ΤΗΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ.

Μενεξές Γεώργιος, Οικονόμου Ανδρέας

ΠΑΤΕΣ/ΣΕΛΕΤΕ Θεσσαλονίκης

ΠΕΡΙΛΗΨΗ

Κατά παράδοση, ο έλεγχος υποθέσεων στην επιστημονική έρευνα έχει να επιδείξει μια σαφή προτίμηση στη χρησιμοποίηση της στατιστικής σημαντικότητας ως κριτηρίου ή προσέγγισης σχετικά με την απόρριψη ή όχι της μηδενικής υπόθεσης. Στο πλαίσιο αυτό η απόφαση του ερευνητή μπορεί να είναι σωστή ή λανθασμένη. Λανθασμένες αποφάσεις μπορούν να ληφθούν κάτω από την επίδραση στοχαστικών και λογικών σφαλμάτων καθώς και λόγω εσφαλμένων αντιλήψεων. Σχετικά με τη διαδικασία του ελέγχου της σημαντικότητας της μηδενικής υπόθεσης έχει ασκηθεί κριτική ήδη από τη δεκαετία του 60' η οποία επαναλαμβάνεται περιοδικά από διάφορους επικριτές. Η διαδικασία αυτή μπορεί αλλά και δίνει λανθασμένες απαντήσεις. Στην εργασία παρουσιάζουμε μερικά ζητήματα που αφορούν στην αξιολόγηση των αποτελεσμάτων των στατιστικών ελέγχων υποθέσεων με σκοπό την αποφυγή του στατιστικού υπερκαταναλωτισμού (βλέπε p-value). Τα ζητήματα αυτά θέτουν τις βάσεις για μια επιστημολογική αναθεώρηση της πληροφορίας, με την έννοια της γνώσης, που παράγεται από τους στατιστικούς ελέγχους υποθέσεων. Προς την κατεύθυνση αυτή μπορεί να συμβάλει η Ανάλυση Δεδομένων με τις μεθόδους της.

ΑΝΤΙ ΕΙΣΑΓΩΓΗΣ

Περίπτωση 1^η. Στοχαστικός Αναλφαβητισμός και Φισεριανή Επιστημολογία.

Οι έλεγχοι σημαντικότητας αν και έχουν επανειλημμένα δεχθεί αυστηρή κριτική από τους ειδικούς μεθοδολόγους, ωστόσο θεωρώ ότι είναι πολύ χρήσιμοι στην ερμηνεία των αποτελεσμάτων μου και δεν έχω σκοπό να τους εγκαταλείψω. Αν κάποιος από τα αποτελέσματα δεν είναι στατιστικά σημαντικό, τότε γνωρίζω ότι πιθανώς να προήλθε στην «τύχη» και ότι το πραγματικό αποτέλεσμα (διαφορά, συσχέτιση, επίδραση) είναι πιθανώς ίσο με μηδέν. Αν τα αποτελέσματα είναι στατιστικά σημαντικά τότε γνωρίζω ότι έχω ένα αξιόπιστο εύρημα. Η παρατηρούμενη στάθμη σημαντικότητας (π.σ.σ.) p του αντίστοιχου στατιστικού ελέγχου δηλώνει το κατά πόσο το δειγματικό μου αποτέλεσμα είναι αρκετά μεγάλο ώστε να θεωρηθεί ως σημαντικό ή όχι από πρακτική ή κλινική σκοπιά. Επίσης, από την τιμή της π.σ.σ. p μπορώ να καθορίσω το πόσο πιθανό είναι αυτά τα ευρήματα να επαναληφθούν σε περίπτωση που πραγματοποιήσω μια νέα έρευνα. Όλα αυτά είναι πολύ χρήσιμα για έναν ερευνητή.

Όσα αναφέρονται παραπάνω θα μπορούσαν να είναι σκέψεις ενός τυπικού ερευνητή. Όλα όμως τα επιχειρήματά του, με τον τρόπο που διατυπώνονται, είναι λανθασμένα (Schmidt, 1996). Πρόκειται για την όχι σπάνια περίπτωση ενός στοχαστικά αναλφάβητου ερευνητή εγκλωβισμένου ταυτόχρονα στην επιστημολογία της Φισεριανής Στατιστικής.

Περίπτωση 2^η. $P(\text{Data} / H_0) = P(H_0 / \text{Data})$; - Η Στατιστική του Ιατροδικαστή.

Ποια είναι η πιθανότητα ένα άτομο να είναι νεκρό με δεδομένο ότι έχει απαγχονιστεί; Προφανώς η πιθανότητα είναι αρκετά υψηλή ας πούμε 0,97 ή υψηλότερη. Ας αντιστρέψουμε την ερώτηση: Ποια είναι η πιθανότητα ένα άτομο να έχει απαγχονιστεί με δεδομένο ότι είναι νεκρό; Αυτή τη φορά η πιθανότητα είναι αρκετά μικρή, ας πούμε 0,01. Είναι φανερό ότι κανείς δε θα έκανε το λάθος να δεχθεί την τιμή 0,97 ως εκτίμηση της πιθανότητας το άτομο να έχει απαγχονιστεί δεδομένου ότι είναι νεκρό.

Όμως, τέτοια λογικά λάθη γίνονται συχνά στην ερμηνεία των αποτελεσμάτων των ελέγχων σημαντικότητας (Carver, 1978), δηλαδή εκτιμήσεις της δεσμευμένης πιθανότητας $P(\text{Δεδομένα} / H_0)$ να εκλαμβάνονται ως εκτιμήσεις της δεσμευμένης πιθανότητας $P(H_0 / \text{Δεδομένα})$.

Περίπτωση 3^η. Ανθολογία του Huck.

Ο Huck (2000a) αναφέρει ότι όταν οι άνθρωποι διαβάζουν ή προετοιμάζουν ερευνητικές αναφορές έχουν ορισμένες κοινές και επικίνδυνες παρανοήσεις για τη

συλλογή, ανάλυση και ερμηνεία των δεδομένων. Παραθέτουμε τις πιο σημαντικές:

1. Ο ερευνητής έχει πλήρη έλεγχο στον καθορισμό της στάθμης σημαντικότητας (σ.σ.) α ενός στατιστικού ελέγχου. Κατά συνέπεια, είναι αντιδεοντολογικό να αλλάξει ο αναγνώστης τη σ.σ. α που έχει θέσει ο ερευνητής (π.χ. από $\alpha=0,05$ σε $\alpha=0,001$) και συνεπώς να καταλήξει σε διαφορετικά συμπεράσματα.
2. Η π.σ.σ. p εκφράζει την πιθανότητα να διαπράξουμε Σφάλμα Τύπου I.
3. Αν κάποιος ερευνητής στην παρουσίαση των αποτελεσμάτων δηλώσει $p<0,05$ αυτό σημαίνει ότι η σ.σ. έχει προκαθοριστεί σε $\alpha=0,05$ και ότι η H_0 έχει απορριφθεί.
4. Η ερευνητική υπόθεση ταυτίζεται πάντα με την εναλλακτική υπόθεση.
5. Πολύ μικρές τιμές της π.σ.σ. p (π.χ. $p=0,001$ ή $p=0,0004$) δηλώνουν και εντονότερες διαφορές, συσχετίσεις, επιδράσεις, κ.λπ.
6. Ένα στατιστικά σημαντικό αποτέλεσμα είναι και πρακτικά ή κλινικά σημαντικό.
7. Οι έλεγχοι υποθέσεων με την προσέγγιση της στατιστικής σημαντικότητας μπορούν να γίνουν και για μη τυχαία δείγματα.

Όλες αυτές οι λανθασμένες αντιλήψεις των τριών περιπτώσεων είναι αποτέλεσμα κυρίως της παράδοσης και των συμβάσεων που έχουν επικρατήσει στο πλαίσιο της «Φισεριανής» Στατιστικής και, γενικότερα, της κυριαρχίας στην εποχή μας της σκέψης που βασίζεται στη λογική του ΝΑΙ-ΟΧΙ (DeBono, 2000). Οι λανθασμένες αντιλήψεις σε συνδυασμό με τον κοινωνικό παράγοντα, κάτω από την επίδραση του οποίου, στις μέρες μας, οι στόχοι για εγκυρότητα και αξιοπιστία των στατιστικών αποτελεσμάτων υποχωρούν σε ισχυρές πιέσεις του τύπου «δείξε ή βρες γρήγορα κάτι στατιστικά σημαντικό», δημιουργούν μια στρεβλή σχέση μεταξύ της Στατιστικής και της Έρευνας που οδηγεί σε στατιστικό υπερκαταναλωτισμό και μοιραία σε λάθη.

ΕΙΔΗ ΣΦΑΛΜΑΤΩΝ ΣΤΟΥΣ ΣΤΑΤΙΣΤΙΚΟΥΣ ΕΛΕΓΧΟΥΣ

Σφάλμα Τύπου I και Σφάλμα Τύπου II.

Σε ένα στατιστικό έλεγχο η απόφαση του ερευνητή σχετικά με την απόρριψη της H_0 μπορεί να είναι σωστή ή λανθασμένη. Λανθασμένη απόφαση έχουμε όταν:

α) απορρίπτουμε την H_0 ενώ στην πραγματικότητα είναι αληθής. Λέμε τότε ότι κάνουμε *Σφάλμα Τύπου I* ή *σφάλμα πρώτου είδους*. Η πιθανότητα να διαπράξουμε Σφάλμα Τύπου I συμβολίζεται με α και είναι η δεσμευμένη πιθανότητα:

$$\alpha = P(\text{απόρριψης της } H_0 / H_0 \text{ αληθής}).$$

β) δεν απορρίπτουμε την H_0 ενώ στην πραγματικότητα είναι ψευδής. Λέμε τότε ότι κάνουμε *Σφάλμα Τύπου II* ή *σφάλμα δεύτερου είδους*. Η πιθανότητα να διαπράξουμε Σφάλμα Τύπου II συμβολίζεται με β και είναι η δεσμευμένη πιθανότητα:

$$\beta = P(\text{μη απόρριψης της } H_0 / H_0 \text{ λανθασμένη}).$$

Όταν ελέγχουμε μια H_0 επιλέγουμε ως α μία τιμή που εκφράζει τη μέγιστη πιθανότητα αποδοχής διάπραξης Σφάλματος Τύπου I. Η πιθανότητα αυτή ονομάζεται *επίπεδο* ή *στάθμη σημαντικότητας* (σ.σ.) και είναι απαραίτητο να καθορίζεται από τον ερευνητή πριν από τη δειγματοληψία ή την εκτέλεση ενός πειράματος, ώστε τα αποτελέσματα των στατιστικών αναλύσεων να μην επηρεάσουν την τιμή της (Hinkle et al. 1988, Kachigan 1991, Pagano και Gauvreau 2000, Cohen 1988). Έτσι, η τιμή του α δε θα πρέπει να καθορίζεται μετά από προκαταρκτικές αναλύσεις των δεδομένων, ούτε θα πρέπει να τροποποιείται έτσι ώστε να εξυπηρετεί την απόρριψη ή μη συγκεκριμένων μηδενικών υποθέσεων. Επίσης, η σ.σ. α εκφράζει την πιθανότητα να διαπραχθεί Σφάλμα Τύπου I μόνο όταν: α) οι μετρήσεις είναι *έγκυρες* και *αξιόπιστες* και β) ισχύουν οι προϋποθέσεις εφαρμογής του αντίστοιχου στατιστικού ελέγχου.

Στην πράξη χρησιμοποιούνται παραδοσιακά οι συμβατικές (αυθαίρετες) τιμές $\alpha = 0,10$ ή $\alpha = 0,05$ ή $\alpha = 0,01$ (Hinkle et al. 1988, Kirk 1995, Hopkins 1997, Hair et al. 1995, Huck 2000b). Αν π.χ. για κάποιον έλεγχο καθορίσουμε ως σ.σ. $\alpha = 0,05$ ή 5% και απορρίψουμε τη H_0 , τότε θεωρητικά σε 100 όμοιες περιπτώσεις ή σε 100 επαναλήψεις του πειράματος μόνο σε 5 αναμένεται να σφάλουμε, δηλαδή να απορρίψουμε τη H_0 ενώ στην πραγματικότητα είναι σωστή. Έτσι, φαίνεται ότι η σ.σ. εκφράζει ένα ρυθμό σφάλματος που συνδέεται κυρίως με τη στατιστική διαδικασία και όχι με την τιμή του στατιστικού (π.χ. t , F και χ^2) του ελέγχου (Lohninger, 1999).

Η πιθανότητα να μην απορρίψουμε μία όντως αληθή H_0 καθορίζεται από τη σ.σ. α :

$$1 - \alpha = P(\text{μη απόρριψης της } H_0 / H_0 \text{ αληθής}) \quad (1)$$

Η πιθανότητα γ να απορρίψουμε μία όντως ψευδή H_0 καθορίζεται από το σφάλμα β και ονομάζεται *ισχύς* (*power*) του στατιστικού ελέγχου:

$$\gamma = 1 - \beta = P(\text{απόρριψης της } H_0 / H_0 \text{ λανθασμένη}) \quad (2)$$

Οι σχέσεις (1) και (2) εκφράζουν την πιθανότητα να έχουμε πάρει σωστή απόφαση σε ένα στατιστικό έλεγχο.

Άρα, για να μπορέσουμε, με βάση τα δεδομένα, να καταλήξουμε σε σχετικά ασφαλή και αξιόπιστα συμπεράσματα, θα πρέπει ο στατιστικός έλεγχος να ελαχιστοποιεί τα σφάλματα α και β . Όμως κάθε προσπάθεια μείωσης του ενός κινδύνου αυξάνει τον άλλο (Κολυβά-Μαχαίρα και Μπόρα-Σέντα 1996, Τσάντας κ.ά. 1999). Σε πρακτικό επίπεδο, προσπαθούμε να μειώσουμε το σπουδαιότερο από τους δύο κινδύνους. Ένας τρόπος να μειωθούν και οι δύο κίνδυνοι ταυτόχρονα είναι να αυξήσουμε το μέγεθος του δείγματος (Pagano και Gauvreau 2000, Χάλκος 2000, Zar 1996), πράγμα όχι πάντοτε εφικτό λόγω φυσικών, τεχνικών, οικονομικών, χρονικών και δεοντολογικών περιορισμών.

Όμως, ποιο από τα δύο σφάλματα είναι σημαντικότερο; Η απάντηση είναι σχετική και εξαρτάται από πολλούς παράγοντες, όπως από το γενικό σκοπό και τους ειδικούς στόχους της έρευνας, από το θεωρητικό πλαίσιο, τις γνώσεις του ερευνητή ή από σκοπιμότητες. Σε κάθε περίπτωση, όμως, στην απόφαση απόρριψης ή όχι μίας υπόθεσης πρέπει να υπολογίζονται και να λαμβάνονται υπόψη τόσο το α όσο και το β .

Οι στατιστικοί έλεγχοι περιλαμβάνουν αρκετές συμβάσεις σε ότι αφορά τον προκαθορισμό του α και του β . Πολλοί ερευνητές π.χ. θέτουν το $\alpha \leq 0,05$ και το $\beta \leq 0,20$. Αυτό σημαίνει ότι θεωρούν ως σοβαρότερο τον κίνδυνο να διαπράξουν Σφάλμα Τύπου I απ' ότι να διαπράξουν Σφάλμα Τύπου II. Αν υπολογίσουμε το κλάσμα:

$$P(\text{να διαπραχθεί Σφάλμα Τύπου II}) / P(\text{να διαπραχθεί Σφάλμα Τύπου I}),$$

για $\alpha=0,05$ και $\beta=0,20$, έχουμε, $0,20 / 0,05 = 4$. Ο ερευνητής, δηλαδή, θεωρεί ότι το Σφάλμα Τύπου I είναι 4 φορές πιο σοβαρό, πιο κρίσιμο, απ' ότι το Σφάλμα Τύπου II. Αν το $\beta = 0,20$ τότε η ισχύς είναι $\gamma = 0,80$. Πολλοί ερευνητές θέτουν ως ελάχιστη αποδεκτή ισχύ ενός στατιστικού ελέγχου την τιμή 0,80 και αν ο στατιστικός έλεγχος έχει ισχύ μικρότερη δεν εκτελούν ή ξανασχεδιάζουν την έρευνα (Kirk, 1995).

Σφάλμα Τύπου II ½.

Ιδιαίτερη προσοχή χρειάζεται για την αποφυγή του λανθασμένου συμπεράσματος ότι η αδυναμία του ελέγχου να αποκαλύψει ένα στατιστικά σημαντικό αποτέλεσμα (π.χ. διαφορά, επίδραση, συσχέτιση) σημαίνει ότι η διαφορά, ή η επίδραση, ή η συσχέτιση δεν υπάρχει στους αντίστοιχους πληθυσμούς. Ο εσφαλμένος αυτός συμπερασμός συχνά ονομάζεται *Σφάλμα Τύπου II ½* (Kritzer, 1996). Το σφάλμα αυτό είναι λογικό (Dometrius 1992, Kargopoulos και Raftopoulos 1998) και αφορά την περίπτωση κατά την οποία, σε έναν υποθετικοπαραγωγικό συλλογισμό, θεωρούμε ότι ισχύει η πρόταση του συμπεράσματος.

Η Παρατηρούμενη Στάθμη Σημαντικότητας (p-value).

Ονομάζουμε παρατηρούμενη στάθμη σημαντικότητας (π.σ.σ.) την πιθανότητα να παρατηρηθεί μια τιμή του στατιστικού μεγαλύτερη ή ίση από αυτήν που έδωσε το δείγμα με δεδομένο ότι η H_0 είναι αληθής, δηλαδή $p = P(Z \geq |z'| / H_0 \text{ είναι αληθής})$, όπου Z είναι η τυχαία μεταβλητή που αντιστοιχεί στο στατιστικό και z' η τιμή του στατιστικού για το συγκεκριμένο δείγμα (π.χ. t , F και χ^2).

Η τιμή της π.σ.σ. που στηρίζεται στα δεδομένα αποτελεί τη βάση πάνω στην οποία θα στηριχθεί η απόφασή μας σχετικά με το αν θα απορρίψουμε την H_0 ή όχι. Αν η π.σ.σ. ενός ελέγχου είναι μικρότερη ή το πολύ ίση με τη σ.σ. α που έχουμε προκαθορίσει, τότε απορρίπτουμε την H_0 σε σ.σ. α (Dometrius 1992, Kirk 1995, Pagano και Gauvreau 2000, Kinnear και Gray 1999). Αν η π.σ.σ. είναι μεγαλύτερη από τη σ.σ. α που προκαθορίσαμε, τότε δεν απορρίπτουμε την H_0 . Η π.σ.σ. εκφράζει την πιθανότητα ένα στατιστικό αποτέλεσμα, τόσο μεγάλο ή μεγαλύτερο από το παρατηρούμενο, θα μπορούσε να συμβεί στην "τύχη" αν η H_0 είναι αληθής (Bryman και Cramer 1999). Η τιμή της π.σ.σ. εκφράζει τη χαμηλότερη σ.σ. στην οποία μπορούμε να απορρίψουμε την H_0 (Τσάντας κ.ά. 1999, Χάλκος 2000). Να τονίσουμε ότι σε κάθε περίπτωση το τι ισχύει στην πραγματικότητα σχετικά με την H_0 μας είναι άγνωστο.

Κατά παράδοση, ο έλεγχος υποθέσεων στην επιστημονική έρευνα έχει να επιδείξει μια σαφή προτίμηση στη χρησιμοποίηση της στατιστικής σημαντικότητας ως κριτηρίου απόρριψης ή όχι της μηδενικής υπόθεσης (Τσάντας κ.ά., 1999) με αποτέλεσμα να δοθεί μεγαλύτερη έμφαση στον έλεγχο και στη διαχείριση του Σφάλματος Τύπου I. Όμως, τα τελευταία χρόνια και ιδιαίτερα μετά τις εργασίες του Cohen (1962, 1965) σχετικά με την ανάλυση ισχύος των στατιστικών ελέγχων στις Επιστήμες της Συμπεριφοράς, η προσοχή των ερευνητών αρχίζει να στρέφεται και στον έλεγχο του Σφάλματος Τύπου II και στην αναγκαιότητα ανάλυσης της ισχύος των στατιστικών ελέγχων (Cohen 1988, Murphy και Myers 1998).

Σφάλμα Τύπου III.

Σφάλμα Τύπου III διαπράττεται όταν μία όντως ψευδής H_0 απορρίπτεται αλλά στη συνέχεια η κατεύθυνση του επαγωγικού συλλογισμού είναι αντίθετη από αυτήν που ισχύει στην πραγματικότητα (Huck, 2000b). Για παράδειγμα, σε δύο διαφορετικά υποστρώματα, Perlite και Zeolite, καλλιεργήθηκαν 100 φυτά ζέρμπερας. Οι διαθέσιμες 100 πειραματικές μονάδες τυχαιοποιήθηκαν στις δύο αγωγές με τον περιορισμό κάθε αγωγή να περιλαμβάνει 50 πειραματικές μονάδες. Στο τέλος της περιόδου ο ερευνητής καταγράφει τη συνολική παραγωγή ανθέων και συγκρίνει τις δύο αγωγές.

Ο στατιστικός έλεγχος που (συνήθως) πραγματοποιείται είναι το δίπλευρο t -test για ανεξάρτητα δείγματα σε σ.σ. (συνήθως) $\alpha = 0,05$.

$$\begin{array}{lll} H_0: \mu_z - \mu_p = 0 & & H_0: \mu_z = \mu_p \\ \text{με} & \text{ή ισοδύναμα} & \text{με} \\ H_1: \mu_z - \mu_p \neq 0 & & H_1: \mu_z \neq \mu_p \end{array}$$

Όπου μ_z και μ_p είναι οι μέσοι όροι της συνολικής παραγωγής ανθέων στους δύο πληθυσμούς που αντιστοιχούν στα υποστρώματα Zeolite και Perlite.

Στο παράδειγμα αυτό ας συζητήσουμε τώρα τα Σφάλματα Τύπου I, II και III. Αν στην πραγματικότητα η H_0 είναι αληθής, δηλαδή τα δύο υποστρώματα είναι εξίσου αποτελεσματικά, και με βάση τα διαθέσιμα πειραματικά δεδομένα η H_0 απορριφθεί και ο ερευνητής υποστηρίξει π.χ. ότι ο Perlite είναι καλύτερος από το Zeolite τότε έχει

διαπραχθεί Σφάλμα Τύπου I. Αντίθετα, αν η H_0 στην πραγματικότητα δεν είναι αληθής, γιατί ο Perlite είναι καλύτερος από το Zeolite, και με βάση τα διαθέσιμα πειραματικά δεδομένα η H_0 δεν απορριφθεί τότε έχει διαπραχθεί Σφάλμα Τύπου II.

Έστω, ότι ο Perlite είναι στην πραγματικότητα πιο αποτελεσματικός από τον Zeolite, δηλαδή $\mu_p > \mu_z$. Όμως, είναι δυνατόν λόγω τυχαίων σφαλμάτων ο δειγματικός μέσος όρος που αντιστοιχεί στο Zeolite να είναι μεγαλύτερος από το δειγματικό μέσο όρο που αντιστοιχεί στον Perlite, δηλαδή $M_{zeolite} > M_{perlite}$. Επίσης, είναι δυνατόν η διαφορά μεταξύ των δύο δειγματικών μέσων όρων να είναι επαρκώς μεγάλη, για το δοσμένο μέγεθος δείγματος, με αποτέλεσμα ο στατιστικός έλεγχος να οδηγήσει στην απόρριψη της μηδενικής υπόθεσης. Ας αναλύσουμε τη νέα κατάσταση:

1. Η H_0 είναι στην πραγματικότητα ψευδής αφού $\mu_p > \mu_z$.
2. Τα πειραματικά δεδομένα δείχνουν ότι $M_{zeolite} > M_{perlite}$.
3. Η διαφορά μεταξύ των δύο δειγματικών μέσων όρων $|\delta| = |M_{zeolite} - M_{perlite}|$ είναι αρκετά μεγάλη, για το δοσμένο μέγεθος δείγματος, ώστε ο στατιστικός έλεγχος να οδηγήσει στην απόρριψη της μηδενικής υπόθεσης.
4. Ο ερευνητής, με βάση τα πειραματικά δεδομένα, καταλήγει στο συμπέρασμα ότι η μέση παραγωγή ανθέων στα φυτά που καλλιεργήθηκαν σε Zeolite είναι στατιστικά σημαντικά μεγαλύτερη απ' ό,τι στα φυτά που καλλιεργήθηκαν σε Perlite.

Τα ερωτήματα που τίθενται είναι τα παρακάτω:

1. Έχει διαπραχθεί Σφάλμα Τύπου I; Όχι, αφού η H_0 είναι ψευδής και εξ ορισμού το Σφάλμα Τύπου I γίνεται όταν απορριφθεί μία όντως αληθής H_0 .
2. Έχει διαπραχθεί Σφάλμα Τύπου II; Όχι, αφού έχει απορριφθεί η H_0 και εξ ορισμού το Σφάλμα Τύπου II γίνεται όταν δεν απορριφθεί μία όντως ψευδής H_0 .
3. Έχει ληφθεί η σωστή απόφαση σχετικά με το πιο υπόστρωμα είναι καλύτερο; Όχι, αφού στην πραγματικότητα ο Perlite είναι πιο αποτελεσματικός από το Zeolite και ο ερευνητής κατέληξε στο αντίθετο συμπέρασμα.

Τότε τι πήγε στραβά; Στο προηγούμενο παράδειγμα η απόφαση σχετικά με την απόρριψη της H_0 ήταν σωστή. Όμως, η κατεύθυνση του επαγωγικού συλλογισμού ήταν αντίθετη από αυτήν που ισχύει στην πραγματικότητα. Ο Perlite είναι στην πραγματικότητα πιο αποτελεσματικός από το Zeolite ενώ με βάση τα πειραματικά δεδομένα το συμπέρασμα ήταν ότι ο Zeolite είναι αποτελεσματικότερος του Perlite.

Το Σφάλμα Τύπου III είναι το σημαντικότερο είδος σφάλματος (Langley 1971, Huck 2000b). Με δεδομένο ότι η μη απόρριψη της H_0 δε σημαίνει την αποδοχή της, είναι τουλάχιστον άστοχο να μην απορριφθεί μία όντως λανθασμένη H_0 . Είναι όμως τραγικό να ισχυριστούμε ότι ένα αποτέλεσμα βαίνει προς μία κατεύθυνση ενώ στην πραγματικότητα βαίνει προς την αντίθετη.

Αθροιστικό Σφάλμα Τύπου I και Αθροιστικό Σφάλμα Τύπου II.

Σε περίπτωση που ελέγχονται πολλές υποθέσεις, δηλαδή πραγματοποιείται μεγάλος αριθμός στατιστικών ελέγχων ταυτόχρονα, υπάρχει περίπτωση να εμφανιστούν στατιστικά σημαντικά αποτελέσματα κάτω από την επίδραση τυχαίων και αστάθμητων παραγόντων. Όσο περισσότερο αναζητούμε την ύπαρξη συσχετίσεων, διαφορών και επιδράσεων τόσο πιθανότερο είναι να βρούμε κάτι που στην πραγματικότητα δεν υπάρχει. Σύμφωνα με τον Hopkins (1997), το φαινόμενο αυτό ονομάζεται *Εξάπλωση του Συνολικού Σφάλματος Τύπου I* ή *Αθροιστικό Σφάλμα Τύπου I*. Το Αθροιστικό Σφάλμα Τύπου I εκφράζει την πιθανότητα, σε μία σειρά στατιστικών ελέγχων, να παρατηρηθεί τουλάχιστον ένα στατιστικά σημαντικό αποτέλεσμα που δεν υπάρχει στην πραγματικότητα. Έτσι, αν π.χ. θέλουμε να εμφανίσουμε στατιστικά σημαντικές συσχετίσεις μεταξύ πολλών μεταβλητών και επιθυμούμε να πείσουμε τον αναγνώστη για τη στατιστική σημαντικότητα των αποτελεσμάτων πρέπει η στάθμη σημαντικότητας α για κάθε έλεγχο να διορθωθεί κατάλληλα προς τα κάτω (Huck, 2000b).

Η πιο απλή διόρθωση που μπορεί να γίνει είναι αυτή που ονομάζεται *Διόρθωση κατά Bonferroni*. Σύμφωνα με τη μέθοδο αυτή η αρχική σ.σ. α διαιρείται δια του αριθμού των στατιστικών ελέγχων που πραγματοποιήθηκαν (Τσάντας κ.ά. 1999, Περσίδης 1997, Coakes και Steed 1999, Bryman και Cramer 1999, Hopkins 1997, Klockars και Sax 1986, Girden 1992, Brown και Melamed 1990, Kirk 1995, Huck 2000b). Ορισμένοι δε συμφωνούν με τη διόρθωση αυτή (Hopkins, 1997) ή, όπως αναφέρει ο Περσίδης (1997), την αντιμετωπίζουν με σκεπτικισμό γιατί: α) με τη διόρθωση κατά Bonferroni οι στατιστικοί έλεγχοι γίνονται πιο συντηρητικοί με αποτέλεσμα να ανιχνεύονται λιγότερα στατιστικά σημαντικά αποτελέσματα (Περσίδης, 1997) και β) με την

ελάττωση της σ.σ. σε κάθε έλεγχο, δηλαδή της πιθανότητας να διαπραχθεί Σφάλμα Τύπου I, αυξάνεται ο κίνδυνος να διαπραχθεί Σφάλμα Τύπου II και συνεπώς ελαττώνεται η ισχύς γ του ελέγχου.

Αντίστοιχα, μπορεί να οριστεί και το *Αθροιστικό Σφάλμα Τύπου II*, το οποίο εκφράζει την πιθανότητα, σε μία σειρά στατιστικών ελέγχων, ένα τουλάχιστον αποτέλεσμα, το οποίο υπάρχει στην πραγματικότητα, να μην ανιχνευθεί ως στατιστικά σημαντικό.

Είναι φανερό, κάτω από αυτές τις θεωρήσεις, ότι η επαγωγική συμπερασματολογία αρχίζει και γίνεται εξαιρετικά επισφαλής.

ΑΝΤΙ ΣΥΜΠΕΡΑΣΜΑΤΟΣ

Στους στατιστικούς ελέγχους υποθέσεων, οι λανθασμένες αντιλήψεις των ερευνητών τους οδηγούν με βεβαιότητα σε λανθασμένες αποφάσεις και λανθασμένα συμπεράσματα. Όμως οι λανθασμένες αντιλήψεις τους δεν είναι η μοναδική πηγή λαθών. Διαπράττουν επίσης και «στοχαστικά» λάθη όπως επίσης και λογικά λάθη.

Στον Πίνακα 1 παρουσιάζουμε μια απλοποιημένη και συνοπτική εκδοχή των επαγωγικών συλλογισμών και των αντίστοιχων Σφαλμάτων I, II, II $\frac{1}{2}$ και III.

Πίνακας 1. Είδη Σφαλμάτων και Αντίστοιχος Επαγωγικός Συλλογισμός.

Διαπράττουμε ...	Όταν ...
Σφάλμα Τύπου I	ισχυριζόμαστε ότι έχουμε βρει κάτι το οποίο στην πραγματικότητα δεν υπάρχει.
Σφάλμα Τύπου II	αποτυγχάνουμε να βρούμε κάτι το οποίο όμως στην πραγματικότητα υπάρχει.
Σφάλμα Τύπου II $\frac{1}{2}$	ισχυριζόμαστε ότι εφόσον δε βρήκαμε κάτι τότε αυτό δεν υπάρχει και στην πραγματικότητα. Ερμηνεύουμε μια μη στατιστικά σημαντική διαφορά ως "καθόλου σημαντική διαφορά" οριστικά.
Σφάλμα Τύπου III	η απόφασή μας σχετικά με την απόρριψη της μηδενικής υπόθεσης είναι σωστή αλλά το τελικό συμπέρασμα είναι προς λάθος κατεύθυνση σε ότι αφορά στην ερευνητική υπόθεση. Με άλλα λόγια, παίρνουμε μια σωστή απόφαση αλλά για μια λάθος ερώτηση.

Ένας έμπειρος στατιστικός μπορεί να σχεδιάσει μια έρευνα με τρόπο ώστε να εξισορροπήσει κατάλληλα τα Σφάλματα Τύπου I και Τύπου II. Μπορεί, δηλαδή, με κατάλληλη επιλογή των α , β και του μεγέθους δείγματος, να προσανατολίσει τα συμπεράσματα προς ορισμένες «επιθυμητές» κατευθύνσεις.

Επίσης, αν ο ερευνητής στηρίζει τις αποφάσεις του και την κρίση του μόνο στους στατιστικούς ελέγχους υποθέσεων και στην προσέγγιση της στατιστικής σημαντικότητας τότε θα πρέπει, στην περίπτωση που πραγματοποιεί ταυτόχρονα μεγάλο αριθμό στατιστικών ελέγχων, να λάβει σοβαρά υπόψη του και τον κίνδυνο για λανθασμένη συμπερασματολογία λόγω των Αθροιστικών Σφαλμάτων Τύπου I και II.

Όλες οι παραπάνω θεωρήσεις φανερώνουν ότι η επαγωγική συμπερασματολογία στον έλεγχο υποθέσεων είναι εξαιρετικά επισφαλής και θα συμφωνήσουμε με τον Gras (1995) ότι *«χρειάζεται, κατά τη διερεύνηση της εγκυρότητας υποθέσεων, να βρεθεί μια σωστή ισορροπία ανάμεσα στην απλοϊκή χρήση των στατιστικών μεθόδων, την άρνηση επένδυσης σε αυτό το πεδίο και την στατιστικομανία που οδηγεί σε μια πληθώρα ανεξερευνήτων αποτελεσμάτων, που συνοδεύονται από μια ψευδαίσθηση διαφάνειας»* (σ.98) και μια έλλειψη επιστημοσύνης και αντικειμενικότητας, θα συμπληρώναμε.

Πράγματι, χρειάζεται μια μεγάλη προσπάθεια στοχαστικού αλφαριθμητισμού που θα αρχίσει με παρεμβάσεις στα προγράμματα της βασικής εκπαίδευσης αλλά και με την αλλαγή του περιεχομένου των μαθημάτων στα Πανεπιστήμια, όπου η έμφαση πρέπει να δοθεί κυρίως στην επιστημολογία των στοχαστικών μεθόδων. Η κλασική Στατιστική θα πρέπει να διδάσκεται ως μία μόνο εκδοχή στη στατιστική επεξεργασία δεδομένων. Υπάρχει ανάγκη να υιοθετηθούν και άλλες προσεγγίσεις. Για παράδειγμα, οι μέθοδοι της Ανάλυσης Δεδομένων είναι κατάλληλες για την ανάλυση ποιοτικών δεδομένων και μπορούν να αναδείξουν *"απρόβλεπτες διαστάσεις και να δημιουργήσουν νέες θεωρητικές προσεγγίσεις και προεκτάσεις"* κατά τη διερεύνηση ενός φαινομένου (Αναστασιάδου και Παπαδημητρίου, 2001, σ.327). Πράγματι, οι μέθοδοι αυτοί:

- Βοηθούν τον αναλυτή να αναδείξει, να περιγράψει και να ερμηνεύσει, με σαφή και κομψό τρόπο, τις αλληλεπιδράσεις μεταξύ ποιοτικών τυχαίων μεταβλητών.
- Δίνουν ιδιαίτερη έμφαση στη γεωμετρική θεώρηση της ερμηνείας των δεδομένων με βασικό σκοπό την ανάδειξη της ενδογενούς δομής που τα χαρακτηρίζει, η οποία, συνήθως, δεν είναι άμεσα αντιληπτή, αλλά βρίσκεται σε λανθάνουσα μορφή.
- Περιγράφουν καθολικά το φαινόμενο με τρόπο που επιτρέπει τη γραφική απεικόνιση των μεταβλητών, δηλαδή την οπτικοποίηση των αλληλεπιδράσεων και των σχέσεων τους, στον ίδιο χώρο (Αναστασιάδου και Παπαδημητρίου, 2001).

- Δεν απαιτούν την a priori παραδοχή ύπαρξης κάποιας θεωρητικής κατανομής ή κάποια υπόθεση σχετικά με τις παραμέτρους του υπό εξέταση πληθυσμού ή πληθυσμών (Clausen 1998, Καραπιστόλης 1999).
- Εκφράζουν μια εναλλακτική μεθοδολογική προσέγγιση αλλά και μια νέα φιλοσοφική θεώρηση της στατιστικής συμπερασματολογίας, η οποία έρχεται σε αντίθεση με την κλασική αγγλοσαξωνική παράδοση του στατιστικού ελέγχου υποθέσεων για την ανάλυση και ερμηνεία κατηγορικών δεδομένων (Gras, 1995).

Τελικά, στις στοχαστικές επιχειρήσεις τα λάθη ελοχεύουν παντού. Τα βήματα των ερευνητών είναι μετέωρα όπως και τα συμπεράσματά τους. Η μόνη βεβαιότητα που μπορούν να έχουν είναι η σχετικότητα των συμπερασμάτων τους και η μεγάλη, τελικά, πιθανότητα να είναι λανθασμένα. Μπορεί, ανά πάσα στιγμή να πάρουν τη θέση του Χριστόφορου Κολόμβου που με την κατάλληλη μέθοδο αλλά λανθασμένο συλλογισμό έφτασε σε μια ήπειρο άλλη από αυτή που ήθελε αλλά η μόνη στην οποία θα μπορούσε να πάει, λάθος που δεν έμαθε ποτέ ότι διέπραξε.

ABSTRACT

The hypotheses testing in science is traditionally considered to be a matter of measuring the statistical significance and using it as a criterion for deciding whether the null-hypothesis should be rejected or not. The decision of the researcher to discard the null-hypothesis could be correct or wrong. Wrong decisions can result from stochastic or logical errors or even from misconceptions. For this reason the usage of the criterion of statistical significance for the rejection or not of the null-hypothesis has been the target of criticism since the early sixties and is still the target of periodical criticisms. In this paper we discuss some issues concerning the evaluation of the results of statistical hypotheses testing with a view to suggest ways of avoiding the so-called statistical consumerism (p-value). These issues lay the foundation for an epistemological revision of information, construed as knowledge, which results from the statistical hypotheses testing. Data-Analysis and its methods could contribute toward this aim.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Αναστασιάδου, Σ., και Παπαδημητρίου, Γ., (2001). Χρήση Μεθόδων της Πολυδιάστατης Στατιστικής Ανάλυσης για τον Προσδιορισμό των Διαθέσεων των Φοιτητών προς τη Στατιστική. *Πρακτικά Τέταρτου Παγκόπιου Συνεδρίου Μαθηματικής Παιδείας και Συμπόσιο Αστροναυτικής και Διαστήματος*. Λάρνακα, σ.327-335.
- Brown, S., and Melamed, L., (1990). *Experimental Design and Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-074. Newbury Park, CA: Sage.
- Bryman, A., and Cramer, D., (1999). *Quantitative Data Analysis with SPSS Release 8 for Windows: A Guide for Social Scientists*. London and New York: Routledge.
- Carver, P., (1978). The case against statistical testing. *Harvard Educational Review*, 48, 378-399.
- Clausen, S.-E., (1998). *Applied Correspondence Analysis: An Introduction*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-121. Thousand Oakes, CA: Sage.
- Coakes, S., and Steed, L., (1999). *SPSS Analysis without Anguish*. Singapore: John Willey & Sons, Inc.

- Cohen, J., (1962). The Statistical Power of Abnormal-Social Psychological Research: A Review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J., (1965). Some Statistical Issues in Psychological Research. In B. B. Wolman (Ed.), *Handbook of Clinical Psychology*. New York: McGraw-Hill, p.p. 95-121.
- Cohen, J., (1988). *Statistical Power Analysis for the Behavioral Sciences*. New Jersey: Lawrence Erlbaum Associates, Inc.
- DeBono, E., (2000). *New Thinking for the New Millennium*, London: Penguin Books.
- Dometrius, N., (1992). *Social Statistics Using SPSS*. New York: HarperCollins Publishers, Inc.
- Girden, E., (1992). *ANOVA Repeated Measures*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-084. Newbury Park, CA: Sage.
- Gras, R., (1995). Ανάλυση ενός Ερωτηματολογίου με τη Συνεπαγωγική Μέθοδο. Στο Α. Γαγάτσης (Ed.), *Διδακτική και Ιστορία των Μαθηματικών*. Θεσσαλονίκη: ERASMUS ICP-94-G-2011/11, σ.97-109.
- Hair, J., Anderson, R., Tatham, R. and Black, W., (1995). *Multivariate Data Analysis With Readings*. Prentice-Hall International, Inc.
- Hinkle, D., Wiersma, W., and Jurs, S., (1988). *Applied Statistics for the Behavioral Sciences*. Boston: Houghton Mifflin Company.
- Hopkins, W., (1997). *A New View of Statistics*. <http://www.sportsci.org/resource/stats/index.html>.
- Huck, S., (2000a). Misconceptions. In *RSR: Reading Statistics & Research-Student Help*, Chapter 9. <http://www.readingstats.com>.
- Huck, S., (2000b). *Reading Statistics and Research*. Addison Wesley Longman, Inc.
- Καραπιστόλης, Δ., (1999). *Ανάλυση Δεδομένων και Έρευνα Αγοράς*. Θεσ/νίκη: Εκδόσεις ANIKOYΛA.
- Kargopoulos P., and Raftopoulos T., (1998). *The Science of Logic & The Art of Thinking*. Thessaloniki: Vanias Publishing House.
- Kachigan, S., (1991). *Multivariate Statistical Analysis: A Conceptual Introduction*. NY: Radius Press.
- Kinnear, P., and Gray, C., (1999). *SPSS for Windows Made Simple*. East Sussex: Psychology Press Ltd.
- Κολυβά-Μαχαίρα, Φ., και Μπόρα-Σέντα, Ε., (1996). *Στατιστική: Θεωρία Εφαρμογές*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.
- Kirk, R., (1995). *Experimental Design: Procedures for the Behavioral Sciences*. Pacific Grove, CA: Brooks/Cole Publishing Company, ITP.
- Klockars A., and Sax, G., (1986). *Multiple Comparisons*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-061. Newbury Park, CA: Sage.
- Kritzer, B., (1996). *Surviving Statistical Spitting Matches*. A Professional Development Seminar presentation for Senior Staff of the National Conference of State Legislatures, Madison, Wisconsin, October 10, 1996. <http://www.polisci.wisc.edu/~kritzer/misc/legstaff/legstaff.htm>
- Langley, R., (1971). *Practical Statistics Simply Explained*. New York: Dover Publications, Inc.
- Lohninger, H., (1999). *Teach Me Data Analysis: Single User Edition*, [Computer program manual]. New York: Springer.
- Murphy, K., and Myers, B., (1998). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Pagano, M. and Gauvreau, K., (2000). *Αρχές Βιοστατιστικής*. Περιστέρι-Αθήνα: Εκδόσεις ΕΛΛΗΝ.
- Περσιδής, Δ., (1997). *Εφαρμοσμένη Στατιστική στην Τεχνολογία Τροφίμων*. Θεσσαλονίκη: Εκδοτική ΟΜΗΡΟΣ.
- Schmidt, L., (1996). Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Τσάντας, Ν., Μωϋσιάδης, Χ., Μπαγιάτης, Ν., και Χατζηπαντελής, Θ., (1999). *Ανάλυση Δεδομένων με τη βοήθεια Στατιστικών Πακέτων*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.
- Χάλκος, Γ., (2000). *Στατιστική Θεωρία Εφαρμογές & Χρήση Στατιστικών Προγραμμάτων σε Η/Υ*. Αθήνα: Τυπωθήτω - ΓΙΩΡΓΟΣ ΔΑΡΔΑΝΟΣ.
- Zar, J., (1996). *Biostatistical Analysis*. Prentice-Hall International, Inc.