Γραμμικά Μοντέλα
Ανάλυση Διασποράς κατά δύο Παράγοντες

Διδάσκουσα: Λουκία Μελιγκοτσίδου
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
Τμήμα Μαθηματικών

May 11, 2020

Ανάλυση Διασποράς με Δυο Παράγοντες (two–factor ANOVA)

Έστω ότι το αποτέλεσμα ενός πειράματος εξαρτάται από 2 παράγοντες, A και B. Αρχικά υποθέτουμε ότι δεν υπάρχει αλληλεπίδραση (interaction) μεταξύ των παραγόντων. Θα μελετήσουμε τον παράγοντα A σε m επίπεδα και τον παράγοντα B σε l επίπεδα. Επομένως υπάρχουν $k = m \cdot l$ συνδυασμοί επιπέδων (treatments) για τους οποίους λαμβάνουμε παρατηρήσεις, έστω μια παρατήρηση σε κάθε treatment. Επομένως, στη συγκεκριμένη περίπτωση, και ο συνολικός αριθμός παρατηρήσεων είναι $n = m \cdot l$.

Δηλαδή

$$
\begin{array}{c|cccc|c|c|}
\multicolumn{1}{c}{} & \multicolumn{4}{c}{B} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\
\cline{2-7}
 & Y_{11} & Y_{12} & \cdots & Y_{1l} & Y_{1\cdot} & \overline{Y}_{1\cdot} \\
 & Y_{21} & Y_{22} & \cdots & Y_{2l} & Y_{2\cdot} & \overline{Y}_{2\cdot} \\
A & & & \vdots & & \vdots & \vdots \\
 & Y_{m1} & Y_{m2} & \cdots & Y_{ml} & Y_{m\cdot} & \overline{Y}_{m\cdot} \\
\cline{2-7}
 & Y_{\cdot 1} & Y_{\cdot 2} & \cdots & Y_{\cdot l} & \multicolumn{2}{c|}{} \\
\cline{2-5}
 & \overline{Y}_{\cdot 1} & \overline{Y}_{\cdot 2} & \cdots & \overline{Y}_{\cdot l} & \multicolumn{2}{c|}{} \\
\cline{2-7}
\end{array}
$$

$$
\begin{aligned}
Y_{i\cdot} &= \sum_{j=1}^{l} Y_{ij} & \qquad Y_{\cdot j} &= \sum_{i=1}^{m} Y_{ij} \\
\overline{Y}_{i\cdot} &= \frac{1}{l} Y_{i\cdot} & \qquad \overline{Y}_{\cdot j} &= \frac{1}{m} Y_{\cdot j}
\end{aligned}
\qquad
\overline{Y}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{l} Y_{ij} = \frac{1}{m} \sum_{i=1}^{m} \overline{Y}_{i\cdot} = \frac{1}{l} \sum_{j=1}^{l} \overline{Y}_{\cdot j}\,.
$$

Μοντέλο ANOVA

$$
\begin{aligned}
Y_{ij} &= \mu_{ij} + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, l \\
\epsilon_{ij} &: \text{τυχαία σφάλματα, } \epsilon_{ij} \overset{\text{ανεξ.}}{\sim} N(0, \sigma^2) \\
\mu_{ij} &= \mu_{\cdot\cdot} + \alpha_i + \beta_j \\
\alpha_i &: \text{επίδραση του } i \text{ επιπέδου του παράγοντα A} \\
\beta_j &: \text{επίδραση του } j \text{ επιπέδου του παράγοντα B} \\
& \quad \text{έτσι ώστε } \sum_{i=1}^{m} \alpha_i = 0 \text{ και } \sum_{j=1}^{l} \beta_j = 0
\end{aligned}
$$

Δηλαδή θεωρούμε ότι το $\mu_{\cdot\cdot}$ περιέχει όλες τις σταθερές (κοινές για όλα τα treatments) επιδράσεις στη μέση τιμή της απαντητικής μεταβλητής, ενώ τα α_i και

β_j είναι οι ειδικές επιδράσεις των επιπέδων των παραγόντων A και B, αντίστοιχα.

Άρα, έχουμε $Y_{ij} \sim N(\mu_{ij}, \sigma^2)$.

Υποθέσεις :

- Κανονικότητα
- Ομοσκεδαστικότητα
- Τυχαιότητα/Ανεξαρτησία

Έχουμε υποθέσει ότι $\sum_{i=1}^{m} \alpha_i = 0$ και $\sum_{j=1}^{l} \beta_j = 0$, δηλαδή ότι οι μέσες επιδράσεις των παραγόντων A και B είναι 0 (ή αλλιώς ότι έχουν συμπεριληφθεί στο $\mu_{..}$). Αυτό φαίνεται αν θεωρήσουμε $\mu_{ij} = \mu'_{..} + \alpha'_i + \beta'_j$ και $\overline{\alpha}' = \frac{1}{m}\sum_{i=1}^{m} \alpha'_i$ και $\overline{\beta}' = \frac{1}{l}\sum_{j=1}^{l} \beta'_j$. Τότε

$$\mu_{ij} = \underbrace{\mu'_{..} + \overline{\alpha}' + \overline{\beta}'}_{\mu_{..}} + \underbrace{\alpha'_i - \overline{\alpha}'}_{\alpha_i} + \underbrace{\beta'_i - \overline{\beta}'}_{\beta_j}$$

Προφανώς $\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m}(\alpha'_i - \overline{\alpha}') = 0$ και $\sum_{i=j}^{l} \beta_j = \sum_{j=1}^{l}(\beta'_j - \overline{\beta}') = 0$.

Έλεγχοι Υποθέσεων

Για να ελέγξουμε αν υπάρχει επίδραση του παράγοντα A κάνουμε τον έλεγχο

$$\begin{aligned} H_0 &: \mu_{11} = \mu_{21} = \cdots = \mu_{m1} = \mu \\ H_1 &: 2 \text{ τουλάχιστον διαφέρουν} \end{aligned}$$

Ισοδύναμα

$$\begin{aligned} H_0 &: \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0 \\ H_1 &: \alpha_i \neq \alpha_j \text{ για τουλάχιστον ένα ζεύγος } i, j \end{aligned}$$

Αντίστοιχα για να ελέγξουμε αν υπάρχει επίδραση του παράγοντα B

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \cdots = \beta_m = 0 \\ H_1 &: \beta_i \neq \beta_j \text{ για τουλάχιστον ένα ζεύγος } i, j \end{aligned}$$

Οι παραπάνω έλεγχοι μπορούν να πραγματοποιηθούν με ανάλυση διασποράς και έλεγχο F σε αναλογία με την ANOVA κατά ένα παράγοντα.

Συνολικό άθροισμα τετραγωνικών αποκλίσεων:

$$\begin{aligned}
SST &= \sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{..})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{l}\big((\overline{Y}_{i.}-\overline{Y}_{..})+(\overline{Y}_{.j}-\overline{Y}_{..})+(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})\big)^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{i.}-\overline{Y}_{..})^2+\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})^2+\sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})^2\,,
\end{aligned}$$

μπορεί να αποδειχθεί ότι όλα τα αθροίσματα διπλασίων γινομένων που σχηματίζονται από τα $(\overline{Y}_{i.}-\overline{Y}_{..})$, $(\overline{Y}_{.j}-\overline{Y}_{..})$ και $(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})$ είναι 0. Επομένως

$$SST = \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{i.}-\overline{Y}_{..})^2}_{SSA} + \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})^2}_{SSB} + \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})^2}_{SSE}$$

SSA **:** άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα A

SSB **:** άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα B

SSE **:** άθροισμα τετραγώνων των τυχαίων σφαλμάτων.

Απόδειξη για $\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})$

$$\begin{aligned}
\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..}) &= \sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})\sum_{i=1}^{m}(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..}) \\
&= \sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})\big(\sum_{i=1}^{m}(Y_{ij}-\overline{Y}_{.j})-(\sum_{i=1}^{m}\overline{Y}_{i.}-m\overline{Y}_{..})\big) \\
&= \sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})\big((Y_{.j}-m\overline{Y}_{.j})-(\sum_{i=1}^{m}\frac{Y_{i.}}{l}-m\frac{Y_{..}}{ml})\big) \\
&= 0
\end{aligned}$$

Σε πλήρη αντιστοιχία με την ANOVA κατά έναν παράγοντα οι συνολικοί βαθμοί ελευθερίας είναι $ml-1$, οι βαθμοί ελευθερίας του παράγοντα A είναι $m-1$ και του παράγοντα B είναι $l-1$. Τέλος οι βαθμοί ελευθερίας των σφαλμάτων είναι $ml-m-l+1=m(l-1)-(l-1)=(m-1)(l-1)$.

Ισχύει πάντα ότι $\frac{SSE}{\sigma^2} \sim X^2_{(m-1)(l-1)}$.

Κάτω από την H_0 **:** $\alpha_1=\alpha_2=\cdots=\alpha_m=0$ ισχύει ότι $\frac{SSA}{\sigma^2} \sim X^2(m-1)$. Άρα, για τον έλεγχο αυτής της H_0 μπορούμε να χρησιμοποιήσουμε την ελεγχοσυνάρτηση $F_A=\frac{SSA/m-1}{SSE/(m-1)(l-1)}$, η οποία κάτω από H_0 ακολουθεί $F(m-1,(m-1)(l-1))$. Απορρίπτουμε την H_0 σε επίπεδο στατιστικής σημαντικότητας α αν $F_A > F_\alpha(m-1,(m-1)(l-1))$.

Κάτω από την H_0 **:** $\beta_1=\beta_2=\cdots=\beta_m=0$ ισχύει ότι $\frac{SSB}{\sigma^2} \sim X^2(l-1)$. Άρα για τον έλεγχο αυτής της H_0 μπορούμε να χρησιμοποιήσουμε την ελεγχοσυνάρτηση $F_B=\frac{SSB/l-1}{SSE/(m-1)(l-1)}$, η οποία κάτω από H_0 ακολουθεί $F(l-1,(m-1)(l-1))$. Απορρίπτουμε την H_0 σε επίπεδο στατιστικής σημαντικότητας α αν $F_B > F_\alpha(l-1,(m-1)(l-1))$.

Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	df	MS	$F=\frac{MSF}{MSE}$	p–value
Παράγοντας A	SSA	$m-1$	$MSA=\frac{SSA}{m-1}$	$F_A=\frac{MSA}{MSE}$	⋆
Παράγοντας B	SSB	$l-1$	$MSB=\frac{SSB}{l-1}$	$F_B=\frac{MSB}{MSE}$	
Σφάλμα(Error)	SSE	$(m-1)(l-1)$	$MSE=\frac{SSE}{(m-1)(l-1)}$		
Σύνολο(Total)	SST	$ml-1$			

Παράδειγμα: Μας ενδιαφέρει να μελετήσουμε την κατανάλωση βενζίνης από αυτοκίνητα μιας συγκεκριμένης κατηγορίας. Έχουμε 3 μάρκες αυτοκινήτων και 4 διαφορετικούς τύπους βενζίνης. Ο αριθμός των χιλιομέτρων (km) ανά γαλόνι βενζίνης για καθέναν από τους συνδυασμούς είναι

	τύπος βενζίνης				
μάρκα αυτοκινήτου	1	2	3	4	$\overline{Y}_{i.}$
1	16	18	21	21	19
2	14	15	18	17	16
3	15	15	18	16	16
$\overline{Y}_{.j}$	15	16	19	18	$\overline{Y}_{..}$=17

Μοντέλο ANOVA: $Y_{ij} = \mu + \alpha_i + \beta_j$, $i = 1, 2, 3$, $j = 1, 2, 3, 4$. Μας ενδιαφέρει να ελέγξουμε κατά πόσο η μέση κατανάλωση βενζίνης διαφέρει για τους διάφορους τύπους βενζίνης σε επίπεδο στατιστικής σημαντικότητας $\alpha = 1\%$. Δηλαδή, θέλουμε να ελέγξουμε την

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \cdots = \beta_m = 0 \\ \text{εναντι} & \\ H_1 &: \beta_i \neq \beta_j \text{ για τουλάχιστον ένα ζεύγος } i, j \end{aligned}$$

Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	df	MS	$F = \frac{MSF}{MSE}$
Παράγοντας A	$SSA = 24$	$m - 1 = 2$	$MSA = \frac{24}{2} = 12$	$F_A = \frac{12}{2/3} = 18$
Παράγοντας B	$SSB = 30$	$l - 1 = 3$	$MSB = \frac{30}{3} = 10$	$F_B = \frac{10}{2/3} = 15$
Σφάλμα(Error)	$SSE = 4$	$(m-1)(l-1) = 6$	$MSE = \frac{4}{6} = \frac{2}{3}$	
Σύνολο(Total)	$SST = 58$	$ml - 1 = 11$		

Σε επίπεδο στατιστικής σημαντικότητας $\alpha = 1\%$ απορίπτουμε την H_0 εαν $F_B > F_{0.01}(3, 6)$. Έχουμε $F_B = 15 > F_{0.01}(3, 6) = 9.78$, άρα απορρίπτουμε την H_0. Δηλαδή η μέση κατανάλωση βενζίνης διαφέρει στατιστικά σημαντικά για τους διάφορους τύπους βενζίνης, στις 3 μάρκες αυτοκινήτων που εξετάστηκαν.

Άσκηση. Ένας κατασκευαστής τηλεοράσεων ενδιαφέρεται να ελέγξει αν η αγωγιμότητα του περιβλήματος τεσσάρων διαφορετικών τύπων καλωδιών τηλεόρασης διαφέρει. Μετρήσεις για την αγωγιμότητα των περιβλημάτων καλωδιών δίνονται στον παρακάτω πίνακα.

Τύπος καλωδίου i	Αγωγιμότητα Y_{ij}	$Y_{i\cdot}$	$\overline{Y}_{i\cdot}$
1	143 141 150 146	580	145
2	152 149 137 143	581	145.25
3	144 146 142 137	569	142.25
4	129 127 132 129	517	129.25
		$Y_{\cdot\cdot} = 2247$	$\overline{Y}_{\cdot\cdot} = 140.43$

(a) Να ελεγχθεί σε $\alpha = 5\%$ αν υπάρχει διαφορά στην αγωγιμότητα του περιβλήματος των τεσσάρων τύπων καλωδιών.
(b) Να κατασκευαστεί διάστημα εμπιστοσύνης συντελεστή 95% για τον μέσο του τέταρτου τύπου καλωδίου.
(c) Ο κατασκευαστής αποφασίζει να χρησιμοποιήσει τον τέταρτο τύπο καλωδίου γιατί υποστηρίζει ότι δίνει τη μικρότερη αγωγιμότητα περιβλήματος σε σύγκριση με τους άλλους τρεις τύπους. Να ελεγχθεί σε $\alpha = 5\%$ αν ισχύει ο ισχυρισμός του κατασκευαστή.

Δίνονται $F_{0.05}(3,12) = 3.49$, $t_{0.025}(12) = 2.179$ και $t_{0.05}(12) = 1.782$.

Λύση:
(a)

$$
\begin{aligned}
SST &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2 = 925.9375 \\
SSF &= \sum_{i=1}^{m} n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 = 689.6875
\end{aligned}
$$

Άρα $SSE = SST - SSF = 236.25$.

Πίνακας ANOVA

Πηγή διασποράς	SS	df	MS	$F = \frac{MSF}{MSE}$
F	$SSF = 689.6875$	$m-1=3$	$229.8958 = \frac{SSF}{m-1}$	11.6772
E	$SSE = 236.25$	$n-m=12$	$19.6875 = \frac{SSE}{n-m}$	
T	$SST = 925.9375$	$n-1=15$		

Θεωρούμε τον έλεγχο της

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ \varepsilon\nu\alpha\nu\tau\iota & \\ H_1 &: \mu_i \neq \mu_j \text{ για ένα τουλάχιστον ζεύγος } i, j \end{aligned}$$

Εφόσον $F = 11.6772 > F_{0.05}(3, 12) = 3.49$ απορρίπτουμε την H_0 σε επίπεδο σημαντικότητας $\alpha = 5\%$. Άρα υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στους μέσους των επιπέδων.

(b) Για τη μέση αγωγιμότητα του τέταρτου τύπου καλωδίου, δηλαδή για το μ_4, έχουμε:

$$\begin{aligned} \overline{Y}_{4\cdot} &= 129.25 \\ S^2_{\overline{Y}_{4\cdot}} &= \frac{MSE}{n_4} = 4.92219 \end{aligned}$$

$$Pr[-t_{0.025}(12) \leq \frac{\overline{Y}_{4\cdot} - \mu_4}{\sqrt{S^2_{\overline{Y}_{4\cdot}}}} \leq t_{0.025}(12)] = 0.95$$

Το αντίστοιχο διάστημα εμπιστοσύνης, συντελεστή 95%, είναι

$$\begin{aligned} \overline{Y}_{4\cdot} - \sqrt{S^2_{\overline{Y}_{4\cdot}}}\, t_{0.025}(12) \leq & \quad \mu_4 \quad \leq \overline{Y}_{4\cdot} + \sqrt{S^2_{\overline{Y}_{4\cdot}}}\, t_{0.025}(12) \\ 129.25 - \sqrt{4.9219}\, 2.179 \leq & \quad \mu_4 \quad \leq 129.25 + \sqrt{4.9219}\, 2.179 \\ 124.4159 \leq & \quad \mu_4 \quad \leq 134.0841. \end{aligned}$$

σε επίπεδο

(c) Θεωρούμε το contrast $L = \frac{\mu_1+\mu_2+\mu_3}{3} - \mu_4$. Έχουμε

$$\begin{aligned} \hat{L} &= \frac{\overline{Y}_{1\cdot} + \overline{Y}_{2\cdot} + \overline{Y}_{3\cdot}}{3} - \overline{Y}_{4\cdot} = 14.9167 \\ \sum_{i=1}^{4} \frac{c_i^2}{n_i} &= \frac{1}{4}[\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + 1] = \frac{1}{3} \\ S^2_{\hat{L}} &= MSE[\sum_{i=1}^{4} \frac{c_i^2}{n_i}] = 6.5625 \end{aligned}$$

Άρα το διάστημα εμπιστοσύνης δίνεται ως

$$\begin{aligned}\hat{L}-\sqrt{S^2_{\hat{L}}}\,t_{0.025}(12)\leq\ & L \leq \hat{L}+\sqrt{S^2_{\hat{L}}}\,t_{0.025}(12)\\ 14.9167-2.5617\cdot 2.179\leq\ & L \leq 14.9167+2.5617\cdot 2.179\\ 9.3348\leq\ & L \leq 20.4986.\end{aligned}$$

Εφόσον το 0 δεν περιέχεται στο διάστημα εμπιστοσύνης η διαφορά του τέταρτου τύπου καλωδίου από τους άλλους τρείς είναι στατιστικά σημαντική σε ε.σ.σ $\alpha = 5\%$.

Ο ισχυρισμός του κατασκευαστή είναι ότι η μέση αγωγιμότητα του περιβλήματος του τέταρτου τύπου καλωδίου είναι μικρότερη από αυτή των τριών άλλων τύπων, δηλαδή $\mu_4 < \frac{\mu_1+\mu_2+\mu_3}{3}$. Ο ισχυρισμός μπορεί να ελεγχθεί μέσω ενός ελέγχου υποθέσεων με μονόπλευρη εναλλακτική. Συγκεκριμένα, ο έλεγχος διατυπώνεται ως

$$\begin{aligned}H_0 &: L=0\\ H_1 &: L>0.\end{aligned}$$

Η ελεγχοσυνάρτηση είναι Συγκεκριμένα είναι $\frac{\hat{L}}{\sqrt{S^2_{\hat{L}}}}$, η οποία κάτω από την ισχύ της H_0 ακολουθεί $t(12)$ κατανομή. Η παρατηρούμενη τιμή της ελεγχοσυνάρτησης είναι $\frac{\hat{L}}{\sqrt{S^2_{\hat{L}}}} = 5.8230 > 1.782 = t_{0.05}(12)$. Επομένως, απορρίπτουμε την H_0 σε επίπεδο σημαντικότητας $\alpha = 5\%$ και άρα ισχύει ο ισχυρισμός του κατασκευαστή.

ANOVA κατά δυο παράγοντες με αλληλεπίδραση

Οι δυο παράγοντες που εξετάζουμε μπορεί να αλληλεπιδρούν μεταξύ τους. Για να ελέγξουμε για πιθανή αλληλεπίδραση (interaction) εξετάζουμε δείγμα μεγάθους c για κάθε συνδυασμό επιπέδων των δυο παραγόντων (treatment). Έχουμε δηλαδή παρατηρήσεις $Y_{ijr} \sim N(\mu_{ij}, \sigma^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, l$ και $r = 1, \ldots, c$.

Μοντέλο ANOVA

$$Y_{ij} \;=\; \mu_{\cdot\cdot} + \alpha_i + \beta_j + \gamma_{ij}$$

έτσι ώστε $\sum_{i=1}^m \alpha_i = 0$, $\sum_{j=1}^l \beta_j = 0$ και $\sum_{i=1}^m \gamma_{ij} = 0$, $j = 1, \ldots, l$ και $\sum_{j=1}^l \gamma_{ij} = 0$, $i = 1, \ldots, m$.

$$\begin{aligned}
\overline{Y}_{ij\cdot} &= \frac{1}{c}\sum_{r=1}^{c} Y_{ijr} & \overline{Y}_{i\cdot r} &= \frac{1}{l}\sum_{j=l}^{c} Y_{ijr} & \overline{Y}_{\cdot jr} &= \frac{1}{m}\sum_{i=l}^{m} Y_{ijr} \\
\overline{Y}_{i\cdot\cdot} &= \frac{1}{lc}\sum_{j=1}^{l}\sum_{r=1}^{c} Y_{ijr} & \overline{Y}_{\cdot j\cdot} &= \frac{1}{mc}\sum_{i=1}^{m}\sum_{r=1}^{c} Y_{ijr} & \overline{Y}_{\cdot\cdot r} &= \frac{1}{ml}\sum_{i=1}^{m}\sum_{j=1}^{l} Y_{ijr} \\
\overline{Y}_{\cdots} &= \frac{1}{mlc}\sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{r=1}^{c} Y_{ijr}
\end{aligned}$$

Συνολική μεταβλητότητα :

$$\begin{aligned}
SST &= \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{r=1}^{c}(Y_{ijr} - \overline{Y}_{\cdots})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{r=1}^{c}\left((\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}) + (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots}) + (Y_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdots}) + (\overline{Y}_{ijr} - \overline{Y}_{ij\cdot})\right)^2 \\
&= \sum_{i=1}^{m} lc(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots})^2 + \sum_{j=1}^{l} mc(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})^2 + \sum_{i=1}^{m}\sum_{j=1}^{l} c(Y_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdots})^2 \\
&\quad + \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{r=1}^{c}(\overline{Y}_{ijr} - \overline{Y}_{ij\cdot})^2\,,
\end{aligned}$$

Ορίζουμε

$SSA = \sum_{i=1}^m lc(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots})^2$

SSA **:** άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα Α

$SSB = \sum_{j=1}^{l} mc(\overline{Y}_{\cdot j \cdot} - \overline{Y}_{\dots})^2$

SSB : άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα B

$SS(AB) = \sum_{i=1}^{m} \sum_{j=1}^{l} c(Y_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\dots})^2$

$SS(AB)$: άθροισμα τετραγώνων της αλληλεπίδρασης

$SSE = \sum_{i=1}^{m} \sum_{j=1}^{l} \sum_{r=1}^{c} (\overline{Y}_{ijr} - \overline{Y}_{ij\cdot})^2$

SSE : άθροισμα τετραγώνων των τυχαίων σφαλμάτων.

Επομένως

$$SST = SSA + SSB + SS(AB) + SSE$$

Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	df	MS	$F = \frac{MSF}{MSE}$
Παράγοντας A	SSA	$m-1$	$MSA = \frac{SSA}{m-1}$	$F_A = \frac{MSA}{MSE}$
Παράγοντας B	SSB	$l-1$	$MSB = \frac{SSB}{l-1}$	$F_B = \frac{MSB}{MSE}$
Παράγοντας AB	$SS(AB)$	$(m-1)(l-1)$	$MS(AB) = \frac{SS(AB)}{(m-1)(l-1)}$	$F_{AB} = \frac{MS(AB)}{MSE}$
Σφάλμα(Error)	SSE	$ml(c-1)$	$MSE = \frac{SSE}{ml(c-1)}$	
Σύνολο(Total)	SST	$mlc-1$		

F–test για αλληλεπίδραση

Θεωρούμε τον έλεγχο

$$\begin{aligned} H_0 &: \gamma_{ij} = 0 \ i = 1, \dots, m, \ j = 1, \dots, l \\ H_1 &: \gamma_{ij} \neq 0 \text{ για ένα τουλαχιστον } i, j \end{aligned}$$

Απορρίπτουμε την H_0 σε επίπεδο στατιστικής σημαντικότητας α αν $F_{AB} = \frac{MS(AB)}{MSE} > F_\alpha((m-1)(l-1), ml(c-1))$.

Ελέγχουμε πρώτα την ύπαρξη στατιστικά σημαντικής αλληλεπίδρασης και έπειτα την στατιστική σημαντικότητα των επιμέρους παραγόντων.