

[Έλεγχος υποθέσεων II]

[ANOVA]

Ε. Κατσανίδης

Μοντέλα στην Επιστήμη Τροφίμων 532Ε

Τομέας Επιστήμης & Τεχνολογίας Τροφίμων

[Έλεγχος υποθέσεων]

- Συνεχή δεδομένα
 - *z-test*
 - *Student's test (t-test)*
 - Ανάλυση παραλλακτικότητας ή ανάλυση διασποράς (ANOVA)



[Γενικά για τον έλεγχο υποθέσεων]

- Για μεγάλο αριθμό παρατηρήσεων ($n > 30$) χρησιμοποιούμε το *z-test* για να ελέγξουμε αν:
 - ο μέσος όρος ενός δείγματος είναι ίσος με μια ορισμένη τιμή
 - οι μέσοι όροι δυο δειγμάτων είναι ίσοι
- Για μικρό αριθμό παρατηρήσεων ($n < 30$) χρησιμοποιούμε το *t-test* για να ελέγξουμε αν:
 - ο μέσος όρος ενός δείγματος είναι ίσος με μια ορισμένη τιμή
 - οι μέσοι όροι δυο δειγμάτων είναι ίσοι
 - έχοντας την ίδια παραλλακτικότητα } *F-test*
 - έχοντας διαφορετική παραλλακτικότητα }
- Χρησιμοποιούμε ανάλυση παραλλακτικότητας (ANOVA) για να ελέγξουμε αν:
 - οι μέσοι όροι πολλών (> 2) δειγμάτων είναι ίσοι ή όχι



Ε. Κατσάνιδης

[Έλεγχος υπόθεσης για το λόγο των διασπορών δύο πληθυσμών]

Μηδενική υπόθεση $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ ή $\sigma_1^2 = \sigma_2^2$

F-test

Εναλλακτική υπόθεση:

Μονόπλευρο test

Δίπλευρο test

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1 \text{ ή } H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

Τιμή του στατιστικού: $f = \frac{s_1^2}{s_2^2}$

Περιοχές Απόρριψης της H_0

Μονόπλευρο test

Δίπλευρο test

$$f > f_{\eta_1-1, \eta_2-1, \alpha} \text{ (} H_1: \sigma_1^2 > \sigma_2^2 \text{)}$$

$$f > f_{\eta_1-1, \eta_2-1, \alpha/2} \text{ ή}$$

$$f < f_{\eta_1-1, \eta_2-1, 1-\alpha} \text{ (} H_1: \sigma_1^2 < \sigma_2^2 \text{)}$$

$$f < f_{\eta_1-1, \eta_2-1, 1-\alpha/2}$$



Ε. Κατσάνιδης

[Για περισσότερα από 2 δείγματα]

- Όταν έχουμε να συγκρίνουμε μ.ο. για $k > 2$ δείγματα εφαρμόζουμε την Ανάλυση παραλλακτικότητας ή ανάλυση διασποράς ή ανάλυση διακύμανσης ή ANOVA (analysis of variance)



Ε. Κατσάνδης

[Ανάλυση παραλλακτικότητας ή διακύμανσης (ANOVA)]

- Αντί να συγκρίνουμε τους μέσους όρους των δειγμάτων ($\mu_1, \mu_2, \dots, \mu_k$), συγκρίνουμε τις διακυμάνσεις των δειγμάτων ($s^2_1, s^2_2, \dots, s^2_k$).
- Θεωρούμε ότι εάν τα δείγματα προέρχονται από τον ίδιο πληθυσμό θα πρέπει να έχουν και ίδια διακύμανση
- Δηλ: εάν $s^2_1 = s^2_2 = \dots = s^2_k = s^2$, τότε και $\mu_1 = \mu_2 = \dots = \mu_k = \mu$



Ε. Κατσάνδης

[Η μηδενική υπόθεση (H_0) και η εναλλακτική υπόθεση (H_a)]

- Στην ANOVA η H_0 είναι :

$$\mu_1 = \mu_2 = \dots = \mu_k$$

και η H_a είναι πάντα:

$$\mu_i \neq \mu_j$$

- Αρκεί μια ανισότητα για να απορρίψει την H_0 (π.χ. $\mu_2 \neq \mu_4$)
- Δεν έχουμε πληροφόρηση για το ποιες ισότητες ισχύουν και ποιες όχι.
- Απαιτούνται έλεγχοι εκ των υστέρων (post hoc) για να αναγνωριστούν συγκεκριμένες διαφορές.



Ε. Κατσάνης

[Ανάλυση διακύμανσης ενός παράγοντα (one-way ANOVA)]

Έστω ότι έχουμε k δείγματα με n παρατηρήσεις

1	2	...	k
x_{11}	x_{21}	...	x_{k1}
x_{12}	x_{22}	...	x_{k2}
...	...	x_{ij}	...
x_{1n}	x_{2n}	...	x_{kn}



Ε. Κατσάνης

[Μονοπαραγοντική ανάλυση]

- Το μοντέλο της ανάλυσης είναι:
 $x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ όπου $i = 1, 2, \dots, n_k$ & $j = 1, 2, \dots, k$
όπου μ : ο γενικός μέσος όρος
 α_i : η επίδραση του παράγοντα στο δείγμα i
($\alpha_i = \mu - \mu_i$) και
 ε_{ij} : τυχαία σφάλματα με κανονική κατανομή $(0, \sigma^2)$
- $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ και $H_a: \alpha_i \neq 0$



Ε. Κατσάνιθης

[Μονοπαραγοντική ανάλυση]

- Εξετάζουμε εάν η διακύμανση μεταξύ των δειγμάτων $(1, 2, \dots, k)$ είναι μεγαλύτερη από τη γενική διακύμανση μέσα σε όλα τα δείγματα (σφάλμα μετρήσεων).
- Γι' αυτό και η ANOVA είναι πάντα μονόπλευρη δοκιμή



Ε. Κατσάνιθης

[Πίνακας one-way ANOVA]

- Για k δείγματα με n παρατηρήσεις σε κάθε δείγμα
- Ως πηγή του σφάλματος θεωρείται η διακύμανση μέσα σε όλα τα δείγματα (MSW)

Πηγή	SS	df	MS	F	P-value	F crit
Μεταξύ δειγμάτων	SSB	k-1	MSB	MSB/MSW		
Μέσα στα δείγματα	SSW	nk-k	MSW			
Σύνολο	SST	nk-1				



Ε. Κατσάνιθης

[Ανάλυση διακύμανσης δύο παραγόντων (two-way ANOVA)]

- Πολλές φορές εξετάζουμε την ταυτόχρονη επίδραση δύο παραγόντων.
- Μπορούμε να έχουμε 1 ή περισσότερες παρατηρήσεις (επαναλήψεις) για κάθε συνδυασμό παραγόντων.
- Έτσι μπορούμε να έχουμε:
 - Ανάλυση παραλλακτικότητας 2 παραγόντων με επανάληψη (2-way ANOVA with replication)
 - Ανάλυση παραλλακτικότητας 2 παραγόντων χωρίς επανάληψη (2-way ANOVA without replication)



Ε. Κατσάνιθης

[Διπαραγοντική ανάλυση]

- Έστω ότι έχουμε 2 παράγοντες (A & B) με κ και λ στάθμες (επίπεδα) αντίστοιχα

	B ₁	B ₂	...	B _λ
A ₁	X ₁₁₁ X ₁₁₂	X ₁₂₁ X ₁₂₂	...	X _{1λ1} X _{1λ2}
A ₂	X ₂₁₁ X ₂₁₂	X ₂₂₁ X ₂₂₂	...	X _{2λ1} X _{2λ2}
...
A _κ	X _{κ11} X _{κ12}	X _{κ21} X _{κ22}	...	X _{κλ1} X _{κλ2}



Ε. Κατσάνιδης

[Διπαραγοντική ανάλυση]

- Το μοντέλο της ανάλυσης είναι:

$$X_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijl} \quad \text{όπου}$$

$$i=1,2,\dots,\kappa \quad j=1,2,\dots,\lambda \quad l=1,2,\dots,n$$

μ : ο γενικός μέσος όρος

α_i : η επίδραση του παράγοντα A στην i στάθμη (γραμμή)

β_j : η επίδραση του παράγοντα B στην j στάθμη (στήλη)

γ_{ij} : η αλληλεπίδραση των παρόντων A & B

ε_{ijl} : τυχαία σφάλματα με κανονική κατανομή (0,σ²)



Ε. Κατσάνιδης

[Οι υποθέσεις που ελέγχονται]

- $H_{0AB}: Y_{11}=Y_{12}=\dots=Y_{\kappa\lambda}=0$ και $H_{\alpha AB}: Y_{ij} \neq 0$
δεν υπάρχουν σημαντικές αλληλεπιδράσεις
- $H_{0A}: \alpha_1=\alpha_2=\dots=\alpha_{\kappa}=0$ και $H_{\alpha A}: \alpha_i \neq 0$
ο παράγοντας A δεν επιδρά σημαντικά
- $H_{0B}: \beta_1=\beta_2=\dots=\beta_{\lambda}=0$ και $H_{\alpha B}: \beta_j \neq 0$
ο παράγοντας B δεν επιδρά σημαντικά



Ε. Κατσάνιδης

[Πίνακας ANOVA (μοντέλο I)]

- Υποθέτοντας ότι τα επίπεδα των παραγόντων A & B επιλέχθηκαν από τον ερευνητή και όχι τυχαία.

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων (SS)	β.ε.	Μέση μεταβλητότητα	Λόγος-F
Παράγοντας A	SSA	$\kappa-1$	$MSA = \frac{SSA}{\kappa-1}$	$F_A = \frac{MSA}{MSE}$
Παράγοντας B	SSB	$\lambda-1$	$MSB = \frac{SSB}{\lambda-1}$	$F_B = \frac{MSB}{MSE}$
Αλληλεπίδραση AxB	SSAB	$(\kappa-1)(\lambda-1)$	$MSAB = \frac{SSAB}{(\kappa-1)(\lambda-1)}$	$F_{AB} = \frac{MSAB}{MSE}$
Σφάλμα	SSE	$\kappa\lambda(n-1)$	$MSE = \frac{SSE}{\kappa\lambda(n-1)}$	



Ε. Κατσάνιδης

[Πίνακας ANOVA (Excel)]

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample	0,936	1	0,936	1,268	0,271	4,260
Columns	1,862	2	0,931	1,261	0,301	3,403
Interaction	2,329	2	1,164	1,577	0,227	3,403
Within	17,716	24	0,738			
Total	22,843	29				



Ε. Κατσάνιδης

[Διπαραγοντική ανάλυση]

- Όταν δεν υπάρχει επανάληψη ($n=1$) δεν μπορούμε να βγάλουμε συμπεράσματα για την αλληλεπίδραση των δυο παραγόντων.
- Η τιμή της αλληλεπίδρασης χρησιμοποιείται ως μέτρο του σφάλματος.



Ε. Κατσάνιδης