ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά –
Χρηματοοικονομικά Μαθηματικά

2017

Υπολογισμοί Ανάλυσης Διακύμανσης

Μοντέλο:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Υπόθεση:

$$\sum_{i,j}^{N} \varepsilon_{ij} = 0$$

$Y_{i\bullet} = \sum_{j=1}^{r_i} Y_{ij}$ το άθροισμα στην i κατηγορία	$Y_{\bullet\bullet} = \sum_{i=1}^{p} \sum_{j=1}^{r_i} Y_{ij}$ το συνολικό άθροισμα
$\bar{Y}_{i\bullet} = \frac{1}{r_i} \sum_{j=1}^{r_i} Y_{ij}$ το μέσο άθροισμα στην i κατηγορία	$\bar{Y}_{\bullet\bullet} = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{r_i} Y_{ij}$ ο γενικός μέσος

Υπολογισμοί Ανάλυσης Διακύμανσης

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{r_i}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^{\;2} - \left(\sum_{i=1}^{p}\frac{Y_{i\bullet}^{\;2}}{r_i}\right)$$

αθ. τετρ. αποκλίσεων από το μέσο κάθε κατηγορίας

$$SSA = \sum_{i=1}^{p} r_i\left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 = \left(\sum_{i=1}^{p}\frac{Y_{i\bullet}^{\;2}}{r_i}\right) - \frac{Y_{\bullet\bullet}^{\;2}}{n}$$

αθ. τετρ. αποκλίσεων με την υπόθεση όλες οι κατηγορίες έχουν τη δική τους τιμή

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{r_i}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^{\;2} - \frac{Y_{\bullet\bullet}^{\;2}}{n}$$

συνολικό αθ. τετρ. αποκλίσεων

SST=SSA+SSE

Αν $\overline{Y}_{i\bullet} = \overline{Y}_{\bullet\bullet}\ ,\ \forall i$ τότε SSA=0, και SST=SSE.	Αν $\overline{Y}_{ij} = \overline{Y}_{i\bullet}\ ,\ \forall i, j$ τότε SSE=0, και SST=SSA.

Υπολογισμοί Ανάλυσης Διακύμανσης

Παρατηρήστε ότι αν συμβεί όλα τα Y_{ij} να είναι ίσα με $\bar{Y}_{i\bullet}$

τότε το **SSE=0** και **SST=SSA**,

δηλαδή στην περίπτωση αυτή όλη η διασπορά οφείλεται στις κατηγορίες του παράγοντα Α. Επομένως οι **ΚΑΤΗΓΟΡΙΕΣ** του παράγοντα Α είναι πολύ σημαντικές

$$\bar{Y}_{i\bullet} = \bar{Y}_{\bullet\bullet}$$

Όμοια αν όλα τα $\bar{Y}_{i\bullet}$ βρεθούν ίσα με $\bar{Y}_{\bullet\bullet}$

τότε το **SSA=0** και **SST=SSE**,

δηλαδή στην περίπτωση αυτή όλη η διασπορά οφείλεται σε καθαρά σφάλματα μετρήσεων. Επομένως οι **ΚΑΤΗΓΟΡΙΕΣ** του παράγοντα Α δεν έχουν καμία σημασία.

$$\bar{Y}_{ij} = \bar{Y}_{i\bullet}$$

Υπολογισμοί Ανάλυσης Διακύμανσης

- Γενικά για τον υπολογισμό των βαθμών ελευθερίας αθροισμάτων με όρους που κατανέμονται με τυπική κανονική κατανομή, ισχύει:

$$\textbf{β.ε. (για SS)} = \left\{ \begin{array}{c} \text{πλήθος ανεξάρτητων} \\ \text{τετραγώνων που} \\ \text{υπολογίζουν το SS} \end{array} \right\} - \left\{ \begin{array}{c} \text{πλήθος ανεξάρτητων} \\ \text{παραμέτρων που} \\ \text{εκτιμώνται} \end{array} \right\}$$

Στο SST έχουμε n τετράγωνα από τα οποία υπολογίζεται μία παράμετρος η $\mathbf{Y_{..}}$ $\Rightarrow$ **β.ε.=n-1**

$$SST = \sum_{i=1}^{p} \sum_{j=1}^{r_i} \left(Y_{ij} - \bar{Y}_{\bullet\bullet} \right)^2$$

Στο SSE έχουμε n τετράγωνα από τα οποία υπολογίζονται p παράμετροι οι $\mathbf{Y_{1.}, Y_{2.}, \ldots, Y_{p.}}$ $\Rightarrow$ **β.ε.=n-p**

$$SSE = \sum_{i=1}^{p} \sum_{j=1}^{r_i} \left(Y_{ij} - \bar{Y}_{i\bullet} \right)^2$$

Στο SSA έχουμε p τετράγωνα από τα οποία υπολογίζονται 1 παράμετρος η $\mathbf{Y_{..}}$ $\Rightarrow$ **β.ε.=p-1**

$$SSA = \sum_{i=1}^{p} r_i \left(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet} \right)^2$$

Υπολογισμοί Ανάλυσης Διακύμανσης

ΘΕΩΡΗΜΑ

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p$

H_1 : όχι η H_0

$$F = \frac{SSA/(p-1)}{SSE/(n-p)} \sim F_{p-1,n-p}$$

Πηγή	Αθροίσμ. Τετραγ.	β.ε.	Μέσα τετρ.	F
Παράγοντας Α (μεταξύ ομάδων)	$SSA = \sum_{i=1}^{p} r_i \left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2 = \left(\sum_{i=1}^{p} \frac{Y_{i\cdot}^2}{r_i}\right) - \frac{Y_{\cdot\cdot}^2}{n}$	p-1	$MSA = \frac{SSA}{p-1}$	$F = \frac{MSA}{MSE}$
Υπόλοιπα (ανάμεσα στις ομάδες)	$SSE = \sum_{i=1}^{p}\sum_{j=1}^{r_i}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^2 - \left(\sum_{i=1}^{p} \frac{Y_{i\cdot}^2}{r_i}\right) = SST - SSA$	n-p	$MSE = \frac{SSE}{n-p}$	
Σύνολο	$SST = \sum_{i=1}^{p}\sum_{j=1}^{r_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{n}$	n-1		

Υπολογισμοί Ανάλυσης Διακύμανσης

Στο μοντέλο $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

η σταθερά μ θεωρείται ότι εκφράζει τον γενικό μέσο όρο των παρατηρήσεων. Άρα

$$E\left(Y_{1j}\right) = \mu + \alpha_1 = \bar{Y}_{1\bullet}$$

$$E\left(Y_{2j}\right) = \mu + \alpha_2 = \bar{Y}_{2\bullet}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$E\left(Y_{1j}\right) = \mu + \alpha_p = \bar{Y}_{p\bullet}$$

$$\Rightarrow \boxed{\bar{Y}_{\bullet\bullet} = \mu + \frac{r_1\alpha_1 + r_2\alpha_2 + \ldots + r_p\alpha_p}{r_1 + r_2 + \ldots + r_p}}$$

Αν $\alpha_1 + \alpha_2 + \ldots + \alpha_p = 0$

$r_1 = r_2 = \ldots = r_p$

$$\Downarrow$$

$$r_1\alpha_1 + r_2\alpha_2 + \ldots + r_p\alpha_p = 0$$

$$\Downarrow$$

$$\boxed{\bar{Y}_{\bullet\bullet} = \mu}$$

Υπολογισμοί Ανάλυσης Διακύμανσης

Αν $$\frac{r_1 \alpha_1 + r_2 \alpha_2 + \ldots + r_p \alpha_p}{r_1 + r_2 + \ldots + r_p} = \alpha_0$$

Θέτοντας $$\mu^* = \mu + \alpha_0 \qquad \alpha_i{}^* = \alpha_i - \alpha_0$$

Άρα

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} = (\mu + \alpha_0) + (\alpha_i - \alpha_0) + \varepsilon_{ij} \quad \Rightarrow \; Y_{ij} = \mu^* + \alpha_i{}^* + \varepsilon_{ij}$$

Επειδή

$$r_1 \alpha_1{}^* + r_2 \alpha_2{}^* + \ldots + r_p \alpha_p{}^* =$$
$$= r_1 \alpha_1 + r_2 \alpha_2 + \ldots + r_p \alpha_p - \alpha_0 \sum_i r_i = 0$$

$$\Downarrow$$

$$\boxed{\bar{Y}_{\bullet\bullet} = \mu^*}$$

Υπολογισμοί Ανάλυσης Διακύμανσης

- Τρεις μέθοδοι διδασκαλίας εφαρμόστηκαν σε24 άτομα (ανά8 η κάθε μία) και μετά εξετάστηκαν σε κοινά θέματα. Η βαθμολογία ήταν:

Μέθ. 1	Μέθ. 2	Μέθ. 3
3	4	6
5	4	7
2	3	8
4	8	6
8	7	7
4	4	9
3	2	10
9	5	9

- Επηρεάζει η μέθοδος τη βαθμολογία?

Άσκηση

Μέθ. 1	Μέθ. 2	Μέθ. 3
3	4	6
5	4	7
2	3	8
4	8	6
8	7	7
4	4	9
3	2	10
9	5	9

Σύν.	38	37	62
$\overline{x}$	4.75	4.625	7.75

Τυχαιοποιημένο σχέδιο

$p=3,\ r_i=8$

Μέθ. 1	Μέθ. 2	Μέθ. 3
$Y_{11}=\mu+\alpha_1+\varepsilon_{11}$ $Y_{12}=\mu+\alpha_1+\varepsilon_{12}$ …………… …………… …… $Y_{18}=\mu+\alpha_1+\varepsilon_{18}$	$Y_{21}=\mu+\alpha_2+\varepsilon_{21}$ $Y_{22}=\mu+\alpha_2+\varepsilon_{22}$ …………… …………… …… $Y_{28}=\mu+\alpha_2+\varepsilon_{28}$	$Y_{31}=\mu+\alpha_3+\varepsilon_{31}$ $Y_{32}=\mu+\alpha_3+\varepsilon_{32}$ …………… …………… …… $Y_{38}=\mu+\alpha_3+\varepsilon_{38}$
$Y_{1.}/8=\mu+\alpha_1$	$Y_{2.}/8=\mu+\alpha_2$	$Y_{3.}/8=\mu+\alpha_3$

Άσκηση

Υπόθεση

$$\sum_{j} \varepsilon_{1j} = 0, \sum_{j} \varepsilon_{2j} = 0, \sum_{j} \varepsilon_{3j} = 0$$

$p=3,\ r_i=8,\ n=24$

$$Y_{1\bullet} = 38, Y_{2\bullet} = 37, Y_{3\bullet} = 62, Y_{\bullet\bullet} = 137$$

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^{\ 2} - \left(\sum_{i=1}^{p} \frac{Y_{i\bullet}^{\ 2}}{r_i}\right) = 919 - \left(\frac{38^2}{8} + \frac{37^2}{8} + \frac{62^2}{8}\right) = 86.875$$

$$SSA = \left(\sum_{i=1}^{p} \frac{Y_{i\bullet}^{\ 2}}{r_i}\right) - \frac{Y_{\bullet\bullet}^{\ 2}}{n} = \left(\frac{38^2}{8} + \frac{37^2}{8} + \frac{62^2}{8}\right) - \frac{137^2}{24} = 50.083$$

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{r_i} Y_{ij}^{\ 2} - \frac{Y_{\bullet\bullet}^{\ 2}}{n} = 919 - \frac{137^2}{24} = 136.958$$

Άσκηση

Πηγή	Αθροίσμ. Τετραγ.	β.ε.	Μέσα τετρ.	F
ΜΕΘΟΔΟΙ (μεταξύ ομάδων)	50.083	2	25.04	6.05
Υπόλοιπα (ανάμεσα στις ομάδες)	86.875	21	4.14	
Σύνολο	136.958	23		

- Υπόθεση

$H_0 : \alpha_1=\alpha_2=\ldots=\alpha_p$

H_1 : όχι η H_0

$F_{2,21;0.01} \approx 5.80$

ΑΠΟΡΡΙΠΤΕΤΑΙ σε $\alpha=0.01$

Άσκηση

(αντοχή υλικών)

A	B	Γ	Δ	
1.700 1.505 1.800	1.725 1.825 1.660 1.640	1.735 1.925 1.815 1.890	2.025 2.075 2.110 1.970	
$Y_{1\cdot} = 5.005$ $r_1 = 3$ $\bar{Y}_{1\cdot} = 1.66833$	$Y_{2\cdot} = 6.850$ $r_2 = 4$ $\bar{Y}_{2\cdot} = 1.71250$	$Y_{3\cdot} = 7.365$ $r_3 = 4$ $\bar{Y}_{3\cdot} = 1.84125$	$Y_{4\cdot} = 8.180$ $r_4 = 4$ $\bar{Y}_{4\cdot} = 2.04500$	$Y_{\cdot\cdot} = 27.4$ $n = 15$ $\bar{Y}_{\cdot\cdot} = 1.826667$

$$SST = 1.700^2 + 1.505^2 + \ldots + 1.970^2 - \frac{27.4^2}{15} = 0.4172333$$

$$SSA = \left(\frac{5.005^2}{3} + \frac{6.850^2}{4} + \frac{7.385^2}{4} + \frac{8.180^2}{4} \right) - \frac{27.4^2}{15} = 0.3188729$$

$$SSE = 0.4172333 - 0.3188729 = 0.0983604$$

Άσκηση

Πηγή	Αθροίσμ. Τετραγ.	β.ε.	Μέσα τετρ.	F
ΥΛΙΚΑ (μεταξύ ομάδων)	0.3188729	3	0.106291	11.8869
Υπόλοιπα (ανάμεσα στις ομάδες)	0.0983604	11	0.00702574	
Σύνολο	0.0417233	14		

$H_0 : \alpha_1=\alpha_2=\alpha_3=\alpha_4$

H_1 : όχι η H_0

$F_{3,11;0.01} < F = 11.8869$

ΑΠΟΡΡΙΠΤΕΤΑΙ σε $\alpha=0.01$

Κύριες Επιδράσεις

$$\hat{\alpha}_1 = -0.1583333333,\ \hat{\alpha}_2 = -0.1141666667,\ \hat{\alpha}_3 = 0.01458333333,\ \hat{\alpha}_4 = 0.2183333333$$

$\bar{Y}_{1.} - \bar{Y}_{..} = 1.668 - 1{,}836$

$$3\alpha_1 + 4\alpha_2 + 4\alpha_3 + 4\alpha_4 = 0$$