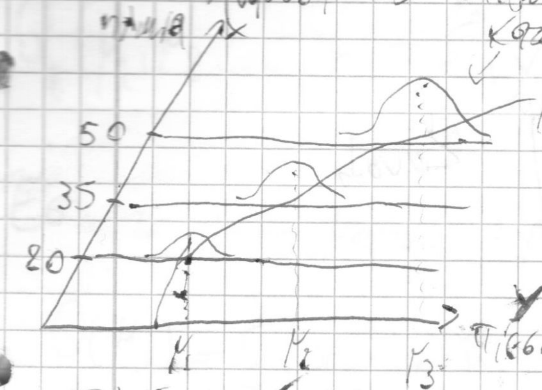
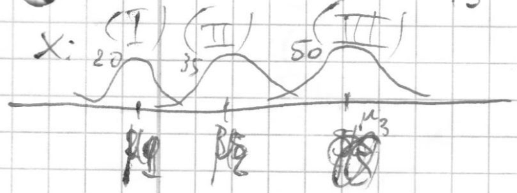
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})}$$

20/1/2016

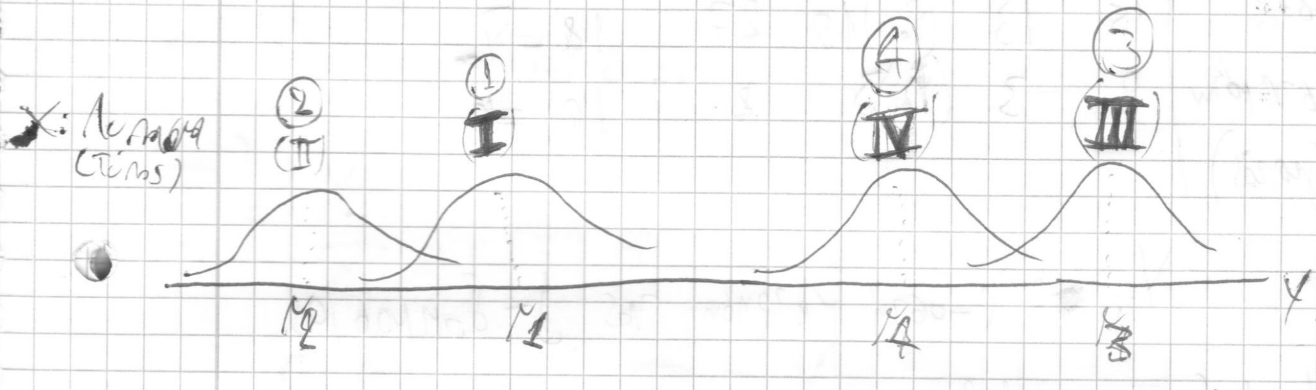
Ανάλυση της διακύμανσης





$\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma^2$

μας ενδιαφέρει αν διαφέρουν τα μ_1, μ_2, μ_3



$H_0: \mu_1=\mu_2=\mu_3=\mu_4$ V H_a: όχι όλα τα μ_i ίσα μεταξύ τους

t-τεστ ανά δύο $\binom{4}{2}=6$ συγκρίσεις

Για $\alpha=0,05$ $\quad H_{0_1}: \mu_1=\mu_2, \ldots, H_{0_6}: \mu_4=\mu_3 \quad : \quad H_0 : \bigcup_{k=1}^{6} H_{0_k}$

$a' = P$(απορρ. H_0 | όταν H_0 αληθής) $= P$(τουλάχιστον μία από τις $H_{0_1},\ldots,H_{0_6}$ | ενώ αληθεύει η H_0) =

$= P$(τουλάχιστον μία επιτυχία σε έξι δοκιμές με $p = P(\text{επιτυχίας}) = a$) =

$= 1 - P$(καμία επιτυχία σε έξι δοκιμές με $p=a$) =

$= 1-\binom{6}{0}a^0(1-a)^6 = 0,265$ για $a=0,05$

Στην ανάλυση της διακύμανσης η ορολογία είναι:

ανεξάρτητη μεταβλητή = παράγοντας

διαφορετικές τιμές της ανεξάρτητης μεταβλητής = επίπεδα του παράγοντα

μετρήσεις κατά παράγοντα = δοκιμασίες

Παράδειγμα

Η εταιρία ξυραφακιών ABEZ

(Δειγματική Μονάδα) Καταστήματα Χονδρικής (i)	(Δοκιμασίες) τύπος ξυραφακιών (j) 1	2	3	4	Σύνολο
1	18	14	19	24	
2	18	12 (Y_{23})	17	30	
3		13	21		
Σύνολο:	30	39	($Y_{\cdot 3}$=) 57	54	$180 = Y_{\cdot\cdot}$
Δειγματικός μέσος:	15	13	($\bar{Y}_{\cdot 3}$=) 19	27	$18 = \bar{Y}_{\cdot\cdot}$
Αριθμός καταστημάτων (Μέγεθος δείγματος)	2	3	(n_3=) 3	2	$10 = n$

Y_{ij}: i-οστή μέτρηση της j-δοκιμασίας

$Y_{ij} = \mu_j + \varepsilon_{ij}$ $\quad i = 1, 2, \dots, n$

$j = 1, 2, \dots, r$

r δοκιμασίες

$n = \sum_{j=1}^{r} n_j$

τυχαία σφάλματα: $\varepsilon_{ij} = Y_{ij} - \mu_j \sim N(0, \sigma^2)$
(ανεξάρτητες τμ)

(ανεξάρτητες τμ) $Y_{ij} \sim N(\mu_j, \sigma^2)$

$H_0: \mu_1 = \mu_2 = \dots = \mu_r \quad V \quad H_a$: όχι όλα τα μ_j ίσα μεταξύ τους

Εκτίμηση των παραμέτρων (με τη μέθοδο των ελαχίστων τετραγώνων)

- $$Q(\mu_1, \mu_2, \dots, \mu_r) = \sum_{j=1}^{r} \sum_{i=1}^{n_j} \varepsilon_{ij}^2 = \sum_{j=1}^{r} \sum_{i=1}^{n_j} (Y_{ij} - \mu_j)^2 \Rightarrow \hat{\mu}_j$$

$$Q_1(\mu_1, \mu_2, \dots, \mu_r) = \sum_{i=1}^{n_1} (Y_{i1} - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_{i2} - \mu_2)^2 + \dots + \sum_{i=1}^{n_r} (Y_{ir} - \mu_r)^2$$

Έστω το j-στού μέρους άθροισμα:

- $$Q(\mu_j) = \sum_{i=1}^{n_j} (Y_{ij} - \mu_j)^2 \rightsquigarrow \frac{dQ}{d\mu_j} = -2 \sum_{i=1}^{n_j} (Y_{ij} - \hat{\mu}_j) = 0 \rightsquigarrow$$

$$n_j \hat{\mu}_j = \sum_{i=1}^{n_j} Y_{ij} \rightsquigarrow \hat{\mu}_j = \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j}$$

$$Y_{\cdot j} = \sum_{i=1}^{n_j} Y_{ij} \quad , j = 1, 2, \dots, r$$

- $$\overline{Y}_{\cdot j} = \frac{Y_{\cdot j}}{n_j}$$: μέσος όρος της j-δειγματοληψίας

$$Y_{\cdot\cdot} = \sum_{j=1}^{r} \sum_{i=1}^{n_j} Y_{ij} \quad , \quad \overline{Y}_{\cdot\cdot} = \frac{Y_{\cdot\cdot}}{n}$$

- $$\hat{\mu}_j = \frac{Y_{\cdot j}}{n_j} = \overline{Y}_{\cdot j}$$

$$\min Q(\mu_1, \mu_2, \dots, \mu_r) = \sum_{j=1}^{r} \sum_{i=1}^{n_j} e_{ij}^2 \Rightarrow \hat{\mu}_j = \overline{Y}_{\cdot j}$$

- $$(\hat{\varepsilon}_{ij} =) \; e_{ij} = Y_{ij} - \hat{\mu}_j = Y_{ij} - \overline{Y}_{\cdot j}$$

Υπόλοιπα

$$\sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_{\cdot j}) = 0$$

$\forall j = 1, 2, \dots$

21/4/2016

Η ανάλυση της διακύμανσης

$$SS_{\underset{\text{total}}{tot}} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2$$

συνολικό άθροισμα τετραγώνων
δίνει τη συνολική μεταβλητότητα

$$Y_{ij} - \overline{Y}_{\cdot\cdot} = \overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot} + \underbrace{Y_{ij} - \overline{Y}_{\cdot j}}_{=e_{ij}} \Longrightarrow$$

$$\sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right)^2 + 2\sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)\bigl(Y_{ij}$$

$$= \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right)^2 + \underbrace{2\sum_{j=1}^{r}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right)}_{= 2\sum_{j=1}^{r}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)\underbrace{\left(Y_{\cdot j} - n_j\overline{Y}_{\cdot j}\right)}_{=0} \;=\; 0}$$

$$\Longrightarrow \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right)^2$$

SS_{tot}		SS_{tre}		SS_{reg}
total		treatment		residual
$n-1$ βαθμοί ελευθερίας	=	$r-1$ βαθμοί ελευθερίας	+	$n-r$ βαθμοί ελευθερίας
γιατί $\sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right) = 0$		γιατί $\sum_{j=1}^{r} n_j\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right) = 0$		γιατί $\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right) = 0 \quad \forall j = 1, 2, \ldots, r$

$$SS_{tre} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 = \sum_{j=1}^{r} n_j(\bar{Y}_{.j}-\bar{Y}_{..})^2 = \sum_{j=1}^{r}\frac{Y_{.j}^2}{n_j} - \frac{Y_{..}^2}{n}$$

$$SS_{tot} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j} Y_{ij}^2 - \frac{Y_{..}^2}{n}$$

$$SS_{reg} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j} Y_{ij}^2 - \sum_{j=1}^{r}\frac{Y_{.j}^2}{n_j}$$

ΑΝΑΛΔΙΑ

Προέλευση Μεταβλητότητας	Αθροίσματα τετραγώνων	Βαθμοί ελευθερίας	Μέσα τετράγωνα	F-τεστ
Δοκιμασίες ή μεταξύ των δοκιμασιών	$SS_{tre} = \sum_{j=1}^{r} n_j(\bar{Y}_{.j}-\bar{Y}_{..})^2$	$r-1$	$MS_{tre} = \frac{SS_{tre}}{r-1}$	$F = \frac{MS_{tre}}{MS_{reg}}$
Υπόλοιπα ή εντός των δοκιμασιών	$SS_{reg} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 = SS_{tot}-SS_{tre}$	$n-r$	$MS_{reg} = \frac{SS_{reg}}{n-r}$	$F \sim F_{r-1,n-r}$ με κρίσιμη περιοχή: $F \geq F_{a,r-1,n-r}$
Ολική μεταβλητότητα	$SS_{tot} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{..})^2$	$n-1$		

Στο παράδειγμα της ΑΒΕΖ

258	3	86	$F = \frac{86}{7.67} = 11.2 > 4.76 = F_{0.05,3,6}$
46	6	7.67	$\Rightarrow$ απορρίπτουμε την H_0
304	9		

Άσκηση 6.1.

$$E(MS_{tre}) = \sigma^2 + \frac{1}{r-1}\sum_{j=1}^{r} n_j(\mu_j-\mu)^2 \quad \text{όπου} \quad \mu = \frac{\sum_{j=1}^{r} n_j \mu_j}{n}$$

$$E(MS_{reg}) = E\left[\frac{1}{n-r}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{\cdot j})^2\right] = E\left[\frac{1}{n-r}\sum_{j=1}^{r}(n_j-1)\underbrace{\frac{1}{n_j-1}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{\cdot j})^2}_{=S_j^2}\right] =$$

$$= \frac{1}{n-r}\sum_{j=1}^{r}(n_j-1)\underbrace{E(S_j^2)}_{=\sigma^2\ \forall j} \overset{\sum_{j=1}^{r} n_j = n}{\underset{\sum_{j=1}^{r}(-1)=-r}{=}} \frac{1}{n-r}(n-r)\sigma^2 = \sigma^2$$

$$\left.\begin{array}{l} \dfrac{SS_{tre}}{\sigma^2} \sim \chi^2_{r-1} \quad \text{όταν } H_0 \text{ αληθής} \\[2ex] \dfrac{SS_{reg}}{\sigma^2} \sim \chi^2_{n-r} \end{array}\right\} \Rightarrow \frac{MS_{tre}}{MS_{reg}} \sim F_{r-1,n-r}$$

$$\frac{e_{ij}}{\sqrt{MS_{reg}}} \approx \frac{\varepsilon_{ij}}{\sigma}$$

Αν απορρίπτουμε την H_0:

$$\overline{Y}_{\cdot j} \sim N\left(\mu_j, \frac{\sigma^2}{n_j}\right) \rightsquigarrow$$

$$\frac{\overline{Y}_{\cdot j} - \mu_j}{\sqrt{MS_{reg}/n_j}} \sim t_{n-r}$$

$$H_0: \mu_u - \mu_v = 0 \quad \vee \quad H_a: \mu_u - \mu_v \neq 0$$

$$\frac{\bar{Y}_{\cdot u} - \bar{Y}_{\cdot v} - (\mu_u - \mu_v)}{\sqrt{MS_{reg}\left(\frac{1}{n_u} + \frac{1}{n_v}\right)}} \sim t_{n-r}$$

$(1-\alpha)100\%$ ΔΕ για $\mu_u - \mu_v$: $\bar{Y}_{\cdot u} - \bar{Y}_{\cdot v} \pm t_{\alpha/2, n-r} \sqrt{MS_{reg}\left(\frac{1}{n_u} + \frac{1}{n_v}\right)}$

Στο παράδειγμα της ΑΒΕΖ μας ενδιαφέρει $H_0: \mu_3 - \mu_4 = 0$ V $H_a: \mu_3 - \mu_4 \neq 0$

$u=3$, $v=4$, $n_3 = 3$, $n_4 = 2$, $\bar{Y}_{\cdot 3} = 19$, $\bar{Y}_{\cdot 4} = 27$

$MS_{reg} = 7{,}67$

$$t = \frac{19-27}{\sqrt{7{,}67\left(\frac{1}{3}+\frac{1}{2}\right)}} = -3{,}164$$

Κρίσιμη περιοχή $|t| \geq t_{\alpha/2, n-r} \overset{\alpha=0{,}05}{=} t_{0{,}025, 6} = 2{,}447$

Άρα απορρίπτουμε την H_0