## Έλεγχοι για δύο δείγματα (Τwo-sample problem) συνέχεια

### Έλεγχοι για ισότητα της θέσης (location) και της διασποράς (dispersion) δυο κατανομών

#### Κλασσικός παραμετρικός έλεγχος: F-test για την ισότητα διακύμανσης δύο κανονικών κατανομών.

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* Θέλουμε να ελέγξουμε ***τη μηδενική ***
* Θέτουμε  και 
* Επίσης θέτουμε . Τότε, υπό  η  έχει κατανομή  με  βαθμούς ελευθερίας.
* Απορρίπτουμε την π.χ. έναντι της εναλλακτικής  για μεγάλες τιμές της , δηλαδή για , όπου  το  ποσοστιαίο σημείο της κατανομής  με  βαθμούς ελευθερίας.
* Υπάρχουν και άλλοι έλεγχοι για περισσότερα δείγματα (π.χ. Bartlett).
* Όλοι οι παραπάνω έλεγχοι για μικρά δείγματα είναι πολύ ***ευαίσθητοι σε αποκλίσεις από την κανονικότητα***.

#### Μη παραμετρικός έλεγχος για την ισότητα διασποράς δύο κατανομών: έλεγχος Ansari-Bradley test

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* Θέλουμε να ελέγξουμε ***τη μηδενική ***
* H **εναλλακτική** θέλουμε να αφορά π.χ. μια κατάσταση όπου **η διακύμανση των  είναι μεγαλύτερη από τη διακύμανση των .**
* Ένα μοντέλο που συχνά χρησιμοποιείται για να περιγράψει μια τέτοια κατάσταση το ***«μοντέλο θέσης-κλίμακας» (“location-scale model”)***:

 και  για κατάλληλη .

Ισοδύναμα θα μπορούσαμε να πούμε ότι υποθέτουμε πως

 έχουν την ίδια κατανομή με τα 

* Κατ’ αρχάς **υποθέτουμε πως  και ελέγχουμε  έναντι, π.χ. .**
* **Αν δε γνωρίζουμε ή δε θέλουμε να υποθέσουμε πως  θα μπορούσαμε να αφαιρέσουμε από κάθε δείγμα ξεχωριστά τη διάμεσό του**.
* Ορίζουμε τη **στατιστική του Ansari-Bradley** ως εξής:

Στο «ενιαίο» δείγμα  αντιστοιχούμε το «βαθμό» (score) 1 στη μεγαλύτερη καθώς και στη μικρότερη των παρατηρήσεων, το «βαθμό» 2 στη δεύτερη μεγαλύτερη καθώς και στη δεύτερη μικρότερη των παρατηρήσεων, κλπ.

Έστω  οι βαθμοί των .

* Θέτουμε

,

και απορρίπτουμε  έναντι π.χ. 

για μεγάλες τιμές της 

* Μπορεί να δειχτεί τότε ότι
* Για  ζυγό έχουμε
  + 
  + 
* Ενώ για  μονό έχουμε
  + 
  + 
* Καθώς και 
* Άρα απορρίπτουμε την  έναντι π.χ.  για  με 

#### Μη παραμετρικός έλεγχος για την ισότητα θέσης και την διασποράς δύο κατανομών: έλεγχος Lepage test

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* Στο «μοντέλο θέσης-κλίμακας» (“location-scale model”):

 και για κατάλληλη .

Θέλουμε να ελέγξουμε την  έναντι, π.χ. .

* Ορίζουμε το στατιστικό



Το οποίο αποδεικνύεται ότι έχει για  κατανομή .(οι δύο ποσότητες που αθροίζουμε είναι ασυμπτωτικά ανεξάρτητες).

* Άρα απορρίπτουμε την μηδενική όταν .

### Έλεγχοι για ισότητα δυο κατανομών (χωρίς περιορισμούς)

#### Kolmogorov-Smirnov έλεγχος για την ισότητα δύο κατανομών.

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* Από το Κεφάλαιο 3 του «κορμού» γνωρίζουμε ότι οι

 και 

είναι εκτιμήτριες των και αντιστοίχως και ότι είναι γνωστή η ασυμπτωτική κατανομή(καθώς και η κατανομή για πεπερασμένο δείγμα) π.χ. της  (που ονομάζεται και στατιστική Kolmogorov**-**Smirnov



* Για να ελέγξουμε την  έναντι της  , εκτιμούμε τις  και  ξεχωριστά και παίρνουμε ως στατιστικό το απόλυτο της διαφοράς τους:



Το οποίο υπό  έχει ασυμπτωτικά την ίδια κατανομή με τη στατιστική **Kolmogorov-Smirnov**, την οποία μπορούμε λοιπόν να χρησιμοποιήσουμε για να βρούμε κρίσιμες τιμές για τον έλεγχο.

* Άλλος παρόμοιος έλεγχος μπορεί να στηριχτεί σα στατιστικό τύπου **Anderson-Darling**: 

### Έλεγχοι για ισότητα της θέσης (location) περισσότερων κατανομών

#### Κλασσικός παραμετρικός έλεγχος: F-test για την ανάλυση διακύμανσης κατά ένα παράγοντα.

* Έστω  i.i.d από , για  ανεξάρτητες, ή, ισοδύναμα,  με .



* Θέλουμε να ελέγξουμε τη μηδενική , ή, ισοδύναμα, .



* Θέτουμε:



* + 



* + 



* + 



* Τότε η  έχει υπό  κατανομή  με  βαθμούς ελευθερίας.



* Απορρίπτουμε την  για , όπου  το  ποσοστιαίο σημείο της κατανομής  με  βαθμούς ελευθερίας.

#### Μη παραμετρικός έλεγχος για τη διαφορά θέσης περισσότερων κατανομών, που είναι μετατοπίσεις κοινής F: Kruskall-Wallis.

* Έστω  i.i.d από συνεχή , για , ανεξάρτητες μεταξύ τους.



* Υποθέτουμε , και θέλουμε να ελέγξουμε τη μηδενική .



* Η ιδέα στο μη-παραμετρικό έλεγχο είναι να μιμηθεί τον παραμετρικό χρησιμοποιώντας τάξεις μεγέθους (Ranks) αντί για τις ίδιες τις παρατηρήσεις.
* Έστω  η τάξη της παρατήρησης  μέσα στο ενιαίο δείγμα όλων των παρατηρήσεων.



* Θέτουμε:





* + 





* + 



* + Τέλος η στατιστική των Kruskal-Wallis ορίζεται ως:





η οποία αν όλα τα  θα τείνει , και απορρίπτουμε την  σε επίπεδο  για μεγάλες τιμές της : .