ANOVA

Blocs complets aleatoris

Generalitza el contrast de mitjanes per a dues mostres aparellades, a k mostres aparellades. L'objectiu és reduir l'efecte de variables estranyes, no contemplades en l'estudi.

Suposem que tenim k tractaments que volem comparar

Escollim blocs: conjunts de k individus relacionats (per exemple, k còpies del mateix individu)

Dins cada bloc, assignam aleatòriament a cada individu un tractament

Blocs complets aleatoris

En els experiments de Blocs complets aleatoris:

- S'han aparellat els individus en blocs (blocs)
- Els tractaments s'assignen de manera aleatòria dins els blocs (aleatoris)
- Cada tractament s'empra exactament una vegada dins cada bloc (complets)
- Pel que fa als tractaments, és d'efectes fixats (la inferència serà vàlida només per als tractaments emprats)
- Pel que fa als blocs, pot ser d'efectes fixats (es trien tots els blocs adients) o aleatori: en aquest cas el model és mixt

Blocs complets aleatoris

Les dades es presenten en una taula:

	Tractaments			
Bloc	Tractament 1	Tractament 2	...	Tractament 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}: valor del tractament i-èsim en l'individu corresponent del bloc j-èsim

ALERTA! A X_{ij}, i hi indica la columna, tractament, i j la filera, bloc

Blocs complets aleatoris

El contrast que es vol realitzar és

$$\left.\begin{array}{l} H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \cdots = \mu_{k\bullet} \\ H_1 : \exists i, j \mid \mu_{i\bullet} \neq \mu_{j\bullet} \end{array}\right\}$$

on cada $\mu_{i\bullet}$ representa la mitjana del tractament i-èsim

Exemple 1

Volem determinar si l'energia que es requereix per dur a terme tres activitats físiques (córrer, passejar i muntar amb bicicleta) és la mateixa o no. Per quantificar aquesta energia, mesuram el nombre de Kca consumides per Km recorregut

Les diferències metabòliques entre els individus poden afectar l'energia requerida per dur a terme una determinada activitat

Per tant, no és aconsellable triar tres grups d'individus i a cada un fer-li fer una de les tres activitats físiques: les diferències metabòliques entre els individus triats podrien afectar els resultats i donar massa variació

Exemple 1

El que fem és seleccionar alguns individus (els blocs), demanar a cadascun que corri, camini i recorri amb bicicleta una distància fixada, i determinar per a cada individu el nombre de Kca consumides per Km durant cada activitat

Cada individu és utilitzat com un bloc. Les activitats es realitzen en ordre aleatori, amb temps de recuperació entre l'una i l'altra.

En aparellar cada individu amb ell mateix, eliminam l'efecte de la variació individual

Disseny de blocs complets aleatoris mixt

Exemple 1

En la taula següent es mostren els resultats obtinguts per a 8 individus:

	Tractament		
Bloc	1 (corrent)	2 (caminant)	3 (pedalejant)
1	1.4	1.1	0.7
2	1.5	1.2	0.8
3	1.8	1.3	0.7
4	1.7	1.3	0.8
5	1.6	0.7	0.1
6	1.5	1.2	0.7
7	1.7	1.1	0.4
8	2.0	1.3	0.6

Exemple 1

El contrast que volem realitzar és

$$\left.\begin{array}{l} H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \mu_{3\bullet} \\ H_1 : \exists i, j \mid \mu_{i\bullet} \neq \mu_{j\bullet} \end{array}\right\}$$

on $\mu_{i\bullet}$, $i = 1, 2, 3$ representa el nombre mitjà de Kca consumides per Km mentre es corre, es passeja o es munta amb bicicleta, respectivament

Model

Expressió matemàtica del model considerat:

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

on:

- X_{ij}: valor del tractament i-èsim en el bloc j-èsim
- μ: mitjana global
- $\mu_{i\bullet}$: mitjana del tractament i-èsim
- $\mu_{\bullet j}$: mitjana del bloc j-èsim
- $\mu_{i\bullet} - \mu$: efecte del tractament i-èsim (efecte tractament)
- $\mu_{\bullet j} - \mu$: efecte de pertànyer al bloc j-èsim (efecte bloc)
- E_{ij}: error residual o aleatori

El model

Les suposicions del model són:

- Les $k \cdot b$ observacions constitueixen mostres aleatòries independents, cadascuna de mida 1, de $k \cdot b$ poblacions
- Aquestes $k \cdot b$ poblacions són totes normals i 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:
  - La diferència de les mitjanes poblacionals de cada parella concreta de blocs és la mateixa a cada tractament
  - La diferència de les mitjanes poblacionals de cada parella concreta de tractaments és la mateixa a cada bloc

No interacció

Mesuram tres variables en homes i dones, i tenim les mitjanes poblacionals de cada variable dins cada grup

No interacció:

	Tractament		
Bloc	A	B	C
Homes	$\mu_{11} = 4$	$\mu_{21} = 5$	$\mu_{31} = 7$
Dones	$\mu_{12} = 3$	$\mu_{22} = 4$	$\mu_{32} = 6$

Sí interacció:

	Tractament		
Bloc	A	B	C
Homes	$\mu_{11} = 4$	$\mu_{21} = 5$	$\mu_{31} = 7$
Dones	$\mu_{12} = 3$	$\mu_{22} = 4$	$\mu_{32} = 2$

Estadístics

- $T_{i\bullet} = \sum_{j=1}^{b} X_{ij}$, suma total del tractament i-èsim, $i = 1, 2, \ldots, k$
- $\overline{X}_{i\bullet} = \dfrac{T_{i\bullet}}{b}$, mitjana mostral del tractament i-èsim, $i = 1, 2, \ldots, k$
- $T_{\bullet j} = \sum_{i=1}^{k} X_{ij}$, suma total del bloc j-èsim, $j = 1, 2, \ldots, b$
- $\overline{X}_{\bullet j} = \dfrac{T_{\bullet j}}{k}$, mitjana mostral del bloc j-èsim, $j = 1, 2, \ldots, b$

Estadístics

- $T_{\bullet\bullet} = \sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij} = \sum_{i=1}^{k} T_{i\bullet} = \sum_{j=1}^{b} T_{\bullet j}$, suma total
- $\overline{X}_{\bullet\bullet} = \frac{T_{\bullet\bullet}}{k \cdot b}$, mitjana mostral global
- $T_{\bullet\bullet}^{(2)} = \sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij}^2$, suma total de quadrats

Exemple 1

Les dades:

	Tractament			
Bloc	1	2	3	
1	1.4	1.1	0.7	$T_{\bullet 1}$
2	1.5	1.2	0.8	$T_{\bullet 2}$
3	1.8	1.3	0.7	$T_{\bullet 3}$
4	1.7	1.3	0.8	$T_{\bullet 4}$
5	1.6	0.7	0.1	$T_{\bullet 5}$
6	1.5	1.2	0.7	$T_{\bullet 6}$
7	1.7	1.1	0.4	$T_{\bullet 7}$
8	2.0	1.3	0.6	$T_{\bullet 8}$
	$T_{1\bullet}$	$T_{2\bullet}$	$T_{3\bullet}$	$T_{\bullet\bullet}$

Exemple 1

Emmagatzemam les dades en una matriu `kilocal`:

```
> kilocal = matrix(c(1.4,1.1,0.7,1.5,1.2,0.8,
 1.8,1.3,0.7,1.7,1.3,0.8,1.6,0.7,0.1,1.5,1.2,
 0.7,1.7,1.1,0.4,2.0,1.3,0.6),8,3,byrow=T)
> kilocal
[,1] [,2] [,3]
[1,] 1.4 1.1 0.7
[2,] 1.5 1.2 0.8
[3,] 1.8 1.3 0.7
[4,] 1.7 1.3 0.8
[5,] 1.6 0.7 0.1
[6,] 1.5 1.2 0.7
[7,] 1.7 1.1 0.4
[8,] 2.0 1.3 0.6
```

Exemple 1

Calculem estadístics

```
> sum.tracts=colSums(kilocal)
> sum.tracts
[1] 13.2  9.2  4.8
> mean.tracts=colMeans(kilocal)
> mean.tracts
[1] 1.65 1.15 0.60
> sum.blocs=rowSums(kilocal)
> sum.blocs
[1] 3.2 3.5 3.8 3.8 2.4 3.4 3.2 3.9
> mean.blocs=rowMeans(kilocal)
> mean.blocs
[1] 1.066667 1.166667 1.266667 1.266667
[5] 0.800000 1.133333 1.066667 1.300000
```

Exemple 1

Calculem estadístics

```
> sum.tot=sum(kilocal)
> sum.tot
[1] 27.2
> mean.tot=mean(kilocal)
> mean.tot
[1] 1.133333
> sum.sq.tot=sum(kilocal^2)
> sum.sq.tot
[1] 36.18
```

Exemple 1

- Sumes totals i mitjanes mostrals dels tractaments:

$T_{1\bullet}$	$T_{2\bullet}$	$T_{3\bullet}$
13.2	9.2	4.8

$\overline{X}_{1\bullet}$	$\overline{X}_{2\bullet}$	$\overline{X}_{3\bullet}$
1.65	1.15	0.6

- Sumes totals dels blocs:

$T_{\bullet 1}$	$T_{\bullet 2}$	$T_{\bullet 3}$	$T_{\bullet 4}$	$T_{\bullet 5}$	$T_{\bullet 6}$	$T_{\bullet 7}$	$T_{\bullet 8}$
3.2	3.5	3.8	3.8	2.4	3.4	3.2	3.9

- Mitjanes mostrals dels blocs:

$\overline{X}_{\bullet 1}$	$\overline{X}_{\bullet 2}$	$\overline{X}_{\bullet 3}$	$\overline{X}_{\bullet 4}$	$\overline{X}_{\bullet 5}$	$\overline{X}_{\bullet 6}$	$\overline{X}_{\bullet 7}$	$\overline{X}_{\bullet 8}$
1.067	1.167	1.267	1.267	0.8	1.133	1.067	1.3

- Suma total: $T_{\bullet\bullet} = 27.2$
- Mitjana mostral: $\overline{X}_{\bullet\bullet} = 1.133$
- Suma de quadrats: $T^{(2)}_{\bullet\bullet} = 36.18$

Identitat de la suma de quadrats

Teorema

$$\begin{aligned}\sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij}-\overline{X}_{\bullet\bullet})^2 \quad &= b\sum_{i=1}^{k}(\overline{X}_{i\bullet}-\overline{X}_{\bullet\bullet})^2 \\ &\quad + k\sum_{j=1}^{b}(\overline{X}_{\bullet j}-\overline{X}_{\bullet\bullet})^2 \\ &\quad + \sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij}-\overline{X}_{i\bullet}-\overline{X}_{\bullet j}+\overline{X}_{\bullet\bullet})^2\end{aligned}$$

Identitat de la suma de quadrats

- $SS_{Total} = \sum_{i=1}^{k} \sum_{j=1}^{b} (X_{ij} - \overline{X}_{\bullet\bullet})^2$, variabilitat total
- $SS_{Tr} = b \sum_{i=1}^{k} (\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2$, variabilitat deguda als tractaments
- $SS_{Blocs} = k \sum_{j=1}^{b} (\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2$, variabilitat deguda als blocs
- $SS_{E} = \sum_{i=1}^{k} \sum_{j=1}^{b} (X_{ij} - \overline{X}_{i\bullet} - \overline{X}_{\bullet j} + \overline{X}_{\bullet\bullet})^2$, variabilitat deguda a factors aleatoris

Identitat de la suma de quadrats

Teorema

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

- $SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{\bullet\bullet})^2$
- $SS_{Tr} = b\sum_{i=1}^{k}(X_{i\bullet} - \overline{X}_{\bullet\bullet})^2$
- $SS_{Blocs} = k\sum_{j=1}^{b}(X_{\bullet j} - \overline{X}_{\bullet\bullet})^2$
- $SS_E = \sum_{i=1}^{k}\sum_{j=1}^{b}(X_{ij} - X_{i\bullet} - X_{\bullet j} + \overline{X}_{\bullet\bullet})^2$

Identitat de la suma de quadrats

Voldrem calcular SS_{Tr}, SS_{Blocs}, SS_E:

- $SS_{Total} = T_{\bullet\bullet}^{(2)} - \dfrac{T_{\bullet\bullet}^2}{k \cdot b}$
- $SS_{Tr} = \displaystyle\sum_{i=1}^{k} \frac{T_{i\bullet}^2}{b} - \frac{T_{\bullet\bullet}^2}{k \cdot b}$
- $SS_{Blocs} = \displaystyle\sum_{j=1}^{b} \frac{T_{\bullet j}^2}{k} - \frac{T_{\bullet\bullet}^2}{k \cdot b}$
- $SS_E = SS_{Total} - SS_{Tr} - SS_{Blocs}$

Exemple 1

$$k = 3,\ b = 8$$

$T_{1\bullet}$	$T_{2\bullet}$	$T_{3\bullet}$
13.2	9.2	4.8

$T_{\bullet 1}$	$T_{\bullet 2}$	$T_{\bullet 3}$	$T_{\bullet 4}$	$T_{\bullet 5}$	$T_{\bullet 6}$	$T_{\bullet 7}$	$T_{\bullet 8}$
3.2	3.5	3.8	3.8	2.4	3.4	3.2	3.9

$$T_{\bullet\bullet} = 27.2, \quad T_{\bullet\bullet}^{(2)} = 36.18$$

$$SS_{Total} = \qquad\qquad SS_{Tr} =$$

$$SS_{Blocs} = \qquad\qquad SS_{E} =$$

Exemple 1

- Variabilitat total:

```
> SS.Tot=sum((kilocal-mean.tot)^2)
> SS.Tot
[1] 5.353333
> sum.sq.tot-sum.tot^2/24
[1] 5.353333
```

- Variabilitat deguda als tractaments:

```
> SS.Tr=8*sum((mean.tracts-mean.tot)^2)
> SS.Tr
[1] 4.413333
> sum(sum.tracts^2)/8-sum.tot^2/24
[1] 4.413333
```

Exemple 1

- Variabilitat deguda als blocs:

```
> SS.Bl=3*sum((mean.blocs-mean.tot)^2)
> SS.Bl
[1] 0.5533333
> sum(sum.blocs^2)/3-sum.tot^2/24
[1] 0.5533333
```

- Variabilitat deguda a factors aleatoris

```
> SSE=SS.Tot-SS.Tr-SS.Bl
> SSE
[1] 0.3866667
```

Contrast

Per realitzar el contrast

$$\left.\begin{array}{l} H_0 : \mu_{1\bullet} = \cdots = \mu_{k\bullet} \\ H_1 : \exists i, j = 1, \ldots, k \mid \mu_{i\bullet} \neq \mu_{j\bullet} \end{array}\right\}$$

emprarem els estadístics següents:

- Quadrat mitjà dels tractaments: $MS_{Tr} = \dfrac{SS_{Tr}}{k-1}$
- Quadrat mitjà de l'error: $MS_E = \dfrac{SS_E}{(b-1)(k-1)}$
- (A més) Quadrat mitjà dels blocs: $MS_{Blocs} = \dfrac{SS_{Blocs}}{b-1}$

Contrast

Es té que

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

Si $H_0 : \mu_{1\bullet} = \cdots = \mu_{k\bullet} = \mu$ és certa,

$$\sum_{i=1}^{k}(\mu_{i\bullet} - \mu)^2 = 0,$$

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

Estadístics del contrast

Considerarem com a estadístic de contrast el quocient

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

Si H_0 és certa:

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

i rebutjarem la hipòtesi nul·la si F és prou més gran que 1

Contrast

- Calculam el quocient

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

- Calculam el p-valor

$$P(F_{k-1,(k-1)(b-1)} \geqslant F)$$

- 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 1

k	b	SS_{Total}	SS_{Tr}	SS_{Blocs}	SS_E
3	8	5.3533	4.4133	0.5533	0.3867

- $MS_{Tr} = \dfrac{SS_{Tr}}{k-1} = 2.2066$
- $MS_{Blocs} = \dfrac{SS_{Blocs}}{b-1} = 0.079$
- $MS_E = \dfrac{SS_E}{(b-1)(k-1)} = 0.0276$
- $F = \dfrac{MS_{Tr}}{MS_E} = 79.949$
- p-valor: $P(F_{2,14} \geqslant 79.949) = 2 \cdot 10^{-8}$
- Conclusió: Rebutjam la hipòtesi nul·la i concloem que les tres mitjanes no són iguals

Taula del contrast

Per realitzar el contrast es construeix la taula següent:

Font de variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	Estadístic	p-valor
Tract.	$k-1$	SS_{Tr}	MS_{Tr}	$F = \frac{MS_{Tr}}{MS_E}$	p-valor
Bloc	$b-1$	SS_{Blocs}	MS_{Blocs}		
Error	$(k-1)\cdot(b-1)$	SS_E	MS_E		

Exemple 1

Font de variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	Estadístic	p-valor
Tract.	2	4.4133	2.2067	79.949	$2 \cdot 10^{-8}$
Bloc	7	0.5533	0.079		
Error	14	0.3867	0.0276		

Exemple 1 amb R

Primer hem de construir un dataframe

```
> kilocal2=c(1.4,1.1,0.7,1.5,1.2,0.8,
  1.8,1.3,0.7,1.7,1.3,0.8,1.6,0.7,0.1,1.5,1.2,
  0.7,1.7,1.1,0.4,2.0,1.3,0.6)
> blocs=as.factor(rep(1:8,each=3))
> blocs
 [1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6 7 7 7
[22] 8 8 8
Levels: 1 2 3 4 5 6 7 8
> tracts=as.factor(rep(1:3,times=8))
> tracts
 [1] 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
[22] 1 2 3
Levels: 1 2 3
```

Exemple 1 amb R

```
> kilocal.df=data.frame(kilocal2,tracts,blocs)
> str(kilocal.df)
'data.frame': 24 obs. of  3 variables:
 $ kilocal2: num  1.4 1.1 0.7 1.5 1.2 0.8 1.8
     1.3 0.7 1.7 ...
 $ tracts  : Factor w/ 3 levels "1","2","3": 1
     2 3 1 2 3 1 2 3 1 ...
 $ blocs   : Factor w/ 8 levels "1","2","3",
    "4",..: 1 1 1 2 2 2 3 3 3 4 ...
> head(kilocal.df,4)
  kilocal2 tracts blocs
1      1.4      1     1
2      1.1      2     1
3      0.7      3     1
4      1.5      1     2
```

Exemple 1 amb R

Aplicam aov "sumant els dos factors"

```
> summary(aov(kilocal2~tracts+blocs,
    data=kilocal.df))
            Df Sum Sq Mean Sq F value  Pr(>F)
tracts       2  4.413  2.2067  79.897 2.2e-08 ***
blocs        7  0.553  0.0790   2.862  0.0446 *
Residuals   14  0.387  0.0276
```

Efectivitat en la construcció dels blocs

Ha tingut èxit la construcció de blocs en el nostre experiment?

En cas afirmatiu, SS_{Blocs} explicaria una part important de SS_{Total} i faria reduir SS_E. Això augmentaria el valor de F i faria més difícil acceptar H_0, la qual cosa milloraria la potència del contrast.

Efectivitat en la construcció dels blocs

- En un ANOVA completament aleatori (CA) amb k tractaments, els graus de llibertat de l'estadístic F són $k-1$ i $N-k$ (N nombre total d'observacions)
- En un ANOVA de blocs complet aleatori (BCA) amb k tractaments i b blocs, els graus de llibertat de l'estadístic F són $k-1$ i $(k-1)(b-1) = N-k-(b-1)$ (on $N = b \cdot k$)

Quan el 2^{on} grau de llibertat de F decreix, el 0.95-quantil de F augmenta (mirau la taula), és més difícil rebutjar H_0, i la potència del contrast disminueix.

Efectivitat en la construcció dels blocs

L'efectivitat en la construcció dels blocs s'estima amb l'eficiència relativa, RE. S'interpreta com la relació entre el nombre d'observacions d'un experiment completament aleatori (CA) i el nombre d'observacions d'un experiment de blocs complet aleatori (BCA) necessària per obtenir tests equivalents.

- $RE = 3$ significa que el disseny CA requereix tres vegades tantes observacions com el disseny de BCA. En aquest cas, ha valgut la pena l'ús de blocs.
- $RE = 0.5$ significa que amb un disseny CA hagués bastat la meitat d'observacions que al disseny BCA. En aquest cas, no era aconsellable l'ús de blocs.

Efectivitat en la construcció dels blocs

RE s'estima amb

$$\widehat{RE} = c + (1-c)\frac{MS_{Blocs}}{MS_E},$$

on $c = \dfrac{b(k-1)}{(bk-1)}$

Per conveni, si $\widehat{RE} > 1.25$, s'entén que la construcció dels blocs ha estat profitosa

Exemple

A l'Exemple 1

$$c = \frac{8 \cdot 2}{24 - 1} = \frac{16}{23}$$

i

$$\widehat{RE} = \frac{16}{23} + \left(1 - \frac{16}{23}\right) \cdot \frac{0.079}{0.0276} = 1.5668$$

ANOVA de dues vies

Les comparacions de mitjanes les podem portar a terme classificant la població mitjançant més d'un factor: s'en diu un experiment factorial

Aquí considerarem només el cas més senzill: ANOVA de dues vies, disseny completament aleatori amb efectes fixats:

- Emprarem dos factors (dues vies)
- Emprarem tots els nivells de cada factor (efectes fixats)
- Prendrem mostres aleatòries independents de la mateixa mida de cada combinació de nivells dels dos factors (completament aleatori)

Format de les dades

Tenim en compte dos factors, A i B. El factor A té a nivells i el factor B, b nivells.

Fem n observacions per a cada combinació de tractaments. El nombre total d'observacions serà $n \cdot a \cdot b$.

La variable aleatòria X_{ijk}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, n$, ens dóna la resposta de la k-èsima unitat experimental al nivell i-èsim del factor A i el nivell j-èsim del factor B

Format de les dades

Donarem les dades en una taula d'aquest estil:

	Factor A			
Factor B	1	2	$\cdots$	a
1	X_{111}	X_{211}	$\cdots$	X_{a11}
	X_{112}	X_{212}	$\cdots$	X_{a12}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{11n}	X_{21n}	$\cdots$	X_{a1n}
2	X_{121}	X_{221}	$\cdots$	X_{a21}
	X_{122}	X_{222}	$\cdots$	X_{a22}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{12n}	X_{22n}	$\cdots$	X_{a2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	X_{1b1}	X_{2b1}	$\cdots$	X_{ab1}
	X_{1b2}	X_{2b2}	$\cdots$	X_{ab2}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	X_{1bn}	X_{2bn}	$\cdots$	X_{abn}

Exemple 2

En un experiment per determinar l'efecte de la llum i la temperatura sobre l'índex gonadosomàtic (GSI; una mesura de creixement de l'ovari) d'una espècie de peix, es van utilitzar dos fotoperíodes (14 hores de llum–10 hores de foscor, i 9 hores de llum–15 hores de foscor) i dues temperatures (16° i 27 °C)

L'experiment es va realitzar sobre 20 femelles. Es van dividir aleatòriament en 4 subgrups de 5 exemplars cadascun. Cada grup va rebre una combinació diferent de llum i temperatura.

Als 3 mesos es mesuraren els GSI dels peixos

Exemple 2

Els resultats es mostren en la taula següent:

	Factor A (fotoperíode)	
Factor B (temperatura)	9 hores	14 hores
27°C	0.90	0.83
	1.06	0.67
	0.98	0.57
	1.29	0.47
	1.12	0.66
16°C	1.30	1.01
	2.88	1.52
	2.42	1.02
	2.66	1.32
	2.94	1.63

El model

Les suposicions del model són:

- Les observacions per a cada combinació de nivells constitueixen mostres aleatòries simples independents, cadascuna de mida n, de $a \cdot b$ poblacions
- Cadascuna de les $a \cdot b$ poblacions és normal
- Totes les $a \cdot b$ poblacions tenen la mateixa variància, σ^2

El model

Els paràmetres que intervindran en el contrast són:

- μ: mitjana poblacional global
- $\mu_{i\bullet\bullet}$: mitjana poblacional del nivell i-èsim del factor A
- $\mu_{\bullet j\bullet}$: mitjana poblacional del nivell j-èsim del factor B
- $\mu_{ij\bullet}$: mitjana poblacional de la combinació (i, j) de nivells de A i B

El model

Expressió matemàtica del model a estudiar en aquest cas:

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}, \\ i = 1, \ldots, a,\ j = 1, \ldots, b,\ k = 1, \ldots, n$$

on

- μ: mitjana global
- $\alpha_i = \mu_{i\bullet\bullet} - \mu$: efecte per pertànyer al nivell i-èsim del factor A
- $\beta_j = \mu_{\bullet j\bullet} - \mu$: efecte per pertànyer al nivell j-èsim del factor B
- $(\alpha\beta)_{ij} = \mu_{ij\bullet} - \mu_{i\bullet\bullet} - \mu_{\bullet j\bullet} + \mu$: efecte de la interacció entre el nivell i-èsim del factor A i el nivell j-èsim del factor B
- $E_{ijk} = X_{ijk} - \mu_{ij\bullet}$: error residual o aleatori

Sumes i mitjanes

Emprarem les sumes i mitjanes següents:

- $T_{ij\bullet} = \sum_{k=1}^{n} X_{ijk}$ $\qquad \overline{X}_{ij\bullet} = \dfrac{T_{ij\bullet}}{n}$
- $T_{i\bullet\bullet} = \sum_{j=1}^{b}\sum_{k=1}^{n} X_{ijk} = \sum_{j=1}^{b} T_{ij\bullet}$ $\qquad \overline{X}_{i\bullet\bullet} = \dfrac{T_{i\bullet\bullet}}{bn}$
- $T_{\bullet j\bullet} = \sum_{i=1}^{a}\sum_{k=1}^{n} X_{ijk} = \sum_{i=1}^{a} T_{ij\bullet}$ $\qquad \overline{X}_{\bullet j\bullet} = \dfrac{T_{\bullet j\bullet}}{an}$
- $T_{\bullet\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} X_{ijk} = \sum_{i=1}^{a} T_{i\bullet\bullet} = \sum_{j=1}^{b} T_{\bullet j\bullet}$
- $\overline{X}_{\bullet\bullet\bullet} = \dfrac{T_{\bullet\bullet\bullet}}{abn}$
- $T^{(2)}_{\bullet\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} X_{ijk}^2$

Exemple 2

Les dades:

	Factor A				
Factor B	9 hores		14 hores		
27°C	0.90		0.83		
	1.06		0.67		
	0.98	$T_{11\bullet}$	0.57	$T_{21\bullet}$	$T_{\bullet 1\bullet}$
	1.29		0.47		
	1.12		0.66		
16°C	1.30		1.01		
	2.88		1.52		
	2.42	$T_{12\bullet}$	1.02	$T_{22\bullet}$	$T_{\bullet 2\bullet}$
	2.66		1.32		
	2.94		1.63		
	$T_{1\bullet\bullet}$		$T_{2\bullet\bullet}$		

Exemple 2

Organitzarem les dades en un dataframe amb 3 variables:

- `GSI`, quantitativa, contendrà el *GSI* de cada peix
- `llum`, factor, contendrà el valor del nivell del factor A per a cada peix
- `temp`, factor, contendrà el valor del nivell del factor B per a cada peix

Exemple 2

```
> GSI= c(0.90,0.83,1.06,0.67,0.98,0.57,
  1.29,0.47,1.12,0.66,1.30, 1.01,2.88,1.52,
  2.42,1.02,2.66,1.32,2.94,1.63)
> llum=as.factor(rep(c(9,14),10))
> llum
[1] 9 14 9 14 9 14 9 14 9 14 9 14 9 14 9
[16] 14 9 14 9 14
Levels: 9 14
> temp=as.factor(c(rep(27,10),rep(16,10)))
> temp
[1] 27 27 27 27 27 27 27 27 27 27 16 16
[13] 16 16 16 16 16 16 16 16
Levels: 16 27
```

Exemple 2

```
> peixos=data.frame(GSI,llum,temp)
> str(peixos)
'data.frame': 20 obs. of  3 variables:
 $ GSI : num  0.9 0.83 1.06 0.67 0.98 0.57
   1.29 0.47 1.12 0.66 ...
 $ llum: Factor w/ 2 levels "9","14": 1 2
   1 2 1 2 1 2 1 2 ...
 $ temp: Factor w/ 2 levels "16","27": 2 2
   2 2 2 2 2 2 2 2 ...
> head(peixos,5)
   GSI llum temp
1 0.90    9   27
2 0.83   14   27
3 1.06    9   27
4 0.67   14   27
5 0.98    9   27
```

Exemple 2

$T_{ij\bullet}$'s:

```
> n=5; a=2; b=2
> T.i.j.bullet=aggregate(GSI~llum+temp,
  data=peixos,sum)
> T.i.j.bullet
  llum temp   GSI
1    9   16 12.20
2   14   16  6.50
3    9   27  5.35
4   14   27  3.20
```

$T_{ij\bullet}$	9	14
27	5.35	3.2
16	12.20	6.5

Exemple 2

$T_{i\bullet\bullet}$'s i $T_{\bullet j\bullet}$'s:

```
> T.A=aggregate(GSI~llum,data=peixos,sum)
> T.A
  llum   GSI
1    9 17.55
2   14  9.70
> T.B=aggregate(GSI~temp,data=peixos,sum)
> T.B
  temp   GSI
1   16 18.70
2   27  8.55
```

$T_{1\bullet\bullet}$	$T_{2\bullet\bullet}$
17.55	9.7

$T_{\bullet 1\bullet}$	$T_{\bullet 2\bullet}$
8.55	18.7

Exemple 2

$T_{\bullet\bullet\bullet}$:

```
> sum(peixos$GSI)
[1] 27.25
```

$\overline{X}_{\bullet\bullet\bullet}$:

```
> sum(peixos$GSI)/(n*a*b)
[1] 1.3625
```

$T^{(2)}_{\bullet\bullet\bullet}$:

```
> sum(peixos$GSI^2)
[1] 48.2619
```

Identitats de sumes de quadrats

Emprarem les sumes de quadrats següents:

- $SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk} - \overline{X}_{\bullet\bullet\bullet})^2$ (Variabilitat total)
- $SS_A = bn\sum_{i=1}^{a}(\overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet\bullet\bullet})^2$ (Variabilitat deguda al factor A)
- $SS_B = an\sum_{j=1}^{b}(\overline{X}_{\bullet j\bullet} - \overline{X}_{\bullet\bullet\bullet})^2$ (Variabilitat deguda al factor B)

Identitats de sumes de quadrats

Emprarem les sumes de quadrats següents:

- $SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{ij\bullet} - \overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet j\bullet} + \overline{X}_{\bullet\bullet\bullet})^2$ (Variabilitat deguda a la interacció dels factors A i B)
- $SS_{Tr} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{X}_{ij\bullet} - \overline{X}_{\bullet\bullet\bullet})^2$ (Variabilitat com si empràssim un sol factor)
- $SS_{E} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk} - \overline{X}_{ij\bullet})^2$ (Variabilitat deguda a l'error aleatori)

Identitats de sumes de quadrats

Teorema

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

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

Càlcul de les sumes de quadrats

- $SS_{Total} = T^{(2)}_{\bullet\bullet\bullet} - \dfrac{T^2_{\bullet\bullet\bullet}}{abn}$
- $SS_A = \dfrac{1}{bn}\displaystyle\sum_{i=1}^{a} T^2_{i\bullet\bullet} - \dfrac{T^2_{\bullet\bullet\bullet}}{abn}$
- $SS_B = \dfrac{1}{an}\displaystyle\sum_{j=1}^{b} T^2_{\bullet j\bullet} - \dfrac{T^2_{\bullet\bullet\bullet}}{abn}$
- $SS_{Tr} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{a}\sum_{j=1}^{b} T^2_{ij\bullet} - \dfrac{T^2_{\bullet\bullet\bullet}}{abn}$
- $SS_{AB} = SS_{Tr} - SS_A - SS_B$
- $SS_E = SS_{Total} - SS_{Tr}$

Exemple 2

$T_{ij\bullet}$	9	14
27	5.35	3.2
16	12.20	6.5

$T_{1\bullet\bullet}$	$T_{2\bullet\bullet}$
17.55	9.7

$T_{\bullet 1\bullet}$	$T_{\bullet 2\bullet}$
8.55	18.7

$T_{\bullet\bullet\bullet}$	$T^{(2)}_{\bullet\bullet\bullet}$
27.25	48.2619

SS_{Total}	SS_A	SS_B	SS_{Tr}	SS_{AB}	SS_E

Exemple 2

$T_{ij\bullet}$	9	14
27	5.35	3.2
16	12.20	6.5

$T_{1\bullet\bullet}$	$T_{2\bullet\bullet}$
17.55	9.7

$T_{\bullet 1\bullet}$	$T_{\bullet 2\bullet}$
8.55	18.7

$T_{\bullet\bullet\bullet}$	$T^{(2)}_{\bullet\bullet\bullet}$
27.25	48.2619

SS_{Total}	SS_A	SS_B	SS_{Tr}	SS_{AB}	SS_E
11.1338	3.0811	5.1511	8.8624	0.6301	2.2714

Quadrats mitjans

Emprarem els quadrats mitjans següents:

- $MS_A = \dfrac{SS_A}{a-1}$
- $MS_B = \dfrac{SS_B}{b-1}$
- $MS_{AB} = \dfrac{SS_{AB}}{(a-1)(b-1)}$
- $MS_{Tr} = \dfrac{SS_{Tr}}{ab-1}$
- $MS_E = \dfrac{SS_E}{ab(n-1)}$

Exemple 1

$n = 5,\ a = b = 2$

SS_{Total}	SS_A	SS_B	SS_{AB}	SS_{Tr}	SS_E
11.1338	3.0811	5.1511	0.6301	8.8624	2.2714

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E

Exemple 1

$n = 5,\ a = b = 2$

SS_{Total}	SS_A	SS_B	SS_{AB}	SS_{Tr}	SS_E
11.1338	3.0811	5.1511	0.6301	8.8624	2.2714

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
3.0811	5.1511	0.6301	2.9541	0.142

Contrastos a realitzar

En una ANOVA de dues vies, ens poden interessar els quatre contrastos següents:

Contrast de mitjanes del factor A: Contrastam si hi ha diferències entre els nivells del factor A:

$$\left\{\begin{array}{l} H_0 : \mu_{1\bullet\bullet} = \mu_{2\bullet\bullet} = \cdots = \mu_{a\bullet\bullet} \\ H_1 : \exists i, i' \mid \mu_{i\bullet\bullet} \neq \mu_{i'\bullet\bullet} \end{array}\right.$$

L'estadístic de contrast és

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

el qual, 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

Contrastos a realitzar

En una ANOVA de dues vies, ens poden interessar els quatre contrastos següents:

Contrast de mitjanes del factor B: Contrastam si hi ha diferències entre els nivells del factor B:

$$\left\{\begin{array}{l} H_0 : \mu_{\bullet 1 \bullet} = \mu_{\bullet 2 \bullet} = \cdots = \mu_{\bullet b \bullet} \\ H_1 : \exists j, j' \mid \mu_{\bullet j \bullet} \neq \mu_{\bullet j' \bullet} \end{array}\right.$$

L'estadístic de contrast és

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

el qual, 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

Contrastos a realitzar

En una ANOVA de dues vies, ens poden interessar els quatre contrastos següents:

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

$$\left\{ \begin{array}{l} H_0 : \forall i, j, i', j' \mid \mu_{ij\bullet} = \mu_{i'j'\bullet} \\ H_1 : \exists i, j, i', j' \mid \mu_{ij\bullet} \neq \mu_{i'j'\bullet} \end{array} \right.$$

L'estadístic de contrast és

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

el qual, 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

Contrastos a realitzar

En una ANOVA de dues vies, ens poden interessar els quatre contrastos següents:

Contrast de no interacció: Contrastam si hi ha interacció entre els factors A i B:

$$\left\{\begin{array}{l} H_0 : \forall i, j \mid (\alpha\beta)_{ij} = 0 \\ H_1 : \exists i, j \mid (\alpha\beta)_{ij} \neq 0 \end{array}\right.$$

L'estadístic de contrast és

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

el qual, 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 a realitzar

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 següent:

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

Exemple 2

$n = 5,\ a = b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
3.0811	5.1511	0.6301	2.9541	0.142

$$\frac{MS_A}{MS_E} = 21.7, \text{ p-valor} = P(F_{1,16} \geqslant 21.7) = 0.0003$$

Al contrast del factor A, rebutjam la hipòtesi nul·la i concloem que hi ha diferències entre els dos nivells

Exemple 2

$n = 5, \ a = b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
3.0811	5.1511	0.6301	2.9541	0.142

$$\frac{MS_B}{MS_E} = 36.3, \text{ p-valor} = P(F_{1,16} \geqslant 36.3) = 2 \cdot 10^{-5}$$

Al contrast del factor B, rebutjam la hipòtesi nul·la i concloem que hi ha diferències entre els dos nivells

Exemple 2

$n = 5,\ a = b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
3.0811	5.1511	0.6301	2.9541	0.142

$$\frac{MS_{Tr}}{MS_E} = 20.8,\ \text{p-valor} = P(F_{3,16} \geqslant 20.8) = 9 \cdot 10^{-6}$$

Al contrast dels tractaments, rebutjam la hipòtesi nul·la i concloem que hi ha diferències entre les parelles (nivell de A, nivell de B)

Exemple 2

$n = 5$, $a = b = 2$

MS_A	MS_B	MS_{AB}	MS_{Tr}	MS_E
3.0811	5.1511	0.6301	2.9541	0.142

$$\frac{MS_{AB}}{MS_E} = 4.437, \text{ p-valor} = P(F_{1,16} \geqslant 4.437) = 0.051$$

Al contrast de no interacció, estam en zona de penombra; amb nivell de significació 0.05, no podem rebutjar la hipòtesi nul·la i hem de concloure que no hi ha interacció entre els factors

Exemple 2

Variació	Graus de llibertat	Suma de quadrats	Quadrats mitjans	F	p-valors
Tract.	3	8.862	2.954	20.8	$9 \cdot 10^{-6}$
A	1	3.081	3.081	21.7	0.0003
B	1	5.151	5.151	36.3	$2 \cdot 10^{-5}$
AB	1	0.630	0.630	4.44	0.051
Error	16	2.271	0.142		
Total	19	11.134			

Exemple 2 amb R

```
> summary(aov(GSI~llum*temp,data=peixos))
            Df Sum Sq Mean Sq F value   Pr(>F)
llum         1  3.081   3.081  21.704 0.000262 ***
temp         1  5.151   5.151  36.285 1.77e-05 ***
llum:temp    1  0.630   0.630   4.439 0.051268 .
Residuals   16  2.271   0.142
```

Exemple 2 amb R

Hi falta la filera dels Tractaments. Aquesta s'obté amb un ANOVA d'un (nou) factor que tengui com a nivells les parelles de nivells de A i B. Es pot realitzar amb

```
> summary(aov(GSI~llum:temp,data=peixos))
            Df Sum Sq Mean Sq F value   Pr(>F)
llum:temp    3  8.862   2.954   20.81 9.06e-06 ***
Residuals   16  2.271   0.142
```