Έλεγχος t για εξαρτημένα δείγματα

Paired samples t-test

Τι ξέρουμε ήδη;

- Να **συγκρίνουμε το μέσο όρο των δύο πληθυσμών**, εάν αυτοί είναι **ανεξάρτητοι**.
- (Independent samples t-test)

Διαφορά με τα προηγούμενα

- Τι γίνεται αν οι καταγραφές δεν είναι ανεξάρτητες;

Παράδειγμα

- Έστω ότι μετρήσαμε τις επιδόσεις πέντε αθλητών πριν και μετά από προπόνηση ενός μηνός.
- Ο λόγος για να δούμε αν το συγκεκριμένο πρόγραμμα προπόνησης επηρέασε τις επιδόσεις των αθλητών.

	X_1 : Πριν	X_2 : Μετά
Αθλ1:	1	2
Αθλ2:	2	4.5
Αθλ3:	2	3
Αθλ4:	2	4.5
Αθλ5:	3	6

Τι βλέπουμε εδώ;

- Οι καταγραφές δεν είναι ανεξάρτητες.
- Κάθε γραμμή του πίνακα αντιστοιχεί **στον ίδιο αθλητή**. Άρα <u>εξαρτημένα</u> δείγματα
- Π.χ. ο *αθλητής 1* είχε επίδοση 1 πριν την προπόνηση και 2 μετά από αυτή.

	X_1 : Πριν	X_2 : Μετά	$(X_1 - \overline{X}_1)^2$	$(X_2 - \overline{X}_2)^2$
Αθλ1:	1	2	1	4
Αθλ2:	2	4.5	0	0.25
Αθλ3:	2	3	0	1
Αθλ4:	2	4.5	0	0.25
Αθλ5:	3	6	1	4
Μέσος όρος	2	4		
	Τυπική απόκλιση	$s_x^2 = \frac{\sum (X - \overline{X})^2}{n} =$	0.4	1.9

One sample t-test

- Μπορούμε να συγκρίνουμε τον μέσο όρο των διαφορών με τιμή ελέγχου το μηδέν (0).
- Αν ο μέσος όρος είναι σημαντικά διαφορετικός σημαίνει ότι η προπόνηση επηρέασε τις επιδόσεις των αθλητών. Αλλιώς δεν είχε αποτέλεσμα.

Τι εξετάζουμε με το paired samples t-test

- **Δε συγκρίνουμε εάν ο μέσος όρος συνολικά αυξήθηκε ή μειώθηκε, αλλά εάν οι επιδόσεις του κάθε αθλητή άλλαξαν**

  π.χ. θα μπορούσε να αυξηθεί ο μέσος όρος γιατί κάποιος αθλητής βελτίωσε πολύ την επίδοσή του ενώ όλοι οι άλλοι χειροτέρευσαν λίγο.

- Και όπως κάθε στατιστικός έλεγχος, δεν επιχειρεί να βγάλει συμπέρασμα για το δείγμα αλλά για τον πληθυσμό.

  Με άλλα λόγια, εάν είχαμε πάρει 5 διαφορετικούς αθλητές πόσο πιθανό θα ήταν να είχαν επίσης βελτίωση στις επιδόσεις τους;

Τι εξετάζουμε με το paired samples t-test

- Δε συγκρίνουμε εάν ο μέσος όρος συνολικά αυξήθηκε ή μειώθηκε, αλλά εάν οι επιδόσεις του κάθε αθλητή άλλαξαν

  π.χ. θα μπορούσε να αυξηθεί ο μέσος όρος γιατί κάποιος αθλητής βελτίωσε πολύ την επίδοσή του ενώ όλοι οι άλλοι χειροτέρευσαν λίγο.

- Και όπως κάθε στατιστικός έλεγχος, δεν επιχειρεί να βγάλει συμπέρασμα για το δείγμα αλλά για τον πληθυσμό. (Με άλλα λόγια, εάν είχαμε πάρει 5 διαφορετικούς αθλητές πόσο πιθανό θα ήταν να είχαν επίσης βελτίωση στις επιδόσεις τους;)

Πως δουλεύει;

	X_1 : Πριν	X_2 : Μετά	$X_1 - X_2$
Αθλ1:	1	2	-1
Αθλ2:	2	4.5	-2.5
Αθλ3:	2	3	-1
Αθλ4:	2	4.5	-2.5
Αθλ5:	3	6	-3
		Μέσος όρος	-2
		Τυπική απόκλιση	0,935

Πως δουλεύει;

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{s_x^2}{n}}} = \frac{-2 - 0}{\sqrt{\frac{0{,}935^2}{5}}} = -4.783$$

Πως δουλεύει;

$t = -4{,}783$, $df=4$, οπότε $p=0.009$

Άρα με **$p=0.009$** μπορούμε να ισχυριστούμε ότι η προπόνηση επηρέασε **σημαντικά** τις επιδόσεις των αθλητών.

Προϋποθέσεις εφαρμογής

- Αντιπροσωπευτικότητα δείγματος
- Κανονικότητα κατανομής **των διαφορών**

- Απαιτείται να υπάρχει εξάρτηση μεταξύ των καταγραφών και οι καταγραφές να αφορούν την ίδια ποσοτική μεταβλητή.

Πότε το χρησιμοποιούμε;

- Η δοκιμασία t για **εξαρτημένα** δείγματα εκτιμά πόσο αξιόπιστα παρατηρούνται οι **διαφορές** ως συνέπεια κάποιου χειρισμού.