

ONE WAY ANOVA

→ Εάν θέλουμε να ελέγξουμε αν η μέση τιμή μιας ποσοτικής μεταβλητής διαφέρει ανάμεσα σε δύο ανεξάρτητα δείγματα, εφαρμόζουμε το t-test.

→ Για να ελέγξουμε αν διαφέρουν οι μέσες τιμές μιας ποσοτικής μεταβλητής ανάμεσα σε περισσότερες από δύο κατηγορίες, χρησιμοποιούμε την Ανάλυση Διασποράς μιας Κατεύθυνσης (one way ANOVA)

→ Αν χρησιμοποιήσουμε t-test σε όλα τα ζεύγη δειγμάτων αυξάνεται η πιθανότητα λάθους.

→ Για να δούμε ποιες ομάδες διαφέρουν κάνουμε post-hoc test

Προϋποθέσεις

- Η ποσοτική μεταβλητή κατανέμεται κανονικά σε κάθε δείγμα
- Οι διασπορές είναι ίσες σε κάθε δείγμα
- Τα δείγματα είναι ανεξάρτητα

(2)

Άσκηση 1

Παρακάτω δίνονται οι χρόνοι που χρειάστηκαν τρεις εργαζόμενοι για να ολοκληρώσουν μια εργασία σε 5 τυχαία επιλεγμένες περιπτώσεις. Οι χρόνοι έχουν στρογγυλοποιηθεί στο πλησιέστερο λεπτό.

Εργάτης 1	Εργάτης 2	Εργάτης 3
8	8	10
10	9	9
9	9	10
11	8	11
10	10	9

Υπάρχει διαφορά στην απόδοση των τριών εργαζομένων;

$$\bar{y}_1 = \frac{8+10+9+11+10}{5} = 9,6$$

$$\bar{y}_2 = 8,8$$

$$\bar{y}_3 = 9,8$$

$$s_1^2 = 1,3$$

$$s_2^2 = 0,7$$

$$s_3^2 = 0,7$$

$$\bar{\bar{y}} = \frac{9,6 + 8,8 + 9,8}{3} = 9,4$$

Κατασκευή πίνακα one way ANOVA.

$$SS_{Tr} = n_1(\bar{y}_1 - \bar{\bar{y}})^2 + n_2(\bar{y}_2 - \bar{\bar{y}})^2 + n_3(\bar{y}_3 - \bar{\bar{y}})^2 = 5(9,6 - 9,4)^2 + 5(8,8 - 9,4)^2 + 5(9,8 - 9,4)^2 = 2,8$$

③

$$SSE = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2 = 4 \cdot 1,3 + 4 \cdot 0,7 + 4 \cdot 0,7 = 10,8$$

$$SST_0 = SST_r + SSE = 2,8 + 10,8 = 13,6$$

$$MST_r = \frac{SST_r}{k-1} = \frac{2,8}{3-1} = 1,4$$

$$MSE = \frac{SSE}{N-k} = \frac{10,8}{15-3} = 0,9$$

$$F = \frac{MST_r}{MSE} = \frac{1,4}{0,9} = 1,5556$$

Ο πίνακας Άνοια.

Πηγή μεταβλησιμότητας	B.E	S.S	MS	F-stat	P-value
Παράχοντας	2	2,8	1,4	1,5556	p-value
Σφάλμα	12	10,8	0,9		0,10
ολική	14	13,6			

Συμπέρασμα

Δεν απορρίπτεται η μηδενική υπόθεση των ίσων μέσων τιμών, δηλ δεν υπάρχει στατιστικά σημαντική διαφορά στην απόδοση των τριών εργασιών

④

Άσκηση 2

Να συμπληρωθούν τα κενά στον παρακάτω πίνακα one way ANOVA

Πηγή μεταβλητών	B.E	SS	MS	F-statistic
Παράγοντας		2.124	0.708	0.75
Σφάλμα	20			
Ολική				

$$MSTr = \frac{SSTr}{df_{Tr}} \Rightarrow df_{Tr} = \frac{SSTr}{MSTr} = \frac{2.124}{0.708} = 3$$

$$F = \frac{MSTr}{MSE} \Rightarrow MSE = \frac{MSTr}{F} = \frac{0.708}{0.75} = 0.944$$

$$SSTo = SSTr + SSE = 2.124 + 18.880 = 21.004$$

$$df_{To} = df_{Tr} + df_E = 3 + 20 = 23$$

Ασκηση 3

Παρακάτω δίνεται η διάρκεια ζωής μπαταριών που κατασκευάζουν 4 διαφορετικές εταιρίες. Για κάθε μάρκα έχουν επιλεγεί τυχαία 5 μπαταρίες και ο χρόνος ζωής τους έχει στρογγυλοποιηθεί στην πλησιέστερη ώρα.

Brand A	Brand B	Brand C	Brand D
42	28	24	20
30	36	36	32
39	31	28	38
28	32	28	28
29	27	33	25

Σε επίπεδο σημαντικότητας 5% υπάρχει σημαντική διαφορά των μέσων χρόνων ζωής ανά εταιρία;

Πρώτο Βήμα Έλεγχος προϋποθέσεων

Δεύτερο Βήμα Υποθέσεις:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

vs

H_1 : Δεν είναι ίσες όλες οι μέσες τιμές

Τρίτο Βήμα Κριτική τιμή

$$F_{0.05, 3, 16} = 3.24 \quad \text{Δηλ. απορ. του } H_0 \text{ εάν } F \geq 3.24 \quad (p\text{-value} \leq 0.05)$$

6

Τέταρτο Βήμα

Κατασκευή one way ANOVA

$$\bar{x}_1 = 33,6 \quad \bar{x}_2 = 30,8 \quad \bar{x}_3 = 29,8$$

$$\bar{x}_4 = 28,6$$

$$\bar{x} = \frac{33,6 + 30,8 + 29,8 + 28,6}{4} = 30,7$$

$$SSTr = 5(33,6 - 30,7)^2 + 5(30,8 - 30,7)^2 + 5(29,8 - 30,7)^2 + 5(28,6 - 30,7)^2 = 68,2$$

$$SSE = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2 + (n_4 - 1)S_4^2 = 492$$

$$SSTo = SSE + SSTr = 492 + 68,2 = 560,2$$

Πηγή	B.E/df	SS	MS = SS/df	Fstat.	p-value
Θεωρίες	3	68,2	22,733	0,7393	> 10
Σχολία	16	492	30,75		
Σύνολο	19	560,2			

Περί το Βήμα Ακού $0,7393 < 3,24$
 ($p\text{-value} > 0,05$) δεν απορρίπτουμε την μηδενική υπόθεση, δηλ δεν έχουμε αρκετά στοιχεία για να ισχυριστούμε ότι ο μέσος χρόνος ζωής των παιδιών διαφέρει ανά εταρεία.