ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΣ ΑΠΑΡΑΜΕΤΡΙΚΩΝ ΜΕΘΟΔΩΝ ΕΛΕΓΧΟΥ ΑΝΑΛΥΣΕΩΣ ΔΙΑΚΥΜΑΝΣΕΩΣ ΚΑΤΑ ΕΝ ΚΡΙΤΗΡΙΟΝ*

ΤΟΥ κ. ΙΩΑΝΝΟΥ Γ. ΧΑΛΙΚΙΑ, M. Sc.

I. Εἰσαγωγὴ

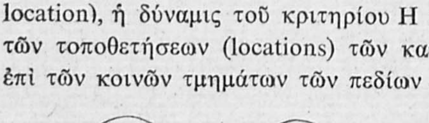
Ὁ προσδιορισμὸς τῆς δυνάμεως ἑνὸς κριτηρίου ἐλέγχου, ἢ ἄλλως τῆς ἀκριβοῦς μορφῆς τῆς δυναμοσυναρτήσεως τούτου, προϋποθέτει τὴν γνῶσιν τῆς κατανομῆς τοῦ πληθυσμοῦ, ἐκ τοῦ ὁποίου λαμβάνεται τὸ δεῖγμα. Εἰς τὰ ἀπαραμετρικὰ καὶ ἐλεύθερα κατανομῶν κριτήρια ἐλέγχων, τοῦτο δὲν εἶναι πάντοτε δυνατὸν διότι ἡ μορφὴ τῶν γεννητόρων πληθυσμῶν εἶναι ἄγνωστος. Εἰς τὴν περίπτωσιν ταύτην μόνον ὡρισμέναι προσεγγίσεις εἶναι δυναταί, ὅταν τὸ μέγεθος τοῦ δείγματος εἶναι ἀρκετὰ μεγάλο, ἐνῶ εἰς τὴν περίπτωσιν τῶν ὀλιγοπληθῶν δειγμάτων ὁ ἐρευνητὴς καταφεύγει εἰς τὰς ἐμπειρικὰς ἐκτιμήσεις (simulation methods). Τοιούτου εἴδους ἐκτιμήσεις παρατίθενται κατωτέρω διὰ τὰς δυνάμεις τριῶν διαφορετικῶν κριτηρίων, τὰ ὁποῖα χρησιμοποιοῦνται πρὸς τὸν ἔλεγχον τῆς ὑποθέσεως μηδὲν ὅτι c εἰς ἀριθμὸν πληθυσμοί, ἐκ τῶν ὁποίων ἐκλέγονται ἀντιστοίχως c τυχαῖα δείγματα, εἶναι ταυτόσημοι. Εἰς τὴν πρᾶξιν ὁ ἀνωτέρω ἔλεγχος εἶναι ἰσοδύναμος μὲ τὸν ἔλεγχον τῆς ὑποθέσεως ὅτι οἱ μέσοι τῶν δειγματιζομένων πληθυσμῶν εἶναι ἴσοι, ἔναντι τῆς ἐναλλακτικῆς ὑποθέσεως ὅτι ἕνας τουλάχιστον μέσος διαφέρει ἀπὸ τοὺς ἄλλους.

Τὰ συγκρινόμενα κατωτέρω, ὡς πρὸς τὰς ἐκτιμωμένας δυνάμεις των, κριτήρια ἐλέγχων εἶναι τὸ κριτήριον τῆς κατανομῆς F, τὸ κριτήριον H τῶν Kruskal καὶ Wallis καὶ τὸ κριτήριον τῆς Διαμέσου. Πλὴν τοῦ κριτηρίου F τὸ ὁποῖον τυγχάνει παραμετρικὸν κριτήριον ἐλέγχου, ὑπὸ τὴν προϋπόθεσιν τῆς κανονικότητος τῶν γεννητόρων πληθυσμῶν, τὰ ἕτερα δύο κριτήρια τυγχάνουν ἀπαραμετρικὰ (καὶ ἐλεύθερα κατανομῶν). Αἱ ἐκτιμήσεις τῶν δυνάμεων τῶν τριῶν ἀνωτέρω κριτηρίων ἐπιτυγχάνονται βάσει ὀλιγοπληθῶν δειγμάτων λαμβανομένων ἀπὸ πληθυσμοὺς διαφορετικῆς ἑκάστοτε μορφῆς καὶ ὑπὸ διαφορε-

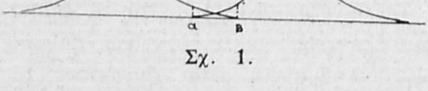
* Ἡ παροῦσα μελέτη βασίζεται εἰς ὑποβληθεῖσαν τὸν Μάϊον τοῦ 1973 ἐργασίαν εἰς London School of Ecomomics πρὸς ἀπόκτησιν πτυχίου M. Sc. εἰς τὴν Στατιστικήν.

τικὴν ἐναλλακτικὴν ὑπόθεσιν. Ἡ δύναμις τοῦ κριτηρίου F ἀκολουθεῖ τὴν μὴ κεντρικὴν κατανομὴν F (non - central F distribution) μόνον ὅταν οἱ πληθυσμοὶ εἶναι κανονικοί, ἐνῶ ἡ δύναμις τοῦ κριτηρίου τῆς διαμέσου διὰ μεγάλα δείγματα προσεγγίζει τὴν μὴ κεντρικὴν κατανομὴν X^2 (non - central chi square distribution) καὶ ἐκτιμᾶται βάσει τῆς μεθόδου τῶν Meng καὶ Chapman.

Τὸ κριτήριον H ἐπροτάθη ὑπὸ τῶν Kruskal καὶ Wallis τὸ 1952 καὶ ἀποτελεῖ ἐπέκτασιν τοῦ κριτηρίου Mann - Whitney εἰς τὴν περίπτωσιν τῶν περισσοτέρων τῶν δύο ἀνεξαρτήτων δειγμάτων. Βασίζεται εἰς τὰς τάξεις (ranks) τῶν ἀρχικῶν παρατηρήσεων καὶ ἔχει τὸ πλεονέκτημα ὅτι ἀπαιτεῖ μόνον πολὺ γενικὰς ὑποθέσεις διὰ τοὺς γεννήτορας πληθυσμούς. Ἀντιθέτως τὸ κριτήριον τῆς κατανομῆς F προϋποθέτει τὴν κανονικότητα τῶν πληθυσμῶν. Αἱ συνήθεις ὑποθέσεις, αἱ ὁποῖαι ἐπιτρέπουν τὴν χρῆσιν τῶν τάξεων γενικῶς, εἶναι ὅτι οἱ δειγματιζόμενοι πληθυσμοὶ εἶναι τῆς αὐτῆς περίπου μορφῆς. Ὅταν οἱ πληθυσμοὶ ἔχουν κατανομὰς τῆς αὐτῆς μορφῆς ἀλλὰ διαφορετικοῦ ἐντοπισμοῦ (shift location), ἡ δύναμις τοῦ κριτηρίου H ἐξαρτᾶται ἀπὸ τὴν πραγματικὴν διαφορὰν τῶν τοποθετήσεων (locations) τῶν κατανομῶν καὶ ἀπὸ τὴν πυκνότητα τούτων ἐπὶ τῶν κοινῶν τμημάτων τῶν πεδίων ὁρισμοῦ των. Διὰ δύο π.χ. πληθυσμούς, ὅσον μικροτέρα εἶναι ἡ πυκνότης τῆς μιᾶς ἢ καὶ τῶν δύο κατανομῶν εἰς τὸ κοινὸν τμῆμα (α, β), τόσον μεγαλυτέρα εἶναι ἡ δύναμις τοῦ κριτηρίου H, ὡς ἐνδεικτικῶς φαίνεται εἰς τὰ σχήματα 1 καὶ 2. Τοῦτο ὀφείλεται εἰς τὸ γεγονὸς ὅτι εἶναι μικρὰ ἡ πιθανότης νὰ περιληφθοῦν εἰς τὸ δεῖγμα παρατηρήσεις ἐκ τῆς πρώτης κατανομῆς, τῶν ὁποίων αἱ τάξεις νὰ εἶναι μεγαλύτεραι τῶν ἀντιστοίχων τοιούτων τῶν παρατηρήσεων ἐκ τῆς δευτέρας κατανομῆς.



Σχ. 1.



Σχ. 2.

Ἐὰν δὲν ὑπάρχουν δεσμοὶ μεταξὺ τῶν παρατηρήσεων, τὸ κριτήριον H ὑπολογίζεται ἐκ τῶν δεδομένων τοῦ δείγματος διὰ τοῦ τύπου:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{c} \frac{S_i^2}{n_i} - 3\,(n+1), \text{ ὅπου}$$

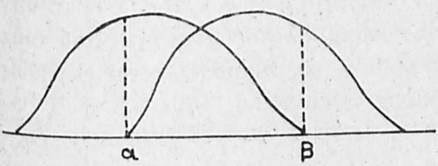
c = ὁ ἀριθμὸς τῶν δειγμάτων,

n_i = τὸ πλῆθος τῶν παρατηρήσεων εἰς τὸ i δεῖγμα,

$n = \sum_{i=1}^{c} n_i$ καὶ

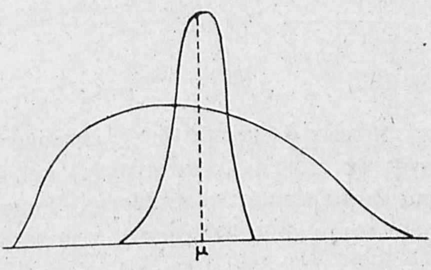
S_i = τὸ ἄθροισμα τῶν τάξεων τῶν παρατηρήσεων τοῦ i δείγματος.

Ἑπομένως εἰς τὴν ἀνωτέρω περίπτωσιν τὸ κριτήριον H θὰ λαμβάνῃ ὡς ἐπὶ τὸ πλεῖστον σημαντικὰς τιμὰς καὶ ἡ δύναμίς του θὰ πλησιάζῃ τὴν μονάδα. Τὸ ἀντίθετον συμβαίνει εἰς τὴν περίπτωσιν τοῦ σχήματος 3 ὅπου αἱ πυκνότητες ἀμφοτέρων τῶν κατανομῶν εἶναι μεγάλαι εἰς τὸ κοινὸν τμῆμα (α, β) τῶν πεδίων ὁρισμοῦ των. Ἡ πιθανότης λήψεως μιᾶς παρατηρήσεως ἐκ τῆς πρώτης κατανομῆς, ἐχούσης μεγαλυτέραν τιμὴν ἀπὸ μίαν ἀντίστοιχον παρατήρησιν ἐκ τῆς δευτέρας κατανομῆς, εἶναι μεγάλη. Ὡς ἐκ τούτου τὸ κριτήριον H θὰ λαμβάνῃ συχνάκις μὴ σημαντικὰς τιμάς.



Σχ. 3.

Ὅταν οἱ πληθυσμοὶ ἔχουν διαφορετικὴν μεταβλητικότητα, ὁπότε δὲν δυνάμεθα νὰ ὑποθέσωμεν ὅτι ἔχουν τὴν αὐτὴν διακύμανσιν, τὸ μόνον συμπέρασμα ἀπὸ μίαν σημαντικὴν τιμὴν τοῦ κριτηρίου H εἶναι ὅτι οἱ πληθυσμοὶ διαφέρουν. Δὲν δυνάμεθα ἀπαραιτήτως νὰ συμπεράνωμεν ὅτι αἱ μέσαι τιμαί των διαφέρουν, ὡς δύναται νὰ δειχθῇ εἰς τὸ σχῆμα 4. Ἑπομένως εἰς τὴν περίπτωσιν αὐτὴν τοῦ εἰς διαφορετικὴν κλίμακα ἐντοπισμοῦ τῶν πληθυσμῶν (scale location), ἡ δύναμις τοῦ κριτηρίου H εἶναι μικρὰ προκειμένου νὰ ἐλέγξωμεν ἐὰν οἱ μέσοι τῶν πληθυσμῶν διαφέρουν. Παρ' ὅλην ὅμως τὴν ἀμφιβολίαν περὶ τῆς συνεπείας τοῦ κριτηρίου Kruskal - Wallis ὑπὸ τὰς ἀνωτέρω συνθήκας, πιστεύωμεν ὅτι εἰς τὴν πρᾶξιν τὸ κριτήριον H δὲν εἶναι εὐαίσθητον εἰς τὰς διαφορὰς τῆς μεταβλητικότητος καὶ ὡς ἐκ τούτου δύναται νὰ χρησιμοποιηθῇ διὰ τὴν σύγκρισιν τῶν μέσων χωρὶς νὰ ὑποθέσωμεν ὅτι οἱ πληθυσμοὶ ἔχουν ἴσας διακυμάνσεις.



Σχ. 4.

Τὸ κριτήριον τῆς διαμέσου ἐσχεδιάσθη διὰ νὰ ἐλέγχῃ ἐὰν τὰ ληφθέντα δείγματα προέρχωνται ἀπὸ πληθυσμοὺς ἔχοντας τὴν αὐτὴν διάμεσον, ἢ ἐὰν οἱ πληθυσμοὶ εἶναι ἀπολύτως ταυτόσημοι ὡς πρὸς τὰς κατανομάς των. Εἰς τὴν πρᾶξιν ὁ ἀνωτέρω ἔλεγχος δύναται νὰ ἑρμηνευθῇ ὡς ἔλεγχος τῆς ἰσότητος τῶν μέσων. Τὸ κριτήριον τῆς διαμέσου ὑπολογίζεται ἐκ τῶν δεδομένων τοῦ δείγματος διὰ τοῦ τύπου:

$$T = \frac{n-1}{\alpha\,(n-\alpha)} \sum_{i=1}^{c} \frac{(n\alpha_i - n_i\alpha)^2}{n n_i}.$$ ὅπου,

$c =$ ὁ ἀριθμὸς τῶν δειγμάτων,

$n_i =$ τὸ πλῆθος τῶν παρατηρήσεων εἰς τὸ i δεῖγμα,

$$n = \sum_{i=1}^{c} n_i,$$

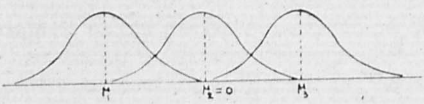
$\alpha_i =$ ὁ ἀριθμὸς τῶν παρατηρήσεων εἰς τὸ i δεῖγμα, αἱ ὁποῖαι ὑπερβαίνουν τὴν κοινὴν διάμεσον (ὑπολογιζομένην δι' ὅλας ὁμοῦ τὰς παρατηρήσεις) καὶ

$$\alpha = \sum_{i=1}^{c} \alpha_i$$

Τὸ κριτήριον T ἀκολουθεῖ κατὰ προσέγγισιν τὴν κατανομὴν X^2 μὲ c - 1 βαθμοὺς ἐλευθερίας. Ἐνῶ τὸ κριτήριον Kruskal - Wallis εἶναι μία συνάρτησις τῶν τάξεων τῶν παρατηρήσεων τοῦ δείγματος, τὸ κριτήριον τῆς διαμέσου ἐξαρτᾶται μόνον ἀπὸ τὴν τοποθέτησιν τῶν παρατηρήσεων ὡς πρὸς τὴν κοινὴν διάμεσον (grand median ἢ combined sample median). Αὐτὸς εἶναι ἕνας ἀπὸ τοὺς λόγους διὰ τοὺς ὁποίους ἡ δύναμις τοῦ κριτηρίου τῆς διαμέσου εἶναι συνήθως μικροτέρα τῆς ἀντιστοίχου δυνάμεως τοῦ κριτηρίου H.

II. Κανονικοὶ πληθυσμοὶ διαφορετικοῦ ἐντοπισμοῦ (shift location)

Ἐξετάζομεν ἐν πρώτοις τὴν συμπεριφορὰν τῶν τριῶν κριτηρίων μὲ κανονικοὺς πληθυσμοὺς τῆς αὐτῆς μεταβλητικότητος (ἔχοντας δηλ. ἴσας διακυμάνσεις), ἀλλὰ μὲ διαφορετικοὺς μέσους ὅρους, ὡς δεικνύεται εἰς τὸ σχῆμα 5. Ἑκάστην φορὰν ἐπιλέγομεν τρία τυχαῖα δείγματα ἐκ πέντε παρατηρήσεων ἕκαστον, ἀπὸ τρεῖς ἀντιστοίχους κανονικοὺς πληθυσμούς. Ἡ ἐπιλογὴ τῶν τυχαίων παρατηρήσεων διεξάγεται δι' ἠλεκτρονικοῦ ὑπολογιστοῦ (πρόγραμμα τυχαίας ἐπιλογῆς ἐκ κανονικῆς κατανομῆς) καὶ ἐπαναλαμβάνεται ἑκατὸ φορὰς διὰ δεκαεπτὰ διαφορετικὰς περιπτώσεις. Ἑκάστη περίπτωσις ἀντιστοιχεῖ εἰς διαφορετικὴν τιμὴν τοῦ ἀθροίσματος τῶν τετραγώνων τῶν ἀποκλίσεων τῶν μέσων τῶν



Σχ. 5.

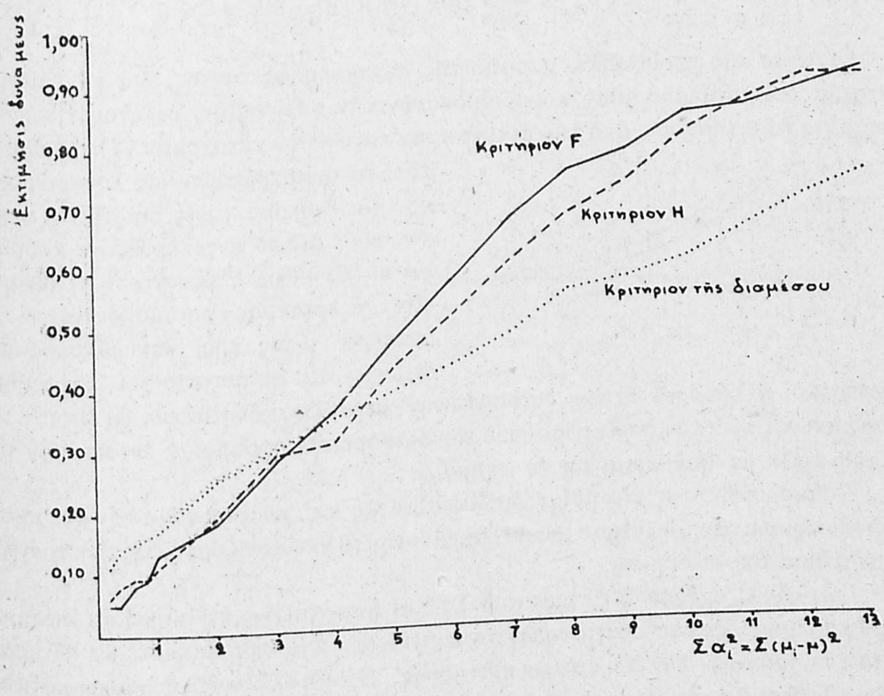
πληθυσμῶν ἀπὸ τὸν κεντρικὸν μέσον μ, ὅπου $\mu = \sum_{i=1}^{c} \mu_i / c$ καὶ $c = 3$. Ἐὰν θέσωμεν $\alpha_i = \mu_i - \mu$, τὸ ἄθροισμα $\sum_{i=1}^{c} \alpha_i^2$ λαμβάνει εἰς ἑκάστην τῶν δεκαεπτὰ περιπτώσεων διαφορετικὴν τιμήν. Αἱ τιμαὶ αὗται εἶναι: 0,2, 0,4, 0,6, 0,8, 1, 2, 3, ...13. Ἐπίσης ὁ μ_2 ἰσοῦται πάντοτε μὲ μηδὲν καὶ οἱ μ_1 καὶ μ_3 λαμβάνουν τιμὰς συμμετρικὰς πέριξ τοῦ μ_2. Ἡ κοινὴ διακύμανσις εἶναι $\sigma^2 = 4$.

Ὁ ἔλεγχος τῆς διαφορᾶς τῶν μέσων βασίζεται καὶ εἰς τὰ τρία κριτήρια, αἱ δὲ ἀντίστοιχοι κριτικαὶ τιμαὶ καὶ τὰ ἐπίπεδα σημαντικότητος παρατίθενται εἰς τὸν κατωτέρω πίνακα καὶ θὰ παραμείνουν τὰ αὐτὰ δι' ὅλα τὰ εἴδη τῶν πληθυσμῶν, οἱ ὁποῖοι θὰ ἐξετασθοῦν περαιτέρω.

Κριτήριον	Κριτικὴ τιμὴ	Ἐπίπεδον σημαντικότητος
Κατανομῆς F	$F_{2.12} = 3{,}88$ (ἀκριβὴς κατανομὴ)	0,05
Kruskal - Wallis *	$H = 5{,}66$ (ἀκριβὴς κατανομὴ)	0,051
Διαμέσου	$X_2^2 = 5{,}991$ (προσεγγιστικὴ κατανομὴ)	0,05

* Ἡ ἀκριβὴς κατανομὴ τοῦ κριτηρίου H ἔχει πινακοποιηθῆ διὰ $c \leqslant 3$ καὶ $n_i \leqslant 5$, $i = 1, 2, 3$.

Μεθ' ἑκάστην ἐκλογὴν τῶν τριῶν τυχαίων δειγμάτων ὑπολογίζομεν τὰς τιμὰς τῶν τριῶν κριτηρίων καὶ τὰς συγκρίνομεν μὲ τὰς ἀντιστοίχους κριτικὰς τιμάς. Ἐφ' ὅσον ἡ ὑπόθεσις μηδὲν τῆς ἰσότητος τῶν μέσων τῶν πληθυσμῶν εἶναι ψευδής, αἱ δυνάμεις τῶν τριῶν κριτηρίων ἐκτιμῶνται ἀπὸ τὰς ἀναλογίας τῶν ὑπολογισθεισῶν τιμῶν των, αἱ ὁποῖαι ὑπερβαίνουν τὰς ἀντιστοίχους κριτικὰς τιμάς. Τὰ προκύπτοντα ἀποτελέσματα ἀπεικονίζονται εἰς τὸ κατωτέρω διάγραμμα.



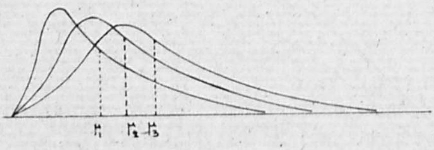
Ὡς προκύπτει, διὰ μικρὰς τιμὰς τοῦ $\Sigma\alpha_i^2$ αἱ δυνάμεις τῶν τριῶν κριτηρίων εἶναι περίπου αἱ αὐταὶ καὶ δὲν ὑφίσταται στατιστικῶς σημαντικὴ διαφορὰ μεταξύ των. Διότι, ἐὰν θεωρήσωμεν τὰς ἀληθεῖς δυνάμεις τῶν τριῶν κριτηρίων ὡς τὰς ἀναλογίας ἑνὸς χαρακτηριστικοῦ A (ἀπόρριψις τῆς ψευδοῦς ὑποθέσεως μηδὲν) εἰς τρεῖς ἀντιστοίχους πληθυσμούς, τότε αἱ ἐκτιμηθεῖσαι δυνάμεις τῶν κριτηρίων λαμβάνονται ὡς ἀναλογίαι δείγματος, καὶ ἑπομένως δυνάμεθα νὰ ἐλέγξωμεν ἐὰν ἡ παρατηρηθεῖσα διαφορὰ μεταξύ των εἶναι στατιστικῶς σημαντική. Διὰ τιμὰς τοῦ $\Sigma\alpha_i^2$ μεγαλυτέρας τοῦ 6 ἡ ἐκτιμηθεῖσα δύναμις τοῦ κριτηρίου τῆς διαμέσου εἶναι σημαντικῶς μικροτέρα τῶν ἐκτιμηθεισῶν δυνάμεων τῶν δύο ἑτέρων κριτηρίων. Ἀντιθέτως, δι' ὅλας τὰς τιμὰς τοῦ $\Sigma\alpha_i^2$ δὲν παρατηρεῖται σημαντικὴ διαφορὰ μεταξὺ τῶν δυνάμεων τῶν κριτηρίων F καὶ H. Παρ' ὅλην ὅμως τὴν φαινομενικὴν ἀνταγωνιστικότητά των, τουλάχιστον διὰ κανονικοὺς γεννήτορας πληθυσμούς, τὸ κριτήριον F εἶναι τὸ πλέον κατάλληλον.

III. Λογαριθμοκανονικοὶ πληθυσμοὶ διαφορετικῆς κλίμακος ἐντοπισμοῦ **(scale location)**

Ἐὰν μία μεταβλητὴ Y κατανέμεται κανονικῶς μὲ μέσον μ καὶ διακύμανσιν σ^2, τότε ἡ μεταβλητὴ $x = e^y$ κατανέμεται λογαριθμοκανονικῶς μὲ μέσον $E(x) = e^{\mu + \frac{1}{2}\sigma^2}$ καὶ διακύμανσιν $\mathrm{var}(x) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. Ἡ συνάρτησις πυκνότητος αὐτῆς εἶναι :

$$dP(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(\ln x - \mu)^2\}\,dx \quad \text{διὰ} \quad x > 0.$$

Λόγω τῆς περιπλόκου μορφῆς τῆς συναρτήσεως ταύτης, διὰ νὰ λάβωμεν τυχαῖα δείγματα ἀπὸ μίαν λογαριθμοκανονικὴν κατανομήν, ἐκλέγομεν πρῶτον τυχαίας παρατηρήσεις ἀπὸ τὴν ἀντίστοιχον κανονικὴν κατανομὴν (Y) τὰς ὁποίας κατόπιν μετατρέπομεν εἰς λογαριθμοκανονικὰς τυχαίας τιμὰς διὰ τῆς σχέσεως $x = e^y$ *. Διὰ τὸ συγκεκριμένον πρόβλημα οἱ γεννήτορες κανονικοὶ πληθυσμοὶ εἶναι οἱ πρότερον χρησιμοποιηθέντες. Ὁ ἀνωτέρω τύπος τῆς διακυμάνσεως μᾶς δεικνύει ὅτι οἱ ἀντίστοιχοι λογαριθμοκανονικοὶ πληθυσμοὶ ἔχουν διαφορετικὴν μεταβλητικότητα καὶ ὡς ἐκ τούτου ἀντιμετωπίζομεν τὴν περίπτωσιν τοῦ εἰς διαφορετικὴν κλίμακα ἐντοπισμοῦ τῶν πληθυσμῶν ὡς δεικνύεται εἰς τὸ σχῆμα 6.



Σχ. 6.

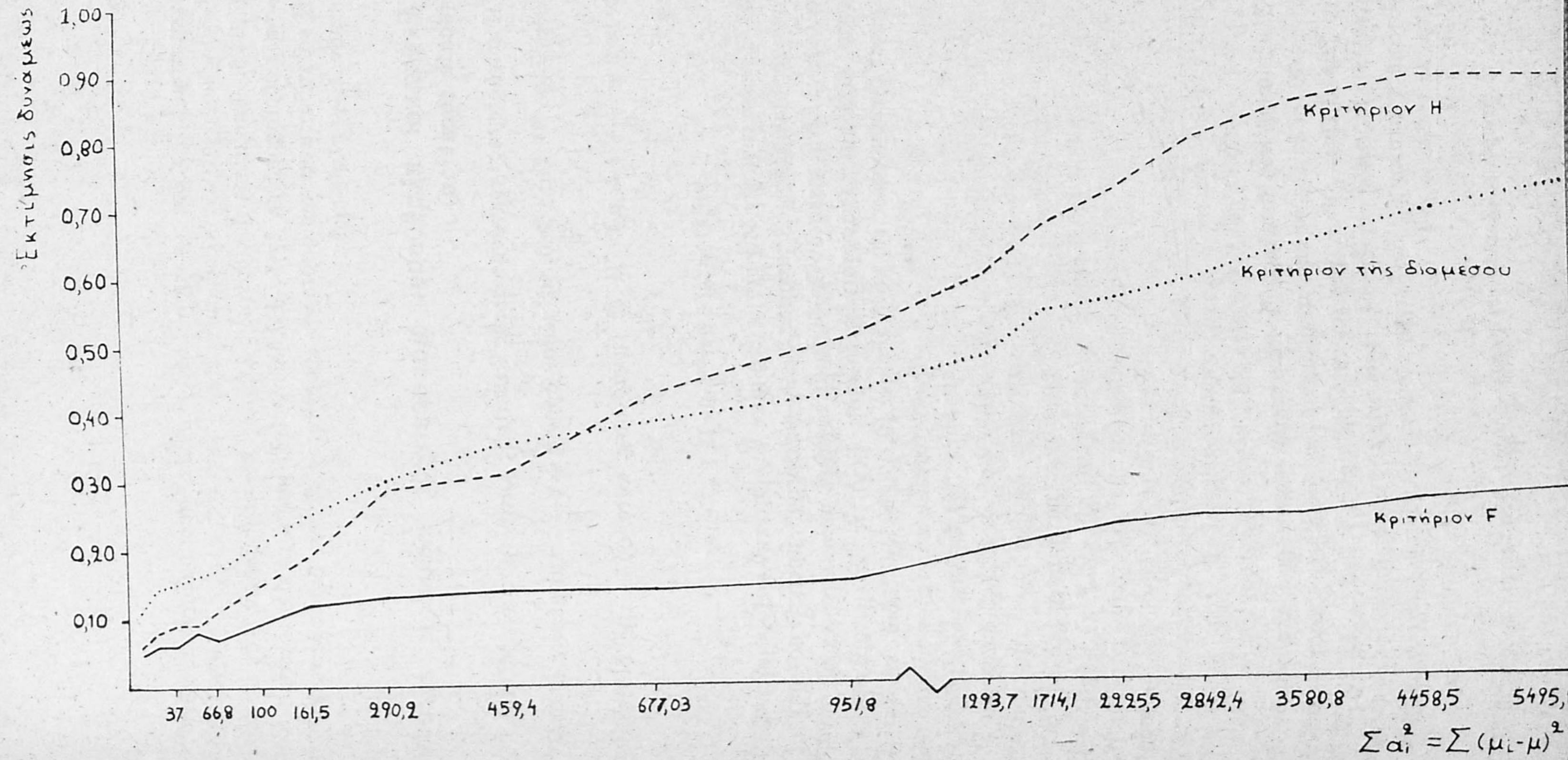
Ἀκολουθοῦντες τὴν ἰδίαν διαδικασίαν ὡς καὶ πρότερον διὰ τὴν ἐκτίμησιν τῶν δυνάμεων τῶν τριῶν κριτηρίων, λαμβάνομεν τὰ ἀπεικονιζόμενα εἰς τὸ κατωτέρω διάγραμμα ἀποτελέσματα.

Δι' ὅλας σχεδὸν τὰς τιμὰς τοῦ $\Sigma\alpha_i^2$ αἱ ἀντιστοίχως ἐκτιμηθεῖσαι δυνάμεις τοῦ κριτηρίου F εἶναι στατιστικῶς σημαντικὰ (εἰς $\alpha = 0{,}05$) μικρότεραι τῶν ἀντιστοίχων τοιούτων τῶν δύο ἑτέρων κριτηρίων. Ἤτοι ἡ ἔλλειψις τῆς κανονικότητος τῶν πληθυσμῶν καὶ ἰδίως ἡ διαφορετικὴ μεταβλητικότης των συνεπάγονται τὴν ἀδυναμίαν τοῦ κριτηρίου F.

Διὰ μεγάλας τιμὰς τοῦ $\Sigma\alpha_i^2$ αἱ προκύψασαι δυνάμεις τοῦ κριτηρίου τῆς διαμέσου εἶναι σημαντικὰ μικρότεραι τῶν ἀντιστοίχων τοιούτων τοῦ κριτηρίου Kruskal - Wallis. Πιθανώτατα, ἡ ὑπεροχὴ αὕτη τοῦ κριτηρίου H ὀφείλεται ἀφ' ἑνὸς μὲν εἰς τὸ ὅτι τοῦτο ἀποτελεῖ πληρεστέραν συνάρτησιν τῶν παρατηρήσεων τοῦ δείγματος, ὡς ἀνεφέρθη προηγουμένως, ἀφ' ἑτέρου δὲ εἰς τὴν λίαν ἀνομοιογενῆ πυκνότητα, τὴν ὁποίαν παρουσιάζουν αἱ δειγματιζόμεναι λογαριθμοκανο-

---

* Εἰς τὴν πραγματικότητα ἡ μέθοδος αὕτη ἐκλογῆς τυχαίων παρατηρήσεων ἀπὸ μίαν λογαριθμοκανονικὴν κατανομὴν δὲν εἶναι ἀμερόληπτος, διότι δίδομεν ἴσην πιθανότητα ἐκλογῆς εἰς ἀμφότερα τὰ μέρη $(0, e^{\mu})$ καὶ (e^{μ}, ∞) τοῦ πεδίου ὁρισμοῦ της $(0, \infty)$, τὰ ὁποῖα ἀντιστοιχοῦν εἰς τὰ δύο μέρη $(-\infty, \mu)$ καὶ (μ, ∞) τοῦ πεδίου ὁρισμοῦ τῆς ἀντιστοίχου κανονικῆς κατανομῆς.


Εκτίμησις δυνάμεως
1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
Κριτήριον H
Κριτήριον τής διαμέσου
Κριτήριον F
37
66,8
100
161,5
290,2
459,4
677,03
951,8
1293,7
1714,1
2225,5
2842,4
3580,8
4458,5
5495,
$\sum a_i^2 = \sum (\mu_i - \mu)^2$

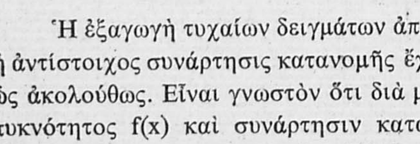
νικαὶ κατανομαί. Π.χ. διὰ τὴν τιμὴν 2.842,4 τοῦ $\Sigma\alpha_i^2$ αἱ διακυμάνσεις τῶν ἀντιστο-
χων λογαριθμοκανονικῶν πληθυσμῶν εἶναι : $\sigma_1^2 = 33,4$, $\sigma_2^2 = 2.926,4$ καὶ $\sigma_3^2 =$ $= 256.184,8$. Ἑπομένως ἡ πιθανότης λήψεως παρατηρήσεων, καὶ ἐκ τῶν τριῶ-
πληθυσμῶν εὑρισκομένων ἐγγὺς ἀλλήλων, εἶναι μικρὰ μὲ ἀποτέλεσμα τὸ κριτή-
ριον H νὰ καθίσταται περισσότερον εὐαίσθητον εἰς τὸν διαφορετικῆς κλίμακο-
ἐντοπισμὸν τῶν πληθυσμῶν.

IV. Ἐκθετικοὶ πληθυσμοὶ διαφορετικῆς κλίμακος ἐντοπισμο- (scale location)

Ἡ συνάρτησις πυκνότητος τῆς ἐκθετικῆς κατανομῆς ἐκφράζεται ὑπὸ τ-
τύπου $f(x) = \alpha e^{-\alpha x}$ διὰ $x \geqslant 0$, μὲ μέσον ὅρον $\mu = \dfrac{1}{\alpha}$ καὶ μέσην ἀπόκλισιν τ-
τραγώνου $\sigma = \dfrac{1}{\alpha}$ $(= \mu)$. Ἡ δὲ συνάρτησις κατανομῆς της δίδεται ὑπὸ τ-
σχέσεως :

$$F(x) = {}_0\!\int^x \alpha e^{-\alpha y}\, dy = 1 - e^{-\alpha x}.$$

Ἡ ἐξαγωγὴ τυχαίων δειγμάτων ἀπὸ μίαν θεωρητικὴν κατανομήν, τῆς ὁπο-
ἡ ἀντίστοιχος συνάρτησις κατανομῆς ἔχει γνωστὴν μαθηματικὴν μορφήν*, γίνε-
ὡς ἀκολούθως. Εἶναι γνωστὸν ὅτι διὰ μίαν τυχαίαν μεταβλητὴν x μὲ συνάρτη-
πυκνότητος f(x) καὶ συνάρτησιν κατανομῆς F(x), ἡ νέα μεταβλητὴ t = F-
κατανέμεται ὁμοιομόρφως εἰς τὸ διάστημα [0,1] Ἑπομένως διὰ νὰ ἐπιλέξω-
μίαν τυχαίαν παρατήρησιν ἀπὸ τὴν -
τανομὴν f(χ), ἐπιλέγομεν πρῶτον ἕνα -
χαῖον ἀριθμὸν ἀπὸ τὸ διάστημα [0,1] -
κατόπιν λαμβάνομεν τὴν ἀντίστοι-
τυχαίαν παρατήρησιν ἀπὸ τὴν ἀντίστ-
φον σχέσιν τῆς $t = F(x)$, τὴν $x = F^{-1}$
ἥτις διὰ τὴν ἐκθετικὴν κατανομὴν εἶ-
$x = -\mu \ln (1 - t)$. Τὴν ἰδίαν μέθοδον -
λογῆς τυχαίων δειγμάτων θὰ χρησι-
ποιήσωμεν καὶ εἰς τοὺς κατωτέρω -
ταζομένους πληθυσμούς.

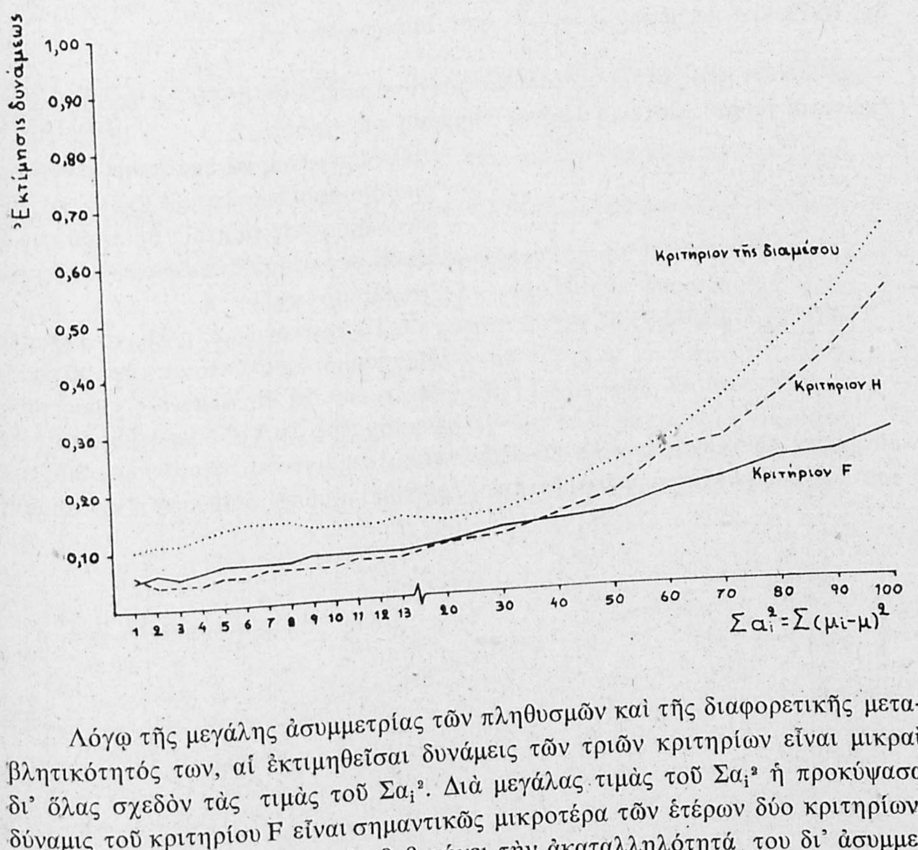


Σχ. 7.

Ὁ μέσος μ_2 τοῦ δευτέρου ἐκ τῶν τριῶν γεννητόρων πληθυσμῶν ἰσοῦ-
πάντοτε μὲ 10, οἱ δὲ ἕτεροι δύο μέσοὶ μ_1 καὶ μ_3 ἰσαπέχουν ἑκάστοτε ἀπὸ τὸ-
ὥστε τὸ $\Sigma\alpha_i^2 = \Sigma(\mu_i - \mu)^2$ νὰ λαμβάνῃ τὰς ἑξῆς τιμάς: 1, 2, 3,. . .13, 20, 30, 40,
100 . Λόγῳ τῆς ἰσότητος μεταξὺ τοῦ μέσου καὶ τῆς μέσης ἀποκλίσεως τετραγώ-
τῆς ἐκθετικῆς κατανομῆς, οἱ δειγματιζόμενοι πληθυσμοὶ ἔχουν διαφορετι-

* Δηλ. τὸ ὡρισμένον ὁλοκλήρωμα $F(x) = \int_0^x f(y)dy$ δύναται νὰ ὑπολογισθῇ.

διακυμάνσεις καὶ ὡς ἐκ τούτου ἀντιμετωπίζομεν καὶ πάλιν τὴν περίπτωσιν τοῦ διαφορετικῆς **κλίμακος** ἐντοπισμοῦ τῶν πληθυσμῶν (σχῆμα 7). Τὰ προκύπτοντα ἀποτελέσματα ἀπεικονίζονται εἰς τὸ κατωτέρω διάγραμμα.



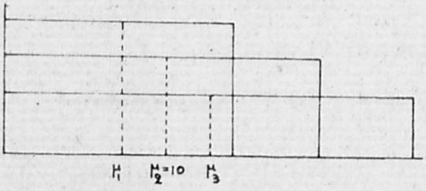
Λόγῳ τῆς μεγάλης ἀσυμμετρίας τῶν πληθυσμῶν καὶ τῆς διαφορετικῆς μεταβλητικότητός των, αἱ ἐκτιμηθεῖσαι δυνάμεις τῶν τριῶν κριτηρίων εἶναι μικραὶ δι' ὅλας σχεδὸν τὰς τιμὰς τοῦ $\Sigma\alpha_i^2$. Διὰ μεγάλας τιμὰς τοῦ $\Sigma\alpha_i^2$ ἡ προκύψασα δύναμις τοῦ κριτηρίου F εἶναι σημαντικῶς μικροτέρα τῶν ἑτέρων δύο κριτηρίων, γεγονὸς τὸ ὁποῖον καὶ πάλιν ἐπιβεβαιώνει τὴν ἀκαταλληλότητά του δι' ἀσυμμετρικοὺς πληθυσμούς.

Ἂν καὶ αἱ ἐκτιμηθεῖσαι δυνάμεις τοῦ κριτηρίου Η δὲν εἶναι σημαντικῶς μικρότεραι τῶν ἀντιστοίχων τοιούτων τοῦ κριτηρίου τῆς διαμέσου (πλὴν τῶν περιπτώσεων τῶν τιμῶν 5, 6, 7 καὶ 8 τοῦ $\Sigma\alpha_i^2$), ὑπάρχει κάποια ἔνδειξις νὰ πιστεύωμεν ὅτι τὸ κριτήριον τῆς διαμέσου ὑπερτερεῖ τοῦ κριτηρίου Kruskal - Wallis διὰ ἐκθετικοὺς πληθυσμοὺς μὲ διαφορετικὴν μεταβλητικότητα. Σημειωτέον ὅτι αἱ διασποραὶ τῶν χρησιμοποιηθέντων πληθυσμῶν δὲν διαφέρουν πολὺ μεταξύ των. Π.χ. διὰ τὴν τιμὴν 70 τοῦ $\Sigma\alpha_i^2$ αἱ ἀντίστοιχοι μέσαι ἀποκλίσεις τετραγώνου τῶν τριῶν πληθυσμῶν εἶναι : $\sigma_1 = 4{,}08$, $\sigma_2 = 10{,}0$ καὶ $\sigma_3 = 15{,}9$. Ὡς ἐκ τούτου διὰ λόγους ἀντιθέτους μὲ ἐκείνους, οἱ ὁποῖοι ἀνεφέρθησαν εἰς τὴν περίπτωσιν τῶν λογαριθμοκανονικῶν κατανομῶν, ἡ πιθανότης λήψεως μιᾶς μὴ σημαντικῆς τιμῆς τοῦ κριτηρίου Η, εἶναι μεγάλη.

V. Ὁμοιόμορφοι πληθυσμοὶ διαφορετικῆς κλίμακος ἐντοπισμοῦ (scale location)

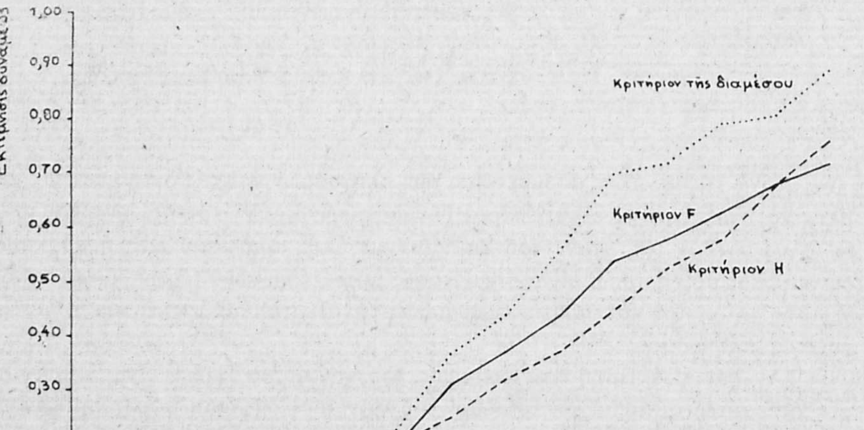
Ἡ συνάρτησις πυκνότητος τῆς ὁμοιομόρφου κατανομῆς εἶναι $f(x) = \frac{1}{\alpha}$ διὰ $0 \leqslant x \leqslant \alpha$, μὲ μέσον $\mu = \frac{\alpha}{2}$ καὶ διακύμανσιν $\sigma^2 = \frac{\alpha^2}{12} = \frac{\mu^2}{3}$.

Ὁ κεντρικὸς μέσος μ_2 ἰσοῦται πάντοτε μὲ 10 ἐνῶ οἱ ἕτεροι δύο ἰσαπέχουν ἑκάστοτε τούτου, ὥστε τὸ $\Sigma\alpha_i^2$ νὰ λαμβάνῃ τὰς τιμὰς 1, 2, 3, . . ., 13, 20, 30, . . ., 100. Ἑπομένως οἱ ἑκάστοτε γεννήτορες ὁμοιόμορφοι πληθυσμοὶ ἔχουν τὸ αὐτὸ κάτω ὅριον εἰς τὰ πεδία ὁρισμοῦ των ἀλλὰ διαφορετικὸν ἄνω ὅριον ὡς δεικνύεται εἰς τὸ σχῆμα 8.



Σχ. 8

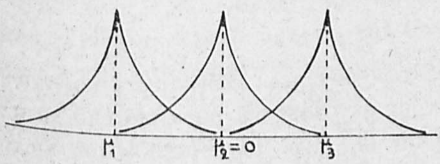
Τὰ ἀπεικονιζόμενα εἰς τὸ κατωτέρω διάγραμμα προκύπτοντα ἀποτελέσματα δεικνύουν ὅτι δι' ὅλας τὰς τιμὰς τοῦ ἀθροίσματος $\Sigma\alpha_i^2$ αἱ διαφοραὶ μεταξὺ τῶν δυνάμεων τῶν κριτηρίων H καὶ F δὲν εἶναι σημαντικαί. Ἀντιθέτως, διὰ τιμὰς τοῦ $\Sigma\alpha_i^2 \geqslant 50$ αἱ ἐκτιμηθεῖσαι δυνάμεις τοῦ κριτηρίου τῆς διαμέσου εἶναι σημαντι-



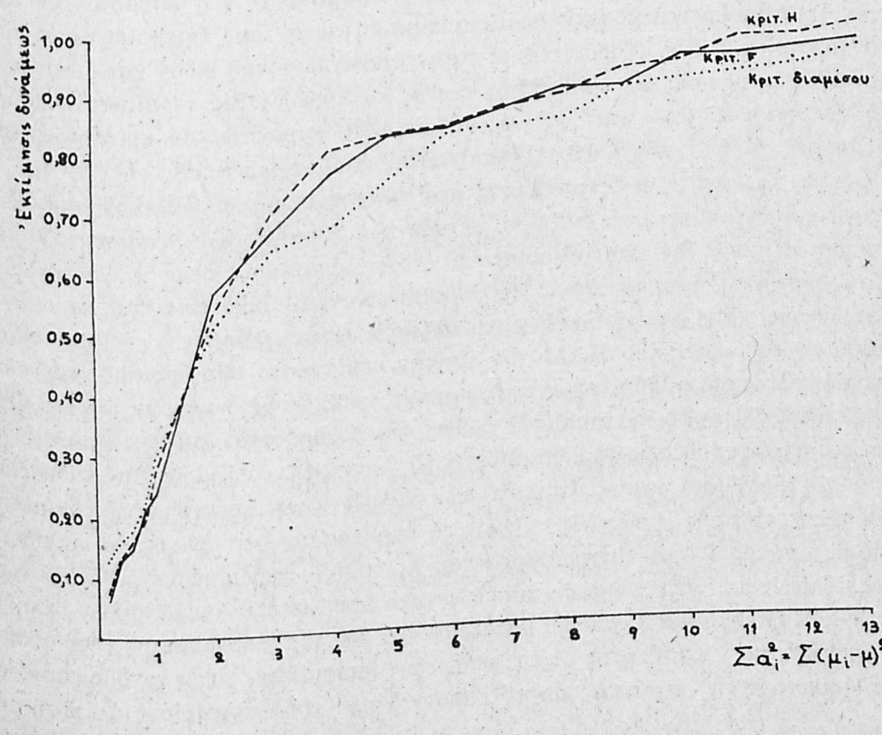
κῶς μεγαλύτεραι τῶν ἀντιστοίχων τοιούτων τοῦ κριτηρίου Kruskal-Wallis. Οἱ προαναφερθέντες εἰς τὴν περίπτωσιν τῶν ἐκθετικῶν πληθυσμῶν λόγοι περὶ τῆς ὑπεροχῆς τοῦ κριτηρίου τῆς διαμέσου ἰσχύουν καὶ διὰ τὴν περίπτωσιν τῶν ὁμοιομόρφων πληθυσμῶν. Π.χ. διὰ τὴν τιμὴν 60 τοῦ $\Sigma\alpha_i^2$ αἱ ἀντίστοιχοι μέσαι ἀποκλίσεις τετραγώνου εἶναι : $\sigma_1 = 2{,}61$, $\sigma_2 = 5{,}77$ καὶ $\sigma_3 = 8{,}93$. Ἡ μικρὰ αὕτη διαφορὰ μεταξὺ τῶν διασπορῶν τῶν πληθυσμῶν σημαίνει ὅτι τὸ κοινὸν τμῆμα τῶν πεδίων ὁρισμοῦ των εἶναι σχετικῶς μέγα μὲ ἀποτέλεσμα τὸ κριτήριον H νὰ καθίσταται ὀλιγώτερον εὐαίσθητον εἰς τοιούτου εἴδους ἐναλλακτικὰς ὑποθέσεις.

VI. Διπλοεκθετικοὶ πληθυσμοὶ διαφορετικοῦ ἐντοπισμοῦ (shift location).

Ἡ συνάρτησις πυκνότητος τῆς τυπικῆς διπλοεκθετικῆς κατανομῆς ἐκφράζεται ὑπὸ τοῦ τύπου $f(x) = \frac{1}{2} e^{-|x|}$ διὰ $-\infty < x < \infty$, μὲ μέσον $\mu = 0$ καὶ διακύμανσιν $\sigma' = 2$. Καὶ οἱ τρεῖς γεννήτορες πληθυσμοὶ ἔχουν τὴν αὐτὴν διπλοεκθετικὴν μορφήν, τὴν αὐτὴν διακύμανσιν, ἀλλὰ διαφορετικὴν τοποθέτησιν ὡς πρὸς τοὺς μέσους των (σχῆμα 9). Ὁ μέσος μ_2 τῆς κεντρικῆς κατανομῆς ἰσοῦται πάντοτε μὲ μηδὲν ἐνῶ οἱ ἕτεροι δύο μέσοι ἰσαπέχουν ἑκάστοτε τούτου, ὥστε τὸ ἄθροισμα $\Sigma(\mu_i - \mu)^2 = \Sigma\alpha_i^2$ νὰ λαμβάνῃ τὰς τιμάς : 0,2, 0,4, 0,6, 0,8, 1, 2, 3, . . ., 13. Αἱ δυνά-



Σχ. 9



μεις τῶν τριῶν κριτηρίων ἐκτιμῶνται, ὡς καὶ πρότερον, ἀπὸ τὰς ἀναλογίας τῶν ὑπολογισθεισῶν τιμῶν τῶν κριτηρίων αἱ ὁποῖαι ὑπερβαίνουν τὰς ἀντιστοίχους κριτικὰς τιμάς. Τὰ προκύπτοντα ἀποτελέσματα ἀπεικονίζονται εἰς τὸ ἀνωτέρω διάγραμμα.

Δι' ὅλας τὰς τιμὰς τοῦ Σa_i^2 αἱ παρατηρούμεναι διαφοραὶ μεταξὺ τῶν ἐκτιμηθεισῶν δυνάμεων τῶν τριῶν κριτηρίων δὲν τυγχάνουν σημαντικαί. Ἡ προκύψασα ἰσοδυναμία μεταξὺ τῶν κριτηρίων F καὶ H δικαιολογεῖται ἀπὸ τὸ γεγονὸς ὅτι οἱ δειγματιζόμενοι διπλοεκθετικοὶ πληθυσμοὶ ἀφ' ἑνὸς μὲν διαφέρουν μόνον εἰς τὴν τοποθέτησιν, ἀφ' ἑτέρου δὲ τυγχάνουν συμμετρικοὶ καὶ λεπτόκυρτοι. Ἐπὶ πλέον ἡ μεγάλη συγκέντρωσις τιμῶν πέριξ τῶν μέσων καθιστᾶ καὶ τὸ κριτήριον τῆς διαμέσου λίαν εὐαίσθητον εἰς ἐναλλακτικὰς ὑποθέσεις περὶ διαφορετικοῦ ἁπλῶς ἐντοπισμοῦ τῶν πληθυσμῶν.

VII. Συμπεράσματα

Ἡ μέχρι τοῦδε ἐμπειρικὴ ἐξέτασις τῶν δυνάμεων τῶν τριῶν κριτηρίων μᾶς ὁδηγεῖ εἰς τὸ συμπέρασμα ὅτι διὰ μικροῦ μεγέθους δείγματα καὶ ὑπὸ καταλλήλους συνθήκας τὸ κριτήριον Kruskal - Wallis εἶναι ἰσχυρότερον τοῦ κριτηρίου τῆς κατανομῆς F. Οἱ Hodges καὶ Lehmann ἀπέδειξαν ὅτι ἐὰν αἱ συναρτήσεις κατανομῶν τῶν πληθυσμῶν εἶναι ταυτόσημοι ὡς πρὸς τὸ σχῆμα των καὶ διαφέρουν μόνον εἰς τὴν τοποθέτησιν (ἐντοπισμόν), τότε ἡ ἀσυμπτωτικὴ σχετικὴ ἀποτελεσματικότης (asymptotic relative efficiency) τοῦ κριτηρίου H ὡς πρὸς τὸ κριτήριον F δὲν λαμβάνει ποτὲ τιμὴν μικροτέραν τοῦ 0,864 καὶ δύναται νὰ ὑπερβῇ τὴν μονάδα δι' ὡρισμένους τύπους κατανομῶν. Εἰς περίπτωσιν καθ' ἣν τὸ κριτήριον F θεωρεῖται ὡς τὸ κατάλληλον παραμετρικὸν κριτήριον, τότε ἡ ἀσυμπτωτικὴ σχετικὴ ἀποτελεσματικότης ἰσοῦται μὲ 0,955.

Εἰς τὰς ἀνωτέρω ἐξετασθείσας περιπτώσεις συμμετρικῶν πληθυσμῶν διαφορετικοῦ ἐντοπισμοῦ, αἱ διαφοραὶ μεταξὺ τῶν ἐκτιμηθεισῶν δυνάμεων τῶν κριτηρίων H καὶ F δὲν εἶναι σημαντικαί. Δι' ἀσυμμετρικοὺς ὅμως πληθυσμοὺς μὲ διαφορετικὴν μεταβλητικότητα (λογαριθμοκανονικοί, ἐκθετικοί), αἱ προκύψασαι δυνάμεις τοῦ κριτηρίου F εἶναι σημαντικῶς μικρότεραι τῶν ἀντιστοίχων τοιούτων τοῦ κριτηρίου H. Μόνον εἰς τὴν περίπτωσιν τῶν ὁμοιομόρφων πληθυσμῶν δὲν παρετηρήθη σημαντικὴ διαφορὰ μεταξὺ τῶν δυνάμεων τῶν ἐν λόγῳ κριτηρίων. Ἤτοι τὸ κριτήριον H καθίσταται ἰσχυρότερον τοῦ κριτηρίου F δι' ἐναλλακτικὰς ὑποθέσεις περὶ διαφορετικῆς κλίμακος ἐντοπισμοῦ τῶν πληθυσμῶν.

Αἱ εὑρεθεῖσαι διαφοραὶ μεταξὺ τῶν δυνάμεων τῶν κριτηρίων τῆς διαμέσου καὶ τῆς κατανομῆς F μᾶς ὁδηγοῦν εἰς τὸ συμπέρασμα ὅτι γενικῶς τὸ κριτήριον τῆς κατανομῆς F εἶναι ἰσχυρότερον τοῦ κριτηρίου τῆς διαμέσου δι' ἐναλλακτικὰς ὑποθέσεις περὶ διαφορετικοῦ ἐντοπισμοῦ μεταξὺ συμμετρικῶν πληθυσμῶν. Οἱ Hodges καὶ Lehmann ἀπέδειξαν ὅτι διὰ μίαν κατανομὴν τῆς ὁποίας ἡ συνάρτησις πυκνότητος ἔχει μεγίστην τεταγμένην εἰς τὴν διάμεσον M, ἡ ἀσυμπτωτικὴ σχετικὴ ἀποτελεσματικότης τοῦ κριτηρίου ἐλέγχου σημείων (τοῦ ὁποίου τὸ κριτήριον τῆς διαμέσου ἀποτελεῖ εἰδικὴν περίπτωσιν) ὡς πρὸς τὸ κριτήριον t τοῦ «Student» (πρὸς τὸ ὁποῖον τὸ κριτήριον F εἶναι ἰσοδύναμον διὰ περισσοτέρους τῶν δύο πληθυσμούς), δὲν λαμβάνει ποτὲ τιμὴν μικροτέραν τοῦ 1/3. Διὰ δὲ κανονικοὺς πληθυσμοὺς ἡ ἀσυμπτωτικὴ σχετικὴ ἀποτελεσματικότης ἰσοῦται μὲ 0,63.

Τουναντίον εἰς ὅλας τὰς περιπτώσεις διαφορετικῆς κλίμακος ἐντοπισμοῦ, εἴτε δι' ἀσυμμετρικοὺς πληθυσμοὺς (λογαριθμοκανονικοί, ἐκθετικοί), εἴτε διὰ συμμετρικοὺς τοιούτους (ὁμοιόμορφοι), αἱ ἐκτιμηθεῖσαι δυνάμεις τοῦ κριτηρίου τῆς διαμέσου εἶναι σημαντικῶς μεγαλύτεραι τῶν ἀντιστοίχων τοιούτων τοῦ κριτηρίου F. Ἤτοι τὸ κριτήριον τῆς κατανομῆς F ὑστερεῖ ἀπὸ ἀπόψεως δυνάμεως τοῦ κριτηρίου τῆς διαμέσου εἰς περιπτώσεις διαφορετικῆς κλίμακος ἐντοπισμοῦ.

Τελικῶς, μόνον εἰς τὴν περίπτωσιν τῶν ὁμοιομόρφων πληθυσμῶν μὲ διαφορετικὴν μεταβλητικότητα, ἡ δύναμις τοῦ κριτηρίου τῆς διαμέσου εὑρέθη σημαντικῶς μεγαλυτέρα τῆς ἀντιστοίχου τοιαύτης τοῦ κριτηρίου Kruskal - Wallis. Εἰς ὅλας τὰς ἄλλας περιπτώσεις (διαφορετικοῦ ἐντοπισμοῦ ἢ διαφορετικῆς κλίμακος ἐντοπισμοῦ) εἴτε αἱ διαφοραὶ μεταξὺ τῶν δυνάμεων τῶν ἐν λόγῳ κριτηρίων εἶναι ἀσήμαντοι, εἴτε ἡ δύναμις τοῦ κριτηρίου H εἶναι σημαντικῶς μεγαλυτέρα τῆς ἀντιστοίχου δυνάμεως τοῦ κριτηρίου τῆς διαμέσου. Ἑπομένως τὸ κριτήριον Kruskal - Wallis καὶ γενικῶς ἡ χρῆσις τῶν τάξεων ἀποτελοῦν ἰσχυρότερον ἀπαραμετρικὸν κριτήριον ἐλέγχου εἰς ἃς περιπτώσεις οὐδὲν δυνάμεθα νὰ ὑποθέσωμεν περὶ τῆς μορφῆς τῶν γεννητόρων πληθυσμῶν.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Conover, W.J., Practical Nonparametric Statistics. J. Wiley and Sons, 1971.
2. Fraser, D.A.S., Nonparametric Methods in Statistics. J. Wiley and Sons, 1966.
3. Johnson, N.I., Continuous Univariate Distributions - 2.
4. Aitckison, and Brown, The Lognormal Distribution.
5. Kendall, M.G. and Stuart A., The Advanced Theory of Statistics, vol. 1, vol. 2, vol. 3. London, Griffin.
6. Kruskal, W.H. and Wallis W.A., Use of Ranks in One-criterion Variance Analysis. Journal of the American Statistical Association, Volume 47, December 1952.
7. Hodges. J.L. Jr. and Lehmann E.L., The Efficiency of some Nonparametric Competitors of the t - Test. The Annals of Mathematical Statistics, Volume 27, 1956.
8. Meng, R.C. and Chapman D.G., The Power of Chi Square Tests for Contingency Tables. Journal of the American Statistical Association, Volume 61, December 1966.