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

* Έχουμε παρατηρήσεις  που προέρχονται από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με αθροιστική συνάρτηση κατανομής  και ανεξάρτητες  που προέρχονται από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με αθροιστική συνάρτηση κατανομής . Και τις δύο cdf τις υποθέτουμε συνεχείς.
* Θέλουμε να κάνουμε **συμπερασματολογία για *κάποια πλευρά της σχέσης των  και , όπως π.χ. αν η «θέση» τους είναι ίδια έναντι της εναλλακτικής να διαφέρουν, ή αν η διακυμάνσεις τους είναι ίδιες***.

### Έλεγχοι για ισότητα της θέσης δυο κατανομών με ίσες διακυμάνσεις

#### Κλασσικός παραμετρικός έλεγχος: t-test για δύο δείγματα με ίσες διακυμάνσεις

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από . (προσοχή **ίδιες διακυμάνσεις** εξ υποθέσεως)
* Θέλουμε να ελέγξουμε τη μηδενική 
* Θέτουμε ,με 
* Τότε αποδεικνύεται ότι  υπό κανονικότητα. Αν δεν ισχύει η κανονικότητα, ακόμα και αν διαφέρουν οι κατανομές των δύο δειγμάτων, ενώ έχουν ίσους μέσους και διακυμάνσεις, τότε  για μεγάλα . Έτσι μπορεί να βρει κανείς κρίσιμες τιμές υπό τη μηδενική.

#### Μη παραμετρικός έλεγχος για τη θέση δύο κατανομών, που η μία είναι μετατόπιση της άλλης: Wilcoxon και Mann-Whitney.

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* Θέλουμε να ελέγξουμε τη μηδενική  έναντι της εναλλακτικής  , δηλαδή ότι η  είναι στοχαστικά μικρότερη της .
* Ένα μοντέλο που περιγράφει μια τέτοια κατάσταση είναι **όταν η  προέρχεται από μετατόπιση (θέσης) της : (location model).** Αυτό ισοδυναμεί με ** έχει την ίδια κατανομή με **. Εδώ ελέγχουμε  έναντι, π.χ. . Αυτό το μοντέλο είναι συγκρίσιμο με εκείνο της προηγούμενης παραγράφου.
* Ορίζουμε τη στατιστική του Wilcoxon ως

,

όπου  η τάξη των  στο «ενιαίο» δείγμα . Όταν  η  θα τείνει να παίρνει μεγαλύτερες τιμές.

* Παρατηρείστε ότι, καθώς , παίρνουμε

.

* Η στατιστική



ονομάζεται στατιστική των Mann-Whitney και είναι προφανώς ισοδύναμο με τη στατιστική του Wilcoxon. Απορρίπτει την  για μεγάλες τιμές της .

* **Πρόταση.**
  + Υπό την  η στατιστική  των Mann-Whitney **είναι ελεύθερη κατανομής**.
  + Η **ακριβής της κατανομή** υπό  **μπορεί να υπολογιστεί** αναδρομικά.
  + Έχουμε υπό :  και .
  + και .
* Άρα μπορεί να βρει κανείς κρίσιμες τιμές  για έναν έλεγχο βασισμένο στην  είτε χρησιμοποιώντας την ακριβή της κατανομή, είτε προσεγγιστικά από 
* Στο μοντέλο του ελέγχου , όπου η  προέρχεται από μετατόπιση (θέσης) της : , **η ARE κατά Pittman** του ελέγχου που βασίζεται στη Mann-Whitney συγκριτικά με τον έλεγχο που βασίζεται στο κλασσικό t-test είναι η ίδια με την ARE τηw Wilcoxon έναντι του t-test στο ένα δείγμα, δηλαδή

*,* όπου .

#### Εκτιμήτρια της διαφοράς θέσης δύο κατανομών βασισμένη στη στατιστική των Mann-Whitney.

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από .
* **Η εκτιμήτρια του ** στο μοντέλο όπου η  προέρχεται από μετατόπιση (θέσης) της : , **βασίζεται στη παρατήρηση ότι η κατανομή των  έχει διάμεσο .**
* Αν, λοιπόν,  το διατεταγμένο δείγμα των  ορίζουμε .
* Διαστήματα εμπιστοσύνης για το **** κατά Moses: Υποθέτουμε ότι  και συνεχείς και . Έστω  ο μεγαλύτερος ακέραιος τέτοιος ώστε . **Τότε το διάστημα  είναι ένα  δ.ε. για το .**
* Προσεγγιστικά μπορούμε να πάρουμε .

### Έλεγχοι για ισότητα της θέσης δυο κατανομών με διαφορετικές διακυμάνσεις

#### Κλασσικός παραμετρικός έλεγχος: t-test για τη θέση δύο κανονικών κατανομών με διαφορετικές διακυμάνσεις (πρόβλημα Behrens-Fisher).

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από . (προσοχή **επιτρέπονται διαφορετικές διακυμάνσεις**)
* Θέλουμε να ελέγξουμε τη μηδενική 
* Θέτουμε 
* Τότε o Welch προτείνει τη προσέγγιση , με «τυχαίους βαθμούς ελευθερίας:

, όπου  και 

Η ιδέα αυτή προκύπτει από εξίσωση της διακύμανσης του παρονομαστή της  με κατάλληλη κατανομή  και επακόλουθη εκτίμηση του  με .

* Ο έλεγχος του Welch έχει καλές ιδιότητες ακόμα και για μικρά μεγέθη δείγματος αν πράγματι οι δύο κατανομές είναι κανονικές. Τι γίνεται όταν αυτό δεν ισχύει?

#### Μη παραμετρικός έλεγχος για τη θέση δύο κατανομών (μη παραμετρικό πρόβλημα Behrens-Fisher).

* Έστω  i.i.d από  και ανεξάρτητες  i.i.d από , που τις υποθέτουμε **και τις δύο** **συνεχείς και συμμετρικές** γύρω από τη διάμεσό τους.
* Έστω  η διάμεσος της  και  η διάμεσος της . ***Θέλουμε να ελέγξουμε τη μηδενική  έναντι της εναλλακτικής  .*** Σημειώστε ότι δεν υποθέτουμε τίποτε άλλο για τη σχέση  και , ούτε π.χ. ότι έχουν ίσες διακυμάνσεις.
* Το στατιστικό που θα χρησιμοποιηθεί μοιάζει με αυτό του Welch, αλλά στηρίζεται στις τάξεις των δειγμάτων: έστω

 και .

* Ως βάση μπορεί να πάρει κανείς τη στατιστική των  των Mann-Whitney που είναι το άθροισμα των τάξεων των  στο δείγμα :



* Τότε υπό  θα έχουμε  συμμετρική και άρα την ίδια αναμενόμενη τιμή όπως υπό , δηλαδή .
* H εκτίμηση της διακύμανσης υπό  δεν είναι τόσο απλή όσο υπό . Κανονικοποιώντας κατάλληλα παίρνουμε τη στατιστική:



* Τότε υπό  θα έχουμε .
* Ο έλεγχος που προκύπτει θα έχει ισχύ που τείνει στο 1 για εναλλακτικές για τις οποίες .