ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Διπλωματική Εργασία

ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΓΙΑ ΤΗΝ ΙΣΟΤΗΤΑ ΤΩΝ ΔΙΑΣΠΟΡΩΝ

ΜΩΡΑΪΤΗΣ ΛΑΜΠΡΟΣ

Επιβλέπων: Κουκουβίνος Χρήστος, Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2011

ΠΡΟΛΟΓΟΣ

Ένα πολύ συνηθισμένο στατιστικό πρόβλημα είναι ο έλεγχος της ισότητας των διασπορών για δυο ή περισσότερους πληθυσμούς. Συχνά θα προτιμούσαμε να γνωρίζουμε αν ένας πληθυσμός έχει μικρότερη διασπορά από τους άλλους. Για παράδειγμα, δυο μέθοδοι παραγωγής ιατρικών γαντιών που χρησιμοποιούνται από γιατρούς μπορούν να συγκριθούν για να δούμε ποια μέθοδος έχει μικρότερη μεταβολή στη δύναμη ή στην ελαστικότητα. Ως δεύτερο παράδειγμα, θεωρούμε ένα ιατρικό πείραμα όπου λάβαμε μετρήσεις εργαστηρίου για να συγκρίνουμε μια ομάδα ασθενών που πάσχουν από μια συγκεκριμένη ασθένεια με μια άλλη ομάδα από υγιείς. Είναι επιθυμητό να διαπιστώσουμε αν η μεταβλητότητα στις μετρήσεις είναι διαφορετική για τα ζητήματα των δυο ομάδων. Η μεταβλητότητα μεταξύ των υγειών περιπτώσεων μπορεί να είναι μεγαλύτερη επειδή μερικές περιπτώσεις μπορεί να είναι υγιέστερες από άλλες, μέσα στην ίδια ομάδα, ή μπορεί να είναι μεγαλύτερη μεταξύ των περιπτώσεων των ασθενών επειδή η ασθένεια ίσως επιδρά σε μερικούς πιο έντονα απ’ ότι σε άλλους.

Άλλη εφαρμογή του ελέγχου της ισότητας των διασπορών βρίσκεται στον έλεγχο της υπόθεσης της κοινής διασποράς δίνοντας ιδιαίτερη σημασία στους παραδοσιακούς ελέγχους συγκρίνοντας την ισότητα των μέσων των πληθυσμών. Για παράδειγμα, στο παραπάνω ιατρικό πείραμα που αναφέρθηκε και περιλαμβάνει ασθενείς και υγιείς, εάν η υπόθεση της κοινής διασποράς απορριφθεί, τότε χρησιμοποιείται ο έλεγχος του Welch ή κάποια άλλη μέθοδος δυο δειγμάτων (two-sample) αντί για τον παραδοσιακό t-έλεγχο.

Στην παρούσα εργασία, περιγράφεται ένα πλήθος μεθόδων με εμβάθυνση σε νέες μεθόδους ανάλυσης της ισότητας των διασπορών. Σκοπός είναι η επιλογή ελέγχων εύρωστων, με μεγάλη ισχύ και σταθερό ρυθμό σφάλματος τύπου I. Το πρώτο κεφάλαιο αποτελεί μια εισαγωγή στο γενικό γραμμικό μοντέλο και την ανάλυση διασποράς για έναν ή και περισσότερους παράγοντες.

Στο δεύτερο κεφάλαιο παρουσιάζουμε πλήθος παραδοσιακών παραμετρικών και μη παραμετρικών ελέγχων που χρησιμοποιούνται για την ισότητα των διασπορών. Αρχικά γίνεται μια αναφορά στις ιδιότητες και στα πλεονεκτήματα των ελέγχων, στις περιπτώσεις που χρησιμοποιούνται, καθώς και μια εισαγωγή στις μεθόδους jackknife και bootstrap. Στη συνέχεια γίνεται μια λεπτομερής περιγραφή για τη μεθοδολογία του κάθε ελέγχου.

Στο τρίτο κεφάλαιο παρουσιάζουμε συγκριτικές μελέτες και προσομοιώσεις βασικών ελέγχων για την ισότητα των διασπορών, από αυτούς που αναφέρθηκαν στο κεφάλαιο 2, και περιγραφή των διαδικασιών που ακολουθούνται. Στη συνέχεια αναλύουμε τα αποτελέσματα των συγκρίσεων και αναφέρουμε τα συμπεράσματα αυτών.

Στο τέταρτο κεφάλαιο περιγράφουμε νέους ελέγχους για την ισότητα των διασπορών και παρουσιάζουμε συγκρίσεις μέσω προσομοιώσεων των νέων με τους πιο βασικούς παραδοσιακούς ελέγχους. Στη συνέχεια αναλύουμε τα αποτελέσματα σχετικά με την απόδοση των νέων ελέγχων.

ABSTRACT

A very common statistical problem is testing equality of variances for two or more populations. Often we wish to know if one population has a smaller variance than others. For example, two methods of producing medical gloves used by doctors might be compared to see which method produces gloves with a smaller variation in their strength or elasticity. As a second example, consider a medical experiment where laboratory measurements are taken to compare a group of patients having a certain disease with another group of healthy subjects. It is desirable to determine if the variation in the measurements is different for the two groups of subjects. Variation among healthy subjects may be larger because some subjects may be healthier than others in the group or variation may be larger among diseased subjects because the disease may affect some patients more adversely than others.

Another application of testing equality of variances lies in testing the assumption of common variance underlying the classical tests comparing the equality of population means. For example, in the above mentioned medical experiment involving patients and healthy subjects if the assumption of common variance is rejected then Welch's t test or some other two-sample method is used instead of the classical t test.

In this work, several methods are described with a thorough examination of new methods in analysis of equality of variances. The goal is the choice of robust tests, with high power and stable Type I error rate. The first chapter is an introduction to the general linear model, one-way analysis of variance or analysis of variance with more factors.

In the second chapter several traditional parametric and nonparametric tests using for testing the equality of variances are presented. Firstly there is a discussion about the properties and the advantages of these tests, the cases these tests are appropriate, as well as an introduction about bootstrap and jackknife methods. Then we give a particular description about the methodology of each test.

In the third chapter comparison studies and simulations about basic tests from chapter two are presented, as well as a description of the procedure we followed. Then we analyze the results of the comparisons and we make some conclusions about them.

In the forth chapter we give a description about new tests for the equality of variances and present some comparisons through simulations between new tests and basic traditional tests. Then we analyze the results related to the performance of new tests.

ΕΥΧΑΡΙΣΤΙΕΣ

Η εκπόνηση και ολοκλήρωση της παρούσας διπλωματικής εργασίας, πραγματοποιήθηκε υπό την επίβλεψη στελεχών από τον χώρο του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα συνεπώς σε αυτό το σημείο να ευχαριστήσω θερμά τον Καθηγητή του Ε.Μ.Π., κ. Χρήστο Κουκουβίνο, για την επίβλεψη και καθοδήγηση του, όπως και για την δυνατότητα που μου προσέφερε να ασχοληθώ με αυτό το θέμα.

Επίσης, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Ανδρουλάκη Εμμανουήλ, για την πολύτιμη βοήθεια και το συνεχές ενδιαφέρον του καθόλη τη διάρκεια της εκπόνησης της εργασίας αυτής.

ΠΕΡΙΕΧΟΜΕΝΑ

Σελίδα

ΚΕΦΑΛΑΙΑ

Κεφάλαιο 1

Εισαγωγή

1.1 Το γενικό γραμμικό μοντέλο παλινδρόμησης

Αρκετές φορές συναντάμε προβλήματα, για τα οποία υπάρχει η υποψία ότι οι τιμές κάποιας μεταβλητής εξαρτώνται από $k \geq 2$ επεξηγηματικές μεταβλητές. Το γενικό γραμμικό μοντέλο, το οποίο περιγράφει αυτή τη σχέση είναι

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \; i = 1, 2, ..., n \qquad (1.1),$$

οπότε και η απόκριση y_i (response) είναι μια γραμμική συνάρτηση των συντελεστών παλινδρόμησης (regression coefficients) β_j, με $j = 1, 2, ..., k$. Ως γνωστόν, έχουμε ότι:

1. y_i, είναι οι τιμές της απόκρισης.
2. x_{ij} είναι οι τιμές των επεξηγηματικών μεταβλητών. Υποθέτουμε, όπως και στο απλό γραμμικό μοντέλο, ότι οι μετρήσεις μας δεν υπόκεινται σε σφάλματα.
3. β_j είναι οι άγνωστες παράμετροι το μοντέλου οι οποίες και πρέπει να εκτιμηθούν.
4. ε_i είναι τα σφάλματα ή υπόλοιπα, τα οποία αποτελούν τυχαίες μεταβλητές και υποθέτουμε ότι ικανοποιούν τις παρακάτω συνθήκες:
   a. $E(\varepsilon_i) = 0 \;\; \forall i$.
   b. $Var(\varepsilon_i) = \sigma^2$, δηλαδή τα σφάλματα ικανοποιούν την υπόθεση της ομοιοσκεδαστικότητας.
   c. $Cov(\varepsilon_i, \varepsilon_j) = 0, \; i \neq j$, δηλαδή τα σφάλματα είναι ασυσχέτιστα.

1.2 Ανάλυση της διασποράς (ANOVA)

Ανάλυση διασποράς είναι η μέθοδος εκείνη με την οποία η ολική μεταβολή που υπάρχει σε ένα σύνολο δεδομένων διαχωρίζεται σε διάφορες συνιστώσες. Σε κάθε μια από αυτές τις συνιστώσες υπάρχει ένα συγκεκριμένο είδος (πηγή) μεταβολής. Έτσι στην ανάλυση είναι δυνατό να διερευνηθεί η σπουδαιότητα (συμβολή) κάθε μιας από τις πηγές μεταβολής στην ολική μεταβολή.

Μια από τις υποθέσεις που απαιτούμε για την προσαρμογή του μοντέλου παλινδρόμησης

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}\,, \tag{1.2}$$

είναι οι μεταβλητές x_i να είναι ποσοτικές. Υπάρχουν όμως αρκετές περιπτώσεις που οι μεταβλητές x_i είναι ποιοτικές ή κατηγορικές. Σε προβλήματα αυτής της μορφής παρ' όλο που η θεωρία της γραμμικής παλινδρόμησης εφαρμόζεται με κατάλληλους μετασχηματισμούς μεταβλητών, εντούτοις η σχετική μέθοδος που κυρίως χρησιμοποιείται είναι η ανάλυση διασποράς (analysis of variance). Δηλαδή η μέθοδος αυτή δεν διαφέρει από την παλινδρόμηση με ανεξάρτητες ποιοτικές μεταβλητές.

Η ανάλυση διασποράς έχει μεγάλη εφαρμογή στην ανάλυση πειραματικών δεδομένων. Η σχέση μεταξύ της θεωρίας "σχεδιασμού πειραμάτων" και "ανάλυσης διασποράς" έγκειται στο ότι όταν γίνεται ο σχεδιασμός πειραμάτων, ο ερευνητής έχει εκ των προτέρων υπόψη του τις μεταβολές εκείνες τις οποίες θεωρεί σημαντικές και έτσι μπορεί να επιλέξει ένα σχεδιασμό για τον οποίο είναι δυνατό να εκφράσει το μέγεθος της συμβολής αυτών των πηγών μεταβολής στην ολική μεταβολή των δεδομένων.

1.2.1 <u>Ανάλυση διασποράς με έναν παράγοντα</u>

Με τον όρο αυτό εννοούμε την ανάλυση προβλημάτων, όπου μια εξαρτημένη παρατηρούμενη ποσοτική μεταβλητή Y, υποτίθεται ότι επηρεάζεται από έναν παράγοντα (factor). Ο παράγοντας είναι μια ποιοτική, ελέγξιμη μεταβλητή, η οποία μπορεί να παίρνει ένα πεπερασμένο πλήθος τιμών που λέγονται στάθμες (levels). Αν είχαμε δύο ή περισσότερους παράγοντες, τότε οι παρατηρήσεις της Y θα γίνονταν για συγκεκριμένους συνδυασμούς από στάθμες των παραγόντων. Κάθε τέτοιος συνδυασμός λέγεται αγωγή (treatment). Στην περίπτωση του ενός παράγοντα οι αγωγές ταυτίζονται με τις στάθμες.

1.2.2 <u>Παραγοντικοί σχεδιασμοί και ανάλυση διασποράς με δύο ή περισσότερους παράγοντες</u>.

Πολλά πειράματα έχουν σαν σκοπό την μελέτη των επιδράσεων δύο ή περισσότερων παραγόντων. Γενικά οι παραγοντικοί σχεδιασμοί είναι πολύ αποδοτικοί για πειράματα αυτού του τύπου. Με τον όρο παραγοντικό σχεδιασμό, εννοούμε ότι σε κάθε πλήρη δοκιμή ή επανάληψη του πειράματος, εξετάζονται όλοι οι δυνατοί συνδυασμοί των επιπέδων (ή σταθμών) των παραγόντων. Για παράδειγμα, αν υπάρχουν a επίπεδα του παράγοντα A και b επίπεδα του παράγοντα B, τότε κάθε επανάληψη περιλαμβάνει όλους τους ab συνδυασμούς των επιπέδων. Όταν οι παράγοντες είναι ταξινομημένοι σε έναν παραγοντικό σχεδιασμό, αυτοί συχνά ονομάζονται εγκάρσιοι ή διασταυρωμένοι (crossed).

Η επίδραση ενός παράγοντα ορίζεται να είναι η αλλαγή που γίνεται στην απόκριση (response), από την αλλαγή στο επίπεδο του παράγοντα. Αυτή συχνά ονομάζεται κύρια επίδραση (main effect), επειδή αναφέρεται στους παράγοντες που είναι πρωταρχικής σημασίας στο πείραμα.

1.2.3 <u>Πολλαπλές συγκρίσεις</u>

Η ανάλυση διασποράς είναι μια μέθοδος που με τη βοήθειά της αποφασίζουμε αν θα δεχθούμε ή θα απορρίψουμε την υπόθεση $H_0 : \mu_1 = \mu_2 = ... = \mu_k$, $k > 2$. Αν η H_0 απορριφθεί σημαίνει ότι υπάρχει ένα τουλάχιστον μ_i διάφορο από τα άλλα, όμως δεν έχουμε καμία πληροφορία για το ποιο ή ποια από τα μ_i είναι διαφορετικά. Απάντηση σε αυτό το ερώτημα μπορεί να μας δώσει ένα πλήθος μεθόδων που ονομάζονται πολλαπλές συγκρίσεις. Οι μέθοδοι των πολλαπλών συγκρίσεων είναι πολλές. Παρακάτω θα παρουσιάσουμε τα κριτήρια που συναντάμε συχνότερα στη βιβλιογραφία, υπενθυμίζοντας ότι αυτά εφαρμόζονται μόνον εφόσον έχει απορριφθεί η υπόθεση $H_0 : \mu_1 = \mu_2 = ... = \mu_k$, στην ανάλυση διασποράς.

1.2.4 Έλεγχοι ομοιογένειας διασπορών

Αν έχουμε k ομάδες παρατηρήσεων και οι διασπορές των σφαλμάτων είναι ίσες μέσα σε κάθε ομάδα, αλλά ίσως διαφέρουν από ομάδα σε ομάδα, τότε εκτιμάμε τη διασπορά σ_j^2 σε κάθε ομάδα με το $s_j^2 = SSE_j \big/ (n_j - 1)$, $j = 1, 2, ..., k$ και κάνουμε έλεγχο ισότητας των διασπορών **:**

H_0**:** $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$

H_1**:** Όχι όλες οι διασπορές ίσες.

Η ανάλυση διασποράς (ANOVA) είναι μια σημαντική τεχνική για να ελεγχθεί η ομοιογένεια των μέσων για αρκετούς πληθυσμούς. Μια από τις βασικές υποθέσεις για την ανάλυση διασποράς (ANOVA) είναι ότι οι διασπορές για πολλούς πληθυσμούς είναι ίσες. Είναι ανάγκη να ελεγχθεί η εγκυρότητα αυτής της υπόθεσης, πριν εφαρμοστεί η τεχνική της ανάλυσης διασποράς (ANOVA). Διάφοροι έλεγχοι έχουν προταθεί για την ομοιογένεια των διασπορών (Conover et al., 1981) και μέχρι τώρα οι πιο συχνές που παρατίθενται είναι οι μέθοδοι που προταθήκαν από τους Bartlett (1937), Cochran (1941) Hartley (1950), Levene (1960) και Box (1953). Όλοι οι έλεγχοι που αναπτυχτήκαν από τους παραπάνω στατιστικούς βασίζονται σε θεωρία μεγάλων δειγμάτων ή αριθμητική προσομοίωση των κρίσιμων σημείων, έτσι ώστε η ισχύς των ελέγχων να βασίζεται στο μέγεθος των δειγμάτων.

Κεφάλαιο 2

Έλεγχοι ισότητας διασπορών

2.1 Εισαγωγή

Οι έλεγχοι για την ομοιογένεια των διασπορών είναι συχνά αντικείμενο ενδιαφέροντος ως προαπαιτούμενα για άλλες αναλύσεις, όπως η ανάλυση διασπορών ή μια συγκέντρωση στοιχείων από διαφορετικές πηγές ώστε να παραχθεί μια βελτιωμένη εκτιμώμενη διασπορά. Στις εργασίες ελέγχου ποιότητας, οι έλεγχοι για την ομοιογένεια των διασπορών είναι συχνά μια χρήσιμη κατάληξη σε μια ανάλυση.

Η κλασσική προσέγγιση στον έλεγχο υποθέσεων αρχίζει συνήθως με τον έλεγχο λόγου πιθανοφάνειας, εάν υποθέσουμε ότι έχουμε κανονικές κατανομές. Εντούτοις, η στατιστική κατανομή του ελέγχου του λόγου πιθανοφάνειας για την ισότητα των διασπορών σε κανονικούς πληθυσμούς εξαρτάται από την κύρτωση της κατανομής (Box 1953), η οποία βοηθά να εξηγήσουμε γιατί εκείνος ο έλεγχος είναι τόσο ευαίσθητος στις παρεκκλίσεις (departures) από την κανονικότητα. Αυτή η μη εύρωστη ιδιότητα (μερικές φορές αποκαλείται «αδύναμη») του ελέγχου του λόγου πιθανοφάνειας έχει οδηγήσει στην κατασκευή πολλών εναλλακτικών ελέγχων των διασπορών. Μερικοί από αυτούς είναι τροποποιήσεις του ελέγχου του λόγου πιθανοφάνειας. Άλλοι είναι προσαρμογές του F-ελέγχου στους ελέγχους των διασπορών και όχι των μέσων. Πολλοί βασίζονται σε μη παραμετρικές μεθόδους, αν και η τροποποίησή τους για την περίπτωση στην οποία οι μέσοι είναι άγνωστοι συχνά καθιστά αυτούς τους ελέγχους εξαρτώμενους από την κατανομή.

Μια μεγάλη ποικιλία από στατιστικούς ελέγχους είναι διαθέσιμη για τη μελέτη της ομοιογένειας των διασπορών για διάφορους πληθυσμούς. Ωστόσο, όταν οι παρατηρήσεις είναι κανονικά κατανεμημένες, χρησιμοποιείται συνήθως ο έλεγχος του Bartlett (μια τροποποίηση του ελέγχου για το λόγο πιθανοφάνειας). Παρόλο που το στατιστικό του ελέγχου του Bartlett ακολουθεί ασυμπτωτικά την κατανομή χ^2, σε μικρά έως μεσαία μεγέθη δειγμάτων, τα πραγματικά επίπεδα αυτού του στατιστικού είναι σχεδόν ίσα με τα ονομαστικά επίπεδα και η ισχύς που εντοπίζει την ετεροσκεδαστικότητα (heteroscedasticity) είναι επίσης αρκετά μεγάλη. Ο Box (1952) επισήμανε ότι για τις μη-κανονικές κατανομές, ο έλεγχος του Bartlett δεν είναι εύρωστος, από τη στιγμή που ο στατιστικός έλεγχος περιλαμβάνει την κύρτωση της κατανομής η οποία ίσως να μην είναι μηδενική. Όταν οι βασικοί πληθυσμοί είναι μη-κανονικοί, το πραγματικό μέγεθος του ελέγχου του Bartlett μπορεί να είναι αρκετές φορές μεγαλύτερο από το ονομαστικό επίπεδο στατιστικής σημαντικότητας.

Σε μεγάλα δείγματα, υπάρχουν μερικοί εύρωστοι και ισχυροί έλεγχοι ικανοί να εξετάσουν την υπόθεση των ίσων διασπορών σε διάφορους μη-κανονικούς πληθυσμούς. Ο Box (1952) πρότεινε τη διαίρεση κάθε δείγματος σε υποδείγματα (περισσότερα από 2) υπολογίζοντας τον λογάριθμο από τις δειγματικές διασπορές σε κάθε ένα από αυτά. Έπειτα πραγματοποιεί ανάλυση διασποράς κατά ένα παράγοντα στις λογαριθμημένες τιμές των διασπορών των υποδειγμάτων. Ο Miller (1968) πρότεινε έναν jackknife έλεγχο για την ομοιογένεια των διασπορών για δυο πληθυσμούς. Συνέστησε να ληφθούν οι ψευδοτιμές σε κάθε ομάδα από τις λογαριθμημένες τιμές των διασπορών με χρήση της μεθόδου jackknife σε μια παρατήρηση την κάθε φορά, και έπειτα την πραγματοποίηση ενός t-ελέγχου δυο δειγμάτων στις ψευδοτιμές. Ο Layard (1973) εισηγήθηκε τον έλεγχο της ισότητας των διασπορών σε διάφορους πληθυσμούς με την πραγματοποίηση ενός F-ελέγχου από την ανάλυση διασποράς κατά ένα παράγοντα στις ψευδοτιμές των ομάδων. Όταν τα μεγέθη των δειγμάτων είναι ίσα σε κάθε ομάδα, ο Levene (1960) πρότεινε την αντικατάσταση κάθε παρατήρησης της ομάδας, με τις απόλυτες αποκλίσεις από τους μέσους τους, και έπειτα την πραγματοποίηση ανάλυσης διασποράς κατά ένα παράγοντα μεταξύ των ομάδων. Οι Brown και Forsythe (1974) πρότειναν την αντικατάσταση της απόκλισης της κάθε παρατήρησης της ομάδας από τη διάμεσο, ή από έναν 10% αποκομμένο μέσο της, αντί για τον μέσο της ομάδας όπως προϋποθέτει ο έλεγχος του Levene.

Υπάρχουν αρκετές μελέτες προσομοίωσης που ανέλαβαν κάποιοι στατιστικοί να εξετάσουν ως προς την απόδοση των παραπάνω ελέγχων. Ο Gartside (1972) πραγματοποίησε συγκρίσεις για οκτώ ελέγχους ως προς την ισχύ και την ευρωστία και συμπέρανε ότι για κανονικά κατανεμημένες παρατηρήσεις το στατιστικό του ελέγχου του Bartlett είναι πολύ ισχυρό και έχει σταθερό σφάλμα τύπου Ι. Όμως, για μη-κανονικά κατανεμημένες παρατηρήσεις, ο έλεγχος βασισμένος σε λογαριθμικό μετασχηματισμό ANOVA (log-ANOVA) είναι εύρωστος αλλά έχει μικρή ισχύ. Ο Layard (1973) πραγματοποίησε συγκρίσεις για τέσσερις ελέγχους και συμπέρανε ότι ο jackknife έλεγχος του Miller και ο χ^2-έλεγχος του Scheffe είναι: πρώτον, εύρωστοι για μικρά δείγματα, δεύτερον, πιο ισχυροί από τον ομαδοποιημένο έλεγχο του Box, και αποδίδουν πανομοιότυπα με τον έλεγχο του Bartlett για κανονικά κατανεμημένες παρατηρήσεις. Ο Hall (1972) εξέτασε έξι ελέγχους και συμπέρανε ότι ο jackknife έλεγχος είναι ο καλύτερος. Οι Keselman, Games και Clinch (1979) πραγματοποίησαν συγκρίσεις για δέκα ελέγχους και συμπέραναν ότι για άνισα μεγέθη δειγμάτων, ο jackknife έλεγχος του Miller έχει ασταθές σφάλμα τύπου Ι. Επιπλέον, τόνισαν ότι οι διαδεδομένοι έλεγχοι για την ομοιογένεια της διασποράς είτε είναι ευαίσθητοι σε παρατηρήσεις μη-κανονικά κατανεμημένες, είτε, αν είναι εύρωστοι, έχουν έλλειψη ισχύος. Οι τέσσερις έλεγχοι που μελετήθηκαν από τον Levy (1978) ήταν όλοι σοβαρά επηρεασμένοι από τις παραβιάσεις της υποκείμενης υπόθεσης της κανονικότητας. Οι Brown και Forsythe (1974) πραγματοποίησαν συγκρίσεις για έξι ελέγχους και συμπέραναν ότι ο έλεγχος του Levene, βασισμένος στις αποκλίσεις από τη διάμεσο ή στον 10% αποκομμένο μέσο, είναι εύρωστος σε συνθήκες μη-κανονικότητας και ότι οι έλεγχοι, του Levene, η jackknife εκδοχή του ελέγχου του Layard και ο χ^2 είναι λιγότερο εύρωστοι σε συνθήκες μη-κανονικότητας. Σε μια εκτενή μελέτη, οι Conover, Johnson και Johnson (1981) πραγματοποίησαν συγκρίσεις για πολλούς παραμετρικούς και μη-παραμετρικούς ελέγχους και τις διασπορές τους (πενήντα έξι έλεγχοι συνολικά) για τον έλεγχο της ισότητας των διασπορών και συμπέραναν ότι μόνο 5 έλεγχοι είναι εύρωστοι. Οι 5 έλεγχοι που επισημάνθηκαν ήταν: πρώτον, μια παραλλαγή του ελέγχου του Bartlett (ο έλεγχος του Bartlett προσαρμοσμένος από μια εκτίμηση της κύρτωσης, λήφθηκε από τη χρήση των διαμέσων αντί για τους μέσους του δείγματος), δεύτερον, ο έλεγχος του Levene βασισμένος στις αποκλίσεις από τη διάμεσο αντί του μέσου (όπως προτάθηκε από τους Brown και Forsythe). Τρίτον, μια τροποποίηση του ελέγχου του Levene βασισμένη στις αποκλίσεις από τη διάμεσο (αντί για αντικατάσταση κάθε παρατήρησης με την απόλυτη απόκλισή της από τη διάμεσο, πρότειναν αντικατάσταση κάθε παρατήρησης με το τετράγωνο των αποκλίσεων από τη διάμεσο) και τέλος, 2 μη-παραμετρικούς ελέγχους. Επιπλέον, συμπέραναν ότι ο έλεγχος του Levene βασισμένος στις αποκλίσεις από τη διάμεσο και οι δυο μη-παραμετρικοί έλεγχοι είναι περισσότερο ισχυροί από τους άλλους δυο.

2.1.1 Εισαγωγή στις μεθόδους Bootstrap και Jackknife

Σ’ αυτό το σημείο θα κάνουμε μια σύντομη εισαγωγή στις μεθόδους jackknife και bootstrap που θα τις συναντήσουμε στα επόμενα κεφάλαια. Οι μη παραμετρικές μέθοδοι jackknife και bootstrap είναι δυο τεχνικές αναδειγματοληψίας που χρησιμοποιούνται ευρέως σήμερα για τον υπολογισμό διάφορων μέτρων ακρίβειας. Τέτοια μέτρα μπορούν να είναι η μεροληψία (bias), η διασπορά (variance) ή το μέσο τετραγωνικό σφάλμα (mean square error). Ο λόγος που καθιστά τις jackknife και bootstrap ιδιαίτερα δημοφιλής σήμερα στους στατιστικούς αναλυτές είναι ότι σε πολύπλοκα στατιστικά προβλήματα η παραδοσιακή προσέγγιση αδυνατεί να εφαρμοστεί και να τα αντιμετωπίσει. Γενικότερα χρησιμοποιούνται όταν είναι δύσκολη η παραμετρική μοντελοποίηση αλλά και η θεωρητική αντιμετώπιση. Αυτές οι δυο μέθοδοι παρουσιάζουν πολλά πλεονεκτήματα σε σχέση με την παραδοσιακή παραμετρική προσέγγιση στατιστικών προβλημάτων αφού μπορούμε εύκολα να τις περιγράψουμε. Εφαρμόζονται σε αυθαίρετες πολύπλοκες καταστάσεις όπου δεν απαιτούνται οι υποθέσεις κανονικής κατανομής. Επίσης, αξίζει να σημειωθεί ότι με τις μεθόδους αυτές επιτυγχάνεται συνήθως σημαντική μείωση της μεροληψίας καθώς και ότι παρόλο που η μέθοδος jackknife προηγείται χρονικά της bootstrap παρουσιάζουν σημαντικές ομοιότητες. Τέλος, με την πρόοδο των ηλεκτρονικών υπολογιστών τις τελευταίες δεκαετίες, οι μέθοδοι αυτές εξελίχθηκαν σε ένα πολύ χρήσιμο εργαλείο στα χέρια των αναλυτών.

Η μέθοδος jackknife προτάθηκε πρώτη φορά από τον Quenouille το 1949 για τον υπολογισμό της μεροληψίας (bias) αλλά πήρε το όνομά της, όπως είναι πλέον γνωστή, από τον Tukey το 1958 ο οποίος την χρησιμοποίησε για τον υπολογισμό της τυπικού σφάλματος (standard error).

Η μέθοδος bootstrap προτάθηκε από τον Efron το 1979 σε μια προσπάθειά του κατασκευάσει διαστήματα εμπιστοσύνης σε περιπτώσεις που οι κλασσικές μέθοδοι δεν μπορούσαν να εφαρμοστούν. Η βασική ιδέα της μεθόδου αυτής είναι ότι κάνει μέσω υπολογιστή ότι θα έκανε ένας πειραματιστής στην πράξη, εάν αυτό ήταν δυνατό, δηλαδή θα επαναλάμβανε αρκετές φορές το πείραμα και οι παρατηρήσεις θα εκχωρούνταν ξανά τυχαία και στη συνέχεια οι εκτιμήσεις θα επαναϋπολογίζονταν. Οι εκχωρήσεις και οι υπολογισμοί αυτοί γίνονται χιλιάδες φορές και αντιμετωπίζονται ως διαρκώς επαναλαμβανόμενα πειράματα.

2.2 Περιγραφή των ελέγχων

Για $i = 1,...,k$, έστω $\{X_{ij}\}$ τυχαίο δείγμα μεγέθους n_i για πληθυσμούς με μέσες τιμές μ_i και διασπορές σ_i^2. Για να εξετάσουμε την υπόθεση των ίσων διασπορών, είναι απαραίτητη μια επιπρόσθετη υπόθεση (Moses 1963). Μια πιθανή υπόθεση είναι ότι τα X_{ij} είναι κανονικά κατανεμημένα. Αυτό οδηγεί σε έναν μεγάλο αριθμό ελέγχων, μερικά από τα οποία έχουν ακριβείς πίνακες, και κάποια άλλα έχουν μόνο ασυμπτωτικές προσεγγίσεις για τις κατανομές των στατιστικών ελέγχων. Μια άλλη πιθανή υπόθεση είναι ότι το X_{ij} κατανέμεται ομοιόμορφα όταν η μηδενική υπόθεση είναι αληθής. Αυτή η υπόθεση μας επιτρέπει διατυπώσουμε διάφορους μη παραμετρικούς ελέγχους. Στην πράξη, καμία υπόθεση δεν είναι εξ ολοκλήρου αληθής, έτσι ώστε όλοι αυτοί οι έλεγχοι των διασπορών είναι μόνο κατά προσέγγιση. Είναι σωστό να εξεταστούν όλοι οι διαθέσιμοι έλεγχοι για την ευρωστία τους στις παραβιάσεις των υποθέσεων. Σε αυτή την ενότητα, παρουσιάζουμε μια λίστα των ελέγχων για τις ίσες διασπορές . Οι περισσότεροι από τους ελέγχους στις παραγράφους 2.2.1 με 2.2.3 είναι βασισμένοι σε κάποια τροποποίηση του στατιστικού ελέγχου λόγου πιθανοφάνειας που προκύπτει βάσει της υπόθεσης της κανονικότητας. Οι έλεγχοι που είναι ουσιαστικά τροποποιήσεις του ελέγχου του λόγου πιθανοφάνειας ή στηρίζονται στην υπόθεση της κανονικότητας, παρουσιάζονται στην παράγραφο 2.2.1. Οι τροποποιήσεις σε αυτούς τους ελέγχους, που υιοθετούν μια εκτίμηση κύρτωσης, εμφανίζονται στην 2.2.2. Δεν έχουν ασυμπτωτική κατανομή για όλους τους αρχικούς πληθυσμούς, ενώ έχουν μόνο δευτερεύοντες περιορισμούς. Η παράγραφος 2.2.3 και ο πίνακας 2.1 παρουσιάζουν τις τροποποιήσεις των μη παραμετρικών ελέγχων. Στην παράγραφο 2.2.5 παρουσιάζονται έλεγχοι με τη χρήση της μεθόδου jackknife. Η τροποποίηση αποτελείται από τη χρήση του δειγματικού μέσου ή της διαμέσου του δείγματος αντί του πληθυσμιακού μέσου κατά τον υπολογισμό του στατιστικού ελέγχου. Μόνο οι μη παραμετρικοί έλεγχοι στην κατηγορία των τάξεων των γραμμικών ελέγχων συμπεριλαμβάνονται εδώ, επειδή αυτή η κατηγορία περιλαμβάνει όλους τους τοπικούς ελέγχους του πιο ισχυρού βαθμού (Hajek και Sidak 1967). Επομένως, στον πίνακα 2.1, παρουσιάζονται μόνο τα αποτελέσματα, $a_{n,i}$ για αυτούς τους ελέγχους. Από αυτά τα αποτελέσματα, μπορούν να πραγματοποιηθούν χ^2 έλεγχοι βάσει του στατιστικού:

$$X^2 = \sum_{i=1}^{k} n_i (\bar{A}_i - \bar{a})^2 \big/ V^2 \,, \tag{2.1}$$

όπου $\bar{A}_i =$ μέσο αποτέλεσμα στο i-οστό δείγμα, $\bar{a} =$ ολικό μέσο αποτέλεσμα $= 1/N \sum_{i=1}^{N} a_{N,i}$, και $V^2 = (1/N-1)\sum_{i=1}^{N}(a_{N,i} - \bar{a})^2$,

το οποίο συγκρίνεται με ποσότητες από την χ^2 κατανομή με $k-1$ βαθμούς ελευθερίας. Εναλλακτικά, το στατιστικό

$$F = \frac{X^2 \big/ (k-1)}{(N-1-X^2)\big/(N-k)}, \tag{2.2}$$

μπορεί να συγκριθεί με τα ποσοστημόρια (quantiles) από την F- κατανομή με ($k-1, N-k$) βαθμούς ελευθερίας. Στις παρακάτω περιγραφές των ελέγχων, αφήνουμε $\bar{X}_i$, $\tilde{X}_i$, και r_i να δείξουν το i-οστό δειγματικό μέσο, διάμεσο και εύρος αντίστοιχα, ενώ το $\bar{X}$ δείχνει το γενικό μέσο όρο. Το i-οστό δείγμα διασποράς με διαιρέτη $n_i - 1$, είναι s_i. Επιπρόσθετα,

$$N = \sum n_i \,, \quad s^2 = \sum (n_i - 1)s_i \big/ (N-k)$$

και

$$F_{ij} = \frac{\sum_i n_i (\bar{X}_i - \bar{X})^2 \big/ (k-1)}{\sum_i \sum_j (X_{ij} - \bar{X}_i)^2 \big/ (N-k)} \quad , \tag{2.3}$$

είναι η συνηθισμένη ανάλυση διασποράς κατά ένα παράγοντα των στατιστικών ελέγχων.

Στους ελέγχους για ίσες διασπορές, ο F-έλεγχος υπολογίζεται σε κάποιο μετασχηματισμό του X_{ij} παρά στο ίδιο το X_{ij} .Οι περιγραφές στους διάφορους ελέγχους παρουσιάζονται τώρα. Η σημείωση *med* αναφέρεται στην αντικατάσταση του $\bar{X}_i$ με το, $\tilde{X}_i$ στον στατιστικό έλεγχου σε μια προσπάθεια να βελτιωθεί η ευρωστία του ελέγχου.

2.2.1 <u>Παραμετρικοί έλεγχοι</u>

2.2.1.1 Ο έλεγχος των Neyman και Pearson

Ο έλεγχος που προτείνεται από τους Neyman και Pearson (1931) είναι ο έλεγχος του λόγου πιθανοφάνειας βάσει της κανονικότητας. Τον συμβολίζουμε με *N-P* με στατιστικό έλεγχο:

$$x_{k-1}^2 = T_1 = N\ln\left(\frac{N-k}{N}s^2\right) - \sum_{i=1}^{k} n_i \ln\left(\frac{n_i-1}{n_i}s_i^2\right). \tag{2.4}$$

Εξετάζουμε, επίσης την τροποποίηση *N-P:med*.

2.2.1.2 Ο έλεγχος του Bartlett (Bartlett's Test)

Ο Bartlett (1937) τροποποίησε το *N-P* για να διορθώσει τη μεροληψία. Ο έλεγχος που προκύπτει είναι πιθανότατα ο πιο συνηθισμένος που χρησιμοποιείται για την ισότητα των διασπορών. Είναι γνωστό πως είναι ευαίσθητος στις αποκλίσεις από την κανονικότητα. Μελέτες από τους Glaser (1976), Chao και Glaser (1978), και τους Dyer και Keating (1980) δίνουν τις μεθόδους εύρεσης της ακριβούς κατανομής για τον στατιστικό έλεγχο. Τον συμβολίζουμε με *Bar*. Εξετάζουμε επίσης την τροποποίηση *Bar:med.*

Ο έλεγχος αυτός χρησιμοποιείται σε όλες τις περιπτώσεις με ίσα ή άνισα δείγματα και υποθέτει ότι τα σφάλματα ακολουθούν την κανονική κατανομή. Το στατιστικό που χρησιμοποιείται είναι :

$$B_1 = \frac{\left(\sum_{i=1}^{k} f_i\right)\ln s^2 - \sum_{i=1}^{k}\left(f_i \ln s_i^2\right)}{C}, \tag{2.5}$$

όπου $f_i = n_i - 1,\ s^2 = \dfrac{\sum_{i=1}^{k} f_i s_i^2}{\sum_{i=1}^{k} f_i},\ C = 1 + \dfrac{\sum_{i=1}^{k} f_i^{-1} - \left(\sum_{i=1}^{k} f_i\right)^{-1}}{3(k-1)}$ με $k \ge 2$

Όταν η H_0 είναι σωστή, το $T_1 \sim X_{k-1}^2$ και η προσέγγιση είναι ικανοποιητική για $f_i \ge 3$. Η H_0 απορρίπτεται σε στάθμη a, όταν $T_1 > X_{k-1,a}^2$.

2.2.1.3 Ο έλεγχος του Cochran (Cochran's Test)

Ο έλεγχος που παρουσιάστηκε από τον Cochran (1941) ήταν αρκετά ευκολότερος στον υπολογισμό σε σχέση με τους μέχρι τότε ελέγχους. Ωστόσο, με τους σημερινούς υπολογιστές η διαφορά στο χρόνο υπολογισμού είναι ελάχιστη. Τον συμβολίζουμε με *Coch*. Εξετάζουμε επίσης το *Coch: med*.

Όταν οι βαθμοί ελευθερίας των s_i^2 είναι ίσοι με $f_1 = f_2 = ... = f_k = v$, χρησιμοποιείται το στατιστικό :

$$T_2 = \frac{\max s_i^2}{\sum_i s_i^2}. \qquad (2.6)$$

βλέπε Pearson & Hartley (1970), p.203 για ειδικούς πίνακες.

Αυτός ο έλεγχος είναι ευαίσθητος και προτιμάται στην περίπτωση που όλες οι διασπορές είναι ίσες εκτός από μια. Όταν ισχύει η H_0 έχει μελετηθεί η κατανομή του στατιστικού T_2 και υπάρχουν πίνακες κρίσιμων τιμών του T_2 για $a = 0.05$, $a = 0.01$.

2.2.1.4 Ο έλεγχος των Bartlett και Kendal

Μια άλλη προσπάθεια να απλοποιηθούν οι υπολογισμοί οδήγησε σε αυτόν τον έλεγχο από τους Bartlett και Kendall (1946), το οποίο στηρίζεται στο γεγονός ότι το $\ln s^2$ είναι περίπου κανονικό, και χρησιμοποιεί τους πίνακες για το ομαλοποιημένο εύρος σε κανονικά δείγματα. Τον συμβολίζουμε με *B-K* με στατιστικό έλεγχο:

$$T_3 = \frac{\ln(\max s_i^2) - \ln(\min s_i^2)}{(n/2)^{1/2}}. \qquad (2.7)$$

βλέπε Pearson & Hartley (1970), p.177 για ειδικούς πίνακες ($n =$μέσο μέγεθος δείγματος)

Δεν εξετάζουμε αυτόν τον έλεγχο λόγω της ισοδυναμίας του με τον ακόλουθο έλεγχο. (Hartley)

2.2.1.5 Ο F-μέγιστος έλεγχος του ***Hartley*** **(** ***Hartley's F-max Test*****)**

Τέσσερα έτη μετά το *B-K,* αυτός ο έλεγχος παρουσιάστηκε από τον Hartley (1950). Γνωστός ως F μέγιστος ($F_{\max}$) έλεγχος, είναι κυρίως ένας εκθετικός μετασχηματισμός του *B-K*. Ένα πλεονέκτημα αυτού του ελέγχου είναι οι ακριβείς πίνακες που είναι διαθέσιμοι για τα ίσα μεγέθη δειγμάτων (David 1952). Τον συμβολίζουμε με *Hart*. Εξετάζουμε επίσης το *Hart :med.*

Συμφώνα με τον F-μέγιστο ($F_{\max}$) έλεγχο του Hartley πρέπει να έχουμε $n_1 = n_2 = ... = n_t = n$ και έστω $\gamma = n-1$. Και αυτός ο έλεγχος χρησιμοποιείται όταν οι βαθμοί ελευθερίας των σ_i^2 είναι ίσοι με $f_1 = f_2 = ... = f_k = v$.

Έτσι ο στατιστικός έλεγχος για την υπόθεση H_0 ως προς H_1 είναι :

$$F_{\max} = \frac{\max_{1 \le i \le t} \hat{\sigma}_i^2}{\min_{1 \le i \le t} \hat{\sigma}_i^2}, \tag{2.8}$$

όπου $\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$ η δειγματική διασπορά για το i-οστό δείγμα και $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$. (βλέπε Pearson & Hartley (1970), p.202 για ειδικούς πίνακες).

Με άλλα λόγια, απορρίπτεται η μηδενική υπόθεση H_0 αν το $F_{\max} \ge c$, όπου c είναι μια σταθερά μου υπολογίζεται από

$$a = \text{Prob(Type I error)}$$

το c υπολογίζεται έτσι ώστε

$$a = \text{Prob(reject } H_0 \mid H_0) = \text{P}(F_{\max} \ge c \mid H_0)$$

Για να αποτιμήσουμε αυτή την πιθανότητα, πρέπει να ξέρουμε την κατανομή του $F_{\max}$ από την υπόθεση H_0 , όμως η κατανομή αυτή είναι περιπλοκή. Ωστόσο, ποσοστιαία σημεία (percentage points) αυτής της κατανομής έχουν υπολογιστεί για διάφορες τιμές του $\gamma = 2, 3, 4, ...$ και $k = t =$ αριθμός αγωγών = αριθμός πληθυσμών $= 2, 3, 4, 5, ...$

Δίνοντας γ, k και a μπορούμε να πάρουμε τις τιμές του c από τον πινάκα που δίνεται από τον Hartley.

Σημείωση: Για διαφορετικά n_i, μπορούμε να πάρουμε $\gamma = \max\{n_1, n_2, ..., n_k\} - 1$. Σε αυτήν την περίπτωση, ο έλεγχος 'χάνει' μια κατάλληλη ρύθμιση (proper control) του σφάλματος Τύπου Ι. Όπως θα δείξουμε στα παρακάτω εμπειρικά πειράματα, ο ρυθμός σφάλματος Τύπου Ι του ελέγχου του Hartley για διαφορετικά μεγέθη δεδομένων μπορεί να είναι μέχρι 0.2 στις περισσότερες προσομοιωμένες περιπτώσεις. Επιπλέον, κάποιος θα πρέπει να βασίζεται σε έναν $F_{\max}$ πινάκα για την κρίσιμη τιμή αυτού του ελέγχου και υπάρχουν διάφορες περιπτώσεις μη διαθέσιμες από έναν F_{max} πινάκα.

2.2.1.6 Ο έλεγχος του Cadwell

Η επιθυμία για απλοποίηση οδήγησε στην αντικατάσταση της διασποράς του *Hart* με το εύρος του δείγματος σε μια μελέτη του Cadwell (1953). Ακριβείς πίνακες για ίσα μεγέθη δειγμάτων δίνονται από τον Harter (1963) για $k = 2$ παράγοντες, και τους Leslie και Brown (1966) για $k \leq 12$. Τον συμβολίζουμε με *Cad* με στατιστικό έλεγχο:

$$T_4 = \frac{\max r_i}{\min r_i}. \tag{2.9}$$

βλέπε Pearson & Hartley (1970), p.264 για ειδικούς πίνακες.

Δεν εξετάζουμε αυτόν τον έλεγχο γιατί πιστεύουμε ότι τα υπολογιστικά πλεονεκτήματα δεν είναι πλέον χρήσιμα με το σημερινό λογισμικό.

2.2.1.7 Ο έλεγχος των Dixon και Massey

Οι Dixon και Massey (1969) ανέφεραν μια παραλλαγή του *Bar* η οποία χρησιμοποιεί την F κατανομή. Τον συμβολίζουμε με *Bar3* με στατιστικό έλεγχο:

$$F_{k-1,w} = \frac{wT_5}{(k-1)(b-T_5)}, \tag{2.10}$$

όπου $w = (k+1)\big/(C-1)^2$, $b = \dfrac{w}{C + 2/w}$ και $T_5 = (N-k)\ln s^2 - \sum_{i=1}^{k}(n_i - 1)\ln s_i^2$.

Επίσης εξετάζουμε την τροποποίηση *Bar3:med.*

2.2.1.8 Μια παραλλαγή του ελέγχου του Bartlett με τη χρήση εύρους

Αυτή η παραλλαγή του ελέγχου του Bartlett, που χρησιμοποιεί το τυποποιημένο εύρος αντί της διασποράς, βρίσκεται στη βιβλιογραφία από τον Patnaik (1950) ο οποίος μελέτησε αυτή τη χρήση του εύρους, αλλά δεν αναφέρεται λεπτομερώς μέχρι τον Gartside (1972). Οι τυποποιημένες σταθερές d_i είναι διαθέσιμες από τους Pearson και Hartley (1970, p. 201). Οι βαθμοί ελευθερίας του X^2 ελέγχου που προκύπτει ρυθμίζονται από το $(k-1)$ έως το v_i, όπου το v_i είναι διαθέσιμο στην ίδια αναφορά. Την συμβολίζουμε με *Bar:range* με στατιστικό έλεγχο:

$$\left[(N-k)\ln\left(\frac{1}{N-k}\sum_i (n_i - 1)(\frac{r_i}{d_i})^2\right) - \sum_i (n_i - 1)\ln(\frac{r_i}{d_i})^2\right]\Big/ C, \tag{2.11}$$

όπου $C = 1 + \dfrac{1}{3(k-1)}\left[\sum_i \dfrac{1}{n_i - 1} - \dfrac{1}{N-k}\right]$.

βλέπε Pearson & Hartley (1970), p.201 για ειδικούς πίνακες.

Δεν εξετάζουμε αυτόν τον έλεγχο γιατί γενικά το εύρος είναι λιγότερο αποδοτικό από την διασπορά του δείγματος.

2.2.1.9 Ο έλεγχος του Samuiddin

Η κυβική ρίζα του s^2 είναι περισσότερο κανονική από το s^2, το οποίο οδηγεί σε αυτόν τον έλεγχο από τον Samuiddin (1976). Τον συμβολίζουμε με *Sam* με στατιστικό έλεγχο:

$$x_{k-1}^2 = \sum_{i=1}^{k} \frac{(m_i - m)^2}{a_i^2}, \tag{2.12}$$

όπου $m_i = \left(1 - \dfrac{2}{9(n_{i-1})}\right) s_i^{-2/3}$, $a_i^2 = 2/[9(n_i - 1)s_i^{4/3}]$ και $m = \dfrac{\sum_i (m_i / a_i^2)}{\sum_j (1/a_j^2)}$

Επίσης εξετάζουμε τον *Sam:med*.

2.2.1.10 Ο έλεγχος του Lehmann

Η προτεινόμενη διαδικασία του Lehmann (1959) είναι ίδια με το *Sch1*, αλλά για $\gamma = 0$, όπως στις κανονικές κατανομές. Την συμβολίζουμε με *Leh1* με στατιστικό έλεγχο.

$$x_{k-1}^2 = T_6/2, \tag{2.13}$$

όπου $T_6 = \sum_i (n_i - 1)(P_i - \dfrac{1}{N-k} \sum_{j=1}^{k} (n_j - 1)P_j)^2$ και $P_j = \ln s_j^2$.

Ο Ghosh (1972) δείχνει ότι ο πολλαπλασιασμός με $(N-k)/(N-2k)$ δίνει μια κατανομή ισοδύναμη με την χ^2. Αυτή την εκδοχή την ονομάζουμε Leh2 με στατιστικό έλεγχο:

$$x_{k-1}^2 = (N-k)T_6/(2N-4k). \tag{2.14}$$

Εξετάζουμε επίσης τις τροποποιήσεις *Leh1:med* και *Leh2:med*.

2.2.1.11 Ο έλεγχος του Box (Box's Test)

Μια ενδιαφέρουσα προσέγγιση για την απόκτηση ενός ισχυρότερου ελέγχου για διασπορές απαιτεί τη χρησιμοποίηση του στατιστικού F , η οποία είναι γνωστό ότι είναι αρκετά ισχυρή. Μια ιδέα που προτάθηκε από τους Barlett και Kendall (1946) αναπτύχθηκε από τον Box (1953) σε έναν έλεγχο ο οποίος είναι γνωστός ως λογαριθμικός anova (log-anova) έλεγχος. Για ένα προεπιλεγμένο τυχαίο ακέραιο $m \geq 2$, κάθε δείγμα διαιρείται σε υποδείγματα μεγέθους m με τυχαίο τρόπο. (Βλ. Martin και Games 1975, 1977 και Martin 1976 για προτεινόμενα δείγματα μεγέθους m.) Οι υπόλοιπες παρατηρήσεις είτε δεν θα χρησιμοποιηθούν είτε θα συμπεριληφθούν σε ένα τελικό υπόδειγμα. Το δείγμα διασποράς s_{ij} υπολογίζεται για κάθε υποδείγμα, $i = 1,...,k$, $j = 1,...,[n_i/m] = J_i$. Ένας λογαριθμικός μετασχηματισμός $Y_{ij} = \ln s_{ij}$ φέρνει τις μεταβλητές πιο κοντά στις κανονικές και ο $F(Y_{ij})$ χρησιμοποιείται ως στατιστικός έλεγχος. Ακόλουθες μελέτες από τους Gartside (1972), Layard (1973) και Levy (1975) επιβεβαίωσαν την ευρωστία της μεθόδου, αλλά επίσης ανακάλυψαν και έλλειψη ισχύος σε σύγκριση με άλλους ελέγχους οι οποίοι είχαν την ίδια ευρωστία.

Μια τροποποίηση που οδηγεί σε ένα πιο κανονικό δείγμα, οφείλεται στον Bargmann με τη βοήθεια του Gartside (1972). Χρησιμοποίει το $W_{ij} = w_i(\ln s_{ij} + c_i)$ όπου w_i και c_i είναι κανονικοποιημένες σταθερές.

Η μέθοδος του ελέγχου του Box:

(i) Διαιρέστε κάθε δείγμα σε ένα αριθμό υποομάδων τυχαία,

(ii) Υπολογίστε τις διασπορές από κάθε υπόδειγμα,

(iii) Υπολογίστε τον λογάριθμο της διασποράς από κάθε υπόδειγμα,

(iv) Διεξάγετε ανάλυση διασποράς με ένα παράγοντα (one-way anova) των καινούργιων δεδομένων, και

(v) Αν το αποτέλεσμα του F-ελέγχου είναι στατιστικά σημαντικό, απορρίπτουμε την H_0.

Η ιδέα πίσω από τον έλεγχο του Box είναι ότι από την στιγμή που οι διασπορές του κάθε υποδείγματος δεν κατανέμονται κανονικά, χρειάζεται να πάρουμε τον λογάριθμο της διασποράς έτσι ώστε να κάνουμε τα νέα στοιχεία προσεγγιστικά κανονικά (approximately normal) και τότε εφαρμόζοντας την τεχνική της ανάλυσης διασποράς στα νέα στοιχεία, θα έχουμε έναν προσεγγιστικό έλεγχο για την ομοιογένεια των διασπορών.

Ο έλεγχος του Box έχει μικρή ισχύ για τα δεδομένα μικρού μεγέθους εξαιτίας της υποδιαίρεσης. Το αποτέλεσμα του ελέγχου είναι ασταθές με την λογική ότι διαφορετικά χωρίσματα οδηγούν σε διαφορετικά υπολογιζόμενες στατιστικές τιμές ελέγχων και ακόμα και αντιφατικά συμπεράσματα. Επίσης, η τυχαία μέθοδος της υποδιαίρεσης δειγμάτων και η πιθανότητα μη χρησιμοποίησης όλων των παρατηρήσεων κάνει αυτές τις διαδικασίες αποτρεπτικές για τον χρήστη. Γι’ αυτό το λόγο, δεν λαμβάνουμε υπόψη τον έλεγχο του Box στην μελέτη προσομοίωσης.

2.2.1.12 Ο έλεγχος του Levene (*Levene's Test*)

Ο Levene (1960), πρότεινε τη χρήση της ανάλυσης της διασποράς κατά ένα παράγοντα (one-way) των μεταβλητών $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$ σαν μια μέθοδο για την ένταξη της ευρωστίας εκείνου του ελέγχου σε έναν έλεγχο για την διασπορά. Τον συμβολίζουμε με *Lev1*. Επιπλέον προτεινόμενες παραλλαγές από τον Levene περιλαμβάνουν $Z_{ij}^{1/2}$ (*Lev2*), $\ln Z_{ij}$ (*Lev3*), και Z_{ij}^{2} (*Lev4*). Επίσης θεωρούμε το *Lev1:med*, συνιστώμενο από τους Brown και Forsythe (1974), και το *Lev4:med*, αλλά δεν εξετάζουμε το *Lev3:med* γιατί το $\ln 0 = -\infty$ προκύπτει με δείγματα περιττού μεγέθους.

Οι στατιστικοί έλεγχοι για τις παραλλαγές του ελέγχου του Levene είναι:

<u>Lev1</u> $$F_{k-1,N-k} = F(|X_{ij} - \bar{X}_i|) \qquad (2.15)$$

<u>Lev2</u> $$F_{k-1,N-k} = F((X_{ij} - \bar{X}_i)^2) \qquad (2.16)$$

<u>Lev3</u> $$F_{k-1,N-k} = F(\ln(X_{ij} - \bar{X}_i)^2) \qquad (2.17)$$

<u>Lev4</u> $$F_{k-1,N-k} = F(|X_{ij} - \bar{X}_i|^{1/2}) \qquad (2.18)$$

Ένας τρόπος για να ελεγχθούν οι πιθανές διαφορές στις διασπορές είναι να κάνουμε έναν F - έλεγχο (test) με τους κατάλληλους βαθμούς ελευθερίας. Αλλά φάνηκε ότι ο F – έλεγχος δεν είναι ο καλύτερος τρόπος για να γίνει αυτό. Στην πραγματικότητα είναι πολύ ευαίσθητος στις παραβιάσεις της υπόθεσης κανονικότητας. Δηλαδή, εάν οι πληθυσμοί εμφανίζονται να είναι μη κανονικοί, τότε ο F - έλεγχος θα τείνει να απορρίψει τη μηδενική υπόθεση στις πληθυσμιακές διασπορές. Υπάρχει ένας καλύτερος τρόπος. Αυτός είναι ο τροποποιημένος έλεγχος Levene. Ο έλεγχος αυτός είναι επίσης λιγότερο ευαίσθητος σε σχέση με τον έλεγχο του Barlett στη μη κανονικότητα των δεδομένων. Παρόλα αυτά αν έχουμε ισχυρές ενδείξεις ότι τα δεδομένα ακολουθούν κανονική κατανομή, τότε το Barlett έχει καλύτερη απόδοση. Πριν τη χρήση του στατιστικού, γίνεται ένας μετασχηματισμός των αρχικών δεδομένων ως εξής:

Έστω Y_{ij} η j παρατήρηση της i ομάδας, $j = 1,...,n_i$ και $i = 1,...,k$, όπου n_i: πλήθος παρατηρήσεων στην i ομάδα, k: πλήθος ομάδων παρατηρήσεων και $N = \sum n_i$ το συνολικό πλήθος παρατηρήσεων. Ορίζουμε ως Z_{ij} ένα από τα παρακάτω:

1. $Z_{ij} = | Y_{ij} - \overline{Y}_{i.} |$, όπου $\overline{Y}_{i.}$ είναι ο μέσος της i ομάδας. (2.19)

2. $Z_{ij} = | Y_{ij} - \tilde{Y}_{i.} |$, όπου $\tilde{Y}_{i.}$ είναι η διάμεσος της i ομάδας. (2.20)

3. $Z_{ij} = | Y_{ij} - \overline{Y}_{i.(tr)} |$, όπου $\overline{Y}_{i.(tr)}$ είναι ο 10% αποκομμένος μέσος της i ομάδας. (2.21)

Το στατιστικό που χρησιμοποιείται είναι:

$$W = \frac{(N-k)\sum_{i=1}^{k} n_i (\overline{Z}_{i.} - \overline{Z}_{..})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i} (Z_{ij} - \overline{Z}_{i.})^2}, \tag{2.22}$$

όπου $\overline{Z}_{i.} = \sum_j Z_{ij} \big/ n_i$ ο καινούργιος μέσος της κάθε ομάδας i

και $\overline{Z}_{..} = \sum_i \sum_j Z_{ij} \big/ N$ ο συνολικός μέσος των μετασχηματισμένων παρατηρήσεων.

Η μηδενική υπόθεση απορρίπτεται αν $W > F_{(k-1, N-k, \alpha)}$.

Όταν προτάθηκε ο έλεγχος από τον Levene (1960), χρησιμοποιήθηκε μόνο η περίπτωση 1. Οι Brown και Forsythe (1974) επέκτειναν τον έλεγχο, προτείνοντας και τη χρήση είτε της διαμέσου είτε του αποκομμένου μέσου. Με Monte Carlo μελέτες που έκαναν κατέληξαν στο ότι έχουμε καλύτερη απόδοση με χρήση της διαμέσου στην περίπτωση όπου τα δεδομένα ακολουθούν χ_4^2 (δηλαδή λοξή (skewed)) κατανομή), ενώ για την περίπτωση της Cauchy κατανομής (δηλαδή κατανομής με παχιές ουρές (heavy-tailed distribution)) καλύτερα αποτελέσματα είχαν με τη χρήση του trimmed μέσου. Αντιθέτως, η χρήση του μέσου ήταν καταλληλότερη για συμμετρικές κατανομές.

Παρόλο λοιπόν που η βέλτιστη επιλογή εξαρτάται από την κατανομή των δεδομένων, συνίσταται η χρήση της διαμέσου (κάτι που γίνεται στα πακέτα Statgraphics και Minitab) ώστε να έχουμε ευρωστία (robustness) του ελέγχου, ιδίως για τις περιπτώσεις μη κανονικών κατανομών.

2.2.2 Έλεγχοι που επιχειρούν να εκτιμήσουν την κύρτωση

2.2.2.1 Τροποποίηση του ελέγχου του Bartlett

Ο Box (1953) έδειξε ότι η ασυμπτωτική κατανομή του *Bar* ήταν εξαρτημένη από την κοινή κύρτωση των δειγματικών κατανομών και διαιρώντας τον *Bar* με $(1+\gamma/2)$, όπου $\gamma = E(X_{ij} - \mu_i)^4 \big/ \sigma_i^4 - 3$, ο έλεγχος δεν θα ακολουθούσε ασυμπτωτική κατανομή, με την προϋπόθεση ότι θα βρίσκαμε την ίδια κύρτωση. Η δική μας μορφή του μετασχηματισμού του *Bar* περιλαμβάνει εκτιμήσεις του γ με μέρη του δείγματος(sample moments), μια πρόταση την οποία ο Layard (1973) οφείλει στον Scheffe (1959). Τον συμβολίζουμε με *Bar1* με στατιστικό έλεγχο:

$$x_{k-1}^2 = \frac{T_5}{C(1+\hat{\gamma}/2)}, \tag{2.23}$$

όπου $\hat{\gamma} = \frac{1}{N}\sum_i n_i^2 \frac{\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^4}{[(n_i-1)s_i^2]^2} - 3$, $\quad C = 1 + \frac{1}{3(k-1)}\left[\sum_i \frac{1}{n_i - 1} - \frac{1}{N-k}\right]$ και

$T_5 = (N-k)\ln s^2 - \sum_{i=1}^{k}(n_i - 1)\ln s_i^2$.

Η παραλλαγή *Bar2* είναι αποτέλεσμα από διαφορετική εκτιμήτρια για το γ, όπως δίνεται από τον Layard. Ο στατιστικός έλεγχος για το *Bar2* είναι:

$$x_{k-1}^2 = \frac{T_5}{C(1+\tilde{\gamma}/2)}, \tag{2.24}$$

όπου $\tilde{\gamma} = \dfrac{N\sum_i \sum_j (X_{ij} - \bar{X}_i)^4}{\left[\sum_i (n_i - 1)s_i^2\right]^2} - 3$.

Επίσης εξετάζουμε τις τροποποιήσεις *Bar1:med*. και *Bar2:med.*

2.2.2.2 Ο έλεγχος του Scheffe

Ο στατιστικός έλεγχος αυτής της παραμετρικής διαδικασίας, που αποδίδεται από τον Layard (1973) στον Scheffe (1959), μοιάζει κατά κάποιον τρόπο με τον αριθμητή μιας F στατιστικής συνάρτησης που υπολογίζει το s_i σταθμισμένο από $n_i - 1$ βαθμούς ελευθερίας. Ο παρανομαστής είναι μία συνάρτηση της (υποτιθέμενης) κοινής κύρτωσης, η οποία πρέπει να υπολογιστεί στην πράξη. Χρησιμοποιούμε το δείγμα κύρτωσης για το γ, και τον συμβολίζουμε με *Sch1* με στατιστικό έλεγχο:

$$x_{k-1}^2 = \frac{T_6}{2 + (1 - k/N)\hat{\gamma}}, \tag{2.25}$$

όπου $\hat{\gamma} = \dfrac{1}{N}\sum_i n_i^2 \dfrac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^4}{[(n_i - 1)s_i^2]^2} - 3$ και $T_6 = \sum_i (n_i - 1)(P_i - \dfrac{1}{N-k}\sum_{j=1}^{k} (n_j - 1)P_j)^2$.

Οι παραλλαγή *Sch2* προκύπτει, όταν χρησιμοποιείται η εκτιμήτρια του Layard για το γ με στατιστικό έλεγχο:

$$x_{k-1}^2 = \frac{T_6}{2 + (1 - k/N)\tilde{\gamma}}, \tag{2.26}$$

όπου $\tilde{\gamma} = \dfrac{N\sum_i \sum_j (X_{ij} - \bar{X}_i)^4}{\left[\sum_i (n_i - 1)s_i^2\right]^2} - 3$.

Επίσης εξετάζουμε τα *Sch1:med* και *Sch2:med*

2.2.2.3 Ο x^2 έλεγχος του Layard

Ο Layard πρότεινε το παρακάτω στατιστικό:

$$LAY = \sum_{i=1}^{9} f_i^2 \left\{ \ln s_i^2 - \frac{\sum_{i=1}^{9} f_i \ln s_i^2}{f} \right\} \Big/ \hat{r}^2 , \tag{2.27}$$

όπου,

$$\hat{r}^2 = 2 + \{1 - (k/n)\}\hat{\gamma} , \tag{2.28}$$

και

$$\hat{\gamma} = \frac{n \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^4}{\left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \right\}^3} - 3, \tag{2.29}$$

είναι η εκτίμηση της κύρτωσης. Ο Layard πρότεινε ότι υπό την υπόθεση H_0 ο στατιστικός έλεγχος LAY ακολουθεί ασυμπτωτικά την χ^2 κατανομή με ($k-1$) βαθμούς ελευθερίας (χ^2_{k-1}).

2.2.3 <u>Μη παραμετρικοί Έλεγχοι</u>

2.2.3.1 Ο έλεγχος του Mood

Ο πρώτος μη παραμετρικός έλεγχος για το πρόβλημα της διασποράς παρουσιάστηκε από τον Mood (1954) και τον συμβολίζουμε με *Mood*. Αυτός, όπως οι περισσότεροι μη παραμετρικοί έλεγχοι, προϋποθέτει ισόνομες (identical) κατανομές βάσει της μηδενικής υπόθεσης. Συγκεκριμένα, απαιτεί ίσους μέσους ή ένα γνωστό μετασχηματισμό που να επιτυγχάνει ίσους μέσους, το οποίο συχνά δεν συναντάται σε εφαρμογές. Για αυτό, έχουμε προσαρμόσει τον έλεγχο Mood και όλους τους μη παραμετρικούς ελέγχους ως εξής: Αντί να αφήσουμε το R_{ij} να πάρει το βαθμό (rank) του

X_{ij} όταν οι μέσοι είναι ίσοι ή του $(X_{ij} - \mu_i)$ όταν οι μέσοι είναι άνισοι αλλά γνωστοί, θέτουμε το R_{ij} να πάρει το βαθμό (rank) του $(X_{ij} - \bar{X}_i)$. Κάθε X_{ij} τότε αντικαθίσταται από την τιμή a_N, R_{ij} βασισμένο σε αυτή την τάξη. Το αποτέλεσμα είναι ένα έλεγχος ο οποίος δεν είναι μη παραμετρικός αλλά μπορεί να είναι εύρωστος και ισχυρός όπως μερικοί παραμετρικοί ανταγωνιστές του. Η χρήση του $\tilde{X}_i$ αντί του $\bar{X}_i$ οδηγεί στο *Mood:med*, το οποίο επίσης θα εξετάσουμε. Η X^2 προσέγγιση και η F προσέγγιση για κάθε έλεγχο οδηγεί σε τέσσερις παραλλαγές (variations), τις οποίες μελετάμε.

2.2.3.2 Ο έλεγχος των Freund, Ansari και Bradley

Παρόλο που ο έλεγχος Mood είναι μια δευτεροβάθμια συνάρτηση του R_{ij}, αυτός ο έλεγχος που παρουσιάστηκε από τους Freund και Ansari (1957) και αναπτύχθηκε περεταίρω από τους Ansari και Bradley (1960) είναι μια γραμμική συνάρτηση του R_{ij}. Ξανά, θέτουμε το R_{ij} να πάρει το βαθμό του (rank) $(X_{ij} - \bar{X}_i)$. Τον συμβολίζουμε με *F-A-B*. Εξετάζουμε τέσσερις διαφοροποιήσεις του F-A-B (βλ. Mood).

2.2.3.3 Οι έλεγχοι B-D και S-T

Ο *B-D* έλεγχος, όπως τον συμβολίζουμε, παρουσιάστηκε από τους Barton και David (1958) λίγο μετά τον έλεγχο *F-A-B* και είναι καταρχήν παρόμοιος με αυτόν. Ενώ οι τιμές του *F-A-B* είναι σε τριγωνικό σχήμα, οι τιμές του *B-D* ακολουθούν ένα σχήμα V με μεγάλες τιμές στα άκρα και μικρές τιμές στη μεγάλη διάμεσο. Το αποτέλεσμα είναι ένας έλεγχος ίδιας ευρωστίας και ισχύος με το *F-A-B*.

Τα ίδια μπορούμε να πούμε για τον έλεγχο *S-T*, όπως τον συμβολίζουμε, που παρουσιάστηκε από τους Siegel και Tukey (1960) περίπου την ίδια περίοδο. Το μόνο πλεονέκτημα του ελέγχου S-T είναι ότι μπορούν να χρησιμοποιηθούν οι πίνακες για τον έλεγχο Mann-Whitney, χωρίς να απαιτούνται ιδιαιτέρως ακριβείς (exact) πίνακες. Δεν εξετάζουμε τους ελέγχους *B-D* και τα *S-T* εδώ γιατί τα αποτελέσματα θα ήταν ουσιαστικά τα ίδια με αυτά που βρήκαμε για το *F-A-B*.

2.2.3.4 Ο έλεγχος του Capon

Αντί για τη χρήση των scores που είναι μια τετραγωνική συνάρτηση των τάξεων (ranks) όπως είχε κάνει ο *Mood*, ο Capon (1961) πρότεινε την επιλογή των scores που δίνουν τη βέλτιστη ισχύ υπό κάποια έννοια. Το αποτέλεσμα είναι αυτός ο κανονικός έλεγχος scores, ο οποίος είναι τοπικά ο ισχυρότερος μεταξύ των βαθμωτών ελέγχων έναντι στις κανονικού τύπου εναλλακτικές, και τοπικά ασυμπτωτικά ο ισχυρότερος μεταξύ όλων των ελέγχων για αυτήν την εναλλακτική.

2.2.3.5 Ο έλεγχος του Klotz

Λίγο αργότερα, ο Klotz (1962) εισήγαγε έναν άλλο κανονικό έλεγχο scores όπου χρησιμοποίησε τα καταλληλότερα κανονικά ποσοστημόρια (quantiles). Το αποτέλεσμα έχει πιθανότατα μικρότερη ισχύ τοπικά για δείγματα μικρού μεγέθους, αλλά έχει τις ίδιες ασυμπτωτικές ιδιότητες όπως το *Capon*. Εξαιτίας αυτής της ευκολίας, εξετάζουμε τον έλεγχο του *Klotz*, αλλά όχι τον αντίστοιχο του *Capon*. Όπως και στο Mood, μελετάμε τέσσερις διαφοροποιήσεις του *Klotz*.

2.2.3.6 Ο έλεγχος των Flinger και Killeen

F-K. Οι Figner και Killeen (1976) πρότειναν ταξινόμηση του $|X_{ij}|$ και καθορισμένη αύξηση των τιμών $a_{N,i} = i$, $a_{N,i} = i^2$, και $a_{N,i} = \Phi^{-1}(1/2 + (i/2(N+1)))$ βάσει αυτών των ταξινομήσεων. Προτείνουμε τη χρήση των βαθμών (ranks) $|X_{ij} - \overline{X}_i|$ και ονομάζουμε τον πρώτο έλεγχο *T-G* από τους Talwar και Gentle (1977) οι οποίοι χρησιμοποίησαν έναν αποκομμένο μέσο αντί του $\overline{X}_i$. Ο δεύτερος έλεγχος, που ονομάζεται squared ranks έλεγχος, *S-R*, είχε μελετηθεί από τους Conover και Iman (1978), αλλά έχει ρίζες σε προγενέστερες δημοσιεύσεις των Shorack (1965), Duran και Mielke (1968), και άλλων. Επισημαίνουμε τον τρίτο έλεγχο, *F-K*, παρόλο που έχουμε αλλάξει τις προτάσεις τους. Επίσης, εξετάζουμε, όπως και με τον έλεγχο Mood, τις τέσσερις παραλλαγές που σχετίζονται με κάθε έλεγχο. Δεν εξετάζουμε τις προτάσεις των Fligner και Killeen για χρησιμοποίηση της μεγάλης διαμέσου στην θέση του $\bar{X}_i$.

<u>Πίνακας 2.1.</u>
Γραμμικοί βαθμωτοί έλεγχοι (τα αποτελέσματα μπορούν να χρησιμοποιηθούν στις εξισώσεις (2. 1), (2.2), ή (2.3))

Abbreviation of Test	Score Function $a_{N,i}$	Score $a_{N,i}$ είναι μια συνάρτηση του R_{ij}, όπου R_{ij} παίρνει τη θέση του:
<u>Mood</u>	$(i-\frac{N+1}{2})^2$	$(X_{ij}-\bar{X}_i)$
<u>F-A-B</u>	$\frac{N+1}{2}-\lvert i-\frac{N+1}{2}\rvert=1,2,3...,3,2,1$	$(X_{ij}-\bar{X}_i)$
<u>B-D</u>	$...,3,2,1,1,2,3,...$	$(X_{ij}-\bar{X}_i)$
<u>S-T</u>	$1,4,5,...,6,3,2$	$(X_{ij}-\bar{X}_i)$
<u>Capon</u>	$[E(Z_{N,i})]^2$ όπου $Z_{N,i}$ είναι το i-οστό στατιστικό στοιχείο διάταξης από ένα κανονικό τυχαίο δείγμα μεγέθους N	$(X_{ij}-\bar{X}_i)$
<u>Klotz</u>	$[\Phi^{-1}(\frac{i}{N+1})]^2$ όπου η $\Phi(x)$ είναι η συνάρτηση της τυποποιημένης κανονικής κατανομής	$(X_{ij}-\bar{X}_i)$
<u>T-G</u>	i	$\lvert X_{ij}-\bar{X}_i\rvert$
<u>S-R</u>	i^2	$\lvert X_{ij}-\bar{X}_i\rvert$
<u>F-K</u>	$\Phi^{-1}(\frac{1}{2}+\frac{i}{2(N+1)})$ (βλέπε <u>Klotz</u> για την Φ)	$\lvert X_{ij}-\bar{X}_i\rvert$

2.2.4 <u>Jackknife έλεγχοι</u>

2.2.4.1 Ο jackknife έλεγχος του Miller

Ο Layard (1973) πρότεινε έναν x^2 στατιστικό έλεγχο ο οποίος είναι μια συνάρτηση της κύρτωσης και ακολουθεί ασυμπτωτικά την κατανομή χ^2_{k-1}. Ο Miller συμπέρανε ότι για πολύ μικρά δείγματα η υψηλή συσχέτιση μεταξύ των αποκλίσεων στο σύνολο δειγμάτων, καταστρέφει την ευρωστία του ελέγχου του Levene. Για να ξεπεράσει αυτό το πρόβλημα, πρότεινε τη χρήση του ελέγχου του Layard σε μια jackknife διαδικασία και υπολόγισε το στατιστικό F για την ανάλυση της διασποράς κατά ένα παράγοντα (one-way ANOVA). Η μέθοδος jackknife στηρίζεται στο χωρισμό των δειγμάτων σε υποδείγματα κάποιου προκαθορισμένου μεγέθους m. Παίρνουμε $m=1$, για να αφαιρέσουμε την τυχαία μεταβλητότητα όταν το $m>1$. Τον συμβολίζουμε με *Mill* με στατιστικό έλεγχο:

$$F_{k-1,N-k} = F(U_{ij}), \tag{2.30}$$

όπου $U_{ij} = n_i \ln s_i^2 - (n_i - 1)\ln s_{ij}^2$ και $s_{ij}^2 = \dfrac{1}{n_i - 2}\left[(n_i-1)s_i^2 - n_i(X_{ij} - \bar{X}_i)^2 \big/ (n_i - 1)\right]$.

(Βλ. εξίσωση (2.3) για F())

Δεν εξετάζουμε το *Mill:med*.

2.2.4.2 Ο τροποποιημένος jackknife έλεγχος του Miller

Ο Layard (1973) επέκτεινε τον jackknife έλεγχο του Miller για να εξετάσει την ομοιογένεια των διασπορών για διάφορους πληθυσμούς.

$$S^2_{i(j)} = \sum_{\ell \neq j=1}^{n_i} (X_{i\ell} - \bar{X}_{i(j)})^2 \Big/ (n_i - 2), \tag{2.31}$$

όπου

$$\bar{X}_{i(j)} = \sum_{\ell \neq j=1}^{n_i} X_{i\ell} \big/ f_i \,,\ i = 1, 2, ..., k \,. \tag{2.32}$$

Λαμβάνουμε τις ψευδοτιμές ως

$$U_{ij} = n_i \ln S_i^2 - (n_i - 1) \ln S_{i(j)}^2 \,, \qquad i = 1, 2, ..., k \,,\ j = 1, 2, ..., n_i \,, \tag{2.33}$$

και τελικά υπολογίζουμε έναν F-έλεγχο από την ανάλυση της διασποράς κατά ένα παράγοντα στην ψευδοτιμή U_{ij}, και συγκεκριμένα

$$JACK = \frac{\sum_{i=1}^{k} n_i (\bar{U}_{i\bullet} - \bar{U}_{\bullet\bullet})^2 \big/ (k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (U_{ij} - \bar{U}_{i\bullet})^2 \big/ f} \,. \tag{2.34}$$

Σημειώνουμε ότι, υπό την υπόθεση H_0, ο έλεγχος $JACK$ είναι ασυμπτωτικά κατανεμημένος ως ένας F -έλεγχος με ($k-1,\ f$) βαθμούς ελευθερίας.

2.2.4.3 Ο νέος jackknife έλεγχος (New Jackknife Test)

Όταν οι πληθυσμοί είναι κανονικοί ή μη-κανονικοί με πεπερασμένο μέσο αλλά και διασπορά, ο νέος έλεγχος *jackknife* για να ελέγξει την μηδενική υπόθεση, $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = k^2$ ή $H_0 : \sigma_i^2 \big/ \sigma^2 = 1$, $i = 1, 2, ..., k$ περιλαμβάνει τα ακόλουθα βήματα. Αρχικά παίρνουμε μια εκτίμηση της διασποράς σ^2 από το συνολικό δείγμα.

$$S^2 = \frac{\sum_{i=1}^{k} (n_i - 1) S_i^2}{\sum_{i=1}^{k} (n_i - 1)} \tag{2.35}$$

όπου,
$$S_i^2 = \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{(n_i - 1)}$$

Μετά, διαγράφουμε την i-οστή ομάδα παρατηρήσεων και έχουμε:

$$S_{(-i)}^2 = \frac{\sum_{\ell \neq i=1}^{k} (n_\ell - 1) S_\ell^2}{\sum_{\ell \neq i=1}^{k} (n_\ell - 1)} \tag{2.36}$$

Έπειτα, υπολογίζουμε την ψευδοτιμή που δίνεται από:

$$\xi_i = k \ln S^2 - (k-1) \ln S_{(-i)}^2, \qquad i = 1, 2, ..., k$$

O Tukey (1958) πρότεινε ότι σε πολλές καταστάσεις οι ψευδοτιμές, ξ_i, μπορούν να θεωρηθούν ως σχεδόν ανεξάρτητες και ισόνομα (identically) κατανεμημένες. Μετά ο νέος έλεγχος *jackknife,* NEWJACK, για τον έλεγχο της υπόθεσης H_0 είναι ένας t-έλεγχος με $(k-1)$ βαθμούς ελευθερίας

$$NEWJACK = \frac{\sqrt{k}(\bar{\xi}_\bullet - \xi)}{\left\{\sum_{i=1}^{g} (\xi_i - \bar{\xi}_\bullet)^2 \big/ (k-2)\right\}^{1/2}}, \tag{2.37}$$

όπου $\bar{\xi}_\bullet = \sum_{i=1}^{k} \xi_i \big/ k$ και $\xi = (\ln \sigma_i^2 \big/ \sigma^2)$. Σημειώνεται ότι για την υπόθεση H_0, είναι $\xi = 0$. Επιπλέον, για τον NEWJACK έλεγχο για να ισχύει η t-κατανομή με λογικούς βαθμούς ελευθερίας, υποθέτουμε ότι $k > 2$.

2.2.4.4 Ο τροποποιημένος x^2 έλεγχος του Layard

Όπως παρατηρήθηκε από τον Layard, η εκτίμηση της κύρτωσης του Scheffe, γ, ήταν εσφαλμένα μεροληπτική (badly biased) στη δειγματοληψία από μη-κανονικούς πληθυσμούς. Ο Layard πρότεινε μια εναλλακτική εκτίμηση του γ, που δίνεται από τη σχέση (2.29) , η οποία είναι λιγότερο μεροληπτική (biased) από την εκτίμηση του Scheffe. Ο Quenouille (1956) αρχικά χρησιμοποίησε την jackknife διαδικασία για να μειώσει τη μεροληψία (bias) μιας εκτιμήτριας. Εδώ προτείνουμε τη χρήση της jackknife εκτίμησης του γ, στον χ^2 έλεγχο του Layard. Λαμβάνουμε την εκτιμήτρια $\hat{\gamma}$ από όλες τις παρατηρήσεις που δίνεται από τη σχέση (2.29). Έπειτα, διαγράφουμε την i-οστή ομάδα των παρατηρήσεων την κάθε φορά και έχουμε την εκτιμήτρια $\hat{\gamma}_{(-i)}$, όπου

$$\hat{\gamma}_{(-i)} = \frac{\left(\sum_{\ell \neq i=1}^{k} n_\ell\right) \sum_{\ell \neq i=1}^{k} \sum_{j=1}^{n_i} (X_{\ell j} - \bar{X}_\ell)^4}{\left\{\sum_{\ell \neq i=1}^{k} \sum_{j=1}^{n_i} (X_{\ell j} - \bar{X}\ell)^2\right\}^2} - 3.$$

Στη συνέχεια, υπολογίζουμε τις ψευδοτιμές,

$$\bar{\gamma}_i = k\hat{\gamma} - (k-1)\hat{\gamma}_{(-i)}, \qquad i = 1, 2, \ldots, k,$$

και τελικά η jackknife εκτίμηση του γ είναι

$$\hat{\gamma}_J = \sum_{i=1}^{k} \bar{\gamma}_i / k .$$

Όταν η εκτιμήτρια $\hat{\gamma}_J$ χρησιμοποιείται στην σχέση (2.27) το στατιστικό που προκύπτει δηλώνεται ως LAYJACK.

Κεφάλαιο 3

Συγκρίσεις ελέγχων

3.1 Σύγκριση βασικών ελέγχων της ισότητας των διασπορών

3.1.1 Περίληψη

7 έλεγχοι ισότητας των διασπορών συγκρίνονται ως προς την ευρωστία και την ισχύ σε μια πειραματική προσομοίωση για μικρά έως μέτρια μεγέθη δειγμάτων. Υποθέτουμε ότι τα δεδομένα έρχονται από μια οικογένεια κατανομών θέσης-κλίμακας (location-scale family) με άγνωστους μέσους, διασπορές και συναρτήσεις πυκνότητας. Οι έλεγχοι που χρησιμοποιούνται είναι, ο έλεγχος του Levene, του Bartlett με και χωρίς προσαρμογή της κύρτωσης, ο έλεγχος των Box-Andersen, και 3 έλεγχοι jackknife. Συγκρίνονται επίσης και οι *bootstrap* εκδοχές αυτών των ελέγχων. Βρέθηκε ότι ο έλεγχος του Levene, και ένας από τους jackknife ελέγχους, όπως επίσης και οι bootstrap εκδοχές του ελέγχου του Levene, του ελέγχου του Bartlett με προσαρμογή της κύρτωσης και των δύο jackknife ελέγχων, είναι εύρωστοι. Μέσα σε όλα αυτά, η *bootstrap* εκδοχή του ελέγχου του Levene δείχνει να έχει τη μεγαλύτερη ισχύ.

3.1.2 Εισαγωγή

Εξετάζουμε το πρόβλημα του ελέγχου της ισότητας των διασπορών των k πληθυσμών με δείγματα $\{x_{ij} : j = 1,\ldots,n_i\}$ από τον i-οστό πληθυσμό με μέσο μ_i, διασπορά σ_i^2 και συνάρτηση κατανομής $F\{\sigma_i^{-1}(x-\mu_i)\}(i=1,...,k)$. Η συνάρτηση F και οι σταθερές $\mu_1,...,\mu_{k,}\sigma_1,...,\sigma_k$ είναι άγνωστες. Η μηδενική υπόθεση είναι

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$

Και η εναλλακτική υπόθεση είναι

$H_1 : \sigma_i^2 \neq \sigma_j^2$ για ένα τουλάχιστον $i \neq j$.

Οι μελέτες προσομοίωσης που καταγράφονται στην βιβλιογραφία (όπως Brown και Forsythe, 1974, Conover et al., 1981, Geng et al., 1979, Hall, 1972, Keselman et al., 1979, Sharma, 1991) δηλώνουν ότι δεν υπάρχει έλεγχος ο οποίος να είναι γενικά καλύτερος για όλες τις κατανομές και για διάφορα μεγέθη δειγμάτων. Ένας έλεγχος ο οποίος ξεχωρίζει όσον αφορά την ισχύ και την ευρωστία έναντι της μη-κανονικότητας είναι ο έλεγχος του Levene χρησιμοποιώντας τη διάμεσο του δείγματος ως μια εκτίμηση της παραμέτρου θέσης.

Ο έλεγχος του Levene είναι η ανάλυση της διασποράς κατά ένα παράγοντα του F – ελέγχου για $|x_{ij}-\overline{x}_{i.}|$, δηλαδή οι απόλυτες αποκλίσεις του x_{ij} από τον μέσο της ομάδας $\overline{x}_{i.}$ (Draper και Hunter, 1969). Διάφορες τροποποιήσεις του ελέγχου του Levene έχουν προταθεί και εξεταστεί. Ο Brown και ο Forsythe (1974) λάβανε υπόψη την διάμεσο και τον 10% αποκομμένο μέσο, οι οποίοι είναι περισσότερο εύρωστες εκτιμήσεις της θέσης, ως εναλλακτικοί του μέσου του δείγματος κατά τον υπολογισμό των απόλυτων αποκλίσεων. Αυτοί βρήκαν ότι οι εναλλακτικές διατυπώσεις του ελέγχου του Levene είναι εύρωστες έναντι της μη-κανονικότητας. Ο Loh (1987) εξετάζει την αποτελεσματικότητα της εφαρμογής της μεθόδου του Satterthwaite των διορθωμένων (correcting) βαθμών ελευθερίας και του δυναμικού μετασχηματισμού που βασίζεται στα δεδομένα, στον έλεγχο του Levene με τις διαμέσους των ομάδων να αντικαθιστούν τους μέσους τους. Ο Loh βρήκε ότι η μέθοδος του Satterthwaite μπορεί να βελτιώσει την ευρωστία του ελέγχου του Levene για μικρά δείγματα, αλλά ισχυροί μετασχηματισμοί που βασίζονται στα δεδομένα αυξάνουν το επίπεδο σημαντικότητας του ελέγχου. Οι Yitnosumarto και O'Neill (1986) δίνουν μια ακόμα μέθοδο για τροποποίηση των βαθμών ελευθερίας του F – ελέγχου.

Οι Boos και Brownie (1989) μελέτησαν τις *bootstrap* εκδοχές του ελέγχου του Bartlett και την γενίκευση k-δειγμάτων του Layard (1973) για τον jackknife έλεγχο του Miller (1968) με δύο παράγοντες. Απέδειξαν ότι για οικογένειες θέσης-κλίμακας (location-scale family) η *bootstrap* εκδοχή του ελέγχου του Bartlett είναι σύμφωνη με την υπόθεση H_0 ως $\min\{n_1,...,n_k\} \to \infty$. Τα αποτελέσματα της προσομοίωσης δείχνουν ότι οι *bootstrap* εκδοχές του ελέγχου του Bartlett και ο jackknife έλεγχος πληρούν τους όρους καλύτερα από ότι οι πρωτότυπες εκδοχές τους. Ωστόσο, όταν τα δεδομένα προέρχονται από μια εκθετική κατανομή, τα μεγέθη των *bootstrap* ελέγχων τείνουν να είναι σχετικά μεγάλα.

Ο σκοπός μας είναι να συγκρίνουμε τους εφτά υπάρχοντες ελέγχους και τις *bootstrap* εκδοχές τους όσον αφορά την ευρωστία του επίπεδου σημαντικότητας (την δυνατότητα να διατηρούν το κανονικό επίπεδο του ελέγχου σε μια ποικιλία από διαφορετικές επιλογές) και ισχύος (την δυνατότητα να απορρίπτουν την H_0 όταν είναι λανθασμένη) σε ένα πείραμα προσομοίωσης.

Το υπόλοιπο του κεφαλαίου είναι οργανωμένο ως εξής: Στην παράγραφο 3.1.3 παρέχεται μια περιγραφή όλων των ελέγχων, στην παράγραφο 3.1.4 παρουσιάζεται ο τρόπος εκτέλεσης της *bootstrap* εκδοχής του κάθε ελέγχου. Στην παράγραφο 3.1.5 καταγράφονται τα αποτελέσματα από ένα πείραμα προσομοίωσης για την απόδοση των ελέγχων σε μικρά έως μέτρια μεγέθη δειγμάτων και στην παράγραφο 3.1.6 παρουσιάζονται κάποια συμπεράσματα.

3.1.3 Περιγραφή των Ελέγχων

Σε αυτήν την παράγραφο θα δώσουμε μια περιγραφή του κάθε ελέγχου. Ορίζουμε τον μέσο (mean) της ομάδας $\overline{x}_{i.} = \sum_{j=1}^{n_i} x_{ij} \big/ n_i$, την διασπορά της ομάδας $s_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_{i.})^2 \big/ (n_i - 1)$, και το συνολικό μέγεθος δείγματος $N = \sum_{i=1}^{k} n_i$.

Στο κεφάλαιο αυτό θα εξετάσουμε την μηδενική υπόθεση $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$ έναντι της εναλλακτικής όπου δεν είναι όλοι οι όροι σ_i^2 ίσοι.

Ο έλεγχος του Levene που θα χρησιμοποιήσουμε σ’ αυτό το κεφάλαιο παρουσιάζεται στο κεφάλαιο 2 (παράγραφος 2.2.1.12).

3.1.3.1 Τροποποιήσεις του ελέγχου του Bartlett

Ο στατιστικός έλεγχος του Bartlett, B_1, παρουσιάζεται στο κεφάλαιο 2 (παράγραφος 2.2.1.2).

Λαμβάνουμε επίσης υπόψη μια τροποποίηση του ελέγχου του Bartlett που ερευνήθηκε από τους Boos και Brownie (1989). Ο τροποποιημένος στατιστικός έλεγχος είναι:

$$B_2 = dB_1 \tag{3.1}$$

όπου,

$$d = 2\big/(\hat{\beta}_2 - 1) \tag{3.2}$$

και

$$\hat{\beta}_2 = \frac{N\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^4}{\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2\right\}^2} \tag{3.3}$$

Η τροποποίηση είναι παρακινούμενη από το γεγονός ότι κάτω από αδύναμες υποθέσεις κανονικότητας *(weak regularity conditions)* $B_1 \to \frac{1}{2}(\beta_2 - 1)\chi^2_{k-1}$ σε κατανομή για την υπόθεση H_0, όπου $\beta_2 = E(X-\mu)^4 \big/ \sigma^4$ είναι η κύρτωση της κατανομής (Box, 1953). Το κρίσιμο σημείο για τον έλεγχο B_2 είναι το ίδιο όπως και στον έλεγχο B_1.

3.1.3.2 Έλεγχος των Box-Andersen

Η εκδοχή που λαμβάνουμε υπόψη είναι αυτή που προτείνεται στη μελέτη του Miller (1986). Ο στατιστικός έλεγχος είναι

$$B_3 = dM \tag{3.4}$$

όπου d δίνεται από την σχέση (3.2) και $M = (N-k)\log\left\{\sum_{i=1}^{k}(n_i-1)s_i^2 \Big/ (N-k)\right\} - \sum_{i=1}^{k}(n_i-1)\log s_i^2$. Η μηδενική υπόθεση απορρίπτεται όταν το B_3 είναι μεγαλύτερο από το $100(1-a)$ ποσοστιαίο στοιχείο (percentile) της χ^2 - κατανομής με $(k-1)$ βαθμούς ελευθερίας.

3.1.3.3 Jackknife έλεγχος

Η γενίκευση m-δειγμάτων του Layard (1973) από τον jackknife έλεγχο δύο δειγμάτων του Miller (1968) είναι η ανάλυση της διασποράς κατά ένα παράγοντα του F –ελέγχου βασιζόμενη στις $\log s_i^2$ ψευδοτιμές. Ο στατιστικός έλεγχος είναι:

$$J_1 = \frac{\sum_{i=1}^{k} n_i(\bar{u}_{i.} - \bar{u}_{..})^2 \Big/ (k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(u_{ij} - \bar{u}_{i.})^2 \Big/ (N-k)}, \tag{5}$$

όπου

$$u_{ij} = n_i \log s_i^2 - (n_i - 1)\log s_{i(j)}^2,$$

$$s_{i(j)}^2 = \sum_{i \neq j}(x_{ik} - \bar{x}_{i(j)})^2 \Big/ (n_i - 2),$$

$$\bar{x}_{i(j)} = \sum_{i \neq j} x_{ik} \Big/ (n_i - 1),$$

$$\bar{u}_{i.} = \sum_{j=1}^{n_i} u_{ij} \Big/ n_i,$$

$$\bar{u}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} u_{ij} \Big/ N ,$$

για $j = 1,...,n_i$, $i = 1,...,k$. Η μηδενική υπόθεση απορρίπτεται όταν το J_1 υπερβαίνει το $100(1-a)$ ποσοστιαίο στοιχείο (percentile) της F-κατανομής με $(k-1)$ και $(N-k)$ βαθμούς ελευθερίας.

Ο O'Brien (1978) προτείνει να χρησιμοποιηθούν οι εναλλακτικές jackknife ψευδοτιμές

$$q_{ij} = n_i s_i^2 - (n_i - 1) s_{i(j)}^2 , \qquad (6)$$

οι οποίες μπορούν να υπολογιστούν από τον τύπο

$$q_{ij} = \{ n_i (x_{ij} - \bar{x}_{i.})^2 - s_i^2 \} \Big/ (n_i - 2) . \qquad (7)$$

Ο στατιστικός έλεγχος που προκύπτει, ο οποίος είναι η ανάλυση της διασποράς κατά ένα παράγοντα του F – στατιστικού και βασίζεται στο q_{ij}, θα ονομάζεται J_2. Το κρίσιμο σημείο είναι το ίδιο όπως και στον έλεγχο J_1.

Ο Sharma (1991) προτείνει τις ακόλουθες τροποποιήσεις στην jackknife διαδικασία του Layard (1973). Βασίζεται στη χρήση της μεθόδου jackknife σε μια ομάδα παρατηρήσεων την κάθε φορά αντί της μιας παρατήρησης σε κάθε ομάδα. Οι ψευδοτιμές ορίζονται ως:

$$\xi_i = k \log s^2 - (k-1) \log s_{-i}^2 , \; i = 1,...,k , \qquad (8)$$

όπου

$$s^2 = \sum_{i=1}^{k} (n_i - 1) s_i^2 \Big/ \sum_{i=1}^{k} (n_i - 1) ,$$

και

$$s_{-i}^2 = \sum_{k \neq i} (n_k - 1) s_k^2 \Big/ \sum_{k \neq i} (n_k - 1) .$$

Ο στατιστικός έλεγχος είναι:

$$J_3 = \frac{k^{1/2}\overline{\xi}_.}{\left\{\sum_{i=1}^{k}(\xi_i - \overline{\xi}_.)^2 \big/ (k-1)\right\}^{1/2}}, \quad (9)$$

όπου $\overline{\xi}_. = \sum_{i=1}^{k} \xi_i / k$. Η μηδενική υπόθεση απορρίπτεται όταν η απόλυτη τιμή του ελέγχου J_3 είναι μεγαλύτερη από το $100(1-a/2)$ ποσοστιαίο στοιχείο (percentile) της t-κατανομής με $(k-1)$ βαθμούς ελευθερίας.

3.1.4 Bootstrap εκδοχές

Σε αυτή την παράγραφο περιγράφουμε την bootstrap διαδικασία της αναδειγματοληψίας. Η διαδικασία είναι παρόμοια με αυτή που χρησιμοποιήθηκε στους Boos και Brownie (1989) και εκτελείται όπως παρακάτω.

1. Υπολογίζεται η τιμή του στατιστικού ελέγχου T από τα παρατηρούμενα δεδομένα.

2. Θέτουμε αρχική τιμή $R = 0$.

3.Υπολογίζονται τα υπόλοιπα $e_{ij} = x_{ij} - \hat{\mu}_i$, $j = 1,...,n_i$, $i = 1,...,k$ όπου $\hat{\mu}_i$ είναι ο 20% κλασματικός αποκομμένος μέσος της ομάδας i.

4. Δημιουργούμε N bootstrap δείγματα e_{ij}^* με επαναθέσεις από το σύνολο των συγχωνευμένων (pooled) υπολοίπων

$$\mathscr{S} = \{e_{ij} : j = 1,\ldots,n_i;\ i = 1,\ldots,I\}.$$

5. Αν $n_i < 10$ για ένα τουλάχιστον δείγμα, ομαλοποιούνται τα bootstrap δείγματα θέτοντας $x_{ij}^* = (12/13)^{1/2}(e_{ij}^* + \upsilon U)$ όπου $\upsilon^2 = N^{-1}\sum\sum(x_{ij} - \overline{x}_{i.})^2$ και U είναι μια ανεξάρτητη ομοιόμορφη τυχαία μεταβλητή στο $(-\frac{1}{2}, \frac{1}{2})$. Διαφορετικά, ορίζεται $x_{ij}^* = e_{ij}^*$.

6. Υπολογίζεται μια πραγματοποίηση του bootstrap στατιστικού T^* από τα bootstrap δείγματα $\{x_{ij}^*\}$. Αν $T^* > T$, αυξάνεται και το R σε $R+1$.

7. Επαναλαμβάνονται τα βήματα 4, 5 και 6 A φορές (χρησιμοποιούμε $A = 1000$ στο πείραμα προσομοίωσης).

8. Η bootstrap P-τιμή δίνεται από το λόγο R/A. Η μηδενική υπόθεση απορρίπτεται αν το R/A είναι μικρότερο από το ονομαστικό επίπεδο (nominal level) a.

Όπως στους Boos και Brownie (1989), η ομαλοποίηση (smoothing) χρησιμοποιείται μεμονωμένα ως μια υπολογιστική συσκευή για να αποφευχθεί η λήψη δειγματικών διασπορών με μηδενικές τιμές. Η κλασματική αποκοπή χρησιμοποιείται για να υπολογιστεί η μέση τιμή $\hat{\mu}_i$ έτσι ώστε ο 20% κλασματικός αποκομμένος μέσος για $n = 5$ είναι $[X_{(2)} + X_{(3)} + X_{(4)}]/3$ ενώ για $n = 4$ είναι $[0.2X_{(1)} + X_{(2)} + X_{(3)} + 0.2X_{(4)}]/2.4$, όπου $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ είναι το διατεταγμένο δείγμα.

3.1.5 Αποτελέσματα της προσομοίωσης

Η ευρωστία του επιπέδου σημαντικότητας και η ισχύς των ελέγχων συγκρίνονται σε ένα πείραμα προσομοίωσης χρησιμοποιώντας 5 κατανομές και 6 διαφορετικά μεγέθη των ομάδων. Οι κατανομές είναι: (i) η κανονική (ii) η t-κατανομή με 5 βαθμούς ελευθερίας, (iii) η ομοιόμορφη, (iv) η εκθετική και (v) η λογαριθμοκανονική. Αυτές οι κατανομές παρουσιάζουν ένα εύρος από συμμετρικές σε ασύμμετρες, από ελαφριές σε παχιές ουρές και από μικρή σε μεγάλη κύρτωση. 3 από τα μεγέθη των ομάδων είναι ισορροπημένα και 3 όχι. Η μηδενική υπόθεση των ίσων διασπορών [όλες είναι ίσες με μονάδα, (1)] μελετήθηκε μαζί με τρεις εναλλακτικές $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (1, 6, 11, 16), (16, 11, 6, 1)$ και $(1, 1, 1, 16)$. Τα αποτελέσματα όσον αφορά την αναλογία των περιπτώσεων που η μηδενική υπόθεση απορρίφθηκε από 1000 Monte Carlo επαναλήψεις συγκεντρώνονται στον Πίνακα 1. Οι εισαγωγές που υπερβαίνουν την τιμή 0.10 είναι υπογραμμισμένες. Δηλώνουμε κάθε *bootstrap* εκδοχή προσθέτοντας το γράμμα “B” μπροστά από το όνομα του ελέγχου. Σημειώνεται ότι η *bootstrap* εκδοχή του ελέγχου Box-Andersen είναι η ίδια με την *bootstrap* εκδοχή του τροποποιημένου ελέγχου του Bartlett (BB_2). Δεν καταγράφουμε τα αποτελέσματα της προσομοίωσης για την *bootstrap* εκδοχή του jackknife ελέγχου J_3 διότι η ισχύς που υπολογίστηκε για όλα τα διαφορετικά μεγέθη που χρησιμοποιήθηκαν είναι σχεδόν μηδέν.

Από την στιγμή πού οι μετρήσεις του επιπέδου σημαντικότητας που δίνονται στον Πίνακα 3.1 αποκομίστηκαν χρησιμοποιώντας 1000 Monte Carlo επαναλήψεις για τον κάθε έλεγχο, το τυπικό σφάλμα είναι περίπου 0.007 όταν η μέτρηση είναι κοντά στο 0.05 και φτάνει σε μέγιστη τιμή 0.016 όταν η μέτρηση είναι κοντά στο 0.5. Ακολουθώντας τον Conover και άλλους (1981), θεωρούμε έναν έλεγχο ότι είναι εύρωστος αν το μέγιστο εκτιμώμενο επίπεδο στατιστικής σημαντικότητάς του, πέρα των 30 μηδενικών υποθέσεων (ίσων διασπορών) που μελετηθήκαν δεν υπερβαίνει την τιμή 0.10, που είναι διπλάσια από το ονομαστικό (nominal) του επίπεδο 0.05. Ο Πίνακας 3.2 δίνει το μέγιστο εκτιμώμενο επίπεδο στατιστικής σημαντικότητας και το πλήθος των περιπτώσεων όπου το εκτιμώμενο επίπεδο δεν υπερβαίνει την τιμή 0.10 πέρα των 30 μηδενικών υποθέσεων για κάθε έλεγχο. Οι Πίνακες 3.3 και 3.4 δίνουν τις μετρήσεις της ισχύος για τους μη-*bootstrap* και *bootstrap* ελέγχους αντίστοιχα.

Πίνακας 3.1.

Monte Carlo εκτιμήσεις του επιπέδου σημαντικότητας ($a = 0,05$) για την μηδενική υπόθεση βασισμένες σε 1000 επαναλήψεις. Είναι υπογραμμισμένα τα εκτιμώμενα επίπεδα σημαντικότητας που υπερβαίνουν την τιμή 0.10.

Distribution	Sample sizes	L	B_1	B_2	B_3	J_1	J_2	J_3	BL	BB_1	BB_2	BJ_1	BJ_2
Normal	(5,5,5,5)	0.000	0.047	0.052	0.073	0.040	0.015	0.050	0.037	0.017	0.037	0.043	0.036
	(5,5,10,10)	0.028	0.049	0.048	0.057	0.053	0.020	0.054	0.052	0.028	0.040	0.042	0.040
	(10,10,10,10)	0.029	0.049	0.045	0.052	0.049	0.023	0.051	0.049	0.028	0.039	0.044	0.045
	(5,10,15,20)	0.028	0.056	0.055	0.067	0.064	0.041	0.063	0.049	0.036	0.051	0.047	0.053
	(5,5,20,20)	0.022	0.069	0.060	0.075	0.083	0.051	0.067	0.047	0.044	0.054	0.048	0.068
	(20,20,20,20)	0.033	0.046	0.049	0.051	0.051	0.041	0.060	0.049	0.038	0.049	0.054	0.048
t_5	(5,5,5,5)	0.005	0.139	0.061	0.080	0.056	0.010	0.059	0.044	0.032	0.043	0.057	0.031
	(5,5,10,10)	0.024	0.156	0.044	0.056	0.059	0.017	0.079	0.056	0.036	0.028	0.034	0.040
	(10,10,10,10)	0.046	0.218	0.044	0.056	0.079	0.026	0.069	0.064	0.043	0.036	0.054	0.049
	(5,10,15,20)	0.024	0.227	0.046	0.053	0.114	0.032	0.083	0.050	0.051	0.044	0.064	0.042
	(5,5,20,20)	0.034	0.194	0.040	0.052	0.100	0.041	0.090	0.051	0.048	0.037	0.053	0.050
	(20,20,20,20)	0.039	0.249	0.043	0.046	0.084	0.035	0.074	0.047	0.043	0.041	0.052	0.046
Uniform	(5,5,5,5)	0.002	0.014	0.055	0.078	0.019	0.010	0.050	0.021	0.010	0.034	0.027	0.049
	(5,5,10,10)	0.018	0.010	0.067	0.085	0.044	0.019	0.041	0.044	0.014	0.050	0.045	0.041
	(10,10,10,10)	0.026	0.005	0.052	0.060	0.028	0.029	0.059	0.050	0.019	0.044	0.038	0.048
	(5,10,15,20)	0.026	0.010	0.073	0.082	0.057	0.033	0.057	0.047	0.018	0.057	0.052	0.057
	(5,5,20,20)	0.033	0.011	0.092	0.115	0.066	0.051	0.058	0.055	0.023	0.061	0.048	0.059
	(20,20,20,20)	0.020	0.002	0.041	0.044	0.022	0.025	0.056	0.038	0.020	0.037	0.034	0.036
Exponential	(5,5,5,5)	0.015	0.298	0.126	0.170	0.099	0.059	0.083	0.083	0.089	0.079	0.088	0.094
	(5,5,10,10)	0.030	0.348	0.102	0.127	0.129	0.047	0.075	0.070	0.086	0.064	0.066	0.064
	(10,10,10,10)	0.043	0.397	0.090	0.107	0.128	0.049	0.077	0.051	0.056	0.049	0.059	0.050
	(5,10,15,20)	0.030	0.419	0.068	0.087	0.147	0.048	0.070	0.045	0.066	0.041	0.061	0.050
	(5,5,20,20)	0.031	0.357	0.057	0.076	0.136	0.070	0.092	0.051	0.071	0.039	0.051	0.062
	(20,20,20,20)	0.046	0.464	0.065	0.069	0.116	0.050	0.068	0.048	0.043	0.036	0.051	0.053
Lognormal	(5,5,5,5)	0.013	0.556	0.129	0.178	0.121	0.042	0.157	0.065	0.169	0.059	0.085	0.064
	(5,5,10,10)	0.036	0.649	0.106	0.132	0.175	0.039	0.155	0.058	0.159	0.058	0.076	0.054
	(10,10,10,10)	0.037	0.705	0.099	0.111	0.171	0.028	0.151	0.049	0.057	0.026	0.048	0.034
	(5,10,15,20)	0.043	0.728	0.074	0.092	0.187	0.062	0.168	0.054	0.129	0.040	0.062	0.051
	(5,5,20,20)	0.033	0.689	0.056	0.070	0.192	0.079	0.175	0.047	0.114	0.035	0.048	0.045
	(20,20,20,20)	0.029	0.814	0.034	0.042	0.183	0.016	0.139	0.039	0.031	0.018	0.041	0.019

Πίνακας 3.2.

Μέγιστο εκτιμώμενο επίπεδα στατιστικής σημαντικότητας και πλήθος των περιπτώσεων όπου τα εκτιμώμενα επίπεδα δεν υπερβαίνουν την τιμή 0.10 για τις 30 μηδενικές υποθέσεις που μελετήθηκαν στον Πίνακα 3.1.

Test	Maximum estimated significance level	Number of times the estimated significance level does not exceed 0.10	Robust?
Non-bootstrap			
L	0.046	30	Yes
B_1	0.814	12	
B_2	0.129	27	
B_3	0.178	22	
J_1	0.192	19	
J_2	0.079	30	Yes
J_3	0.175	24	
Bootstrap			
BL	0.083	30	Yes
BB_1	0.169	26	
BB_2	0.079	30	Yes
BJ_1	0.088	30	Yes
BJ_2	0.094	30	Yes

3.1.5.1 Μη- Bootstrap έλεγχοι

Τα αποτελέσματα δείχνουν ότι ο έλεγχος του Levene (L) είναι εύρωστος. Ωστόσο είναι λίγο συντηρητικός, επειδή το εκτιμώμενο επίπεδο σημαντικότητας είναι χαμηλότερο από το ονομαστικό (nominal) επίπεδο 0.05 για όλες τις συνθήκες με μέγιστη τιμή 0.046. Είναι εξαιρετικά συντηρητικό για μεγέθη δειγμάτων $(5,5,5,5)$ για όλες τις κατανομές, ένα φαινόμενο το οποίο έχει παρατηρηθεί προηγουμένως από τους Boos και Brownie (1989) και Conover et al. (1981). Όπως σημειώνεται στον Loh (1987) η ισχύς είναι μεγάλη όταν μεγάλα n_i σχετίζονται με μικρά σ_i^2 και μικρή όταν μεγάλα n_i σχετίζονται με μεγάλα σ_i^2.

Ο έλεγχος του Bartlett (B_1) είναι εξαιρετικά μη-εύρωστος. Το μέγιστο εκτιμώμενο επίπεδο στατιστικής σημαντικότητας γι' αυτόν τον έλεγχο είναι 0.814. Ωστόσο, για την κανονική κατανομή, τα επίπεδα σημαντικότητάς του είναι αρκετά μικρότερα του 0.10 και η ισχύς του είναι αρκετά μεγάλη. Είναι πολύ συντηρητικός μολονότι είναι ακόμη ισχυρός για την ομοιόμορφη κατανομή (μικρή κύρτωση) αλλά υπερβολικά εύχρηστος για τις κατανομές t_5, εκθετική και λογαριθμοκανονική (μεγάλη κύρτωση). Ο τροποποιημένος έλεγχος του Bartlett (B_2) λειτουργεί καλύτερα. Παρόλο που είναι ακόμη μη-εύρωστος, το μέγιστο εκτιμώμενο επίπεδο σημαντικότητάς του είναι μόλις 0.129. Για την κανονική κατανομή τα επίπεδα σημαντικότητάς του είναι αρκετά μικρότερα του 0.10 αλλά η ισχύς του δεν είναι τόσο μεγάλη όπως του (B_1). Για άλλες συμμετρικές κατανομές, τα επίπεδα σημαντικότητας του (B_2) είναι μικρότερα του 0.10. Όταν το μέγεθος του δείγματος είναι μεγάλο, τα εκτιμώμενα επίπεδα σημαντικότητάς του δεν υπερβαίνουν την τιμή 0.10 για ασύμμετρες κατανομές. Σε αυτές τις περιπτώσεις όπου τα επίπεδα σημαντικότητάς του δεν υπερβαίνουν το 0.10, η ισχύς του είναι συνήθως μεγαλύτερη από αυτή του ελέγχου του Levene. Ο έλεγχος των Box-Andersen (B_3), έτερη παραλλαγή του ελέγχου του Bartlett, είναι επίσης μη-εύρωστος με μέγιστο εκτιμώμενο επίπεδο σημαντικότητας 0.178. Είναι πολύ εύχρηστος για ασύμμετρες κατανομές εκτός όταν το μέγεθος του δείγματος είναι μεγάλο. Τα επίπεδα σημαντικότητάς του είναι μικρότερα του 0.10 για την κανονική κατανομή, την t_5 και την ομοιόμορφη, εκτός από τα μεγέθη δειγμάτων (5,5,20,20) για την ομοιόμορφη κατανομή. Η ισχύς του (B_3) είναι ελάχιστα μεγαλύτερη από ότι αυτή του (B_2) σε αυτές τις περιπτώσεις όπου το επίπεδο σημαντικότητας του (B_3) είναι μικρότερο από 0,10.

Μεταξύ των τριών jackknife ελέγχων που εξετάστηκαν, μόνο ο jackknife έλεγχος J_2 είναι εύρωστος με μέγιστο εκτιμώμενο επίπεδο σημαντικότητας 0.079. Είναι σχετικά συντηρητικός με μικρά μεγέθη δειγμάτων που ακολουθούν συμμετρικές κατανομές. Τα εκτιμώμενα επίπεδα σημαντικότητας των jackknife ελέγχων J_1 και J_3 είναι μικρότερα από 0,10 για συμμετρικές κατανομές [εκτός του J_1 που ακολουθεί την t_5 κατανομή με μεγέθη δειγμάτων $(5,10,15,20)$] αλλά είναι μεγαλύτερα από 0,10 για την λογαριθμοκανονική κατανομή. Τα επίπεδα σημαντικότητας του J_3, αλλά όχι του J_1, είναι μικρότερα από 0,10 για την εκθετική κατανομή. Σημειώνεται ότι ο J_2 είναι συνήθως λιγότερο ισχυρός από τον έλεγχο του Levene. Όπως ισχύει και στην περίπτωση του ελέγχου του Levene, η ισχύς του J_2 είναι πολύ μικρή όταν μεγάλα n_i σχετίζονται με μεγάλες διασπορές σ_i^2. Η ισχύς του J_1 είναι γενικά ισοδύναμη με αυτή του B_2 αλλά η ισχύς του J_3 διαφέρει αρκετά. Ο έλεγχος J_3 έχει μικρή ισχύ στην εναλλακτική $(\sigma_1^2, \sigma_2^2, \sigma_2^3, \sigma_4^2) = (1,1,1,16)$ στην περίπτωση με ίσα μεγέθη δειγμάτων. Ωστόσο συχνά έχει την υψηλότερη ισχύ στο εναλλακτικό $(\sigma_1^2, \sigma_2^2, \sigma_2^3, \sigma_4^2) = (1,6,11,16)$.

Βασιζόμενοι στα αποτελέσματα της προσομοίωσης, συμπεραίνουμε ότι ανάμεσα στους 7 μη-*bootstrap* ελέγχους, ο έλεγχος του Levene και ο jackknife έλεγχος J_2 είναι οι καλύτεροι όσον αφορά την ευρωστία του επιπέδου σημαντικότητας, με τον έλεγχο του Levene να είναι, επιπλέον, ο πιο ισχυρός.

__Πίνακας 3.3.__

Μη-bootstrap έλεγχοι. Monte Carlo εκτιμήσεις της ισχύος ($a = 0,05$ ***) βασισμένες σε 1000 επαναλήψεις.***

Distribution	Sample sizes	Variances	L	B_1	B_2	B_3	J_1	J_2	J_3
Normal	(5,5,5,5)	(1,6,11,16)	0.041	0.527	0.367	0.462	0.312	0.054	0.744
		(1,1,1,16)	0.294	0.794	0.291	0.362	0.398	0.217	0.006
	(5,5,10,10)	(1,6,11,16)	0.216	0.617	0.474	0.536	0.494	0.070	0.939
		(16,11,6,1)	0.347	0.852	0.683	0.748	0.697	0.285	0.652
	(10,10,10,10)	(1,6,11,16)	0.648	0.959	0.838	0.861	0.862	0.337	0.927
		(1,1,1,16)	0.918	0.985	0.817	0.834	0.907	0.816	0.000
	(5,10,15,20)	(1,6,11,16)	0.401	0.735	0.637	0.674	0.711	0.136	0.983
		(16,11,6,1)	0.893	0.994	0.963	0.973	0.980	0.795	0.542
	(5,5,20,20)	(1,6,11,16)	0.332	0.681	0.587	0.643	0.694	0.078	0.996
		(16,11,6,1)	0.851	0.990	0.943	0.958	0.970	0.787	0.454
	(20,20,20,20)	(1,6,11,16)	0.995	1.000	0.998	0.998	1.000	0.916	0.985
		(1,1,1,16)	1.000	1.000	0.993	0.994	1.000	0.994	0.000
t_5	(5,5,5,5)	(1,6,11,16)	0.059		0.315	0.397	0.305	0.039	0.627
		(1,1,1,16)	0.244		0.268	0.346	0.375	0.157	0.015
	(5,5,10,10)	(1,6,11,16)	0.153		0.352	0.416	0.453	0.036	0.844
		(16,11,6,1)	0.280		0.525	0.584	0.594	0.174	0.500
	(10,10,10,10)	(1,6,11,16)	0.491		0.632	0.664	0.713	0.184	0.828
		(1,1,1,16)	0.861		0.709	0.747	0.788	0.636	0.008
	(5,10,15,20)	(1,6,11,16)	0.258		0.390	0.418		0.068	0.943
		(16,11,6,1)	0.763		0.797	0.829		0.561	0.348
	(5,5,20,20)	(1,6,11,16)	0.187		0.319	0.371	0.557	0.038	0.974
		(16,11,6,1)	0.719		0.773	0.804	0.799	0.611	0.270
	(20,20,20,20)	(1,6,11,16)	0.942		0.911	0.916	0.943	0.568	0.938
		(1,1,1,16)	0.997		0.945	0.952	0.955	0.933	0.003
Uniform	(5,5,5,5)	(1,6,11,16)	0.046	0.399	0.479	0.551	0.330	0.108	0.860
		(1,1,1,16)	0.335	0.837	0.361	0.437	0.447	0.344	0.000
	(5,5,10,10)	(1,6,11,16)	0.343	0.508	0.753	0.807	0.639	0.163	0.989
		(16,11,6,1)	0.356	0.899	0.863	0.892	0.908	0.429	0.821
	(10,10,10,10)	(1,6,11,16)	0.726	0.983	0.971	0.978	0.981	0.633	0.980
		(1,1,1,16)	0.978	0.998	0.931	0.947	0.993	0.955	0.000

Πίνακας 3.3. (συνέχεια)

Distribution	Sample sizes	Variances	L	B_1	B_2	B_3	J_1	J_2	J_3
Uniform	(5,10,15,20)	(1,6,11,16)	0.647	0.709	0.933	0.945	0.883	0.528	0.999
		(16,11,6,1)	0.942	1.000	0.998	0.999	0.999	0.964	0.722
	(5,5,20,20)	(1,6,11,16)	0.634	0.609	0.925		0.855	0.389	1.000
		(16,11,6,1)	0.917	0.999	0.995		1.000	0.946	0.620
	(20,20,20,20)	(1,6,11,16)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1,1,1,16)	1.000	1.000	1.000	1.000	1.000	1.000	0.000
Exponential	(5,5,5,5)	(1,6,11,16)	0.057				0.279	0.076	0.462
		(1,1,1,16)	0.181				0.314	0.169	0.029
	(5,5,10,10)	(1,6,11,16)	0.116					0.053	0.746
		(16,11,6,1)	0.188					0.180	0.394
	(10,10,10,10)	(1,6,11,16)	0.324		0.549			0.191	0.698
		(1,1,1,16)	0.663		0.574			0.483	0.012
	(5,10,15,20)	(1,6,11,16)	0.129		0.298	0.333		0.038	0.889
		(16,11,6,1)	0.577		0.641	0.683		0.471	0.288
	(5,5,20,20)	(1,6,11,16)	0.150		0.323	0.360		0.054	0.873
		(16,11,6,1)	0.554		0.593	0.654		0.511	0.232
	(20,20,20,20)	(1,6,11,16)	0.765		0.746	0.764		0.359	0.867
		(1,1,1,16)	0.970		0.828	0.837		0.791	0.003
Lognormal	(5,5,5,5)	(1,6,11,16)	0.045					0.078	
		(1,1,1,16)	0.106					0.127	
	(5,5,10,10)	(1,6,11,16)	0.069					0.046	
		(16,11,6,1)	0.140					0.099	
	(10,10,10,10)	(1,6,11,16)	0.169		0.308			0.091	
		(1,1,1,16)	0.418		0.443			0.245	
	(5,10,15,20)	(1,6,11,16)	0.060		0.158	0.182		0.034	
		(16,11,6,1)	0.322		0.364	0.398		0.250	
	(5,5,20,20)	(1,6,11,16)	0.050		0.125	0.150		0.034	
		(16,11,6,1)	0.308		0.353	0.392		0.330	
	(20,20,20,20)	(1,6,11,16)	0.373		0.414	0.431		0.136	
		(1,1,1,16)	0.785		0.573	0.584		0.443	

3.1.5.2 Bootstrap έλεγχοι

Τα επίπεδα σημαντικότητας των *bootstrap* εκδοχών όλων των ελέγχων, με εξαίρεση αυτόν του Bartlett, BB_1, είναι εύρωστοι. Το μέγιστο εκτιμώμενο επίπεδο σημαντικότητας των ελέγχων BL, BB_2, BJ_1 και BJ_2 είναι 0.083, 0.079, 0.088 και 0.094, αντίστοιχα. Τα εκτιμώμενα επίπεδα σημαντικότητας του BB_1 είναι μικρότερα από 0,090 εκτός όταν η κατανομή είναι λογαριθμοκανονική και το μέγεθος του δείγματος είναι μικρό, ή όταν ο σχεδιασμός είναι μη-ισορροπημένος.

Η ισχύς του BB_1 είναι μεγάλη, αρκετά συχνά γίνεται η μεγαλύτερη. Η ισχύς του BL είναι μεγαλύτερη από αυτές των BB_2 και BJ_1 στην εναλλακτική υπόθεση διασπορών $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (1,1,1,16)$ για όλες τις κατανομές και τα μεγέθη των δειγμάτων και έχουν ισοδύναμη ισχύ με όλους τους εναλλακτικούς ελέγχους. Ο BJ_2, όπως και στην μη-bootstrap εκδοχή του, έχει πολύ μικρή ισχύ όταν μεγάλα n_i σχετίζονται με μεγάλα σ_i^2.

Καταλήγουμε ότι ο BL είναι προτιμότερος για όλες τις κατανομές και τα διαφορετικά μεγέθη των δειγμάτων. Όταν η κατανομή δεν έχει πολύ μεγάλη κύρτωση, ο BB_1 είναι μια ακόμη επιλογή.

Ερευνήσαμε επίσης την bootstrap διαδικασία αναδειγματοληψίας η οποία απεικονίζει τα *bootstrap* δείγματα από το σύνολο των συγχωνευμένων (pooled) τυποποιημένων υπολοίπων $\mathcal{S} = \{e_{ij} = (x_{ij} - \bar{x}_{i.})/s_i :\ j = 1,\ldots,n_i;\ i = 1,\ldots,I\}$. Ωστόσο δεν καταγράψαμε τα αποτελέσματα της προσομοίωσης από την στιγμή που αποδεικνύεται ότι πολλοί από τους ελέγχους δεν είναι εύρωστοι με τη χρήση της μεθόδου *bootstrap*.

<u>*Πίνακας 3.4.*</u>

Έλεγχοι Bootstrap. Monte Carlo εκτιμήσεις της ισχύος ($a = 0.05$ ***) βασισμένες σε 1000 επαναλήψεις. Εκτιμώμενες τιμές ισχύος για τις οποίες το αντίστοιχο εκτιμώμενο επίπεδο σημαντικότητας υπερβαίνει το 0.10, παραλείπονται.***

Distribution	Sample sizes	Variances	BL	BB_1	BB_2	BJ_1	BJ_2
Normal	(5,5,5,5)	(1,6,11,16)	0.251	0.248	0.252	0.287	0.149
		(1,1,1,16)	0.545	0.495	0.163	0.313	0.308
	(5,5,10,10)	(1,6,11,16)	0.370	0.395	0.403	0.416	0.132
		(16,11,6,1)	0.511	0.670	0.611	0.577	0.438
	(10,10,10,10)	(1,6,11,16)	0.708	0.790	0.781	0.788	0.470
		(1,1,1,16)	0.939	0.888	0.734	0.799	0.864
	(5,10,15,20)	(1,6,11,16)	0.536	0.618	0.611	0.629	0.176
		(16,11,6,1)	0.928	0.975	0.964	0.946	0.846
	(5,5,20,20)	(1,6,11,16)	0.492	0.560	0.556	0.567	0.108
		(16,11,6,1)	0.904	0.962	0.955	0.930	0.826
	(20,20,20,20)	(1,6,11,16)	0.996	1.000	0.998	0.999	0.952
		(1,1,1,16)	1.000	1.000	0.994	1.000	0.997
t_5	(5,5,5,5)	(1,6,11,16)	0.208	0.279	0.199	0.265	0.104
		(1,1,1,16)	0.497	0.480	0.124	0.303	0.245
	(5,5,10,10)	(1,6,11,16)	0.287	0.375	0.280	0.346	0.086
		(16,11,6,1)	0.437	0.575	0.461	0.443	0.309
	(10,10,10,10)	(1,6,11,16)	0.557	0.632	0.541	0.590	0.270
		(1,1,1,16)	0.893	0.816	0.559	0.629	0.716
	(5,10,15,20)	(1,6,11,16)	0.349	0.494	0.365	0.431	0.091
		(16,11,6,1)	0.827	0.893	0.812	0.725	0.626
	(5,5,20,20)	(1,6,11,16)	0.284	0.441	0.312	0.396	0.048
		(16,11,6,1)	0.804	0.859	0.792	0.663	0.633
	(20,20,20,20)	(1,6,11,16)	0.954	0.952	0.917	0.902	0.650
		(1,1,1,16)	0.999	0.987	0.943	0.904	0.957
Uniform	(5,5,5,5)	(1,6,11,16)	0.236	0.233	0.374	0.339	0.230
		(1,1,1,16)	0.604	0.533	0.234	0.384	0.472
	(5,5,10,10)	(1,6,11,16)	0.530	0.484	0.656	0.605	0.324
		(16,11,6,1)	0.557	0.781	0.831	0.833	0.573
	(10,10,10,10)	(1,6,11,16)	0.789	0.921	0.953	0.961	0.727
		(1,1,1,16)	0.989	0.987	0.917	0.979	0.980

Πίνακας 3.4. *(συνέχεια)*

Distribution	Sample sizes	Variances	BL	BB_1	BB_2	BJ_1	BJ_2
Uniform	(5,10,15,20)	(1,6,11,16)	0.759	0.816	0.913	0.864	0.609
		(16,11,6,1)	0.971	0.999	0.999	0.999	0.981
	(5,5,20,20)	(1,6,11,16)	0.776	0.774	0.903	0.800	0.463
		(16,11,6,1)	0.964	0.996	0.998	0.997	0.959
	(20,20,20,20)	(1,6,11,16)	1.000	1.000	1.000	1.000	1.000
		(1,1,1,16)	1.000	1.000	1.000	1.000	1.000
Exponential	(5,5,5,5)	(1,6,11,16)	0.185	0.332	0.212	0.236	0.142
		(1,1,1,16)	0.385	0.506	0.202	0.240	0.230
	(5,5,10,10)	(1,6,11,16)	0.203	0.342	0.228	0.259	0.072
		(16,11,6,1)	0.293	0.539	0.321	0.293	0.238
	(10,10,10,10)	(1,6,11,16)	0.356	0.494	0.362	0.374	0.194
		(1,1,1,16)	0.714	0.617	0.407	0.401	0.511
	(5,10,15,20)	(1,6,11,16)	0.177	0.381	0.228	0.288	0.038
		(16,11,6,1)	0.646	0.786	0.596	0.482	0.466
	(5,5,20,20)	(1,6,11,16)	0.199	0.408	0.262	0.311	0.053
		(16,11,6,1)	0.632	0.754	0.578	0.438	0.480
	(20,20,20,20)	(1,6,11,16)	0.790	0.802	0.702	0.633	0.368
		(1,1,1,16)	0.978	0.949	0.804	0.730	0.850
Lognormal	(5,5,5,5)	(1,6,11,16)	0.151		0.137	0.152	0.116
		(1,1,1,16)	0.251		0.174	0.211	0.169
	(5,5,10,10)	(1,6,11,16)	0.121		0.114	0.162	0.060
		(16,11,6,1)	0.223		0.214	0.186	0.155
	(10,10,10,10)	(1,6,11,16)	0.191	0.264	0.135	0.187	0.088
		(1,1,1,16)	0.492	0.458	0.219	0.234	0.251
	(5,10,15,20)	(1,6,11,16)	0.071		0.097	0.124	0.029
		(16,11,6,1)	0.387		0.287	0.204	0.238
	(5,5,20,20)	(1,6,11,16)	0.072		0.078	0.111	0.024
		(16,11,6,1)	0.357		0.310	0.220	0.250
	(20,20,20,20)	(1,6,11,16)	0.434	0.450	0.292	0.291	0.151
		(1,1,1,16)	0.829	0.686	0.456	0.358	0.485

3.1.6 Συμπεράσματα

Έχουμε συγκρίνει 7 ελέγχους της ισότητας των διασπορών και τις *bootstrap* εκδοχές τους με όρους την ευρωστία του επίπεδου σημαντικότητας (την δυνατότητα δηλαδή να διατηρούν το ονομαστικό (nominal) επίπεδο του ελέγχου κάτω από μια ποικιλία από διαφορετικές επιλογές) και την ισχύ (την δυνατότητα να απορρίπτουν την υπόθεση H_0 όταν είναι λανθασμένη) σε ένα πείραμα προσομοίωσης. Λάβαμε υπόψη ότι ένας έλεγχος είναι εύρωστος αν το μέγιστο εκτιμώμενο επίπεδο σημαντικότητάς του για πάνω από 30 μηδενικές υποθέσεις (με ίσες διασπορές) που μελετήθηκαν δεν υπερβαίνει την τιμή 0,10.

Τα αποτελέσματα της προσομοίωσης μπορούν να ανακεφαλαιωθούν παρακάτω:

1. Ο έλεγχος Levene, ο jackknife έλεγχος J_2 και οι *bootstrap* εκδοχές του ελέγχου Levene, του ελέγχου Bartlett με προσαρμογή της κύρτωσης, και των jackknife ελέγχων J_1 και J_2 είναι εύρωστοι. Μεταξύ των εύρωστων ελέγχων, η bootstrap εκδοχή του ελέγχου του Levene τείνει να έχει την μεγαλύτερη ισχύ.

2. Μεταξύ των 7 μη-*bootstrap* ελέγχων, ο έλεγχος του Levene και ο jackknife έλεγχος J_2 είναι οι καλύτεροι όσον αφορά την ευρωστία του επιπέδου σημαντικότητας, με τον έλεγχο του Levene να είναι ο πιο ισχυρός. Οι άλλοι 5 έλεγχοι δεν μπορούν να προταθούν για γενική χρήση.

3. Ο έλεγχος του Levene είναι σχετικά συντηρητικός, ειδικά με μικρές, ίσες, και περιττές τιμές του n_i. Η μέθοδος bootstrap βοηθάει στο να φέρει το επίπεδο σημαντικότητας του ελέγχου του Levene πιο κοντά στο ονομαστικό (nominal) επίπεδο και να αυξήσει την ισχύ του ελέγχου.

4. Η χρήση της μεθόδου bootstrap στον έλεγχο Bartlett προσφέρει μια σημαντική βελτίωση όσον αφορά την ευρωστία σε σχέση από την αρχική εκδοχή του ελέγχου, ειδικά για συμμετρικές κατανομές. Επίσης, με τη χρήση της μεθόδου bootstrap στον έλεγχο Bartlett με προσαρμογή κύρτωσης (και γι’ αυτό το λόγο επίσης χρήση της μεθόδου bootstrap στον έλεγχο των Box-Andersen) προκύπτει ένας εύρωστος και πιο ισχυρός έλεγχος.

5. Ο jackknife έλεγχος J_1 μπορεί να γίνει εύρωστος με τη βοήθεια της μεθόδου bootstrap. Η bootstrap εκδοχή του jackknife ελέγχου BJ_2 είναι επίσης εύρωστη. Σε αντίθεση με τη μη-bootstrap εκδοχή, η ισχύς του BJ_2 είναι πολύ μικρή όταν μεγάλα n_i σχετίζονται με μεγάλα σ_i^2. Από την άλλη πλευρά, η χρήση της μεθόδου bootstrap δεν βοηθάει τον jackknife έλεγχο J_3.

Επειδή τα συμπεράσματά μας βασίζονται σε ένα σύνολο ενός πειράματος προσομοίωσης, οι γενικεύσεις χρειάζονται προσοχή. Για παράδειγμα, τα λειτουργικά χαρακτηριστικά των ελέγχων μπορεί να αλλάξουν όταν υπάρχουν παραπάνω από τέσσερις ομάδες. Επίσης, τα συμπεράσματα μπορεί να μην είναι εφαρμόσιμα αν οι αληθείς διασπορές είναι πολύ διαφορετικές από αυτές που εξετάστηκαν εδώ. Τελικά, τα αποτελέσματά μας ίσως να μην επεκτείνονται σε περιπτώσεις όπου τα δεδομένα δεν προέρχονται από μια οικογένεια θέσης-κλίμακας (location-scale family).

Εντός των ορίων της μελέτης, προτείνουμε την *bootstrap* εκδοχή του ελέγχου του Levene με τη διάμεσο του δείγματος ως μια εκτίμηση της παραμέτρου θέσης. Αν η κατανομή δεν έχει πολύ μεγάλη κύρτωση, η *bootstrap* εκδοχή του ελέγχου του Bartlett είναι μια άλλη εναλλακτική. Για να εφαρμόσουμε τους ελέγχους στην πράξη, είναι αρκετά χρήσιμο να έχουμε ένα ισορροπημένο σχεδιασμό με μεγάλα μεγέθη δειγμάτων (π.χ., τουλάχιστον 10 σε κάθε ομάδα) για να πετύχουμε ικανοποιητική ισχύ.

3.2 Ένας jackknife έλεγχος για την ομοιογένεια των διασπορών

3.2.1 Περίληψη

Ένας διαφορετικός jackknife έλεγχος θα μελετηθεί για τον έλεγχο της ισότητας των διασπορών σε διάφορους πληθυσμούς. Αυτός ο έλεγχος παρουσιάστηκε στο κεφάλαιο 2 (παρ. 2.2.4.3) και είναι βασισμένος στη χρήση της μεθόδου jackknife σε μια ομάδα παρατηρήσεων την κάθε φορά, αντί για μια παρατήρηση σε κάθε ομάδα όπως έχει προταθεί από τον Miller για την περίπτωση που έχουμε δυο δείγματα και από τον Layard για αρκετά δείγματα. Εξετάζεται και συγκρίνεται με άλλους ελέγχους όσο αφορά την ισχύ και την ευρωστία, λαμβάνοντας υπόψη και μια μεγάλη ποικιλία από μη-κανονικές κατανομές. Βρέθηκε ότι ο έλεγχος αυτός είναι εύρωστος και έχει μεγάλη ισχύ για κανονικές καθώς επίσης και για μη-κανονικές παρατηρήσεις, ανεξάρτητα από το μέγεθος του δείγματος. Επιπλέον, είναι καλύτερος από όλους τους άλλους ελέγχους που εξετάζονται σ’ αυτή τη συγκριτική μελέτη για μικρά έως μέτρια μεγέθη δειγμάτων και είναι ίσως και καλύτερος ακόμη και για μεγάλα δείγματα έναντι των άλλων ελέγχων, ανεξάρτητα από τις κατανομές που ακολουθούν οι παρατηρήσεις των δειγμάτων.

3.2.2 Εισαγωγή

Οι jackknife έλεγχοι των Miller και Layard είναι βασισμένοι στη χρήση της μεθόδου jackknife σε μια παρατήρηση την κάθε φορά, π.χ., εάν υπάρχουν τέσσερις ομάδες από είκοσι παρατηρήσεις η καθεμία, τότε μια χρειάζεται να υπολογίσει ογδόντα δειγματικές διασπορές και ογδόντα ψευδοτιμές. Ένας διαφορετικός jackknife έλεγχος εξετάζεται, βασισμένος στη χρήση της μεθόδου jackknife σε μια ομάδα παρατηρήσεων την κάθε φορά, αντί για μια παρατήρηση σε κάθε ομάδα. Ο προτεινόμενος έλεγχος εξετάζεται και συγκρίνεται με: τον έλεγχο του Bartlett χρησιμοποιώντας κρίσιμες τιμές από την ασυμπτωτική κατανομή του στατιστικού ελέγχου, τον έλεγχο του Bartlett χρησιμοποιώντας κρίσιμες τιμές που δίνονται από τους Dyer και Keating, την jackknife εκδοχή του ελέγχου του Layard, την εκδοχή του Layard για τον χ^2-έλεγχο του Scheffe, τον έλεγχο του Levene και τον έλεγχο Levene βασισμένο στις αποκλίσεις από τη διάμεσο αντί για το μέσο (όπως προτάθηκε από τους Brown και Forsythe) ως προς την ισχύ και την ευρωστία σε μια ευρεία ποικιλία από μη-κανονικές κατανομές.

Το υπόλοιπο του κεφαλαίου είναι οργανωμένο ως εξής: Στην παράγραφο 3.2.3 παρέχεται μια περιγραφή όλων των ελέγχων, στην παράγραφο 3.2.4 η περιγραφή της μελέτης προσομοίωσης. Στην παράγραφο 3.2.5 καταγράφονται τα αποτελέσματα από ένα πείραμα προσομοίωσης για την απόδοση των ελέγχων σε μικρά έως μεγάλα μεγέθη δειγμάτων και διάφορες κατανομές και στην παράγραφο 3.2.6 παρουσιάζονται κάποια συμπεράσματα.

3.2.3 Περιγραφή των ελέγχων

Από το κεφάλαιο 2 θα χρησιμοποιήσουμε τους ελέγχους των: Bartlett (παράγραφος 2.2.1.2) που θα τον συμβολίζουμε με BART, Levene (παράγραφος 2.2.1.12) και εδώ τον συμβολίζουμε ως LEVENE, Layard (παράγραφος 2.2.2.3) και τον συμβολίζουμε ως LAY καθώς και μια εκδοχή του jackknife (παράγραφος 2.2.4.2) που τον συμβολίζουμε ως JACK, τον προτεινόμενο έλεγχο NEWJACK (παράγραφος 2.2.4.3), και τον τροποποιημένο χ^2 έλεγχο του Layard που συμβολίζεται ως LAYJACK.

3.2.3.1 Ακριβής (exact) έλεγχος του Bartlett

Οι Dyer και Keating (1980) εξέτασαν τον έλεγχο του λόγου πιθανοφάνειας με μια τροποποίηση του ελέγχου του Bartlett και έλαβαν τις ακριβείς κρίσιμες τιμές από τον στατιστικό έλεγχο που προέκυψε, τον οποίο ονομάζουμε BARTE, όπου

$$BARTE = \prod_{i=1}^{k} (S_i^2)^{a_i} \Big/ \sum_{i=1}^{k} a_i S_i^2$$

3.2.4 Περιγραφή της προσομοίωσης

Για να εξετάσουμε την ευρωστία και την ισχύ των ελέγχων που αναφέραμε θα χρησιμοποιήσουμε μερικές κατανομές. Η κανονική κατανομή θεωρείται ως σημείο αναφοράς των υπολογισμών. Μαζί με την κανονική, θα μελετήσουμε ακόμη έξι μη-κανονικές κατανομές. Αυτές είναι: (i) μια t-κατανομή με 5 βαθμούς ελευθερίας, $t_{(5)}$, (ii) μια χ^2-κατανομή με 4 βαθμούς ελευθερίας, $\chi^2_{(4)}$, (iii) μια διπλά εκθετική κατανομή, που συμβολίζεται ως $DEXP$, (iv) μια τυπική εκθετική κατανομή, που συμβολίζεται ως EXP, (v) μια βήτα κατανομή με παραμέτρους 3 και 2, $B(3,2)$, και (vi) μια ομοιόμορφη κατανομή με εύρος 0 έως 1, $U(0,1)$.

Σημειώνουμε ότι οι $t_{(5)}$, $\chi^2_{(4)}$, $DEXP$ και EXP είναι κατανομές με παχιές ουρές (heavy-tailed) και οι $B(3,2)$ και $U(0,1)$ είναι κατανομές με λεπτές ουρές (light-tailed). Επιπλέον, η $B(3,2)$ είναι λοξή (skewed) προς τα αρνητικά, οι $t_{(5)}$, $DEXP$ και $U(0,1)$ είναι συμμετρικές, και οι $\chi^2_{(4)}$ και EXP είναι λοξές (skewed) προς τα δεξιά.

Εξετάζουμε μεγέθη δειγμάτων $n = 20$, 40, και 80 για καθεμία από τις κατανομές. Αυτά χωρίζονται σε τέσσερις ομάδες ίσων μεγεθών, (π.χ., $n_1 = n_2 = n_3 = n_4$). Έτσι έχουμε, $n_i = 5$, 10, και 20 για τα $n = 20$, 40, και 80 αντίστοιχα. Για την μηδενική υπόθεση, θέτουμε όλες τις διασπορές των πληθυσμών ίσες με 1 (μονάδα), π.χ., $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$. Θεωρούμε τέσσερις εναλλακτικές υποθέσεις $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (1,1,2,2)$, $(1,2,3,4)$, $(1,1,4,4)$, και $(1,4,9,6)$, που συμβολίζονται αντίστοιχα με H_1, H_2, H_3, και H_4.

Για κάθε στατιστική συνάρτηση η μέθοδος Monte Carlo επαναλαμβάνεται 10.000 φορές και καταγράφεται το ποσοστό των περιπτώσεων όπου ο στατιστικός έλεγχος ανήκει στο πεδίο απόρριψης στο 5% ονομαστικό (nominal) επίπεδο σημαντικότητας. Για τη μηδενική υπόθεση, αυτό το ποσοστό είναι το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας ή το πραγματικό επίπεδο του ελέγχου, και για την εναλλακτική υπόθεση, αυτό το ποσοστό είναι η εκτιμώμενη ισχύς του ελέγχου. Για όλους τους ελέγχους χρησιμοποιείται η δεξιά μονόπλευρη εναλλακτική (one-sided right alternative). Σημειώνουμε ότι για 10.000 προσομοιώσεις, το τυπικό σφάλμα του εκτιμώμενου επιπέδου στατιστικής σημαντικότητας είναι μόλις 0.2%. Γι' αυτό το λόγο, σε μια αυστηρή λογική, ένας έλεγχος με εκτιμώμενο επίπεδο στατιστικής σημαντικότητας μεγαλύτερο από 5.6% δεν είναι ικανοποιητικός. Άλλοι ερευνητές έχουν χρησιμοποιήσει διαφορετικά κριτήρια για τον ορισμό της ευρωστίας. Για παράδειγμα, οι Conover, Johnson and Johnson (1981) θεώρησαν ότι ένας έλεγχος είναι εύρωστος εάν το πραγματικό σφάλμα τύπου I είναι μικρότερο ή ίσο από το 10%, ενώ οι Sharma και Giacotto (1991) θεώρησαν ότι ένας έλεγχος είναι εύρωστος αν το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας είναι μικρότερο ή ίσο από το 7%. Σ' αυτή τη μελέτη, θεωρούμε ότι ένας έλεγχος είναι εύρωστος αν το πραγματικό επίπεδο στατιστικής σημαντικότητας είναι ίσο ή μικρότερο από το 7%.

3.2.5 Αποτελέσματα της προσομοίωσης

Το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας και ισχύος των ελέγχων (σε ποσοστά) για $n = 20, 40$, και 80 αναφέρονται στον πίνακα 3.5. Οι στατιστικοί έλεγχοι που θα αναφερθούμε στη συνέχεια γράφονται για συντομία ως εξής: NEWJACK είναι ο έλεγχος που προτείναμε, JACK είναι ο jackknife έλεγχος του Layard, BARTE είναι ο έλεγχος του Bartlett όταν χρησιμοποιούνται οι κρίσιμες τιμές που δίνονται από τους Dyer και Keating, BART είναι ο έλεγχος του Bartlett όταν οι κρίσιμες τιμές λαμβάνονται από την χ^2 κατανομή, LAY είναι ο χ^2 έλεγχος του Layard, LAYJACK είναι ο χ^2 έλεγχος του Layard όταν χρησιμοποιείται η jackknife εκτίμηση της κύρτωσης, LEVENE και LEVMED είναι ο έλεγχος του Levene όταν οι απόλυτες αποκλίσεις προέρχονται από τον μέσο και τη διάμεσο αντίστοιχα. Για δεδομένο μέγεθος δείγματος, εξετάζουμε όλους τους ελέγχους όσο αφορά την ευρωστία και την ισχύ για κανονικές όπως επίσης και για διαφορετικές μη-κανονικές κατανομές.

Όταν το μέγεθος του δείγματος είναι μικρό, π.χ., $n = 20$, κάτω από ιδανικές συνθήκες δειγματοληψίας από μια κανονική κατανομή, οι έλεγχοι NEWJACK, JACK, BARTE, και BART είναι εύρωστοι, με ακριβή επίπεδα στατιστικής σημαντικότητας 3.60, 3.83, 5.04 και 4.95% αντίστοιχα. Όπως αναμενόταν, τα πραγματικά επίπεδα των ελέγχων BARTE και BART είναι σχεδόν 5%. Οι έλεγχοι LAY, LAYJACK και LEVENE δεν είναι εύρωστοι, και ο LEVMED είναι αρκετά συντηρητικός. Μεταξύ των εύρωστων ελέγχων, ο NEWJACK έχει την μεγαλύτερη ισχύ. Σημειώνουμε ότι η ισχύς του ελέγχου NEWJACK είναι ακόμη μεγαλύτερη από αυτές των ελέγχων BARTE και BART.

Όταν οι παρατηρήσεις προέρχονται από κατανομές με παχιές ουρές (heavy-tailed), όπως οι $t_{(5)}$, $\chi^2_{(4)}$, $DEXP$, και EXP, ο NEWJACK είναι ο μοναδικός έλεγχος που παραμένει εύρωστος για όλες τις κατανομές. Ωστόσο, ο έλεγχος NEWJACK απορρίπτει την μηδενική υπόθεση λιγότερο συχνά από ότι θα έπρεπε, με τα πραγματικά επίπεδα στατιστικής σημαντικότητας να ανήκουν στο εύρος μεταξύ 2.25% και 2.60%. Εκτός από τον NEWJACK, ο έλεγχος JACK είναι επίσης εύρωστος για τρεις από τις τέσσερις κατανομές με παχιές ουρές. Κανένας άλλος από τους ελέγχους δεν είναι εύρωστος. Όσο αφορά την ισχύ, ο έλεγχος NEWJACK έχει μεγαλύτερη ισχύ από τον JACK για όλες τις εναλλακτικές υποθέσεις. Σε αρκετές περιπτώσεις η ισχύς του NEWJACK είναι ακόμη και διπλάσια από αυτή του JACK. Όταν οι παρατηρήσεις προέρχονται από κατανομές με λεπτές ουρές (light-tailed), όπως οι $B(3,2), U(0,1)$, οι έλεγχοι NEWJACK, JACK, BARTE, και BART είναι εύρωστοι. Ωστόσο, μεταξύ των τεσσάρων εύρωστων ελέγχων ο NEWJACK έχει πολύ μεγαλύτερη ισχύ από τους υπόλοιπους. Σημειώνουμε ότι για την υπόθεση H_2, η ισχύς του NEWJACK είναι περίπου 50% με 60%, συγκρινόμενη με το 8% του ελέγχου JACK και το 7% με 10% των BARTE και BART.

Γι' αυτό το λόγο όταν το μέγεθος του δείγματος είναι μικρό, π.χ., $n = 20$, και οι παρατηρήσεις είναι από την κανονική κατανομή, ο έλεγχος NEWJACK είναι καλύτερος από τους BARTE και BART. Όμως, όταν οι παρατηρήσεις προέρχονται από μια μη-κανονική κατανομή (με παχιές ή λεπτές ουρές) ο NEWJACK είναι η μοναδική επιλογή.

Όταν το μέγεθος του δείγματος είναι ικανοποιητικό, π.χ., $n = 40$, και οι παρατηρήσεις είναι από την κανονική κατανομή, ουσιαστικά όλοι έλεγχοι είναι εύρωστοι εκτός από τον LAY. Τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας των ελέγχων JACK, BARTE, και BART είναι περίπου 5% αλλά για τον NEWJACK είναι περίπου 4%. Γενικά, ο NEWJACK έχει πολύ μεγαλύτερη ισχύ από όλους τους εύρωστους ελέγχους. Μόνο για την υπόθεση H_3, οι έλεγχοι BARTE και BART έχουν μεγαλύτερα ποσοστά ισχύος από τον NEWJACK.

Για ικανοποιητικά μεγέθη δειγμάτων, όταν οι παρατηρήσεις προέρχονται από κατανομές με παχιές ουρές (heavy-tailed), οι NEWJACK και LEVMED είναι οι μόνοι εύρωστοι έλεγχοι. Και οι δυο είναι αρκετά συντηρητικοί με τα πραγματικά επίπεδα στατιστικής σημαντικότητας να ανήκουν στο εύρος μεταξύ 2.3% και 3% για τον NEWJACK, και 3.25% και 4.6% για τον LEVMED. Ωστόσο, η ισχύς του ελέγχου NEWJACK είναι πολύ μεγαλύτερη από αυτή του LEVMED. Σημειώνουμε ότι για την υπόθεση H_2, η ισχύς του NEWJACK είναι τριπλάσια από αυτή του LEVMED, και για την υπόθεση H_4 η ισχύς του NEWJACK είναι διπλάσια από αυτή του LEVMED. Από την άλλη μεριά για κατανομές με λεπτές ουρές (light-tailed), όλοι οι έλεγχοι είναι εύρωστοι. Τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας του ελέγχου NEWJACK είναι περίπου 5% αλλά ο LEVMED απορρίπτει την μηδενική υπόθεση λιγότερο συχνά και δεν είναι πάντα στατιστικά σημαντικός. Όσο αφορά την ισχύ, ο έλεγχος NEWJACK είναι καλύτερος από τον LEVMED και για τις τέσσερις εναλλακτικές υποθέσεις. Ωστόσο, σε σύγκριση με άλλους εύρωστους ελέγχους, ο NEWJACK έχει μεγαλύτερη ισχύ για τρεις (H_1, H_2 και H_4) από τις τέσσερις εναλλακτικές υποθέσεις. Για την υπόθεση H_3, η ισχύς του NEWJACK είναι ελαφρώς μικρότερη από τους άλλους ελέγχους, αλλά η διαφορά είναι αμελητέα για πρακτικούς λόγους.

Γι’ αυτό το λόγο, παρατηρούμε ότι γενικά για ικανοποιητικά μεγέθη δειγμάτων, με $n = 40$, ο έλεγχος NEWJACK είναι και πάλι ο καλύτερος μεταξύ όλων των ελέγχων που εξετάστηκαν. Για παρατηρήσεις που ακολουθούν κανονική κατανομή ο NEWJACK είναι τόσο καλός, ίσως και καλύτερος από τους BARTE και BART. Για μη-κανονικές παρατηρήσεις ο NEWJACK είναι σαφώς καλύτερος από τον κοντινότερο ανταγωνιστή του, τον έλεγχο LEVMED.

Όταν το μέγεθος του δείγματος είναι μεγάλο, π.χ., $n = 80$, όλοι οι έλεγχοι είναι εύρωστοι για κανονικές παρατηρήσεις. Τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας των ελέγχων JACK, BARTE, και BART είναι περίπου 5% και για τον NEWJACK είναι περίπου 4.4%. Η ισχύς του NEWJACK είναι τόσο μεγάλη, ίσως και μεγαλύτερη από τους άλλους ελέγχους για τις υποθέσεις H_2 και H_4, αλλά μικρότερη για τις υποθέσεις H_1 και H_3.

Για τις τέσσερις κατανομές με παχιές ουρές (heavy -tailed), οι NEWJACK και LEVMED είναι εύρωστοι έλεγχοι, με τα πραγματικά επίπεδα στατιστικής σημαντικότητας να ανήκουν στο εύρος μεταξύ 2.4% και 3.2% για τον NEWJACK, και 3.65% και 4.4% για τον LEVMED. Γι’ αυτή την περίπτωση, ο έλεγχος LAYJACK (που παρουσιάζεται στο κεφ.2) είναι επίσης εύρωστος για τις κατανομές $t_{(5)}$, $\chi^2_{(4)}$ και $DEXP$, αλλά όχι και για την EXP. Σημειώνουμε πως όταν το μέγεθος του δείγματος είναι μεγάλο, για το 5% ονομαστικό επίπεδο, τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας των ελέγχων BARTE και BART είναι περίπου 30% για τις κατανομές $t_{(5)}$, $\chi^2_{(4)}$ και $DEXP$, και περίπου 50% για την EXP. Όσο αφορά την ισχύ, ο έλεγχος NEWJACK έχει τόσο μεγάλη, ίσως και μεγαλύτερη ισχύ από τους LAYJACK και LEVMED για τις υποθέσεις H_1, H_2, και H_4 αλλά μικρότερη για την H_3.

Για κατανομές με λεπτές ουρές, και πάλι, όλοι οι έλεγχοι είναι εύρωστοι. Τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας του NEWJACK είναι περίπου 5.5%, αυτών του LEVMED ανήκουν στο εύρος μεταξύ 2.64% και 3.15% και αυτών του LAYJACK μεταξύ 5% και 6%. Σημειώνουμε ότι οι έλεγχοι BARTE και BART είναι πολύ συντηρητικοί με πραγματικά επίπεδα σημαντικότητας κοντά στο 1%. Η ισχύς του NEWJACK είναι τόσο μεγάλη, ίσως και μεγαλύτερη από τους άλλους εύρωστους ελέγχους για τις υποθέσεις H_2 και H_4, αλλά μικρότερη για τις υποθέσεις H_1 και H_3.

Συμπερασματικά λοιπόν, όταν το μέγεθος του δείγματος είναι μεγάλο και οι παρατηρήσεις προέρχονται από (μια) κανονική κατανομή όλοι οι έλεγχοι που εξετάστηκαν είναι σχεδόν ισοδύναμοι. Ωστόσο, όταν οι παρατηρήσεις είναι μη-κανονικές, οι έλεγχοι NEWJACK, LAYJACK και LEVMED έχουν επίσης πολλές ομοιότητες και δεν χρειάζεται πολύ επεξεργασία για να επιλέξουμε μεταξύ αυτών. Όμως, ο NEWJACK είναι ευκολότερος όσο αφορά τους υπολογισμούς, σε σχέση με τους άλλους δυο, και γι' αυτό θα πρέπει να προτιμάται.

Πίνακας 3.5.

Εκτιμώμενο επίπεδο στατιστικής σημαντικότητας και ισχύς των ελέγχων (σε ποσοστά)

Dist.	Hyp.	NEWJACK	JACK	BARTE	BART	LAY	LAYJACK	LEVENE	LEVMED
		$n = 20$							
Normal	H_0	3.60	3.83	5.04	4.95	11.95	10.80	9.32	0.38
	H_1	14.66	5.63	9.40	9.24	16.52	14.28	14.28	0.87
	H_2	45.35	8.26	15.49	15.20	23.00	19.62	19.31	1.55
	H_3	32.06	13.22	26.75	26.48	31.31	25.30	28.88	3.13
	H_4	87.54	28.94	53.29	52.72	59.00	47.55	42.57	6.91
$t_{(5)}$	H_0	2.26	5.12	13.06	12.85	13.35	11.57	10.24	0.57
	H_1	8.42	7.26	18.76	18.52	18.12	14.75	14.19	1.22
	H_2	31.12	10.15	25.64	25.32	24.86	19.66	19.21	1.67
	H_3	22.64	14.35	36.01	35.56	31.11	23.82	26.19	3.33
	H_4	77.19	28.56	57.60	57.18	56.09	43.08	38.66	6.25
$\chi^2_{(4)}$	H_0	2.34	5.78	16.40	16.17	17.39	14.62	17.35	0.83
	H_1	9.23	7.95	21.30	21.11	21.89	17.44	21.67	1.44
	H_2	30.41	11.12	29.73	29.37	28.53	22.58	26.83	2.32
	H_3	22.71	15.35	38.66	38.39	34.90	27.06	33.96	3.71
	H_4	75.36	27.86	59.37	59.10	57.26	44.02	46.06	6.21
DEXP	H_0	2.57	6.12	18.72	18.55	17.05	14.30	11.98	0.89
	H_1	8.79	8.25	24.66	24.34	21.91	17.96	15.23	1.57
	H_2	30.66	10.83	31.00	30.70	27.49	21.98	19.37	2.07
	H_3	22.68	15.22	41.78	41.46	35.01	26.76	26.17	4.29
	H_4	73.61	26.83	62.57	62.16	55.97	42.79	37.06	6.71
EXP	H_0	2.20	8.08	29.77	29.55	25.16	20.05	24.16	1.17
	H_1	6.92	10.76	36.05	35.75	29.58	23.38	28.39	1.82
	H_2	22.46	13.41	41.52	41.19	34.80	26.87	33.40	2.57
	H_3	17.69	17.19	49.63	49.43	41.03	30.31	39.36	4.08
	H_4	64.19	25.87	65.93	65.67	57.99	44.04	50.82	5.84
B(3,2)	H_0	4.67	3.00	3.08	3.02	11.70	11.24	9.81	0.19
	H_1	17.45	5.37	6.74	6.55	15.36	13.68	15.44	0.78
	H_2	52.10	8.13	10.73	10.43	22.80	19.92	20.94	1.02
	H_3	37.42	13.03	22.23	21.95	31.69	27.11	32.04	2.77
	H_4	90.17	29.85	49.75	49.30	61.68	50.45	47.63	6.86
U(0,1)	H_0	6.29	2.88	1.65	1.57	10.52	10.41	9.23	0.09
	H_1	20.31	4.86	3.83	3.70	14.87	13.88	17.09	0.39
	H_2	60.32	7.85	7.37	7.24	22.64	20.26	24.26	0.85
	H_3	43.64	12.61	16.21	15.85	32.47	28.65	37.22	2.29
	H_4	93.66	30.69	46.34	45.90	65.41	54.86	52.89	6.05

(continued)

<u>Πίνακας 3.5.</u> (συνέχεια)

		$n=40$							
Dist.	Hyp.	NEWJACK	JACK	BARTE	BART	LAY	LAYJACK	LEVENE	LEVMED
Normal	H_0	3.90	4.94	4.85	4.80	7.70	6.59	6.77	3.30
	H_1	20.14	13.72	17.92	17.88	18.91	15.65	18.69	10.60
	H_2	67.92	28.63	35.87	35.80	37.18	31.33	31.53	18.31
	H_3	46.60	48.29	64.64	64.58	56.71	47.94	55.91	38.81
	H_4	97.65	84.73	94.25	94.21	90.73	80.21	82.34	66.37
$t_{(5)}$	H_0	2.32	7.98	20.36	20.33	8.03	6.33	7.61	3.26
	H_1	12.43	15.94	34.46	34.40	16.58	12.83	16.68	8.62
	H_2	48.17	25.01	49.10	49.05	28.45	22.71	26.07	14.71
	H_3	32.04	39.97	67.91	67.85	43.82	35.36	45.01	29.24
	H_4	93.24	69.41	91.87	91.84	77.66	64.01	70.63	52.11
$\chi^2_{(4)}$	H_0	2.74	9.64	23.31	23.30	11.61	9.61	15.65	4.13
	$H1$	13.49	17.03	37.03	36.96	20.40	16.56	27.22	9.64
	H_2	47.07	26.12	51.06	51.00	32.10	26.77	37.01	14.54
	H_3	33.13	39.23	68.99	68.95	45.75	37.45	54.88	27.73
	H_4	91.58	65.25	91.70	91.67	76.65	64.32	78.06	48.47
DEXP	H_0	2.98	8.12	26.18	26.12	10.40	8.70	8.00	3.25
	H_1	13.34	14.47	40.17	40.12	17.81	13.81	16.69	8.41
	H_2	46.39	23.72	53.89	53.82	29.08	23.59	25.19	13.56
	H_3	32.91	35.05	70.04	70.02	41.66	33.30	40.86	25.22
	H_4	91.26	61.06	91.30	91.29	73.17	60.49	63.43	44.53
EXP	H_0	2.51	12.45	40.67	40.63	15.39	12.84	22.99	4.60
	H_1	10.23	17.39	50.95	50.94	22.09	18.32	32.55	8.22
	H_2	37.02	24.31	60.93	60.90	31.01	25.48	41.43	12.28
	H_3	26.10	33.18	73.35	73.29	42.14	34.08	55.34	21.58
	H_4	84.52	52.96	90.14	90.13	67.23	54.57	74.57	35.88
B(3,2)	H_0	4.78	4.17	1.90	1.89	7.98	7.49	7.52	3.13
	H_1	23.42	14.75	12.09	12.04	22.12	19.03	22.86	11.62
	H_2	74.86	32.76	28.59	28.56	43.48	37.71	38.07	21.58
	H_3	52.90	57.93	62.94	62.86	65.59	57.75	63.65	44.51
	H_4	98.84	92.29	96.20	96.18	95.46	88.15	88.90	73.03
U(0,1)	H_0	5.34	2.78	0.48	0.48	6.54	6.65	6.91	2.57
	H_1	25.69	16.08	5.50	5.49	24.80	22.40	26.02	12.22
	H_2	83.67	42.08	20.16	20.09	53.97	48.72	47.46	25.18
	H_3	61.36	73.48	60.70	60.64	77.69	72.34	73.50	50.64
	H_4	99.54	97.88	97.84	97.84	98.50	95.40	94.46	80.23

Πίνακας 3.5. (συνέχεια)

Dist.	Hyp.	NEWJACK	JACK	BARTE	BART	LAY	LAYJACK	LEVENE	LEVMED
		$n = 80$							
Normal	H_0	4.38	4.96	4.97	4.96	6.14	5.36	5.38	3.38
	H_1	24.26	33.65	39.14	39.06	35.84	30.78	34.92	27.38
	H_2	85.49	67.16	73.74	73.74	70.94	64.43	62.92	53.45
	H_3	61.26	90.87	95.29	95.29	91.53	86.38	90.28	85.36
	H_4	99.91	99.94	99.99	99.99	99.88	98.89	99.78	99.47
$t_{(5)}$	H_0	2.36	7.60	26.41	26.41	5.27	3.81	6.26	3.65
	H_1	16.21	25.60	53.53	53.50	21.93	17.09	26.45	19.92
	H_2	66.05	48.23	75.96	75.95	46.02	38.29	48.33	38.79
	H_3	42.68	69.33	92.21	92.20	68.91	59.13	77.96	70.26
	H_4	98.50	92.51	99.70	99.70	94.78	86.26	97.52	95.63
$\chi^2_{(4)}$	H_0	3.16	9.79	28.87	28.85	7.94	6.20	14.61	4.12
	H_1	17.89	25.65	55.98	55.98	24.53	20.32	38.82	19.16
	H_2	67.06	44.80	76.35	76.32	46.05	38.96	60.54	36.28
	H_3	44.64	64.44	91.28	91.28	67.25	57.67	83.82	66.22
	H_4	98.32	90.55	99.61	99.61	94.04	85.45	98.29	93.02
DEXP	H_0	3.28	7.82	30.70	30.68	6.77	5.51	6.77	4.13
	H_1	18.74	22.01	56.04	56.03	20.88	16.44	22.35	16.12
	H_2	65.56	41.62	77.58	77.57	43.07	35.59	42.73	33.89
	H_3	44.69	61.86	91.78	91.77	63.31	53.69	70.47	61.87
	H_4	98.03	90.48	99.58	99.58	93.50	84.86	94.43	90.69
EXP	H_0	2.85	11.27	47.61	47.58	9.67	7.92	21.68	4.33
	H_1	14.98	22.15	65.72	65.70	21.52	17.53	41.31	14.98
	H_2	55.21	35.25	79.74	79.71	37.16	31.37	58.29	26.25
	H_3	36.48	50.93	89.97	89.97	53.51	44.93	78.10	48.29
	H_4	95.32	78.71	99.08	99.08	84.34	73.06	95.87	79.15
B(3,2)	H_0	5.48	3.86	1.28	1.28	6.29	5.90	6.42	3.15
	H_1	27.05	42.20	31.10	31.07	46.29	42.09	41.71	30.27
	H_2	90.66	82.27	73.37	73.33	84.03	79.61	73.91	62.84
	H_3	69.09	97.53	97.05	97.03	97.08	95.27	94.96	91.09
	H_4	99.94	99.99	100.00	100.00	99.98	99.89	99.92	99.78
U(0,1)	H_0	5.49	2.53	0.11	0.11	4.93	5.01	5.93	2.64
	H_1	28.47	63.62	20.74	20.70	63.38	60.05	51.49	36.96
	H_2	96.01	96.41	73.25	73.18	95.32	94.18	84.74	73.87
	H_3	79.46	99.90	98.74	98.74	99.71	99.43	98.38	95.99
	H_4	99.99	100.00	100.00	100.00	100.00	100.00	100.00	99.99

3.2.6 Συμπεράσματα

Ένας jackknife έλεγχος, NEWJACK, προτείνεται για τον έλεγχο της ομοιογένειας των διασπορών σε διάφορους κανονικούς ή μη-κανονικούς πληθυσμούς. Για να έχει ο προτεινόμενος έλεγχος μια έγκυρη (valid) κατανομή με λογικούς βαθμούς ελευθερίας, υποθέτουμε ότι ο αριθμός των πληθυσμών είναι μεγαλύτερος από δυο. Ο έλεγχος αυτός είναι βασισμένος στη χρήση της μεθόδου jackknife σε μια ομάδα παρατηρήσεων την κάθε φορά αντί για μια παρατήρηση σε κάθε ομάδα όπως έχει προταθεί από τον Miller και αργότερα από τον Layard. Ο προτεινόμενος έλεγχος εξετάστηκε και συγκρίθηκε με άλλους επτά ελέγχους όσο αφορά το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας και την ισχύ για τέσσερις κανονικές υποθέσεις, για την κανονική κατανομή καθώς επίσης και για μια μεγάλη ποικιλία μη-κανονικών κατανομών.

Για μια γρήγορη αναφορά η ευρωστία των ελέγχων σε ποιοτικά επίπεδα και τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας βρίσκονται μαζί στον Πίνακα 3.6. Όταν οι παρατηρήσεις είναι κανονικές οι έλεγχοι, NEWJACK, JACK, BARTE, και BART είναι εύρωστοι ανεξάρτητα από το μέγεθος του δείγματος. Οι έλεγχοι LAYJACK, LEVENE και LEVMED είναι εύρωστοι για μεσαία ($n = 40$) και μεγάλα δείγματα ($n = 80$). Ωστόσο όταν οι παρατηρήσεις είναι μη-κανονικές, ο NEWJACK είναι ο μόνος έλεγχος ο οποίος είναι εύρωστος ανεξάρτητα από το μέγεθος του δείγματος. Ο έλεγχος LEVMED είναι εύρωστος για μεσαία και μεγάλα δείγματα και ο LAYJACK μόνο για μεγάλα δείγματα.

Σε ότι αφορά την ισχύ, για κανονικές παρατηρήσεις ο έλεγχος NEWJACK είναι πολύ καλύτερος από τους συγκρινόμενους ελέγχους για μικρά ($n = 20$) και μεσαία ($n = 40$) μεγέθη δειγμάτων. Σημειώνουμε ότι σ' αυτές τις περιπτώσεις η ισχύς του NEWJACK είναι ακόμη μεγαλύτερη από την ισχύ των ελέγχων BARTE και BART. Σε μεγαλύτερα δείγματα, $n = 80$, όλοι οι έλεγχοι έχουν σχεδόν ισοδύναμη ισχύ. Όταν οι παρατηρήσεις είναι μη-κανονικές, ο NEWJACK, έχει ασφαλώς μεγαλύτερη ισχύ από τους άλλους εύρωστους ελέγχους, JACK και LEVMED, σε μικρά και μεσαία μεγέθη δειγμάτων. Ωστόσο, όταν το μέγεθος του δείγματος είναι μεγάλο οι έλεγχοι NEWJACK, LAYJACK και LEVMED έχουν σχεδόν ισοδύναμη ισχύ.

Γι' αυτό το λόγο, αυτά τα αποτελέσματα υποδεικνύουν ότι ο NEWJACK είναι σαφώς καλύτερος από τους άλλους ελέγχους όταν τα μεγέθη των δειγμάτων είναι μικρά ($n = 20$) και μεσαία ($n = 40$) και οι παρατηρήσεις ακολουθούν κανονική ή μη-κανονική κατανομή (με παχιές ή μακριές ουρές). Ωστόσο, σε μεγάλα δείγματα, για κανονικές κατανομές δεν υπάρχει μεγάλη ποικιλία επιλογής μεταξύ των ελέγχων που εξετάζονται εδώ. Για μη-κανονικές παρατηρήσεις, οι NEWJACK, LAYJACK και LEVMED είναι ισοδύναμοι έλεγχοι. Όμως, ο NEWJACK είναι λιγότερο πολύπλοκος όσο αφορά τους υπολογισμούς, σε σχέση με τους άλλους δυο, και γι' αυτό θα πρέπει να είναι η επιλογή μας.

Συνεπώς, για τον έλεγχο της ισότητας των διασπορών σε διάφορους πληθυσμούς, κανονικούς ή μη, ο NEWJACK είναι ο έλεγχος που θα χρησιμοποιούμε ανεξάρτητα από το μέγεθος του δείγματος και την εναλλακτική υπόθεση. Σημειώνουμε ότι τα παραπάνω συμπεράσματα είναι βάσιμα όταν τα μεγέθη των δειγμάτων είναι ίσα σε κάθε ομάδα, και ίσως δεν ισχύουν για άνισα μεγέθη δειγμάτων.

Ο Layard (1973, p. 190) τόνισε ότι οι έλεγχοι jackknife και χ^2 (που συμβολίζονται εδώ ως JACK και LAY) είναι έλεγχοι για μεγάλα δείγματα. Όμως, βασισμένοι στις προσομοιώσεις αυτής μελέτης, οι έλεγχοι JACK και LAY δεν συνιστώνται για μεγέθη δειγμάτων 20, 40, ακόμη και 80 παρατηρήσεων.

Πίνακας 3.6.

Ευρωστία των ελέγχων σε ποιοτικά επίπεδα και τα εκτιμώμενα επίπεδα στατιστικής σημαντικότητας.

	NEWJACK			JACK			BARTE			BART		
	n			n			n			n		
Distributions	20	40	80	20	40	80	20	40	80	20	40	80
Normal	yes 3.60	yes 3.90	yes 4.38	yes 3.83	yes 4.94	yes 4.96	yes 5.04	yes 4.85	yes 4.97	yes 4.95	yes 4.80	yes 4.96
$t_{(5)}$	yes 2.26	yes 2.32	yes 2.36	yes 5.12	no 7.98	no 7.60	no 13.06	no 20.36	no 26.41	no 12.85	no 20.33	no 26.41
$\chi^2_{(4)}$	yes 2.34	yes 2.74	yes 3.16	yes 5.78	no 9.64	no 9.79	no 16.40	no 23.31	no 28.87	no 16.17	no 23.30	no 28.85
DEXP	yes 2.57	yes 2.98	yes 3.28	yes 6.12	no 8.12	no 7.82	no 18.72	no 26.18	no 30.70	no 18.55	no 26.12	no 30.68
EXP	yes 2.20	yes 2.51	yes 2.85	no 8.08	no 12.45	no 11.27	no 29.77	no 40.67	no 47.61	no 29.55	no 40.63	no 47.58
B(3,2)	yes 4.67	yes 4.78	yes 5.48	yes 3.0	yes 4.17	yes 3.86	yes 3.08	yes 1.90	yes 1.28	yes 3.02	yes 1.89	yes 1.28
U(0,1)	yes 6.29	yes 5.34	yes 5.49	yes 2.88	yes 2.78	yes 2.53	yes 1.65	* 0.48	* 0.11	yes 1.57	* 0.48	* 0.11

(continued)

__Πίνακας 3.6.__ (συνέχεια)

Distributions	LAY n 20	40	80	LAYJACK n 20	40	80	LEVENE n 20	40	80	LEVMED n 20	40	80
Normal	no 11.95	no 7.70	yes 6.14	no 10.80	yes 6.59	yes 5.36	no 9.32	yes 6.77	yes 5.38	* 0.38	yes 3.30	yes 3.38
$t_{(5)}$	no 13.35	no 8.03	yes 5.27	no 11.57	yes 6.33	yes 3.81	no 10.24	no 7.61	yes 6.26	* 0.57	yes 3.26	yes 3.65
$\chi^2_{(4)}$	no 17.39	no 11.61	no 7.94	no 14.62	no 9.61	yes 6.20	no 17.35	no 15.65	no 14.61	* 0.83	yes 4.13	yes 4.12
DEXP	no 17.05	no 10.40	yes 6.77	no 14.30	no 8.70	yes 5.51	no 11.98	no 8.00	yes 6.77	* 0.89	yes 3.25	yes 4.13
EXP	no 25.16	no 15.39	no 9.67	no 20.05	no 12.84	no 7.92	no 24.16	no 22.99	no 21.68	yes 1.17	yes 4.60	yes 4.33
B(3,2)	no 11.70	no 7.98	yes 6.29	no 11.24	no 7.49	yes 5.90	no 9.81	no 7.52	yes 6.42	* 0.19	yes 3.13	yes 3.15
U(0,1)	no 10.52	yes 6.54	yes 4.93	no 10.41	yes 6.65	yes 5.01	no 9.23	yes 6.91	yes 5.93	* 0.09	yes 2.57	yes 2.64

yes: σημαίνει ότι ο έλεγχος είναι εύρωστος, π.χ., για 5% ονομαστικό επίπεδο (nominal level) το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας είναι μικρότερο από 7%.

no: σημαίνει ότι ο έλεγχος δεν είναι εύρωστος, π.χ., για 5% ονομαστικό επίπεδο το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας είναι μεγαλύτερο από 7%.

*: σημαίνει ότι για 5% ονομαστικό επίπεδο το εκτιμώμενο επίπεδο στατιστικής σημαντικότητας είναι μικρότερο από 1%.

Κεφάλαιο 4

Νέοι έλεγχοι

4.1 Ένας εναλλακτικός έλεγχος για την ισότητα των διασπορών για αρκετούς πληθυσμούς όταν οι κατανομές είναι κανονικές

4.1.1 Εισαγωγή

Ο έλεγχος της ομοιογένειας των διασπορών έχει μελετηθεί από τους Bartlett (1937), Hartley (1950) και Levene (1960) μεταξύ άλλων. Οι έλεγχοι που αναπτύχτηκαν από τους παραπάνω στατιστικούς, είναι είτε προσεγγιστικοί ή έλεγχοι που χρησιμοποιούν αριθμητική προσομοίωση των κρίσιμων σημείων, έτσι ώστε η ισχύς των ελέγχων να βασίζεται στο μέγεθος των δειγμάτων. Έχουμε αναπτύξει έναν έλεγχο, τον αποκαλούμενο νέο έλεγχο (New-test), για την ισότητα των διασπορών, του οποίου το σφάλμα Τύπου Ι (Type I) είναι καλά ελεγχόμενο και του οποίου η ισχύς είναι ανταγωνίσιμη του βέλτιστου εναλλακτικού ελέγχου. Η ισχύς του δηλαδή υπερνικά τους άλλους ελέγχους όπως επίσης υπερνικά την ισχύ του ελέγχου του Bartlett τις περισσότερες φορές για κανονικά δείγματα. Εκτεταμένα εμπειρικά πειράματα έχουν διεξαχθεί για να συγκριθεί η απόδοση του Νέου ελέγχου (New-test) με τις τρεις κλασσικές μεθόδους.

Ένα πείραμα με εκθετικά δεδομένα επίσης έγινε μέσω προσομοίωσης. Φαίνεται ότι, υπό εκθετική κατανομή, το σφάλμα Τύπου Ι δεν είναι ελεγχόμενο όπως στην περίπτωση της κανονικής κατανομής.

Με σχετικά μεγαλύτερο και ακριβή έλεγχο του σφάλματος Τύπου Ι, ο Νέος έλεγχος (New-test) μπορεί να προταθεί για μελλοντική χρήση από τους μελετητές όταν τα βασικά δεδομένα ακολουθούν κανονική κατανομή.

Ο Νέος-έλεγχος (New-test) είναι στατιστικά επωφελής για τους παρακάτω λόγους **:**

- Η μέθοδος είναι εύκολο να εκτελεστεί. Είναι κατανοητή από τους μελετητές με ελάχιστη στατιστική προετοιμασία και μπορεί να εκτελεστεί με το χέρι.

- Δεν υπάρχει περιοριστική απαίτηση για το μέγεθος του δείγματος για κάθε ομάδα σε ένα σύνολο δεδομένων. Συνεπώς, αυτός ο έλεγχος μπορεί να πραγματοποιηθεί για τα μη ισορροπημένα δεδομένα στα οποία μια ομάδα έχει λιγότερες παρατηρήσεις σε σχέση με τις άλλες ομάδες.

- Αυτός ο έλεγχος μπορεί να προσφέρει ως ένας έλεγχος γενικής χρήσης ώστε να αναγνωρίσει κατά πόσον οι διασπορές είναι ομοιογενείς ή όχι. Εάν είναι, μπορούμε διαδοχικά να κάνουμε σύγκριση κατά ζεύγη με την κλασσική προσαρμογή του Bonferroni για περισσότερη έρευνα. Ο έλεγχος γενικής χρήσης και η επιπρόσθετη σύγκριση κατά ζεύγη βασίζονται στο ίδιο παράδειγμα F – ελέγχου (F -test).

- Συγχωνεύοντας τα υπόλοιπα δεδομένα, ο Νέος-έλεγχος (New-test) έχει την δυνατότητα να εντοπίζει την ετερογένεια και να περιορίζει το μέγεθος των συγκρίσεων σε μια λογική κλίμακα. Η ισχύς του Νέου-έλεγχου είναι μεγαλύτερη από τον μέγιστο F – έλεγχο (F_{max} - test) και τον έλεγχο του Levene και είναι ανταγωνίσιμος με αυτόν του Barlett.

Στην παράγραφο 4.1.2 περιγράφουμε τον έλεγχο μας. Στην παράγραφο 4.1.3 συγκρίνουμε το σφάλμα Τύπου Ι και την ισχύ του ελέγχου μας με άλλους ελέγχους μέσω προσομοίωσης για την περίπτωση της κανονικής κατανομής και επίσης μια σύγκριση γίνεται για την περίπτωση εκθετικής κατανομής.

4.1.2 Περιγραφή των Ελέγχων

Σ' αυτή τη μελέτη θα χρησιμοποιήσουμε διαφορετικούς ελέγχους για την ισότητα των διασπορών για διαφόρους πληθυσμούς που αναπτύχτηκαν από διαφορετικούς στατιστικούς και προτείνουμε τον έλεγχο μας. Οι έλεγχοι αυτοί έχουν παρουσιαστεί στο κεφάλαιο 2 και είναι του Bartlett (παράγραφος 2.2.1.2), του Hartley (παράγραφος 2.2.1.5), του Levene (παράγραφος 2.2.1.12) και του Box (παράγραφος 2.2.1.11).

Έστω ότι έχουμε t αγωγές και η αγωγή i εφαρμόζεται σε n_i πειραματικές μονάδες, $i = 1, 2, ..., t$ και τα στοιχεία είναι όπως παρακάτω:

	Tr. 1	*Tr. 2*	...	*Tr. t*
	y_{11}	y_{21}	$\cdots$	y_{t1}
	y_{11}	y_{21}	$\cdots$	y_{t2}
Αποκρίσεις	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	y_{1n_1}	y_{2n_2}	$\cdots$	y_{tn_t}

Το μοντέλο που χρησιμοποιούμε σε αυτήν την περίπτωση για να αναλύσουμε τα στοιχεία είναι όπως παρακάτω:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{array}{l} i = 1, ..., t \\ j = 1, ..., n_i \end{array}$$

όπου $y_{ij} =$ η απόκριση που αντιστοιχεί στην i-οστή αγωγή και την j-οστή πειραματική μονάδα, $\mu =$ συνολικός αριθμητικός μέσος, $\tau_i =$ αποτέλεσμα που οφείλεται στην i-οστή αγωγή και $\varepsilon_{ij} =$ τυχαίο σφάλμα στην παρατήρηση y_{ij}.

Οι υποθέσεις στο παραπάνω μοντέλο μπορούν να γραφούν ως εξής **:**

(i) $\sum_{i=1}^{t} \tau_i = 0$

(ii) $\varepsilon_{ij} \sim N(0, \sigma_i^2) \qquad \begin{array}{l} i = 1, ..., t \\ j = 1, ..., n_i \end{array}$

Προτού αναλυθούν τα στοιχεία, πρέπει να ελεγχθεί η ομοιογένεια των διασπορών εξετάζοντας τις υποθέσεις:

$H_o : \sigma_1^2 = \sigma_2^2 = ... = \sigma_t^2$ ως προς H_1: όχι H_0.

4.1.2.1 Νέος Έλεγχος (New-test)

Σε αυτή την παράγραφο περιγράφουμε τον έλεγχο που δημιουργήθηκε από εμάς.

Βήμα 1. Παίρνουμε τα δεδομένα στον i-οστό πληθυσμό $(i=1,2,3,...,t)$ και τα συγχωνεύουμε στους υπόλοιπους πληθυσμούς. Έστω $S_i^2 = \dfrac{\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{ij})^2}{n_i-1}$ οι δειγματικές διασπορές στον i-οστό πληθυσμό και

$$S_{p,i}^2 = \frac{(n_1-1)S_1^2+...+(n_{i-1}-1)S_{i-1}^2+(n_{i+1}-1)S_{i+1}^2...+(n_t-1)S_t^2}{r_i}$$

είναι οι συγχωνευμένες (pooled) δειγματικές διασπορές στα συγχωνευμένα δεδομένα με $r_i=(n_1-1)+...+(n_{i-1}-1)+(n_{i+1}-1)+...+(n_t-1)$ βαθμούς ελευθερίας. Ορίζουμε τους F-ελέγχους να είναι $F_{i,calc.} = \dfrac{S_i^2}{S_{p,i}^2}$ και $\overline{F}_{i,calc.} = \dfrac{S_{p,i}^2}{S_i^2}$.

Βήμα 2. Διαμορφώνουμε τις P- τιμές να είναι η δεξιά ουρά πιθανότητας (right tail probability) των στατιστικών δεδομένων που υπολογίστηκαν στο Βήμα 1 έτσι ώστε $P_i = P(X > F_{i,calc.})$ όπου $X \sim F_{n_i-1,r_i}$ και $P_i = P(\overline{X} > \overline{F}_{i.calc.})$ όπου $\overline{X} \sim F_{r_i,n_i-1}$.

Βήμα 3. Τοποθετούμε τις P- τιμές, που λάβαμε από το Βήμα 2, σε ανιούσα τάξη και τις συμβολίζουμε ως $P_{(1)},P_{(2)},...,P_{(2t)}$. Απορρίπτουμε την υπόθεση H_0 αν $P_{(i)} < \dfrac{i}{2t}a$ για κάποια $i \in \{1,2,...,2t\}$.

Σημείωση : Για να είναι η υπόθεση H_0 στατιστικά μη σημαντική, χρειάζονται τα εξής:

$$P_{(1)} \geq \frac{1}{2t}a, P_{(2)} \geq \frac{2}{2t}a, ..., P_{(2t)} \geq a$$

Έτσι, αν $P_{(i)} < \dfrac{i}{2t}a$ έχουμε το δεδομένο ότι οι διασπορές που αντιστοιχούν στο $P_{(1)}$ είναι στατιστικά μη σημαντικές, αυτές που αντιστοιχούν στο $P_{(2)}$ είναι στατιστικά μη σημαντικές,…., αυτές που αντιστοιχούν στο $P_{(i-1)}$ είναι μη σημαντικές και τέλος αυτές που αντιστοιχούν στο $P_{(i)}$ είναι σημαντικές και ο συνολικός έλεγχος είναι στατιστικά σημαντικός.

Τώρα παρουσιάζουμε μια εξήγηση της διαδικασίας του Νέου ελέγχου (New-test). Από την υπόθεση $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_t^2 = \sigma^2$ είναι προφανές ότι $\frac{r_i S_{p,i}^2}{\sigma^2} \sim \chi_{r,i}^2$. Ο λόγος (ratio) μεταξύ της i-στης δειγματικής διασποράς και της συγχωνευμένης διασποράς ακολουθεί την F κατανομή υπό την υπόθεση H_0 , δηλ. $F_{i,calc.} = \frac{S_i^2}{S_{p,i}^2} \sim F_{n_i - 1, r_i}$ και $\overline{F}_{i,calc.} = \frac{S_{p,i}^2}{S_i^2} \sim F_{r_i, n_i - 1}$. Έτσι οι P τιμές των δεξιών μονοπλευρων F ελέγχων (right-sided F tests) μπορούν να υπολογιστούν.

Ο Simes (1986) πρότεινε την διαδικασία με την προσαρμογή τύπου Bonferroni στις διατεταγμένες P-τιμές ώστε να ελεγχθεί το σφάλμα Τύπου Ι. Αυτή η προσαρμογή είναι ειδικά σχεδιασμένη για πολλαπλούς ελέγχους στην ίδια μηδενική υπόθεση. Ο Simes απέδειξε ότι αν $P_1, ..., P_n$ είναι ανεξάρτητα ομοιόμορφοι (0, 1) υπό την υπόθεση H_0, τότε $P\left(\bigcup_{k=1}^{n} P_{(k)} < \frac{k}{n} a \mid H_0 \right) = a$. Από την στιγμή που οι έλεγχοι σε σχέση με την κοινή μηδενική υπόθεση συχνά συσχετίζονται, οι προσομοιώσεις διεξάγονται για να ερευνηθεί η απόδοση της διαδικασίας του Simes σε πολυ-κανονικές (multi-normal) και χ^2 κατανεμημένες μεταβλητές με διάφορα επίπεδα συσχέτισης. Έτσι προκύπτει ότι αυτές οι συσχετίσεις δεν διακινδυνεύουν την απόδοση των ελέγχων: το σφάλμα Τύπου Ι είναι μικρότερο από τον ονομαστικό ρυθμό (nominal rate) και η ισχύς αυτής της προσαρμοσμένης, στον τύπο Bonferroni, διαδικασίας βελτιώθηκε όταν οι έλεγχοι συσχετίστηκαν. Συγκεκριμένα, η ισχύς είναι δυνατή για συγκρίσεις μεγάλης ακριβείας οι οποίες έχουν διάφορες μικρές έως μέτριες ετερογένειες. Αυτό είναι επειδή οι παραπάνω δύο μονόπλευροι (one-sided) έλεγχοι είναι αρνητικά συσχετιζόμενοι, οπότε η διαδικασία του Simes είναι καλύτερη και μπορεί να τους ενώσει σε αμφίπλευρους ελέγχους (two-sided). Όταν δυο έλεγχοι είναι θετικά συσχετιζόμενοι, η ταύτιση μεταξύ των δύο *p*-τιμών μας βοηθά να εξάγουμε μια σωστά συμπεράσματα.

Όπως οι $F_{i,calc.}$ και $\overline{F}_{i,calc.}$ είναι σε αντίστροφη σχέση (reverse association), οι προσαρμογές του Simes μετασχηματίζουν έναν μονόπλευρο (one-sided) έλεγχο στην αναλογία μεταξύ των S_i^2 και $S_{p,i}^2$ σε αμφίπλευρο (two-sided). Εξαιτίας αυτών, θα απορρίψουμε την υπόθεση H_0 όταν ο λόγος $S_i^2 / S_{p,i}^2$ είναι πολύ μεγάλος ή πολύ μικρός. Η προσομοίωση που κάνουμε στους συσχετιζόμενους F-ελέγχους επιβεβαιώνουν τον ισχυρισμό του Simes και δείχνουν ότι μια μικρή παρέκκλιση (departure) από τις ανεξάρτητες υποθέσεις δεν επηρεάζει τον έλεγχο του σφάλματος Τύπου Ι από την προσαρμογή του Simes.

Συγκρίνουμε τις αποδόσεις όλων των ελέγχων που περιγράφηκαν παραπάνω μέσω μελετών προσομοιώσεων. Στην παράγραφο 4.1.3, δίνεται η σύγκριση ισχύος όλων των ελέγχων και βρέθηκε ότι ο προτεινόμενος μας νέος έλεγχος (new-test) είναι ο πιο ισχυρός σε σύγκριση με τους άλλους ελέγχους.

4.1.3 Σύγκριση όλων των ελέγχων μέσω μελετών προσομοιώσεων

Είναι σημαντική η σύγκριση της απόδοσης του προτεινόμενου Νέου Ελέγχου (new-test) με τους συνηθισμένους ελέγχους σε μικρά και μεσαία μεγέθη δειγμάτων. Τα δεδομένα με ομοιογενείς και ετερογενείς ομάδες διασπορών προέρχονται από κανονικές κατανομές. Ο Νέος Έλεγχος, ο έλεγχος του Bartlett, του Levene και ο μέγιστος F – έλεγχος ($F_{\max}$ - test) του Hartley εκτελούνται. Αυτή η διαδικασία επαναλαμβάνεται για 10,000 φορές για να προσδιοριστεί ο ρυθμός σφάλματος Τύπου Ι και η ισχύς.

Εμείς μελετάμε την απόδοση αυτών των ελέγχων υπό κανονικές συνθήκες. Από την στιγμή που όλοι οι έλεγχοι παραμένουν σταθεροί στην αλλαγή θέσης (location), θέτουμε τους κανονικούς μέσους να είναι 0. Για τον προσδιορισμό σφάλματος Τύπου Ι, δημιουργούμε έξι ομάδες παρατηρήσεων που ακολουθούν κανονική κατανομή. Μελετούνται ισορροπημένα δεδομένα συμπεριλαμβάνοντας 5, 10, 21, 31 και 61 παρατηρήσεις (subjects) ανά ομάδα καθώς επίσης και μη ισορροπημένα δεδομένα.

Τα αποτελέσματα στον πίνακα 4.1 δείχνουν πως ο Νέος Έλεγχος (New-Test) και ο έλεγχος του Bartlett διατηρούν σε καλό επίπεδο τον ρυθμό σφάλματος Τύπου Ι, ακόμη και για μικρού μεγέθους ή μη ισορροπημένα δεδομένα. Στον μέγιστο F – έλεγχο ($F_{\max}$ - test), οι βαθμοί ελευθερίας καθορίζονται από το μέγεθος της μεγαλύτερης ομάδας. Επομένως, ο μέγιστος F – έλεγχος ($F_{\max}$ - test) του Hartley δεν μπορεί να εφαρμοστεί σε μη ισορροπημένα δεδομένα. Οι προσομοιώσεις δείχνουν ότι ο ρυθμός σφάλματος Τύπου Ι του μέγιστου F – ελέγχου ($F_{\max}$ - test) αυξάνεται γρήγορα όσο η ανομοιογένεια μεταξύ των μεγεθών των δύο αυτών ομάδων μεγαλώνει. Σε σύγκριση με τον Νέο Έλεγχο (New-Test) και τον έλεγχο του Bartlett, ο έλεγχος του Levene απαιτεί μεγαλύτερο μέγεθος δείγματος, τουλάχιστον 30 παρατηρήσεις (subjects) ανά ομάδα, για να εγγυηθεί ότι το σφάλμα Τύπου Ι είναι σε καλό επίπεδο.

Μετά αποτιμούμε την ικανότητα να εντοπίζουμε την ετερογένεια της διασποράς για κανονικές περιπτώσεις. Όπως φαίνεται στον πίνακα 4.2, ο Νέος Έλεγχος υπερτερεί των τριών εναλλακτικών ελέγχων όταν η διασπορά μιας ομάδας είναι διαφορετική από αυτές των υπόλοιπων. Η αύξηση της αποτελεσματικότητας αποδίδεται στην συγχώνευση των δεδομένων, η οποία δημιουργεί περισσότερους βαθμούς ελευθερίας. Για διαφορετικές διασπορές μεταξύ των ομάδων, η ισχύς του Νέου Έλεγχου είναι μεγαλύτερη από αυτή του μέγιστου F - ελέγχου του Hartley, πολύ κοντά σ' αυτή του ελέγχου του Levene και ελαφρώς μικρότερη από αυτή του ελέγχου του Bartlett.

<u>Πίνακας 4.1.</u>

Ακριβής ρυθμός απόρριψης για ομογενείς κανονικές διασπορές (με $a = 0.05$)

Sample size	Hartley	Bartlett	Levene	New-test
(5, 5, 5, 5, 5, 5)	0.0504	0.0470	0.1044	0.0500
(10, 10, 10, 10, 10, 10)	0.0493	0.0512	0.0729	0.0476
(21, 21, 21, 21, 21, 21)	0.0455	0.0459	0.0600	0.0460
(31, 31, 31, 31, 31, 31)	0.0486	0.0491	0.0571	0.0484
(61, 61, 61, 61, 61, 61)	0.0484	0.0487	0.0521	0.0485
(11, 21, 31, 11, 11, 11)	0.4486	0.0502	0.0619	0.0509
(5, 21, 31, 11, 11, 11)	0.5786	0.0509	0.0664	0.0512

Πίνακας 4.2.

Ικανότητα εντοπισμού της ετερογένειας σε κανονικές διασπορές

Sample size	Standard deviation	Hartley	Bartlett	Levene	New-test
(5, 5, 5, 5, 5, 5)	(2, 1, 1, 1, 1, 1)	0.1748	0.2584	0.3161	0.3199
(5, 5, 5, 5, 5, 5)	(3, 1, 1, 1, 1, 1)	0.4154	0.5976	0.5818	0.6802
(5, 5, 5, 5, 5, 5)	(5, 1, 1, 1, 1, 1)	0.7988	0.9075	0.8583	0.9380
(5, 5, 5, 5, 5, 5)	(1, 2, 3, 1, 1, 1)	0.4219	0.6150	0.5829	0.5589
(5, 5, 5, 5, 5, 5)	(1, 1.5, 5, 5, 1, 1)	0.8724	0.9685	0.8992	0.8850
(5, 5, 5, 5, 5, 5)	(1, 1, 1, 2, 2, 2)	0.2132	0.3370	0.3643	0.2412
(5, 5, 5, 5, 5, 5)	(1, 1.5, 2, 2.5, 3, 3.5)	0.3549	0.4637	0.4371	0.3485
(10, 10, 10, 10, 10, 10)	(2, 1, 1, 1, 1, 1)	0.4723	0.5672	0.5416	0.6357
(10, 10, 10, 10, 10, 10)	(3, 1, 1, 1, 1, 1)	0.8794	0.9294	0.8897	0.9493
(10, 10, 10, 10, 10, 10)	(1, 2, 3, 1, 1, 1)	0.8846	0.9518	0.9054	0.9091
(10, 10, 10, 10, 10, 10)	(1, 1.5, 3, 1, 1, 1)	0.8603	0.9165	0.8742	0.9156
(21, 21, 21, 21, 21, 21)	(1.5, 1, 1, 1, 1, 1)	0.3932	0.4462	0.4213	0.5023
(21, 21, 21, 21, 21, 21)	(2, 1, 1, 1, 1, 1)	0.8730	0.9082	0.8733	0.9376
(21, 21, 21, 21, 21, 21)	(1, 1.5, 2, 1, 1, 1)	0.8670	0.9230	0.8808	0.8832
(31, 31, 31, 31, 31, 31)	(0.5, 1, 1, 1, 1, 1)	0.9832	0.9710	0.8960	0.9843
(31, 31, 31, 31, 31, 31)	(1.5, 1, 1, 1, 1, 1)	0.5684	0.6103	0.5704	0.6643
(31, 31, 31, 31, 31, 31)	(1, 1.3, 1.5, 1, 1, 1)	0.5828	0.6712	0.6115	0.6056
(61, 61, 61, 61, 61, 61)	(1.3, 1, 1, 1, 1, 1)	0.4966	0.5263	0.4816	0.5734
(61, 61, 61, 61, 61, 61)	(1.5, 1, 1, 1, 1, 1)	0.8954	0.9131	0.8787	0.9369
(61, 61, 61, 61, 61, 61)	(0.8, 1, 1.5, 1, 1, 1)	0.9845	0.9850	0.9692	0.9853
(11, 21, 31, 11, 11, 11)	(1.5, 1, 1, 1, 1, 1)	N/A	0.2477	0.2680	0.2861
(11, 21, 31, 11, 11, 11)	(1, 1.5, 1, 1, 1, 1)	N/A	0.3951	0.3688	0.4461
(11, 21, 31, 11, 11, 11)	(1, 1, 1.5, 1, 1, 1)	N/A	0.4825	0.4318	0.5336
(11, 21, 31, 11, 11, 11)	(1, 1.5, 2, 1, 1, 1)	N/A	0.8903	0.8157	0.8226
(11, 21, 31, 11, 11, 11)	(2, 1, 1.5, 1, 1, 1)	N/A	0.7336	0.6729	0.6157
(11, 21, 31, 11, 11, 11)	(2, 1.5, 1, 1, 1, 1)	N/A	0.7428	0.6976	0.6752
(5, 21, 31, 11, 11, 11)	(2, 1, 1, 1, 1, 1)	N/A	0.3463	0.3934	0.4135
(5, 21, 31, 11, 11, 11)	(1, 1.5, 1, 1, 1, 1)	N/A	0.3800	0.3670	0.4343
(5, 21, 31, 11, 11, 11)	(1, 1, 1.5, 1, 1, 1)	N/A	0.4528	0.4113	0.5105
(5, 21, 31, 11, 11, 11)	(1, 1.5, 2, 1, 1, 1)	N/A	0.8500	0.7680	0.7822
(5, 21, 31, 11, 11, 11)	(2, 1, 1.5, 1, 1, 1)	N/A	0.5921	0.5565	0.5365

Πίνακας 4.3.

Ακριβής ρυθμός απόρριψης για ομογενείς εκθετικές διασπορές (όταν $a = 0.05$ και οι διασπορές είναι 1)

Sample size	Hartley	Bartlett	Levene	New-test
(5, 5, 5)	0.053	0.05	0.079	0.05
(10, 10, 10)	0.049	0.051	0.058	0.049
(21, 21, 21)	0.057	0.058	0.056	0.054
(31, 31, 31)	0.067	0.058	0.064	0.056
(61, 61, 61)	0.056	0.062	0.069	0.06
(11, 21, 31)	0.189	0.043	0.047	0.04
(21, 31, 61)	0.224	0.045	0.044	0.039

Πίνακας 4.4.

Ικανότητα εντοπισμού της ετερογένειας σε εκθετικές διασπορές

Sample size	Standard deviation	Hartley	Bartlett	Levene	New-test
(5, 5, 5)	(2, 1, 1)	0.203	0.249	0.274	0.264
(5, 5, 5)	(3, 1, 1)	0.493	0.595	0.485	0.62
(5, 5, 5)	(5, 1, 1)	0.802	0.874	0.677	0.887
(5, 5, 5)	(1, 2, 3)	0.346	0.35	0.316	0.339
(5, 5, 5)	(1, 3, 5)	0.729	0.709	0.5	0.673
(5, 5, 5)	(1, 2, 5)	0.698	0.717	0.57	0.709
(10, 10, 10)	(2, 1, 1)	0.48	0.546	0.469	0.564
(10, 10, 10)	(3, 1, 1)	0.871	0.903	0.814	0.907
(10, 10, 10)	(1, 2, 3)	0.817	0.815	0.68	0.796
(10, 10, 10)	(1, 1.5, 3)	0.773	0.799	0.715	0.795
(21, 21, 21)	(1.5, 1, 1)	0.415	0.464	0.413	0.452
(21, 21, 21)	(2, 1, 1)	0.848	0.878	0.817	0.881
(21, 21, 21)	(1, 1.5, 2)	0.764	0.756	0.666	0.742
(31, 31, 31)	(1.5, 1, 1)	0.592	0.618	0.55	0.613
(31, 31, 31)	(2, 1, 1)	0.974	0.979	0.953	0.979
(31, 31, 31)	(1, 1.5, 2)	0.95	0.948	0.873	0.935
(61, 61, 61)	(1.3, 1, 1)	0.515	0.547	0.489	0.546
(61, 61, 61)	(1.5, 1, 1)	0.896	0.91	0.879	0.909
(61, 61, 61)	(0.8, 1, 1.5)	0.919	0.927	0.893	0.923
(11, 21, 31)	(1.5, 1, 1)	N/A	0.332	0.295	0.34
(11, 21, 31)	(1, 1.5, 1)	N/A	0.458	0.41	0.454
(11, 21, 31)	(1, 1, 1.5)	N/A	0.462	0.378	0.451
(11, 21, 31)	(2, 1, 1)	N/A	0.741	0.704	0.758
(11, 21, 31)	(1, 2, 1)	N/A	0.914	0.852	0.91
(11, 21, 31)	(1, 1, 2)	N/A	0.924	0.867	0.915

Σημείωση: Μελετούμε την ευρωστία των παραπάνω ελέγχων ως εξής:

Στον Πίνακα 4.3 παρουσιάζουμε την ίδια ανάλυση όπως και στον Πίνακα 4.1 με τη διαφορά ότι η βασική κατανομή είναι εκθετική. Όπως επίσης και στον Πίνακα 4.4 παρουσιάζουμε την ίδια ανάλυση με τον Πίνακα 4.2 με τη διαφορά ότι και εδώ η βασική κατανομή είναι εκθετική. Φαίνεται ότι στην περίπτωση της εκθετικής κατανομής το σφάλμα Τύπου Ι (Type I error) δεν είναι τόσο ελεγχόμενο όσο στην περίπτωση της κανονικής κατανομής.

4.2 Έλεγχος για την ομοιογένεια των διασπορών χρησιμοποιώντας εύρωστες σταθμισμένες εκτιμήσεις πιθανοφάνειας.

4.2.1 Εισαγωγή

Έχει παρατηρηθεί υπέρμετρη ευαισθησία όσο αφορά τον F - έλεγχο και τον έλεγχο του Bartlett για απόκλιση από την κανονικότητα, από μελέτες των Levene (1960), Brown και Forsythe (1974), Conover, Johnson and Johnson (1981), Tiku και Balakrishnan (1984) μεταξύ άλλων. Ωστόσο αυτό το πρόβλημα της μη-ευρωστίας των εκτιμήσεων παραδοσιακής κλίμακας και των αντίστοιχων ελέγχων, έναντι της απόκλισης από την υπόθεση της κανονικότητας ήταν γνωστό πριν την μελέτη του Levene. Μετά από ευρείας κλίμακας έρευνα με τη χρήση της μεθόδου Monte Carlo συγκρίνοντας πενήντα έξι ελέγχους, ο Conover et al., πρότειναν τρεις ελέγχους μεταξύ των οποίων ο έλεγχος των Brown-Forsythe υπολογίστηκε ως ο ευκολότερος. Μέσω αυτής της μελέτης με τη βοήθεια του ελέγχου των Brown και Forsythe παραπέμπουμε στον έλεγχο που χρησιμοποιεί τη διάμεσο και όχι τον αποκομμένο μέσο, για τον προσδιορισμό της έννοιας των απόλυτων αποκλίσεων. Επίσης, για συντομία θα αναφέρουμε παρακάτω τον έλεγχο των Brown-Forsythe ως ‘B-F έλεγχο’. Σ’ αυτή τη μελέτη, προτείνουμε μια εναλλακτική εύρωστη διαδικασία με τη χρήση της τροποποιημένης στατιστικής συνάρτησης του ελέγχου του Levene (1960). Στην τροποποίηση χρησιμοποιούμε τις σταθμισμένες εκτιμήτριες πιθανοφάνειας των μέσων των πληθυσμών (Markatou, Basu και Lindsay, 1996, 1997) χρησιμοποιώντας την ελάχιστη εκτίμηση απόστασης Hellinger (Hellinger distance) αντί για τους δειγματικούς μέσους οι οποίοι είναι μέγιστες εκτιμήτριες πιθανοφάνειας για το κανονικό μοντέλο. Οι σταθμισμένες εξισώσεις εκτίμησης της πιθανοφάνειας είναι ανάλογες με τις εξισώσεις μέγιστης πιθανοφάνειας με αρκετούς τρόπους. Σε διακεκριμένα μοντέλα αυτή η αναλογία αντιστοιχεί σε ένα πρόβλημα βελτιστοποίησης (optimisation). Χρησιμοποιώντας αυτή την προσέγγιση, οι Basu, Sarkar και Basu (1997) ανέπτυξαν μια χρήσιμη εύρωστη διαδικασία για τον έλεγχο της ισότητας των μέσων για δυο κανονικούς πληθυσμούς με άγνωστες αλλά ίσες διασπορές. Στην παρούσα μελέτη διεξάγουμε ένα πείραμα προσομοίωσης για τον έλεγχο της ισότητας των διασπορών δυο και τριών πληθυσμών και επεξηγείται η σχετική παρουσίαση του προτεινόμενου ελέγχου. Η διαδικασία του ελέγχου συγκρίνεται με τον B-F έλεγχο που συνιστάται, όσο αφορά το επίπεδο και την ισχύ, για κατανομές συμμετρικές και με παχιές ουρές (thick-tailed). Ωστόσο, για την λοξή (skewed) κατανομή ο B-F έλεγχος εμφανίζεται να έχει τιμές εμπειρικού μεγέθους (empirical size) πιο κοντά στο ονομαστικό επίπεδο (nominal level) από τον προτεινόμενο έλεγχο, ενώ ο τελευταίος εμφανίζει τιμές μεγαλύτερης ισχύος.

Το υπόλοιπο της μελέτης είναι οργανωμένο ως εξής: Στην παράγραφο 4.2.2 παρέχεται περιγραφή όλων των ελέγχων που θα χρησιμοποιήσουμε, στην παράγραφο 4.2.3 παρουσιάζεται και περιγράφεται ο προτεινόμενος εύρωστος έλεγχος. Στην παράγραφο 4.2.4 παρουσιάζονται μερικά Monte Carlo αποτελέσματα της συγκριτικής εκτέλεσης του προτεινόμενου ελέγχου και στην παράγραφο 4.2.5 παρουσιάζονται κάποια συμπεράσματα.

4.2.2 Περιγραφή των Ελέγχων

Έστω $\{x_{ij}, j=1,2,...,n_i\}$, $i=1,2,...,k$ είναι οι ανεξάρτητα τυχαία δείγματα από k πληθυσμούς. Έστω επίσης σ_i^2, $\overline{x}_i$ και $s_i^2=(n_i-1)^{-1}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i)^2$ αντιπροσωπεύουν την διασπορά του i-οστού πληθυσμού, τον δειγματικό μέσο και τη διασπορά του δείγματος αντίστοιχα. Όταν $k = 2$, για τον έλεγχο της μηδενικής υπόθεσης έχουμε $H_0 : \sigma_1^2 = \sigma_2^2$ για συνθήκες κανονικότητας και έναντι στην εναλλακτική υπόθεση κατά ένα παράγοντα (one-sided) π.χ., $H_a : \sigma_1^2 > \sigma_2^2$, ο πιο ομοιόμορφα ισχυρός έλεγχος βασίζεται στην στατιστική συνάρτηση $F = s_1^2 / s_2^2$. Ο Bartlett (1937) γενίκευσε αυτό τον έλεγχο για το λόγο πιθανοφάνειας στην περίπτωση όπου $k \geq 2$ πληθυσμούς. Ο F έλεγχος ή ο έλεγχος του Bartlett δεν είναι εξ' ολοκλήρου εύρωστοι ακόμη και ασυμπτωτικά λαμβάνοντας υπόψη και την απόκλιση από την κανονικότητα. Σε μικρά δείγματα, αυτοί οι έλεγχοι δεν είναι εύρωστοι ακόμη και για μικρές αποκλίσεις από την κανονικότητα. Οι πραγματικοί ρυθμοί σφαλμάτων τύπου Ι των ελέγχων μπορεί να είναι αρκετές φορές μεγαλύτεροι από το ονομαστικό επίπεδο (Box, 1953. Gartside, 1972). Αρκετοί εύρωστοι έλεγχοι έχουν μελετηθεί, μεταξύ άλλων, από τους Levene (1960), Miller (1968), Layard (1973), Flinger και Killeen (1976), Gartside (1972), Brown και Forsythe (1974), Conover (1981). Ο Conover πρότεινε τρεις ελέγχους όσο αφορά την ευρωστία και την ισχύ, οι οποίοι είναι ο B-F έλεγχος, ο χ^2 έλεγχος των Flinger και Killeen, και ο F έλεγχος των Flinger και Killeen χρησιμοποιώντας τη διάμεσο του δείγματος αντί για τον δειγματικό μέσο.

Στα ακόλουθα, το εύρος των τιμών του i θα είναι από $1, 2, \ldots, k$ και για το j από $1, 2, \ldots, n_i$ για k πληθυσμούς. Ο έλεγχοι που θα χρησιμοποιήσουμε, και παρουσιάζονται στο κεφάλαιο 2, είναι ο έλεγχος του Bartlett (παράγραφος 2.2.1.2) και ο έλεγχος του Levene (παράγραφος 2.2.1.12).

Οι Brown και Forsythe τροποποίησαν τον έλεγχο του Levene αντικαθιστώντας τον δειγματικό μέσο $\overline{Y}_{i.}$ με τον αποκομμένο μέσο ως μια εύρωστη εκτίμηση της κεντρικής θέσης (central location) του πληθυσμού. Αφού το στατιστικό του ελέγχου του Levene ορίζεται από τις σχέσεις (2.19) και (2.22) χρησιμοποιώντας τον μη αποκομμένο μέσο, οι Brown και Forsythe ονόμασαν αυτό το στατιστικό ως W_0. Στη συνέχεια με τη χρήση των σχέσεων (2.20) και (2.22) όρισαν το στατιστικό που προκύπτει ως W_{50}, όπου $\tilde{Y}_{i.}$ είναι η διάμεσος του δείγματος (50% αποκομμένος μέσος). Όμοια, όρισαν και το W_{10}.

4.2.3 Διαδικασία του προτεινόμενου στατιστικού ελέγχου.

4.2.3.1 Ορισμός

Υποθέτουμε ότι έχουμε ανεξάρτητα τυχαία δείγματα από k κανονικούς πληθυσμούς, το i-οστό δείγμα μεγέθους n_i προέρχεται από την $N(\mu_i, \sigma_i^2)$ κανονική κατανομή, με $i = 1, 2, ...k$. Οι σταθμισμένες εκτιμήτριες πιθανοφάνειας λαμβάνονται ως εξής: για το i-οστό δείγμα, έστω $\theta_i = (\mu_i, \sigma_i^2)'$ και F_{n_i} είναι το εμπειρικό της σωρευτικής συνάρτησης κατανομής (CDF). Για έναν κατάλληλο πυρήνα $k(x, y, h_i)$, η εκτίμηση πυκνότητας του πυρήνα $f_{n_i}^*(x)$ και το ομαλοποιημένο μοντέλο πυκνότητας του πυρήνα $m_{\theta_i}^*(x)$ είναι:

$$f_{n_i}^*(x) = \int k(x, y, h_i) dF_{n_i}(y), \qquad m_{\theta_i}^*(x) = \int k(x, y, h_i) dM_{\theta_i}(y).$$

Η συνάρτηση του πυρήνα $k(\cdot, y, h_i)$,που χρησιμοποιήθηκε στη μελέτη μας, είναι η πυκνότητα $N(y, h_i^2)$, όπου h_i είναι η παράμετρος ομαλοποίησης. Η ομαλοποιημένη πυκνότητα $m_{\theta_i}^*(x)$ είναι στην ουσία η κανονική πυκνότητα (ως συνέπεια της χρήσης ενός κανονικού πυρήνα). Οι Basu και Lindsay (1994) όρισαν το υπόλοιπο Pearson $\delta_i^*(x)$, μια τυποποιημένη εκδοχή του υπολοίπου, στο σημείο x ως

$$\delta_i^*(x) = \frac{f_{n_i}^*(x) - m_{\theta_i}^*(x)}{m_{\theta_i}^*(x)}.$$

Έπειτα με δοσμένη μια κυρτή συνάρτηση τρεις φορές διαφορήσιμη συνάρτηση G με πραγματικές τιμές στο διάστημα $[-1, \infty)$ με $G(0) = 0$, κατασκεύασαν την ανομοιότητα (disparity) που ορίζεται ως

$$\rho_G(f_{n_i}^*, m_{\theta_i}^*) = \int G(\delta_i^*(x)) m_{\theta_i}^*(x) dx.$$

Σημειώνουμε ότι για την απόσταση Hellinger (HD, Hellinger distance), $G(\delta^*) = 2[\sqrt{(\delta^* - 1)} - 1]^2$. Έστω ότι ο όρος ∇_θ δηλώνει τον (επί μέρους) διαφορήσιμο τελεστή όσο αφορά το θ. Σε αναλογία με τις εκτιμώμενες εξισώσεις οι οποίες αποτελούν λύση για τους ελαχιστοποιητές (minimizers) του $\rho_G(f_{n_i}^*, m_{\theta_i}^*)$ όσο αφορά το θ_i, μια μπορεί να κατασκευάσει και να επιλύσει τις ακόλουθες, 2 παραγόντων k, εξισώσεις σταθμισμένης πιθανοφάνειας (Markatou, Basu και Lindsay, 1996, 1997. Basu, Harris και Basu, 1997)

$$\mu_i : \int w_i(\delta_i^*(x))\nabla_{\mu_i}\{\log m_{\theta_i}\}dF_{n_i}(x) = \frac{1}{n_i}\sum_j w_i(\delta_i^*(x_{ij}))u_{\mu_i}(x_{ij},\theta_i) = 0,$$

$$\sigma_i^2 : \int w_i(\delta_i^*(x))\nabla_{\sigma_i^2}\{\log m_{\theta_i}\}dF_{n_i} = \frac{1}{n_i}\sum_j w_i(\delta_i^*(x_{ij}))u_{\sigma_i^2}(x_{ij},\theta_i) = 0,$$

για $i = 1,2,...,k$ όπου

$$w_i(\delta_i^*(x)) = \frac{A(\delta_i^*(x))+1}{\delta_i^*(x)+1}, \qquad A(\delta^*) \equiv (\delta^*+1)\dot{G}(\delta^*) - G(\delta^*). \tag{4.1}$$

$\dot{G}(\delta^*)$ δηλώνει την πρώτη παράγωγο του $G(\delta^*)$, και $u_{\mu_i}(x)$ είναι η προκείμενη συνάρτηση πιθανοφάνειας $\nabla_{\mu_i}\log m_{\theta_i}$. Ομοίως για $u_{\sigma_i^2}(x)$. Αυτές οι εκτιμώμενες εξισώσεις 2 παραγόντων k μπορούν να συνοψιστούν ως εξής:

$$\mu_i : \sum_j w_i(\delta_i^*(x_{ij}))(x_{ij}-\mu_i) = 0, \qquad i = 1,2,...,k,$$

$$\sigma_i^2 : \sum_j w_i(\delta_i^*(x_{ij}))[(x_{ij}-\mu_i)^2 - \sigma_i^2] = 0, \qquad i = 1,2,...,k. \tag{4.2}$$

Οι παραπάνω εξισώσεις επιλύονται επαναληπτικά από ένα απλό αλγόριθμο σταθερού σημείου *(fixed point)* δημιουργώντας νέα βάρη (weights) $w_i(\cdot)$ σε κάθε στάδιο. Οι λύσεις από αυτές τις εξισώσεις είναι της μορφής

$$\hat{\mu}_i = \frac{\sum_j w_i(\delta_i^*(x_{ij}))x_{ij}}{\sum_j w_i(\delta_i^*(x_{ij}))}, \qquad i = 1,2,...,k,$$

$$\hat{\sigma}_i^2 = \frac{\sum_j w_i(\delta_i^*(x_{ij}))(x_{ij}-\hat{\mu}_i)^2}{\sum_j w(\delta_i^*(x_{ij}))}, \qquad i = 1,2,...,k.$$

Στη συνέχεια ορίζουμε τον προτεινόμενο στατιστικό έλεγχο. Έστω $\hat{\mu}_i,\ i=1,2,...,k$ οι σταθμισμένες εκτιμήτριες πιθανοφάνειας από τους k μέσους των πληθυσμών, όπου τα βάρη (weights) από αυτές υπολογίζονται χρησιμοποιώντας τη συνάρτηση της απόστασης Hellinger με προσαρμοσμένο υπόλοιπο *(the Hellinger distance residual adjustment function).*

$$A(\delta^*) = 2[\sqrt{\delta^* + 1} - 1]. \tag{4.3}$$

Έπειτα ο τροποποιημένος έλεγχος του Levene με σταθμισμένες εκτιμήτριες πιθανοφάνειας ορίζεται από την

$$z_{ij} = | x_{ij} - \hat{\mu}_i |, \tag{4.4}$$

και συμβολίζεται με W_{HD}. Η προσέγγιση των κρίσιμων σημείων του W_{HD} λαμβάνεται από την $F^{k-1}_{\sum_i (n_i - 1)}$ κατανομή. (Lindsay, 1994. Basu και Lindsay, 1994)

4.2.3.2 Μεθοδολογία

Οι Brown και Forsythe βρήκαν, κατά τη διάρκεια της δειγματοληψίας τους, ότι ο έλεγχος του Levene δεν ήταν εύρωστος όταν οι σημαντικοί πληθυσμοί ακολουθούσαν την λοξή (skewed) κατανομή. Αυτό τους βοήθησε να σκεφτούν περισσότερο για εύρωστες εκτιμήσεις την κεντρικής θέσης *(central location)* από ότι τον δειγματικό μέσο στον υπολογισμό των απόλυτων αποκλίσεων. Θεώρησαν τη διάμεσο και τον 10% αποκομμένο μέσο ως εύρωστες εναλλακτικές του δειγματικού μέσου, οι οποίοι αποτελούν τη μέγιστη εκτιμήτρια πιθανοφάνειας (ΕΜΠ). Αλλά αυτές οι εκτιμήτριες δεν είναι ασυμπτωτικά τόσο αποδοτικές όσο η ΕΜΠ για κανονικές συνθήκες, π.χ. θυσίασαν μέγιστη ασυμπτωτική αποδοτικότητα για να επιτύχουν ευρωστία. Είναι γνωστό ότι οι ελάχιστες εκτιμήτριες της απόστασης Hellinger (Hellinger distance) των παραμέτρων του πληθυσμού έχουν ισχυρές ιδιότητες ευρωστίας ενώ παράλληλα είναι ασυμπτωτικά αποδοτικές στο μοντέλο (Beran, 1977. Basu και Lindsay, 1994). Συνεπώς, μια ίσως προσδοκία είναι ότι, για συνθήκες κανονικότητας, η χρήση της ελάχιστης εκτιμήτριας της απόστασης Hellinger αντί για τη διάμεσο του δείγματος στους υπολογισμούς των απόλυτων αποκλίσεων του ελέγχου του Levene θα οδηγούσε σε καλύτερη απόδοση του μοντέλου όσο αφορά το εμπειρικό επίπεδο (empirical level) από ότι στον B-F έλεγχο, διατηρώντας το παράλληλα εύρωστο έναντι της μη-κανονικότητας. Από τη στιγμή που ο υπολογισμός αυτών των εκτιμητριών είναι αρκετά πολύπλοκος, στρεφόμαστε στις σταθμισμένες εκτιμήτριες πιθανοφάνειας της Markatou (1996,1997) που βασίζονται στην απόσταση Hellinger.

4.2.3.3 Η επιλογή της σταθμισμένης πιθανοφάνειας

Όπως αναφέρθηκε νωρίτερα, οι σταθμισμένες εξισώσεις εκτίμησης της πιθανοφάνειας είναι, με διάφορους τρόπους, ανάλογες με τη score εξίσωση μέγιστης πιθανοφάνειας. Σε διακεκριμένα μοντέλα, όλες οι ελάχιστης ανομοιότητας (minimum disparity) μέθοδοι των Cressie και Read (1984) και Lindsay (1994) – από τις οποίες η μέγιστη πιθανοφάνεια είναι ειδική περίπτωση – συμπίπτουν ακριβώς με τις διαδικασίες σταθμικής πιθανοφάνειας (Markatou et al., 1997). Στην περίπτωση της μέγιστης πιθανοφάνειας, η οποία ελαχιστοποιεί μια εκδοχή της απόκλισης (divergence) των Kullback-Leibler, τα βάρη (weights) καταλήγουν να είναι ίσα με μονάδα (το οποίο είναι στην ουσία ο λόγος που παρουσιάζουν έλλειψη ευρωστίας. Σε συνεχή μοντέλα αυτό δεν είναι εφικτό όσο τα δεδομένα είναι διακριτά (discrete) και το μοντέλο είναι συνεχές. Αυτό απαιτεί τη χρήση μερικών μη παραμετρικών εκτιμητριών πυκνότητας όπως η εκτιμήτρια πυκνότητας του πυρήνα. Ωστόσο η προσέγγιση της σταθμισμένης εξίσωσης πιθανοφάνειας σε συνεχή μοντέλα στην πραγματικότητα αντιπροσωπεύει μια σημαντική βελτίωση όσο αφορά την υλοποίηση του χρόνου της πραγματικής προσέγγισης ελάχιστης ανομοιότητας (disparity) από τη στιγμή που οι εκτιμώμενες εξισώσεις είναι αθροίσματα των παρατηρούμενων δεδομένων και όχι ολοκληρώματα ολικής ενίσχυσης (support) της κατανομής.

Μέθοδοι ελάχιστης απόστασης όπως αυτοί που προτείνονται από τους Beran (1977), Tamura και Boos (1986), Simpson (1987, 1989), Basu και Lindsay (1994) και άλλους έχουν τη δική τους εξέχουσα θέση στη βιβλιογραφία εφ’ όσον πέτυχαν αρτια, πλήρη ασυμπτωτική αποδοτικότητα μαζί με σημαντικές ιδιότητες ευρωστίας αντίθετα με τις Μ-εκτιμήτριες οι οποίες τυπικά πέτυχαν την ευρωστία με κόστος την χαμηλή αποδοτικότητα του μοντέλου. Παρόμοιες ιδέες έχουν επίσης χρησιμοποιηθεί για να αναζητηθούν εύρωστες και αποδοτικές διαδικασίες ελέγχων σε γενικότερο πλαίσιο. (όπως για παράδειγμα, οι Simpson, 1989, Basu, 1996, 1997 και άλλοι).

Η χρήση της μεθόδου σταθμισμένης πιθανοφάνειας στο γενικότερο πλαίσιο όπως προτάθηκε στην παράγραφο 4.2.3.1 έχει αρκετά πλεονεκτήματα. Αρχικά, αυτές οι εξισώσεις δεν απαιτούν αριθμητικές λύσεις από ολοκληρώματα, όπως είναι αναγκαίο στις άμεσες μεθόδους ελάχιστης ανομοιότητας. Δεύτερον, η ομαλοποιημένη πυκνότητα δεδομένων $f^*_{n_i}$ υπολογίζεται μόνο μια φορά στην αρχή της επαναληπτικής διαδικασίας και δεν χρειάζεται να ανανεώνεται σε κάθε στάδιο της επανάληψης. Στην ουσία αυτό είναι το μόνο σημείο στη διαδικασία εκτίμησης που η ομαλοποίηση με ένα πυρήνα είναι άμεσα συνδεδεμένη με την υλοποίηση. Τρίτον, το ομαλοποιημένο i-οστό μοντέλο πυκνότητας $m^*_{\theta_i}$ σ' αυτό πλαίσιο είναι απλά το $N(\mu_i, \sigma_i^2 + h_i^2)$, έτσι ώστε να μην υπάρχουν επιπρόσθετες επιπλοκές σχετικές με την ομαλοποίηση όταν οι παράμετροι του μοντέλου είναι ενημερωμένες σε κάθε στάδιο της επανάληψης. Τέταρτον, από τη στιγμή που το μοντέλο $\delta^*(x)$ συγκλίνει στο (σημείο) μηδέν με πιθανότητα 1, όλα τα βάρη (weights) συγκλίνουν στο 1, έτσι ώστε για μεγάλα n οι εξισώσεις σταθμισμένης εκτίμησης πιθανοφάνειας συμπεριφέρονται όπως οι score εξισώσεις μέγιστης πιθανοφάνειας *(maximum likelihood score equations).* Από την άλλη μεριά, τα βάρη σ' αυτή την περίπτωση δικαιολογούν την ευρωστία τη διαδικασίας, από τη στιγμή που για μεγάλες απόμακρες (outlying) τιμές μπορούν να παρέχουν ένα σταθερό αποτέλεσμα, μια συνέπεια της αμβλυμμένης (dampened) απόκρισης της συνάρτησης προσαρμοσμένου υπολοίπου της απόστασης Hellinger. Τελικά, από τη στιγμή που ομαλοποιούμε τη συνοχή των δεδομένων και του μοντέλου της εκτιμήτριας (σε αντίθεση με τις προσεγγίσεις των Beran, 1977. Tamura και Boos, 1986. Simpson, 1989), λαμβάνουμε (για) μια σταθερή τιμή της παραμέτρου ομαλοποίησης h, και αυτή η παράμετρος δεν χρειάζεται να συγκλίνει στο μηδέν αφού το μέγεθος του δείγματος συγκλίνει στο ∞ (άπειρο).

4.2.4 Αποτελέσματα προσομοίωσης

Από τη στιγμή που ο B-F έλεγχος είναι ένας από τους πλέον προτεινόμενους ελέγχους μεταξύ άλλων και εύκολος στον υπολογισμό, εμείς θα ερευνήσουμε την απόδοση του προτεινόμενου ελέγχου W_{HD} σχετικό με το W_{50} και του F ή του ελέγχου του Bartlett. Μελετούμε τον λόγο του F ελέγχου στην περίπτωση που έχουμε δυο πληθυσμούς και τον έλεγχο του Bartlett στην περίπτωση τριών πληθυσμών. Τέσσερις διαφορετικές κατανομές συμπεριλαμβάνονται στις προσομοιώσεις. Αυτοί είναι (i) η κανονική κατανομή $N(0,1)$, (ii) μια μικτή κατανομή: $0.9N(0,1)+0.1N(0,8)$, (iii) μια κατανομή με μακριά ουρά (long-tailed): $t(4)$ και (iv) μια λοξή (skew) κατανομή: χ^2. Οι κατανομές (i), (iii) και (iv) μελετήθηκαν από τους Brown και Forsythe στην προσομοίωσή τους. Εμείς μελετήσαμε επιπρόσθετα την συνδυαστική κατανομή (ii). Για δυο πληθυσμούς τα μεγέθη δειγμάτων που μελετήθηκαν είναι $(n_1,n_2)=(40,40)$, $(10,10)$, $(10,20)$ και $(20,40)$ όπως και από τους Brown και Forsythe. Στη μελέτη μας, αντίθετα με τους παραπάνω, πραγματοποιήθηκαν επίσης προσομοιώσεις για τρεις πληθυσμούς και τα αντίστοιχα μεγέθη δειγμάτων που χρησιμοποιήθηκαν είναι $(n_1,n_2,n_3)=(40,40,40)$, $(10,10,10)$, $(10,10,40)$, $(20,20,40)$ και $(10,20,40)$.

Πίνακας 4.5.

Εμπειρικό μέγεθος και ισχύς των ελέγχων για την ισότητα των διασπορών των δυο πληθυσμών αντίστοιχα για την κανονική κατανομή $N(0,1)$ ***και την συνδυαστική*** $0.9N(0,1)+0.1N(0,8)$ ***για*** $a=0.05$ ***και μεγέθη δειγμάτων*** $(n_1,n_2)=(40,40)$, $(10,10)$, $(10,20)$ ***και*** $(20,40)$.

	$N(0,1)$					
	F	W_{50}	W_{HD}	F	W_{50}	W_{HD}
$(\sigma_1^2 : \sigma_2^2)$	(40, 40)			(10, 10)		
(1 : 1)	0.046	0.043	0.051	0.055	0.035	0.054
(1 : 2)	0.561	0.463	0.495	0.151	0.100	0.144
(1 : 4)	0.990	0.968	0.973	0.510	0.315	0.405
(1 : 8)	1.000	1.000	1.000	0.834	0.593	0.681
	(10, 20)			(20, 40)		
(1 : 1)	0.051	0.044	0.057	0.052	0.041	0.051
(1 : 2)	0.190	0.113	0.159	0.376	0.295	0.322
(2 : 1)	0.216	0.175	0.203	0.410	0.340	0.370
(1 : 4)	0.603	0.399	0.487	0.934	0.843	0.865
(4 : 1)	0.664	0.543	0.584	0.927	0.865	0.882
(1 : 8)	0.937	0.748	0.823	1.000	0.996	0.997
(8 : 1)	0.931	0.839	0.864	0.999	0.995	0.996

	$0.9N(0,1)+0.1N(0,8)$					
	F	W_{50}	W_{HD}	F	W_{50}	W_{HD}
$(\sigma_1^2 : \sigma_2^2)$	(40, 40)			(10, 10)		
(1 : 1)	0.233	0.043	0.047	0.154	0.038	0.054
(1 : 2)	0.552	0.346	0.362	0.256	0.083	0.111
(1 : 4)	0.921	0.874	0.880	0.502	0.232	0.298
(1 : 8)	0.997	0.996	0.996	0.762	0.489	0.551
	(10, 20)			(20, 40)		
(1 : 1)	0.181	0.040	0.049	0.224	0.043	0.049
(1 : 2)	0.304	0.077	0.107	0.463	0.200	0.217
(2 : 1)	0.280	0.152	0.167	0.433	0.273	0.290
(1 : 4)	0.619	0.277	0.338	0.834	0.658	0.680
(4 : 1)	0.586	0.445	0.467	0.818	0.753	0.764
(1 : 8)	0.845	0.556	0.618	0.977	0.938	0.944
(8 : 1)	0.840	0.754	0.766	0.979	0.969	0.973

<u>Πίνακας 4.6.</u>

Εμπειρικό μέγεθος και ισχύς των ελέγχων για την ισότητα των διασπορών των δυο πληθυσμών αντίστοιχα για τις κατανομές $t(4)$ και χ^2 για $a = 0.05$ και μεγέθη δειγμάτων $(n_1, n_2) = (40, 40)$, $(10,10)$, $(10, 20)$ και $(20, 40)$.

$t(4)$						
	F	W_{50}	W_{HD}	F	W_{50}	W_{HD}
$(\sigma_1^2 : \sigma_2^2)$	(40, 40)			(10, 10)		
(1 : 1)	0.237	0.042	0.047	0.152	0.036	0.053
(1 : 2)	0.546	0.321	0.337	0.258	0.087	0.115
(1 : 4)	0.913	0.854	0.862	0.501	0.235	0.289
(1 : 8)	0.993	0.992	0.993	0.761	0.450	0.519
	(10, 20)			(20, 40)		
(1 : 1)	0.163	0.040	0.053	0.220	0.045	0.052
(1 : 2)	0.314	0.082	0.111	0.467	0.181	0.200
(2 : 1)	0.307	0.166	0.177	0.444	0.265	0.280
(1 : 4)	0.604	0.251	0.306	0.835	0.624	0.644
(4 : 1)	0.588	0.433	0.454	0.825	0.735	0.752
(1 : 8)	0.862	0.537	0.596	0.977	0.922	0.927
(8 : 1)	0.845	0.738	0.752	0.974	0.966	0.968

χ_4^2						
	(40, 40)			(10, 10)		
$(\sigma_1^2 : \sigma_2^2)$	F	W_{50}	W_{HD}	F	W_{50}	W_{HD}
(1 : 1)	0.191	0.055	0.086	0.142	0.078	0.090
(1 : 2)	0.550	0.353	0.408	0.238	0.140	0.156
(1 : 4)	0.942	0.879	0.897	0.509	0.331	0.343
(1 : 8)	0.998	0.997	0.997	0.768	0.555	0.554
	(10, 20)			(20, 40)		
(1 : 1)	0.151	0.067	0.084	0.182	0.056	0.083
(1 : 2)	0.295	0.121	0.163	0.447	0.213	0.274
(2 : 1)	0.278	0.198	0.195	0.421	0.272	0.308
(1 : 4)	0.608	0.320	0.392	0.849	0.665	0.716
(4 : 1)	0.599	0.480	0.467	0.843	0.752	0.766
(1 : 8)	0.872	0.610	0.664	0.988	0.948	0.958
(8 : 1)	0.859	0.776	0.764	0.984	0.966	0.970

Για δυο πληθυσμούς, δύο τυχαία μεγέθη n_1, n_2 χρησιμοποιήθηκαν στον υπολογισμό των στατιστικών ελέγχων. Οι παρατηρήσεις του δεύτερου δείγματος αναδημιουργήθηκαν σε μικρότερη κλίμακα (rescaled) για να εκφράσουν το λόγο των διασπορών των πληθυσμών των δυο ομάδων. Οι λόγοι διασπορών $(\sigma_1^2 : \sigma_2^2)$ που μελετήθηκαν είναι $(1:1)$, $(1:2)$, $(1:4)$ και $(1:8)$ για την περίπτωση ίσων μεγεθών των δειγμάτων και $(1:2)$, $(1:4)$, $(1:8)$, $(2:1)$, $(4:1)$ και $(8:1)$ για την περίπτωση διαφορετικών μεγεθών των δειγμάτων.

Για την περίπτωση των τριών πληθυσμών, μελετήσαμε τους λόγους διασπορών $(\sigma_1^2 : \sigma_2^2 : \sigma_3^2) = (1:1:1)$, $(1:1:2)$, $(1:1:4)$, $(1:2:4)$, $(4:2:1)$, $(4:1:1)$ και $(2:1:1)$. Όλοι οι στατιστικοί έλεγχοι υπολογίστηκαν στο ίδιο σύνολο δυο ή τριών ομάδων παρατηρήσεων. Αυτοί επαναλαμβάνονται πέντε χιλιάδες φορές για κάθε κατανομή. Για ονομαστικά (nominal) επίπεδα $\alpha = 0.05$ και 0.01 υπολογίζουμε το ποσοστό των απορρίψεων για κάθε διαδικασία ελέγχου χρησιμοποιώντας τις κατάλληλες κρίσιμες τιμές από την χ^2 ή την F κατανομή. Αυτό το ποσοστό δείχνει το πρακτικό μέγεθος των ίσων διασπορών και την πρακτική διασπορά των άνισων διασπορών. Υπολογίζοντας τη σταθμισμένη εκτίμηση πιθανοφάνειας της απόστασης Hellinger *(HDWLE)* για τις $2k$ παραμέτρους πληθυσμού χρησιμοποιήσαμε bandwidth $h_i = 0.5 \times MAD_i$ για το i-οστό δείγμα, όπου

$$MAD_i = median\{|x_{ij} - \tilde{x}_i|, j = 1, 2, ..., n_i\}$$

και $\tilde{x}_i = median\{x_{ij}, j = 1, 2, ..., n_i\}$. Οι σταθμισμένες εκτιμήσεις πιθανοφάνειας της απόστασης Hellinger *(HDWLEs)* των διασπορών υπολογίζονται για να λύσουμε τις εξισώσεις (4.2). Για συντομία τα αποτελέσματα των προσομοιώσεων δίνονται μόνο από το ονομαστικό επίπεδο 0.05 (τα οποία είναι όμοια για την περίπτωση όπου το κατηγορικό επίπεδο (nominal level) είναι ίσο με 0.01). Αυτά παρουσιάζονται στους πίνακες 4.5 και 4.6 για δυο πληθυσμούς για τις 4 κατανομές (i)-(iv). Για τρεις πληθυσμούς, για λόγους συντομίας, παρουσιάζουμε τα αποτελέσματα μόνο για τις δυο πρώτες κατανομές (i)-(ii) στους πίνακες 4.7 και 4.8. Όπως περιμέναμε, όταν οι προκείμενες κατανομές είναι μη-κανονικές, απορρίπτονται από τον παραδοσιακό F έλεγχο και τον έλεγχο του Bartlett. Για συμμετρικές κατανομές ο B-F έλεγχος είναι συντηρητικός όσο αφορά το εμπειρικό επίπεδο όταν οι προκείμενες κατανομές είναι κανονικές, όπως αποδείχθηκε από τους Brown και Forsythe. Ο δικός μας W_{HD} έλεγχος έχει παρουσιάσει μια εξαιρετική απόδοση για κατανομές μεικτές (mixed) και με μακριές ουρές (long-tailed) πετυχαίνοντας πολύ πιο σταθερό ρυθμό σφάλματος τύπου I και την ίδια στιγμή έχει μεγαλύτερη ισχύ από τον W_{50} έλεγχο. Το πραγματικό σφάλμα τύπου I αυτού του ελέγχου είναι συχνά σημαντικά μικρότερο από τα ονομαστικά επίπεδά του. Για τη λοξή (skew) κατανομή χ_4^2, ο W_{HD} έλεγχος έχει υψηλότερο εμπειρικό επίπεδο από ότι ονομαστικό επίπεδο, και εξαιτίας αυτού έχει μεγαλύτερη ισχύ από τον W_{50}.

<u>Πίνακας 4.7.</u>

Εμπειρικό μέγεθος και ισχύς των ελέγχων για την ισότητα των διασπορών των τριών πληθυσμών σχετικά με την κανονική κατανομή $N(0,1)$ ***για*** $a = 0.05$ ***και μεγέθη δειγμάτων*** $(n_1, n_2, n_3) = (40,40,40)$, $(10,10,10)$, $(10,10,40)$, $(20,20,40)$ και $(10,20,40)$.

	B	W_{50}	W_{HD}	B	W_{50}	W_{HD}
$(\sigma_1^2 : \sigma_2^2 : \sigma_3^2)$		(40, 40, 40)			(10, 10, 10)	
(1 : 1 : 1)	0.054	0.047	0.060	0.051	0.034	0.061
(1 : 1 : 2)	0.605	0.512	0.546	0.159	0.113	0.158
(1 : 1 : 4)	0.994	0.984	0.987	0.546	0.365	0.455
(1 : 2 : 4)	0.978	0.934	0.942	0.389	0.235	0.323
		(10, 10, 40)			(20, 20, 40)	
(1 : 1 : 1)	0.053	0.042	0.057	0.047	0.041	0.055
(1 : 1 : 2)	0.269	0.170	0.228	0.447	0.332	0.378
(1 : 1 : 4)	0.841	0.660	0.741	0.973	0.920	0.937
(1 : 2 : 4)	0.612	0.415	0.506	0.880	0.760	0.802
(4 : 2 : 1)	0.702	0.617	0.640	0.892	0.796	0.826
(4 : 1 : 1)	0.692	0.606	0.633	0.910	0.846	0.872
(2 : 1 : 1)	0.215	0.185	0.211	0.387	0.307	0.347
		(10, 20, 40)				
(1 : 1 : 1)	0.046	0.044	0.059			
(1 : 1 : 2)	0.376	0.272	0.326			
(1 : 1 : 4)	0.946	0.848	0.884			
(1 : 2 : 4)	0.669	0.490	0.563			
(4 : 2 : 1)	0.727	0.633	0.668			
(4 : 1 : 1)	0.707	0.641	0.663			
(2 : 1 : 1)	0.229	0.208	0.222			

Πίνακας 4.8.

Εμπειρικό μέγεθος και ισχύς των ελέγχων για την ισότητα των διασπορών των τριών πληθυσμών σχετικά με την συνδυαστική κατανομή $0.9N(0,1)+0.1N(0,8)$ ***για*** $a=0.05$ ***και μεγέθη δειγμάτων*** $(n_1,n_2,n_3)=(40,40,40)$, $(10,10,10)$, $(10,10,40)$, $(20,20,40)$ και $(10,20,40)$.

	B	W_{50}	W_{HD}	B	W_{50}	W_{HD}
$(\sigma_1^2:\sigma_2^2:\sigma_3^2)$		(40,40,40)			(10,10,10)	
(1:1:1)	0.333	0.041	0.047	0.221	0.037	0.057
(1:1:2)	0.664	0.367	0.388	0.315	0.083	0.118
(1:1:4)	0.962	0.928	0.935	0.584	0.281	0.342
(1:2:4)	0.916	0.795	0.814	0.489	0.178	0.232
		(10,10,40)			(20,20,40)	
(1:1:1)	0.246	0.040	0.046	0.307	0.043	0.052
(1:1:2)	0.466	0.093	0.121	0.593	0.223	0.247
(1:1:4)	0.829	0.403	0.460	0.927	0.764	0.786
(1:2:4)	0.718	0.237	0.293	0.834	0.538	0.563
(4:2:1)	0.627	0.490	0.504	0.805	0.634	0.666
(4:1:1)	0.650	0.495	0.504	0.851	0.708	0.735
(2:1:1)	0.340	0.164	0.175	0.495	0.225	0.249
		(10,20,40)				
(1:1:1)	0.282	0.046	0.054			
(1:1:2)	0.544	0.159	0.184			
(1:1:4)	0.894	0.643	0.680			
(1:2:4)	0.746	0.312	0.361			
(4:2:1)	0.675	0.496	0.515			
(4:1:1)	0.669	0.520	0.529			
(2:1:1)	0.398	0.169	0.175			

4.2.5. Συμπεράσματα της μελέτης

Είναι επιθυμητό, για τον έλεγχο της ομοιογένειας των διασπορών, να έχουμε μια διαδικασία ελέγχου που έχει την ιδιότητα να γίνεται σταθερή σε μεταβλητότητες για την προκείμενη κατανομή (ευρωστία), και την ίδια στιγμή ευαίσθητη στις αποκλίσεις από την υπόθεση ίσων διασπορών (ισχύς). Ο Conover et al. (1981) πραγματοποίησαν μια πολύ μεγάλη πρακτική μελέτη περιλαμβάνοντας διάφορες κατανομές, μεγέθη δειγμάτων, μέσους και διασπορές ώστε να βρει τέτοιους ελέγχους που να εμφανίζουν σταθερό ρυθμό σφάλματος Ι και λογική ισχύ. Αυτοί βρήκαν τρεις ελέγχους να αποτελούν τις καλύτερες επιλογές όσο αφορά την ευρωστία και την ισχύ, με τον ευκολότερο στον υπολογισμό να είναι ο W_{50} στατιστικός έλεγχος των Brown και Forsythe. Ο Conover et al. δεν πρότειναν τη χρήση του ελέγχου W_{10}, περιλαμβάνοντας τον αποκομμένο μέσο, όπως έκαναν οι Brown και Forsythe σε μεγάλο βαθμό στην περίπτωση των συμμετρικών αποκλίσεων (departures) από την κανονικότητα, επειδή τα αποτελέσματα δεν παρουσίασαν σημαντικά πλεονεκτήματα με τη χρήση της μεταβλητότητας. Οι Brown και Forsythe δήλωσαν ότι οι μελλοντικοί εύρωστοι έλεγχοι για την ομοιογένεια της διασποράς πρέπει να συμπεριλάβουν τις εύρωστες εκτιμήσεις της κεντρικής θέσης (central location). Επομένως, για την βελτίωση του ελέγχου W_{50}, θεωρήσαμε σ' αυτή τη μελέτη την σταθμισμένη εκτίμηση πιθανοφάνειας της απόστασης Hellinger ως μια εύρωστη εκτίμηση του πληθυσμού κεντρικής θέσης (central location), όπως η δειγματική διάμεσος. Αλλά αντίθετα με τη διάμεσο, είναι ασυμπτωτικά τόσο αποδοτική όσο η εκτιμήτρια μέγιστης πιθανοφάνειας (ΕΜΠ) όταν οι προκείμενες κατανομές του πληθυσμού είναι κανονικές. Αυτό είναι που μας οδήγησε να ορίσουμε τον στατιστικό έλεγχο W_{HD} και να εξετάσουμε την απόδοσή του σε σχέση με τον W_{50} και τον τυπικό F έλεγχο ή τον έλεγχο του Bartlett. Τα αποτελέσματα της προσομοίωσής μας έδειξαν ότι ο έλεγχος W_{HD} είναι πιο ισχυρός από τον W_{50} όταν οι προκείμενες κατανομές είναι μικτές (mixed) και με μακριές (long-tailed) ουρές και εμφανίζει πιο σταθερό σφάλμα τύπου Ι σε αυτές τις συνθήκες. Ωστόσο, για την λοξή (skew) κατανομή χ_4^2, ο W_{HD} έλεγχος εμφανίζει λιγότερο σταθερό ρυθμό σφάλματος τύπου Ι (μεγαλύτερο από το ονομαστικό επίπεδο) ενώ πετυχαίνει μεγαλύτερη ισχύ από τον W_{50} έλεγχο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ansari, A. R., and Bradley, R. A., 1960: Rank-Sum Tests for Dispersion. *Annals of Mathematical Statistics*, **31**, pp. 1174-1189.

Bartlett, M. S., 1937: Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society Series A*, **160**, pp. 268-282.

Bartlett, M. S., and Kendall, D. G., 1946: The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation. *Journal of the Royal Statistical Society*, **8**, pp. 128-138.

Barton, D. E., and David, F. N., 1958: A Test for Birth Order Effects, *Annals of Eugenics*, **22**, pp. 250-257.

Bhandary, Madhusudan and Dai, Hongying, 2009: An Alternative Test for the Equality of Variances for Several Populations When the Underlying Distributions are Normal. *Communications in Statistics: Simulation and Computation*, **38**, pp. 109-117.

Box, G. E. P., 1953: Non-normality and tests on variances. *Biometrika*, **40**, pp. 318-335.

Box, G. E. P., and Anderson, S. L., 1955: Permutation Theory in the Derivation of Robust Criteria and the Study of Departures From Assumption. *Journal of the Royal Statistical Society, Ser. B*, **17**, pp. 1-26.

Basu, A., Harris, I. R., and Basu, S., 1996: Tests of hypothesis in discrete models based on the penalized Hellinger distance. *Statistics & Probability Letters*, **27**, pp. 367-373.

Basu, A., Harris, I. R., and Basu, S., 1997: Minimum distance estimation: The approach using density based distances. In: G. S. Maddala and C. R. Rao, (eds.): *Handbook of Statistics*, **15**. Elsevier Science B. V., Amsterdam, The Netherlands, pp. 21-48.

Basu, A. and Lindsay, B. G., 1994: Minimum Disparity estimation in the continuous Case: Efficiency, Distributions and Robustness. *Annals of Institute of Statistical Mathematics*, **46**, pp. 683-705.

Basu, S., Sarkar, S., and Basu, A., 1997: Robust tests for equality of two population means under the normal model. *Communications in Statistics: Simulation and Computation*, **26**, pp. 333-353.

Beran, R. J., 1977: Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, **5**, pp. 445-463.

Brown, M. B. and Forsythe, A. B., 1974: Robust test for the equality of variance. *Journal of American Statistical Association*, **69**, pp. 364-367.

Cadwell, J. H., 1953: Approximating to the Distributions of Measures of Dispersion by a Power of X^2. *Biometrika*, **40**, pp. 336-346.

Capon, J., 1961: Asymptotic Efficiency of Certain Locally Most Powerful Ranks Tests. *Annals of Mathematical Statistics*, **32**, pp. 88-100.

Chao, M., and Glaser, R. E., 1978: The Exact Distribution of Bartlett's Test Statistic for Homogeneity of Variances With Unequal Sample Sizes. *Journal of the American Statistical Association*, **73**, pp. 422-426.

Cochran, W. G., 1941: The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total. *Annals of Eugenics, London*, **11**, pp. 47-52.

Conover, W. J., and Iman, R. L., 1978: Some Exact Tables for the Squared Ranks Test. *Communications in Statistics B: Simulation and Computation*, **7**, pp. 491-513.

Conover, W. J., Johnson, M., Johnson, M. (1981). A comparative study of tests for homogeneity of variances, with application to the outer continental shelf bidding data. *Technometrics,* **23**, pp. 315-361.

Cressie, N. and Read, T. R. C., 1984: Multivariate goodness-of-fit tests. *Journal of the Royal Statistical Society B*, **46**, pp. 440-464.

David, H. A., 1952: Upper 5 and 1% Points of the Maximum F-Ratio. *Biometrika*, **39**, pp. 422-424.

Dixon, W. L., and Massey, F. J., JR, 1969: *Introduction to Statistical Analysis,* New York: McGraw-Hill.

Duran, B. S., and Mielke, P. W., JR, 1968: Robustness of Sum of Squared Ranks Test. *Journal of the American Statistical Association*, **63**, pp. 338-344.

Dyer, D.D., and Keating, J.P., 1980: On the Determination of Critical Values for Bartlett's Test. *Journal of the American Statistical Association*, **75**, pp. 313-319.

Flinger, M. A. and Killen, T. J., 1976: Distribution-free two sample tests for scale. *Journal of American Statistical Association*, **71**, pp. 210-213.

Freund, J. E., and Ansari, A. R., 1957: Two-Way Rank Sum Test for Variances. Technical Report No.34, Virginia Polytechnic Institute, Blacksburg, Virginia.

Gartside, P. S., 1972: A study of methods for comparing several variances. *Journal of American Statistical Association*, **67**, pp. 342-346.

Glaser, R. E., 1976: Exact Critical Values for Bartlett;s Test for Homogeneity of Variances. *Journal of American Statistical Association*, **71**, pp. 488-490.

Hajek, J., and Sidak, Z., 1967: *Theory of Rank Tests*, New York: Academic Press.

Hall, I. J., 1972: Some Comparisons of Tests for Equality of variances. *Journal of Statistical Computations and Simulations*, **1**, pp. 183-194.

Harter, H. L., 1963: Percentage Points of the Ratio of Two Ranges and Power of the Associated Test. *Biometrika*, **50**, pp. 187-194.

Hartley, H. O., 1950: The maximum F ratio as a short-cut test for heterogeneity of variance. *Biometrika*, **37**, pp. 308-312.

Keselman, H.J., P.A., Games and J.J. Clinch, 1979: Tests for homogeneity of variance. *Communications in Statistics: Simulation and Computation*, **8**, pp. 113-129.

Klotz, J., 1962: Nonparametric Tests for Scale. *Annals of Mathematical Statistics*, **33**, pp. 498-512.

Κουκουβίνος, Χ. *"Γραμμικά μοντέλα και Σχεδιασμοί"*, Ε.Μ.Π. 2005.

Layard, M. W. J., 1973: Robust large-sample tests for homogeneity of variance. *Journal of American Statistical Association*, **68**, pp. 195-198.

Lehmann, E. L., 1959: *Testing Statistical Hypothesis*, New York: John Wiley.

Leslie, R. T., and Brown, B. M., 1966: Use of Range in Testing Heterogeneity of Variance. *Biometrika*, **53**, pp. 221-227.

Levene, H., 1960: Robust tests for equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotellings.* In: Olkin, I., Churye, S. G., Hoeffding, W., Madow, W. G., and Mann, H. B. eds. Stanford University Press, Stanford, California, pp. 278-292.

Levy, K. J., 1975: An Empirical Comparison of Several Range Tests for Variances. *Journal of American Statistical Association*, **70**, pp. 180-183.

& 1978: An Empirical Study of the Cube-Root Test for Homogeneity of Variances With Respect to the Effects of Non-normality and Power. *Journal of Statistical. Comput. Simulation*, **7**, pp. 71-78.

Lim, Tjen-Sien, and Loh, Wei-Yin, 1996: A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*, **22**, pp. 287-301.

Lindsay, B. G., 1994: Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Annals of Statistics*, **22**, pp. 1081-1114.

Loh, W.-Y., 1987: Some modifications of Levene's test of variance homogeneity. *Journal of Statist. Comput. Simulation*, **28**, pp. 213-226.

Markatou, M., Basu, A., and Lindsay, B. G., 1996: Weighted likelihood estimating equations: The continuous case. *Technical Report No 323*, Department of Statistics, Stanford University, California, U.S.A.

Markatou, M., Basu, A., and Lindsay, B. G., 1997: Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning an Inference*, **57**, pp. 215-232.

Martin, C. G., 1976: Comment on Levy's An Empirical Comparison of the Z-Variance and Box-Scheffe Tests for Homogeneity of Variance. *Psychometrika*, 41, pp. 551-556.

Martin, C. G., and Games, P. A., 1975: Selection of Subsample Sizes for the Bartlett and Kendall Test of Homogeneity of Variance, paper presented at the meeting of the American Educational Research Association, Washington, D.C., April 1975 (ERIC Document Reproduction Service No. ED 117 150).

& 1977: ANOVA Tests for Homogeneity of Variance: Non-normality and Unequal Samples. *Journal of Educ. Statist.*, **2**, 187-206.

Miller, R. G., 1968: Jackknifing variance. *Annals of Mathematical Statistics*, **39**, pp. 567-582.

Miller, R. G., 1986: *Beyond ANOVA, basics of applied statistics*. Wiley, New York.

Milliken, G. A. and Johnson, D. E., 1984: *Analysis of Messy Data, Volume 1: Designed Experiments*. Van Nostrand Reinhold Company, New York.

Mood, A. M., 1954: On the Asymptotic Efficiency of Certain Non-parametric Two-Saple Tests. *Annals of Mathematical Statistics*, **25**, pp. 514-533.

Moses, L. E., 1963: Rank Tests for Dispersion. *Annals of Mathematical Statistics*, **34**, pp. 973-983.

Neyman, J., and Pearson, E. S., 1931: On the Problem of k Samples. *Bull. Acad. Polon. Sci. et Lettres*, **Ser. A**, pp. 460-481.

O'Brien, R. G., 1978: Robust techniques for testing heterogeneity of variance effects in factorial designs, *Psychometrika*, **43**, pp. 327-342.

Patnaik, P. B., 1950: The Use of Mean Range as an Estimator of Variance in Statistical Tests. *Biometrika*, 37, pp. 78-87.

Pearson, E. S., and Hartley, H. O., 1970: *Biometrika Tables for Statisticians*, (Vol. 1, 3rd ed.), (reprinted with corrections, 1976), Cambridge University Press.

Quenouille, M. H., 1956: Notes on Bias in Estimation. *Biometrika*, **43**, pp. 353-360.

Samuiddin, M., 1976: Bayesian Test of Homogeneity of Variance, *Journal of American Statistical Association*, **71**, pp. 515-517.

Sarkar, C., Kim, C., and Basu, A., 1999: Tests for Homogeneity of Variances Using Robust Weighted Likelihood Estimates. *Biometrical Journal*, **41**, pp. 857-871.

Scheffe, H., 1959: *The Analysis of Variance*, New York: John Wiley.

Schneider, H. and Carroll, R. J., 1985: A note on Levene's tests for equality of variances. *Statistics and Probability Letters*, **3**, pp.191-194.

Sharma, S. C., 1991: A new jackknife test for homogeneity of variances. *Communications in Statistics: Simulation and Computation*, **20**, pp. 479-495.

Sharma, S. C., and Giaccotto, G., 1991: Power and Robustness of Jackknife and Likelihood Ratio Tests for Grouped Heteroscedasticity, Forthcoming. *Journal of Econometrics*.

Shorack, G. R., 1965: Nonparametric Tests and Estimation of Scale in the Two Sample Problem, Technical Report No. 10, Stanford University.

Siegel, S., and Tukey, J. W., 1960: A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples, *Journal of American Statistical Association*, **55**, pp. 429-444.

Simes, J. R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, pp. 751-754.

Simpson, D. G., 1987: Minimum Hellinger distance estimation for analysis of count data. *Journal of American Statistical Association*, **82**, pp. 802-807.

Simpson, D. G., 1989: Hellinger deviance tests: Efficiency, breakdown points and examples. *Journal of American Statistical Association*, **84**, pp. 107-113.

Talwar, P. P., and Gentle, J. E., 1977: A Robust Test for the Homogeneity of Scales. *Communications in Statistics A: Theory and Methods*, **6**, pp. 363-369.

Tamura, R. N. and Boos, D. D., 1986: Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of American Statistical Association*, **81**, pp. 223-229.

Tiku, M. L. and Balakrishnan, N., 1984: Testing equality of population variance the robust way. *Communications in Statistics: Theory and Methods*, **13**, pp. 2143-2159.

Tukey, J. W., 1958: Bias and Confidence in Not-quite Large Samples, *Annals of Mathematical Statistics*, **29**, pp. 614.

Yitnosumarto, S. and M. E. O'Neil, 1986: On Levene's test of variance homogeneity, *Austral. J. Statist.*, **28**, pp. 230-241.