

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ



● Ανάλυση Διασποράς

ANOVA : Είναι μια στατιστική μεθοδολογία για τον έλεγχο διαφορών ανάμεσα στους μέσους πολλών πληθυσμών.

ΠΙΝΑΚΑΣ ANOVA (Μονοπαραγοντική ANOVA)

	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσο άθροισμα	F
Πηγή Ομάδες	SSTR	$r-1$	MSTR	
Κατάλοιπα	SSE	$n-r$	MSE	
Συνολικά	SST	$n-1$	MST	

● SSTR : μετράει το ερμηνεύσιμο υατάλοιπο

Είναι η απρύλιση μεταξύ των ομάδων στους πληθυσμούς και ονομάζεται μεταβλητότητα μεταξύ πληθυσμών

SSE : μετράει το μη ερμηνεύσιμο υατάλοιπο

Είναι η απρύλιση σε κάθε ομάδα η οποία δεν ερμηνεύεται από τις ενδεχόμενες απρύψεις ανάμεσα στις ομάδες και ονομάζεται εσωτερική μεταβλητότητα.

SST: μετράει το συνολικό υατάλοιο

Είναη η απόκλιση υαύε σημείου από το γενικό μέσο και ονομάζεται ολική μεταβλητότητα

$$SST = SSTR + SSE$$



● n : συνολικές παρατηρήσεις

Γ : ανεξάρτητα τυχαία δείγματα

● MSTR: μέση μεταβλητότητα μεταξύ πληθυσμίων

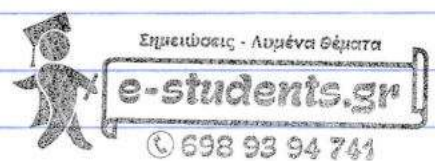
$$MSTR = \frac{SSTR}{\Gamma - 1}$$

MSE: μέση εσωτερική μεταβλητότητα

$$MSE = \frac{SSE}{n - \Gamma}$$

MST: μέση ολική μεταβλητότητα

$$MST = \frac{SST}{n - 1}$$



$$MST = MSTR + MSE$$

● Στατιστικός δείκτης για την ουσιαστική διακύμανση

$$F_{(r-1, n-r)} = \frac{MSTR}{MSE}$$

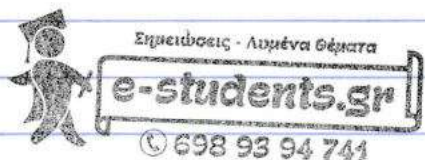


ΔΕΔΟΜΕΝΑ → ANOVA

- ▶ Δεν απορρίπτεται την H_0
- ▶ Απορρίπτεται την H_0

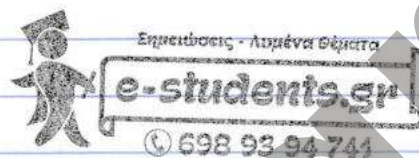
● Όταν το F τείνει στο 1 η μηδενική υπόθεση είναι αληθής

● Όταν το $F > 1$ η μηδενική υπόθεση είναι ψευδής



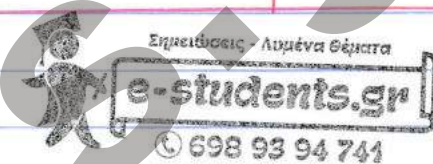
Διαστήματα εμπιστοσύνης για τους μέσους

$$\bar{X}_i \pm t_{\alpha/2} \frac{\sqrt{MSE}}{\sqrt{n_i}}$$



ΠΙΝΑΚΑΣ ANOVA (Παραγοντική ANOVA)

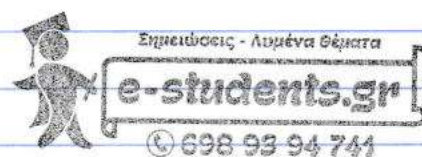
Πηγή	Αιτία παρα- τηρήσεων	Βαθμοί Ελευθερίας	Μεσ. τετράγ.	F
Παράγ. A	SSA	a-1	$MSA = \frac{SSA}{a-1}$	$F = \frac{MSA}{MSE}$
Παράγ. B	SSB	b-1	$MSB = \frac{SSB}{b-1}$	$F = \frac{MSB}{MSE}$
Αλληλ. AB	SS(AB)	(a-1)(b-1)	$MS(AB) = \frac{SS(AB)}{(a-1)(b-1)}$	$F = \frac{MS(AB)}{MSE}$
Σφάλμα	SSE	ab(n-1)	$MSE = \frac{SSE}{ab(n-1)}$	
Σύνολο	SST	abn-1		



Έλεγχος επίδρασης A: $F(a-1, ab(n-1))$

Έλεγχος επίδρασης B: $F(b-1, ab(n-1))$

Έλεγχος (AB) αλληλεπίδρασης: $F((a-1)(b-1), ab(n-1))$



Αιτή χραμμική πάλινδρόμηση

Με την αιτή χραμμική πάλινδρόμηση απεικονίζουμε την σχέση μεταξύ δύο μεταβλητών

Y: εξαρτημένη μεταβλητή

X: ανεξάρτητη μεταβλητή



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

Μοντέλο αιτής χραμμικής πάλινδρόμησης

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

→ $\beta_0 + \beta_1 X$: Μη τυχαίος ή συστηματικός όρος

β_0 : σταθερός όρος

β_1 : κλίση

→ ε : τυχαίος όρος

σφάλμα ή κατάλοιπο



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

● Εξίσωση της γραμμικής παλινδρόμησης

$$Y = b_0 + b_1 X$$

b_0, b_1 : συντελεστές

όπου $b_1 = \frac{SS_{xy}}{SS_x}$



υ' $b_0 = \bar{y} - b_1 \bar{x}$

$\bar{y} = \frac{\sum y_i}{n}$, $\bar{x} = \frac{\sum x_i}{n}$

και $SS_x = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$$SS_y = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$



● $S = \sqrt{MSE}$, S^2 : διακύμανση.

$$MSE = \frac{SSE}{n-2}$$

όπου $SSE = SS_y - b_1 SS_{xy}$

Εφαρμοχή



Εμπειρίες - Λυμένα θέματα

e-students.gr

© 698 93 94 741

Σε 5 δοχεία που περιέχεται η ίδια ποσότητα νερού στο μαθήμα, ρίχνουμε τριχό και μετράμε την πτώση της θερμοκρασίας

Δοχείο	Τριχός (X)	Θερμοκρασία (Y)
1	7,2	3,7
2	9,6	4,5
3	12,2	6,4
4	15,7	7,6
5	18,4	9,2

Να βρεθούν οι ελάχιστες τετραγώνων των σταθερών b_0, b_1 της ευθείας παλινδρόμησης $\hat{y} = b_0 + b_1 x$

Λύση



Εμπειρίες - Λυμένα θέματα

e-students.gr

© 698 93 94 741

$$\bar{x} = \frac{\sum x_i}{n} = \frac{7,2 + 9,6 + 12,2 + 15,7 + 18,4}{5} = 12,62$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{3,7 + 4,5 + 6,4 + 7,6 + 9,2}{5} = 6,28$$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$SS_{xy} = (7,2 - 12,62)(3,7 - 6,28) + (9,6 - 12,62)(4,5 - 6,28) + (12,2 - 12,62)(6,4 - 6,28) + (15,7 - 12,62)(7,6 - 6,28) + (18,4 - 12,62)(9,2 - 6,28) = 41,452$$

$$SS_{x..} = \sum (x_i - \bar{x})^2 = 81,568$$



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

$$b_1 = \frac{SS_{xy}}{SS_{x..}} = \frac{41,452}{81,568} = 0,508$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$\Rightarrow b_0 = 6,28 - 0,508 \cdot 12,62 = -0,131$$

$$y = -0,131 + 0,508x$$



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

● Σφάλματα στους συντελεστές της γραμμικής παλινδρόμησης



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

→ Σφάλμα για το b_0

$$S(b_0) = \frac{S \sqrt{\sum x^2}}{\sqrt{n} S_{xx}}, \quad S = \sqrt{MSE}$$

→ Σφάλμα για το b_1

$$S(b_1) = \frac{S}{\sqrt{SS_{xx}}}$$



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

● Διαστήματα εμπιστοσύνης για τις παραμέτρους στη γραμμική παλινδρόμηση

→ Διάστημα εμπιστοσύνης για το b_0

$$b_0 \pm t_{(\alpha/2, n-2)} S(b_0)$$

→ Διάστημα εμπιστοσύνης για το b_1

$$b_1 \pm t_{(\alpha/2, n-2)} S(b_1)$$

- Λιάστημα Εμπιστοσύνης για
την μέση τιμή του Y για
δεδομένη τιμή του x

$$\hat{y}_0 \pm t_{\alpha/2, n-2} S(\hat{y}_0)$$



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

όπου $S(\hat{y}_0) = S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$

- Λιάστημα Πρόβλεψης για το Y
για δεδομένη τιμή του x

$$\hat{y}_0 \pm t_{\alpha/2, n-2} S(y_0 - \hat{y}_0)$$

όπου $S(y_0 - \hat{y}_0) = S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$



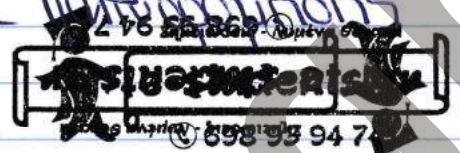
Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

Συντελεστής συσχέτισης r

Είναι ένα ποσοτικό μέτρο της ισχύος της σχέσης παλινδρόμησης, ένα μέτρο για το πόσο καλά προσαρμόζεται η ευθεία παλινδρόμησης στα δεδομένα.



$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{ή} \quad r^2 = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} \quad \text{Συν. συσχ. στο δείγμα}$$

r^2 : συντελεστής προσδιορισμού

όπου $SST = \sum (y - \bar{y})^2$ Συνολική διακύμανση

$SSE = \sum (y - \hat{y})^2$ Ανεξήγητη διακύμανση

$SSR = \sum (\hat{y} - \bar{y})^2$ Ερμηνεύσιμη διακύμανση

Έλεγχος υπόθεσης για τον συντελεστή συσχέτισης

$H_0: \rho = 0$ (Μη γραμμική συσχέτιση)

$H_1: \rho \neq 0$ (Γραμμική συσχέτιση)



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$\text{όπου } r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}}$$

Αν $t_{\frac{\alpha}{2}, n-2} < t$ απορρίπτεται η H_0

● Ελέγχος υπόθεσης για την σχέση τριγωνομέτρησης

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

$$t = \frac{b_1}{S(b_1)}$$

b_1 : υπόθεση της ευθείας
 $S(b_1)$: σφάλμα του b_1

Αν $t_{\frac{\alpha}{2}, n-2} < |t|$ απορρίπτεται η H_0



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

● Εφαρμοχή

Η ελαστικότητα Y ενός υλινού συνδέεται γραμμικά με το μήκος X της παραμόρφωσης. Έπτα μετρήσεις έδωσαν:

X	5	8	12	14	19	25	29
Y	69	67	65	56	42	41	38

α) Να βρούν τα a, b του μοντέλου $Y = a + bx$

β) Να ελεγχθεί η υπόθεση $H_0: b = 0$ έναντι της εναλλακτικής $H_1: b \neq 0$ με $\sigma.σ = 0.05$

Λύση

$$a) \bar{x} = \frac{\sum x_i}{n} = 16, \quad \bar{y} = \frac{\sum y_i}{n} = 54$$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = -570$$

$$SS_x = \sum (x_i - \bar{x})^2 = 464$$

$$\hat{b} = \frac{SS_{xy}}{SS_x} = -1,23$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \Rightarrow \hat{a} = 73,68$$

$$\begin{aligned} \text{B)} \quad H_0: b &= 0 \\ H_1: b &\neq 0 \end{aligned}$$

$$t = \frac{|b|}{S(b)} = \frac{|-1,23|}{0,4085} = 3,011$$

$$S(b) = \frac{S}{\sqrt{SS_x}}$$

$$S = \sqrt{\frac{SSE}{n-2}}$$

$$SSE = SS_y - b \cdot SS_{xy}$$

$$t_{\alpha/2, n-2} = t_{0,95, 5} = 2,015$$

$$t_{\alpha/2, n-2} < t$$

απορριπτεται
η H_0



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741



Σημειώσεις - Λυμένα θέματα

e-students.gr

© 698 93 94 741

● Μοντέλο πολλαπλής παλινδρόμησης

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

β_0 : ο Y -σταθερός όρος της επιφάνειας παλινδρόμησης

και $\beta_1, \beta_2, \dots, \beta_k$: είναι η υψωση της επιφάνειας παλινδρόμησης ως προς την αντίστοιχη μεταβλητή x_i



Σημειώσεις - Ασμένα Θέματα

e-students.gr

© 698 93 94 741

● Εξίσωση $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$$\sum Y = n\beta_0 + \beta_1 \sum x_1 + \beta_2 \sum x_2$$

$$\sum x_1 Y = \beta_0 \sum x_1 + \beta_1 \sum x_1 x_1 + \beta_2 \sum x_1 x_2$$

$$\sum x_2 Y = \beta_0 \sum x_2 + \beta_1 \sum x_1 x_2 + \beta_2 \sum x_2 x_2$$



Σημειώσεις - Ασμένα Θέματα

e-students.gr

© 698 93 94 741

Έλεγχος F στο μοντέλο

Πολυπληθούς Παλινδρόμησης



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

	Αθροίσματα Τετράγωνων	Βαθμοί Ελευθερίας	Μέσα αθροίσματα	F
Πηγή Regression	SSR	k	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Error	SSE	$n - (k + 1)$	$MSE = \frac{SSE}{n - (k + 1)}$	
total	SST	$n - 1$	$MST = \frac{SST}{n - 1}$	

$$SST = SSR + SSE$$

Συνολικη διασπο.	Ληψ. παιχν.	Ληψήμανση Σφάλματος
---------------------	----------------	------------------------



Σημειώσεις - Λυμένα θέματα

e-students.gr

☎ 698 93 94 741

● Πολλαπλός συντελεστής προσδιορισμού R^2

Μετράει το ποσοτό της διακύμανσης της εξαρτημένης μεταβλητής που ερμηνεύεται από το συνδυασμό των ανεξάρτητων μεταβλητών στο μοντέλο της πολλαπλής παλινδρόμησης

$$R^2 = \frac{SSR}{SSE} = 1 - \frac{SSE}{SST}$$



● Προσαρμοσμένος συντελεστής πολλαπλού προσδιορισμού $\bar{R}^2$

$$\bar{R}^2 = 1 - \frac{\frac{SSE}{(n - (k + 1))}}{\frac{SST}{n - 1}}$$

