



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Στατιστική ανάλυση δεδομένων που αφορούν την απήχηση του
επιστημονικού έργου
των Τμημάτων Πληροφορικής των ελληνικών Πανεπιστημίων**

Ελένη Τσολάκου

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων

Ιωάννης Σ. Τριανταφύλλου
Λέκτορας Πανεπιστημίου Θεσσαλίας

Λαμία, 2016



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Στατιστική ανάλυση δεδομένων που αφορούν την απήχηση του
επιστημονικού έργου
των Τμημάτων Πληροφορικής των ελληνικών Πανεπιστημίων**

Ελένη Τσολάκου

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων

Ιωάννης Σ. Τριανταφύλλου
Λέκτορας Πανεπιστημίου Θεσσαλίας

Λαμία, 2016

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια.
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

**Στατιστική ανάλυση δεδομένων που αφορούν την απήχηση του
επιστημονικού έργου
των Τμημάτων Πληροφορικής των ελληνικών Πανεπιστημίων**

Ελένη Τσολάκου

Τριμελής Επιτροπή:

Ιωάννης Τριανταφύλλου, Λέκτορας (επιβλέπων)

Μαρία Αδάμ, Επίκουρος Καθηγήτης

Βασίλειος Δρακόπουλος, Επίκουρος Καθηγητής

Περιεχόμενα

Λίστα Πινάκων.....	2
Λίστα Γραφημάτων	3
Λίστα Εικόνων	4
Περίληψη	5
ΚΕΦΑΛΑΙΟ 1	6
ΒΙΒΛΙΟΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ.....	6
1.1 Εισαγωγή.....	6
1.2 Βιβλιομετρικές μέθοδοι και δείκτες απήχησης	9
1.2.1 Κανονικοποιημένη βάσει πηγής δεδομένων απήχηση ανά άρθρο (<i>Source Normalized Impact per Paper (SNIP)</i>)	9
1.2.2 Απήχηση ανά δημοσίευση (<i>Impact Per Publication (IPP)</i>).....	10
1.2.3 Βαθμίδα περιοδικού SCImago (SCImago Journal Rank (SJR)).....	11
1.2.4 Παράγοντας απήχησης (<i>Impact Factor (IF)</i>)	13
1.2.5 Δείκτης αμεσότητας περιοδικού (<i>Journal Immediacy Index</i>)	14
1.2.6 Δείκτης Journal Cited Half-life.....	15
1.2.7 Δείκτης επιστημονικής ποιότητας <i>h</i> (<i>h-index</i>).....	15
1.2.8 Δείκτης επιστημονικής ποιότητας <i>p</i> (<i>p-index</i>).....	17
1.2.9 Μετρικές του Μελετητή της Google (<i>Google Scholar Metrics</i>)	18
1.2.10 Ιδιοπαράγοντας (<i>Eigenfactor</i>)	19
1.2.11 Επιρροή άρθρου (<i>Article Influence (AI)</i>).....	20
ΚΕΦΑΛΑΙΟ 2	21
ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ.....	21
2.1 Εισαγωγή.....	21
2.2 Περιγραφική στατιστική	21
2.2.1 Μέτρα κεντρικής τάσης	22
2.2.2 Μέτρα μεταβλητότητας	23
2.2.3 Μέτρα ασυμμετρίας.....	24
2.2.4 Μέτρα κυρτότητας.....	24
2.3 Παραμετρικοί και μη παραμετρικοί έλεγχοι υποθέσεων	25
2.3.1 Το <i>t</i> στατιστικό τεστ ενός δείγματος (παραμετρικός έλεγχος)	25

2.3.2 Το t στατιστικό τεστ ανεξάρτητων δειγμάτων (παραμετρικός έλεγχος)	26
2.3.3 Το t στατιστικό τεστ εξαρτημένων δειγμάτων (παραμετρικός έλεγχος)	26
2.3.4 Το χ^2 στατιστικό τεστ (μη παραμετρικός έλεγχος)	27
2.3.5 Το Kolmogorov-Smirnov τεστ (μη παραμετρικός έλεγχος)	28
2.4 Ανάλυση παλινδρόμησης	31
2.4.1 Ευθεία παλινδρόμησης	31
2.4.2 Ο συντελεστής προσδιορισμού	33
2.5 Ανάλυση Διακύμανσης	36
2.5.1 Γενικά	36
2.5.2 Ανάλυση διακύμανσης προς έναν παράγοντα (<i>One-Way ANOVA</i>)	37
ΚΕΦΑΛΑΙΟ 3	42
ΑΠΟΤΕΛΕΣΜΑΤΑ	42
3.1 Εισαγωγή	42
3.2 Δείγμα και καταγραφή δεδομένων	42
3.3 Σύγκριση σε μη ατομικό επίπεδο βάσει του δείκτη h_1	44
3.3.1 Συντελεστές συσχέτισης Spearman	44
3.3.2 Εξάρτηση δεικτών από κατηγορικά χαρακτηριστικά των Τμημάτων	45
3.3.3 Πρότυπο Γραμμικής Παλινδρόμησης για την πρόβλεψη του δείκτη h_1	45
3.3.4 Περιγραφικά στοιχεία για τους δείκτες	49
3.3.5 Περιγραφικά στοιχεία για τις λοιπές μεταβλητές	50
3.3.6 Ταξινόμηση των τμημάτων ανάλογα με την τιμή του δείκτη h_1	57
3.3.7 Έλεγχος για το αν οι σχηματιζόμενες ομάδες (βάσει του δείκτη h_1) διαφέρουν και ως προς άλλα χαρακτηριστικά	60
3.4 Σύγκριση σε ατομικό επίπεδο βάσει του δείκτη h_i	62
3.4.1 Γραφήματα	62
3.4.2 Περιγραφικά στατιστικά μέτρα για τους δείκτες <i>Publications</i> , <i>Citations</i> , <i>h-index</i> , <i>Citations per publication</i> και <i>p-index</i>	67
3.4.3 Σύνδεση age_Uni, age_Dep, size, location με τους δείκτες <i>Publications</i> , <i>Citations</i> , <i>h-index</i> , <i>Citations per publication</i> , <i>p-index</i>	72
3.4.4 Πρόσθετα γραφήματα και αριθμητικά στοιχεία	75
3.4.5 Αριθμητικές συγκρίσεις μεταξύ των Τμημάτων	76
3.4.6 Στατιστικές Συγκρίσεις μεταξύ των Τμημάτων	80

ΚΕΦΑΛΑΙΟ 4	87
ΣΥΜΠΕΡΑΣΜΑΤΑ	87
ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ	92
ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ	93
ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ	94
ΕΥΡΕΤΗΡΙΟ ΟΡΩΝ	95

Λίστα Πινάκων

Πίνακας 1: Πίνακας ανάλυσης διακύμανσης κατά έναν παράγοντα	40
Πίνακας 2: Συντελεστής συσχέτισης Spearman δείκτη h_1	44
Πίνακας 3: Περιγραφικά στατιστικά μέτρα για τον δείκτη h_1	45
Πίνακας 4: Ανάλυση διακύμανσης για τον δείκτη h_1	45
Πίνακας 5: Βηματική γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή τον δείκτη h_1	46
Πίνακας 6: Γραμμικό μοντέλο για τις μεταβλητές ratio και size	46
Πίνακας 7	47
Πίνακας 8: Kolmogorov-Smirnov τεστ	47
Πίνακας 9: Runs τεστ	48
Πίνακας 10: Τεστ ομογένειας για τη διασπορά	48
Πίνακας 11: Περιγραφικά στατιστικά μέτρα για τον δείκτη h_1	49
Πίνακας 12: Περιγραφικά στατιστικά μέτρα για τις λοιπές μεταβλητές	50
Πίνακας 13: Ομαδοποίηση Τμημάτων σε 2 ομάδες βάσει του δείκτη h_1	58
Πίνακας 14: Στατιστικά μέτρα των δύο ομάδων ως προς τον δείκτη h_1	58
Πίνακας 15: Independent samples t-test για τις σχηματιζόμενες ομάδες	58
Πίνακας 16: Ομαδοποίηση Τμημάτων σε 3 ομάδες βάσει του δείκτη h_1	59
Πίνακας 17: Στατιστικά μέτρα των τριών ομάδων ως προς τον δείκτη h_1	59
Πίνακας 18: Ανάλυση διακύμανσης για τις τρεις ομάδες	60
Πίνακας 19: Independent samples t-test βάση του δείκτη h_1 για τις δύο ομάδες	60
Πίνακας 20: Ανάλυση διακύμανσης βάση του δείκτη h_1 για τις τρεις ομάδες	61
Πίνακας 21: Πολλαπλές συγκρίσεις μεταξύ των τριών ομάδων	62
Πίνακας 22: Περιγραφικά στατιστικά μέτρα για τους λοιπούς δείκτες	67
Πίνακας 23: Συντελεστής συσχέτισης Spearman για την μεταβλητή age_uni	72
Πίνακας 24: Συντελεστής συσχέτισης Spearman για την μεταβλητή age_dep	73
Πίνακας 25: Συντελεστής συσχέτισης Spearman για την μεταβλητή size	73
Πίνακας 26: Ανάλυση Διακύμανσης κατά έναν παράγοντα για την μεταβλητή location	74
Πίνακας 27: Αριθμός δημοσιεύσεων και αναφορών ανά Τμήμα	76
Πίνακας 28: Μέση τιμή (και τυπική απόκλιση) αριθμού δημοσιεύσεων και αναφορών ανά μέλος Δ.Ε.Π για κάθε Τμήμα	77
Πίνακας 29: Μέση τιμή και τυπική απόκλιση του αριθμού των δημοσιεύσεων για τα 11 Τμήματα	77
Πίνακας 30: Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών για τα 11 Τμήματα	78

Πίνακας 31: Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών ανά δημοσίευση και των δεικτών <i>h-index</i> και <i>p-index</i> για κάθε Τμήμα	78
Πίνακας 32: : Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών ανά δημοσίευση για τα 11 Τμήματα	79
Πίνακας 33: : Μέση τιμή και τυπική απόκλιση του δείκτη <i>p-index</i> για τα 11 Τμήματα	79
Πίνακας 34: : Μέση τιμή και τυπική απόκλιση του δείκτη <i>h-index</i> για τα 11 Τμήματα	80
Πίνακας 35: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των δημοσιεύσεων	80
Πίνακας 36: Συγκριτικός πίνακας ως προς τον αριθμό των δημοσιεύσεων	81
Πίνακας 37: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των αναφορών	82
Πίνακας 38: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των αναφορών ανά δημοσίευση.....	82
Πίνακας 39: Συγκριτικός πίνακας ως προς τον αριθμό των αναφορών ανά δημοσίευση	83
Πίνακας 40: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον δείκτη <i>h-index</i>	84
Πίνακας 41: Συγκριτικός πίνακας ως προς τον αριθμό τον δείκτη <i>h-index</i>	84
Πίνακας 42: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον δείκτη <i>p-index</i>	85
Πίνακας 43: Συγκριτικός πίνακας ως προς τον αριθμό τον δείκτη <i>p-index</i>	86

Λίστα Γραφημάτων

Γράφημα 1: Κυκλικό διάγραμμα βάσει της τοποθεσίας των Τμημάτων.....	56
Γράφημα 2: Θηκόγραμμα για τις τιμές του δείκτη h_1 βάσει και της τοποθεσίας των Τμημάτων.....	57
Γράφημα 3: Ιστόγραμμα συχνοτήτων για τον δείκτη publications	63
Γράφημα 4: Ιστόγραμμα συχνοτήτων για τον δείκτη citations	63
Γράφημα 5: Ιστόγραμμα συχνοτήτων για τον δείκτη <i>h-index</i>	63
Γράφημα 6: Ιστόγραμμα συχνοτήτων για τον δείκτη citations_per_publication	64
Γράφημα 7: Ιστόγραμμα συχνοτήτων για τον δείκτη <i>p-index</i>	64
Γράφημα 8: Θηκόγραμμα για τον δείκτη publications.....	65
Γράφημα 9: Θηκόγραμμα για τον δείκτη citations	65
Γράφημα 10: Θηκόγραμμα για τον δείκτη <i>h-index</i>	65
Γράφημα 11: Θηκόγραμμα για τον δείκτη citations_per_publication.....	66
Γράφημα 12: Θηκόγραμμα για τον δείκτη <i>p-index</i>	66

Γράφημα 13: Απόλυτος αριθμός δημοσιεύσεων ανά Τμήμα 75

Γράφημα 14: Απόλυτος αριθμός αναφορών ανά Τμήμα 75

Λίστα Εικόνων

Εικόνα 1 : Παράδειγμα υπολογισμού δείκτη h-index 17

Περίληψη

Σκοπός της παρούσας πτυχιακής εργασίας είναι η στατιστική ανάλυση δεδομένων που αφορούν την απήχηση του επιστημονικού έργου των Τμημάτων Πληροφορικής και Επιστήμης Υπολογιστών των ελληνικών Πανεπιστημίων. Πιο συγκεκριμένα, πραγματοποιήθηκε συγκριτική αξιολόγηση: α) σε ατομικό επίπεδο, μεταξύ των μελών Δ.Ε.Π του εκάστοτε τμήματος Πληροφορικής ή Επιστήμης Υπολογιστών και β) σε μη ατομικό επίπεδο, μεταξύ των Τμημάτων Πληροφορικής και Επιστήμης Υπολογιστών λαμβάνοντας υπ' όψιν το συνολικό επιστημονικό έργο όλων των μελών Δ.Ε.Π κάθε τμήματος.

Για την πραγματοποίηση αυτής της συγκριτικής αξιολόγησης, χρησιμοποιήθηκαν βιβλιομετρικές και στατιστικές μέθοδοι, καθώς και βιβλιομετρικοί δείκτες, με τη βοήθεια των οποίων έγινε η ανάλυση των δεδομένων που αφορούν την απήχηση του επιστημονικού έργου και διεξήχθησαν αποτελέσματα.

Στο πρώτο κεφάλαιο γίνεται αναφορά σε βασικές βιβλιομετρικές μεθόδους και βιβλιομετρικούς δείκτες που αποτελούν εργαλεία για τη στατιστική μελέτη του συνόλου δεδομένων. Στην παρούσα πτυχιακή χρησιμοποιήθηκαν εκτενώς οι δείκτες επιστημονικής ποιότητας p (p -index) και h (h -index) για σύγκριση σε ατομικό επίπεδο, καθώς και μια επέκταση του δείκτη h για σύγκριση σε μη ατομικό επίπεδο, ο h_1 (h_1 -index).

Στο δεύτερο κεφάλαιο γίνεται εισαγωγή στους βασικότερους ορισμούς και έννοιες της Στατιστικής Ανάλυσης και περιγράφονται διάφορες στατιστικές μέθοδοι εκ των οποίων οι περισσότερες εφαρμόζονται στο τρίτο κεφάλαιο με σκοπό την εξαγωγή των αποτελεσμάτων. Οι βασικότερες μέθοδοι που εφαρμόστηκαν είναι : Kolmogorov-Smirnov τεστ για μη παραμετρικό έλεγχο, t - στατιστικό τεστ ανεξάρτητων δειγμάτων για παραμετρικό έλεγχο, One-Way ANOVA για ανάλυση διακύμανσης προς έναν παράγοντα και αλγόριθμος k -means για ταξινόμηση. Εκτός των άλλων, χρησιμοποιήθηκαν και ο συντελεστής γραμμικής συσχέτισης Spearman και το πρότυπο Γραμμικής Παλινδρόμησης για την πρόβλεψη του δείκτη h_1 .

Στο τρίτο κεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν μετά την εφαρμογή των βιβλιομετρικών και στατιστικών μεθόδων που αναφέρθηκαν παραπάνω. Η εφαρμογή των μεθόδων αυτών βασίστηκε κυρίως στην ανάλυση του δείκτη h_1 -index σε ατομικό και μη ατομικό επίπεδο.

Τέλος, στο τέταρτο κεφάλαιο, μετά από αξιολόγηση των αποτελεσμάτων που προέκυψαν κατά τη διάρκεια της έρευνας, γίνεται παρουσίαση των συμπερασμάτων και προτείνονται ιδέες για μελλοντική χρήση της έρευνας αυτής.

Λέξεις κλειδιά: βιβλιομετρική ανάλυση, δείκτης h_1 -index, σύγκριση, δείκτες απήχησης επιστημονικού έργου, ελληνικά Πανεπιστήμια, Πληροφορική, Επιστήμη Υπολογιστών

ΚΕΦΑΛΑΙΟ 1

ΒΙΒΛΙΟΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ

1.1 Εισαγωγή

Τα τελευταία χρόνια, λόγω της ραγδαίας εξέλιξης της τεχνολογίας και κατ' επέκταση της χρήσης του διαδικτύου το πλήθος των διαδικτυακών επιστημονικών περιοδικών αλλά και των έντυπων επιστημονικών περιοδικών αυξήθηκε. Για τον λόγο αυτό, οι ποσοτικές βιβλιομετρικές μέθοδοι που χρησιμοποιούνται για την αξιολόγηση του ερευνητικού έργου σε εθνική, σε παγκόσμια ή ακόμα και σε ιδρυματική ή ατομική κλίμακα παρουσιάζουν όλο και μεγαλύτερη διάδοση. Οι ποσοτικές αυτές μέθοδοι χρησιμοποιούνται στον κλάδο της βιβλιομετρικής ανάλυσης. Πιο συγκεκριμένα, ο κλάδος αυτός ασχολείται με τη συλλογή, την καταγραφή και την επεξεργασία δεδομένων που αφορούν επιστημονικά δημοσιεύματα, όπως το πλήθος των δημοσιεύσεων, οι βιβλιογραφικές αναφορές των δημοσιεύσεων, οι αναφορές σε αυτές από άλλες δημοσιεύσεις (*citations*), η κατανομή τους ανά συγγραφέα, φορέα, επιστημονικό πεδίο, χώρα κ.λπ. Επιπλέον, στόχος της βιβλιομετρικής ανάλυσης είναι η εξαγωγή σημαντικών βιβλιομετρικών δεικτών με σκοπό την αξιολόγηση ερευνητικών και επιστημονικών δημοσιεύσεων.

Εκτός από τα παραπάνω, ο κλάδος της βιβλιομετρικής ανάλυσης έχει ως στόχο την εκτίμηση της απήχησης και της πρωτοτυπίας των επιστημονικών έργων που προέρχονται από τα διάφορα επιστημονικά πεδία της ερευνητικής κοινότητας. Συνεπώς, με την εφαρμογή βιβλιομετρικής ανάλυσης και κατ' επέκταση με τη χρήση βιβλιομετρικών δεικτών είναι δυνατόν να διαμορφωθεί μια κατά το δυνατόν αντικειμενική εικόνα για τη ροή της πληροφορίας μεταξύ επιστημονικών πεδίων ή τομέων.

Οι παραπάνω διαδικασίες έχουν τη δυνατότητα να εφαρμοστούν είτε σε επίπεδο τομέα, είτε σε επίπεδο μιας ομάδας ερευνητών ή επιστημόνων είτε σε ατομικό επίπεδο για κάποιον μεμονωμένο ερευνητή είτε σε επίπεδο χώρας ή ομάδας χωρών (π.χ. χώρες εντός Ευρωπαϊκής Ένωσης). Αυτό σημαίνει πως μπορούν εύκολα να εξαχθούν συμπεράσματα για την απήχηση του επιστημονικού έργου και να αξιολογηθεί κατάλληλα η εκάστοτε επιστημονική δραστηριότητα.

Παρόλα αυτά, έχουν αναφερθεί κατά καιρούς στη διεθνή βιβλιογραφία προβλήματα που σχετίζονται με τον "καθαρισμό" πρωτογενών δεδομένων όπως αυτά συλλέγονται καθώς επίσης και περιορισμοί που πρέπει οπωσδήποτε να λαμβάνονται υπόψη στον υπολογισμό των βιβλιομετρικών δεικτών και ειδικά στον υπολογισμό και τη χρήση του αριθμού των αναφορών. Οι περιορισμοί αυτοί αφορούν κυρίως το επιστημονικό πεδίο των αναφορών, το χρονικό διάστημα ανάλυσης αναφορών και το είδος των επιστημονικών δημοσιεύσεων [29].

Οι μέθοδοι που χρησιμοποιούνται κυρίως κατά τη βιβλιομετρική ανάλυση χωρίζονται σε ποσοτικές και σε ποιοτικές. Τα τελευταία χρόνια, οι ποσοτικές βιβλιομετρικές μέθοδοι παρουσιάζουν μεγαλύτερη απήχηση στην επιστημονική κοινότητα για την αξιολόγηση του επιστημονικού έργου, σε σχέση με τις ποιοτικές μεθόδους, όπως είναι η «αξιολόγηση από ομότιμους» (*peer review*). Τόσο όμως η μέθοδος αξιολόγησης από ομότιμους όσο και οι ποσοτικές μέθοδοι (με

σημαντικότερη την «ανάλυση αναφορών» (*citation analysis*) παρουσιάζουν πλεονεκτήματα αλλά και μειονεκτήματα και αντίστοιχα έχουν τους υποστηρικτές τους και τους πολέμιους τους στην κοινότητα των ερευνητών. Ας εξετάσουμε πιο αναλυτικά τα υπέρ και τα κατά της κάθε μεθόδου [24].

Πλεονεκτήματα και μειονεκτήματα της μεθόδου αξιολόγησης από ομότιμους

Το βασικότερο πλεονέκτημα της μεθόδου της αξιολόγησης από ομότιμους είναι πως η αποτίμηση γίνεται από ειδικούς που έχουν γνώση και του επιστημονικού αντικειμένου αλλά και του ιδιαίτερου προτύπου των δημοσιεύσεων (*publication pattern*) σε κάθε επιστημονικό κλάδο. Ως μειονεκτήματα θεωρούνται: (α) οι χρονοβόρες και με υψηλό κόστος διαδικασίες που απαιτούνται [7], (β) η αναπόφευκτη προκατάληψη και μεροληψία των κριτών που καθορίζεται από το συγκεκριμένο κοινωνικοπολιτικό πλαίσιο που εργάζονται, αλλά και (γ) η ενδεχόμενη ατελής ενημέρωση για κάθε υπό κρίση εξειδικευμένο ερευνητικό έργο [9].

Πλεονεκτήματα και μειονεκτήματα της μεθόδου ανάλυσης αναφορών

Για τις ποσοτικές βιβλιομετρικές μεθόδους και ειδικότερα για την ανάλυση αναφορών ως πλεονεκτήματα θεωρούνται: (α) η αντικειμενικότητα των μετρήσεων, (β) η ευκολία και το χαμηλό κόστος, (γ) η δυνατότητα μετρήσεων για έναν απεριόριστο αριθμό δημοσιεύσεων και (δ) η διαβαθμισιμότητα από το μικρό-επίπεδο (άτομα) στο μακρό-επίπεδο (παγκόσμιες μετρήσεις) [20]. Ως μειονεκτήματα της μεθόδου ανάλυσης αναφορών θεωρούνται: (α) η αδυναμία ποιοτικής διαφοροποίησης των αναφορών (αρνητικές και θεωρητικές αναφορές εξομοιώνονται, συμπεριλαμβάνονται και αυτοαναφορές), (β) η έλλειψη σταθερών και τυποποιημένων δεικτών [20] και (γ) η δυσκολία σύγκρισης μεταξύ διαφορετικών επιστημονικών πεδίων επειδή το κάθε επιστημονικό πεδίο έχει ιδιαίτερη δομή δημοσιεύσεων και αναφορών (*publication and citation pattern*) [18]. Τα περισσότερα όμως και μεγαλύτερα προβλήματα της ανάλυσης αναφορών σχετίζονται με τις πηγές άντλησης δεδομένων (*ISI Web of Science, Scopus, Google Scholar*) και αφορούν περιορισμούς στην κάλυψη των βάσεων δεδομένων (ευρητηριάζεται ένας μικρός αριθμός του συνόλου των επιστημονικών περιοδικών με χρονολογικά κενά) ή στην ομοιογένεια κάλυψης (διαφορετική κάλυψη μεταξύ των επιστημονικών πεδίων, μεταξύ των γλωσσών και του τύπου των δημοσιεύσεων καθώς και των χωρών προέλευσης των συγγραφέων).

Για όλους τους παραπάνω λόγους οι περισσότεροι ερευνητές τονίζουν πως οι βιβλιομετρικοί δείκτες είναι ενδεικτικοί της απήχησης (*impact*) και της ορατότητας (*visibility*) των δημοσιεύσεων και σε καμία περίπτωση δεν μπορούν να αποτιμήσουν την ερευνητική ποιότητα [24]. Σε κάθε περίπτωση θα πρέπει να χρησιμοποιείται συνδυαστικά ένα πλήθος βιβλιομετρικών δεικτών με σκοπό τη σύγκλισή τους [9]. Πρόσθετα, συστήνεται οι δείκτες ανάλυσης αναφορών να υφίστανται κανονικοποίηση (*normalization*) ως προς τον αριθμό των συγγραφέων και ως προς τον επιστημονικό κλάδο της κάθε δημοσίευσης [20], ενώ ταυτόχρονα, ασχέτως από οποιαδήποτε βάση δεδομένων παραθέσεων ή πηγή άντλησης δεδομένων έχει χρησιμοποιηθεί, οι μετρήσεις θα πρέπει να πραγματοποιούνται από επαγγελματίες με θεωρητική κατανόηση και ολοκληρωμένη τεχνική κατάρτιση για τις βάσεις δεδομένων, τις τεχνικές ανάκτησης πληροφοριών, τις συντομογραφίες, τις έννοιες και την ορολογία των επιστημών που μελετώνται και σύμφωνα με τις αρχές “βέλτιστης πρακτικής” της βιβλιομετρικής [24].

Διεθνώς, οι πλέον καθιερωμένες βάσεις δεδομένων που περιλαμβάνουν βιβλιογραφικές εγγραφές επιστημονικών δημοσιεύσεων σε παγκόσμιο επίπεδο και στοιχεία για τις αναφορές μεταξύ τους, είναι τα συστήματα *Web of Science* (της εταιρίας Thomson Reuters), *Scopus* (της Elsevier) και *Google Scholar* (της Google).

Στο σύστημα *Web of Science (WoS)* ευρετηριάζονται πάνω από 12.000 περιοδικά τα οποία υπόκεινται σε αξιολόγηση κριτών (*peer-review*) ενώ από το 1990 προστέθηκε και η καταγραφή ορισμένων πρακτικών συνεδρίων. Δεν ευρετηριάζονται μονογραφίες και βιβλία. Είναι η παλαιότερη βάση δεδομένων επιστημονικών δημοσιεύσεων με βιβλιογραφικές εγγραφές και αναφορές που χρονολογούνται, για ορισμένους επιστημονικούς κλάδους, από το 1900. Σημαντικό πλεονέκτημα του *Web of Science* θεωρείται η αξιοπιστία του, αποτέλεσμα της αυστηρής αξιολόγησης των εκδόσεων και κυρίως των περιοδικών που εισάγονται στο σύστημα βάσει συγκεκριμένων κριτηρίων, μεταξύ των οποίων και η επιστημονική τους απήχηση. Στα μειονεκτήματα αναφέρονται κυρίως: η ανισομερής κάλυψη των επιστημονικών δημοσιεύσεων, τόσο γεωγραφικά όπου η συντριπτική πλειοψηφία του υλικού προέρχεται από αγγλόφωνες χώρες και κυρίως τις ΗΠΑ, όσο και θεματικά -υστερούν κυρίως περιοχές των επιστημονικών πεδίων “Engineering and Technology”, “Social Sciences” και “Humanities”, ενώ υπερτερούν άλλες του πεδίου “Natural Sciences [24] [36].

Στο σύστημα *Scopus* ευρετηριάζονται 18.500 τίτλοι περιοδικών, πρακτικά συνεδρίων και βιβλία. Όπως και στο σύστημα *WoS*, τα μεταδεδομένα περιέχουν αναλυτικά στοιχεία για άρθρα, συγγραφείς και οργανισμούς, ενώ η εισαγωγή των εκδόσεων πραγματοποιείται μετά από ποιοτική αξιολόγηση βάσει κριτηρίων. Το σύστημα *Scopus* παρέχει, σε σχέση με το σύστημα *WoS*, πιο ισορροπημένη γεωγραφική κάλυψη, υστερεί όμως χρονικά: οι παλαιότερες βιβλιογραφικές εγγραφές ξεκινούν το 1966, ενώ οι αναφορές σε δημοσιεύσεις καταγράφονται μετά το 1995. Επιπλέον, όπως και στο *WoS*, διαπιστώνεται ανισομερής κάλυψη ορισμένων επιστημονικών περιοχών - υστερούν περιοχές των επιστημονικών πεδίων “Social Sciences” και “Humanities”, ενώ υπερτερούν αυτές των “Medical Sciences” [37].

Το σύστημα *Google Scholar* περιλαμβάνει έναν τεράστιο αριθμό πηγών, πολλές από τις οποίες αφορούν δημοσιεύματα «γκρίζας βιβλιογραφίας» (π.χ. ακαδημαϊκές εργασίες, διατριβές, τεχνικές εκθέσεις, παραδοτέα ερευνητικών έργων). Παρά το μεγάλο εύρος σε περιεχόμενο, τα μεταδεδομένα των εγγραφών είναι περιορισμένα και απουσιάζουν κρίσιμα στοιχεία για τη γεωγραφική προέλευση ή τη διεύθυνση των οργανισμών στους οποίους ανήκουν οι συγγραφείς. Επιπλέον, τα κριτήρια εισαγωγής των πηγών του περιεχομένου είναι περιορισμένα, ενώ δεν διατίθενται ακριβή στοιχεία για τη γεωγραφική ή τη θεματική κάλυψη που παρέχει το σύστημα [38].

Με βάση τα παραπάνω χαρακτηριστικά, το σύστημα *Google Scholar*, παρά τον τεράστιο αριθμό πηγών που περιλαμβάνει, δεν είναι κατάλληλο για βιβλιομετρικές αναλύσεις που αναφέρονται σε επίπεδο χωρών ή οργανισμών, λόγω της έλλειψης μεταδεδομένων που απαιτούνται για την ταυτοποίηση των δημοσιεύσεων και της απουσίας κριτηρίων που διασφαλίζουν την ποιότητα των δημοσιεύσεων που περιλαμβάνονται στο σύστημα [10][21][29].

1.2 Βιβλιομετρικές μέθοδοι και δείκτες απήχησης

Στην παρούσα ενότητα, θα παρουσιασθούν οι κυριότερες βιβλιομετρικές μέθοδοι και οι σημαντικότεροι δείκτες απήχησης που έχουν προταθεί στη διεθνή βιβλιογραφία. Για κάθε περίπτωση, θα δοθεί αναλυτικά ο ορισμός και ο τρόπος εφαρμογής της μεθόδου, η προέλευση και λεπτομερές παράδειγμα προσδιορισμού του εκάστοτε δείκτη κάτω από συγκεκριμένα πραγματικά δεδομένα. Για αναλυτική παρουσίαση των βασικών βιβλιομετρικών τεχνικών και των δεικτών απήχησης, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο σύγγραμμα των Vinkler (2010) ή Andres (2009) [1] [2].

1.2.1 Κανονικοποιημένη βάση πηγής δεδομένων απήχηση ανά άρθρο (*Source Normalized Impact per Paper (SNIP)*)

Ο δείκτης *SNIP* δημιουργήθηκε από τον καθηγητή Henk Moed στο Κέντρο Επιστήμης & Τεχνολογίας Σπουδών (*CTWS*), Πανεπιστήμιο του Leiden. Η μετρική *SNIP* μετρά την επίπτωση συναφών παραπομπών σταθμίζοντας αναφορές με βάση το συνολικό αριθμό των παραπομπών σε ένα θεματικό πεδίο. Το αντίκτυπο μιας μεμονωμένης παραπομπής δίνει μεγαλύτερη αξία σε θεματικούς τομείς όπου οι παραπομπές είναι λιγότερο πιθανό να συμβούν, και το αντίστροφο. Όπως εξήγησε ο Moed (2010) μετρώντας την επίπτωση συναφών αναφορών σε επιστημονικά περιοδικά:

"Αναπτύσσει περαιτέρω τις έννοιες του Eugene Garfield του «ενδεχόμενου της παραπομπής (citation potential)» ενός πεδίου που ορίζεται ως το μέσο μήκος της λίστας αναφορών σε ένα πεδίο και προσδιορίζει την πιθανότητα να αναφερθούν, καθώς και την ανάγκη για δίκαιη αξιολόγηση της απόδοσης για τη διόρθωση των διαφορών μεταξύ των γνωστικών αντικειμένων."

Ο δείκτης *SNIP* ορίζεται ως ο λόγος του αριθμού των αναφορών ενός περιοδικού ανά επιστημονικό άρθρο και του ενδεχόμενου παραπομπής στο θεματικό πεδίο που ανήκει. Έχει ως στόχο να επιτρέψει την άμεση σύγκριση πηγών μεταξύ διαφορετικών θεματικών πεδίων. Το ενδεχόμενο παραπομπής φαίνεται να ποικίλλει όχι μόνο μεταξύ θεματικών κατηγοριών ενός περιοδικού αλλά και μεταξύ των περιοδικών μέσα στην ίδια θεματική κατηγορία [11]. Για παράδειγμα, θεωρητικά περιοδικά τείνουν να εμφανίζουν υψηλότερο ενδεχόμενο παραπομπής από ότι εφαρμοσμένα ή κλινικά περιοδικά, ενώ ταυτόχρονα περιοδικά που καλύπτουν επείγοντα θέματα υψηλότερα από ότι περιοδικά κλασικών θεμάτων ή πιο γενικά περιοδικά. Η μετρική *SNIP* διορθώνει τέτοιες διαφορές.

Ο δείκτης *SNIP* ορίζεται ως το πηλίκο δύο άλλων δεικτών. Πιο συγκεκριμένα, στον αριθμητή του δείκτη, τοποθετείται η ακατέργαστη επίδραση ανά δημοσίευση που δημοσιεύεται στο περιοδικό (*Raw Impact per Publication (RIP)*), ενώ στον αντίστοιχο παρονομαστή η σχετική βάση δεδομένων του ενδεχομένου αναφοράς (*Relative Database Citation Potential (RDCP)*) στο υποπεδίο του περιοδικού. Αξίζει να σημειωθεί ότι τόσο ο αριθμητής όσο και ο παρονομαστής του δείκτη *SNIP* είναι πηλικά [33]. Πιο συγκεκριμένα, ο τύπος υπολογισμού του δείκτη *SNIP* δίνεται ως εξής

$$SNIP = \frac{RIP}{RDCP}.$$

Όλα τα εμπειρικά αποτελέσματα προέρχονται από τη βιβλιογραφική βάση δεδομένων *Scopus*. Οι τιμές του δείκτη *SNIP* ενημερώνονται μία φορά το έτος, παρέχοντας μια ανανεωμένη από πλευράς ημερομηνίας άποψη του ερευνητικού τοπίου.

Η μετρική *SNIP* παρέχει εναλλακτικές τιμές που στα πλαίσια μιας βιβλιομετρικής ανάλυσης είναι χρήσιμες για να δημιουργήσουν μια πιο αμερόληπτη και αντικειμενική ανάλυση. Βοηθά τους συντάκτες να αξιολογήσουν το περιοδικό τους και να κατανοήσουν τον τρόπο που αποδίδει στον ανταγωνισμό. Η μετρική *SNIP* παρέχει περισσότερες συναφείς πληροφορίες και μπορεί να δώσει μια καλύτερη εικόνα για συγκεκριμένους τομείς, όπως η Μηχανική, η Επιστήμη των Υπολογιστών και οι Κοινωνικές Επιστήμες. Μπορεί επίσης να βοηθήσει όλους τους ακαδημαϊκούς να προσδιορίσουν ποια περιοδικά έχουν την καλύτερη απόδοση στο θεματικό τους πεδίο, ώστε να προσδιορίζουν κατάλληλα τους ερευνητικούς τους στόχους.

Παράδειγμα 1. Το επιστημονικό περιοδικό *Applied Bioinformatics*, το οποίο εκδίδεται από τον Οίκο *Springer* και ανήκει στη θεματική κατηγορία *Computer Science*, παρουσίασε κατά το έτος 2009 δείκτη *SNIP* ίσο με 0,382. Αυτό πρακτικά σημαίνει ότι η απήχηση (μετρημένη σε αριθμό αναφορών που λαμβάνει) που παρουσιάζει ένα άρθρο που δημοσιεύεται στο περιοδικό *Applied Bioinformatics* είναι 0,382 φορές μικρότερη (περίπου υποτετραπλάσια) από την αντίστοιχη απήχηση που αναμένεται να έχει ένα άρθρο που δημοσιεύεται σε τυχαία επιλεγόμενο περιοδικό της ίδιας θεματικής κατηγορίας. Αντίθετα, για το επιστημονικό περιοδικό *Applied Intelligence*, το οποίο εκδίδεται από τον Οίκο *Kluwer Academic Publishers* και ανήκει στη θεματική κατηγορία *Computer Science*, παρουσίασε κατά το έτος 2009 δείκτη *SNIP* ίσο με 1,282. Αυτό πρακτικά σημαίνει ότι η απήχηση (μετρημένη σε αριθμό αναφορών που λαμβάνει) που παρουσιάζει ένα άρθρο που δημοσιεύεται στο περιοδικό *Applied Intelligence* είναι 1,282 φορές μεγαλύτερη από την αντίστοιχη απήχηση που αναμένεται να έχει ένα άρθρο που δημοσιεύεται σε τυχαία επιλεγόμενο περιοδικό της ίδιας θεματικής κατηγορίας.

1.2.2 Απήχηση ανά δημοσίευση (*Impact Per Publication (IPP)*)

Η απήχηση ανά δημοσίευση (*Impact Per Publication (IPP)*) για ένα έτος (Y) είναι ο αριθμός των αναφορών (A) που έλαβαν κατά τον χρόνο αυτό ακαδημαϊκά άρθρα που δημοσιεύτηκαν σε ένα συγκεκριμένο περιοδικό κατά τα τρία προηγούμενα έτη (Y_1, Y_2, Y_3), διαιρούμενος με τον αριθμό των ακαδημαϊκών άρθρων (B) που δημοσιεύτηκαν σε αυτό το περιοδικό σε αυτά τα τρία έτη (Y_1, Y_2, Y_3). Ο τύπος υπολογισμού του δείκτη είναι *IPP* είναι

$$IPP = \frac{\text{πλήθος αναφορών } A}{\text{πλήθος άρθρων } B} .$$

Για παράδειγμα αν θέλουμε να υπολογίσουμε τον δείκτη *IPP* για ένα συγκεκριμένο περιοδικό για το έτος 2011, θα πρέπει να απαριθμήσουμε τόσο το πλήθος των άρθρων

που δημοσιεύτηκαν στο συγκεκριμένο περιοδικό κατά τα έτη 2010, 2009, 2008, όσο και το πλήθος αναφορών που τα συγκεκριμένα άρθρα έλαβαν κατά το έτος 2011.

Η μετρική *IPP* χρησιμοποιεί μια χρονική περίοδο διάρκειας τριών ετών για τις παραπομπές, η οποία θεωρείται ότι είναι η βέλτιστη χρονική περίοδος για την ακριβή μέτρηση αναφορών στα περισσότερα θεματικά πεδία. Ο δείκτης δεν είναι κανονικοποιημένος για το θεματικό πεδίο και, ως εκ τούτου δίνει μια ακατέργαστη ένδειξη του μέσου αριθμού των παραπομπών που είναι πιθανό να λάβει μια δημοσίευση που δημοσιεύθηκε σε ένα περιοδικό. Μετά από κανονικοποίηση στο πεδίο του θέματος, η ακατέργαστη απήχηση ανά δημοσίευση μετατρέπεται σε *SNIP* [32]. Σημειώνεται ότι στο πλαίσιο του υπολογισμού των *SNIP*, η ακατέργαστη απήχηση ανά δημοσίευση συνήθως αναφέρεται ως *RIP*. Όπως και το *SNIP*, έτσι και το *RIP* αναπτύχθηκε επίσης από το Κέντρο Επιστήμης & Τεχνολογίας Σπουδών (*CWTS*) του Πανεπιστημίου του Leiden. Ο δείκτης απήχησης ανά δημοσίευση υπολογίζεται από τη βάση δεδομένων *Scopus*.

Στη συνέχεια, παρατίθεται σχετικό παράδειγμα υπολογισμού του δείκτη *IPP*.

Παράδειγμα 2. Το επιστημονικό περιοδικό *Applied Mathematics & Information Sciences*, το οποίο εκδίδεται από τον Οίκο *Natural Sciences Publishing Corporation* και ανήκει στη θεματική κατηγορία *Computer Science*, παρουσίασε κατά το έτος 2011 δείκτη *IPP* ίσο με 0,619. Αυτό πρακτικά σημαίνει ότι το πλήθος των άρθρων που δημοσιεύτηκαν στο περιοδικό *Applied Mathematics & Information Sciences* κατά τα έτη 2008, 2009 και 2010 είναι 1,6 φορές μεγαλύτερο από το πλήθος αναφορών που έλαβαν τα συγκεκριμένα άρθρα κατά το έτος 2011. Ισοδύναμα, μπορούμε να διατυπώσουμε το συμπέρασμα ότι το πλήθος αναφορών που έλαβαν κατά το έτος 2011, τα άρθρα που δημοσιεύτηκαν στο περιοδικό *Applied Mathematics & Information Sciences* κατά τα έτη 2008, 2009 και 2010, είναι 0,619 φορές μικρότερο από το πλήθος των άρθρων αυτών.

1.2.3 Βαθμίδα περιοδικού *SCImago* (*SCImago Journal Rank* (*SJR*))

Αναπτύχθηκε από τους Félix de Moya, καθηγητή στο Ανώτατο Συμβούλιο Επιστημονικής Έρευνας και Vicente Guerrero Bote καθηγητή στο Πανεπιστήμιο Extremadura. Ο δείκτης *SCImago Journal Rank* (*SJR*) είναι μια μετρική κύρους η οποία βασίζεται στην ιδέα ότι «όλες οι αναφορές δεν είναι ίσες» [4]. Με την τάξη περιοδικού *SCImago*, το θεματικό πεδίο, η ποιότητα και η φήμη του περιοδικού έχουν άμεση επίδραση στην αξία που έχει μια παραπομπή. Η τάξη περιοδικού *SCImago* υπολογίζεται όπως και οι δείκτες *SNIP* και *IPP* από την ευρετηριακή βάση δεδομένων *Scopus* [34].

Ο δείκτης *SJR* είναι μέτρο της επιστημονικής επιρροής των ακαδημαϊκών περιοδικών και αντιπροσωπεύει τόσο τον αριθμό των αναφορών που λαμβάνονται από ένα περιοδικό όσο και τη σημασία ή το κύρος των περιοδικών απ' όπου τέτοιες αναφορές προέρχονται. Πρόκειται για μια παραλλαγή του μέτρου της κεντρικότητας του ιδιοδιανύσματος που χρησιμοποιείται στη Θεωρία των Δικτύων [34]. Τέτοια μέτρα αποδεικνύουν τη σημασία ενός κόμβου σε ένα δίκτυο που βασίζεται στην αρχή ότι οι συνδέσεις σε κόμβους υψηλής βαθμολογίας συμβάλλουν περισσότερο στη βαθμολογία του κόμβου. Είναι ένας ανεξάρτητου μεγέθους δείκτης που κατατάσσει

τα περιοδικά με βάση το κατά μέσο όρο κύρος κατ' άρθρο και μπορεί να χρησιμοποιηθεί για συγκρίσεις περιοδικών σε διαδικασίες επιστημονικής αξιολόγησης [19].

Ο υπολογισμός του δείκτη SJR δεν είναι τόσο απλός και εύκολος όσο στους προηγούμενους δείκτες που αναφέρθηκαν. Στην πραγματικότητα δεν υπάρχει απλός μαθηματικός τύπος υπολογισμού αλλά πραγματοποιείται χρησιμοποιώντας έναν επαναληπτικό αλγόριθμο που διανέμει τιμές κύρους μεταξύ των περιοδικών, έως ότου επιτευχθεί σταθερή λύση. Ο αλγόριθμος SJR αρχίζει θέτοντας τη ίδια τιμή κύρους σε κάθε περιοδικό, στη συνέχεια, χρησιμοποιώντας μια επαναληπτική διαδικασία, το κύρος αναδιανέμεται, δηλαδή τα περιοδικά μεταφέρουν το κύρος τους το ένα στο άλλο μέσα από αναφορές. Η διαδικασία ολοκληρώνεται όταν η διαφορά μεταξύ των τιμών κύρους του περιοδικού μετά από διαδοχικές επαναλήψεις δεν υπερβαίνει μια ελάχιστη τιμή. Ο υπολογισμός του τελικού κύρος ενός περιοδικού είναι μια επαναληπτική διαδικασία, κατά την οποία το κύρος στο στάδιο i ενός περιοδικού εξαρτάται από το κύρος του συνόλου των περιοδικών στο στάδιο $i-1$ [17]. Ο υπολογισμός του δείκτη αυτού δίνεται από την παρακάτω σχέση:

$$PSJR_i = \frac{(1-d-e)}{N} + e \cdot \frac{Art_i}{\sum_{j=1}^N Art_j} + d \cdot \left[\sum_{j=1}^N C_{ji} \cdot \frac{PSJR_j}{C_j} \cdot CF + \frac{Art_i}{\sum_{j=1}^N Art_j} \cdot \sum_{k \in DN} PSJR_k \right]$$

$PSJR_i$: τάξη μεγέθους $SCImago$ του περιοδικού i

C_{ji} : αναφορές από το περιοδικό j στο περιοδικό i

C_j : πλήθος αναφορών του περιοδικού j

d : σταθερά ($d = 0,9$)

e : σταθερά ($e = 0,0999$)

N : πλήθος περιοδικών που περιλαμβάνονται στη βάση δεδομένων

Art_j : πλήθος πρωτευόντων αντικειμένων (άρθρων, ανασκοπήσεων, εργασιών Συνεδρίων) του περιοδικού j .

Η διαδικασία αναπτύσσεται σε δύο φάσεις: (α) ο υπολογισμός του κύρους SJR ($PSJR$) για κάθε περιοδικό. Πρόκειται για ένα μέτρο που εξαρτάται από το μέγεθος και αντικατοπτρίζει το συνολικό κύρος του περιοδικού, και (β) η κανονικοποίηση του μέτρου αυτού για να επιτευχθεί ένα μέτρο κύρους ανεξάρτητου μεγέθους, ο δείκτης SJR [17]. Δεν είναι εύκολο να δοθεί αριθμητικό παράδειγμα καθώς όπως αναφέρθηκε παραπάνω απαιτείται αλγόριθμος για τον υπολογισμό της τάξης περιοδικού.

Επιπλέον αξίζει να αναφερθεί ότι ο δείκτης είναι εμπνευσμένος από τον αλγόριθμο PageRank που χρησιμοποιείται από τη Google. Το PageRank είναι μία αριθμητική τιμή, η οποία αντιπροσωπεύει το πόσο σημαντική και σπουδαία είναι μία ιστοσελίδα στο διαδίκτυο. Το PageRank επηρεάζει σημαντικά, μαζί με κάποιους επιπλέον παράγοντες, την κατάταξη των αποτελεσμάτων αναζήτησης. Για τον υπολογισμό και προσδιορισμό του PageRank, η Google χρησιμοποιεί έναν πολύπλοκο αλγόριθμο με εκατομμύρια μεταβλητές και όρους και ο υπολογισμός του γίνεται με αδιάβλητο αυτοματοποιημένο τρόπο. Όσο πιο πολλούς συνδέσμους έχει μία ιστοσελίδα από άλλες ιστοσελίδες, τόσο πιο σημαντική και σπουδαία

χαρακτηρίζεται. Εάν δηλαδή οι σύνδεσμοι προς την ιστοσελίδα προέρχονται από σημαντικές ιστοσελίδες μεγάλης σπουδαιότητας, η αξία η οποία προσδίδεται είναι ακόμα μεγαλύτερη [40].

1.2.4 Παράγοντας απήχησης (*Impact Factor (IF)*)

Ο παράγοντας απήχησης επινοήθηκε από τον Eugene Garfield, ιδρυτή του επιστημονικού Οργανισμού *Institute for Scientific Information*. Ο παράγοντας απήχησης (*IF*) ενός ακαδημαϊκού περιοδικού είναι ένα μέτρο που αντικατοπτρίζει το μέσο αριθμό των παραπομπών σε πρόσφατα άρθρα που δημοσιεύτηκαν στο εν λόγω περιοδικό. Συχνά χρησιμοποιείται ως υποκατάστατο για τη σχετική σημασία ενός περιοδικού στο πεδίο του, με περιοδικά με υψηλότερους συντελεστές απήχησης να θεωρούνται πιο σημαντικά από αυτά με χαμηλότερους. Οι παράγοντες απήχησης υπολογίζονται ετησίως, αρχής γενομένης από το 1975 για τα εν λόγω περιοδικά που είναι καταχωρημένα στο *Journal Citation Reports* [37].

Σε οποιοδήποτε δεδομένο έτος, ο παράγοντας απήχησης ενός περιοδικού είναι ο μέσος αριθμός των αναφορών που λαμβάνει ένα επιστημονικό άρθρο που δημοσιεύτηκε στο περιοδικό κατά τη διάρκεια των δύο προηγούμενων ετών [3]. Για παράδειγμα, αν θέλουμε να υπολογίσουμε τον παράγοντα απήχησης ενός περιοδικού για το έτος i , θα πρέπει να απαριθμήσουμε τόσο τα άρθρα που δημοσιεύτηκαν στο συγκεκριμένο περιοδικό κατά τα έτη $i-1$ και $i-2$ (πλήθος άρθρων A), όσο και τις αναφορές που έλαβαν τα άρθρα αυτά κατά το έτος i (πλήθος αναφορών I).

$$IF = \frac{\text{πλήθος αναφορών } I}{\text{πλήθος άρθρων } A}.$$

Ο παράγοντας απήχησης χρησιμοποιείται κυρίως για τη σύγκριση διαφορετικών περιοδικών μέσα σε ένα συγκεκριμένο πεδίο. Η βάση δεδομένων του *ISI Web of Knowledge* περιέχει περισσότερα από 11.000 επιστημονικά περιοδικά.

Είναι δυνατόν να εξεταστεί ο παράγοντας απήχησης περιοδικών στα οποία ένα συγκεκριμένο πρόσωπο έχει δημοσιεύσει άρθρα. Αυτή η χρήση είναι διαδεδομένη, αλλά αμφιλεγόμενη. Ο Garfield προειδοποιεί για την "κατάχρηση κατά την αξιολόγηση ατόμων" επειδή υπάρχει μεγάλη διαφοροποίηση μεταξύ άρθρων εντός ενός περιοδικού [5]. Οι παράγοντες απήχησης έχουν μεγάλη, αλλά αμφιλεγόμενη, επιρροή στον τρόπο που η δημοσιευμένη επιστημονική έρευνα γίνεται αντιληπτή και αξιολογείται [41]. Δυστυχώς, ορισμένες εταιρίες παράγουν ψευδείς παράγοντες απήχησης. Είναι σημαντικό να σημειωθεί ότι ο παράγοντας απήχησης είναι ένας δείκτης για τα περιοδικά και δεν πρέπει να χρησιμοποιείται για την αξιολόγηση μεμονωμένων ερευνητών ή φορέων.

Μια αξιοσημείωτη παραλλαγή του παράγοντα απήχησης (*Impact Factor*) είναι ο πενταετής παράγοντας απήχησης (*5-Year Impact Factor*) με τη μόνη διαφορά ότι ο δεύτερος έχει χρονική περίοδο υπολογισμού πέντε χρόνια αντί για δύο. Πιο συγκεκριμένα ο πενταετής παράγοντας απήχησης ορίζεται ως ο αριθμός των παραπομπών που έγιναν σε άρθρα που δημοσιεύθηκαν τα πέντε προηγούμενα χρόνια, διαιρούμενος με το συνολικό αριθμό των άρθρων που δημοσιεύτηκαν αυτά τα πέντε χρόνια [3]. Και οι δύο δείκτες υπολογίζονται από τη βάση δεδομένων του *ISI Web of Knowledge*. Στη συνέχεια, παρατίθεται σχετικό παράδειγμα υπολογισμού του *IF*.

Παράδειγμα 3. Το επιστημονικό περιοδικό *Computers & Security*, το οποίο εκδίδεται από τον Οίκο *Elsevier* και ανήκει στη θεματική κατηγορία *Computer Science*, παρουσίασε κατά το έτος 2014 δείκτη *IF* ίσο με 1,031. Αυτό πρακτικά σημαίνει ότι τα άρθρα που δημοσιεύτηκαν στο περιοδικό *Computers & Security* κατά τα έτη 2012 και 2013 έλαβαν κατά το έτος 2014 κατά μέσο όρο 1,031 αναφορές. Πρόσθετα, για το ίδιο περιοδικό ο δείκτης *5-Year Impact Factor* υπολογίστηκε για το έτος 2014 ίσος με 1,386. Αυτό πρακτικά σημαίνει ότι τα άρθρα που δημοσιεύτηκαν στο περιοδικό *Computers & Security* κατά τα έτη 2009, 2010, 2011, 2012 και 2013 έλαβαν κατά το έτος 2014 κατά μέσο όρο 1,386 αναφορές.

1.2.5 Δείκτης αμεσότητας περιοδικού (*Journal Immediacy Index*)

Ο δείκτης αμεσότητας ενός επιστημονικού περιοδικού (*Journal Immediacy Index*) είναι ο αριθμός των παραπομπών που έγιναν το τρέχον έτος σε άρθρα που έχουν δημοσιευθεί στο περιοδικό το τρέχον έτος, διαιρούμενος με το συνολικό αριθμό των άρθρων που δημοσιεύτηκαν το έτος αυτό. Είναι ένα μέτρο που δείχνει πόσο επίκαιρο είναι το έργο που δημοσιεύτηκε σε ένα επιστημονικό περιοδικό [37].

Επειδή αντιπροσωπεύει το μέσο πλήθος αναφορών ανά άρθρο, ο δείκτης αμεσότητας τείνει να μειώσει το πλεονέκτημα που έχουν τα μεγάλα περιοδικά σε σχέση με τα μικρά. Ωστόσο, συχνά εκδίδονται περιοδικά που μπορεί να έχουν πλεονέκτημα, διότι ένα άρθρο τους που δημοσιεύθηκε στις αρχές του έτους έχει καλύτερη ευκαιρία να παρατεθεί από εκείνο το οποίο δημοσιεύτηκε αργότερα μέσα στο έτος. Πολλά περιοδικά που δημοσιεύουν λίγα τεύχη ή στο τέλος του έτους τείνουν να έχουν χαμηλούς δείκτες αμεσότητας [37]. Για τη σύγκριση των περιοδικών που ειδικεύονται στην έρευνα αιχμής, ο δείκτης αμεσότητας μπορεί να προσφέρει μια χρήσιμη προοπτική. Μαζί με τον παράγοντα απήχησης (*IF*) υπολογίζονται κάθε έτος από το *Institute for Scientific Information (ISI)*. Οι δύο αυτοί παράγοντες δημοσιεύονται ετησίως στο *Journal Citation Reports*.

Ο τύπος υπολογισμού του δείκτη αμεσότητας είναι σχεδόν ίδιος με τον τύπο του παράγοντα απήχησης με τη μόνη διαφορά ότι ο πρώτος υπολογίζεται για διάστημα ενός έτους ενώ ο δεύτερος για διάστημα δύο ή πέντε ετών. Για παράδειγμα, αν θέλουμε να βρούμε τον δείκτη αμεσότητας ενός επιστημονικού περιοδικού το έτος 2015, θα πρέπει να απαριθμήσουμε τόσο τα άρθρα που δημοσιεύθηκαν κατά το έτος 2015 στο συγκεκριμένο περιοδικό, όσο και τις αναφορές που τα άρθρα αυτά έλαβαν κατά το ίδιο έτος. Στη συνέχεια, παρατίθεται σχετικό παράδειγμα υπολογισμού του παράγοντα αμεσότητας.

Παράδειγμα 4. Έστω ότι ένα επιστημονικό περιοδικό δημοσιεύει κατά το έτος 2014 50 άρθρα, τα οποία λαμβάνουν κατά το ίδιο έτος 20 αναφορές. Τότε, ο δείκτης αμεσότητας (*JII*) του συγκεκριμένου περιοδικού για το έτος 2014 είναι ίσος με

$$JII = \frac{20}{50} = 0,4 .$$

Όπως και στον υπολογισμό του παράγοντα απήχησης έτσι και εδώ στον παρονομαστή του κλάσματος λαμβάνονται υπόψη μόνο άρθρα και κριτικές άρθρων. Είναι άμεσο ότι

σε περίπτωση που το άρθρο ενός περιοδικού δεν έχει αναφερθεί καμία φορά μέσα στον χρόνο που δημοσιεύτηκε από κανένα άλλο άρθρο τότε ο δείκτης αμεσότητας είναι 0.

1.2.6 Δείκτης *Journal Cited Half-life*

Ένα άρθρο μπορεί να αναφερθεί εκατοντάδες ή χιλιάδες φορές ανά έτος, αλλά γενικότερα τα άρθρα που λαμβάνουν αναφορές ενδέχεται να διαφέρουν σημαντικά ως προς την ηλικία (χρόνο δημοσίευσης τους). Μερικά θα είναι μόνο λίγων μηνών, δημοσιευμένα μέσα στο ίδιο ημερολογιακό έτος, όπως και η αναφορά ενώ άλλα άρθρα μπορεί να έχουν δημοσιευθεί πριν από δεκαετίες. Ο δείκτης *Journal cited half-life* μετρά όλες τις αναφορές ενός περιοδικού κατά τη διάρκεια ενός ημερολογιακού έτους i , προσδιορίζει το πλήθος των άρθρων που έχουν δημοσιευθεί στο συγκεκριμένο περιοδικό και έχουν λάβει αναφορά κατά το έτος i και υπολογίζει την ενδιάμεση ημερομηνία δημοσίευσης των άρθρων του περιοδικού που έχουν λάβει τουλάχιστον μία αναφορά κατά το έτος i . Τα μισά από τα αναφερόμενα άρθρα δημοσιεύθηκαν πριν από αυτή την ημερομηνία και τα άλλα μισά έχουν δημοσιευθεί αργότερα από αυτήν.

Ο δείκτης *Journal cited half-life* υπολογίζει τον αριθμό των συνολικών αναφορών που έλαβε το περιοδικό το έτος έκδοσης ανεξάρτητα από το πότε δημοσιεύθηκαν τα αναφερόμενα άρθρα. Είναι διαθέσιμο μόνο αν το περιοδικό έχει αναφερθεί περισσότερες από 100 φορές κατά το έτος έκδοσης [37].

Παράδειγμα 5. Έστω ότι ανάμεσα στα άρθρα που έχουν δημοσιευθεί σε ένα επιστημονικό περιοδικό, υπάρχουν 120 άρθρα που λαμβάνουν τουλάχιστον μια αναφορά κατά το έτος 2014. Προκειμένου να υπολογισθεί ο δείκτης *Journal cited half-life*, για το συγκεκριμένο περιοδικό για το έτος 2014, θα πρέπει να καταγραφεί ο χρόνος δημοσίευσης των 120 άρθρων που έχουν δημοσιευθεί στο συγκεκριμένο περιοδικό και τα οποία έλαβαν κατά το έτος 2014 τουλάχιστον μία αναφορά και να υπολογισθεί η διάμεσος των χρόνων δημοσίευσης τους. Αν η διάμεσος των χρόνων δημοσίευσης των 120 άρθρων, είναι ίση με 4,5 έτη, αυτό πρακτικά σημαίνει ότι οι μισές αναφορές που γίνονται (κατά το έτος 2014) σε άρθρα του συγκεκριμένου περιοδικού, αναφέρονται σε άρθρα του περιοδικού που έχουν δημοσιευθεί τα τελευταία 4,5 έτη.

1.2.7 Δείκτης επιστημονικής ποιότητας h (*h-index*)

Ο δείκτης επιστημονικής ποιότητας h , ή απλά δείκτης h (*h-index*) είναι ένας αριθμητικός δείκτης που στοχεύει στη μέτρηση της ποιότητας των ακαδημαϊκών δημοσιεύσεων ενός επιστήμονα. Ονομάζεται επίσης δείκτης Hirsch ή αριθμός Hirsch από τον επιστήμονα που αρχικά πρότεινε τη χρήση του. Ο δείκτης h αναπτύχθηκε από τον καθηγητή Φυσικής Jorge Hirsch από το Πανεπιστήμιο της Καλιφόρνια του Σαν Ντιέγκο και δημοσιεύθηκε στα Πρακτικά της Εθνικής Ακαδημίας Επιστημών των Ηνωμένων Πολιτειών της Αμερικής το Νοέμβριο του 2005. Στο εν λόγω άρθρο του, ο Hirsch αναφέρει τα εξής: «Προτείνω το δείκτη h , ο οποίος ορίζεται ως ο αριθμός των επιστημονικών άρθρων με αριθμό αναφορών $\geq h$, ως ένα χρήσιμο εργαλείο για τον προσδιορισμό της επιστημονικής ποιότητας και απήχησης ενός ερευνητή» [6].

Ο δείκτης h είναι ίσος με τον αριθμό των επιστημονικών άρθρων που έχει δημοσιεύσει ένας ερευνητής, τα οποία έχουν τουλάχιστον h αναφορές από άλλους επιστήμονες. Για παράδειγμα, αν γνωρίζουμε ότι για έναν επιστήμονα ισχύει ότι ο δείκτης $h=20$, τότε αυτό πρακτικά σημαίνει ότι έχει δημοσιεύσει μεταξύ άλλων, 20 επιστημονικά άρθρα, τα οποία έχουν 20 ή περισσότερες αναφορές από άλλους επιστήμονες. Αν έχει $h=100$, αυτό σημαίνει ότι έχει δημοσιεύσει μεταξύ άλλων, 100 επιστημονικά άρθρα, το καθένα από τα οποία έχει τουλάχιστον 100 αναφορές από άλλους επιστήμονες. Από την άλλη, αν κάποιος έχει δημοσιεύσει 100 άρθρα αλλά μόνο ένα εκ των άρθρων του, έχει τουλάχιστον 100 αναφορές από άλλους επιστήμονες τότε $h=1$. Πρόσθετα, αν κάποιος ερευνητής έχει δημοσιεύσει ένα άρθρο με 100 αναφορές από άλλους επιστήμονες, τότε και πάλι $h=1$ [42]. Γίνεται σαφές λοιπόν, ότι μέσω του δείκτη h δεν έχει σημασία τόσο η ποσότητα των άρθρων αλλά η ποιότητα και απήχησή τους στην επιστημονική κοινότητα.

Αυτός ο τρόπος μέτρησης της ποιότητας και απήχησης ενός επιστήμονα είναι χρήσιμος επειδή δεν εμπεριέχει το δυσανάλογο ειδικό βάρος (α) άρθρων με εξαιρετικά μεγάλο αριθμό αναφορών, και (β) άρθρων τα οποία δεν έχουν ακόμη αναφερθεί στη βιβλιογραφία από άλλους επιστήμονες.

Ο δείκτης μπορεί επίσης να χρησιμοποιηθεί για να μετρήσει την ποιότητα ενός επιστημονικού περιοδικού καθώς και μια ομάδα επιστημόνων, ενός Πανεπιστημιακού Τμήματος, ενός Πανεπιστημίου ή ακόμα και μιας ολόκληρης χώρας ή ηπείρου, όσον αφορά την παραγωγικότητά της στην υψηλής ποιότητας έρευνα [24].

Σε μη ατομικό επίπεδο λοιπόν, (π.χ. σε επίπεδο Τμήματος πανεπιστημίου) ο h -index μπορεί να υπολογιστεί είτε σφαιρικά (δηλ. το Τμήμα να θεωρηθεί ως ένας συγγραφέας και ο hg -index να υπολογιστεί στο σύνολο των δημοσιεύσεων του Τμήματος) είτε με διαδοχικό (*successive*) τρόπο [8][13], όπως αρχικά πρότειναν οι Prathar (2006) και Schumbert (2007) και σήμερα έχει γίνει ευρέως αποδεκτό.

Σύμφωνα με αυτή τη μεθοδολογία ένα πανεπιστημιακό Τμήμα έχει h_1 -index εάν στο σύνολο των μελών του (N_s) έχει h_1 μέλη με h -index τουλάχιστον h_1 το καθένα και τα υπόλοιπα μέλη του Τμήματος (N_s-h_1) έχουν το μέγιστο h -index ίσο με h_1 το καθένα».

Μειονεκτήματα του h -index είναι ότι δεν λαμβάνει υπόψη του: την ηλικία και τον τύπο των δημοσιεύσεων, το μέγεθος της ακαδημαϊκής μονάδας και τη θεματική περιοχή. Ένα επιπλέον μειονέκτημα είναι ότι ως φυσικός αριθμός δεν έχει μεγάλη διαβαθμισιμότητα.

Για να αντιπαρέλθουν το πρόβλημα αυτό οι Ruane & Tol (2008) πρότειναν μια σειρά από ρητούς (*rational*) διαδοχικούς (*successive*) h -indices (h_{1+} , $h_{1Δ}$, h_{1*}) [16]. Ο h_{1+} index λαμβάνει υπόψη το συνολικό αριθμό μελών Δ.Ε.Π που έχουν h -index ίσο με h_1 σύμφωνα με τον τύπο:

$$h_{1+} = h_1 + \frac{(n - h_1)}{(s - h_1)}$$

όπου n είναι ο συνολικός αριθμός των μελών Δ.Ε.Π με h -index ίσο με h_1 και s είναι ο συνολικός αριθμός μελών Δ.Ε.Π. Ο $h_{1Δ}$ index λαμβάνει υπόψη τις μονάδες (m) h -index των μελών Δ.Ε.Π που υπολείπονται για να γίνει ο h_1 -index h_1+1 , σύμφωνα με τον τύπο:

$$h_{1\Delta} = h_1 + 1 - \frac{m}{(2h+1)}$$

Ο h_1^* index είναι παρόμοιος με τον $h_{1\Delta}$ index με τη μόνη διαφορά πως δεκαδικοποιείται πρώτα ο h -index του κάθε μέλους Δ.Ε.Π (με βάση τις μονάδες που υπολείπονται για να γίνει $h+1$) και υπολογίζεται μετά με διαδοχικό τρόπο.

Τέλος, αξίζει να αναφερθεί ότι ο δείκτης επιστημονικής ποιότητας h υπολογίζεται από τις βάσεις δεδομένων *Scopus*, *ISI Web Of Knowledge* και *Google Scholar* αλλά με διαφορετικό τρόπο για το ίδιο άτομο, καθώς το περιεχόμενο σε κάθε βάση δεδομένων είναι μοναδικό και διαφορετικό. Στη συνέχεια, παρατίθεται σχετικό παράδειγμα υπολογισμού του δείκτη h .

Παράδειγμα 6. Ας υποθέσουμε ότι ένας συγγραφέας έχει δημοσιεύσει 8 άρθρα με 33, 30, 20, 15, 7, 6, 5 και 4 αναφορές το καθένα αντίστοιχα. Για τον υπολογισμό του δείκτη h πρέπει να τοποθετηθούν τα άρθρα του συγγραφέα (στήλη *Articles*) σε φθίνουσα διάταξη σύμφωνα με τον αριθμό των φορών που έχουν αναφερθεί (στήλη *Citation numbers*). Τότε, όπως φαίνεται από τον παρακάτω πίνακα, ο δείκτης h είναι ίσος με 6 γιατί ο συγγραφέας έχει δημοσιεύσει 6 άρθρα που έχουν λάβει το λιγότερο 6 αναφορές το καθένα απ' αυτά.

<u>Articles</u>	<u>Citation numbers</u>
1	33
2	30
3	20
4	15
5	7
<u>6</u>	<u>6</u> = h-index
7	5
8	4

Εικόνα 1 : Παράδειγμα υπολογισμού δείκτη h -index

(<http://subjectguides.uwaterloo.ca/content.php?pid=84805&sid=1885850>)

1.2.8 Δείκτης επιστημονικής ποιότητας p (p -index)

Πριν από λίγα χρόνια ο Prathap (2009a,b) σε μια προσπάθεια του να επεκτείνει την αξιολόγηση του ερευνητικού έργου πρότεινε ένα νέο δείκτη ο οποίος ορίζεται ως εξής:

$$p\text{-index} = \left(\frac{c^2}{p} \right)^{1/3}$$

Το πλεονέκτημα αυτού του δείκτη σε σχέση με τον αριθμό αναφορών ανά δημοσίευση είναι ότι δίνοντας μεγαλύτερη βαρύτητα στον αριθμό των αναφορών σε σχέση με τον αριθμό των δημοσιεύσεων, δεν επηρεάζεται τόσο από έναν μικρό αριθμό δημοσιεύσεων [14][15]. Έτσι, π.χ. για ένα σύνολο 50 δημοσιεύσεων με 200

αναφορές, οι «αναφορές ανά δημοσίευση» είναι 4, όπως ακριβώς και για ένα σύνολο 5 δημοσιεύσεων με 20 αναφορές. Αντίθετα, ο *p-index* ενώ στην πρώτη περίπτωση είναι ίσος με 9,28 στη δεύτερη περίπτωση είναι μόλις 4,31. Μειονέκτημα του δείκτη αυτού σε σχέση με τον προηγούμενο είναι πως δεν είναι αμερόληπτος ως προς το μέγεθος του εκάστοτε συνόλου που εξετάζουμε. Για παράδειγμα, αν εξετάζουμε Πανεπιστημιακά Τμήματα μεταξύ τους, τότε θεωρητικά για ισοδύναμα σε παραγωγικότητα και απήχηση Τμήματα ο δείκτης ευνοεί Τμήματα με μεγαλύτερο αριθμό μελών Δ.Ε.Π τα οποία λογικά θα έχουν και μεγαλύτερο απόλυτο αριθμό αναφορών [24].

1.2.9 Μετρικές του Μελετητή της Google (*Google Scholar Metrics*)

Οι μετρικές του Μελετητή της Google παρέχουν έναν εύκολο τρόπο για τους συγγραφείς έτσι ώστε να μπορούν να μετρήσουν γρήγορα την προβολή και την επιρροή πρόσφατων άρθρων τους σε επιστημονικά δημοσιεύματα. Επίσης, συνοψίζουν πρόσφατες αναφορές σε διάφορες δημοσιεύσεις, για να βοηθήσουν τους συγγραφείς να αποφασίσουν πού να δημοσιεύσουν τη νέα έρευνα τους.

Το σύστημα *Google Scholar* περιλαμβάνει έναν τεράστιο αριθμό πηγών, πολλές από τις οποίες αφορούν δημοσιεύματα «γκρίζας βιβλιογραφίας» (π.χ. ακαδημαϊκές εργασίες, διατριβές, τεχνικές εκθέσεις, παραδοτέα ερευνητικών έργων). Παρά το μεγάλο εύρος σε περιεχόμενο, τα μετα-δεδομένα των εγγραφών είναι περιορισμένα και απουσιάζουν κρίσιμα στοιχεία για τη γεωγραφική προέλευση ή τη διεύθυνση των οργανισμών στους οποίους ανήκουν οι συγγραφείς. Επιπλέον, τα κριτήρια εισαγωγής των πηγών του περιεχομένου είναι περιορισμένα, ενώ δεν διατίθενται ακριβή στοιχεία για τη γεωγραφική ή τη θεματική κάλυψη που παρέχει το σύστημα.

Οι μετρικές ή αλλιώς οι δείκτες που παρέχει ο Μελετητής της *Google* είναι τρεις και ορίζονται ως ακολούθως.

1. **Δείκτης *h* (*h-index*):** Ο δείκτης *h* ενός συγγραφέα είναι ο μεγαλύτερος αριθμός, ώστε τουλάχιστον *h* άρθρα του συγγραφέα να έχουν αναφερθεί τουλάχιστον *h* φορές το κάθε ένα.
2. **Πυρήνας *h* (*h-core*):** Ο πυρήνας *h* ενός συγγραφέα είναι ένα σύνολο από κορυφαία αναφερόμενα *h* άρθρα του συγγραφέα. Αυτά είναι τα άρθρα στα οποία βασίζεται ο δείκτης *h*.
3. **Ενδιάμεσος *h* (*h-median*):** Ο ενδιάμεσος *h* ενός συγγραφέα είναι η ενδιάμεση τιμή των παραπομπών του πυρήνα *h*. Είναι ένα μέτρο της κατανομής των παραπομπών στα άρθρα του *h-core*.

Επίσης, ορίζονται ανάλογα και ο δείκτης *h-5*, ο πυρήνας *h-5*, και ο ενδιάμεσος *h-5*, όπου πλέον λαμβάνονται υπόψη μόνο τα άρθρα εκείνα που δημοσιεύθηκαν στα τελευταία πέντε έτη. Συνήθως, για τα περισσότερα άρθρα που είναι καταχωρημένα στη βάση δεδομένων του μελετητή της *Google* υπολογίζεται ο πενταετής δείκτης *h* (*h-5 index*) και ο πενταετής ενδιάμεσος *h* (*h-5 median*).

Οι μετρικές αυτές καλύπτουν άρθρα που δημοσιεύθηκαν μεταξύ των ετών 2010 και 2014 και περιλαμβάνουν άρθρα περιοδικών από ιστοσελίδες που ακολουθούν τις κατευθυντήριες γραμμές του μελετητή, επιλεγμένα άρθρα συνεδρίων για την Επιστήμη των Υπολογιστών και των Ηλεκτρολόγων Μηχανικών και προδημοσιεύσεις από το arXiv, AMKA, NBER, και RePEc. Επίσης, βασίζονται σε αναφορές από όλα τα άρθρα τα οποία περιλήφθηκαν στο *Google Scholar*, τον Ιούνιο του 2015. Το *Google Scholar* περιλαμβάνει επίσης αναφορές από άρθρα που δεν καλύπτονται από τις μετρικές της Google [43]. Στη συνέχεια, παρατίθεται παράδειγμα για τον υπολογισμό του δείκτη h , του πυρήνα h και του ενδιάμεσου h .

Παράδειγμα 7. Έστω ότι έχουμε 5 άρθρα ενός συγγραφέα με αναφορές αντίστοιχα το καθένα 17, 9, 6, 3, 2. Τότε ο δείκτης h θα είναι ίσος με 3 επειδή από τα 5 άρθρα υπάρχουν τουλάχιστον 3 με τουλάχιστον 3 αναφορές το καθένα. Ο πυρήνας h θα είναι επίσης ίσος με 3 καθώς από τα 5 άρθρα τα 3 έχουν τις περισσότερες αναφορές, δηλαδή τα άρθρα με 17, 9 και 6 αναφορές. Τέλος, ο ενδιάμεσος h θα είναι η ενδιάμεση τιμή των παραπομπών του πυρήνα h . Ο πυρήνας h αποτελείται από 3 άρθρα με παραπομπές 17, 9 και 6 αντίστοιχα. Άρα η τιμή του ενδιάμεσου h θα είναι το 9.

1.2.10 Ιδιοπαράγοντας (*Eigenfactor*)

Η βαθμολογία *Eigenfactor* αναπτύχθηκε από τους Kevin West και Carl Bergstrom στο Πανεπιστήμιο της Ουάσιγκτον. Είναι μια εκτίμηση της συνολικής σημαντικότητας ενός επιστημονικού περιοδικού. Τα περιοδικά βαθμολογούνται ανάλογα με τον αριθμό των εισερχόμενων αναφορών, όπου αναφορές από περιοδικά με μεγάλη βαθμολογία συμβάλλουν περισσότερο στη βαθμολογία του περιοδικού από εκείνα με χαμηλή βαθμολογία. Ως μέτρο της σημαντικότητας, η βαθμολογία *Eigenfactor* μετρά την συνολική απήχηση ενός περιοδικού. Αν όλες οι άλλες συνθήκες είναι ίσες, τα περιοδικά τείνουν να έχουν μεγαλύτερο αντίκτυπο στον τομέα που έχουν τις μεγαλύτερες βαθμολογίες *Eigenfactor* [12].

Οι βαθμολογίες *Eigenfactor* υπολογίζονται και μπορούν να προβληθούν ελεύθερα από το *eigenfactor.org*. Η βαθμολογία αυτή προορίζεται για τη μέτρηση της σημαντικότητας ενός περιοδικού στην επιστημονική κοινότητα, λαμβάνοντας υπόψη την προέλευση των εισερχόμενων αναφορών και θεωρείται ότι αντικατοπτρίζει το πόσο συχνά ένας μέσος ερευνητής θα έχει πρόσβαση σε περιεχόμενο από αυτό το περιοδικό. Ωστόσο, επηρεάζεται από το μέγεθος του περιοδικού, έτσι ώστε η βαθμολογία διπλασιάζεται όταν το περιοδικό διπλασιάζεται σε μέγεθος (μετριέται με βάση τον αριθμό των δημοσιευμένων άρθρων ετησίως) [35].

Αυτό το μέτρο παίρνει τιμές έτσι ώστε η βαθμολογία *Eigenfactor* όλων των περιοδικών που απαριθμούνται στο *Journal Citation Reports (JCR)* του εκδοτικού οίκου *Thomson Reuters* να αθροίζεται στο 100. Έτσι, αν ένα περιοδικό έχει βαθμολογία ίση με 1,0, αυτό σημαίνει ότι έχει το 1% της συνολικής επίδρασης όλων των καταχωρημένων δημοσιεύσεων. Για παράδειγμα, για το 2013, το περιοδικό *Nature* είχε την υψηλότερη βαθμολογία, με τιμή ίση με 1,603.

Η προσέγγιση της βαθμολογίας *Eigenfactor* θεωρείται ότι είναι πιο ισχυρή από τον δείκτη του παράγοντα απήχησης ο οποίος μετρά εισερχόμενες αναφορές χωρίς να λάβει υπόψη τη σημασία των εν λόγω αναφορών. Αξίζει να σημειωθεί ότι

για ένα δεδομένο αριθμό παραπομπών, παραπομπές από τα σημαντικότερα περιοδικά θα οδηγήσουν σε μια υψηλότερη βαθμολογία.

Αρχικά τα αποτελέσματα της βαθμολογίας αυτής ήταν μέτρα της σημαντικότητας ενός περιοδικού. Σήμερα, έχουν επεκταθεί και σε επίπεδο συγγραφέα και μπορούν να χρησιμοποιηθούν σε συνδυασμό με το *h-index* για να αξιολογήσουν το έργο μεμονωμένων επιστημόνων.

1.2.11 Επιρροή άρθρου (*Article Influence (AI)*)

Η επιρροή άρθρου καθορίζει τη μέση επίδραση των άρθρων ενός περιοδικού για τα πρώτα πέντε έτη μετά τη δημοσίευσή τους. Υπολογίζεται διαιρώντας τη βαθμολογία *Eigenfactor* του περιοδικού με το συνολικό ποσοστό όλων των άρθρων που καταγράφονται στο *Journal Citation Reports* και δημοσιεύθηκαν σε ένα συγκεκριμένο περιοδικό. Στην πραγματικότητα είναι το ίδιο με τη βαθμολογία *Eigenfactor* αλλά σε κανονικοποιημένη μορφή [35]. Το μέτρο αυτό είναι κατά προσέγγιση ανάλογο με τον πενταετή συντελεστή απήχησης (*5-Year Impact Factor*) ενός περιοδικού. Στην ουσία είναι ο πενταετής συντελεστής απήχησης διαιρούμενος με το 2.

Η μέση επιρροή άρθρου είναι 1.00. Επομένως, βαθμολογία μεγαλύτερη από 1.00 δηλώνει ότι κάθε άρθρο στο περιοδικό έχει επιρροή άνω του μέσου όρου. Αντιθέτως, βαθμολογία μικρότερη από 1.00 δηλώνει ότι κάθε άρθρο στο περιοδικό έχει επιρροή κάτω του μέσου όρου.

Ο δείκτης της επιρροής άρθρου είναι ευρέως γνωστός και χρησιμοποιείται αρκετά καθώς έχει πολλά πλεονεκτήματα. Ενδεικτικά αναφέρουμε τα εξής :

- Παρέχεται δωρεάν και είναι προσβάσιμος από παντού.
- Η περίοδος υπολογισμού είναι πέντε χρόνια, χρονική διάρκεια κατάλληλη και δίκαιη για αξιολόγηση.
- Προσπαθεί να δώσει μια πιο ακριβή αναπαράσταση της αξίας των παραπομπών από ότι η ακατέργαστη μέτρηση παραπομπών.

Ο δείκτης επιρροής υπολογίζεται όπως και η βαθμολογία *Eigenfactor* από τη βάση δεδομένων *ISI Web Of Knowledge* του *Thomson Reuters*. Στη συνέχεια, παρατίθεται παράδειγμα για τον υπολογισμό του δείκτη *AI*.

Παράδειγμα 8. Έστω ότι για ένα συγκεκριμένο περιοδικό, η βαθμολογία *Eigenfactor* για το έτος 2014 είναι ίση με 0.009 και ότι το ποσοστό των άρθρων που δημοσιεύτηκαν στο περιοδικό είναι 5,6 % δηλαδή 0,056 τότε ο δείκτης απήχησης άρθρου είναι:

$$AI = \frac{0.009}{0.056} = 1.607 .$$

ΚΕΦΑΛΑΙΟ 2

ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ

2.1 Εισαγωγή

Στο παρόν κεφάλαιο πραγματοποιείται μία επισκόπηση βασικών και ευρέως γνωστών μεθόδων της στατιστικής που θα χρησιμοποιηθούν στην επιστημονική μας έρευνα με σκοπό τον υπολογισμό των αποτελεσμάτων καθώς και την εξαγωγή των συμπερασμάτων μας. Αν και η Στατιστική ως επιστήμη προτείνει ένα μεγάλο σύνολο από τεχνικές για την οργάνωση και την ανάλυση των δεδομένων, αυτές οι τεχνικές μπορούν να κατηγοριοποιηθούν σε δύο γενικές κατηγορίες.

Η πρώτη κατηγορία, η Περιγραφική Στατιστική (*Descriptive Statistics*), περιλαμβάνει μεθόδους για την οργάνωση, απλοποίηση και συνοπτική παρουσίαση των δεδομένων. Αν και υπάρχουν πολλές τεχνικές που ανήκουν σε αυτή την κατηγορία η πιο διαδεδομένη είναι ο υπολογισμός των μέτρων κεντρικής τάσης και των μέτρων διασποράς. Άλλοι περιγραφικοί στατιστικοί δείκτες θα παρουσιαστούν συνοπτικά στη συνέχεια του κεφαλαίου.

Η δεύτερη κατηγορία, η Επαγωγική Στατιστική (*Inferential Statistics*), περιλαμβάνει τεχνικές που επιτρέπουν την ανάλυση δεδομένων που προέρχονται από ένα υποσύνολο του πληθυσμού που ονομάζεται δείγμα έτσι ώστε να γίνει εφικτή η εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων για τους πληθυσμούς, με βάση τις πληροφορίες που συλλέγονται από τα δείγματα.

Οι δύο αυτές κατηγορίες περιλαμβάνουν το μεγαλύτερο τμήμα στατιστικών μεθόδων που χρησιμοποιεί η Στατιστική ως επιστήμη, ένα μέρος των οποίων θα αναλυθεί στη συνέχεια με απλό και κατανοητό τρόπο [23].

2.2 Περιγραφική στατιστική

Τα σύνολα δεδομένων που χρησιμοποιούνται για την διεξαγωγή μιας έρευνας είναι συνήθως αρκετά μεγάλα σε αριθμό με αποτέλεσμα να δυσχεραίνεται η διαδικασία διεξαγωγής αποτελεσμάτων ενός πειράματος. Κρίνεται λοιπόν αναγκαία η εύρεση διαδικασιών, με τις οποίες τα αποτελέσματα αυτά μπορούν να οργανωθούν και να παρουσιαστούν με απλό και εύληπτο τρόπο. Τη λύση στο πρόβλημα δίνουν τα περιγραφικά στατιστικά μέτρα τα οποία παρέχουν μεθόδους που απλοποιούν και διευκολύνουν την οργάνωση και την παρουσίαση των αποτελεσμάτων [23]. Στη συνέχεια θα δούμε ενδεικτικά κάποια από τα πιο γνωστά περιγραφικά στατιστικά μέτρα που χρησιμοποιούνται σήμερα στις περισσότερες έρευνες.

Τα μέτρα αυτά υπολογίζονται συνήθως για ποσοτικές μεταβλητές και όχι για ποιοτικές. Με τον όρο ποσοτικές μεταβλητές εννοούμε τις μεταβλητές των οποίων η τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μια μονάδα μέτρησης. Παραδείγματα τέτοιων μεταβλητών είναι το ύψος, το βάρος, το εισόδημα. Οι ποσοτικές μεταβλητές διακρίνονται με την σειρά τους σε συνεχείς και διακριτές μεταβλητές. Οι συνεχείς (*continuous*) ποσοτικές μεταβλητές είναι εκείνες οι μεταβλητές όπου το σύνολο των δυνατών τιμών που μπορούν να πάρουν είναι

συνεχές υποσύνολο των πραγματικών αριθμών. Οι διακριτές (*discrete*) ποσοτικές μεταβλητές έχουν σύνολο τιμών το υποσύνολο των φυσικών αριθμών. Με τον όρο ποιοτικές μεταβλητές εννοούμε τις μεταβλητές που αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό, του οποίου οι τιμές δεν έχουν αριθμητικές ιδιότητες. Τέτοιες μεταβλητές είναι για παράδειγμα, το φύλο, το μορφωτικό επίπεδο, η περιοχή κατοικίας ενός ατόμου. Μια ποιοτική μεταβλητή μπορεί να διακριθεί επιπλέον σε μια απ' τις δύο κατηγορίες: κατηγορική (*nominal*) και διάταξης (*ordinal*).

- **Κατηγορική (*nominal*):** Η μόνη προσδιορισμένη σχέση μεταξύ των κατηγοριών είναι η ύπαρξη διαφοράς. Το σύνολο τιμών τέτοιων μεταβλητών δεν έχει καμία ιδιότητα.
- **Διάταξης (*ordinal*):** Η διάταξη το μόνο που εξασφαλίζει είναι ο προσδιορισμός της 'καλύτερης', της 'προτιμότερης', της 'μεγαλύτερης' κατηγορίας αλλά όχι και το πόσο 'καλύτερη', 'προτιμότερη', 'μεγαλύτερη' είναι η κατηγορία σε σχέση με τις υπόλοιπες [31].

2.2.1 Μέτρα κεντρικής τάσης

Μια από τις πιο γνωστές κατηγορίες μέτρων περιγραφικής στατιστικής είναι τα μέτρα κεντρικής τάσης. Ο σκοπός των μέτρων κεντρικής τάσης (*measures of central tendency*) είναι να προσδιοριστεί ένα στατιστικό μέγεθος το οποίο μπορεί να χρησιμοποιηθεί για να αντιπροσωπεύσει ένα σύνολο δεδομένων. Για την επίτευξη αυτού του στόχου χρησιμοποιούνται συνήθως τρία μέτρα κεντρικής τάσης: η μέση τιμή (*mean*), η διάμεσος (*median*) και η επικρατούσα τιμή (*mode*) [30].

1. Μέση Τιμή (*mean*): Είναι το συνηθέστερο μέτρο κεντρικής τάσης καθώς δίνει την πιο αντιπροσωπευτική εικόνα για ένα σύνολο δεδομένων. Υπολογίζεται συνήθως για ποσοτικές μεταβλητές που δεν έχουν ελλειπείς τιμές. Ορίζεται ως το αλγεβρικό άθροισμα των τιμών μιας μεταβλητής διαιρούμενο δια του πλήθους τους.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

2.Διάμεσος (*median*): Όταν υπάρχουν απροσδιόριστες τιμές στα δεδομένα, ή έχουμε διαθέσιμα ποιοτικά δεδομένα ή η μέση τιμή δεν παρέχει μια αντιπροσωπευτική εικόνα για τα δεδομένα χρησιμοποιείται η διάμεσος ως μέτρο κεντρικής τάσης. Είναι η τιμή η οποία βρίσκεται στο μέσο των τιμών της κατανομής όταν αυτές διαταχθούν σε αύξουσα ή φθίνουσα διάταξη. Πιο αναλυτικά, για τον υπολογισμό της διαμέσου διατάσσουμε τις παρατηρήσεις σε αύξουσα ή φθίνουσα σειρά, και αν το πλήθος τους είναι περιττό επιλέγουμε τη μεσαία παρατήρηση, ενώ αν είναι άρτιο η τιμή της διαμέσου δίνεται από το άθροισμα των μεσαίων παρατηρήσεων.

3.Επικρατούσα Τιμή (*mode*): Ένα άλλο μέτρο κεντρικής τάσης είναι η επικρατούσα τιμή η οποία λειτουργεί συμπληρωματικά με τα δύο προηγούμενα μέτρα. Επίσης είναι ένα μέτρο που υπολογίζεται όταν είναι διαθέσιμα ποιοτικά δεδομένα με τη μορφή ονομαστικών κατηγοριών, όπου δεν έχει νόημα ο υπολογισμός της μέσης τιμής ή της διαμέσου. Ορίζεται ως η παρατήρηση με τη μεγαλύτερη συχνότητα εμφάνισης.

2.2.2 Μέτρα μεταβλητότητας

Εκτός από την κεντρική τάση μας ενδιαφέρει επίσης και η μεταβλητότητα ή διασπορά των παρατηρήσεων. Όταν τα δεδομένα είναι συγκεντρωμένα γύρω από μια κεντρική τιμή, δηλαδή η διασπορά των δεδομένων είναι μικρή, τότε η κεντρική τιμή αντιπροσωπεύει ικανοποιητικά τα δεδομένα. Από την άλλη, όταν τα δεδομένα είναι πολύ σκορπισμένα τα μέτρα κεντρικής τιμής δε δίνουν καλή περιληπτική περιγραφή των δεδομένων. Επίσης, διαφορετικά δείγματα από τον ίδιο πληθυσμό μπορεί να έχουν το ίδιο μέτρο κεντρικής τάσης αλλά να διαφέρουν κατά κάποιο σημαντικό τρόπο ως προς τη διασπορά των παρατηρήσεων. Τα κυριότερα μέτρα διασποράς ή αλλιώς μεταβλητότητας είναι το εύρος (*range*), η τυπική απόκλιση (*standard deviation*) και η διακύμανση (*variance*) [30].

1. Εύρος (*range*): Το εύρος ορίζεται ως η διαφορά της μεγαλύτερης παρατήρησης από τη μικρότερη. Το εύρος θεωρείται ως ένα πρόχειρο μέτρο διασποράς, καθώς βασίζεται μόνο στις ακραίες παρατηρήσεις και επηρεάζεται σημαντικά από το μέγεθος του δείγματος, καθώς κάθε νέα παρατήρηση μπορεί να αλλάξει την τιμή της μικρότερης ή της μεγαλύτερης παρατήρησης στο δείγμα. Δίνεται από τον τύπο:

$$R = x_{\max} - x_{\min}$$

2. Διακύμανση (*variance*): Η διασπορά ή διακύμανση μετράει τη μεταβλητότητα των παρατηρήσεων γύρω από τη μέση τιμή. Αν ορίσουμε την απόκλιση μιας παρατήρησης x_i από τη μέση τιμή ως $x_i - \bar{x}$, είναι φανερό πως το άθροισμα όλων αυτών των αποκλίσεων είναι 0 γιατί χρησιμοποιώντας τον ορισμό της δειγματικής μέσης τιμής έχουμε :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

3. Τυπική απόκλιση (*standard deviation*): Η διακύμανση είναι μια αξιόπιστη παράμετρος διασποράς, αλλά έχει ένα μειονέκτημα. Δεν εκφράζεται με τις μονάδες με τις οποίες εκφράζονται οι παρατηρήσεις. Για παράδειγμα, αν οι παρατηρήσεις εκφράζονται σε cm, η διακύμανση εκφράζεται σε cm^2 . Αν όμως πάρουμε τη θετική τετραγωνική ρίζα της διακύμανσης, θα έχουμε ένα μέτρο διασποράς που θα εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού, όπως ακριβώς είναι και όλα τα άλλα μέτρα θέσης, που εξετάσαμε έως τώρα. Η ποσότητα αυτή λέγεται τυπική απόκλιση (*standard deviation*), συμβολίζεται με s και δίνεται από τη σχέση:

$$s = \sqrt{s^2}$$

4. Συντελεστής Μεταβλητότητας (*coefficient of variation*): Ένα άλλο εξίσου σημαντικό μέτρο μεταβλητότητας είναι ο συντελεστής μεταβλητότητας ο οποίος εκφράζει την διασπορά των τιμών σε σχέση με τη μέση τιμή. Είναι καθαρός αριθμός, απαλλαγμένος από τις μονάδες μέτρησης της μεταβλητής για την οποία υπολογίζεται. Επιπλέον, χρησιμοποιείται για συγκρίσεις ομάδων μεταξύ τους αλλά και για την εξέταση της ομοιογένειας μέσα στην ίδια ομάδα. Συμβολίζεται με CV και όταν δεν ξεπερνά το 10% ($CV < 10\%$) λέμε ότι το δείγμα είναι ομοιογενές [31]. Δίνεται από τις

εξής σχέσεις ανάλογα με το αν αναφερόμαστε σε ολόκληρο τον πληθυσμό ή σε ένα δείγμα αυτού:

$$CV = \frac{s}{\bar{x}} \cdot 100 \quad \text{ή} \quad CV = \frac{\sigma}{\mu} \cdot 100$$

2.2.3 Μέτρα ασυμμετρίας

Η κατανομή ενός πληθυσμού μπορεί να είναι είτε συμμετρική είτε ασύμμετρη. Στην πρώτη περίπτωση η κορυφή, η διάμεσος και η μέση τιμή συμπίπτουν. Σε οποιαδήποτε άλλη περίπτωση ένα από τα τμήματα στα οποία χωρίζει την κατανομή η κορυφή έχει περισσότερες παρατηρήσεις απ' το άλλο. Οι ασυμμετρίες που εμφανίζονται είναι δύο ειδών. Η θετική ασυμμετρία, όπου οι περισσότερες παρατηρήσεις καθώς και η μέση τιμή και η διάμεσος βρίσκονται δεξιά της κορυφής και η αρνητική ασυμμετρία όπου σε αυτήν την περίπτωση οι περισσότερες παρατηρήσεις καθώς και η μέση τιμή και η διάμεσος βρίσκονται αριστερά της κορυφής. Ως μέτρο υπολογισμού της ασυμμετρίας χρησιμοποιείται συνήθως ο συντελεστής ασυμμετρίας με βάση τις ροπές και δίνεται ως εξής:

$$\gamma = \frac{\frac{1}{n} \cdot \sum_1^n (x_i - \bar{x})^3}{\left\{ \sqrt{\frac{1}{n} \cdot \sum_1^n (x_i - \bar{x})^2} \right\}^3}$$

Όταν $\gamma > 0$ έχουμε θετική ασυμμετρία, όταν $\gamma < 0$ έχουμε αρνητική ασυμμετρία ενώ για $\gamma = 0$ έχουμε συμμετρία [31].

2.2.4 Μέτρα κυρτότητας

Μια κατανομή η οποία έχει σχετικά μεγάλη μέγιστη συχνότητα και επομένως μεγάλη συγκέντρωση τιμών γύρω απ' τη μέση τιμή λέγεται λεπτόκυρτη (*leptokurtic*) ενώ αν η μέγιστη συχνότητα είναι σχετικά μικρή λέγεται πλατύκυρτη (*platykurtic*). Κατανομές που προσεγγίζουν την κανονική κατανομή λέγονται μεσόκυρτες (*mesokurtic*). Ένα μέτρο που εκφράζει τον βαθμό κυρτότητας μιας κατανομής είναι ο συντελεστής κύρτωσης του *Pearson* ο οποίος δίνεται από τον τύπο:

$$a = \frac{\frac{1}{n} \cdot \sum_1^n (x_i - \bar{x})^4}{\left\{ \sqrt{\frac{1}{n} \cdot \sum_1^n (x_i - \bar{x})^2} \right\}^4}$$

Επειδή για κανονικές κατανομές έχουμε $a = 3$ συνηθίζεται να μετράμε την κυρτότητα συγκρίνοντας την με αυτή την τιμή. Για λεπτόκυρτες κατανομές έχουμε $a > 3$ (θετική κύρτωση) ενώ για πλατύκυρτες $a < 3$ (αρνητική κύρτωση) [31].

2.3 Παραμετρικοί και μη παραμετρικοί έλεγχοι υποθέσεων

Μια στατιστική υπόθεση είναι μια δήλωση για μια ή περισσότερες κατανομές του πληθυσμού η οποία ελέγχεται με βάση τις παρατηρήσεις ενός ή περισσότερων αντίστοιχα τυχαίων δειγμάτων. Ανάλογα με το είδος των υποθέσεων οι οποίες μπορούν να γίνουν δεκτές, οι στατιστικοί έλεγχοι διακρίνονται σε δύο γενικές κατηγορίες.

Παραμετρικοί έλεγχοι ονομάζονται οι έλεγχοι των υποθέσεων για μία ή περισσότερες παραμέτρους του πληθυσμού οι οποίοι γίνονται κάτω απ' την υπόθεση ότι η κατανομή του πληθυσμού είναι γνωστή (συνήθως κανονική). Οι έλεγχοι αυτοί μπορούν να εφαρμοστούν και σε πληθυσμούς με άγνωστη κατανομή αρκεί να έχουμε μεγάλο πλήθος δείγματος.

Μη παραμετρικοί έλεγχοι ονομάζονται όλοι οι υπόλοιποι έλεγχοι, όπως οι έλεγχοι για τη θέση, τη διασπορά μιας ή περισσότερων κατανομών ενός πληθυσμού. Στην κατηγορία αυτοί υπάγονται και έλεγχοι για τη διάμεσο. Οι μη παραμετρικοί έλεγχοι εφαρμόζονται όταν η κατανομή του πληθυσμού δεν είναι γνωστή και το μέγεθος του δείγματος είναι μικρό [36].

Σε έναν έλεγχο υποθέσεων έχουμε την μηδενική υπόθεση η οποία περιγράφει την υπάρχουσα κατάσταση και συμβολίζεται με H_0 και την εναλλακτική υπόθεση η οποία είναι αντίθετη της μηδενικής και υπαγορεύει αλλαγή ενεργειών. Η εναλλακτική υπόθεση συμβολίζεται με H_1 . Η λογική ελέγχου υποθέσεων μπορεί να συνοψιστεί στα εξής τέσσερα βήματα:

1. Διατύπωση υποθέσεων
2. Διαμόρφωση των κριτηρίων για την λήψη μιας απόφασης
3. Συλλογή δεδομένων από δείγματα
4. Αξιολόγηση της μηδενικής υπόθεσης

Στη συνέχεια της παραγράφου θα δούμε βασικές κατηγορίες παραμετρικών και μη παραμετρικών ελέγχων και το πώς αυτοί εφαρμόζονται στον έλεγχο υποθέσεων.

2.3.1 Το t στατιστικό τεστ ενός δείγματος (παραμετρικός έλεγχος)

Το t στατιστικό τεστ ενός δείγματος (*single sample t-test*) χρησιμοποιεί δεδομένα που προέρχονται από ένα δείγμα για να ελέγξει υποθέσεις που σχετίζονται με τη μέση τιμή ενός πληθυσμού όταν η διακύμανση του αρχικού πληθυσμού είναι άγνωστη. Σε αυτή τη στατιστική μέθοδο γίνεται μία εκτίμηση της διακύμανσης του πληθυσμού με τη βοήθεια της διακύμανσης του δείγματος. Η μηδενική υπόθεση δηλώνει μία συγκεκριμένη τιμή για την μέση τιμή του άγνωστου πληθυσμού που προκύπτει από την αλλαγή της ανεξάρτητης μεταβλητής. Για την εύρεση των κρίσιμων περιοχών ο ερευνητής επιλέγει ένα επίπεδο σημαντικότητας, υπολογίζει του βαθμούς ελευθερίας που δίνονται από τον τύπο $df = n-1$ (όπου n το μέγεθος τους δείγματος), και συμβουλευέται το στατιστικό πίνακα κατανομής t . Ο υπολογισμός της τιμής του στατιστικού δείκτη ελέγχου t δίνεται από την παρακάτω σχέση:

$$t = \frac{\bar{X} - \mu}{s_{\bar{x}}}, \quad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Το t στατιστικό τεστ ενός δείγματος μπορεί να χρησιμοποιηθεί μόνο όταν τηρούνται ορισμένες προϋποθέσεις οι οποίες είναι:

- a) Το δείγμα είναι αντιπροσωπευτικό και οι τιμές που το απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις
- b) Η κατανομή των τιμών του δείγματος είναι κανονική [23]

2.3.2 Το t στατιστικό τεστ ανεξάρτητων δειγμάτων (παραμετρικός έλεγχος)

Το t στατιστικό τεστ ανεξάρτητων δειγμάτων (*independent measures t-test*) χρησιμοποιεί δεδομένα που προέρχονται από δύο ξεχωριστά δείγματα για να ελέγξει υποθέσεις που σχετίζονται με την διαφορά των μέσων τιμών δύο πληθυσμών. Σε αυτή την περίπτωση η μηδενική υπόθεση δηλώνει ότι δεν υφίσταται διαφορά ανάμεσα στις μέσες τιμές των δύο πληθυσμών. Για τον υπολογισμό της τιμής του στατιστικού δείκτη ελέγχου t και του βαθμού ελευθερίας που καθορίζει τις κρίσιμες περιοχές στο στατιστικό πίνακα κατανομής t , χρησιμοποιούνται οι εξής μαθηματικές σχέσεις:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}-\bar{x}}}, s_{\bar{x}-\bar{x}} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}, s_p^2 = \sqrt{\frac{SS_1 + SS_2}{df_1 + df_2}}$$

$$df = df_1 + df_2 = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

Οι προϋποθέσεις που πρέπει να ισχύουν για να χρησιμοποιηθεί το t στατιστικό τεστ ανεξάρτητων δειγμάτων είναι:

- a) Τα δείγματα είναι αντιπροσωπευτικά και οι τιμές που τα απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις
- b) Η κατανομή των τιμών των δειγμάτων είναι κανονική
- c) Οι δύο πληθυσμοί από τους οποίους έχουν επιλεγεί τα δύο δείγματα έχουν την ίδια διακύμανση (δηλαδή $\sigma_1 \approx \sigma_2$) [23].

2.3.3 Το t στατιστικό τεστ εξαρτημένων δειγμάτων (παραμετρικός έλεγχος)

Εύκολα γίνεται αντιληπτό ότι η χρήση ανεξάρτητων δειγμάτων δεν είναι ο μοναδικός τρόπος για να σχεδιαστεί ένα πείραμα το οποίο εξετάζει την επίδραση μιας ελεγχόμενης μεταβλητής σε μία παρατηρούμενη. Μια διαφορετική προσέγγιση αποτελεί η χρήση ενός μόνο δείγματος υποκειμένων, το οποίο εξετάζεται σε όλες τις διαφορετικές τιμές της ελεγχόμενης μεταβλητής. Με άλλα λόγια, επαναλαμβάνοντας τις μετρήσεις στο ίδιο σύνολο υποκειμένων και για διαφορετικές συνθήκες, προσπαθούμε να εντοπίσουμε διαφορές στη συμπεριφορά τους. Οι ερευνητικές μελέτες που βασίζονται σε πειράματα σχεδιασμένα με αυτή την προσέγγιση ονομάζονται μελέτες επαναλαμβανόμενων μετρήσεων (*repeated-measures studies*).

Το σημαντικό πλεονέκτημα αυτών των μελετών, συγκριτικά με τις μελέτες που βασίζονται σε πειράματα με ανεξάρτητα δείγματα, είναι το γεγονός ότι δεν επιτρέπουν στη διαφορετικότητα των υποκειμένων να επιδράσει στα αποτελέσματα της μελέτης, αφού χρησιμοποιούν το ίδιο σύνολο υποκειμένων σε όλα τα πειράματα. Επίσης, οι μελέτες επαναλαμβανόμενων μετρήσεων, λόγω του τρόπου σχεδιασμού τους, χρειάζονται μικρότερο συνολικό αριθμό υποκειμένων, κάτι που μπορεί να είναι πολύ σημαντικό σε ορισμένες περιπτώσεις. Το μόνο μειονέκτημα αυτών των μελετών είναι ότι, επειδή ακριβώς χρησιμοποιούν τον ίδιο αριθμό υποκειμένων σε διαφορετικά πειράματα, υπάρχει ο κίνδυνος να επιδράσουν άλλοι ανεπιθύμητοι

παράγοντες (*carry-over effects and progressive error*) στα αποτελέσματα. Το τεστ αυτό χρησιμοποιεί δύο σύνολα μετρήσεων που προέρχονται από ένα δείγμα για να ελέγξει υποθέσεις που σχετίζονται με τη διαφορά των μέσων τιμών τους. Σε αυτήν την περίπτωση, η μηδενική υπόθεση δηλώνει ότι δεν υφίσταται διαφορά ανάμεσα στις μέσες τιμές των δύο συνόλων μετρήσεων. Για την απλοποίηση των μαθηματικών τύπων που επιτρέπουν τον υπολογισμό της τιμής του στατιστικού ελέγχου t , συνήθως εισάγεται ο όρος "διαφορά μετρήσεων D " (*difference score D*) που εκφράζει για κάθε υποκείμενο τη διαφορά ανάμεσα στις μετρήσεις κάθε πειράματος. Οι μαθηματικές σχέσεις που χρειάζονται σε ένα t στατιστικό τεστ με επαναλαμβανόμενες μετρήσεις είναι οι εξής:

$$t = \frac{\bar{D} - \mu_D}{s_{\bar{D}}}, \quad D = X_1 - X_2$$

$$\bar{D} = \frac{\sum_n D}{n}, \quad \text{με } s_{\bar{D}} = \frac{s^2}{n}$$

$$df = n - 1$$

Οι προϋποθέσεις που πρέπει να ισχύουν για να χρησιμοποιηθεί το t στατιστικό τεστ με επαναλαμβανόμενες μετρήσεις είναι:

- a) Το δείγμα είναι αντιπροσωπευτικό και οι τιμές οφείλονται σε ανεξάρτητες παρατηρήσεις
- b) Η κατανομή των τιμών του D είναι κανονική [23]

Όλα τα τεστ που αναφέρθηκαν στις προηγούμενες ενότητες αφορούν παραμετρικό έλεγχο ποσοτικών δεδομένων. Στις επόμενες ενότητες θα γίνει ανάλυση μη παραμετρικών τεστ που αφορούν ποσοτικά αλλά και ποιοτικά δεδομένα.

2.3.4 Το χ^2 στατιστικό τεστ (μη παραμετρικός έλεγχος)

Τα στατιστικά τεστ που παρουσιάστηκαν παραπάνω ανήκουν στην κατηγορία των παραμετρικών τεστ (*parametric tests*). Τα παραμετρικά τεστ χρησιμοποιούν ποσοτικά δεδομένα και μπορούν να εφαρμοστούν μόνο όταν ικανοποιούνται ορισμένες προϋποθέσεις. Σε περιπτώσεις όπου ο ερευνητής εξετάζει ποιοτικά δεδομένα, χρησιμοποιεί ένα άλλο σύνολο στατιστικών ελέγχου υποθέσεων, τα οποία ονομάζονται μη παραμετρικά (*non-parametric tests*). Όταν λοιπόν ένα παραμετρικό τεστ δε μπορεί να χρησιμοποιηθεί, συνήθως υπάρχει ένα μη παραμετρικό τεστ που μπορεί να το αντικαταστήσει. Αν και τα μη παραμετρικά τεστ μπορούν να χρησιμοποιηθούν και σε περιπτώσεις που είναι δυνατή η εφαρμογή ενός παραμετρικού τεστ, συνήθως αυτό αποφεύγεται γιατί τα παραμετρικά τεστ δίνουν πιο αξιόπιστα αποτελέσματα.

Το στατιστικό τεστ χ^2 (*Chi-Square test*) είναι ίσως το πιο δημοφιλές μη παραμετρικό τεστ. Στην περίπτωση αυτή, ο ερευνητής έχει στη διάθεση του ένα δείγμα ποιοτικών δεδομένων οργανωμένο σε ονομαστικές κατηγορίες και ο στόχος είναι να χρησιμοποιηθούν τα δεδομένα αυτά ώστε να προσδιοριστεί η αναλογία (ή το ποσοστό) του πληθυσμού που ανήκει στην κάθε κατηγορία. Για την επίτευξη αυτού του στόχου διατυπώνεται μια μηδενική υπόθεση, που είτε δηλώνει ότι δεν υπάρχει κάποια συγκεκριμένη προτίμηση στις διαθέσιμες ονομαστικές κατηγορίες (*no-*

preference null-hypothesis), είτε δηλώνει ότι τα ποσοστά που προτιμώνται από τα υποκείμενα δε διαφέρουν από τα ποσοστά άλλων πληθυσμών οι οποίοι αποτελούν το σημείο αναφοράς (*no difference from a comparison population*).

Και στις δύο περιπτώσεις, αυτό που προσδιορίζει η μηδενική υπόθεση είναι ο αναμενόμενος αριθμός (*expected frequency –E*) των υποκειμένων που ανήκει σε κάθε ονομαστική κατηγορία. Ο έλεγχος υποθέσεων που ακολουθεί αξιολογεί αυτή τη μηδενική υπόθεση, συγκρίνοντας τον αριθμό των αναμενόμενων υποκειμένων με τον αριθμό των παρατηρούμενων υποκειμένων που παρατηρείται ότι ανήκει σε κάθε ονομαστική κατηγορία (*observed frequency –O*), με βάση τις μετρήσεις του δείγματος. Ο στατιστικός δείκτης ελέγχου που χρησιμοποιείται για αυτή την αξιολόγηση είναι το X^2 [23]. Οι κρίσιμες τιμές για το στατιστικό δείκτη ελέγχου δίνονται από την κατανομή X^2 , ενώ οι τύποι υπολογισμού δίνονται παρακάτω:

$$X^2 = \frac{(O - E)^2}{E}, \text{ όπου } O, E \text{ η παρατηρούμενη και η αναμενόμενη συχνότητα}$$

$$df = C - 1, \text{ όπου } C \text{ ο αριθμός των ονομαστικών κατηγοριών}$$

Αν και όπως αναφέραμε, τα μη παραμετρικά τεστ δε χρειάζεται να ικανοποιούν συγκεκριμένες προϋποθέσεις, τα συμπεράσματα που προκύπτουν από την εφαρμογή του X^2 στατιστικού τεστ είναι πιο αξιόπιστα όταν χρησιμοποιείται αντιπροσωπευτικό δείγμα και η αναμενόμενη συχνότητα όλων των κατηγοριών είναι μεγαλύτερη από πέντε.

2.3.5 Το Kolmogorov-Smirnov τεστ (μη παραμετρικός έλεγχος)

Για να ελέγξουμε αν η κατανομή μιας μεταβλητής είναι συμβατή με την κανονική εφαρμόζουμε το τεστ *Kolmogorov-Smirnov*. Στη βιβλιογραφία συχνά συμβολίζεται και με τα αρχικά K-S. Το τεστ K-S χρησιμοποιείται για το έλεγχο καλής προσαρμογής ενός τυχαίου δείγματος σε μία δεδομένη συνεχή κατανομή. Το κριτήριο K-S βασίζεται στην διαφορά της εμπειρικής συνάρτησης κατανομής (που προέρχεται από το δείγμα) και της αναμενόμενης F_0 . Πιο συγκεκριμένα, αν X_1, X_2, \dots, X_n είναι ένα τυχαίο δείγμα η εμπειρική συνάρτηση κατανομής (ΕΣΚ) του δείγματος αυτού είναι:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad I_C(X_i) = \begin{cases} 1, & X_i \in C \\ 0, & X_i \notin C \end{cases}$$

Επομένως, η μηδενική υπόθεση H_0 υποθέτει ότι η ΕΣΚ πρέπει να είναι κοντά στην F_0 ενώ αντίθετα η εναλλακτική υπόθεση H_1 υποθέτει ότι πρέπει να υπάρχει σημαντική απόκλιση της ΕΣΚ από την F_0 . Για να κατασκευάσουμε έναν έλεγχο με βάση αυτόν τον συλλογισμό, θα πρέπει να ορίσουμε μία «απόσταση» μεταξύ των δύο κατανομών (της ΕΣΚ και της F_0) και να απορρίπτουμε H_0 την όταν αυτή η απόσταση γίνεται «μμεγάλη» [27]. Σχετικά έχουμε τον επόμενο ορισμό:

Ορισμός: Αν F, G είναι δύο συναρτήσεις κατανομής στον R , τότε η ποσότητα

$$d_k(F, G) = \sup_{x \in R} \{ |F(x) - G(x)| \}$$

καλείται απόσταση *Kolmogorov* μεταξύ της F και της G.

Σύμφωνα με τα παραπάνω, θα απορρίπτουμε την $H_0 : X_i \sim F_0$ όταν η στατιστική συνάρτηση

$$D_n = d_k(\hat{F}_n, F_0) = \sup_{x \in R} \{ |\hat{F}_n(x) - F_0(x)| \}$$

λαμβάνει «ασυνήθιστα» μεγάλες τιμές, δηλαδή όταν $D_n > c$. Το κριτήριο αυτό είναι γνωστό ως κριτήριο *Kolmogorov-Smirnov* (και η στατιστική συνάρτηση D_n καλείται ελεγχосυνάρτηση *Kolmogorov-Smirnov*).

Έστω λοιπόν X_1, X_2, \dots, X_n ένα τυχαίο δείγμα μεγέθους n από κάποια άγνωστη κατανομή με αθροιστική συνάρτηση κατανομής $F(x)$. Έστω, επίσης, $F_0(x)$ μία σαφώς καθορισμένη συνάρτηση κατανομής. Οι υποθέσεις που ενδιαφέρει να ελεγχθούν διακρίνονται στις εξής τρεις κατηγορίες:

1. Αμφίπλευρη εναλλακτική υπόθεση

$$H_0 : F(x) = F_0(x), \text{ για κάθε } x \text{ στο } (-\infty, +\infty)$$

$$H_1 : F(x) \neq F_0(x), \text{ για τουλάχιστον μια τιμή του } x$$

2. Μονόπλευρη εναλλακτική υπόθεση

$$H_0 : F(x) \geq F_0(x), \text{ για κάθε } x \text{ στο } (-\infty, +\infty)$$

$$H_1 : F(x) < F_0(x), \text{ για τουλάχιστον μια τιμή του } x$$

3. Μονόπλευρη εναλλακτική υπόθεση

$$H_0 : F(x) \leq F_0(x), \text{ για κάθε } x \text{ στο } (-\infty, +\infty)$$

$$H_1 : F(x) > F_0(x), \text{ για τουλάχιστον μια τιμή του } x$$

Έστω $F_n(x)$ η εμπειρική συνάρτηση κατανομής με βάση το τυχαίο δείγμα X_1, X_2, \dots, X_n . Η στατιστική συνάρτηση ελέγχου ορίζεται διαφορετικά για τα τρία διαφορετικά σύνολα υποθέσεων ως εξής:

1. **Αμφίπλευρος έλεγχος:** Η κατάλληλη στατιστική συνάρτηση ελέγχου για την περίπτωση αυτή είναι η μέγιστη κατακόρυφη απόσταση μεταξύ των συναρτήσεων $F_n(x)$ και $F_0(x)$. Συμβολικά, γράφουμε:

$$T = \sup_x |F_0(x) - F_n(x)|$$

2. **Μονόπλευρος έλεγχος:** Είναι προφανές ότι, στην περίπτωση αυτή, η κατάλληλη στατιστική συνάρτηση έλεγχου είναι η μέγιστη κατακόρυφη απόσταση που έχει η $F_0(x)$ υπεράνω της $F_n(x)$. Δηλαδή:

$$T^+ = \sup_{x:F_0(x)>F_n(x)} |F_0(x) - F_n(x)|$$

Η στατιστική αυτή συνάρτηση είναι παρόμοια με την D , με την διαφορά ότι, τώρα, θεωρούμε μόνο την μέγιστη διαφορά όταν η συνάρτηση $F_0(x)$ είναι υπεράνω της $F_n(x)$.

3. **Μονόπλευρος έλεγχος:** Με ανάλογο τρόπο, η κατάλληλη στατιστική συνάρτηση έλεγχου ορίζεται ως η μέγιστη κατακόρυφη απόσταση που έχει η $F_n(x)$ υπεράνω της $F_0(x)$. Επομένως:

$$T^- = \sup_{x:F_0(x)<F_n(x)} |F_n(x) - F_0(x)|$$

Είναι προφανές ότι, και στις τρεις περιπτώσεις, οι μεγάλες τιμές της στατιστικής συνάρτησης είναι αυτές που συνηγορούν υπέρ της απόρριψης της μηδενικής υπόθεσης αφού αντανακλούν χαμηλό βαθμό εγγύτητας των τιμών των συναρτήσεων $F_n(x)$ και $F_0(x)$ [27]. Η ακριβής κατανομή των στατιστικών συναρτήσεων T , T^+ , T^- έχει μελετηθεί από τον *Kolmogorov*.

Συγκεντρωτικά, θα μπορούσαμε να συνοψίσουμε τις ακόλουθες παρατηρήσεις για το τεστ *Kolmogorov-Smirnov*:

- Το *Kolmogorov-Smirnov* τεστ καλής προσαρμογής υποθέτει ότι η μηδενική υπόθεση είναι πλήρως ορισμένη, δηλαδή ελέγχουμε μία απλή υπόθεση. Με άλλα λόγια θεωρεί ότι η $F_0(x)$ είναι πλήρως ορισμένη χωρίς την «παρουσία» άγνωστων παραμέτρων.
- Ο στατιστικός έλεγχος των *Kolmogorov-Smirnov* εφαρμόζεται σε κατανομές με συνεχή αθροιστική κατανομή και τείνει να είναι πιο ευαίσθητος στο κέντρο της κατανομής από ότι στις άκρες.

Στη βιβλιογραφία έχουν μελετηθεί αρκετές παραλλαγές του έλεγχου *Kolmogorov-Smirnov*, οι οποίες επιτρέπουν τη χρήση του σε περιπτώσεις όπου οι παράμετροι εκτιμώνται από τα δεδομένα. Στην πραγματικότητα, η στατιστική συνάρτηση παραμένει η ίδια, αλλά χρησιμοποιούνται διαφορετικοί πίνακες ποσοστιαίων σημείων και κρίσιμων τιμών. Οι πίνακες αυτοί δεν είναι οι ίδιοι για όλες τις κατανομές, αλλά εξαρτώνται από τη μηδενική υπόθεση.

Τα τεστ που αναφέρθηκαν στις παραγράφους 2.3.4 και 2.3.5 αποτελούν τεστ καλής προσαρμογής. Οι έλεγχοι αυτοί εφαρμόζονται σε περιπτώσεις όπου θέλουμε να εξετάσουμε αν οι παρατηρήσεις προσαρμόζονται σε κάποιο συγκεκριμένο μοντέλο. Στο πλαίσιο αυτό έχουν αναπτυχθεί πολλές στατιστικές μεθοδολογίες. Στις προηγούμενες παραγράφους εξετάσαμε δύο αρκετά γνωστές από αυτές; Το X^2 τεστ καλής προσαρμογής και το *Kolmogorov-Smirnov* τεστ. Παρόλο που αυτά τα τεστ

χρησιμοποιούνται για παρόμοιο σκοπό και έχουν παρόμοιες ιδιότητες, δεν παύουν να έχουν και αρκετές αξιοσημείωτες διαφορές.

Ο έλεγχος *Kolmogorov* συχνά προτιμάται από τον έλεγχο χ^2 ως έλεγχος καλής προσαρμογής, όταν το μέγεθος του δείγματος είναι μικρό. Ο έλεγχος *Kolmogorov* είναι ακριβής ακόμη και για μικρά δείγματα, ενώ ο έλεγχος χ^2 υποθέτει ότι ο αριθμός των παρατηρήσεων είναι αρκετά μεγάλος, ώστε η κατανομή χ^2 να παρέχει μία ικανοποιητική προσέγγιση της κατανομής της στατιστικής συνάρτησης. Υπάρχουν πολλές αντιφάσεις όσο αφορά το ποιός έλεγχος είναι περισσότερο ισχυρός, αλλά η γενική εντύπωση φαίνεται να είναι ότι ο έλεγχος *Kolmogorov* είναι ενδεχομένως περισσότερο ισχυρός από τον έλεγχο χ^2 στις περισσότερες περιπτώσεις.

Ένας προβληματισμός για τον έλεγχο *Kolmogorov* αναφέρεται στο κατά πόσο αυτός αγνοεί πληροφορίες με το να χρησιμοποιεί μόνο την διαφορά με το μεγαλύτερο μέγεθος σε αντίθεση με τους ελέγχους οι οποίοι λαμβάνουν υπόψη τους όλες τις διαφορές. Τα πλεονεκτήματα των τελευταίων δεν είναι τόσα πολλά γιατί η τιμή της συνάρτησης $F_n(x)$ εξαρτάται σε κάθε στάδιο από το πόσες παρατηρήσεις είναι μικρότερες από την συγκεκριμένη τιμή x . Επομένως, οι συγκρίσεις γίνονται με βάση τις συσσωρευμένες ενδείξεις μέχρι το συγκεκριμένο στάδιο [27]. Μέχρι στιγμής αναφερθήκαμε στο τεστ *Kolmogorov-Smirnov* για ένα δείγμα. Όμως, το τεστ *Kolmogorov-Smirnov* μπορεί να εφαρμοστεί και για δύο ανεξάρτητα δείγματα. Επιπλέον, υπάρχουν πολλές μεθοδολογίες που ανήκουν στην κατηγορία των τεστ καλής προσαρμογής τα όποια όμως δε θα αναφερθούν σε αυτό το κεφάλαιο καθώς είναι απαραίτητο να επεκταθούμε σε άλλες αξιοσημείωτες μεθοδολογίες [28]. Για περισσότερες πληροφορίες για αυτά ο αναγνώστης μπορεί να ανατρέξει στο σύγγραμμα των Παπαϊωάννου & Σωτήρης (2002).

2.4 Ανάλυση παλινδρόμησης

Σε περιπτώσεις όπου θέλουμε να εξετάσουμε τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με σκοπό την πρόβλεψη της μιας απ' αυτές με χρήση των τιμών μιας ή περισσότερων άλλων εφαρμόζουμε ανάλυση παλινδρόμησης. Σε κάθε πρόβλημα παλινδρόμησης διακρίνουμε συνήθως δύο είδη μεταβλητών: τις ανεξάρτητες (*independent*) και τις εξαρτημένες μεταβλητές ή μεταβλητές απόκρισης (*dependent, response variables*). Ανεξάρτητες μεταβλητές, οι οποίες συμβολίζονται συνήθως με X , είναι εκείνες στις οποίες μπορούμε να δίνουμε μια συγκεκριμένη τιμή ή παίρνουν τιμές που μπορούμε να παρατηρήσουμε αλλά όχι να ελέγξουμε. Η εξαρτημένη μεταβλητή, η οποία συμβολίζεται συνήθως με Y , είναι εκείνη η οποία αντανακλά το αποτέλεσμα μεταβολών στις ελεγχόμενες μεταβλητές. Αξίζει να σημειωθεί ότι η διάκριση μεταξύ ανεξαρτήτων και εξαρτημένων μεταβλητών δεν είναι πάντοτε σαφής. Για παράδειγμα αν οι μεταβλητές που μας ενδιαφέρουν είναι το βάρος και το ύψος ενός ατόμου δεν είναι ξεκάθαρο ποια από τις δύο μεταβλητές θα πάρουμε ως ανεξάρτητη και ποια ως εξαρτημένη [26].

2.4.1 Ευθεία παλινδρόμησης

Για να αποδώσουμε μορφή εξάρτησης μεταξύ των μεταβλητών Y και X θα γράφουμε:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

όπου x είναι η συγκεκριμένη τιμή που πήρε η μη τυχαία μεταβλητή X , Y η τυχαία μεταβλητή που αντιστοιχεί στην τιμή x της X και ε μια τυχαία μεταβλητή που περιγράφει την «απόκλιση» της Y από το γραμμικό όρο $\beta_0 + \beta_1 x$.

Προκειμένου να διαμορφώσουμε ένα στοχαστικό μοντέλο δίνουμε στη μεταβλητή X τις τιμές x_1, x_2, \dots, x_n και καταγράφουμε τις αντίστοιχες τιμές y_1, y_2, \dots, y_n που λαμβάνει η μεταβλητή Y . Εφαρμόζοντας τον προηγούμενο τύπο για τα ζεύγη (x_i, y_i) , $i = 1, 2, \dots, n$ μπορούμε να γράψουμε

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

Προκειμένου να μπορέσουμε να διαμορφώσουμε ένα αποτελεσματικό μοντέλο πρόβλεψης της Y μέσω της τιμής x που έλαβε η μεταβλητή X , θα πρέπει οι διαφορές

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

να λαμβάνουν μικρές (θετικές ή αρνητικές) τιμές, ή ισοδύναμα να ισχύει η προσεγγιστική σχέση

$$y_i \cong \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, n$$

Αν τα ζεύγη (x_i, y_i) , $i = 1, 2, \dots, n$ τοποθετηθούν σε ένα ορθογώνιο σύστημα συντεταγμένων, τα σημεία που αντιστοιχούν σε αυτά θα πρέπει να βρίσκονται αρκετά κοντά σε μια ευθεία. Το διάγραμμα που απεικονίζει τη σχέση των ζευγών αυτών ονομάζεται διάγραμμα διασποράς (*scatter diagram*). Η ευθεία που προσεγγίζει καλύτερα τα σημεία αυτά θα λέγεται ευθεία παλινδρόμησης της Y πάνω στη X . Η εξίσωση της ευθείας που προσεγγίζει τα δεδομένα έχει τη γενική μορφή:

$$y = \beta_0 + \beta_1 x$$

Η παράμετρος β_0 θα μας δίνει τότε τη θέση όπου η ευθεία τέμνει τον άξονα $y'y$, ενώ η παράμετρος β_1 θα παριστάνει το συντελεστή διεύθυνσης της ευθείας. Για να βρεθεί η εξίσωση της ευθείας πρέπει να καθοριστούν οι παράμετροι β_0 και β_1 . Η διαδικασία καθορισμού των β_0 και β_1 λέγεται εκτίμηση των παραμέτρων ενώ οι τιμές που προκύπτουν για αυτές με την υλοποίηση της διαδικασίας εκτίμησης λέγονται εκτιμήτριες. Αν και υπάρχουν πολλές διαδικασίες με τις οποίες μπορεί να γίνει η εκτίμηση των παραμέτρων αυτών εμείς θα αναφερθούμε στην πιο διαδεδομένη μέθοδο η οποία είναι γνωστή στην βιβλιογραφία ως η μέθοδος των ελαχίστων τετραγώνων (*Least Squares Method*).

Σύμφωνα με τη μέθοδο αυτή, αν είναι γνωστές οι τιμές x_i , $i = 1, 2, \dots, n$ της μεταβλητής X και οι αντίστοιχες παρατηρηθείσες τιμές y_i , $i = 1, 2, \dots, n$ της μεταβλητής Y τότε οι εκτιμήτριες ελαχίστων τετραγώνων για τις παραμέτρους β_0 , β_1 της ευθείας

$y = \beta_0 + \beta_1 x$, με βάση n ζεύγη σημείων (x_i, y_i) , $i=1,2,\dots,n$ δίνονται από τους τύπους:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta}_1 \cdot \frac{1}{n} \sum_{i=1}^n x_i$$

Η αντίστοιχη ευθεία $y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$ καλείται ευθεία ελαχίστων τετραγώνων ή ευθεία παλινδρόμησης της Y (πάνω) στη X . Για την απόδειξη, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Κούτρα & Ευαγγελάρα (2010).

Επιπλέον, αν συμβολίσουμε με

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

τους αριθμητικούς μέσες τιμές των παρατηρήσεων x_1, x_2, \dots, x_n και y_1, y_2, \dots, y_n αντίστοιχα, ο τύπος που δόθηκε προηγουμένως για την εκτιμήτρια του β_0 , γράφεται ισοδύναμα ως εξής

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

ενώ αν αντικαταστήσουμε την έκφραση $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ στην εξίσωση της ευθείας παλινδρόμησης, η τελευταία γράφεται στη μορφή

$$y = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x \Leftrightarrow y - \bar{y} = \hat{\beta}_1 (x - \bar{x}),$$

η οποία δείχνει ότι η ευθεία ελαχίστων τετραγώνων $y = \hat{\beta}_0 + \hat{\beta}_1 x$ διέρχεται πάντοτε από το σημείο με συντεταγμένες (\bar{x}, \bar{y}) . Τέλος αναφέρουμε ότι, η εκτιμήτρια ελαχίστων τετραγώνων της παραμέτρου β_1 μπορεί να γραφεί εναλλακτικά στη μορφή:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

2.4.2 Ο συντελεστής προσδιορισμού

Αν υποθέσουμε ότι, με βάση τα n ζεύγη (x_i, y_i) $i=1,2,\dots,n$, έχουμε υπολογίσει τις εκτιμήτριες ελαχίστων τετραγώνων $\hat{\beta}_0$ και $\hat{\beta}_1$ και αν συμβολίσουμε με \hat{y}_i τις τιμές που

προκύπτουν για τη μεταβλητή απόκρισης Y χρησιμοποιώντας την ευθεία παλινδρόμησης $y = \hat{\beta}_0 + \hat{\beta}_1 x$, δηλαδή

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x}), \quad i = 1, 2, \dots, n.$$

τότε οι ποσότητες αυτές είναι οι εκτιμήσεις μας για την τιμή της μεταβλητής απόκρισης στις θέσεις $x_i, i = 1, 2, \dots, n$ και θα λέγονται εκτιμημένες τιμές (ή εκτιμώμενες τιμές) της Y στις θέσεις $x_i, i = 1, 2, \dots, n$. Οι διαφορές

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, 2, \dots, n$$

εκφράζουν τις αποκλίσεις που έχουν οι παρατηρηθείσες τιμές y_i της μεταβλητής απόκρισης για την τιμή x_i της ανεξάρτητης μεταβλητής X από τις εκτιμημένες τιμές. Τα ε_i ονομάζονται εκτιμημένα σφάλματα (ή εκτιμώμενα σφάλματα) για την τιμή x_i της ανεξάρτητης μεταβλητής X

$$g(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2$$

είναι η ελάχιστη τιμή του αθροίσματος $g(\beta_0, \beta_1)$ για όλες τις δυνατές επιλογές των παραμέτρων β_0 και β_1 . Η ποσότητα αυτή λέγεται άθροισμα τετραγώνων των (εκτιμημένων) σφαλμάτων (*Sum of Squares of Errors* ή *Error Sum of Squares*) και θα συμβολίζεται με SSE , δηλαδή:

$$SSE = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Είναι φανερό ότι όταν το SSE λαμβάνει μικρές (θετικές) τιμές, όλες οι διαφορές $\hat{\varepsilon}_i = y_i - \hat{y}_i$, θα είναι μικρές (θετικές ή αρνητικές) και επομένως η ευθεία παλινδρόμησης περνάει «κοντά» στα σημεία $(x_i, y_i) \quad i = 1, 2, \dots, n$ ενώ όταν το SSE λαμβάνει μεγάλες (θετικές) τιμές, όλες οι διαφορές $\hat{\varepsilon}_i = y_i - \hat{y}_i$, θα είναι μεγάλες (κατ' απόλυτη τιμή) και επομένως η ευθεία παλινδρόμησης δεν βρίσκεται «κοντά» στα σημεία $(x_i, y_i) \quad i = 1, 2, \dots, n$ [26].

Αν κάποιος εξετάζε μόνο τις τιμές των y_1, y_2, \dots, y_n χωρίς να λαμβάνει υπόψη ότι αυτές αντιστοιχούν σε διαφορετικές (πιθανώς) τιμές x_1, x_2, \dots, x_n της ανεξάρτητης μεταβλητής X , θα μπορούσε να χρησιμοποιήσει ως ένα μέτρο διασποράς (μεταβλητότητας) αυτών γύρω από τη μέση τους τιμή \bar{y} , το συνολικό άθροισμα τετραγώνων (*Total Sum of Squares*).

$$SSTO = \sum_{i=1}^n (y_i - \bar{y})^2$$

Το συνολικό άθροισμα τετραγώνων γράφεται στη μορφή

$$SSTO = SSR + SSE, \quad SSR = \sum_{i=1}^{\nu} (\hat{y}_i - \bar{y})^2$$

είναι το άθροισμα τετραγώνων της παλινδρόμησης (Regression Sum of Squares) και SSE το άθροισμα τετραγώνων των σφαλμάτων [25]. Για την απόδειξη, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Κούτρα & Ευαγγελάρα (2010).

Αξίζει να αναφερθεί πως θα μπορούσαμε να χρησιμοποιήσουμε το λόγο $SSR/SSTO$ ως ένα δείκτη ποιότητας του μοντέλου της γραμμικής παλινδρόμησης $y = \hat{\beta}_0 + \hat{\beta}_1 x$. Η ποσότητα

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

λέγεται συντελεστής προσδιορισμού (*coefficient of determination*) του γραμμικού μοντέλου, λαμβάνει τιμές μεταξύ του 0 και του 1 και για τις δύο ακραίες τιμές του έχουμε τα εξής:

- Αν $R^2 = 1$, θα ισχύει $SSR = SSTO$ όποτε $SSE = 0$, δηλαδή

$$\hat{\epsilon}_i = y_i - \hat{y}_i = 0 \Leftrightarrow \hat{y}_i = y_i.$$

Τότε, το γραμμικό μοντέλο περιγράφει τέλεια τα διαθέσιμα δεδομένα (έχουμε τέλεια προσαρμογή των δεδομένων στην ευθεία παλινδρόμησης).

- Αν $R^2 = 0$, θα ισχύει $SSR = 0$, δηλαδή

$$\sum_{i=1}^{\nu} (\hat{y}_i - \bar{y})^2 = 0 \Leftrightarrow \hat{y}_i = \bar{y} \text{ για κάθε } i = 1, 2, \dots, \nu$$

Τότε, θα ισχύει $\hat{\beta}_1 = 0$, $\hat{\beta}_0 = \bar{y}$ και η ευθεία παλινδρόμησης παίρνει τη μορφή

$$y = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{y}$$

Στην περίπτωση αυτή, η πρόβλεψή μας για την τιμή της εξαρτημένης μεταβλητής Y είναι συνεχώς η ίδια ($y = \bar{y}$), χωρίς να επηρεάζεται από την τιμή x που πήρε η ανεξάρτητη μεταβλητή X .

Συνοψίζοντας, μπορούμε να πούμε ότι ο συντελεστής R^2 εκφράζει το ποσοστό της συνολικής διασποράς (των τιμών της εξαρτημένης μεταβλητής Y) η οποία εξηγείται από την ανεξάρτητη μεταβλητή X (μέσω της ευθείας παλινδρόμησης). Το υπόλοιπο ποσοστό

$$1 - R^2 = \frac{SSE}{SSTO}$$

είναι το ποσοστό της συνολικής διασποράς που παραμένει ανεξήγητο από τη X και θα πρέπει να αποδοθεί σε άλλες ανεξάρτητες μεταβλητές που επηρεάζουν την τιμή της Y είτε σε φυσική μεταβλητότητα (στατιστική τυχαιότητα) των τιμών της Y .

Η ανάλυση παλινδρόμησης ως στατιστική μέθοδος αφορά από μόνη της ένα μεγάλο κεφάλαιο της Στατιστικής Επιστήμης το οποίο περιέχει πολλές επιμέρους μεθόδους και επιδέχεται περαιτέρω ανάλυση. Στις προηγούμενες παραγράφους,

πραγματοποιήθηκε αναφορά βασικών μεθόδων παλινδρόμησης με έμφαση στη γραμμική παλινδρόμηση. Για περισσότερες πληροφορίες, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Κούτρα & Ευαγγελάρα (2010).

2.5 Ανάλυση Διακύμανσης

2.5.1 Γενικά

Η ανάλυση της διακύμανσης (*Analysis Of Variance – ANOVA*) είναι μία στατιστική μέθοδος με την οποία η μεταβλητότητα που υπάρχει σ' ένα σύνολο δεδομένων διασπάται στις επιμέρους συνιστώσες της με στόχο την κατανόηση της σημαντικότητας των διαφορετικών πηγών προέλευσής της. Η ανάπτυξη της μεθοδολογίας οφείλεται στον θεμελιωτή της σύγχρονης στατιστικής επιστήμης, Άγγλο στατιστικό Sir Ronald Aylmer Fisher (1890-1962). Στην πραγματικότητα η ANOVA περιλαμβάνει μία ομάδα στατιστικών μεθόδων κατάλληλων για την ανάλυση δεδομένων που προκύπτουν από πειραματικούς σχεδιασμούς.

Τα δεδομένα ενός δείγματος ανάλογα με την προέλευσή τους διακρίνονται σε παρατηρήσεις (*observational sampling*) ή σε πειραματικά (*designed sampling*). Στην πρώτη κατηγορία ο στατιστικός ερευνητής απλά παρατηρεί τις τιμές που εμφανίζονται χωρίς να έχει τη δυνατότητα επέμβασης στις αντίστοιχες μεταβλητές. Αντίθετα στη δεύτερη κατηγορία ο στατιστικός ερευνητής προσπαθεί να ελέγξει τα επίπεδα μιας ή περισσότερων ανεξάρτητων (*independent*) μεταβλητών προκειμένου να προσδιορίσει την επίδραση που έχουν πάνω στην υπό μελέτη μεταβλητή που καλείται εξαρτημένη (*dependent*) ή απόκριση (*response*).

Στόχος κάθε στατιστικού πειράματος είναι ο προσδιορισμός της επίδρασης μιας ή περισσότερων ανεξάρτητων μεταβλητών πάνω στην απόκριση. Οι μεταβλητές αυτές αναφέρονται συνήθως ως παράγοντες (*factors*) και μπορεί να είναι είτε ποσοτικές είτε ποιοτικές. Για παράδειγμα θα ήταν ενδιαφέρον να διερευνήσουμε την επίδραση που έχει ο ποιοτικός παράγοντας “φύλο” στη βαθμολογία της στατιστικής ή ο ποσοτικός παράγοντας “πλήθος καταστημάτων πώλησης” στον όγκο των πωλήσεων. Σε άλλες πάλι περιπτώσεις, ενδεχομένως να ενδιαφερόμαστε για την επίδραση που έχουν πάνω στην απόκριση περισσότερες της μιας ανεξάρτητες μεταβλητές όπως ο ποσοτικός παράγοντας “πλήθος εργαζομένων” και ο ποιοτικός παράγοντας “πόλη διαμονής” πάνω στο οικογενειακό εισόδημα.

Οι τιμές του παράγοντα που προσδιορίζονται στο πείραμα λέγονται επίπεδα (*levels*). Σε ένα πείραμα με ένα παράγοντα οι μεταχειρίσεις (*treatments*) όπως ονομάζονται του πειράματος είναι τα επίπεδα του παράγοντα. Για παράδειγμα αν στο πείραμα βαθμολογία της στατιστικής μας ενδιαφέρει η επίδραση του παράγοντα “φύλο” τότε οι θεραπείες του πειράματος είναι αρσενικό – θηλυκό. Σε ένα πείραμα με δύο ή περισσότερους παράγοντες οι μεταχειρίσεις είναι οι συνδυασμοί παραγόντων-επιπέδων [26]. Για παράδειγμα αν μας ενδιαφέρει η επίδραση των παραγόντων “φύλο”, “ηλικία” στη βαθμολογία της στατιστικής, τότε οι μεταχειρίσεις είναι οι συνδυασμοί των επιπέδων φύλου και ηλικίας π.χ. (αρσενικό, 21), (θηλυκό, 19).

2.5.2 Ανάλυση διακύμανσης προς έναν παράγοντα (One-Way ANOVA)

Έστω μία μεταβλητή απόκρισης Y και ο παράγοντας X με k επίπεδα που πιθανόν να επηρεάζουν την τιμή της Y . Σκοπός της μελέτης είναι η αξιολόγηση των επιδράσεων που ασκούν οι μεταβολές στα επίπεδα του παράγοντα X στη μεταβλητή απόκριση. Πιο συγκεκριμένα, ελέγχοντας αν οι μέσες τιμές της μεταβλητής απόκρισης στα k επίπεδα του παράγοντα είναι στατιστικά ίσες ή όχι, θα εξαχθούν συμπεράσματα για το βαθμό επίδρασης του παράγοντα X στην τιμή της. Με άλλα λόγια, αν συμβολίσουμε με $\mu_1, \mu_2, \dots, \mu_k$ τις μέσες τιμές της μεταβλητής απόκρισης στα k επίπεδα του παράγοντα X , τότε η ανάλυση διακύμανσης κατά έναν παράγοντα είναι ο στατιστικός έλεγχος που αφορά τις ακόλουθες υποθέσεις:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ για τουλάχιστον ένα ζεύγος } (i,j), \text{ όπου } i \neq j \text{ [26]}$$

Ας υποθέσουμε ότι n_1, n_2, \dots, n_k είναι ο αριθμός των παρατηρήσεων σε κάθε επίπεδο του παράγοντα X και Y_{ij} , $j=1,2,\dots,n_i$ οι παρατηρηθείσες τιμές της μεταβλητής απόκρισης στο i επίπεδο του παράγοντα, όπου $i=1,2,\dots,k$. Στη συνέχεια θα συμβολίζουμε με

$$n = n_1 + n_2 + \dots + n_k$$

το συνολικό πλήθος παρατηρήσεων που έχουν γίνει για τη μεταβλητή απόκρισης Y .

Για την περιγραφή των δεδομένων από ένα πλήρως τυχαίο πείραμα χρησιμοποιούμε το ακόλουθο μοντέλο:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

όπου ε_{ij} είναι το τυχαίο σφάλμα που αντιστοιχεί στην πρόβλεψη για τη j -οστή παρατήρηση του οστού επιπέδου του παράγοντα. Με άλλα λόγια, σύμφωνα με το παραπάνω μοντέλο Ανάλυσης Διακύμανσης, η τυχαία παρατήρηση Y_{ij} είναι ίση με τη μέση τιμή που αντιστοιχεί στο i -οστό επίπεδο του παράγοντα συν κάποιο τυχαίο σφάλμα ε_{ij} . Οι βασικές υποθέσεις που πρέπει να ισχύουν είναι οι εξής:

- Οι τυχαίες μεταβλητές ε_{ij} ακολουθούν, κανονική κατανομή με την ίδια διασπορά σ^2
- Οι τυχαίες μεταβλητές ε_{ij} που αντιστοιχούν σε κάθε ένα επίπεδο του παράγοντα είναι ανεξάρτητες από τις αντίστοιχες μεταβλητές των άλλων επιπέδων.

Δεδομένου ότι η μέση τιμή των μεταβλητών Y_{ij} είναι ίση με μ_i για κάθε επίπεδο $i=1,2,\dots,k$, προκύπτει άμεσα ότι η μέση τιμή των μεταβλητών ε_{ij} είναι ίση με μηδέν για κάθε $i=1,2,\dots,k$ και $j=1,2,\dots,n_i$. Επομένως η υπόθεση που τελικά γίνεται για τις τυχαίες μεταβλητές ε_{ij} είναι ότι:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

και ότι όλα τα ε_{ij} για $j=1,2,\dots, n_i$, $i=1,2,\dots,k$ είναι ανεξάρτητα μεταξύ τους.

Επομένως, αφού η μέση τιμή μ_i των τιμών της μεταβλητής απόκρισης σε κάθε επίπεδο του παράγοντα είναι παράμετρος (και όχι τυχαία μεταβλητή), από τον τύπο προκύπτει ότι οι τυχαίες μεταβλητές Y_{ij} έχουν την ίδια διασπορά με τις μεταβλητές ε_{ij} . Επομένως:

$$Y_{ij} \sim N(\mu_i, \sigma^2)$$

και όλα τα Y_{ij} για $j=1,2,\dots, n_i$, $i=1,2,\dots,k$ είναι ανεξάρτητα μεταξύ τους.

Προκειμένου να μελετήσουμε διεξοδικότερα τη διάσπαση της συνολικής μεταβλητότητας της μεταβλητής απόκρισης, ώστε να προχωρήσουμε στη διαμόρφωση του κανόνα απόφασης για την απόρριψη ή μη της H_0 , θα χρειαστούμε τους ακόλουθους συμβολισμούς:

- $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$, $i=1,2,\dots,k$: εκφράζει το άθροισμα των παρατηρήσεων της μεταβλητής απόκρισης στο i -οστό επίπεδο του παράγοντα.
- $\bar{Y}_{i\bullet} = \frac{Y_{i\bullet}}{n_i} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$, $i=1,2,\dots,k$: εκφράζει το δειγματικό μέσο (μέσο όρο) των παρατηρήσεων της μεταβλητής απόκρισης στο i -οστό επίπεδο του παράγοντα. Όπως θα δούμε στη συνέχεια, αποτελεί την εκτιμήτρια ελαχίστων τετραγώνων για το μέσο μ_i του i -οστού επιπέδου του παράγοντα.
- $Y_{\bullet\bullet} = \sum_{i=1}^k Y_{i\bullet} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^k n_i \bar{Y}_{i\bullet}$: εκφράζει το συνολικό άθροισμα των παρατηρήσεων της μεταβλητής απόκρισης.
- $\bar{Y} = \frac{Y_{\bullet\bullet}}{n}$: εκφράζει το συνολικό μέσο όρο των παρατηρήσεων της μεταβλητής απόκρισης.

Η συνολική μεταβλητότητα $SSTO$ των τιμών της μεταβλητής απόκρισης Y μπορεί να μετρηθεί με τη βοήθεια του ολικού αθροίσματος των τετραγώνων των αποστάσεων των τιμών Y_{ij} από τον ολικό μέσο \bar{Y} , όπως φαίνεται ακολούθως

$$SSTO = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$$

Προσθαφαιρώντας μέσα στην παρένθεση της παραπάνω έκφρασης τον μέσο $\bar{Y}_{i\bullet}$ του i -οστού επιπέδου του παράγοντα, προκύπτει η ακόλουθη σχέση

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2$$

ή ισοδύναμα

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^k n_i (\bar{Y}_{i\cdot} - \bar{Y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$$

(για την επαλήθευση της ισότητας $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, θα πρέπει να ληφθεί υπόψη

ότι ισχύει η ισότητα $\sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i\cdot} - \bar{Y})(Y_{ij} - \bar{Y}_{i\cdot}) = 0$)

Το άθροισμα $SSTR = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i\cdot} - \bar{Y})^2$ είναι γνωστό ως άθροισμα των τετραγώνων των θεραπειών (*Sum of Squares of Treatments*) και εκφράζει τη μεταβλητότητα που εμφανίζεται μεταξύ των παρατηρήσεων της μεταβλητής απόκρισης που οφείλεται στις διαφορετικές επιδράσεις των θεραπειών.

Το άθροισμα $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$ είναι γνωστό ως άθροισμα των τετραγώνων των σφαλμάτων (*Sum of Squares of Errors*) και εκφράζει τη μη ελεγχόμενη μεταβλητότητα που εμφανίζεται μεταξύ των παρατηρήσεων που ανήκουν στο ίδιο επίπεδο του παράγοντα [21]. Τα αθροίσματα αυτά συνδέονται με την παρακάτω σχέση:

$$SSTO = SSTR + SSE$$

Για τον προσδιορισμό των βαθμών ελευθερίας (*degrees of freedom*) των παραπάνω ποσοτήτων, θα πρέπει να ληφθεί υπόψη ότι οι βαθμοί ελευθερίας ενός αθροίσματος τετραγώνων εκφράζουν το πλήθος των "ανεξάρτητων" στοιχείων πληροφορίας που απαιτούνται για να υπολογιστεί το άθροισμα. Συνεπώς, το άθροισμα των τετραγώνων των θεραπειών $SSTR$ έχει $k - 1$ βαθμούς ελευθερίας, μιας και ο υπολογισμός του μπορεί να γίνει με βάση οποιεσδήποτε $k - 1$ από τις k διαφορές $\bar{Y}_{i\cdot} - \bar{Y}$ (δεν χρειάζονται όλες, αφού η μία μπορεί να βρεθεί από τις υπόλοιπες καθώς το άθροισμα τους είναι ίσο με μηδέν). Επιπρόσθετα, για τον υπολογισμό του ολικού αθροίσματος $SSTO$ απαιτείται η γνώση των $n - 1$ διαφορών $Y_{ij} - \bar{Y}$ (από το σύνολο των n διαφορών), καθώς ισχύει η σχέση:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}) = 0$$

Από την σχέση $SSTO = SSTR + SSE$ μπορούν εύκολα να υπολογιστούν οι βαθμοί ελευθερίας των αθροισμάτων που εμφανίζονται σε αυτή ως εξής:

$$\begin{aligned} \beta.ε.(SSE) &= \beta.ε.(SSTO) - \beta.ε.(SSTR) \\ &= (n - 1) - (k - 1) \\ &= n - k. \end{aligned}$$

Συνεπώς, τα μέσα αθροίσματα τετραγώνων των θεραπειών και των σφαλμάτων δίνονται ως ακολούθως:

$$MSTR = \frac{SSTR}{k - 1} \quad MSE = \frac{SSE}{n - k}$$

Για να ελεγχθεί η μηδενική υπόθεση, είναι απαραίτητο να υπάρχει μια στατιστική συνάρτηση ελέγχου η κατανομή της οποίας πρέπει να είναι γνωστή όταν αληθεύει η μηδενική υπόθεση. Για το μοντέλο σταθερών επιδράσεων, η στατιστική συνάρτηση για τον έλεγχο της παραπάνω μηδενικής υπόθεσης σχηματίζεται από το λόγο δύο μέσων τετραγώνων, δηλαδή δύο αθροισμάτων διαιρούμενων με τους αντίστοιχους βαθμούς ελευθερίας τους. Πιο συγκεκριμένα, η στατιστική συνάρτηση για είναι ο λόγος του μέσου αθροίσματος τετραγώνων των θεραπειών προς το μέσο άθροισμα τετραγώνων των σφαλμάτων, όπως φαίνεται στην ακόλουθη σχέση

$$F^* = \frac{MSTR}{MSE}$$

Δεδομένου ότι, όταν $Y_{ij} \sim N(\mu_i, \sigma^2)$, οι ποσότητες $SSTR/\sigma^2$ και SSE/σ^2 είναι ανεξάρτητες και ακολουθούν κατανομή χ^2 με $k-1$ και $n-1$ βαθμούς ελευθερίας αντίστοιχα, συμπεραίνουμε άμεσα ότι ο λόγος F^* ακολουθεί την κατανομή F με $k-1$ και $n-1$ βαθμούς ελευθερίας για αριθμητή και παρονομαστή αντίστοιχα. Πράγματι, έχουμε τα εξής:

$$F^* = \frac{MSTR}{MSE} = \frac{SSTR/(k-1)}{SSE/(n-k)} = \frac{SSTR/\sigma^2}{SSE/\sigma^2} \frac{n-k}{k-1} \sim F_{k-1, n-k}$$

Συνεπώς, η κρίσιμη περιοχή για την απόρριψη της μηδενικής υπόθεσης σε επίπεδο σημαντικότητας $\alpha\%$ ορίζεται από τη σχέση:

$$F^* > F_{k-1, n-k}(\alpha)$$

Η προαναφερθείσα διαδικασία ελέγχου της μηδενικής υπόθεσης συνοψίζεται στον ακόλουθο πίνακα, που είναι γνωστός ως πίνακας Ανάλυσης Διακύμανσης κατά έναν παράγοντα (*One-way Analysis of Variance Table*).

Πίνακας 1: Πίνακας ανάλυσης διακύμανσης κατά έναν παράγοντα

Πηγή Μεταβλητότητας	Άθροισμα Τετραγώνων (Sum of Squares)	Βαθμοί ελευθερίας (Degrees of freedom)	Μέσα τετράγωνα (Mean Squares)	Στατιστικό κριτήριο F^*
Θεραπείες (μεταξύ των Αγωγών)	SSTR	$k-1$	$MSTR = \frac{SSTR}{k-1}$	$F^* = \frac{MSTR}{MSE}$
Σφάλματα (εντός των αγωγών)	SSE	$n-k$	$MSE = \frac{SSE}{n-k}$	
Σύνολο	SSTO	$n-1$		

Η ανάλυση διακύμανσης μπορεί να εφαρμοστεί και για δύο ή περισσότερους παράγοντες, γεγονός που αυξάνει την πολυπλοκότητα και τον βαθμό δυσκολίας της μεθόδου. Στο κεφάλαιο αυτό αναλύσαμε μόνο την μέθοδο ANOVA για έναν

παράγοντα. Αν ο αναγνώστης επιθυμεί περισσότερες πληροφορίες και περαιτέρω ανάλυση μπορεί να συμβουλευτεί το βιβλίο του Γναρδέλλη Χ. (2013).

ΚΕΦΑΛΑΙΟ 3

ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 Εισαγωγή

Στο παρόν κεφάλαιο θα παρουσιαστούν αναλυτικά τόσο τα αποτελέσματα όσο και τα συμπεράσματα που προκύπτουν και αφορούν την έρευνα που πραγματοποιήθηκε με σκοπό την συγκριτική αξιολόγηση των μελών Δ.Ε.Π όλων των Τμημάτων Πληροφορικής Δημοσίων Πανεπιστημίων στην Ελλάδα. Για τη σύγκριση αυτή χρησιμοποιήθηκαν ποιοτικοί και ποσοτικοί δείκτες όπως αυτοί που χρησιμοποιούνται στις έρευνες της επιστήμης της βιβλιομετρίας και αναφέρθηκαν αναλυτικά στο Κεφάλαιο 1. Πιο συγκεκριμένα, χρησιμοποιώντας τους δείκτες αυτούς πραγματοποιήθηκε σύγκριση και σε ατομικό επίπεδο αλλά και σε επίπεδο Τμήματος ή Πανεπιστημίου. Ένα μεγαλύτερο μέρος της μελέτης περιλαμβάνει την παρουσίαση και αξιολόγηση του ερευνητικού και επιστημονικού έργου των τμημάτων Πληροφορικής των ελληνικών Α.Ε.Ι. Πληροφορικής που αφορούν τα δεδομένα και το πως αυτά συλλέχθηκαν δημιουργώντας ένα επαρκές δείγμα θα παρουσιαστούν σε επόμενη παράγραφο λεπτομερώς. Το σύνολο των δεδομένων που δημιουργήθηκε, καταχωρήθηκε στο ευρέως γνωστό στατιστικό πακέτο *IBM SPSS Statistics* με σκοπό την καλύτερη και πιο λεπτομερή ανάλυση του. Στη συνέχεια, εφαρμόζοντας στατιστικές μεθόδους και τεχνικές όπως αυτές που αναφέρθηκαν στο Κεφάλαιο 2 και με τη χρήση του εργαλείου *IBM SPSS Statistics* εξήχθησαν αποτελέσματα τα οποία αναλύθηκαν περαιτέρω με σκοπό την εξαγωγή των αντίστοιχων συμπερασμάτων. Στις επόμενες ενότητες θα γίνει επεξήγηση τόσο αυτών όσο και της σημασίας τους στην εν λόγω έρευνα.

3.2 Δείγμα και καταγραφή δεδομένων

Τα μέλη Δ.Ε.Π των Τμημάτων Πληροφορικής των ελληνικών Α.Ε.Ι. αποτελούν το δείγμα της παρούσας μελέτης. Στην Ελλάδα λειτουργούν σήμερα δεκατέσσερα πανεπιστημιακά Τμήματα του κλάδου Πληροφορικής τα οποία σύμφωνα με Προεδρικό Διάταγμα παρέχουν στους αποφοίτους τους κοινά επαγγελματικά δικαιώματα. Τα δώδεκα από αυτά θεωρούνται με βάση σχετικές υπουργικές αποφάσεις «αντίστοιχα» και παρέχεται το δικαίωμα μετεγγραφής των φοιτητών από το ένα στο άλλο. Τα τμήματα που συμμετέχουν στην παρούσα έρευνα είναι τα εξής:

1. Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
2. Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Πελοποννήσου
3. Τμήμα Πληροφορικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
4. Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών
5. Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς
6. Τμήμα Πληροφορικής, Ιόνιο Πανεπιστήμιο
7. Τμήμα Πληροφορικής με εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας

8. Τμήμα Εφαρμοσμένης Πληροφορικής, Πανεπιστήμιο Μακεδονίας Οικονομικών και Κοινωνικών Επιστημών
9. Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
10. Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
11. Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο Πανεπιστήμιο.

Στην παρούσα μελέτη έχουν συμπεριληφθεί μόνο τα παραπάνω Τμήματα εξαιρουμένου του Τμήματος Πληροφορικής του Πανεπιστημίου Θεσσαλίας εξαιτίας του μικρού αριθμού τακτικών μελών Δ.Ε.Π.

Για την εύρεση και καταγραφή των μελών Δ.Ε.Π πραγματοποιήθηκε αναζήτηση στην ιστοσελίδα κάθε Τμήματος χωριστά. Στη συνέχεια για κάθε μέλος Δ.Ε.Π, καταγράφηκαν οι βιβλιομετρικές τους μετρήσεις με την βοήθεια της βάσης δεδομένων *Scopus*. Η βάση αυτή όπως έχει προαναφερθεί έχει επιλεγεί ως καταλληλότερη λόγω προσβασιμότητας, μετά από συγκριτική αξιολόγησή της με τις βάσεις *Web of Science* και *Google Scholar*. Οι μετρήσεις αφορούν δημοσιεύσεις και αναφορές σε αυτές των μελών Δ.Ε.Π κάθε Τμήματος, όπως αυτά προκύπτουν από τις αντίστοιχες ιστοσελίδες των Τμημάτων. Επιπλέον, υπολογίστηκαν και καταγράφηκαν οι δείκτες επιστημονικής ποιότητας *h-index*, *p-index* καθώς και ο λόγος του αριθμού των αναφορών ανά δημοσίευση (*citations per publication*).

Πρόσθετα, για κάθε Τμήμα Πληροφορικής καταγράφηκαν τα εξής στοιχεία: τα χρόνια λειτουργίας του Τμήματος καθώς και τα χρόνια λειτουργίας του Πανεπιστημίου στο οποίο ανήκει το εκάστοτε Τμήμα, το μέγεθος του Πανεπιστημίου υπολογισμένο σε Τμήματα που διαθέτει το κάθε Πανεπιστήμιο, η τοποθεσία στην οποία βρίσκεται το Τμήμα, το πλήθος των μελών Δ.Ε.Π και του διοικητικού προσωπικού, ο αριθμός των εισακτέων του Τμήματος κατά προσέγγιση και ο λόγος του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π. Τέλος υπολογίστηκε ο δείκτης *h₁-index* σε επίπεδο Τμήματος.

3.3 Σύγκριση σε μη ατομικό επίπεδο βάσει του δείκτη h_1

3.3.1 Συντελεστές συσχέτισης Spearman

Πίνακας 2: Συντελεστής συσχέτισης Spearman δείκτη h_1

			Correlations							
			h1_index	size	age_Uni	age_Dep	Teaching_staff	Admin_staff	students	ratio
Spearman's rho	h1_index	Correlation Coefficient	1,000	,636	,738	,583	,748	,721	,147	-,839
		Sig. (2-tailed)	.	,035	,010	,060	,008	,012	,666	,001
			N	11	11	11	11	11	11	11
	size	Correlation Coefficient	,636	1,000	,407	,320	,265	,475	-,204	-,384
		Sig. (2-tailed)	,035	.	,214	,337	,431	,139	,546	,244
			N	11	11	11	11	11	11	11
	age_Uni	Correlation Coefficient	,738	,407	1,000	,690	,756	,843	,139	-,706
		Sig. (2-tailed)	,010	,214	.	,019	,007	,001	,685	,015
			N	11	11	11	11	11	11	11
	age_Dep	Correlation Coefficient	,583	,320	,690	1,000	,624	,873	,368	-,486
		Sig. (2-tailed)	,060	,337	,019	.	,040	,000	,266	,129
			N	11	11	11	11	11	11	11
	Teaching_staff	Correlation Coefficient	,748	,265	,756	,624	1,000	,722	,573	-,864
		Sig. (2-tailed)	,008	,431	,007	,040	.	,012	,065	,001
			N	11	11	11	11	11	11	11
	Admin_staff	Correlation Coefficient	,721	,475	,843	,873	,722	1,000	,215	-,708
		Sig. (2-tailed)	,012	,139	,001	,000	,012	.	,526	,015
			N	11	11	11	11	11	11	11
	students	Correlation Coefficient	,147	-,204	,139	,368	,573	,215	1,000	-,151
		Sig. (2-tailed)	,666	,546	,685	,266	,065	,526	.	,658
			N	11	11	11	11	11	11	11
	ratio	Correlation Coefficient	-,839	-,384	-,706	-,486	-,864	-,708	-,151	1,000
		Sig. (2-tailed)	,001	,244	,015	,129	,001	,015	,658	.
			N	11	11	11	11	11	11	11

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Ο δείκτης h_1 σχετίζεται ισχυρά και θετικά σε επίπεδο σημαντικότητας 5% με το μέγεθος του Πανεπιστημίου (*size*, $r = 0.636$, $p = 0.035$), την ηλικία (χρονική διάρκεια λειτουργίας) του Πανεπιστημίου (*age_Uni*, $r = 0.737$, $p = 0.010$) υπολογισμένη σε έτη λειτουργίας, το πλήθος των μελών Δ.Ε.Π. (*Teaching_staff*, $r = 0.748$, $p = 0.008$), το πλήθος των διοικητικών υπαλλήλων (*admin_staff*, $r = 0.721$, $p = 0.012$), ενώ σχετίζεται ισχυρά και αρνητικά με τον λόγο των φοιτητών ανά μέλος Δ.Ε.Π (*ratio*, $r = -0.839$, $p = 0.01$). Αυτό πρακτικά σημαίνει ότι ο δείκτης επιστημονικής ποιότητας h_1 επηρεάζεται άμεσα και θετικά από τον αριθμό των Τμημάτων που έχει κάθε Πανεπιστήμιο καθώς και από τον αριθμό των μελών Δ.Ε.Π και του διοικητικού προσωπικού που διαθέτει κάθε τμήμα του Πανεπιστημίου. Για την ακρίβεια όσο περισσότερα τμήματα και μέλη Δ.Ε.Π. και διοικητικού προσωπικού έχει ένα τμήμα τόσο μεγαλύτερες τιμές λαμβάνει ο δείκτης h_1 . Αυτό φαίνεται να είναι λογικό καθώς όσο περισσότερα τμήματα υπάρχουν τόσο περισσότερες δημοσιεύσεις θα πραγματοποιηθούν από περισσότερα μέλη Δ.Ε.Π.

Επιπλέον, ο δείκτης αυτός επηρεάζεται από τα χρόνια λειτουργίας του Πανεπιστημίου. Σύμφωνα με τα παραπάνω, όσα περισσότερα χρόνια λειτουργεί ένα Τμήμα ή ένα Πανεπιστήμιο τόσο μεγαλύτερη πιθανότητα υπάρχει να λαμβάνει ο δείκτης επιστημονικής ποιότητας h_1 μεγαλύτερες τιμές. Όσον αφορά τον λόγο των φοιτητών ανά μέλος Δ.Ε.Π, σχετίζεται αντιστρόφως ανάλογα με τον δείκτη. Στο σημείο αυτό, μπορεί να δοθεί η λογική εξήγηση ότι όσους περισσότερους φοιτητές επιβλέπει ένας καθηγητής τόσο περισσότερο χρόνο «ξοδεύει» στην επίβλεψη τους, με αποτέλεσμα να μειώνεται ο χρόνος που αφιερώνει στη δημοσίευση επιστημονικών εργασιών, συνεπώς και η τιμή του h_1 .

3.3.2 Εξάρτηση δεικτών από κατηγορικά χαρακτηριστικά των Τμημάτων

- Μεταβλητή *Location* (1 = Αθήνα, 2 = Θεσ/νίκη, 3 = άλλη πόλη)

Πίνακας 3: Περιγραφικά στατιστικά μέτρα για τον δείκτη h_1

Descriptives

h1_index

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	4	10,25	3,304	1,652	4,99	15,51	6	14
2	2	11,50	4,950	3,500	-32,97	55,97	8	15
3	5	8,40	2,074	,927	5,83	10,97	6	11
Total	11	9,64	3,009	,907	7,61	11,66	6	15

Πίνακας 4: Ανάλυση διακύμανσης για τον δείκτη h_1

ANOVA

h1_index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16,095	2	8,048	,865	,457
Within Groups	74,450	8	9,306		
Total	90,545	10			

Στον Πίνακα 3, παρατηρούμε ότι από τα 11 τμήματα που μελετήθηκαν τα 4 βρίσκονται στην Αθήνα, τα 2 στην Θεσσαλονίκη και τα υπόλοιπα 5 σε άλλες πόλεις της Ελλάδας. Για κάθε μία από αυτές τις 3 κατηγορίες υπολογίζεται η μέση τιμή, η τυπική απόκλιση, 95% διάστημα εμπιστοσύνης καθώς και η μέγιστη και η ελάχιστη τιμή της. Στον Πίνακα 4, βλέπουμε τα αποτελέσματα ενός *ANOVA test* για τις 3 ομάδες που αναφέρθηκαν. Παρατηρώντας την τιμή του *sig* βλέπουμε ότι ο δείκτης h_1 δεν εξαρτάται από την τοποθεσία των τμημάτων ($p=0,457$). Θα περίμενε ίσως κανείς ότι σε μεγάλες πόλεις όπως είναι η Αθήνα και η Θεσσαλονίκη, θα υπήρχαν περισσότερες δημοσιεύσεις ανά μέλος Δ.Ε.Π. και επομένως η τιμή του δείκτη h_1 θα ήταν μεγαλύτερη, ωστόσο αυτό δεν επιβεβαιώνεται στατιστικά.

3.3.3 Πρότυπο Γραμμικής Παλινδρόμησης για την πρόβλεψη του δείκτη h_1

Ορίζουμε ως εξαρτημένη μεταβλητή την τιμή του δείκτη h_1 , ενώ τα υπόλοιπα στοιχεία χρησιμοποιούνται ως επεξηγηματικοί παράγοντες. Εφαρμόζοντας *stepwise linear regression* (βηματική γραμμική παλινδρόμηση) έχουμε τα εξής αποτελέσματα:

Πίνακας 5: Βηματική γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή τον δείκτη h_1

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	ratio	.	Stepwise (Criteria: Probability-of- F-to-enter <= , 050, Probability-of- F-to-remove >= ,100).
2	size	.	Stepwise (Criteria: Probability-of- F-to-enter <= , 050, Probability-of- F-to-remove >= ,100).

a. Dependent Variable: h1_index

Οι μεταβλητές *ratio* και *size* κρίνονται στατιστικά σημαντικές και εισέρχονται στο γραμμικό μοντέλο, όπως φαίνεται στον Πίνακα 5. Πρόσθετα, το γραμμικό μοντέλο που προκύπτει έχει υψηλό συντελεστή προσδιορισμού. Πιο συγκεκριμένα,

Πίνακας 6: Γραμμικό μοντέλο για τις μεταβλητές *ratio* και *size*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,827 ^a	,684	,649	1,783
2	,917 ^b	,842	,802	1,338

a. Predictors: (Constant), ratio

b. Predictors: (Constant), ratio, size

ο συντελεστής προσδιορισμού είναι ίσος με 0,842, που πρακτικά σημαίνει ότι το 84,2% της συνολικής μεταβλητότητας που παρουσιάζει ο δείκτης h_1 ερμηνεύεται (οφείλεται) στους παράγοντες *ratio* και *size*, ενώ το υπόλοιπο 15,7% δεν ερμηνεύεται από το μοντέλο που σχηματίστηκε καθώς οφείλεται σε παράγοντες που δεν έχουν ληφθεί υπόψη. Όπως φαίνεται από τον παρακάτω πίνακα

Πίνακας 7

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	16,260	1,594		10,202	,000	12,655	19,866
	DEP_to_Students	-,723	,164	-,827	-4,414	,002	-1,093	-,352
2	(Constant)	12,299	1,844		6,671	,000	8,048	16,551
	DEP_to_Students	-,493	,147	-,564	-3,341	,010	-,833	-,153
	size	,122	,043	,476	2,824	,022	,022	,222

a. Dependent Variable: h1_index

οι παράγοντες *ratio* και *size* είναι στατιστικά σημαντικοί καθώς τα αντίστοιχα *p-values* είναι 0.01 και 0.022 αντίστοιχα. Πρόσθετα, η γραμμική εξίσωση πρόβλεψης της τιμής του δείκτη h_1 για ένα τμήμα του οποίου γνωρίζουμε τις τιμές των χαρακτηριστικών *ratio* και *size*, δίνεται ως ακολούθως

$$h_1 = 12,299 - 0,493 \cdot ratio + 0,122 \cdot size .$$

Μάλιστα, το 95% διάστημα εμπιστοσύνης για τον συντελεστή του παράγοντα *ratio* δίνεται ως εξής (-0.833,-0.153). Αυτό πρακτικά σημαίνει ότι εκτιμάμε με πιθανότητα 95% πως σε κάθε αύξηση του *ratio* ενός Τμήματος κατά μία μονάδα, ο δείκτης h_1 θα μειώνεται τουλάχιστον κατά 0.153 και το πολύ κατά 0.833 μονάδες. Μάλιστα, όπως φαίνεται από τα παρακάτω αποτελέσματα, η κανονικότητα, η ανεξαρτησία και η ομοσκεδαστικότητα των σφαλμάτων δεν παραβιάζεται καθώς τα *p-values* των αντίστοιχων ελέγχων (*Kolmogorov - Smirnov test*, *Runs test* και *Homogeneity test*) είναι μεγαλύτερα από 5% και πιο συγκεκριμένα ίσα με 0.2, 0.502 και 0.555 αντίστοιχα.

Πίνακας 8: Kolmogorov-Smirnov τεστ

One-Sample Kolmogorov-Smirnov Test

		Studentized Residual
N		11
Normal Parameters ^{a,b}	Mean	-,0477199
	Std. Deviation	1,05348887
Most Extreme Differences	Absolute	,168
	Positive	,106
	Negative	-,168
Test Statistic		,168
Asymp. Sig. (2-tailed)		,200 ^{c,d}

- a. Test distribution is Normal.
- b. Calculated from data.
- c. Lilliefors Significance Correction.
- d. This is a lower bound of the true significance.

Πίνακας 9: Runs τεστ

Runs Test

	Studentized Residual
Test Value ^a	-,07870
Cases < Test Value	5
Cases >= Test Value	6
Total Cases	11
Number of Runs	8
Z	,671
Asymp. Sig. (2-tailed)	,502

a. Median

Πίνακας 10: Τεστ ομογένειας για τη διασπορά

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Studentized Residual	Based on Mean	,406	1	9	,540
	Based on Median	,375	1	9	,555
	Based on Median and with adjusted df	,375	1	7,039	,559
	Based on trimmed mean	,364	1	9	,561

3.3.4 Περιγραφικά στοιχεία για τους δείκτες

- Περιγραφικά στοιχεία για τον δείκτη h_1

Πίνακας 11: Περιγραφικά στατιστικά μέτρα για τον δείκτη h_1

		Statistics	
		h1_index	h1_plus_inde x
N	Valid	11	11
	Missing	0	0
Mean		9,64	10,64
Median		10,00	11,00
Mode		6 ^a	7 ^a
Std. Deviation		3,009	3,009
Skewness		,535	,535
Std. Error of Skewness		,661	,661
Kurtosis		-,568	-,568
Std. Error of Kurtosis		1,279	1,279
Range		9	9
Minimum		6	7
Maximum		15	16
Percentiles	25	7,00	8,00
	50	10,00	11,00
	75	11,00	12,00

a. Multiple modes exist. The smallest value is shown

Ο δείκτης h_1 plus δίνει ποιοτικά τα ίδια αποτελέσματα με τον δείκτη h_1 και συνεπώς δεν θα σχολιασθεί περαιτέρω.

- Η μέση τιμή του δείκτη h_1 είναι ίση με 9.64, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή του δείκτη h_1 για ένα Τμήμα Πληροφορικής είναι ίση με 9.64.
- Η διάμεσος του δείκτη h_1 είναι ίση με 10, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν δείκτη h_1 το πολύ ίσο με 10, ενώ το υπόλοιπο 50% των τμημάτων έχουν h_1 μεγαλύτερο ή ίσο με 10.
- Η τυπική απόκλιση του δείκτη h_1 είναι ίση με 3.009 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο τμήμα Πληροφορικής αναμένεται να έχει δείκτη h_1 που απέχει 3.009 μονάδες από τον αναμενόμενο δείκτη που είναι ίσος με 9.64.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 0.535 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή

την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη h_1 είναι συμμετρικό.

- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με -0.568 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη h_1 είναι μεσόκυρτο.
- Το εύρος των τιμών του δείκτη h_1 είναι ίσο με 9 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο τμημάτων Πληροφορικής ως προς την τιμή του δείκτη h_1 είναι ίση με 9 μονάδες.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του δείκτη h_1 είναι ίσο με 7 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 τμήματα έχει δείκτη h_1 μέχρι το πολύ 7 μονάδες, ενώ τα υπόλοιπα 3 στα 4 τμήματα έχουν δείκτη h_1 μεγαλύτερο ή ίσο από 7 μονάδες.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του δείκτη h_1 είναι ίσο με 11 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 τμήματα έχει δείκτη h_1 μέχρι το πολύ 11 μονάδες, ενώ τα υπόλοιπα 1 στα 4 τμήματα έχουν δείκτη h_1 μεγαλύτερο ή ίσο από 11 μονάδες.

3.3.5 Περιγραφικά στοιχεία για τις λοιπές μεταβλητές

Πίνακας 12: Περιγραφικά στατιστικά μέτρα για τις λοιπές μεταβλητές

		Statistics					
		age_Uni	age_Dep	Teaching_staff	Admin_staff	students	DEP_to_Students
N	Valid	11	11	11	11	11	11
	Missing	0	0	0	0	0	0
Mean		64,55	21,64	24,55	4,45	193,18	9,1631
Median		52,00	24,00	26,00	5,00	225,00	8,6538
Mode		32	12 ^a	11 ^a	5	225	4,31 ^a
Std. Deviation		46,708	8,346	11,579	1,753	40,452	3,44282
Skewness		1,554	-,353	,667	-,316	-,847	,005
Std. Error of Skewness		,661	,661	,661	,661	,661	,661
Kurtosis		2,909	-1,490	-,368	,411	-,764	-1,382
Std. Error of Kurtosis		1,279	1,279	1,279	1,279	1,279	1,279
Range		163	23	36	6	100	9,75
Minimum		16	9	11	1	125	4,31
Maximum		179	32	47	7	225	14,06
Percentiles	25	32,00	12,00	14,00	3,00	175,00	5,7692
	50	52,00	24,00	26,00	5,00	225,00	8,6538
	75	91,00	27,00	31,00	5,00	225,00	12,5000

a. Multiple modes exist. The smallest value is shown

1. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει τα χρόνια λειτουργίας του Πανεπιστημίου στο οποίο ανήκει το Τμήμα (*age_Uni*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει τα χρόνια λειτουργίας του Πανεπιστημίου είναι ίση με 64.55, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για το έτος λειτουργίας του Πανεπιστημίου για ένα Τμήμα Πληροφορικής είναι ίση με 64.55.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει τα έτη λειτουργίας ενός Πανεπιστημίου είναι ίση με 52 έτη, που πρακτικά σημαίνει ότι το 50% των Πανεπιστημίων στα οποία ανήκουν τα Τμήματα Πληροφορικής έχουν χρόνο λειτουργίας το πολύ ίσο με 52 χρόνια, ενώ το υπόλοιπο 50% των Πανεπιστημίων έχουν χρόνο λειτουργίας μεγαλύτερο ή ίσο με 52 χρόνια.
- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τα έτη λειτουργίας ενός Πανεπιστημίου είναι ίση με 46.708 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 1.554 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των ετών λειτουργίας ενός Πανεπιστημίου είναι μη συμμετρικό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 2.909 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των ετών λειτουργίας ενός Πανεπιστημίου είναι μεσόκυρτο.
- Το εύρος των τιμών της «ηλικίας» του Πανεπιστημίου είναι ίσο με 163 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς την «ηλικία» του Πανεπιστημίου είναι ίση με 163 μονάδες.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων της «ηλικίας» ενός Πανεπιστημίου είναι ίσο με 32 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Πανεπιστήμια έχει «ηλικία» μέχρι το πολύ 32 μονάδες, ενώ τα υπόλοιπα 3 στα 4 Πανεπιστήμια έχουν «ηλικία» μεγαλύτερη ή ίση από 7 μονάδες.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων της «ηλικίας» ενός Πανεπιστημίου είναι ίσο με 91 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Πανεπιστήμια έχουν «ηλικία» μέχρι το πολύ 91 μονάδες, ενώ τα υπόλοιπα 1 στα 4 τμήματα έχουν «ηλικία» μεγαλύτερη ή ίση από 91 μονάδες.

2. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει τα χρόνια λειτουργίας του Τμήματος Πληροφορικής (*age_Dep*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει τα χρόνια λειτουργίας του Τμήματος Πληροφορικής είναι ίση με 21.64, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για το έτος λειτουργίας του Τμήματος είναι ίση με 21.64.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει τα έτη λειτουργίας ενός Τμήματος Πληροφορικής είναι ίση με 24 έτη, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν χρόνο λειτουργίας το πολύ ίσο με

24 χρόνια, ενώ το υπόλοιπο 50% των Τμημάτων έχουν χρόνο λειτουργίας μεγαλύτερο ή ίσο με 24 χρόνια.

- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τα έτη λειτουργίας ενός Τμήματος Πληροφορικής είναι ίση με 8.346 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με -0.353 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των ετών λειτουργίας ενός Τμήματος Πληροφορικής είναι συμμετρικό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με -1.490 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των ετών λειτουργίας ενός Τμήματος Πληροφορικής είναι μεσόκυρτο.
- Το εύρος των τιμών της «ηλικίας» του Τμήματος είναι ίσο με 23 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς την «ηλικία» του Τμήματος είναι ίση με 23 μονάδες.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων της «ηλικίας» ενός Τμήματος Πληροφορικής είναι ίσο με 12 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Τμήματα Πληροφορικής έχει «ηλικία» μέχρι το πολύ 12 μονάδες, ενώ τα υπόλοιπα 3 στα 4 Τμήματα έχουν «ηλικία» μεγαλύτερη ή ίση από 12 μονάδες.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων της «ηλικίας» ενός Τμήματος Πληροφορικής είναι ίσο με 27 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Τμήματα έχουν «ηλικία» μέχρι το πολύ 27 μονάδες, ενώ τα υπόλοιπα 1 στα 4 Τμήματα έχουν «ηλικία» μεγαλύτερη ή ίση από 27 μονάδες.

3. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών Δ.Ε.Π του Τμήματος Πληροφορικής (*Teaching_staff*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών Δ.Ε.Π του Τμήματος Πληροφορικής είναι ίση με 24.55, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για τον αριθμό των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 24.55.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 26 μονάδες, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν αριθμό μελών Δ.Ε.Π το πολύ ίσο με 26, ενώ το υπόλοιπο 50% των Τμημάτων έχουν αριθμό μελών Δ.Ε.Π μεγαλύτερο ή ίσο με 26.
- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τον αριθμό μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 11.579 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 0.667 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι συμμετρικό.

- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με -0.368 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι μεσόκυρτο.
- Το εύρος των τιμών του πλήθους των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 36 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς τον αριθμό μελών Δ.Ε.Π του Τμήματος είναι ίση με 36 μονάδες.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 14 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Τμήματα Πληροφορικής έχουν αριθμό μελών Δ.Ε.Π το πολύ μέχρι 14, ενώ τα υπόλοιπα 3 στα 4 Τμήματα έχουν αριθμό μελών Δ.Ε.Π μεγαλύτερο ή ίσο από 14.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων των μελών Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 31 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Τμήματα έχουν αριθμό μελών Δ.Ε.Π το πολύ μέχρι 31, ενώ τα υπόλοιπα 1 στα 4 Τμήματα έχουν αριθμό μελών Δ.Ε.Π μεγαλύτερο ή ίσο από 31.

4. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών διοικητικού προσωπικού του Τμήματος Πληροφορικής (*Admin_staff*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών διοικητικού προσωπικού του Τμήματος Πληροφορικής είναι ίση με 4.45, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για τον αριθμό των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίση με 4.45.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει το πλήθος των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίση με 5 μονάδες, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν αριθμό μελών διοικητικού προσωπικού το πολύ ίσο με 5, ενώ το υπόλοιπο 50% των Τμημάτων έχουν αριθμό μελών διοικητικού προσωπικού μεγαλύτερο ή ίσο με 5.
- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τον αριθμό μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίση με 1.753 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με -0.316 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι συμμετρικό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 0.411 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την

τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι μεσόκυρτο.

- Το εύρος των τιμών του πλήθους των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίσο με 6 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς τον αριθμό μελών διοικητικού προσωπικού του Τμήματος είναι ίση με 6 μονάδες.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίσο με 3 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Τμήματα Πληροφορικής έχουν αριθμό μελών διοικητικού προσωπικού το πολύ μέχρι 3, ενώ τα υπόλοιπα 3 στα 4 Τμήματα έχουν αριθμό μελών διοικητικού προσωπικού μεγαλύτερο ή ίσο από 3.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων των μελών διοικητικού προσωπικού ενός Τμήματος Πληροφορικής είναι ίσο με 5 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Τμήματα έχουν αριθμό μελών διοικητικού προσωπικού το πολύ μέχρι 5, ενώ τα υπόλοιπα 1 στα 4 Τμήματα έχουν αριθμό μελών διοικητικού προσωπικού μεγαλύτερο ή ίσο από 5.

5. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει τον αριθμό των εισακτέων του Τμήματος Πληροφορικής (*students*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει τον αριθμό των εισακτέων του Τμήματος Πληροφορικής είναι ίση με 193.18, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για τον αριθμό των εισακτέων ενός Τμήματος Πληροφορικής είναι ίση με 193.18.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει τον αριθμό των εισακτέων ενός Τμήματος Πληροφορικής είναι ίση με 225 μονάδες, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν αριθμό εισακτέων το πολύ ίσο με 225, ενώ το υπόλοιπο 50% των Τμημάτων έχουν αριθμό εισακτέων μεγαλύτερο ή ίσο με 225.
- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τον αριθμό των εισακτέων ενός Τμήματος Πληροφορικής είναι ίση με 40.452 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με -0.847 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των εισακτέων ενός Τμήματος Πληροφορικής είναι συμμετρικό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με -0.764 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των εισακτέων ενός Τμήματος Πληροφορικής είναι μεσόκυρτο.
- Το εύρος των τιμών του αριθμού των εισακτέων ενός Τμήματος Πληροφορικής είναι ίσο με 100 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς τον αριθμό των εισακτέων του Τμήματος είναι ίση με 100 μονάδες.

- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του αριθμού των εισακτέων ενός Τμήματος Πληροφορικής είναι ίσο με 175 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Τμήματα Πληροφορικής έχουν αριθμό εισακτέων το πολύ μέχρι 175, ενώ τα υπόλοιπα 3 στα 4 Τμήματα έχουν αριθμό εισακτέων μεγαλύτερο ή ίσο από 175.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του αριθμού των εισακτέων ενός Τμήματος Πληροφορικής είναι ίσο με 225 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Τμήματα έχουν αριθμό εισακτέων το πολύ μέχρι 225, ενώ τα υπόλοιπα 1 στα 4 Τμήματα έχουν αριθμό εισακτέων μεγαλύτερο ή ίσο από 225.

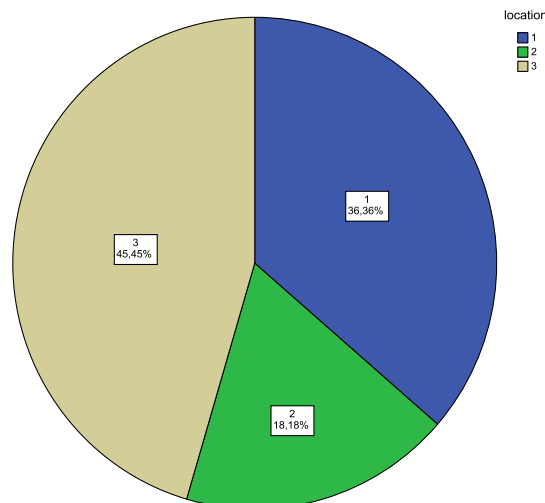
6. Περιγραφικά στοιχεία της μεταβλητής που αντιπροσωπεύει τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π για κάθε Τμήμα Πληροφορικής (*DEP_to_Students*)

- Η μέση τιμή της μεταβλητής που αντιπροσωπεύει τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π για κάθε Τμήμα Πληροφορικής είναι ίση με 9.1631, που πρακτικά σημαίνει ότι μία αντιπροσωπευτική τιμή για τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 9.1631.
- Η διάμεσος της μεταβλητής που αντιπροσωπεύει τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 8.6538 μονάδες, που πρακτικά σημαίνει ότι το 50% των Τμημάτων Πληροφορικής έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π το πολύ ίσο με 8.6538, ενώ το υπόλοιπο 50% των Τμημάτων έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π μεγαλύτερο ή ίσο με 8.6538.
- Η τυπική απόκλιση της μεταβλητής που αντιπροσωπεύει τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίση με 3.44282 μονάδες.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 0.005 και 0.661 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι συμμετρικό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με -1.382 και 1.279 αντίστοιχα. Εφόσον το πηλίκο τους δεν υπερβαίνει κατά απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι μεσόκυρτο.
- Το εύρος των τιμών του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 9.75 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο Τμημάτων Πληροφορικής ως προς τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π του Τμήματος είναι ίση με 9.75 μονάδες.

- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 5.7692 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 Τμήματα Πληροφορικής έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π το πολύ μέχρι 5.7692, ενώ τα υπόλοιπα 3 στα 4 Τμήματα έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π μεγαλύτερο ή ίσο από 5.7692.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π ενός Τμήματος Πληροφορικής είναι ίσο με 12,50 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 Τμήματα έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π το πολύ μέχρι 12.50, ενώ τα υπόλοιπα 1 στα 4 Τμήματα έχουν αριθμό φοιτητών ανά μέλος Δ.Ε.Π μεγαλύτερο ή ίσο από 12.50.

3.3.6 Γραφήματα

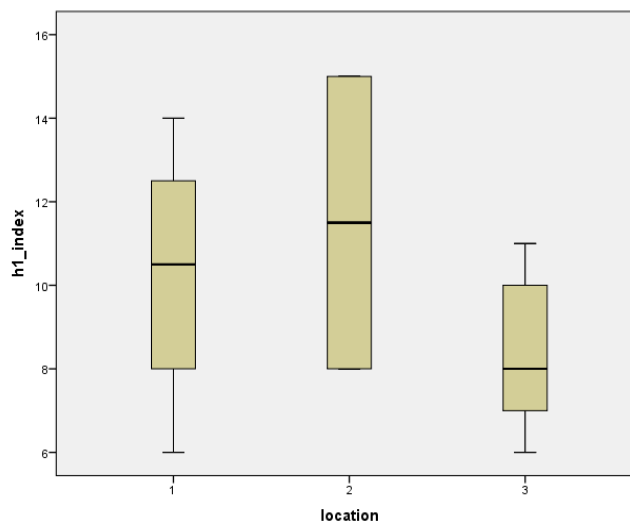
- **Κυκλικό διάγραμμα βάσει της τοποθεσίας των τμημάτων**



Γράφημα 1: Κυκλικό διάγραμμα βάσει της τοποθεσίας των Τμημάτων

Το παραπάνω κυκλικό διάγραμμα (*pie chart*) που απεικονίζεται στο *Γράφημα 1* αντιπροσωπεύει το ποσοστό των Τμημάτων που ανήκουν στην Αθήνα (ομάδα 1), στην Θεσσαλονίκη (ομάδα 2) και σε άλλες πόλεις της Ελλάδας (ομάδα 3). Τα τμήματα που ανήκουν στην Αθήνα καταλαμβάνουν το 35,36% του συνόλου, αυτά που ανήκουν στην Θεσσαλονίκη το 18,18% ενώ αυτά που ανήκουν σε άλλες πόλεις το 45,45%. Αυτό πρακτικά σημαίνει ότι σχεδόν το 50% των Τμημάτων Πληροφορικής είναι συγκεντρωμένα σε άλλες πόλεις και όχι στις δύο μεγάλες, την Αθήνα και την Θεσσαλονίκη.

- **Θηκόγραμμα για τις τιμές του δείκτη h_1 βάσει και της τοποθεσίας των Τμημάτων**



Γράφημα 2: Θηκόγραμμα για τις τιμές του δείκτη h_1 βάσει και της τοποθεσίας των Τμημάτων

Στο Γράφημα 2, παρουσιάζονται τα ποσοστημόρια 25%, 75% και η διάμεσος των τιμών του δείκτη h_1 για κάθε ομάδα Τμημάτων χωριστά. Παρατηρείται εντονότερη ομοιογένεια στην ομάδα των Τμημάτων που βρίσκονται σε πόλεις εκτός Αθήνας και Θεσσαλονίκης (δηλαδή στα τμήματα που ανήκουν στην ομάδα 3).

3.3.6 Ταξινόμηση των τμημάτων ανάλογα με την τιμή του δείκτη h_1

Εφαρμόζουμε τη μέθοδο *k-means Cluster analysis* ώστε να ταξινομήσουμε τα τμήματα Πληροφορικής σε ομοιογενείς ομάδες βάσει της τιμής h_1 που παρουσιάζουν.

- Ταξινόμηση σε 2 ομάδες

Πίνακας 13: Ομαδοποίηση Τμημάτων σε 2 ομάδες βάσει του δείκτη h_1

Case Number	Cluster	Distance
1	2	1,250
2	2	2,250
3	2	1,750
4	2	1,750
5	1	2,143
6	1	1,857
7	1	,143
8	1	1,857
9	1	2,143
10	1	,857
11	1	,143

Όπως φαίνεται από τον Πίνακα 13, τα Τμήματα 1,2,3,4 σχηματίζουν την ομάδα 2, ενώ τα υπόλοιπα Τμήματα την ομάδα 1. Το κριτήριο ταξινόμησης των Τμημάτων στις 2 ομάδες είναι να σχηματισθούν όσο το δυνατόν πιο ομοιογενείς ομάδες, οι οποίες μάλιστα να απέχουν το δυνατόν περισσότερο η μία από την άλλη. Πράγματι, όπως βλέπουμε ακολούθως στον Πίνακα 14.

Πίνακας 14: Στατιστικά μέτρα των δύο ομάδων ως προς τον δείκτη h_1

Cluster Number of Case	N	Mean	Std. Deviation	Std. Error Mean
h1_index 1	7	7,86	1,676	,634
2	4	12,75	2,062	1,031

Οι μέσες τιμές των δύο ομάδων ως προς τον δείκτη h_1 διαφέρουν σημαντικά (η 2^η ομάδα έχει μεγαλύτερη τιμή στον δείκτη h_1 έναντι της 1^{ης} ομάδας). Μάλιστα, με κατάλληλο στατιστικό έλεγχο (*independent samples t-test*) όπως φαίνεται παρακάτω στον Πίνακα 15, η διαφορά κρίνεται στατιστικά σημαντική ($p\text{-value} = 0,002$).

Πίνακας 15: Independent samples t-test για τις σχηματιζόμενες ομάδες

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
h1_index	Equal variances assumed	,833	,385	-4,304	9	,002	-4,893	1,137	-7,465	-2,321
	Equal variances not assumed			-4,044	5,315	,009	-4,893	1,210	-7,948	-1,837

- Ταξινόμηση σε 3 ομάδες (clusters)

Πίνακας 16: Ομαδοποίηση Τμημάτων σε 3 ομάδες βάσει του δείκτη h_1

Case Number	Cluster	Distance
1	2	,500
2	2	,500
3	3	,500
4	3	,500
5	3	,500
6	1	1,000
7	1	1,000
8	1	1,000
9	3	,500
10	1	,000
11	1	1,000

Όπως φαίνεται από τον Πίνακα 16, τα τμήματα 6,7,8,10,11 σχηματίζουν την 1^η ομάδα, τα τμήματα 1,2 τη 2^η ομάδα, ενώ τα υπόλοιπα τμήματα τη 3^η ομάδα. Το κριτήριο ταξινόμησης των τμημάτων στις 3 ομάδες είναι τα σχηματισθούν όσο το δυνατόν πιο ομοιογενείς ομάδες, οι οποίες μάλιστα να απέχουν το δυνατόν περισσότερο η μία από την άλλη. Πράγματι, όπως βλέπουμε ακολούθως στον Πίνακα 17

Πίνακας 17: Στατιστικά μέτρα των τριών ομάδων ως προς τον δείκτη h_1

h1_index	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					1	5		
2	2	14,50	,707	,500	8,15	20,85	14	15
3	4	10,50	,577	,289	9,58	11,42	10	11
Total	11	9,64	3,009	,907	7,61	11,66	6	15

Οι μέσες τιμές των τριών ομάδων ως προς τον δείκτη h_1 διαφέρουν σημαντικά. Μάλιστα σύμφωνα με τον ακόλουθο πίνακα ο κατάλληλος στατιστικός έλεγχος (*independent samples t-test*) δείχνει ότι η διαφορά κρίνεται στατιστικά σημαντική ($p\text{-value}=0,000$).

Πίνακας 18: Ανάλυση διακύμανσης για τις τρεις ομάδες

ANOVA

h1_index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	85,045	2	42,523	61,851	,000
Within Groups	5,500	8	,687		
Total	90,545	10			

3.3.7 Έλεγχος για το αν οι σχηματιζόμενες ομάδες (βάσει του δείκτη h_1) διαφέρουν και ως προς άλλα χαρακτηριστικά

- Για την ταξινόμηση που έγινε σε 2 ομάδες

Πίνακας 19: Independent samples t-test βάση του δείκτη h_1 για τις δύο ομάδες

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
size	Equal variances assumed	17,112	,003	-2,319	9	,046	-14,250	6,144	-28,148	-,352	
	Equal variances not assumed			-1,800	3,389	,159	-14,250	7,916	-37,883	9,383	
age_Uni	Equal variances assumed	2,166	,175	-2,801	9	,021	-63,179	22,557	-114,206	-12,151	
	Equal variances not assumed			-2,253	3,637	,094	-63,179	28,038	-144,187	17,830	
age_Dep	Equal variances assumed	9,299	,014	-1,891	9	,091	-8,821	4,665	-19,373	1,731	
	Equal variances not assumed			-2,363	8,404	,044	-8,821	3,733	-17,359	-,284	
Teaching_staff	Equal variances assumed	,045	,837	-2,328	9	,045	-14,071	6,043	-27,743	-,400	
	Equal variances not assumed			-2,383	6,818	,050	-14,071	5,904	-28,109	-,034	
Admin_staff	Equal variances assumed	,022	,884	-2,932	9	,017	-2,429	,828	-4,302	-,555	
	Equal variances not assumed			-3,104	7,497	,016	-2,429	,782	-4,254	-,603	
students	Equal variances assumed	,280	,610	-,405	9	,695	-10,714	26,486	-70,631	49,202	
	Equal variances not assumed			-,372	5,012	,725	-10,714	28,794	-84,678	63,250	
DEP_to_Students	Equal variances assumed	,571	,469	2,850	9	,019	4,69988	1,64912	,96930	8,43046	
	Equal variances not assumed			3,204	8,600	,011	4,69988	1,46701	1,35758	8,04218	

Όπως φαίνεται από τον παραπάνω πίνακα, οι δύο σχηματιζόμενες ομάδες (βάσει του δείκτη h_1) παρουσιάζουν τα εξής αποτελέσματα:

- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,021$) ως προς το χρόνο που λειτουργούν τα αντίστοιχα Πανεπιστήμια (η 2^η ομάδα έχει μεγαλύτερο χρόνο λειτουργίας έναντι της 1^{ης} ομάδας).
- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,044$) ως προς το χρόνο που λειτουργούν τα Τμήματα (η 2^η ομάδα έχει μεγαλύτερο χρόνο λειτουργίας έναντι της 1^{ης} ομάδας).
- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,045$) ως προς το πλήθος των μελών Δ.Ε.Π που έχουν τα Τμήματα (η 2^η ομάδα έχει μεγαλύτερο πλήθος των μελών Δ.Ε.Π έναντι της 1^{ης} ομάδας).

- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,017$) ως προς το πλήθος των διοικητικών υπαλλήλων που έχουν τα Τμήματα (η 2^η ομάδα έχει μεγαλύτερο πλήθος διοικητικών υπαλλήλων έναντι της 1^{ης} ομάδας).
- Διαφέρουν στατιστικά σημαντικά ($p=0,019$) ως προς τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π που έχουν τα Τμήματα (η 2^η ομάδα έχει μικρότερο αριθμό φοιτητών ανά μέλος Δ.Ε.Π έναντι της 1^{ης} ομάδας).

- Για την ταξινόμηση που έγινε σε 3 ομάδες

Πίνακας 20: Ανάλυση διακύμανσης βάση του δείκτη h_1 για τις τρεις ομάδες

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
size	Between Groups	1163,686	2	581,843	21,357	,001
	Within Groups	217,950	8	27,244		
	Total	1381,636	10			
age_Uni	Between Groups	13146,777	2	6573,389	6,065	,025
	Within Groups	8669,950	8	1083,744		
	Total	21816,727	10			
age_Dep	Between Groups	111,245	2	55,623	,760	,499
	Within Groups	585,300	8	73,163		
	Total	696,545	10			
Teaching_staff	Between Groups	540,777	2	270,389	2,704	,127
	Within Groups	799,950	8	99,994		
	Total	1340,727	10			
Admin_staff	Between Groups	8,777	2	4,389	1,600	,260
	Within Groups	21,950	8	2,744		
	Total	30,727	10			
students	Between Groups	2488,636	2	1244,318	,717	,517
	Within Groups	13875,000	8	1734,375		
	Total	16363,636	10			
DEP_to_Students	Between Groups	69,425	2	34,713	5,655	,029
	Within Groups	49,104	8	6,138		
	Total	118,530	10			

Όπως φαίνεται από τον Πίνακα 20, οι τρεις σχηματιζόμενες ομάδες (βάσει του δείκτη h_1) παρουσιάζουν τα εξής αποτελέσματα

- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,001$) ως προς το πλήθος των Τμημάτων που λειτουργούν στα αντίστοιχα Πανεπιστήμια.
- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,025$) ως προς το χρόνο λειτουργίας των αντίστοιχων Πανεπιστημίων.
- Διαφέρουν στατιστικά σημαντικά ($p\text{-value}=0,025$) ως προς τον αριθμό των φοιτητών ανά μέλος Δ.Ε.Π που έχουν τα Τμήματα.

Μάλιστα, όπως φαίνεται από τον παρακάτω πίνακα

Πίνακας 21: Πολλαπλές συγκρίσεις μεταξύ των τριών ομάδων

Multiple Comparisons

Bonferroni

Dependent Variable	(I) Cluster Number of Case	(J) Cluster Number of Case	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
size	1	2	-26,600	4,367	,001	-39,77	-13,43
		3	,150	3,501	1,000	-10,41	10,71
	2	1	26,600 *	4,367	,001	13,43	39,77
		3	26,750 *	4,520	,001	13,12	40,38
	3	1	-,150	3,501	1,000	-10,71	10,41
		2	-26,750 *	4,520	,001	-40,38	-13,12
age_Uni	1	2	-95,600 *	27,543	,025	-178,66	-12,54
		3	-21,350	22,084	1,000	-87,95	45,25
	2	1	95,600 *	27,543	,025	12,54	178,66
		3	74,250	28,510	,094	-11,73	160,23
	3	1	21,350	22,084	1,000	-45,25	87,95
		2	-74,250	28,510	,094	-160,23	11,73
DEP_to_Students	1	2	6,88259 *	2,07283	,032	,6314	13,1338
		3	2,79654	1,66197	,393	-2,2156	7,8086
	2	1	-6,88259 *	2,07283	,032	-13,1338	-,6314
		3	-4,08605	2,14559	,280	-10,5566	2,3845
	3	1	-2,79654	1,66197	,393	-7,8086	2,2156
		2	4,08605	2,14559	,280	-2,3845	10,5566

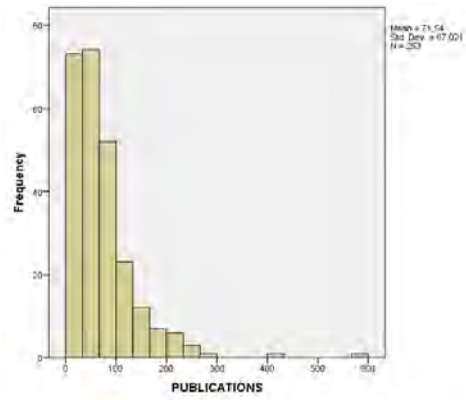
*. The mean difference is significant at the 0.05 level.

- Η 1^η ομάδα παρουσιάζει σημαντικά μικρότερο πλήθος Τμημάτων που λειτουργούν στα αντίστοιχα Πανεπιστήμια έναντι της 2^{ης} ομάδας ($p\text{-value}=0,001$).
- Η 2^η ομάδα παρουσιάζει σημαντικά μεγαλύτερο πλήθος Τμημάτων που λειτουργούν στα αντίστοιχα Πανεπιστήμια έναντι της 3^{ης} ομάδας ($p\text{-value}=0,001$).
- Η 1^η ομάδα παρουσιάζει σημαντικά μικρότερο χρόνο λειτουργίας των αντίστοιχων Πανεπιστημίων έναντι της 2^{ης} ομάδας ($p\text{-value}=0,025$).
- Η 1^η ομάδα παρουσιάζει σημαντικά μεγαλύτερο αριθμό φοιτητών ανά μέλος Δ.Ε.Π των αντίστοιχων Τμημάτων έναντι της 2^{ης} ομάδας ($p\text{-value}=0,032$).

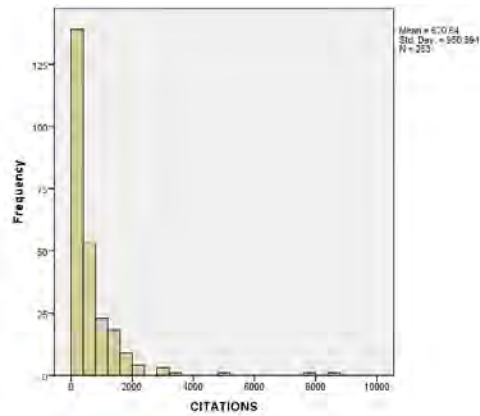
3.4 Σύγκριση σε ατομικό επίπεδο βάσει του δείκτη h_1

3.4.1 Γραφήματα

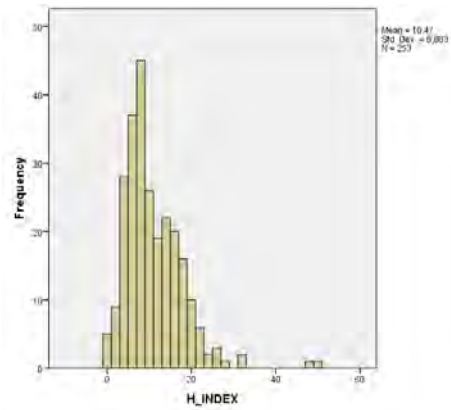
- **Ιστογράμματα για τους δείκτες**



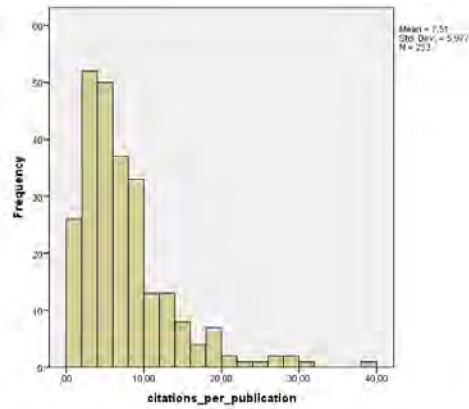
Γράφημα 3: Ιστόγραμμα συχνοτήτων για τον δείκτη publications



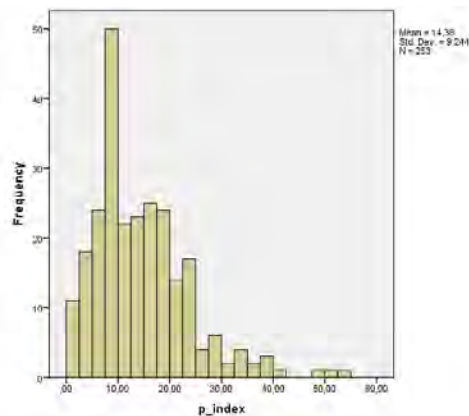
Γράφημα 4: Ιστόγραμμα συχνοτήτων για τον δείκτη citations



Γράφημα 5: Ιστόγραμμα συχνοτήτων για τον δείκτη h-index



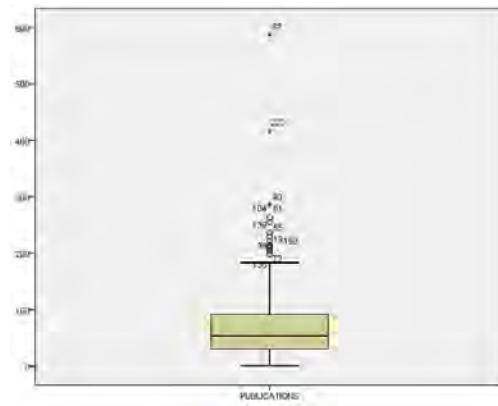
Γράφημα 6: Ιστόγραμμα συχνοτήτων για τον δείκτη citations_per_publication



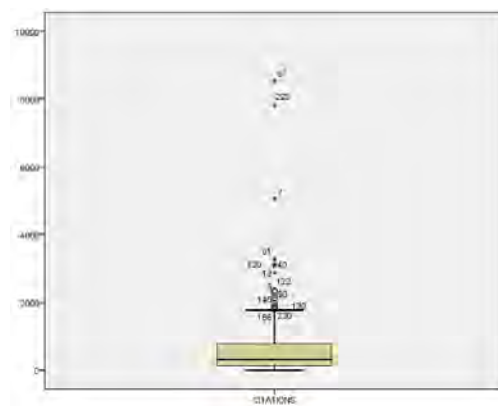
Γράφημα 7: Ιστόγραμμα συχνοτήτων για τον δείκτη p-index

Τα ιστογράμματα αποκαλύπτουν ότι όλοι οι δείκτες (κυρίως οι δείκτες *Publications*, *Citations*, *Citations per publications*) παρουσιάζουν έντονη ασυμμετρία προς τις χαμηλές τιμές τους. Επιπλέον, τα ιστογράμματα παρουσιάζουν τη συχνότητα εμφάνισης όλων των παραπάνω μεταβλητών. Οι δείκτες *Citations* και *Publications* παρουσιάζουν υψηλότερη συχνότητα εμφάνισης στις χαμηλότερες τιμές τους όπως και ο δείκτης *Citations per publications*. Οι δείκτες *p-index* και *h-index* φαίνεται να παρουσιάζουν την μεγαλύτερη συχνότητα στις μεσαίες τιμές και μικρότερη συχνότητα στις χαμηλές και στις υψηλές τιμές τους. Με άλλα λόγια ακολουθούν κατά προσέγγιση την κανονική κατανομή.

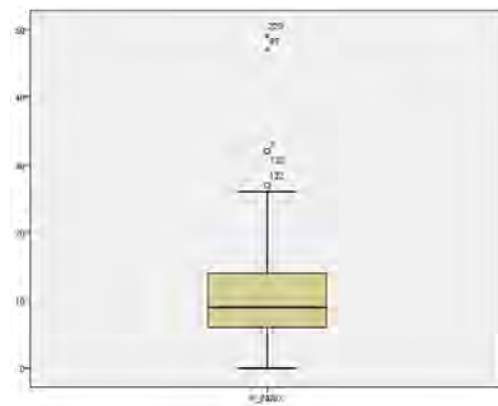
- Θηκογράμματα για τους δείκτες



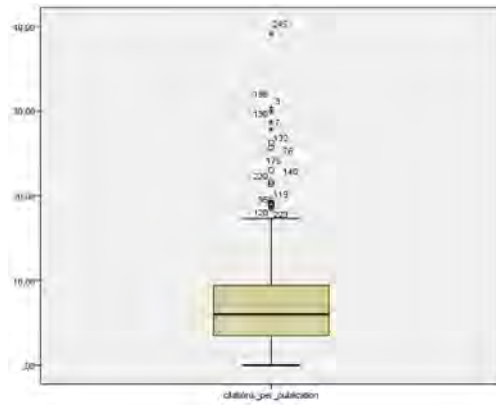
Γράφημα 8: Θηκογράμμα για τον δείκτη publications



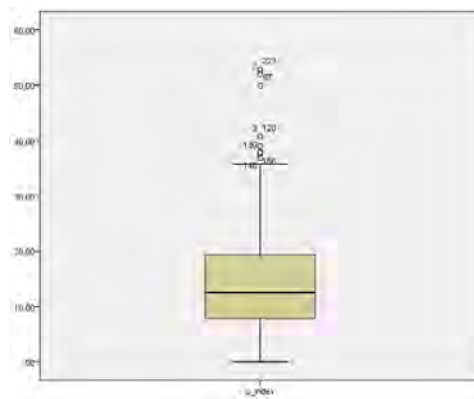
Γράφημα 9: Θηκογράμμα για τον δείκτη citations



Γράφημα 10: Θηκογράμμα για τον δείκτη h-index



Γράφημα 11: Θηκόγραμμα για τον δείκτη citations_per_publication



Γράφημα 12: Θηκόγραμμα για τον δείκτη p-index

Τα θηκογράμματα αποκαλύπτουν ότι υπάρχουν ορισμένες παρατηρήσεις που απέχουν σημαντικά από την αντιπροσωπευτική περιοχή τιμών του εκάστοτε δείκτη. Πρόσθετα, κανένα εκ των χαρακτηριστικών δεν φαίνεται να παρουσιάζει υψηλή μεταβλητότητα.

3.4.2 Περιγραφικά στατιστικά μέτρα για τους δείκτες *Publications*, *Citations*, *h-index*, *Citations per publication* και *p-index*

Πίνακας 22: Περιγραφικά στατιστικά μέτρα για τους λοιπούς δείκτες

		Statistics				
		PUBLICATIO NS	CITATIONS	H_INDEX	citations_per _publication	p_index
N	Valid	253	253	253	253	253
	Missing	27	27	27	27	27
Mean		71,54	620,64	10,47	7,5085	14,3766
Median		54,00	306,00	9,00	5,9828	12,5365
Std. Deviation		67,021	950,594	6,883	5,97728	9,24358
Skewness		3,034	4,840	1,750	1,908	1,274
Std. Error of Skewness		,153	,153	,153	,153	,153
Kurtosis		16,108	32,738	6,210	4,988	2,344
Std. Error of Kurtosis		,305	,305	,305	,305	,305
Range		586	8531	49	39,13	52,73
Minimum		1	0	0	,00	,00
Maximum		587	8531	49	39,13	52,73
Percentiles	25	30,00	127,50	6,00	3,4476	7,9081
	50	54,00	306,00	9,00	5,9828	12,5365
	75	92,50	792,00	14,00	9,5053	19,3709

1. Περιγραφικά στατιστικά μέτρα για τον δείκτη *Publications*

- Η μέση τιμή του δείκτη *Publications* είναι ίση με 71.54, που πρακτικά σημαίνει ότι ένας αντιπροσωπευτικός αριθμός δημοσιεύσεων για ένα μέλος Δ.Ε.Π που υπηρετεί σε ένα Τμήμα Πληροφορικής είναι ίσος με 71.54.
- Η διάμεσος του δείκτη *Publications* είναι ίση με 54, που πρακτικά σημαίνει ότι το 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν το πολύ 54 δημοσιεύσεις, ενώ το υπόλοιπο 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν τουλάχιστον 54 δημοσιεύσεις.
- Η τυπική απόκλιση του δείκτη *Publications* είναι ίση με 67.021 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο μέλος που υπηρετεί σε τμήμα Πληροφορικής αναμένεται να έχει αριθμό δημοσιεύσεων που απέχει περίπου κατά 67 από τον αναμενόμενο αριθμό δημοσιεύσεων που είναι ίσος με 71.54.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 3.034 και 0.153 αντίστοιχα. Εφόσον το ηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *Publications* είναι ασύμμετρο, δηλαδή φαίνεται να υπάρχουν περισσότερα μέλη Δ.Ε.Π που δεν υπερβαίνουν τον μέσο όρο των 71.54 δημοσιεύσεων και λιγότερα μέλη Δ.Ε.Π που υπερβαίνουν τον μέσο όρο αυτό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 4.84 και 0.153 αντίστοιχα. Εφόσον το ηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2,

έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *Publications* είναι λεπτόκυρτο, δηλαδή παρουσιάζει έντονη συγκέντρωση γύρω από την επικρατούσα τιμή του.

- Το εύρος των τιμών του δείκτη *Publications* είναι ίσο με 586 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο μελών Δ.Ε.Π τμημάτων Πληροφορικής ως προς την τιμή του δείκτη *Publications* είναι ίση με 586 δημοσιεύσεις.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του δείκτη *Publications* είναι ίσο με 30 δημοσιεύσεις, που πρακτικά σημαίνει ότι 1 στα 4 μέλη Δ.Ε.Π έχει δείκτη *Publications* μέχρι το πολύ 30 δημοσιεύσεις, ενώ τα υπόλοιπα 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Publications* μεγαλύτερο ή ίσο από 30 δημοσιεύσεις.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του δείκτη *Publications* είναι ίσο με 92.5 δημοσιεύσεις, που πρακτικά σημαίνει ότι 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Publications* μέχρι το πολύ 92.5 δημοσιεύσεις, ενώ τα υπόλοιπα 1 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Publications* μεγαλύτερο ή ίσο από 92.5 δημοσιεύσεις.

2. Περιγραφικά στατιστικά μέτρα για τον δείκτη *Citations*

- Η μέση τιμή του δείκτη *Citations* είναι ίση με 620.64, που πρακτικά σημαίνει ότι ένας αντιπροσωπευτικός αριθμός αναφορών σε μια δημοσίευση για ένα μέλος Δ.Ε.Π που υπηρετεί σε ένα Τμήμα Πληροφορικής είναι ίσος με 620.64.
- Η διάμεσος του δείκτη *Citations* είναι ίση με 306, που πρακτικά σημαίνει ότι το 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν το πολύ 306 αναφορές στις δημοσιεύσεις τους, ενώ το υπόλοιπο 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν τουλάχιστον 306 αναφορές στις δημοσιεύσεις τους.
- Η τυπική απόκλιση του δείκτη *Citations* είναι ίση με 950.59 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο μέλος που υπηρετεί σε Τμήμα Πληροφορικής αναμένεται να έχει αριθμό αναφορών που απέχει περίπου κατά 950.59 από τον αναμενόμενο αριθμό αναφορών που είναι ίσος με 620.64 .
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 4.840 και 0.153 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *Citations* είναι ασύμμετρο, δηλαδή φαίνεται να υπάρχουν περισσότερα μέλη Δ.Ε.Π που δεν υπερβαίνουν τον μέσο όρο των 620.64 αναφορών και λιγότερα μέλη Δ.Ε.Π που υπερβαίνουν τον μέσο όρο αυτό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 32.73 και 0.305 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *Citations* είναι λεπτόκυρτο, δηλαδή παρουσιάζει έντονη συγκέντρωση γύρω από την επικρατούσα τιμή του.

- Το εύρος των τιμών του δείκτη *Citations* είναι ίσο με 8.531 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο μελών Δ.Ε.Π τμημάτων Πληροφορικής ως προς την τιμή του δείκτη *Citations* είναι ίση με 8.531 αναφορές.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του δείκτη *Citations* είναι ίσο με 127.50 αναφορές, που πρακτικά σημαίνει ότι 1 στα 4 μέλη Δ.Ε.Π έχει δείκτη *Citations* μέχρι το πολύ 128 αναφορές, ενώ τα υπόλοιπα 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Citations* μεγαλύτερο ή ίσο από 128 αναφορές.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του δείκτη *Citations* είναι ίσο με 792 αναφορές, που πρακτικά σημαίνει ότι 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Citations* μέχρι το πολύ 792 αναφορές, ενώ τα υπόλοιπα 1 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Citations* μεγαλύτερο ή ίσο από 792 αναφορές.

3. Περιγραφικά στατιστικά μέτρα για τον δείκτη *h-index*

- Η μέση τιμή του δείκτη *h-index* είναι ίση με 10.47, που πρακτικά σημαίνει ότι ένας αντιπροσωπευτικός αριθμός του δείκτη *h-index* για ένα μέλος Δ.Ε.Π που υπηρετεί σε ένα Τμήμα Πληροφορικής είναι ίσος με 10.47.
- Η διάμεσος του δείκτη *h-index* είναι ίση με 9, που πρακτικά σημαίνει ότι το 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν δείκτη *h-index* το πολύ ίσο με 9, ενώ το υπόλοιπο 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν δείκτη *h-index* τουλάχιστον ίσο με 9.
- Η τυπική απόκλιση του δείκτη *h-index* είναι ίση με 6.883 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο μέλος που υπηρετεί σε Τμήμα Πληροφορικής αναμένεται να έχει δείκτη *h-index* που απέχει περίπου κατά 6,883 από την αναμενόμενη τιμή για τον δείκτη *h-index* που είναι ίσος με 10.47.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 1.750 και 0.153 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *h-index* είναι ασύμμετρο, δηλαδή φαίνεται να υπάρχουν περισσότερα μέλη Δ.Ε.Π που δεν υπερβαίνουν τον μέσο όρο του του δείκτη *h-index* ο οποίος ισούται με 10.47 και λιγότερα μέλη Δ.Ε.Π που υπερβαίνουν τον μέσο όρο αυτό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 6.210 και 0.305 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *h-index* είναι λεπτόκυρτο, δηλαδή παρουσιάζει έντονη συγκέντρωση γύρω από την επικρατούσα τιμή του.
- Το εύρος των τιμών του δείκτη *h-index* είναι ίσο με 49 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο μελών Δ.Ε.Π Τμημάτων Πληροφορικής ως προς την τιμή του δείκτη *h-index* είναι ίση με 49.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του δείκτη *h-index* είναι ίσο με 6 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 μέλη Δ.Ε.Π έχει δείκτη

h-index μέχρι το πολύ 6, ενώ τα υπόλοιπα 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *h-index* μεγαλύτερο ή ίσο από 6.

- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του δείκτη *h-index* είναι ίσο με 14 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *h-index* μέχρι το πολύ 14, ενώ τα υπόλοιπα 1 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *h-index* μεγαλύτερο ή ίσο από 14.

4. Περιγραφικά στατιστικά μέτρα για τον δείκτη *Citations per publication*

- Η μέση τιμή του δείκτη *Citations per publication* είναι ίση με 7.50, που πρακτικά σημαίνει ότι ένας αντιπροσωπευτικός αριθμός αναφορών ανά δημοσίευση για ένα μέλος Δ.Ε.Π που υπηρετεί σε ένα Τμήμα Πληροφορικής είναι ίσος με 7.50.
- Η διάμεσος του δείκτη *Citations per publication* είναι ίση με 5.98, που πρακτικά σημαίνει ότι το 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν αριθμό αναφορών ανά δημοσίευση το πολύ ίσο με 5.98, ενώ το υπόλοιπο 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν αριθμό αναφορών ανά δημοσίευση τουλάχιστον ίσο με 5.98.
- Η τυπική απόκλιση του δείκτη *Citations per publication* είναι ίση με 5.97 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο μέλος που υπηρετεί σε Τμήμα Πληροφορικής αναμένεται να έχει δείκτη *Citations per publication* που απέχει περίπου κατά 5.97 μονάδες από την αναμενόμενη τιμή για τον δείκτη *Citations per publication* που είναι ίσος με 7.50.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 1.908 και 0,153 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των αναφορών ανά δημοσίευση για ένα μέλος Δ.Ε.Π είναι ασύμμετρο, δηλαδή φαίνεται να υπάρχουν περισσότερα μέλη Δ.Ε.Π που δεν υπερβαίνουν τον μέσο όρο του του δείκτη *Citations per publication* ο οποίος ισούται με 7.50, και λιγότερα μέλη Δ.Ε.Π που υπερβαίνουν τον μέσο όρο αυτό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 4.988 και 0.305 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ' απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του αριθμού των αναφορών ανά δημοσίευση για ένα μέλος Δ.Ε.Π είναι λεπτόκυρτο, δηλαδή παρουσιάζει έντονη συγκέντρωση γύρω από την επικρατούσα τιμή του.
- Το εύρος των τιμών του αριθμού των αναφορών ανά δημοσίευση είναι ίσο με 39.13 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο μελών Δ.Ε.Π Τμημάτων Πληροφορικής ως προς την τιμή του δείκτη *Citations per publication* είναι ίση με 39.13.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του αριθμού των αναφορών ανά δημοσίευση για ένα μέλος Δ.Ε.Π είναι ίσο με 3.44 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 μέλη Δ.Ε.Π έχει δείκτη *Citations per*

publication μέχρι το πολύ 3.44, ενώ τα υπόλοιπα 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Citations per publication* μεγαλύτερο ή ίσο από 3.44.

- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του αριθμού των αναφορών ανά δημοσίευση για ένα μέλος Δ.Ε.Π είναι ίσο με 9.50 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *Citations per publication* μέχρι το πολύ 9.50, ενώ τα υπόλοιπα 1 στα 4 μέλη Δ.Ε.Π έχουν *Citations per publication* μεγαλύτερο ή ίσο από 9.50.

5. Περιγραφικά στατιστικά μέτρα για τον δείκτη *p-index*

- Η μέση τιμή του δείκτη *p-index* είναι ίση με 14.37, που πρακτικά σημαίνει ότι ένας αντιπροσωπευτικός αριθμός του δείκτη *p-index* για ένα μέλος Δ.Ε.Π που υπηρετεί σε ένα Τμήμα Πληροφορικής είναι ίσος με 14.37.
- Η διάμεσος του δείκτη *p-index* είναι ίση με 12.53, που πρακτικά σημαίνει ότι το 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν δείκτη *p-index* το πολύ ίσο με 12.53, ενώ το υπόλοιπο 50% των μελών Δ.Ε.Π των Τμημάτων Πληροφορικής έχουν δείκτη *p-index* τουλάχιστον ίσο με 12.53.
- Η τυπική απόκλιση του δείκτη *p-index* είναι ίση με 9.24 μονάδες, που πρακτικά σημαίνει ότι ένα τυχαία επιλεγόμενο μέλος που υπηρετεί σε Τμήμα Πληροφορικής αναμένεται να έχει δείκτη *p-index* που απέχει περίπου κατά 9.24 από την αναμενόμενη τιμή για τον δείκτη *p-index* που είναι ίσος με 14.37.
- Ο συντελεστής και το τυπικό σφάλμα ασυμμετρίας είναι ίσα με 1.274 και 0.153 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ'απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *p-index* είναι ασύμμετρο, δηλαδή φαίνεται να υπάρχουν περισσότερα μέλη Δ.Ε.Π που δεν υπερβαίνουν τον μέσο όρο του δείκτη *p-index* ο οποίος ισούται με 14.37 και λιγότερα μέλη Δ.Ε.Π που υπερβαίνουν τον μέσο όρο αυτό.
- Ο συντελεστής και το τυπικό σφάλμα κύρτωσης είναι ίσα με 2.344 και 0.305 αντίστοιχα. Εφόσον το πηλίκο τους υπερβαίνει κατ'απόλυτη τιμή την τιμή 2, έχουμε ενδείξεις ότι το δείγμα των μετρήσεων του δείκτη *p-index* είναι λεπτόκυρτο, δηλαδή παρουσιάζει έντονη συγκέντρωση γύρω από την επικρατούσα τιμή του.
- Το εύρος των τιμών του δείκτη *p-index* είναι ίσο με 52.73 μονάδες, που πρακτικά σημαίνει ότι η μεγαλύτερη απόσταση μεταξύ δύο μελών Δ.Ε.Π Τμημάτων Πληροφορικής ως προς την τιμή του δείκτη *p-index* είναι ίση με 52.73.
- Το ποσοστημόριο 25% του δείγματος των μετρήσεων του δείκτη *p-index* είναι ίσο με 7.90 μονάδες, που πρακτικά σημαίνει ότι 1 στα 4 μέλη Δ.Ε.Π έχει δείκτη *p-index* μέχρι το πολύ 7.90, ενώ τα υπόλοιπα 3 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *p-index* μεγαλύτερο ή ίσο από 7.90.
- Το ποσοστημόριο 75% του δείγματος των μετρήσεων του δείκτη *p-index* είναι ίσο με 19.37 μονάδες, που πρακτικά σημαίνει ότι 3 στα 4 μέλη Δ.Ε.Π έχουν

δείκτη *p-index* μέχρι το πολύ 19.37, ενώ τα υπόλοιπα 1 στα 4 μέλη Δ.Ε.Π έχουν δείκτη *p-index* μεγαλύτερο ή ίσο από 19.37.

3.4.3 Σύνδεση *age_Uni*, *age_Dep*, *size*, *location* με τους δείκτες *Publications*, *Citations*, *h-index*, *Citations per publication*, *p-index*

- *age_Uni* (υπολογισμός συντελεστών συσχέτισης Spearman)

Πίνακας 23: Συντελεστής συσχέτισης Spearman για την μεταβλητή *age_uni*

			Correlations					
			PUBLICATIO NS	CITATIONS	H_INDEX	citations_per _publication	p_index	how many years does the Uni oprerates
Spearman's rho	PUBLICATIONS	Correlation Coefficient	1,000	,847**	,850**	,351**	,697**	,169**
		Sig. (2-tailed)	.	,000	,000	,000	,000	,007
		N	253	253	253	253	253	253
	CITATIONS	Correlation Coefficient	,847**	1,000	,960**	,763**	,967**	,184**
		Sig. (2-tailed)	,000	.	,000	,000	,000	,003
		N	253	253	253	253	253	253
	H_INDEX	Correlation Coefficient	,850**	,960**	1,000	,692**	,911**	,175**
		Sig. (2-tailed)	,000	,000	.	,000	,000	,005
		N	253	253	253	253	253	253
	citations_per_publication	Correlation Coefficient	,351**	,763**	,692**	1,000	,892**	,103
		Sig. (2-tailed)	,000	,000	,000	.	,000	,101
		N	253	253	253	253	253	253
	p_index	Correlation Coefficient	,697**	,967**	,911**	,892**	1,000	,165**
		Sig. (2-tailed)	,000	,000	,000	,000	.	,009
		N	253	253	253	253	253	253
	how many years does the Uni oprerates	Correlation Coefficient	,169**	,184**	,175**	,103	,165**	1,000
		Sig. (2-tailed)	,007	,003	,005	,101	,009	.
		N	253	253	253	253	253	280

** . Correlation is significant at the 0.01 level (2-tailed).

Οι δείκτες *Publications*, *Citations*, *h_index*, *p-index* για ένα μέλος Δ.Ε.Π σχετίζονται ισχυρά και θετικά σε επίπεδο σημαντικότητας 5% με το χρόνο λειτουργίας του αντίστοιχου Πανεπιστημίου στο οποίο ανήκει το μέλος Δ.Ε.Π, καθώς τα *p-values* είναι αντίστοιχα ίσα με 0.007, 0.003, 0.005 και 0.009. Οι αντίστοιχοι συντελεστές συσχέτισης Pearson είναι ίσοι με 0.169, 0.184, 0.175 και 0.165. Αυτό πρακτικά σημαίνει ότι, όσα περισσότερα έτη λειτουργεί το Πανεπιστήμιο στο οποίο ανήκει το μέλος Δ.Ε.Π, τόσο μεγαλύτερη πιθανόν να είναι ερευνητική του δραστηριότητα με αποτέλεσμα οι τιμές των παραπάνω δεικτών να αυξάνονται και να επηρεάζονται σημαντικά.

- **age_Dep (υπολογισμός συντελεστών συσχέτισης Spearman)**

Πίνακας 24: Συντελεστής συσχέτισης Spearman για την μεταβλητή age_dep

			Correlations					
			PUBLICATIO NS	CITATIONS	H_INDEX	citations_per publication	p_index	how many years does the Department operates
Spearman's rho	PUBLICATIONS	Correlation Coefficient	1,000	,847**	,850**	,351**	,697**	,106
		Sig. (2-tailed)	.	,000	,000	,000	,000	,094
		N	253	253	253	253	253	253
	CITATIONS	Correlation Coefficient	,847**	1,000	,960**	,763**	,967**	,172**
		Sig. (2-tailed)	,000	.	,000	,000	,000	,006
		N	253	253	253	253	253	253
	H_INDEX	Correlation Coefficient	,850**	,960**	1,000	,692**	,911**	,141
		Sig. (2-tailed)	,000	,000	.	,000	,000	,025
		N	253	253	253	253	253	253
	citations_per_publication	Correlation Coefficient	,351**	,763**	,692**	1,000	,892**	,159
		Sig. (2-tailed)	,000	,000	,000	.	,000	,012
		N	253	253	253	253	253	253
	p_index	Correlation Coefficient	,697**	,967**	,911**	,892**	1,000	,177**
		Sig. (2-tailed)	,000	,000	,000	,000	.	,005
		N	253	253	253	253	253	253
	how many years does the Department operates	Correlation Coefficient	,106	,172**	,141*	,159	,177**	1,000
		Sig. (2-tailed)	,094	,006	,025	,012	,005	.
		N	253	253	253	253	253	280

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Οι δείκτες *Citations*, *h_index*, *citations_per_publication*, *p-index* για ένα μέλος Δ.Ε.Π σχετίζονται ισχυρά και θετικά σε επίπεδο σημαντικότητας 5% με το χρόνο λειτουργίας του αντίστοιχου Τμήματος στο οποίο ανήκει το μέλος Δ.Ε.Π, καθώς τα *p-values* είναι αντίστοιχα ίσα με 0.006, 0.025, 0.012 και 0.005. Οι αντίστοιχοι συντελεστές συσχέτισης είναι 0.172, 0.141, 0.159 και 0.177. Όπως και προηγουμένως, έτσι και σε αυτή την περίπτωση, όσο περισσότερα είναι τα χρόνια που λειτουργεί το τμήμα στο οποίο ανήκει το μέλος Δ.Ε.Π τόσο μεγαλύτερη είναι η ερευνητική δραστηριότητα που αναπτύσσει το μέλος και επομένως τόσο μεγαλύτερες τιμές λαμβάνουν οι δείκτες αυτοί.

- **Size (υπολογισμός συντελεστών συσχέτισης Spearman)**

Πίνακας 25: Συντελεστής συσχέτισης Spearman για την μεταβλητή size

			Correlations					
			PUBLICATIO NS	CITATIONS	H_INDEX	citations_per publication	p_index	how many Departments does the Uni have
Spearman's rho	PUBLICATIONS	Correlation Coefficient	1,000	,847**	,850**	,351**	,697**	,390**
		Sig. (2-tailed)	.	,000	,000	,000	,000	,000
		N	253	253	253	253	253	253
	CITATIONS	Correlation Coefficient	,847**	1,000	,960**	,763**	,967**	,389**
		Sig. (2-tailed)	,000	.	,000	,000	,000	,000
		N	253	253	253	253	253	253
	H_INDEX	Correlation Coefficient	,850**	,960**	1,000	,692**	,911**	,414**
		Sig. (2-tailed)	,000	,000	.	,000	,000	,000
		N	253	253	253	253	253	253
	citations_per_publication	Correlation Coefficient	,351**	,763**	,692**	1,000	,892**	,210**
		Sig. (2-tailed)	,000	,000	,000	.	,000	,001
		N	253	253	253	253	253	253
	p_index	Correlation Coefficient	,697**	,967**	,911**	,892**	1,000	,349**
		Sig. (2-tailed)	,000	,000	,000	,000	.	,000
		N	253	253	253	253	253	253
	how many Departments does the Uni have	Correlation Coefficient	,390**	,389**	,414**	,210**	,349**	1,000
		Sig. (2-tailed)	,000	,000	,000	,001	,000	.
		N	253	253	253	253	253	280

** Correlation is significant at the 0.01 level (2-tailed).

Οι δείκτες *Publications*, *Citations*, *h_index*, *citations_per_publication*, *p-index* για ένα μέλος Δ.Ε.Π σχετίζονται ισχυρά και θετικά σε επίπεδο σημαντικότητας 5% με το πλήθος των Τμημάτων που λειτουργούν στο αντίστοιχο Πανεπιστήμιο στο οποίο υπηρετεί το μέλος Δ.Ε.Π, καθώς τα *p-values* είναι αντίστοιχα ίσα με <0.001, <0.001, <0.001, 0.001, <0.001. Οι αντίστοιχοι συντελεστές συσχέτισης Pearson είναι 0,390, 0.389, 0.414, 0.210 και 0.349. Με άλλα λόγια προκύπτει το συμπέρασμα ότι οι τιμές των δεικτών που μετρούν την ερευνητική δραστηριότητα ενός μέλους Δ.Ε.Π αυξάνονται και μάλιστα σημαντικά όσο περισσότερα Τμήματα διαθέτει το Πανεπιστήμιο στο οποίο ανήκει το εκάστοτε Τμήμα Πληροφορικής και κατ' επέκταση το μέλος Δ.Ε.Π.

- **location (με εφαρμογή Ανάλυση Διακύμανσης κατά έναν παράγοντα)**

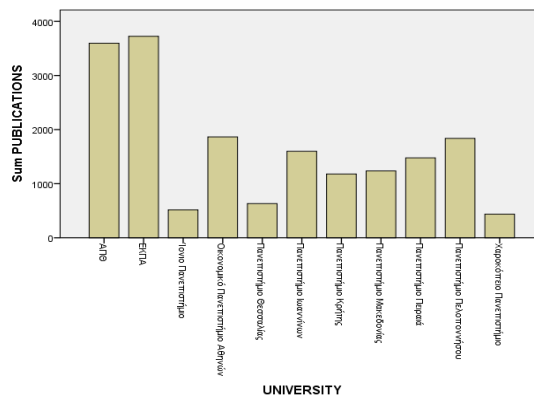
Πίνακας 26: Ανάλυση Διακύμανσης κατά έναν παράγοντα για την μεταβλητή location

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
PUBLICATIONS	Between Groups	6420,153	2	3210,077	,713	,491
	Within Groups	1125518,740	250	4502,075		
	Total	1131938,893	252			
CITATIONS	Between Groups	17695,294	2	8847,647	,010	,990
	Within Groups	227696802,7	250	910787,211		
	Total	227714498,0	252			
H_INDEX	Between Groups	21,132	2	10,566	,222	,801
	Within Groups	11917,896	250	47,672		
	Total	11939,028	252			
citations_per_publication	Between Groups	135,955	2	67,977	1,916	,149
	Within Groups	8867,465	250	35,470		
	Total	9003,420	252			
p_index	Between Groups	229,931	2	114,966	1,349	,261
	Within Groups	21301,886	250	85,208		
	Total	21531,817	252			

Η τοποθεσία δεν επιδρά σημαντικά στους δείκτες *Publications*, *Citations*, *h-index*, *citations per publication* και *p-index*, καθώς τα αντίστοιχα *p-values* είναι μεγαλύτερα από 5%. Αυτό πρακτικά σημαίνει ότι οι τιμές των δεικτών αυτών είναι ανεξάρτητες της τοποθεσίας του Τμήματος στο οποίο δραστηριοποιείται το μέλος Δ.Ε.Π με αποτέλεσμα να μην υπάρχει κάποια σημαντική αλλαγή στις τιμές αυτές όταν πραγματοποιείται σύγκριση μεταξύ μελών που ανήκουν σε Τμήματα που βρίσκονται σε διαφορετική τοποθεσία.

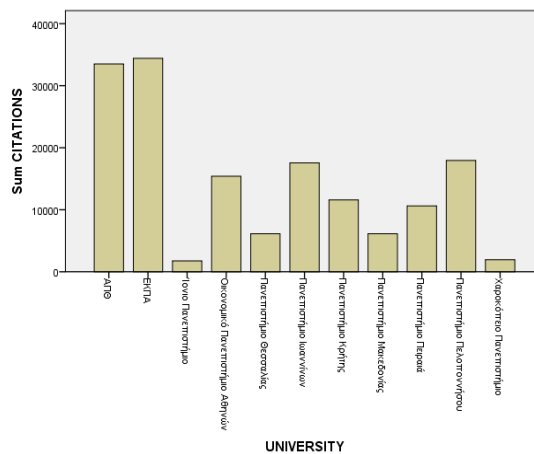
3.4.4 Πρόσθετα γραφήματα και αριθμητικά στοιχεία

- **Απόλυτος αριθμός δημοσιεύσεων**



Γράφημα 13: Απόλυτος αριθμός δημοσιεύσεων ανά Τμήμα

- **Απόλυτος αριθμός αναφορών**



Γράφημα 14: Απόλυτος αριθμός αναφορών ανά Τμήμα

- **Απόλυτος αριθμός δημοσιεύσεων και αναφορών ανά Τμήμα**

Ο Πίνακας 27 που ακολουθεί στη συνέχεια προκύπτει από τα προηγούμενα ραβδογράμματα τα οποία απεικονίζουν τον συνολικό αριθμό αναφορών και δημοσιεύσεων ανά Τμήμα. Σύμφωνα με αυτόν, το δημοφιλέστερο Τμήμα ως προς τον αριθμό των δημοσιεύσεων είναι το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ. Το Τμήμα αυτό όχι μόνο έχει τις περισσότερες δημοσιεύσεις αλλά λαμβάνει και τις περισσότερες αναφορές. Το Τμήμα αυτό ήταν αναμενόμενο να έχει τις περισσότερες δημοσιεύσεις καθώς και τις περισσότερες αναφορές αφού όπως είδαμε προηγουμένως διαθέτει τα περισσότερα μέλη Δ.Ε.Π.

Πίνακας 27: Αριθμός δημοσιεύσεων και αναφορών ανά Τμήμα

ΙΔΡΥΜΑ	ΔΗΜΟΣΙΕΥΣΕΙΣ	ΑΝΑΦΟΡΕΣ
ΕΚΠΑ	2872	29634
ΑΠΘ	1072	5374
ΟΠΑ	2266	20056
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ	2460	21444
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ	1384	11133
ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	1167	9920
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ	835	4954
ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	390	1650
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ	1815	19891
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ	1081	7709
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ	2757	25258

Ωστόσο, δεν μπορούμε να βγάλουμε εύκολα το συμπέρασμα ποιο τμήμα υπερτερεί από άποψη επιστημονικής και ερευνητικής δραστηριότητας μετρώντας μόνο τον συνολικό αριθμό των δημοσιεύσεων και των αναφορών. Για τον λόγο αυτό, θα αναζητηθούν και θα εξετασθούν και άλλοι παράγοντες που επηρεάζουν την αξιολόγηση ενός τμήματος.

3.4.5 Αριθμητικές συγκρίσεις μεταξύ των Τμημάτων

Στον παρακάτω πίνακα (Πίνακας 28) έχουν υπολογιστεί και καταγραφεί για κάθε Τμήμα ξεχωριστά η μέση τιμή και η τυπική απόκλιση του αριθμού των δημοσιεύσεων και των αναφορών ανά μέλος Δ.Ε.Π. Ο Πίνακας 28 προκύπτει από τους πίνακες 29 και 30 αντίστοιχα που φαίνονται στη συνέχεια

Πίνακας 28: Μέση τιμή (και τυπική απόκλιση) αριθμού δημοσιεύσεων και αναφορών ανά μέλος Δ.Ε.Π για κάθε Τμήμα

ΙΔΡΥΜΑ	Αριθμός δημοσιεύσεων ανά μέλος Δ.Ε.Π	Αριθμός αναφορών ανά μέλος Δ.Ε.Π
ΕΚΠΑ	66.79 (45.65)	689.16 (708.70)
ΑΠΘ	38.29 (26.80)	191.93 (206.34)
ΟΠΑ	75.53 (61.40)	668.53 (749.01)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ	102.50 (119.41)	893.50 (1733.17)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ	55.36 (44.73)	445.32 (470.32)
ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	97.25 (88.32)	826.67 (677.49)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ	59.64 (37.06)	353.86 (266.64)
ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	48.75 (37.07)	206.25 (183.28)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ	82.50 (58.50)	904.14 (1204.92)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ	90.08 (57.38)	642.42 (511.84)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ	78.77 (82.88)	721.66 (1353.02)

Πίνακας 29: Μέση τιμή και τυπική απόκλιση του αριθμού των δημοσιεύσεων για τα 11 Τμήματα

Descriptive Statistics
Dependent Variable: PUBLICATIONS

VAR00001	Mean	Std. Deviation	N
1	66,79	45,653	43
10	90,08	57,385	12
11	78,77	82,889	35
2	38,29	26,808	28
3	75,53	61,400	30
4	102,50	119,418	24
5	55,36	44,736	25
6	97,25	88,332	12
7	59,64	37,069	14
8	48,75	37,078	8
9	82,50	58,505	22
Total	71,54	67,021	253

Πίνακας 30: Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών για τα 11 Τμήματα

Descriptive Statistics
Dependent Variable: CITATIONS

VAR00001	Mean	Std. Deviation	N
1	689,16	708,700	43
10	642,42	511,845	12
11	721,66	1353,024	35
2	191,93	206,343	28
3	668,53	749,012	30
4	893,50	1733,176	24
5	445,32	470,320	25
6	826,67	677,499	12
7	353,86	266,264	14
8	206,25	183,288	8
9	904,14	1204,926	22
Total	620,64	950,594	253

Επιπρόσθετα, έχουν υπολογιστεί και καταγραφεί για κάθε Τμήμα χωριστά ο συνολικός αριθμός των αναφορών ανά δημοσίευση και οι δείκτες *p-index* και *h-index*. Όπως φαίνεται και στον Πίνακα 31 το Τμήμα με τον μεγαλύτερο αριθμό αναφορών ανά δημοσίευση (10.26) είναι το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου, ενώ ακολουθούν τα Τμήματα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ (9.75) και του Πανεπιστημίου Πελοποννήσου (8.53). Επιπλέον, το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου έχει τους μεγαλύτερους δείκτες *p-index* (18.42) και *h-index* (12.75) το οποίο είναι αναμενόμενο καθώς οι δείκτες αυτοί επηρεάζονται άμεσα και σε μεγάλο βαθμό από τον αριθμό των αναφορών ανά δημοσίευση.

Πίνακας 31: Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών ανά δημοσίευση και των δεικτών *h-index* και *p-index* για κάθε Τμήμα

ΙΔΡΥΜΑ	Αναφορές ανά δημοσίευση	<i>p-index</i>	<i>h-index</i>
ΕΚΠΑ	9,75 (8,49)	16,67 (10,19)	11,14 (6,51)
ΑΠΘ	4,69 (3,26)	8,72 (4,84)	6,29 (3,16)
ΟΠΑ	7,39 (4,43)	15,10 (8,71)	11,77 (6,80)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ	6,38 (5,36)	14,83 (10,94)	12,08 (9,54)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ	7,97 (5,52)	13,35 (7,20)	9,08 (5,03)
ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	10,26 (6,03)	18,42 (8,02)	12,75 (6,63)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ	7,01 (5,10)	12,15 (5,47)	8,79 (4,24)
ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	3,68 (1,97)	8,23 (4,62)	6,50 (4,07)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ	8,53 (7,58)	17,17 (12,52)	12,41 (7,89)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

6,79 (3,57)

15.16 (6.88)

11,92 (5,45)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

7,26 (5,10)

14.73 (9.91)

10,86 (8,33)

Πίνακας 32: : Μέση τιμή και τυπική απόκλιση του αριθμού των αναφορών ανά δημοσίευση για τα 11 Τμήματα

Descriptive Statistics

Dependent Variable: citations_per_publication

VAR00001	Mean	Std. Deviation	N
1	9,7526	8,49354	43
10	6,7946	3,57006	12
11	7,2621	5,10335	35
2	4,6901	3,26959	28
3	7,3985	4,43644	30
4	6,3823	5,36818	24
5	7,9790	5,52434	25
6	10,2641	6,03498	12
7	7,0165	5,10680	14
8	3,6899	1,97866	8
9	8,5336	7,58535	22
Total	7,5085	5,97728	253

Πίνακας 33: : Μέση τιμή και τυπική απόκλιση του δείκτη p-index για τα 11 Τμήματα

Descriptive Statistics

Dependent Variable: p_index

VAR00001	Mean	Std. Deviation	N
1	16,6727	10,19646	43
10	15,1689	6,88602	12
11	14,7395	9,91841	35
2	8,7214	4,84432	28
3	15,1039	8,71571	30
4	14,8376	10,94244	24
5	13,3545	7,20792	25
6	18,4293	8,02421	12
7	12,1592	5,47313	14
8	8,2325	4,62649	8
9	17,1787	12,52013	22
Total	14,3766	9,24358	253

Πίνακας 34: : Μέση τιμή και τυπική απόκλιση του δείκτη *h-index* για τα 11 Τμήματα

Descriptive Statistics

Dependent Variable: H_INDEX

VAR00001	Mean	Std. Deviation	N
1	11,14	6,516	43
10	11,92	5,452	12
11	10,86	8,332	35
2	6,29	3,161	28
3	11,77	6,801	30
4	12,08	9,541	24
5	9,08	5,033	25
6	12,75	6,635	12
7	8,79	4,246	14
8	6,50	4,071	8
9	12,41	7,896	22
Total	10,47	6,883	253

3.4.6 Στατιστικές Συγκρίσεις μεταξύ των Τμημάτων

- Δείκτης *Publications*

Πίνακας 35: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των δημοσιεύσεων

Tests of Between-Subjects Effects

Dependent Variable: PUBLICATIONS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	84629,284 ^a	10	8462,928	1,956	,039
Intercept	1035472,630	1	1035472,630	239,265	,000
VAR00001	84629,284	10	8462,928	1,956	,039
Error	1047309,610	242	4327,726		
Total	2426697,000	253			
Corrected Total	1131938,893	252			

a. R Squared = ,075 (Adjusted R Squared = ,037)

Εφαρμόζοντας Ανάλυση Διακύμανσης κατά έναν παράγοντα, προκύπτει το συμπέρασμα ότι υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στα Τμήματα Πληροφορικής ως προς τον αριθμό δημοσιεύσεων που έχουν τα μέλη Δ.Ε.Π τους, καθώς το αντίστοιχο *p-value* είναι ίσο με $0.039 < 0.05$. Πιο συγκεκριμένα, όπως βλέπουμε από στον Πίνακα 36 προκύπτει ότι:

Πίνακας 36: Συγκριτικός πίνακας ως προς τον αριθμό των δημοσιεύσεων

Σύγκριση	Μέση διαφορά	p-value	95% Διάστημα εμπιστοσύνης για τη διαφορά
ΕΚΠΑ vs Παν. Ιωαννίνων	-35.709	0.034	(-68.727, -2.691)
ΑΠΘ vs Παν. Θεσσαλίας	-51.798	0.023	(-95.509, -7.086)
ΑΠΘ vs Παν. Μακεδονίας	-40.486	0.016	(-73.342, -7.630)
ΑΠΘ vs ΟΠΑ	-37.248	0.032	(-71.299, -3.197)
ΑΠΘ vs Παν. Ιωαννίνων	-64.214	0.001	(-100.262, -28.167)
ΑΠΘ vs Ιόνιο Παν.	-58.964	0.010	(-103.675, -14.253)
ΑΠΘ vs Παν. Πελοποννήσου	-44.214	0.019	(-81.133, -7.295)
Παν. Πειραιώς vs Παν. Ιωαννίνων	-47.140	0.013	(-84.172, -10.108)
Χαροκόπειο Παν. vs Παν Ιωαννίνων	-53.750	0.046	(-106.653, -0.847)

- Το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ παρουσιάζει σημαντικά μικρότερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π σε σχέση με το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων ($p\text{-value}=0,034$)
- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Θεσσαλίας ($p\text{-value}=0.023$), του Πανεπιστημίου Μακεδονίας ($p\text{-value}=0.016$), του ΟΠΑ ($p\text{-value}=0.032$), του Ιονίου Πανεπιστημίου ($p\text{-value}=0.010$), του Πανεπιστημίου Ιωαννίνων ($p\text{-value}=0.001$) και του Πανεπιστημίου Πελοποννήσου ($p\text{-value}=0.019$).
- Το Τμήμα Πληροφορικής του Πανεπιστημίου Πειραιώς έχει σημαντικά μικρότερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π συγκριτικά με το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων ($p\text{-value}=0.013$)
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου έχει σημαντικά μικρότερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π συγκριτικά με το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων ($p\text{-value}=0.046$)

- Δείκτης Citations

Πίνακας 37: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των αναφορών

Tests of Between-Subjects Effects

Dependent Variable: CITATIONS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12982752,0... ^a	10	1298275,209	1,463	,154
Intercept	70065566,3...	1	70065566,3...	78,963	,000
VAR00001	12982752,0...	10	1298275,209	1,463	,154
Error	214731745,...	242	887321,264		
Total	325169923,...	253			
Corrected Total	227714497,...	252			

a. R Squared = ,057 (Adjusted R Squared = ,018)

Εφαρμόζοντας Ανάλυση Διακύμανσης κατά έναν παράγοντα, προκύπτει το συμπέρασμα ότι δεν υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στα Τμήματα Πληροφορικής ως προς τον αριθμό αναφορών που έχουν τα μέλη Δ.Ε.Π τους, καθώς το αντίστοιχο *p-value* είναι ίσο με $0.154 > 0.05$. Παρόλο που υπάρχουν στατιστικά σημαντικές διαφορές ως προς τον αριθμό των δημοσιεύσεων ανά μέλος Δ.Ε.Π όπως είδαμε προηγουμένως, δεν συμβαίνει το ίδιο με τον αριθμό των αναφορών. Αυτό πρακτικά σημαίνει ότι ο αριθμός των αναφορών ανά μέλος Δ.Ε.Π είναι ανεξάρτητος και δεν επηρεάζεται σημαντικά από τον αριθμό των δημοσιεύσεων που έχει κάθε Τμήμα.

- Δείκτης Citations per publication

Πίνακας 38: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον αριθμό των αναφορών ανά δημοσίευση

Tests of Between-Subjects Effects

Dependent Variable: citations_per_publication

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	717,820 ^a	10	71,782	2,097	,025
Intercept	10411,196	1	10411,196	304,083	,000
VAR00001	717,820	10	71,782	2,097	,025
Error	8285,600	242	34,238		
Total	23267,069	253			
Corrected Total	9003,420	252			

a. R Squared = ,080 (Adjusted R Squared = ,042)

Εφαρμόζοντας Ανάλυση Διακύμανσης κατά έναν παράγοντα, προκύπτει το συμπέρασμα ότι υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στα Τμήματα Πληροφορικής ως προς τον αριθμό αναφορών ανά δημοσίευση που έχουν τα μέλη Δ.Ε.Π τους, καθώς το αντίστοιχο *p-value* είναι ίσο με $0.025 < 0.05$. Πιο συγκεκριμένα, όπως βλέπουμε από τον παρακάτω πίνακα προκύπτει ότι:

Πίνακας 39: Συγκριτικός πίνακας ως προς τον αριθμό των αναφορών ανά δημοσίευση

Σύγκριση	Μέση διαφορά	<i>p-value</i>	95% Διάστημα εμπιστοσύνης για τη διαφορά
ΕΚΠΑ vs ΑΠΘ	5.063	< 0.001	(2.264, 7.861)
ΕΚΠΑ vs Παν. Ιωαννίνων	3.370	0.025	(0.433, 6.307)
ΕΚΠΑ vs Χαροκόπειο Παν.	6.063	0.008	(1.625, 10.501)
Παν. Πειραιώς vs ΑΠΘ	3.289	0.042	(0.117, 6.460)
Ιόνιο Παν. vs ΑΠΘ	5.574	0.006	(1.597, 9.551)
Παν. Πελοποννήσου vs ΑΠΘ	3.843	0.022	(0.560, 7.127)
Ιόνιο Παν. vs Χαροκόπειο Παν.	6.574	0.015	(1.313, 11.835)
Παν. Πελοποννήσου vs Χαροκόπειο Παν.	4.844	0.046	(0.085, 9.602)

- Το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ παρουσιάζει σημαντικά μεγαλύτερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του ΑΠΘ (*p-value*<0,001), του Πανεπιστημίου Ιωαννίνων (*p-value*=0.025) και του Χαροκόπειου Πανεπιστημίου (*p-value*=0.008).
- Το Τμήμα Πληροφορικής του Πανεπιστημίου Πειραιώς παρουσιάζει σημαντικά μεγαλύτερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με το Τμήμα Πληροφορικής του ΑΠΘ (*p-value*=0.042).
- Το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου παρουσιάζει σημαντικά μεγαλύτερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του ΑΠΘ (*p-value*=0.006) και του Χαροκόπειου Πανεπιστημίου (*p-value*=0.015).
- Το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Πελοποννήσου παρουσιάζει σημαντικά μεγαλύτερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του ΑΠΘ (*p-value*=0.022) και του Χαροκόπειου Πανεπιστημίου (*p-value*=0.046).

- Δείκτης *h-index*

Πίνακας 40: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον δείκτη *h-index*

Tests of Between-Subjects Effects

Dependent Variable: H_INDEX

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1011,983 ^a	10	101,198	2,241	,016
Intercept	21108,175	1	21108,175	467,480	,000
VAR00001	1011,983	10	101,198	2,241	,016
Error	10927,045	242	45,153		
Total	39675,000	253			
Corrected Total	11939,028	252			

a. R Squared = ,085 (Adjusted R Squared = ,047)

Εφαρμόζοντας Ανάλυση Διακύμανσης κατά έναν παράγοντα, προκύπτει το συμπέρασμα ότι υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στα Τμήματα Πληροφορικής ως προς τον δείκτη *h-index* που έχουν τα μέλη Δ.Ε.Π τους, καθώς το αντίστοιχο *p-value* είναι ίσο με $0.016 < 0.05$. Πιο συγκεκριμένα, όπως βλέπουμε από τον ακόλουθο πίνακα:

Πίνακας 41: Συγκριτικός πίνακας ως προς τον αριθμό τον δείκτη *h-index*

Σύγκριση	Μέση διαφορά	<i>p-value</i>	95% Διάστημα εμπιστοσύνης για τη διαφορά
ΕΚΠΑ vs ΑΠΘ	4.854	0.003	(1.64, 8.068)
Παν. Θεσσαλίας vs ΑΠΘ	5.631	0.016	(1.064, 10.198)
Παν. Μακεδονίας vs ΑΠΘ	4.571	0.008	(1.215, 7.927)
ΟΠΑ vs ΑΠΘ	5.481	0.002	(2.003, 8.959)
Παν Ιωαννίνων vs ΑΠΘ	5.798	0.002	(2.116, 9.480)
Παν Ιωαννίνων vs Χαροκόπειο Παν.	5.583	0.043	(0.180, 10.987)
Ιόνιο Παν. vs ΑΠΘ	6.464	0.006	(1.897, 11.031)
Ιόνιο Παν. vs Χαροκόπειο Παν.	6.250	0.043	(0.208, 12.292)
Παν. Πελοποννήσου vs ΑΠΘ	6.123	0.002	(2.352, 9.894)
Παν. Πελοποννήσου vs Χαροκόπειο Παν.	5.909	0.034	(0.444, 11.374)

- Το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ παρουσιάζει σημαντικά μεγαλύτερο δείκτη *h-index* σε σχέση με το Τμήμα Πληροφορικής του ΑΠΘ (*p-value*=0,003).
- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο δείκτη *h-index* συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Θεσσαλίας (*p-value*=0.016), του Πανεπιστημίου Μακεδονίας (*p-value*=0.008), του ΟΠΑ (*p-value*=0.002), του Πανεπιστημίου Ιωαννίνων (*p-value*=0.002), του Ιονίου Πανεπιστημίου (*p-value*=0.006) και του Πανεπιστημίου Πελοποννήσου (*p-value*=0.002).
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου παρουσιάζει σημαντικά μικρότερο δείκτη *h-index* συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Ιωαννίνων (*p-value*=0.043), του Ιονίου Πανεπιστημίου (*p-value*=0.043) και του Πανεπιστημίου Πελοποννήσου (*p-value*=0.034).
- **Δείκτης *p-index***

Πίνακας 42: Ανάλυση Διακύμανσης κατά έναν παράγοντα ως προς τον δείκτη *p-index*

Tests of Between-Subjects Effects

Dependent Variable: *p_index*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1922,073 ^a	10	192,207	2,372	,011
Intercept	39111,306	1	39111,306	482,665	,000
VAR00001	1922,073	10	192,207	2,372	,011
Error	19609,744	242	81,032		
Total	73823,714	253			
Corrected Total	21531,817	252			

a. R Squared = ,089 (Adjusted R Squared = ,052)

Εφαρμόζοντας Ανάλυση Διακύμανσης κατά έναν παράγοντα, προκύπτει το συμπέρασμα ότι υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στα Τμήματα Πληροφορικής ως προς τον δείκτη *p-index* που έχουν τα μέλη Δ.Ε.Π τους, καθώς το

αντίστοιχο p -value είναι ίσο με $0.011 < 0.05$. Πιο συγκεκριμένα, όπως βλέπουμε από τον ακόλουθο πίνακα:

Πίνακας 43: Συγκριτικός πίνακας ως προς τον αριθμό τον δείκτη p -index

Σύγκριση	Μέση διαφορά	p -value	95% Διάστημα εμπιστοσύνης για τη διαφορά
ΕΚΠΑ vs ΑΠΘ	7.951	< 0.001	(3.645, 12.257)
ΕΚΠΑ vs Χαροκόπειο Παν.	8.440	0.016	(1.613, 15.268)
Παν. Θεσσαλίας vs ΑΠΘ	6.447	0.039	(0.329, 12.566)
ΟΠΑ vs ΑΠΘ	6.382	0.007	(1.723, 11.042)
Παν Ιωαννίνων vs ΑΠΘ	6.116	0.015	(1.184, 11.049)
Ιόνιο Παν. vs ΑΠΘ	9.708	0.002	(3.590, 15.826)
Ιόνιο Παν. vs Χαροκόπειο Παν.	10.197	0.014	(2.103, 18.290)
Παν. Πελοποννήσου vs ΑΠΘ	8.457	0.001	(3.406, 13.509)
Παν. Πελοποννήσου vs Χαροκόπειο Παν.	8.946	0.017	(1.625, 16.267)

- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο δείκτη p -index συγκριτικά με τα αντίστοιχα Τμήματα του ΕΚΠΑ (p -value<0.001), του Πανεπιστημίου Θεσσαλίας (p -value=0.039), του ΟΠΑ (p -value=0.007), του Πανεπιστημίου Ιωαννίνων (p -value=0.015), του Ιονίου Πανεπιστημίου (p -value=0.002) και του Πανεπιστημίου Πελοποννήσου (p -value=0.001).
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου παρουσιάζει σημαντικά μικρότερο δείκτη p -index συγκριτικά με τα αντίστοιχα Τμήματα του ΕΚΠΑ (p -value=0.016), του Ιονίου Πανεπιστημίου (p -value=0.014) και του Πανεπιστημίου Πελοποννήσου (p -value=0.017).

ΚΕΦΑΛΑΙΟ 4

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα μελέτη φιλοδοξεί να καταγράψει και να αναλύσει αξιόπιστα δεδομένα που σχετίζονται με τα μέλη Δ.Ε.Π των Τμημάτων Πληροφορικής των ελληνικών Πανεπιστημίων με σκοπό τη συγκριτική αξιολόγηση αυτών, τόσο σε ατομικό επίπεδο όσο και σε επίπεδο Τμήματος ή/και Πανεπιστημίου. Η έρευνα που πραγματοποιήθηκε βασίστηκε σε γνωστούς επιστημονικούς δείκτες που χρησιμοποιούνται στον κλάδο της βιβλιομετρικής ανάλυσης, οι οποίοι μετρούν τόσο την παραγωγικότητα όσο και την απήχηση των δημοσιεύσεων ενός επιστήμονα ή ενός ακαδημαϊκού ιδρύματος. Πιο συγκεκριμένα οι δείκτες που χρησιμοποιήθηκαν ήταν ο δείκτης h_1 -index σε μη ατομικό επίπεδο καθώς και οι δείκτες h -index και p -index σε ατομικό επίπεδο. Άλλοι ενδιαφέροντες δείκτες που καταγράφηκαν και χρησιμοποιήθηκαν στην εν λόγω έρευνα είναι ο αριθμός των δημοσιεύσεων, ο αριθμός των αναφορών και ο αριθμός των αναφορών ανά δημοσίευση.

Σε μη ατομικό επίπεδο, παρατηρήθηκε ότι ο δείκτης h_1 ενός τμήματος σχετίζεται ισχυρά και θετικά με τον χρόνο λειτουργίας και το μέγεθος (υπολογισμένο σε τμήματα) του Πανεπιστημίου που ανήκει το εκάστοτε τμήμα. Αυτό φαίνεται να είναι λογικό, καθώς όσο μεγαλύτερο αριθμό τμημάτων διαθέτει ένα Πανεπιστήμιο και όσο περισσότερα χρόνια ύπαρξης έχει, τόσο μεγαλύτερος θα είναι ο αριθμός των δημοσιεύσεων που έχουν πραγματοποιήσει τα μέλη Δ.Ε.Π του και τόσο περισσότερες πιθανόν να είναι οι αναφορές σε αυτές. Συνεπώς, ως επέκταση της παραπάνω διαπίστωσης, θα υπάρχει αύξηση του δείκτη h_1 για το εν λόγω τμήμα αφού ο δείκτης επηρεάζεται τόσο από τον αριθμό των δημοσιεύσεων όσο και από τον αριθμό των αναφορών.

Επιπρόσθετα, ο δείκτης h_1 επηρεάζεται θετικά και ισχυρά από τον αριθμό των μελών Δ.Ε.Π και των μελών διοικητικού προσωπικού που υπηρετούν στο εκάστοτε τμήμα. Επομένως, μπορεί εύκολα να βγει το συμπέρασμα ότι ένα τμήμα που απασχολεί μεγάλο αριθμό μελών Δ.Ε.Π και διοικητικού προσωπικού θα παράγει μεγάλο αριθμό επιστημονικών δημοσιεύσεων και συνεπώς ο δείκτης h_1 θα αυξάνεται θετικά και σημαντικά.

Ωστόσο, ο δείκτης h_1 σχετίζεται ισχυρά και αρνητικά με τον λόγο (αναλογία) του αριθμού των φοιτητών ανά μέλος Δ.Ε.Π. Επομένως ο λόγος αυτός και ο δείκτης h_1 είναι αντιστρόφως ανάλογα ποσά, καθώς όσο περισσότεροι είναι οι φοιτητές ανά μέλος Δ.Ε.Π τόσο μικρότερη τιμή θα λαμβάνει ο δείκτης h_1 . Μια λογική εξήγηση που μπορεί να δοθεί για την αρνητική αυτή σχέση είναι ότι όσους περισσότερους φοιτητές επιβλέπει ένα μέλος Δ.Ε.Π τόσο περισσότερο χρόνο καταναλώνει στην επίβλεψή τους με αποτέλεσμα να μην απομένει αρκετός χρόνος για να αφιερώσει στην προετοιμασία ή/και δημοσίευση επιστημονικών εργασιών. Τέλος, μετά από ανάλυση που πραγματοποιήθηκε για τον αν ο δείκτης h_1 εξαρτάται από την τοποθεσία στην οποία βρίσκεται το τμήμα παρατηρήθηκε ότι η τοποθεσία δεν επηρεάζει σε καμία περίπτωση την τιμή του δείκτη αυτού.

Στη συνέχεια, έγινε προσπάθεια ταξινόμησης των τμημάτων σε δύο και τρεις ομάδες βάσει του δείκτη h_1 εφαρμόζοντας την μέθοδο *k-means Cluster analysis*. Από

την ταξινόμηση που πραγματοποιήθηκε σε 2 ομάδες προέκυψαν οι εξής ομαδοποιήσεις:

- 1^η Ομάδα: Περιλαμβάνει το τμήμα Πληροφορικής & Τηλεπικοινωνιών του ΕΚΠΑ, Πληροφορικής του ΑΠΘ και του ΟΠΑ και Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
- 2^η Ομάδα: Περιλαμβάνει τα τμήματα Πληροφορικής του Πανεπιστημίου Πειραιώς και του Ιονίου Πανεπιστημίου, το τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης, το τμήμα Πληροφορικής & Τηλεματικής του Χαροκόπειου Πανεπιστημίου, το τμήμα Πληροφορικής & Τηλεπικοινωνιών του Πανεπιστημίου Πελοποννήσου, το τμήμα Πληροφορικής με εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας καθώς και το τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας.

Μετά από κατάλληλο στατιστικό έλεγχο, βρέθηκε ότι η 1^η ομάδα διαφέρει στατιστικά σημαντικά από την 2^η και μάλιστα είναι καλύτερη ως προς τον δείκτη h_I . Από την ταξινόμηση που πραγματοποιήθηκε σε 3 ομάδες προέκυψαν οι παρακάτω ομάδες:

- 1^η Ομάδα: Περιλαμβάνει το τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου, το τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης, το τμήμα Πληροφορικής & Τηλεματικής του Χαροκόπειου Πανεπιστημίου, το τμήμα Πληροφορικής με εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας καθώς και το τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας.
- 2^η Ομάδα: Περιλαμβάνει το τμήμα Πληροφορικής & Τηλεπικοινωνιών του ΕΚΠΑ και το τμήμα Πληροφορικής του ΑΠΘ.
- 3^η Ομάδα: Περιλαμβάνει τα τμήματα Πληροφορικής του ΟΠΑ και του Πανεπιστημίου Πειραιώς, το τμήμα Πληροφορικής & Τηλεπικοινωνιών του Πανεπιστημίου Πελοποννήσου και το τμήμα Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.

Μετά από κατάλληλο στατιστικό έλεγχο, βρέθηκε ότι οι 3 ομάδες διαφέρουν στατιστικά σημαντικά μεταξύ τους. Πιο συγκεκριμένα, παρατηρήθηκε ότι η 2^η Ομάδα είναι καλύτερη από την 3^η Ομάδα, η οποία με την σειρά της είναι καλύτερη από την 1^η Ομάδα ως προς τις τιμές που λαμβάνει ο δείκτης h_I . Επομένως, και στις δύο περιπτώσεις παρατηρείται ότι το τμήμα Πληροφορικής & Τηλεπικοινωνιών του ΕΚΠΑ και το τμήμα Πληροφορικής του ΑΠΘ εμφανίζουν τον μεγαλύτερο δείκτη h_I .

Σε ατομικό επίπεδο, μετρήθηκαν ο αριθμός των δημοσιεύσεων, ο αριθμός των αναφορών που έλαβαν οι δημοσιεύσεις στο σύνολο τους, ο αριθμός των αναφορών ανά δημοσίευση καθώς και οι δείκτες h -index και p -index για κάθε μέλος Δ.Ε.Π. Συγκεντρωτικά, παρατηρήθηκε ότι ένα μέλος Δ.Ε.Π έχει δημοσιεύσει κατά μέσο όρο περίπου 72 δημοσιεύσεις ενώ έχει λάβει κατά μέσο όρο 640 αναφορές. Επιπλέον, παρατηρήθηκε ότι ένα μέλος Δ.Ε.Π που ανήκει σε ένα Τμήμα Πληροφορικής έχει κατά μέσο όρο 8 αναφορές ανά δημοσίευση, ενώ αξιοσημείωτο γεγονός είναι ότι το 50% των μελών Δ.Ε.Π έχει το πολύ 6 αναφορές ανά δημοσίευση ενώ το υπόλοιπο 50% έχει τουλάχιστον 6. Όσον αφορά τους δείκτες h -index και p -index, μια αντιπροσωπευτική τιμή που λαμβάνουν τα μέλη Δ.Ε.Π των 11 τμημάτων Πληροφορικής είναι η τιμή 10 για τον δείκτη h -index και 14 για τον p -index. Αυτό πρακτικά σημαίνει ότι ένα μέλος Δ.Ε.Π έχει δημοσιεύσει τουλάχιστον 10

επιστημονικά άρθρα τα οποία έχουν λάβει 10 ή και περισσότερες αναφορές από άλλους ερευνητές.

Στη συνέχεια, πραγματοποιήθηκε σύγκριση τμημάτων μετρώντας τον απόλυτο αριθμό αναφορών και δημοσιεύσεων για κάθε τμήμα χωριστά. Το τμήμα με τη μεγαλύτερη επιστημονική δραστηριότητα αλλά και απήχηση είναι το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ. Αυτό φαίνεται να είναι λογικό καθώς το τμήμα αυτό διαθέτει τα περισσότερα μέλη Δ.Ε.Π (47 μέλη) ανάμεσα στα 11 υπό μελέτη ακαδημαϊκά τμήματα.

Αν και ο απόλυτος αριθμός των αναφορών και των δημοσιεύσεων είναι ένα κριτήριο για την συγκριτική αξιολόγηση των τμημάτων, ωστόσο δεν είναι επαρκές. Με άλλα λόγια, για να εξετασθεί αν ένα τμήμα έχει αξιόλογη επιστημονική δραστηριότητα πρέπει να συγκριθεί με βάση την επιστημονική απήχηση ανά μέλος Δ.Ε.Π που ανήκει σε αυτό. Για τον λόγο αυτό πραγματοποιήθηκε σύγκριση τμημάτων εξετάζοντας και άλλους παράγοντες όπως είναι ο αριθμός των δημοσιεύσεων και των αναφορών ανά μέλος Δ.Ε.Π. Σύμφωνα με τις μετρήσεις αυτές το Τμήμα Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων παρουσιάζει τον μεγαλύτερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π (περίπου 103). Ωστόσο, το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Πελοποννήσου παρουσιάζει τον μεγαλύτερο αριθμό αναφορών ανά μέλος Δ.Ε.Π παρόλο που έχει λιγότερες δημοσιεύσεις ανά μέλος (περίπου 83).

Αξιοσημείωτο είναι το γεγονός ότι το τμήμα με τον μεγαλύτερο αριθμό αναφορών ανά δημοσίευση είναι το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου. Το τμήμα αυτό δεν έχει τον μεγαλύτερο αριθμό αναφορών και δημοσιεύσεων ανά μέλος Δ.Ε.Π, έχει όμως όπως φαίνεται περισσότερο ποιοτικές δημοσιεύσεις στις οποίες αξίζει να αναφερθεί κανείς. Αυτό πρακτικά σημαίνει ότι δεν είναι απαραίτητο το τμήμα με τις περισσότερες δημοσιεύσεις να έχει και τις περισσότερες αναφορές.

Στη συνέχεια από στατιστικές συγκρίσεις που έγιναν μεταξύ των τμημάτων εφαρμόζοντας ανάλυση διακύμανσης παρατηρήθηκε ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των τμημάτων ως προς τις δημοσιεύσεις και ως προς τον αριθμό των αναφορών ανά δημοσίευση ενώ αντίθετα δεν υπάρχουν αξιοσημείωτες διαφορές ως προς τον αριθμό των αναφορών. Επιπλέον, στατιστικά σημαντικές κρίνονται και οι διαφορές ως προς τους δείκτες *h-index* και *p-index*. Ενδεικτικά αναφέρονται οι πιο αξιοσημείωτες διαφορές:

Ως προς τον αριθμό των δημοσιεύσεων

- Το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων παρουσιάζει σημαντικά μεγαλύτερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π συγκριτικά με τα αντίστοιχα Χαροκόπειου Πανεπιστημίου, του ΕΚΠΑ, του Πανεπιστημίου Πειραιώς του ΑΠΘ.
- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο αριθμό δημοσιεύσεων ανά μέλος Δ.Ε.Π συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Θεσσαλίας, του Πανεπιστημίου Μακεδονίας, του ΟΠΑ, του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.

✚ Ως προς τον αριθμό των αναφορών ανά δημοσίευση

- Το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ παρουσιάζει σημαντικά μεγαλύτερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του ΑΠΘ, του Πανεπιστημίου Ιωαννίνων και του Χαροκόπειου Πανεπιστημίου.
- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Πειραιώς, του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου παρουσιάζει σημαντικά μικρότερο αριθμό αναφορών ανά δημοσίευση συγκριτικά με τα αντίστοιχα Τμήματα του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.

✚ Ως προς τον δείκτη *h-index*

- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο δείκτη *h-index* συγκριτικά με τα αντίστοιχα Τμήματα του ΕΚΠΑ, του Πανεπιστημίου Θεσσαλίας, του Πανεπιστημίου Μακεδονίας, του ΟΠΑ, του Πανεπιστημίου Ιωαννίνων, του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου παρουσιάζει σημαντικά μικρότερο δείκτη *h-index* συγκριτικά με τα αντίστοιχα Τμήματα του Πανεπιστημίου Ιωαννίνων, του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.

✚ Ως προς τον δείκτη *p-index*

- Το Τμήμα Πληροφορικής του ΑΠΘ παρουσιάζει σημαντικά μικρότερο δείκτη *p-index* συγκριτικά με τα αντίστοιχα Τμήματα του ΕΚΠΑ, του Πανεπιστημίου Θεσσαλίας, του ΟΠΑ, του Πανεπιστημίου Ιωαννίνων, του Ιονίου και του Πανεπιστημίου Πελοποννήσου.
- Το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου παρουσιάζει σημαντικά μικρότερο δείκτη *p-index* συγκριτικά με τα αντίστοιχα Τμήματα του ΕΚΠΑ, του Ιονίου Πανεπιστημίου και του Πανεπιστημίου Πελοποννήσου.

Τα συμπεράσματα της εν λόγω έρευνας θα μπορούσαν να χρησιμοποιηθούν ώστε να δημιουργηθεί μελλοντικά μια αξιόπιστη κατάταξη Πανεπιστημίων ή Τμημάτων. Η μελέτη αυτή μπορεί να θεωρηθεί ως η βάση για τη δημιουργία μιας κατάταξης Πανεπιστημίων ή Τμημάτων όχι μόνο του κλάδου της Πληροφορικής αλλά και οποιουδήποτε άλλου επιστημονικού κλάδου. Ως κριτήριο κατάταξης για τη διαμόρφωση της λίστας αυτής μπορεί να χρησιμοποιηθεί οποιοσδήποτε βιβλιομετρικός δείκτης που μετράει την επιστημονική απήχηση ενός ακαδημαϊκού ιδρύματος όπως για παράδειγμα ο δείκτης h_1 που χρησιμοποιήθηκε στην μελέτη. Με άλλα λόγια, η μελέτη θα μπορούσε να επεκταθεί περισσότερο μετρώντας και άλλους επιστημονικούς δείκτες που θεωρούν το ακαδημαϊκό ίδρυμα ως μια οντότητα (βλ. κεφάλαιο 1). Συνεπώς, καταγράφοντας και συγκρίνοντας όσο το δυνατόν περισσότερους δείκτες μπορεί να προκύψει μια πιο ολοκληρωμένη κατάταξη ή οποία

δε θα έχει σαν κριτήριο μόνο τον αριθμό των δημοσιεύσεων και των αναφορών αλλά θα προκύπτει ως αποτέλεσμα πολλών παραγόντων.

Μελλοντικά, μια τέτοια κατάταξη θα μπορούσε να αποτελεί σημαντικό εργαλείο για τους βιβλιομετρικούς αναλυτές αλλά πιθανόν και για τους ίδιους τους καθηγητές των Πανεπιστημίων. Το σημαντικότερο όμως όλων είναι ότι θα αποτελεί ένα αξιόπιστο κριτήριο επιλογής από τους μελλοντικούς φοιτητές που επιθυμούν να φοιτήσουν σε τμήμα Πληροφορικής σε προπτυχιακό, μεταπτυχιακό ή και διδακτορικό επίπεδο.

ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Andres, A. (2009). *Measuring academic research: how to undertake a bibliometric study*, Chandos Publishing, Oxford.
- [2] Vinkler, P. (2010). *Evaluation of research by scientometric indicators*, Chandos Publishing, Oxford.
- [3] Chang, C-L, McAleer, M., Oxley, L. (2010). *Article Influence Score = 5YIF divided by 2*. Department of Economics and Finance.
- [4] De Moya-Anegon, F. et al. (2007). *Coverage analysis of Scopus: a journal metric approach*. *Scientometrics*, 73, pg.53-78.
- [5] Garfield E. (1972). *Citation analysis as a tool in journal evaluation*. *Essays of an information Scientist*, Vol 1, pg.527-544.
- [6] Hirsch, J.E. (2005). *An index to quantify an individual's scientific research output*. *Proceedings of the National Academy of Science*, 102, pg.16569-16572.
- [7] Holmes, A., Oppenheim, C. (2001). *Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and information management*. *Information Research*, 6 (2). Available at: <http://InformationR.net/ir/6-2/paper103.html>.
- [8] András Schubert (2007). Successive h-indices. *Scintometrics*, Vol 70, Iss. 1, DOI: 10.1007/s11192-007-0112-x
- [9] Martin, B.R. (1996). *The use of multiple indicators in the assessment of basic research*. *Scientometrics*, 36, pg.343-362.
- [10] Meho, L.I. & Yang, K. (2007). *Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google Scholar*. *Journal of the American Society for Information Science and Technology*, 58, 2105-2125.
- [11] Moed (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*. Vol. 4, Iss. 3, July 2010, Pg.265–277.
- [12] Bergstrom, C. T., West, J. D., Wiseman, M. A. (2008). *"The Eigenfactor™ Metrics"*. *Journal of Neuroscience* **28** (45): 11433–11434. doi: 10.1523/JNEUROSCI.0003-08.2008
- [13] Prathap, G. (2006). *Hirsch-type indices for ranking institutions' scientific research output*. *Current Science*, 91, 1439.
- [14] Prathap, G. (2009a). *Is there a place for a moch h-index?* *Scientometrics* [published online 19 June 2009]. doi: 10.1007/s11192-009-0066-2 43.

- [15] Prathap, G. (2009b). *The 100 most prolific economists using the p-index*. *Scientometrics* [published online 19 June 2009]. doi: 10.1007/s11192-009-0068-0.
- [16] Ruane, F. & Tol, R.S.J. (2008). *Rational (successive) h-indices: An application to economics in the Republic of Ireland*. *Scientometrics*, 75, 395-405.
- [17] Scimago Research Group. (2007). *Description of SCImago Journal Rank indicator*. Available at: <http://www.scimagojr.com/SCImagoJournalRank.pdf>.
- [18] Seglen, P.O. (1998). *Citation rates and journal impact factors are not suitable for evaluation of research*. *Acta Orthopaedica Scandinavica*, 69, 224-229.
- [19] González-Pereira, B., Guerrero-Bote, V.P, Moya-Anegón, F. (2010). *A new approach to the metric of journals' scientific prestige: The SJR indicator*. *Journal of Informetrics*. Volume 4, Issue 3, July 2010, 379–391. doi:10.1016/j.joi.2010.03.002
- [20] Wallin, J.A. (2005). *Bibliometric methods: Pitfalls and possibilities*. *Basic and Clinical Pharmacology and Toxicology*, 97, 261-275.
- [21] Yang, K. & Meho, L.I., (2006). *Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science*. *Proceedings 69th Annual Meeting of the American Society for Information Science and Technology (ASIST)*. Available at: http://eprints.rclis.org/archive/00008121/01/Yang_citation.pdf, Vol. 43.

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- [22] Γναρδέλλης, Χ. (2013). *Ανάλυση δεδομένων με το IBM SPSS Statistics 21*, Εκδόσεις Παπαζήση
- [23] Κατσάνος Χ, Αβούρης Ν. (2008). *Στατιστικές μέθοδοι ανάλυσης πειραματικών δεδομένων συνεργασίας*. Στο Ν. Αβούρης, Χ. Καραγιαννίδης, Β. Κόμης (Επιμέλεια Έκδοσης) *Συνεργατική τεχνολογία, συστήματα, και μοντέλα συνεργασίας για εργασία, μάθηση, κοινότητες πρακτικής και δημιουργία γνώσης* (σελ. 483-516). Εκδόσεις Κλειδάριθμος, Αθήνα, 2008, ISBN: [978-960-461-232-1](https://doi.org/10.1007/978-960-461-232-1).
- [24] Ζωντανός Κ., Κατρανίδης Σ. (2010), *Συγκριτική Αξιολόγηση Ερευνητικού Έργου των Τμημάτων Πληροφορικής των Ελληνικών Πανεπιστημίων*, Πανεπιστήμιο Μακεδονίας, Θεσσαλονίκη.
- [25] Κούτρας Μ., Ευαγγελάρας Χ. (2010). *Ανάλυση Παλινδρόμησης: Θεωρία και Εφαρμογές*. Εκδόσεις Σταμούλη
- [26] Κούτρας, Μ. Β. (2014). *Ανάλυση Παλινδρόμησης και Ανάλυση Διακύμανσης*, Ακαδημαϊκές Σημειώσεις.
- [27] Μπούτσικας Μ. Β. (2004), *Στατιστικά Προγράμματα*. Ακαδημαϊκές Σημειώσεις. Τμήμα Στατ. & Ασφ. Επιστήμης, Πανεπιστήμιο Πειραιώς

- [28] Παπαϊωάννου Π., Λούκας Σ. (2002). *Εισαγωγή στη Στατιστική*, Εκδόσεις Σταμούλη
- [29] Σαχίνη Ε., Μάλλιου Ν., Χούσος Ν. (2012), *Βιβλιομετρική Ανάλυση: Μεθοδολογική Προσέγγιση ΕΚΤ*, Εθνικό Κέντρο Τεκμηρίωσης.
- [30] Τριανταφύλλου, Ι. Σ. (2015). *Στατιστικές μέθοδοι στην Επιστήμη των Υπολογιστών*, Ακαδημαϊκές Σημειώσεις.
- [31] Φουσκάκης, Δ. (2016). *Περιγραφική στατιστική*, Ακαδημαϊκές Σημειώσεις.

ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ

- [32] <https://www.journalmetrics.com/ipp.php>
- [33] <https://www.journalmetrics.com/snip.php>
- [34] <https://www.journalmetrics.com/sjr.php>
- [35] <http://www.eigenfactor.org/about.php>
- [36] <http://statistics.scientist.gr/5.pdf>
- [37] <http://wokinfo.com/essays/impact-factor/>
- [38] <https://www.scopus.com/>
- [39] <https://scholar.google.com/intl/en/scholar/about.html>
- [40] <https://en.wikipedia.org/wiki/PageRank>
- [41] https://en.wikipedia.org/wiki/Impact_factor
- [42] <https://en.wikipedia.org/wiki/H-index>
- [43] <https://scholar.google.com/intl/en/scholar/metrics.html#coverage>

ΕΥΡΕΤΗΡΙΟ ΟΡΩΝ

Google Scholar.....	7
h_1^* index	16
h_1+ index	15
h_1 -index	15
$h_{1\Delta}$ index.....	16
Kolmogorov-Smirnov τεστ.....	28
Raw Impact per Publication (RIP)	8
Scopus	7
t στατιστικό τεστ ανεξάρτητων δειγμάτων	26
t στατιστικό τεστ ενός δείγματος	25
t στατιστικό τεστ εξαρτημένων δειγμάτων	26
Web of Science	7
Άθροισμα Τετραγώνων.....	40
Αλγόριθμος PageRank	12
Ανάλυση διακύμανσης προς έναν παράγοντα	37
Ανάλυση παλινδρόμησης.....	31
Ανεξάρτητες μεταβλητές	31
Αξιολόγηση από ομότιμους (peer-review)	6
Απήχηση ανά δημοσίευση (Impact Per Publication (IPP)).....	9
Βιβλιογραφικές αναφορές.....	5
Βιβλιομετρικές μέθοδοι	5
Βιβλιομετρικής ανάλυσης.....	5
Δείκτες απήχησης	8
Δείκτης Journal Cited Half-life.....	14
Δείκτης αμεσότητας περιοδικού (Journal Immediacy Index).....	13
Δείκτης επιστημονικής ποιότητας h (h-index).....	14
Δείκτης επιστημονικής ποιότητας p (p-index).....	16
Διάγραμμα διασποράς (scatter diagram)	32
Διακριτές ποσοτικές μεταβλητές	22
Διακύμανση.....	23
Διάμεσος	22
Εναλλακτική υπόθεση.....	25
Ενδιάμεσος h (h-median)	18
Εξαρτημένες μεταβλητές	31
Επαγωγική Στατιστική	21
Επικρατούσα Τιμή(Article Influence (AI)).....	22
Επιρροή άρθρου	19
Ευθεία παλινδρόμησης	31
Εύρος.....	23
Ιδιοπαράγοντας (Eigenfactor).....	18
Κανονικοποιημένη βάσει πηγής δεδομένων απήχηση ανά δημοσίευση (Source Normalized Impact per Paper (SNIP)).....	8
Κατηγορικές ποιοτικές μεταβλητές	22
Μέθοδος ανάλυσης αναφορών	6
Μέθοδος ταξινόμησης k-means	57

Μέθοδος των ελαχίστων τετραγώνων (Least Squares Method)	32
Μέσα τετράγωνα.....	40
Μέση Τιμή	22
Μέτρα ασυμμετρίας	24
Μέτρα κεντρικής τάσης	22
Μέτρα κυρτότητας	24
Μέτρα μεταβλητότητας	23
Μετρικές του Μελετητή της Google (Google Scholar Metrics).....	17
Μη παραμετρικοί έλεγχοι υποθέσεων	25
Μηδενική υπόθεση.....	25
Παράγοντας απήχησης (Impact Factor (IF))	12
Παραμετρικοί έλεγχοι υποθέσεων	25
Πενταετής παράγοντας απήχησης (5-Year Impact Factor)	13
Περιγραφική Στατιστική	21
Ποιοτικές μεταβλητές	22
Ποιοτικές μεταβλητές διάταξης	22
Ποσοτικές μεταβλητές	21
Πυρήνας h (h-core)	17
Στατιστικό κριτήριο F*	40
Συνεχείς ποσοτικές μεταβλητές	21
Συντελεστής Μεταβλητότητας	23
Συντελεστής προσδιορισμού.....	33
Συντελεστής συσχέτισης Spearman	44
Τάξη περιοδικού SCImago (SCImago Journal Rank (SJR))	10
Τυπική απόκλιση.....	23
X ² στατιστικό τεστ.....	27

