Στατιστικός έλεγχος υποθέσεων (Μέρος 3ο)

Εξαρτημένα δείγματα / Ζευγαρωτές παρατηρήσεις (1)

- Έστω η τ.μ. **X** και η τ.μ. **Y** που αντιστοιχούν σε δύο πληθυσμούς μετρήσεων
- Έστω ότι έχουμε δύο δείγματα $\mathbf{X_1, X_2, \ldots, X_n}$ και $\mathbf{Y_1, Y_2, \ldots, Y_n}$
- Θεωρούμε τα ζεύγη $\mathbf{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)}$ τα οποία είναι ανεξάρτητα το ένα από το άλλο ενώ τα $\mathbf{X_i}$ και $\mathbf{Y_i}$ ενός του ίδιου ζεύγους δεν είναι ανεξάρτητα (δεν μπορούν να θεωρηθούν ανεξάρτητα)
- Για κάθε ζεύγος σχηματίζουμε τη διαφορά $\mathbf{D_i = X_i - Y_i}$ και πλέον μπορούμε να εργαστούμε με ένα δείγμα, αυτό των διαφορών $\mathbf{D_1, D_2, \ldots, D_n}$ το οποίο θεωρούμε ότι προέρχεται από ένα θεωρητικό πληθυσμό (τον πληθυσμό των διαφορών) με μέση τιμή $\boldsymbol{\mu_D = \mu_A - \mu_B}$, όπου $\boldsymbol{\mu_A}$ η μέση τιμή της **X** και $\boldsymbol{\mu_\beta}$ η μέση τιμή της **Y**

Εξαρτημένα δείγματα / Ζευγαρωτές παρατηρήσεις (2)

- Έτσι, ο έλεγχος της μηδενικής υπόθεσης $\mathbf{H_0: \mu_D = 0}$ είναι ισοδύναμος με τον έλεγχο μηδενικής υπόθεσης $\mathbf{H_0: \mu_A - \mu_B = 0}$
- Γενικότερα ο έλεγχος της μηδενικής υπόθεσης $\mathbf{H_0: \mu_D = \delta}$ είναι ισοδύναμος με τον έλεγχο μηδενικής υπόθεσης $\mathbf{H_0: \mu_A - \mu_B = \delta}$
- Αν η διακύμανση του πληθυσμού των διαφορών, $\boldsymbol{\sigma_D^2}$, είναι άγνωστη (που είναι το πιο συνηθισμένο) ως στατιστική συνάρτηση ελέγχου μπορούμε να χρησιμοποιήσουμε την $\mathbf{T} = \frac{(\bar{\mathbf{D}}-\delta)\sqrt{n}}{S_D}$, όπου $\bar{\mathbf{D}}$ και $\mathbf{S_D^2}$ οι τ.μ. που αντιστοιχούν στο μέσο και στη διακύμανση του δείγματος των διαφορών αντίστοιχα

Εξαρτημένα δείγματα / Ζευγαρωτές παρατηρήσεις (3)

- Αν επομένως το δείγμα των διαφορών προέρχεται από κανονικό πληθυσμό ισχύουν τα εξής:

H_0: μ_D = δ

Περιοχή απόρριψης της H_0

H_1: $\mu_D \neq \delta$	H_1: $\mu_D > \delta$	H_1: $\mu_D < \delta_0$
$\lvert T \rvert = \dfrac{\lvert \bar{d} - \delta \rvert}{s_d/\sqrt{n}} \geq t_{n-1,a/2}$	$T = \dfrac{\bar{d} - \delta}{s_d/\sqrt{n}} \geq t_{n-1,a}$	$T = \dfrac{\bar{d} - \delta}{s_d/\sqrt{n}} \leq -t_{n-1,a}$

Με $\bar{d}$ και s_d^2 συμβολίζεται, αντίστοιχα, η τιμή της τ.μ. $\bar{D}$ και η τιμή της τ.μ. S_D^2 για τη συγκεκριμένη πραγματοποίηση του δείγματος των διαφορών

Παράδειγμα 4 (1)

- Ένας ερευνητής θέλει να συγκρίνει τις αποδόσεις (ανά στρέμμα) δύο ποικιλιών σταριού στον κάμπο της Θεσσαλίας
- Για το σκοπό αυτό σχεδίασε ένα πείραμα ως εξής:
- Επέλεξε 10 αγρούς σε δέκα διαφορετικές τοποθεσίες του Θεσσαλικού κάμπου και κάθε αγρό τον χώρισε σε δύο αγροτεμάχια ίδιου σχήματος και ίδιου εμβαδού
- Στο ένα αγροτεμάχιο κάθε αγρού καλλιέργησε στάρι της μιας ποικιλίας, έστω Α, και στο άλλο αγροτεμάχιο καλλιέργησε στάρι της άλλης ποικιλίας, έστω Β
- Σε ποιο από τα δύο αγροτεμάχια καλλιέργησε την ποικιλία Α και σε ποιο την ποικιλία Β το αποφάσισε με τυχαίο τρόπο (π.χ. με τη ρίψη ενός νομίσματος)

Παράδειγμα 4 (2)

- Επίσης φρόντισε στα δύο αγροτεμάχια κάθε αγρού να υπάρχουν ίδιες καλλιεργητικές συνθήκες και ίδιες συνθήκες συγκομιδής (γονιμότητα εδάφους, υγρασία, προσανατολισμός, χρόνος σποράς, καλλιεργητική μέθοδος, λίπανση, ημέρα θερισμού, κλπ.)
- Στον παρακάτω Πίνακα φαίνεται η απόδοση των δύο ποικιλιών σε καθέναν από τους 10 πειραματικούς αγρούς

Αγρός (i)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Απόδοση ποικιλίας Α (Kg / στρέμμα) (x_i)	500	650	490	570	555	545	535	635	625	540
Απόδοση ποικιλίας Β (Kg / στρέμμα) (y_i)	455	620	455	610	505	495	515	600	600	510

- Τα ευρήματα στα δύο δείγματα μαρτυρούν άραγε ότι οι μέσες αποδόσεις των δύο ποικιλιών στον κάμπο της Θεσσαλίας διαφέρουν; (α = 5%)