ANOVA de blocs

Exemple

En un estudi es volgué determinar quina varietat de mongeta (d'entre 6 varietats) era la més adient per cultivar en una granja, atès el seu tipus de terra.

Per evitar que els llocs on se sembrassin les diferents mongeteres poguessin afectar els resultats, en prengueren 4 quadrats (blocs) de terreny de la granja, cada un es quadriculà en 6 parts i a cada bloc se li assignà de manera aleatòria a cada quadrat un tipus diferent de mongetera.

Exemple

Les produccions per quadrat (en kg)

	Varietat de mongeta					
Bloc	1	2	3	4	5	6
1	15.9	15.5	13.0	10.3	10.0	6.4
2	16.9	18.1	12.3	13.6	10.1	10.0
3	15.0	16.8	9.0	11.2	8.2	6.5
4	13.9	13.0	11.3	11.4	8.6	8.7

La producció mitjana de les diferents varietats de mongeta en aquest tipus de terra és la mateixa?

Exemple

Variable poblacional global:

- X: Prenc un quadrat sembrat de mongeteres (de les característiques dels quadrats emprats en aquest experiment) i mesur la seva producció (en kg de mongetes)

Subpoblacions:

- Les 6 varietats de mongeta (1,...,6)

Variables d'interès:

- X_i: Prenc un quadrat sembrat amb mongeteres de varietat i i mesur la seva producció ($i = 1, \ldots, 6$)

Contrast:

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 = \cdots = \mu_6 \\ H_1 : \text{Hi ha } i, j \text{ tals que } \mu_i \neq \mu_j \end{array} \right.$$

ANOVA de blocs

En un experiment amb disseny d'ANOVA de blocs:

- Empram els nivells (tractaments) d'un únic factor per classificar la població en $k \geqslant 3$ subpoblacions
  - El tipus de mongetera, amb $k = 6$ nivells
- Prenem una mostra aleatòria de b blocs: conjunts de k subjectes aparellats
  - Els $b = 4$ terrenys que quadriculam en $k = 6$ quadrats
- Dins cada bloc, assignam aleatòriament a cada subjecte un tractament, de manera que cada tractament s'empri exactament un cop dins cada bloc (complet aleatori)
  - Dins cada bloc, assignam aleatòriament a cada quadrat un tractament diferent

L'ANOVA de blocs generalitza el contrast de 2 mitjanes amb mostres aparellades a k mitjanes amb mostres aparellades

Situació general

Mesuram una variable X sobre k subpoblacions (tractaments)

- μ: mitjana poblacional de X en tota la població
- X_i: Prenc un individu del nivell i-èsim i hi mesur la X, $i = 1, \dots, k$; la seva mitjana és μ_i
- $X_{\bullet j}$: Prenc un individu del bloc j-èsim i hi mesur la X, $j = 1, \dots, b$; la seva mitjana és $\mu_{\bullet j}$
- X_{ij}: Prenc l'individu del nivell i-èsim del bloc j-èssim i hi mesur la X; la seva mitjana és μ_{ij}

Situació general

Mesuram una variable X sobre k subpoblacions (tractaments)

- μ: mitjana poblacional de X en tota la població
- X_i: Prenc un individu del nivell i-èsim i hi mesur la X, $i = 1, \ldots, k$; la seva mitjana és μ_i
- $X_{\bullet j}$: Prenc un individu del bloc j-èsim i hi mesur la X, $j = 1, \ldots, b$; la seva mitjana és $\mu_{\bullet j}$
- X_{ij}: Prenc l'individu del nivell i-èsim del bloc j-èssim i hi mesur la X; la seva mitjana és μ_{ij}

Contrast:

$$\left\{\begin{array}{l} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 : \text{Hi ha } i, j \text{ tals que } \mu_i \neq \mu_j \end{array}\right.$$

Situació general

Mesuram una variable X sobre k subpoblacions (tractaments)

- μ: mitjana poblacional de X en tota la població
- X_i: Prenc un individu del nivell i-èsim i hi mesur la X, $i = 1, \ldots, k$; la seva mitjana és μ_i
- $X_{\bullet j}$: Prenc un individu del bloc j-èsim i hi mesur la X, $j = 1, \ldots, b$; la seva mitjana és $\mu_{\bullet j}$
- X_{ij}: Prenc l'individu del nivell i-èsim del bloc j-èssim i hi mesur la X; la seva mitjana és μ_{ij}

Contrast:

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu \\ H_1 : \text{Hi ha } i, j \text{ tals que } \mu_i \neq \mu_j \end{array} \right.$$

Situació general

Les dades es presenten en una taula:

	Tractaments			
Bloc	Tract. 1	Tract. 2	...	Tract. k
1	X_{11}	X_{21}	...	X_{k1}
2	X_{12}	X_{22}	...	X_{k2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	X_{1b}	X_{2b}	...	X_{kb}

X_{ij}:

- La i representa la columna: el tractament
- La j representa la filera: el bloc

Situació general

- $\overline{X}_i$: mitjana mostral del tractament i-èsim

$$\overline{X}_i = \frac{\sum_{j=1}^{b} X_{ij}}{b}$$

- $\overline{X}_{\bullet j}$: mitjana mostral del bloc j-èsim

$$\overline{X}_{\bullet j} = \frac{\sum_{i=1}^{k} X_{ij}}{k}$$

- $\overline{X}$: mitjana mostral global

$$\overline{X} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij}}{k \cdot b}$$

Situació general

Tractaments

Bloc	Tract. 1	Tract. 2	$\ldots$	Tract. k	$\overline{X}_{\bullet j}$
1	X_{11}	X_{21}	$\ldots$	X_{k1}	$\overline{X}_{\bullet 1}$
2	X_{12}	X_{22}	$\ldots$	X_{k2}	$\overline{X}_{\bullet 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	X_{1b}	X_{2b}	$\ldots$	X_{kb}	$\overline{X}_{\bullet b}$
$\overline{X}_i$	$\overline{X}_1$	$\overline{X}_2$	$\ldots$	$\overline{X}_k$	

$$\overline{X}$$

Exemple

Emmagatzemam les dades en un *dataframe* amb tres variables:

- `Prod`: la producció
- `Mong`: la varietat de mongetera (un factor)
- `Bloc`: el bloc (un factor)

	Var. mongetes					
Bloc	1	2	3	4	5	6
1	15.9	15.5	13.0	10.3	10.0	6.4
2	16.9	18.1	12.3	13.6	10.1	10.0
3	15.0	16.8	9.0	11.2	8.2	6.5
4	13.9	13.0	11.3	11.4	8.6	8.7

Exemple

```
> Prod=c(15.9,15.5,13.0,10.3,10.0,6.4,
      16.9,18.1,12.3,13.6,10.1,10.0,
      15.0,16.8,9.0,11.2,8.2,6.5,
      13.9,13.0,11.3,11.4,8.6,8.7)
> Mong=as.factor(rep(1:6,times=4))
> Mong
 [1] 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4
    5 6
Levels: 1 2 3 4 5 6
> Bloc=as.factor(rep(1:4,each=6))
> Bloc
 [1] 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4
    4 4
Levels: 1 2 3 4
```

Exemple

```
> Dades=data.frame(Prod,Mong,Bloc)
> str(Dades)
'data.frame': 24 obs. of  3 variables:
 $ Prod: num  15.9 15.5 13 10.3 10 6.4 16.9 18.1
    12.3 13.6 ...
 $ Mong: Factor w/ 6 levels "1","2","3","4",..:
   1 2 3 4 5 6 1 2 3 4 ...
 $ Bloc: Factor w/ 4 levels "1","2","3","4": 1 1
    1 1 1 1 2 2 2 2 ...
```

Exemple

```
> Xb.i=aggregate(Prod~Mong,data=Dades,mean)
> Xb.i
  Mong   Prod
1    1 15.425
2    2 15.850
3    3 11.400
4    4 11.625
5    5  9.225
6    6  7.900
> Xb.bj=aggregate(Prod~Bloc,data=Dades,mean)
> Xb.bj
  Bloc     Prod
1    1 11.85000
2    2 13.50000
3    3 11.11667
4    4 11.15000
> Xb=mean(Prod)
> Xb
[1] 11.90417
```

Exemple

- Mitjanes mostrals dels tractaments:

$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_4$	$\overline{X}_5$	$\overline{X}_6$
15.425	15.85	11.4	11.625	9.225	7.9

- Mitjanes mostrals dels blocs:

$\overline{X}_{\bullet 1}$	$\overline{X}_{\bullet 2}$	$\overline{X}_{\bullet 3}$	$\overline{X}_{\bullet 4}$
11.85	13.5	11.12	11.15

- Mitjana mostral global: $\overline{X} = 11.9$

Condicions necessàries

Per poder fer una ANOVA de blocs, cal que:

- Les $k \cdot b$ observacions constitueixen mostres aleatòries, cadascuna de mida 1, de les $k \cdot b$ variables X_{ij}
- Les variables X_{ij} són totes normals amb la mateixa variància σ^2
- L'efecte dels blocs i els tractaments és additiu: no hi ha interacció entre els blocs i els tractaments:

  Per a cada parell de nivells i_1, i_2 i per a cada parell de blocs j_1, j_2

  $$\mu_{i_1 j_1} - \mu_{i_2 j_1} = \mu_{i_1 j_2} - \mu_{i_2 j_2}$$

Cap d'elles no es pot contrastar, per tant l'experimentador decideix si se satisfan o no segons la seva experiència

Interacció?

Mesuram l'efecte d'un analgèsic A (en disminució del grau de dolor) en homes i dones

- μ_B: grau mitjà de dolor abans de prendre A
- μ_{BD}: grau mitjà de dolor de les dones abans de prendre A
- μ_{BH}: grau mitjà de dolor dels homes abans de prendre A
- μ_A: grau mitjà de dolor després de prendre A
- μ_{AD}: grau mitjà de dolor de les dones després de prendre A
- μ_{AH}: grau mitjà de dolor dels homes després de prendre A

Interacció?

Mesuram l'efecte d'un analgèsic A (en disminució del grau de dolor) en homes i dones

- μ_B: grau mitjà de dolor abans de prendre A
- μ_{BD}: grau mitjà de dolor de les dones abans de prendre A
- μ_{BH}: grau mitjà de dolor dels homes abans de prendre A
- μ_A: grau mitjà de dolor després de prendre A
- μ_{AD}: grau mitjà de dolor de les dones després de prendre A
- μ_{AH}: grau mitjà de dolor dels homes després de prendre A

- No interacció: $\mu_{BD} - \mu_{AD} = \mu_{BH} - \mu_{AH}$

Interacció?

Mesuram l'efecte d'un analgèsic A (en disminució del grau de dolor) en homes i dones

- μ_B: grau mitjà de dolor abans de prendre A
- μ_{BD}: grau mitjà de dolor de les dones abans de prendre A
- μ_{BH}: grau mitjà de dolor dels homes abans de prendre A
- μ_A: grau mitjà de dolor després de prendre A
- μ_{AD}: grau mitjà de dolor de les dones després de prendre A
- μ_{AH}: grau mitjà de dolor dels homes després de prendre A

- No interacció: $\mu_{BD} - \mu_{AD} = \mu_{BH} - \mu_{AH} = \mu_B - \mu_A$

Interacció?

Mesuram l'efecte d'un analgèsic A (en disminució del grau de dolor) en homes i dones

- μ_B: grau mitjà de dolor abans de prendre A
- μ_{BD}: grau mitjà de dolor de les dones abans de prendre A
- μ_{BH}: grau mitjà de dolor dels homes abans de prendre A
- μ_A: grau mitjà de dolor després de prendre A
- μ_{AD}: grau mitjà de dolor de les dones després de prendre A
- μ_{AH}: grau mitjà de dolor dels homes després de prendre A

- No interacció: $\mu_{BD} - \mu_{AD} = \mu_{BH} - \mu_{AH} = \mu_B - \mu_A$
- Sí interacció: $\mu_{BD} - \mu_{AD} \neq \mu_{BH} - \mu_{AH}$

Interacció?

Exercici: En un contrast obteniu un p-valor 0.0001. Amb $\alpha = 0.05$:

1. **(1 punt)** Acceptau o rebutjau H_0?
2. **(1 punt)** Quin tipus d'error pot ser que cometeu?

Interacció?

Exercici: En un contrast obteniu un p-valor 0.0001. Amb $\alpha = 0.05$:

1. **(1 punt)** Acceptau o rebutjau H_0? Rebutjau
2. **(1 punt)** Quin tipus d'error pot ser que cometeu? Tipus I

Interacció?

Exercici: En un contrast obteniu un p-valor 0.0001. Amb $\alpha = 0.05$:

1. **(1 punt)** Acceptau o rebutjau H_0? Rebutjau
2. **(1 punt)** Quin tipus d'error pot ser que cometeu? Tipus I

No interacció:

- Si teniu bé les dues, 2 punts
- Si en teniu una bé i una malament, 1 punt
- Si teniu malament les dues, 0 punts

Interacció?

Exercici: En un contrast obteniu un p-valor 0.0001. Amb $\alpha = 0.05$:

1. **(1 punt)** Acceptau o rebutjau H_0? Rebutjau
2. **(1 punt)** Quin tipus d'error pot ser que cometeu? Tipus I

Interacció

- Si teniu bé les dues, 2 punts
- Si teniu bé la 1a i malament la 2a, 1 punt
- Si en teniu bé la 2a i malament la 1a, 0 punts (respostes inconsistents)
- Si teniu malament les dues, 1 punt (respostes consistents)

Model

$$X_{ij} - \mu = (\mu_i - \mu) + (\mu_{\bullet j} - \mu) + E_{ij}, \;\; i = 1, \ldots, k, \;\; j = 1, \ldots, b$$

on:

- $\mu_i - \mu$: Efecte del tractament i-èsim
- $\mu_{\bullet j} - \mu$: Efecte del bloc j-èsim
- E_{ij} $(= X_{ij} - \mu_i - \mu_{\bullet j} + \mu)$: Residu, Error aleatori

Identitat de les sumes de quadrats

Teorema

$SS_{Total} = SS_{Tr} + SS_{Blocs} + SS_E$

- $SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij} - \overline{X})^2$, és la Suma de Total de Quadrats: variabilitat global de la mostra
- $SS_{Tr} = b\sum_{i=1}^{k}(\overline{X}_i - \overline{X})^2$, és la Suma de Quadrats dels Tractaments: variabilitat de les mitjanes dels tractaments
- $SS_{Blocs} = k\sum_{j=1}^{b}(\overline{X}_{\bullet j} - \overline{X})^2$, és la Suma de Quadrats dels Blocs: variabilitat de les mitjanes dels blocs
- $SS_E = \sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij} - \overline{X}_i - \overline{X}_{\bullet j} + \overline{X})^2$, és la Suma de Quadrats dels Residus o dels Errors: variabilitat deguda a factors aleatoris

Exemple

```
> k=6; b=4
> SS.Tot=sum((Prod-Xb)^2)
> SS.Tot
[1] 252.2496
> SS.Tr=b*sum((Xb.i[,2]-Xb)^2)
> SS.Tr
[1] 206.0371
> SS.Bl=k*sum((Xb.bj[,2]-Xb)^2)
> SS.Bl
[1] 22.43125
> SSE=sum((Prod-Xb.i[,2]-rep(Xb.bj[,2],each=k)+
   Xb)^2)
> SSE
[1] 23.78125
```

Identitat de les sumes de quadrats?

```
> SS.Tr+SS.Bl+SSE
[1] 252.2496
```

Contrast

- Quadrat mitjà dels tractaments:

$$MS_{Tr} = \frac{SS_{Tr}}{k-1}$$

- Quadrat mitjà dels errors:

$$MS_E = \frac{SS_E}{(b-1)(k-1)}$$

- (A més, R calcula) Quadrat mitjà dels blocs:

$$MS_{Blocs} = \frac{SS_{Blocs}}{b-1}$$

Contrast

Si se satisfan les condicions necessàries per fer una ANOVA de blocs:

$$E(MS_{Tr}) = \sigma^2 + \frac{b}{k-1}\sum_{i=1}^{k}(\mu_i - \mu)^2$$
$$E(MS_E) = \sigma^2$$

En particular, MS_E estima la variància comuna σ^2

Si $H_0 : \mu_1 = \cdots = \mu_k (= \mu)$ és certa,

$$\frac{b}{k-1}\sum_{i=1}^{k}(\mu_i - \mu)^2 = 0,$$

i si H_0 no és certa, aquesta quantitat és > 0

Contrast

Prenem com a estadístic de contrast

$$F = \frac{MS_{Tr}}{MS_E}$$

Si H_0 és certa:

- la seva distribució és $F_{k-1,(b-1)(k-1)}$ (F de Fisher-Snedecor amb $k-1$ i $(b-1)(k-1)$ graus de llibertat)
- el seu valor serà proper a 1

A més, si $k = 2$, F és igual al quadrat de l'estadístic del test t de 2 mostres aparellades

Rebutjarem la hipòtesi nul·la si F és molt gran:

$$\text{p-valor} = P(F_{k-1,(b-1)(k-1)} \geqslant F)$$

Contrast

1. Calculam
$$SS_{Tr},\ SS_E$$
2. Calculam
$$MS_{Tr}=\frac{SS_{Tr}}{k-1},\ MS_E=\frac{SS_E}{(b-1)(k-1)}$$
3. Calculam
$$F=\frac{MS_{Tr}}{MS_E}$$
4. Calculam el p-valor
$$P(F_{k-1,(b-1)(k-1)}\geqslant F)$$
5. Si el p-valor és més petit que el nivell de significació α, rebutjam H_0 i concloem que no totes les mitjanes són iguals. En cas contrari, acceptam H_0.

Exemple

k	b	SS_{Total}	SS_{Tr}	SS_{Blocs}	SS_E
6	4	252.24	206.04	22.43	23.78

- $MS_{Tr} = \frac{SS_{Tr}}{k-1} = \frac{206.04}{5} = 41.21$
- $MS_E = \frac{SS_E}{(b-1)(k-1)} = \frac{23.78}{3 \cdot 5} = 1.59$
- $MS_{Blocs} = \frac{SS_{Blocs}}{b-1} = \frac{23.78}{3} = 7.48$
- $F = \frac{MS_{Tr}}{MS_E} = 26$
- p-valor: $P(F_{5,15} \geqslant 26) = \texttt{1-pf(26,5,15)} = 7 \cdot 10^{-7}$
- Conclusió: Hem obtingut evidència estadística que les produccions mitjanes per als diferents tipus de mongetera no són totes iguals (ANOVA de blocs, p-valor)

Taula ANOVA

Una ANOVA de blocs es resumeix en una taula ANOVA:

Origen variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	Estadístic	p-valor
Tracts.	$k-1$	SS_{Tr}	MS_{Tr}	F	p-valor
Blocs	$b-1$	SS_{Blocs}	MS_{Blocs}		
Errors	$(b-1)(k-1)$	SS_E	MS_E		

Taula ANOVA

Una ANOVA de blocs es resumeix en una taula ANOVA:

Origen variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	Estadístic	p-valor
Tracts.	$k-1$	SS_{Tr}	MS_{Tr}	F	p-valor
Blocs	$b-1$	SS_{Blocs}	MS_{Blocs}		
Errors	$(b-1)(k-1)$	SS_E	MS_E		

Exemple:

Origen variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	Estadístic	p-valor
Tracts.	5	206.04	41.21	25.99	$7 \cdot 10^{-7}$
Blocs	3	22.43	7.48		
Errors	15	23.78	1.59		

Amb R

S'aplica summary(aov( )) a la fórmula que separa la variable numèrica per la suma (+) dels tractaments i els blocs

```
> summary(aov(Prod~Mong+Bloc, data=Dades))
            Df Sum Sq Mean Sq F value   Pr(>F)    
Mong         5 206.04   41.21  25.992 6.89e-07 ***
Bloc         3  22.43    7.48   4.716   0.0164 *  
Residuals   15  23.78    1.59                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 
    '.' 0.1 ' ' 1
```

El p-valor de la filera Bloc contrasta si hi ha diferències entre les mitjanes dels blocs.

Comparacions posteriors per parelles

Si rebutjam H_0, podem demanar-nos quins tractaments donen mitjanes diferents

Podem emprar un test t, fent cada comparació per a mostres aparellades i emprant un ajust del p-valor (Bonferroni, Holm, ...)

Amb R es fa amb `pairwise.t.test` indicant-hi que `paired=TRUE`

Exemple

```
> pairwise.t.test(Dades$Prod, Dades$Mong,
  paired=TRUE, p.adjust.method="bonferroni")

  Pairwise comparisons using paired t tests

data:  Dades$Prod and Dades$Mong

  1      2      3      4      5
2 1.0000 -      -      -      -
3 0.2208 0.7870 -      -      -
4 0.1542 0.2775 1.0000 -      -
5 0.0068 0.1052 0.3139 0.6583 -
6 0.0614 0.1310 0.6459 0.0432 1.0000

P value adjustment method: bonferroni
```

Només trobam evidència que $\mu_1 \neq \mu_5$ i $\mu_4 \neq \mu_6$

Contrast no paramètric

Si no podem aplicar ANOVA de blocs perquè sospitem que no se satisfan les condicions necessàries, cal emprar un test no paramètric

El més popular és el test de Friedman (generalitza el test de Wilcoxon amb mostres aparellades a més de 2 mostres), implementat a la funció `friedman.test` (cal substituir + per | a la fórmula)

```
> friedman.test(Prod~Mong|Bloc, data=Dades)

  Friedman rank sum test

data:  Prod and Mong and Bloc
Friedman chi-squared = 18.571, df = 5,
  p-value = 0.002309
```

ANOVA de 2 vies

ANOVA de 2 vies

Ens pot interessar comparar les mitjanes d'una variable sobre subpoblacions definides per més d'un factor a partir de mostres d'aquestes subpoblacions: se'n diu un experiment factorial

Aquí considerarem només el cas més senzill: el disseny d'ANOVA de 2 vies (completament aleatori):

- Empram els nivells (tractaments) de dos factors (2 vies) per classificar
- Prenem mostres aleatòries independents de la mateixa mida de cada combinació de nivells dels dos factors (completament aleatori)

Exemple

En un experiment per determinar l'atracció de moscards per colors i tipus de mel, s'han emprat 2 colors (vermell i verd) i 3 esquers (mel de taronger, mel de romaní i aigua), s'han oferit pots amb l'esquer tenyit del color a grups de moscards i s'ha mesurat el percentatge de moscards que s'han vist atrets pel pot. S'ha repetit 4 vegades per combinació (esquer, color) amb grups independents de moscards. Resultats:

	Esquer					
Color	Taronger		Romaní		Aigua	
Verd	65	42	67	73	35	37
	53	37	67	70	43	43
Vemell	57	38	60	42	35	33
	45	51	41	68	41	21

El color i el tipus de mel, afecten el percentatge mitjà de moscards atrets?

Exemple

Variable poblacional global:

- X: Prenc un esbart de moscards i mir quin percentatge és atret per un pot contenint un esquer colorejat

Subpoblacions: Definides per les combinacions de

- Els dos colors (G: Verd, R: Vermell)
- Els tres esquers (MT: Mel de taronger, MR: Mel de romaní, A: Aigua)

Exemple

Variables d'interès:

- X_G, X_R: Prenc un esbart de moscards i mir quin percentatge és atret per un pot contenint un esquer colorejat de verd (G) o de vermell (R)
- X_{MT}, X_{MR}, X_A: Prenc un esbart de moscards i mir quin percentatge és atret per un pot contenint mel de taronger (MT), mel de romaní (MR) o aigua (A)
- $X_{MT,G}, X_{MR,G}, X_{A,G}, X_{MT,R}, X_{MR,R}, X_{A,R}$: Prenc un esbart de moscards i mir quin percentatge és atret per un pot contenint l'esquer corresponent tenyit del color corresponent

Exemple

- Hi ha diferència segons el color?

$$\begin{cases} H_0 : \mu_G = \mu_R \\ H_1 : \mu_G \neq \mu_R \end{cases}$$

Exemple

- Hi ha diferència segons el color?

$$\begin{cases} H_0 : \mu_G = \mu_R \\ H_1 : \mu_G \neq \mu_R \end{cases}$$

- Hi ha diferència segons l'esquer?

$$\begin{cases} H_0 : \mu_{MT} = \mu_{MR} = \mu_A \\ H_1 : \text{ No és veritat que } \ldots \end{cases}$$

Exemple

- Hi ha diferència segons el color?

$$\begin{cases} H_0 : \mu_G = \mu_R \\ H_1 : \mu_G \neq \mu_R \end{cases}$$

- Hi ha diferència segons l'esquer?

$$\begin{cases} H_0 : \mu_{MT} = \mu_{MR} = \mu_A \\ H_1 : \text{ No és veritat que } \dots \end{cases}$$

- Hi ha diferència segons la combinació d'esquer i color?

$$\begin{cases} H_0 : \mu_{E,C} = \mu_{E'C'} \text{ per a tots esquers } E, E' \text{ i colors } C, C' \\ H_1 : \text{ No és veritat que } \dots \end{cases}$$

Exemple

- Hi ha diferència segons el color?

$$\begin{cases} H_0 : \mu_G = \mu_R \\ H_1 : \mu_G \neq \mu_R \end{cases}$$

- Hi ha diferència segons l'esquer?

$$\begin{cases} H_0 : \mu_{MT} = \mu_{MR} = \mu_A \\ H_1 : \text{ No és veritat que } \ldots \end{cases}$$

- Hi ha diferència segons la combinació d'esquer i color?

$$\begin{cases} H_0 : \mu_{E,C} = \mu_{E'C'} \text{ per a tots esquers } E, E' \text{ i colors } C, C' \\ H_1 : \text{ No és veritat que } \ldots \end{cases}$$

- Hi ha interacció entre colors i esquers?

$$\begin{cases} H_0 : \text{No hi ha interacció entre els esquers i els colors} \\ H_1 : \text{Hi ha interacció entre alguns esquers i alguns colors} \end{cases}$$

Situació general

Tenim una v.a. X definida sobre una població

Classificam la població en subpoblacions segons dos factors, A i B. El factor A té a nivells i el factor B, b nivells.

Al nostre exemple:

- Factor A: Esquer, $a = 3$
- Factor B: Color, $b = 2$

Situació general

- μ: mitjana poblacional de X global
- $X_{i\bullet}$: Prenc un individu del nivell i-èsim del factor A i hi mesur X, $i = 1, \ldots, a$; la seva mitjana és $\mu_{i\bullet}$
- $X_{\bullet j}$: Prenc un individu del nivell j-èsim del factor B i hi mesur X, $j = 1, \ldots, b$; la seva mitjana és $\mu_{\bullet j}$
- X_{ij}: Prenc un individu del nivell i-èsim del factor A i el nivell j-èsim del factor B i hi mesur X, de mitjana μ_{ij}

Situació general

Prenem mostres de mida n de cada X_{ij}

	Factor A			
Factor B	1	2	$\cdots$	a
1	X_{111}	X_{211}	$\cdots$	X_{a11}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{11n}	X_{21n}	$\cdots$	X_{a1n}
2	X_{121}	X_{221}	$\cdots$	X_{a21}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{12n}	X_{22n}	$\cdots$	X_{a2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	X_{1b1}	X_{2b1}	$\cdots$	X_{ab1}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{1bn}	X_{2bn}	$\cdots$	X_{abn}

- Les mostres de cada $X_{i\bullet}$ tenen mida bn
- Les mostres de cada $X_{\bullet j}$ tenen mida an
- El nombre total d'observacions és $n \cdot a \cdot b$.

Exemple

	Esquer		
Color	MT	MR	A
G	65	67	35
	42	73	37
	53	67	43
	37	70	43
R	57	60	35
	38	42	33
	45	41	41
	51	68	21

Situació general

- $\overline{X}_{i\bullet}$: mitjana mostral del nivell i-èsim d'A

$$\overline{X}_i = \frac{\sum_{j=1}^{b} X_{ij}}{b \cdot n}$$

- $\overline{X}_{\bullet j}$: mitjana mostral del nivell j-èsim de B

$$\overline{X}_{\bullet j} = \frac{\sum_{i=1}^{k} X_{ij}}{a \cdot n}$$

- $\overline{X}_{ij}$: mitjana mostral de la combinació del nivell i-èsim del factor A i el nivell j-èsim del factor B

$$\overline{X}_{ij} = \frac{\sum_{i=1}^{k} X_{ij}}{n}$$

- $\overline{X}$: mitjana mostral global

$$\overline{X} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij}}{a \cdot b \cdot n}$$

Exemple

	Esquer			
Color	Taronger	Romaní	Aigua	
Verd	65	67	35	
	42	73	37	
	53	67	43	
	37 $\overline{X}_{11}$	70 $\overline{X}_{21}$	43 $\overline{X}_{31}$	$\overline{X}_{\bullet 1}$
Vemell	57	60	35	
	38	42	33	
	45	41	41	
	51 $\overline{X}_{12}$	68 $\overline{X}_{22}$	21 $\overline{X}_{32}$	$\overline{X}_{\bullet 2}$
	$\overline{X}_{1\bullet}$	$\overline{X}_{2\bullet}$	$\overline{X}_{3\bullet}$	

$$\underbrace{\overline{X}_{1\bullet} \quad \overline{X}_{2\bullet} \quad \overline{X}_{3\bullet}}_{\overline{X}}$$

Exemple

Organitzarem les dades del nostre exemple en un dataframe amb 3 variables:

- `Percent`, quantitativa, el percentatge de moscards atrets
- `Esquer`, un factor que contendrà el valor del nivell del factor A (esquer) per a cada grup de moscards: MT, MR, A
- `Color`, un factor que contendrà el valor del nivell del factor B (color) per a cada grup de moscards: G, R

Exemple

	Esquer		
Color	MT	MR	A
G	65	67	35
	42	73	37
	53	67	43
	37	70	43
R	57	60	35
	38	42	33
	45	41	41
	51	68	21

Exemple

```
> Percent=c(65,67,35,42,73,37,
   53,67,43,37,70,43,57,60,35,
   38,42,33,45,41,41,51,68,21)
> Esquer=factor(rep(c("MT","MR","A"),times=8),
               levels=c("MT","MR","A"))
> Esquer
 [1] MT MR A  MT MR A  MT MR A  MT MR
[12] A  MT MR A  MT MR A  MT MR A  MT
[23] MR A
Levels: MT MR A
> Color=factor(rep(c("G","R"),each=12),
              levels=c("G","R"))
> Color
 [1] G G G G G G G G G G G G R R R R
[17] R R R R R R R R
Levels: G R
```

Exemple

```
> str(Moscards)
'data.frame': 24 obs. of  3 variables:
 $ Percent: num  65 67 35 42 73 37 53 67 43 37
   ...
 $ Esquer : Factor w/ 3 levels "MT","MR","A": 1
   2 3 1 2 3 1 2 3 1 ...
 $ Color  : Factor w/ 2 levels "G","R": 1 1 1 1
   1 1 1 1 1 1 ...
```

Exemple

```
> Xb.i.b=aggregate(Percent~Esquer,data=Moscards,
   mean)
> Xb.i.b
  Esquer Percent
  Esquer Percent
1     MT    48.5
2     MR    61.0
3      A    36.0
> Xb.b.j=aggregate(Percent~Color,data=Moscards,
   mean)
> Xb.b.j
  Color  Percent
1     G 52.66667
2     R 44.33333
> Xb=mean(Percent); Xb
[1] 48.5
```

$\overline{X}_{MT}$	$\overline{X}_{MR}$	$\overline{X}_{A}$	$\overline{X}_{G}$	$\overline{X}_{R}$	$\overline{X}$
48.5	61	36	52.67	44.33	48.5

Exemple

```
> Xb.i.j=aggregate(Percent~Esquer+Color,
   data=Moscards,mean)
> Xb.i.j
  Esquer Color Percent
1     MT     G   49.25
2     MR     G   69.25
3      A     G   39.50
4     MT     R   47.75
5     MR     R   52.75
6      A     R   32.50
```

$\overline{X}_{ij}$	MT	MR	A
G	49.25	69.25	39.50
R	47.75	52.75	32.50

Condicions necessàries

Per poder fer una ANOVA de 2 vies, cal que:

- Les observacions per a cada combinació de nivells (i, j) constitueixin m.a.s. independents de les variables X_{ij}, totes de la mateixa mida n
- Les variables X_{ij} siguin totes normals
- Homocedasticitat: Les variables X_{ij} tenguin totes la mateixa variància, σ^2

Model

$$X_{ij} = \mu + (\mu_{i\bullet} - \mu) + (\mu_{\bullet j} - \mu) + (\alpha\beta)_{ij} + E_{ij}$$

on

- $\mu_{i\bullet} - \mu$: Efecte del factor A
- $\mu_{\bullet j} - \mu$: Efecte del factor B
- $(\alpha\beta)_{ij}$ $(= \mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu)$: Efecte de la interacció entre el nivell i-èsim del factor A i el nivell j-èsim del factor B
- E_{ij} $(= X_{ij} - \mu_{ij})$: Residu, Error aleatori

Identitats de les sumes de quadrats

Teorema

$SS_{Total} = SS_{Tr} + SS_E$

$SS_{Tr} = SS_A + SS_B + SS_{AB}$

- $SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk} - \overline{X}_{\bullet\bullet})^2$: Suma de Total de Quadrats, representa la variabilitat global de la mostra
- $SS_{Tr} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{ij} - \overline{X}_{\bullet\bullet})^2$: Suma de Quadrats dels Tractaments, representa la variabilitat de les mitjanes de les combinacions de tractaments d'A i B
- $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk} - \overline{X}_{ij})^2$: Suma de Quadrats dels Errors, representa la variabilitat deguda a factors aleatoris

Identitats de les sumes de quadrats

Teorema

$SS_{Total} = SS_{Tr} + SS_E$

$SS_{Tr} = SS_A + SS_B + SS_{AB}$

- $SS_A = bn\sum_{i=1}^{a}(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2$: Suma de Quadrats del factor A, representa la variabilitat de les mitjanes dels nivells d'A
- $SS_B = an\sum_{j=1}^{b}(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2$: Suma de Quadrats del factor B, representa la variabilitat de les mitjanes dels nivells de B
- $SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{ij} - \overline{X}_{i\bullet} - \overline{X}_{\bullet j} + \overline{X}_{\bullet\bullet})^2$: Suma de Quadrats de la Interacció, representa la variabilitat deguda a la interacció dels nivells d'A i B

Identitats de les sumes de quadrats

$$SS_{Total} = SS_{Tr} + SS_E$$

$$SS_{Tr} = SS_A + SS_B + SS_{AB}$$

Per tant

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_E$$

Exemple

```
> n=4; a=3; b=2
> SS.Tot=sum((Percent-Xb)^2); SS.Tot
[1] 4636
> SS.A=n*b*sum((Xb.i.b[,2]-Xb)^2); SS.A
[1] 2500
> SS.B=n*a*sum((Xb.b.j[,2]-Xb)^2); SS.B
[1] 416.6667
> SS.Tr=n*sum((Xb.i.j[,3]-Xb)^2); SS.Tr
[1] 3147
> SS.AB=n*sum((Xb.i.j[,3]-Xb.i.b[,2]
   -rep(Xb.b.j[,2],each=a)+Xb)^2)
> SS.AB
[1] 230.3333
> SS.E=sum((Percent-c(rep(Xb.i.j[1:a,3],n),
   rep(Xb.i.j[(a+1):(2*a),3],n)))^2)
> SS.E
[1] 1489
```

Exemple

SS_{Total}	SS_A	SS_B	SS_{AB}	SS_{Tr}	SS_E
4636	2500	416.67	230.33	3147	1489

```
> SS.A+SS.B+SS.AB
[1] 3147
> SS.Tr+SS.E
[1] 4636
```

Quadrats mitjans

- Quadrat mitjà del factor A: $MS_A = \frac{SS_A}{a-1}$
- Quadrat mitjà del factor B: $MS_B = \frac{SS_B}{b-1}$
- Quadrat mitjà dels tractaments: $MS_{Tr} = \frac{SS_{Tr}}{ab-1}$
- Quadrat mitjà de la interacció: $MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$
- Quadrat mitjà dels errors: $MS_E = \frac{SS_E}{ab(n-1)}$

Exemple

$n = 4,\ a = 3,\ b = 2$

SS_{Total}	SS_A	SS_B	SS_{AB}	SS_{Tr}	SS_E
4636	2500	416.67	230.33	3147	1489

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
1250	416.67	115.17	629.4	82.72

```
> MS.B=SS.B/(b-1)
> MS.AB=SS.AB/((a-1)*(b-1))
> MS.Tr=SS.Tr/(a*b-1)
> MS.E=SS.E/(a*b*(n-1))
> round(c(MS.A,MS.B,MS.AB,MS.Tr,MS.E),2)
[1] 1250.00  416.67  115.17  629.40  82.72
```

Contrast de mitjanes del factor A

Contrastam si hi ha diferències entre les mitjanes dels nivells del factor A:

$$\left\{\begin{array}{l} H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \cdots = \mu_{a\bullet} \\ H_1 : \text{Hi ha } i, i' \text{ tals que } \mu_{i\bullet} \neq \mu_{i'\bullet} \end{array}\right.$$

L'estadístic de contrast és

$$F_A = \frac{MS_A}{MS_E},$$

Si H_0 és certa, té distribució F de Fisher amb $a-1$ i $ab(n-1)$ graus de llibertat i valor proper a 1

Contrast de mitjanes del factor B

Contrastam si hi ha diferències entre les mitjanes dels nivells del factor B:

$$\left\{\begin{array}{l} H_0 : \mu_{\bullet 1} = \mu_{\bullet 2} = \cdots = \mu_{\bullet b} \\ H_1 : \text{Hi ha } j, j' \text{ tals que } \mu_{\bullet j} \neq \mu_{\bullet j'} \end{array}\right.$$

L'estadístic de contrast és

$$F_B = \frac{MS_B}{MS_E},$$

Si H_0 és certa, té distribució F de Fisher amb $b-1$ i $ab(n-1)$ graus de llibertat i valor proper a 1

Contrast dels tractaments

Contrastam si hi ha diferències entre les mitjanes de les parelles (nivell de A, nivell de B):

$$\left\{\begin{array}{l} H_0 : \text{Per a tots } i, j, i', j',\ \mu_{ij} = \mu_{i'j'} \\ H_1 : \text{Hi ha } i, j, i', j' \text{ tals que } \mu_{ij} \neq \mu_{i'j'} \end{array}\right.$$

L'estadístic de contrast és

$$F_{Tr} = \frac{MS_{Tr}}{MS_E},$$

Si H_0 és certa, té distribució F de Fisher amb $ab-1$ i $ab(n-1)$ graus de llibertat i valor proper a 1

Contrast de no interacció

Contrastam si hi ha interacció entre els factors A i B:

$$\begin{cases} H_0 : \text{Per a tots } i, j, (\alpha\beta)_{ij} = 0 \\ H_1 : \text{Hi ha } i, j \text{ tals que } (\alpha\beta)_{ij} \neq 0 \end{cases}$$

L'estadístic de contrast és

$$F_{AB} = \frac{MS_{AB}}{MS_E},$$

Si H_0 és certa, té distribució F de Fisher amb $(a-1)(b-1)$ i $ab(n-1)$ graus de llibertat i valor proper a 1

Contrastos

En els quatre casos, el p-valor és

$$P(F_{x,y} \geqslant \text{valor de l'estadístic})$$

on $F_{x,y}$ representa la distribució F de Fisher amb els graus de llibertat que pertoquin.

Taula ANOVA

Els contrastos anteriors es resumeixen en la taula ANOVA:

Font de variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	F	p-
Tract.	$ab-1$	SS_{Tr}	MS_{Tr}	F_{Tr}	p-valor
A	$a-1$	SS_A	MS_A	F_A	p-valor
B	$b-1$	SS_B	MS_B	F_B	p-valor
AB	$(a-1)(b-1)$	SS_{AB}	MS_{AB}	F_{AB}	p-valor
Error	$ab(n-1)$	SS_E	MS_E		

Exemple

$n = 4,\ a = 3,\ b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
1250	416.67	115.17	629.4	82.72

$$\frac{MS_A}{MS_E} = \frac{1250}{82.72} = 15.11$$

p-valor$= P(F_{2,18} \geqslant 15.11) = \texttt{1-pf(15.11,2,18)} = 10^{-4}$

Hem trobat evidència estadística que els percentatges mitjans d'esbarts moscards que són atrets pels diferents tipus d'esquer no són tots iguals (ANOVA de 2 vies, p-valor 10^{-4})

Exemple

$n = 4,\ a = 3,\ b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
1250	416.67	115.17	629.4	82.72

$$\frac{MS_B}{MS_E} = \frac{416.67}{82.72} = 5.04$$

p-valor$= P(F_{1,18} \geqslant 5.04) = \texttt{1-pf(5.04,1,18)} = 0.038$

Hem trobat evidència estadística que els percentatges mitjans d'esbarts moscards que són atrets pels colors verd i vemell són diferents (ANOVA de 2 vies, p-valor 0.038)

Exemple

$n = 4,\ a = 3,\ b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
1250	416.67	115.17	629.4	82.72

$$\frac{MS_{Tr}}{MS_E} = \frac{629.4}{82.72} = 7.61$$

p-valor$= P(F_{5,18} \geqslant 7.61) = \texttt{1-pf(7.61,5,18)} = 5 \cdot 10^{-4}$

Hem trobat evidència estadística que els percentatges mitjans d'esbarts moscards que són atrets per les diferents combinacions d'esquer i color no són tots iguals (ANOVA de 2 vies, p-valor $5 \cdot 10^{-4}$)

Exemple

$n = 4,\ a = 3,\ b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
1250	416.67	115.17	629.4	82.72

$$\frac{MS_{AB}}{MS_E} = \frac{115.17}{82.72} = 1.39$$

p-valor$= P(F_{2,18} \geqslant 1.39) = \texttt{1-pf(1.39,2,18)} = 0.274$

No hem trobat evidència estadística d'interacció entre els tipus d'esquer i els colors (ANOVA de 2 vies, p-valor 0.274)

Exemple

Font de variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	F	p-
Combs.	5	3147	629.4	7.61	$5 \cdot 10^{-4}$
Esquer	2	2500	1250	15.11	10^{-4}
Color	1	416.67	416.67	5.04	0.038
Interacció	2	230.33	115.17	1.39	0.274
Error	30	1489	82.72		

Amb R

```
> summary(aov(Percent~Esquer*Color,
  data=Moscards))
             Df Sum Sq Mean Sq F value    Pr(>F)
Esquer        2 2500.0  1250.0  15.111 0.000141
   ***
Color         1  416.7   416.7   5.037 0.037622
   *
Esquer:Color  2  230.3   115.2   1.392 0.274038
Residuals    18 1489.0    82.7
```

Amb R

```
> summary(aov(Percent~Esquer*Color,
  data=Moscards))
             Df Sum Sq Mean Sq F value   Pr(>F)
Esquer        2 2500.0  1250.0  15.111 0.000141
   ***
Color         1  416.7   416.7   5.037 0.037622
   *
Esquer:Color  2  230.3   115.2   1.392 0.274038
Residuals    18 1489.0    82.7
```

```
> summary(aov(Percent~Esquer:Color,
  data=Moscards))
             Df Sum Sq Mean Sq F value   Pr(>F)
Esquer:Color  5   3147   629.4   7.609 0.000538
   ***
Residuals    18   1489    82.7
```

Podíem emprar una ANOVA?

1. La variable d'interès sobre cada combinació de nivells és normal?

```
> es.normal=function(x){shapiro.test(x)$p.value}
> aggregate(Percent~Esquer+Color, data=Moscards,
    es.normal)
  Esquer Color   Percent
1     MT     G 0.7506899
2     MR     G 0.2724532
3      A     G 0.1612393
4     MT     R 0.9772990
5     MR     R 0.2674892
6      A     R 0.6499652
```

Podíem emprar una ANOVA?

2. I totes amb la mateixa variància?

```
> bartlett.test(Percent~interaction(Esquer,Color
    ),data=Moscards)

  Bartlett test of homogeneity of variances

data:  Percent by interaction(Esquer, Color)
Bartlett's K-squared = 7.6201, df = 5,
p-value = 0.1785
```

Alerta

Les condicions sobre les combinacions de nivells no impliquen les condicions sobre cada nivell per separat

```
> set.seed(100)
> mostra.X.1=rnorm(10,0,1)
> mostra.X.2=rnorm(10,5,1)
> mostra.Y.1=rnorm(10,0,1)
> mostra.Y.2=rnorm(10,10,1)
> x=c(mostra.X.1,mostra.X.2)
> y=c(mostra.Y.1,mostra.Y.2)
> shapiro.test(x)$p.value
[1] 0.003534705
> shapiro.test(y)$p.value
[1] 0.0001874323
> fligner.test(list(x,y))$p.value
[1] 1.245335e-06
```

Comparacions posteriors per parelles

Les comparacions posteriors per parelles de les mitjanes de les combinacions de nivells les podeu fer amb `pairwise.t.test`

```
> pairwise.t.test(Moscards$Percent,
  Moscards$Esquer:Moscards$Color, paired=FALSE)

...
     MT:G    MT:R    MR:G    MR:R    A:G
MT:R 1.00000 -       -       -       -
MR:G 0.06665 0.04706 -       -       -
MR:R 1.00000 1.00000 0.17930 -       -
A:G  0.88126 1.00000 0.00294 0.37892 -
A:R  0.17930 0.23273 0.00031 0.06665 1.00000

P value adjustment method: holm
```

Comparacions posteriors per parelles

Les comparacions posteriors per parelles de les mitjanes de les combinacions de nivells les podeu fer amb `pairwise.t.test`

```
> pairwise.t.test(Moscards$Percent,
  Moscards$Esquer:Moscards$Color, paired=FALSE)


...
     MT:G    MT:R    MR:G    MR:R    A:G
MT:R 1.00000 -       -       -       -
MR:G 0.06665 0.04706 -       -       -
MR:R 1.00000 1.00000 0.17930 -       -
A:G  0.88126 1.00000 0.00294 0.37892 -
A:R  0.17930 0.23273 0.00031 0.06665 1.00000

P value adjustment method: holm
```

La resta, ho deixarem córrer