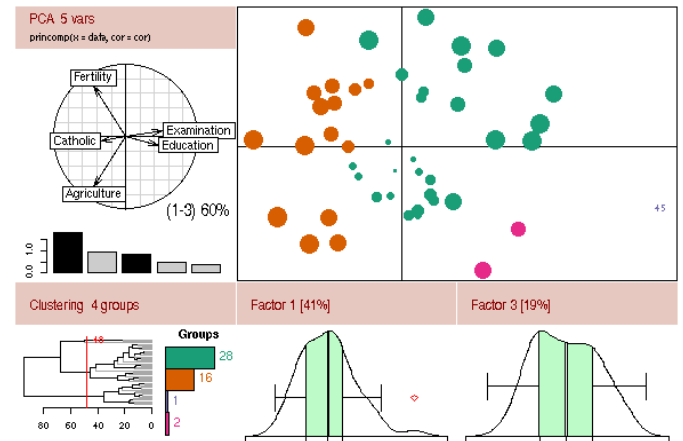


Comparació de tractaments



$$y_{ij} = \mu + \alpha_i + e_{ij}$$



Contrastos Ortogonals

És una tècnica per a dividir la suma de quadrats de l'ANOVA en graus de llibertat individuals

Definició

Si x_1, x_2, \dots, x_p són p valors numèrics obtinguts de les dades.(poden ser estadístics o bé les observacions)

Una **combinació lineal** , L , de x_1, x_2, \dots, x_p és :

$$L = c_1x_1 + c_2x_2 + \dots + c_px_p$$

On els **coeficients** c_1, c_2, \dots, c_p són valors predeterminats

Definició

Si els coeficients c_1, c_2, \dots, c_p satisfan:

$$c_1 + c_2 + \dots + c_p = 0,$$

Lavors la combinació lineal

$$L = c_1 X_1 + c_2 X_2 + \dots + c_p X_p$$

S'anomena **contrast lineal**.

Exemples

1.
$$L = \bar{x} = \frac{x_1 + x_2 + \cdots + x_p}{p} = \left(\frac{1}{p}\right)x_1 + \left(\frac{1}{p}\right)x_2 + \cdots + \left(\frac{1}{p}\right)x_p$$

Combinació lineal

2.
$$L = \frac{x_1 + x_2 + x_3}{3} - \frac{x_4 + x_5}{2}$$

$$= \left(\frac{1}{3}\right)x_1 + \left(\frac{1}{3}\right)x_2 + \left(\frac{1}{3}\right)x_3 + \left(-\frac{1}{2}\right)x_4 + \left(-\frac{1}{2}\right)x_5$$

Contrast lineal

3.
$$L = x_1 - 4x_2 + 6x_3 - 4x_4 + x_5$$

$$= (1)x_1 + (-4)x_2 + (6)x_3 + (-4)x_4 + (1)x_5$$

Contrast lineal

Definició

Siguin $A = a_1x_1 + a_2x_2 + \dots + a_px_p$ y $B = b_1x_1 + b_2x_2 + \dots + b_px_p$ són dos contrastos lineals de les quantitats x_1, x_2, \dots, x_p . Llavors A y B s'anomenen **Contrastos Ortogonals** si a la condició:

$$a_1 + a_2 + \dots + a_p = 0$$

$$b_1 + b_2 + \dots + b_p = 0,$$

S'afegeix:

$$a_1b_1 + a_2b_2 + \dots + a_pb_p = 0.$$

.

Exemple

$$A = \frac{x_1 + x_2 + x_3}{3} - \frac{x_4 + x_5}{2}$$

$$= \left(\frac{1}{3}\right)x_1 + \left(\frac{1}{3}\right)x_2 + \left(\frac{1}{3}\right)x_3 + \left(-\frac{1}{2}\right)x_4 + \left(-\frac{1}{2}\right)x_5$$

$$B = \left[\frac{x_1 + x_2}{2} - x_3 \right] + [x_4 - x_5]$$

$$= \left(\frac{1}{2}\right)x_1 + \left(\frac{1}{2}\right)x_2 + (-1)x_3 + (1)x_4 + (-1)x_5$$

Notem:

$$\left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)(-1) + \left(\frac{1}{2}\right)(1) + \left(\frac{1}{2}\right)(-1) = 0$$

Definició

Siguin

$$A = a_1x_1 + a_2x_2 + \dots + a_px_p,$$

$$B = b_1x_1 + b_2x_2 + \dots + b_px_p,$$

..., i

$$L = l_1x_1 + l_2x_2 + \dots + l_px_p$$

Un conjunt de contrastos lineals de les quantitats x_1, x_2, \dots, x_p .

El conjunt constitueix un conjunt de **Contrastos mutuament ortogonals** si cada contrast lineal del conjunt és ortogonal a qualsevol altre del mateix conjunt.

Teorema:

El nombre màxim de contrastos lineals d'un conjunt de contrastos ortogonals de les quantitats x_1, x_2, \dots, x_p és **$p - 1$** .

$p - 1$ són els graus de llibertat de la comparació de les quantitats x_1, x_2, \dots, x_p .

Comentaris

1. Els contrastos lineals permeten fer **comparacions** entre els p valors x_1, x_2, \dots, x_p
2. Els contrastos ortogonals permeten fer **comparacions independents** entre els p valors x_1, x_2, \dots, x_p .
3. El nombre de **comparacions independents** entre els p valors x_1, x_2, \dots, x_p és $p - 1$.

Definició

$$L = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_p \bar{x}_p$$

Representa un contrast lineal de: $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$

On cada mitjana, \bar{x}_i , està calculada amb n observacions.

La **Suma de quadrats per a verificar el contrast lineal L**, es defineix com:

$$SS_L = \frac{n L^2}{a_1^2 + a_2^2 + \dots + a_p^2}$$

En el cas de dissenys no balancejats:

$$SS_L = \frac{L^2}{\sum_{j=1}^p a_j^2 / n_j}$$

Els graus de llibertat del test per a L és 1

El valor de l'estadístic **F del test pel Contrast L** , és:

$$F = \frac{SS_L / 1}{MS_{Error}}$$

Teorema

Siguin

$$L_1, L_2, \dots, L_{p-1}$$

Un conjunt de $p-1$ contrastos ortogonals per a comparar p mitjanes.

Lavors la Suma de Quadrats entre les mitjanes basada en $p - 1$ graus de llibertat, $SS_{Between}$, satisfà:

$$SS_{Between} = SS_{L_1} + SS_{L_2} + \dots + SS_{L_{p-1}}$$

Comentari

La definició d'un conjunt de contrastos ortogonals per a comparar p mitjanes

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$$

Permet a l'investigador "subdividir" la Suma de Quadrats , $SS_{Between}$, en tests individuals per a cada Contrast.

Exemple: Estudi de guany de pes

- En un estudi es volen comparar 6 dietes per a engreixar animals (A,B,C,D,E,F) amb les següents característiques:
 1. A,B,C tenen un percentatge elevat de proteïnes mentre que D,E,F el tenen baix.
 2. A i D estan preparades amb carn de bou mentre que C i F ho esta amb carn de porc
 3. B i E estan preparades amb un contingut elevat de cereals

$$\bar{x}_1 = 100.0, \bar{x}_2 = 85.9, \bar{x}_3 = 99.5,$$

$$\bar{x}_4 = 79.2, \bar{x}_5 = 83.9, \bar{x}_6 = 78.7$$

La Suma de Quadrats per a comparar les 6 mitjanes es trova a la taula Anova :

Analysis of Variance Table

Source :	DF :	Sum Squares :	Mean Square :	F-test :
Between groups	5	4612.933	922.587	4.3
Within groups	54	11586	214.556	p = .0023
Total	59	16198.933		

Alguns contrastos ortogonals associats a les 6 mitjanes podrien ser:

$$L_1 = \frac{1}{4} [\bar{x}_1 + \bar{x}_3 + \bar{x}_4 + \bar{x}_6] - \frac{1}{2} [\bar{x}_2 + \bar{x}_5]$$

(Comparació de proteïnes de la carn(Bou + Porc) respecte a les de cereals)

$$L_2 = \frac{1}{2} [\bar{x}_1 + \bar{x}_4] - \frac{1}{2} [\bar{x}_3 + \bar{x}_6]$$

(Comparació de dietes amb bou i porc)

$$L_3 = \bar{x}_1 - \bar{x}_4$$

(Comparació d'alta i baixa concentració de proteïnes per a les dietes amb bou)

$$L_4 = \bar{x}_2 - \bar{x}_5$$

(Comparació d'alta i baixa concentració de proteïnes per a les dietes amb cereals)

$$L_5 = \bar{x}_3 - \bar{x}_6$$

(Comparació d'alta i baixa concentració de proteïnes per a les dietes amb porc)

La taula ANOVA per a verificar el conjunt de contrastos ortogonals és:

Source:	DF:	Sum Squares:	Mean Square:	F-test:
Beef vs Pork (L_1)	1	2.500	2.500	0.012
Meat vs Cereal (L_2)	1	264.033	264.033	1.231
High vs Low for Beef (L_3)	1	2163.200	2163.200	10.082
High vs Low for Cereal (L_4)	1	20.000	20.000	0.093
High vs Low for Pork (L_5)	1	2163.200	2163.200	10.082
Error	54	11586.000	214.556	

Contrastos Lineals Ortogonals

Regressió Polinòmica

Coeficients dels contrastos

k	Polynomial	1	2	3	4	5	6	7	8	9	10	Σa_i^2
3	Linear	-1	0	1								2
	Quadratic	1	-2	1								6
4	Linear	-3	-1	1	3							20
	Quadratic	1	-1	-1	1							4
	Cubic	-1	3	-3	1							20
5	Linear	-2	-1	0	1	2						10
	Quadratic	2	-1	-2	-1	2						14
	Cubic	-1	2	0	-2	1						10
	Quartic	1	-4	6	-4	1						70
6	Linear	-5	-3	-1	1	3	5					70
	Quadratic	5	-1	-4	-4	-1	5					84
	Cubic	-5	7	4	-4	-7	5					180
	Quartic	1	-3	2	2	-3	1					28
7	Linear	-3	-2	-1	0	1	2	3				28
	Quadratic	5	0	-3	-4	-3	0	5				84
	Cubic	-1	1	1	0	-1	-1	1				6
	Quartic	3	-7	1	6	1	-7	3				154

Coeficients dels contrastos (2)

k	Polynomial	1	2	3	4	5	6	7	8	9	10	Σa_i^2
8	Linear	-7	-5	-3	-1	1	3	5	7			168
	Quadratic	7	1	-3	-5	-5	-3	1	7			168
	Cubic	-7	5	7	3	-3	-7	-5	7			264
	Quartic	7	-13	-3	9	9	-3	-13	7			616
	Quintic	-7	23	-17	-15	15	17	-23	7			2184
9	Linear	-4	-3	-2	-1	0	1	2	3	4		20
	Quadratic	28	7	-8	-17	-20	-17	-8	7	28		2772
	Cubic	-14	7	13	9	0	-9	-13	-7	14		990
	Quartic	14	-21	-11	9	18	9	-11	-21	14		2002
	Quintic	-4	11	-4	-9	0	9	4	-11	4		468
10	Linear	-9	-7	-5	-3	-1	1	3	5	7	9	330
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6	132
	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42	8580
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18	2860
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6	780

Exemple

S'ha mesurat la vida d'una component electrònica per a veure com depèn de la ***temperatura d'activació***

Table

Activation Temperature	0	25	50	75	100	
	53	60	67	65	58	
	50	62	70	68	62	
	47	58	73	62	60	T..
Ti.	150	180	210	195	180	915
Mean	50	60	70	65	60	

$$\Sigma\Sigma y_{ij}^2 = 56545$$

$$\Sigma T_{i.}^2/n = 56475$$

$$T_{..}^2/nt = 55815$$

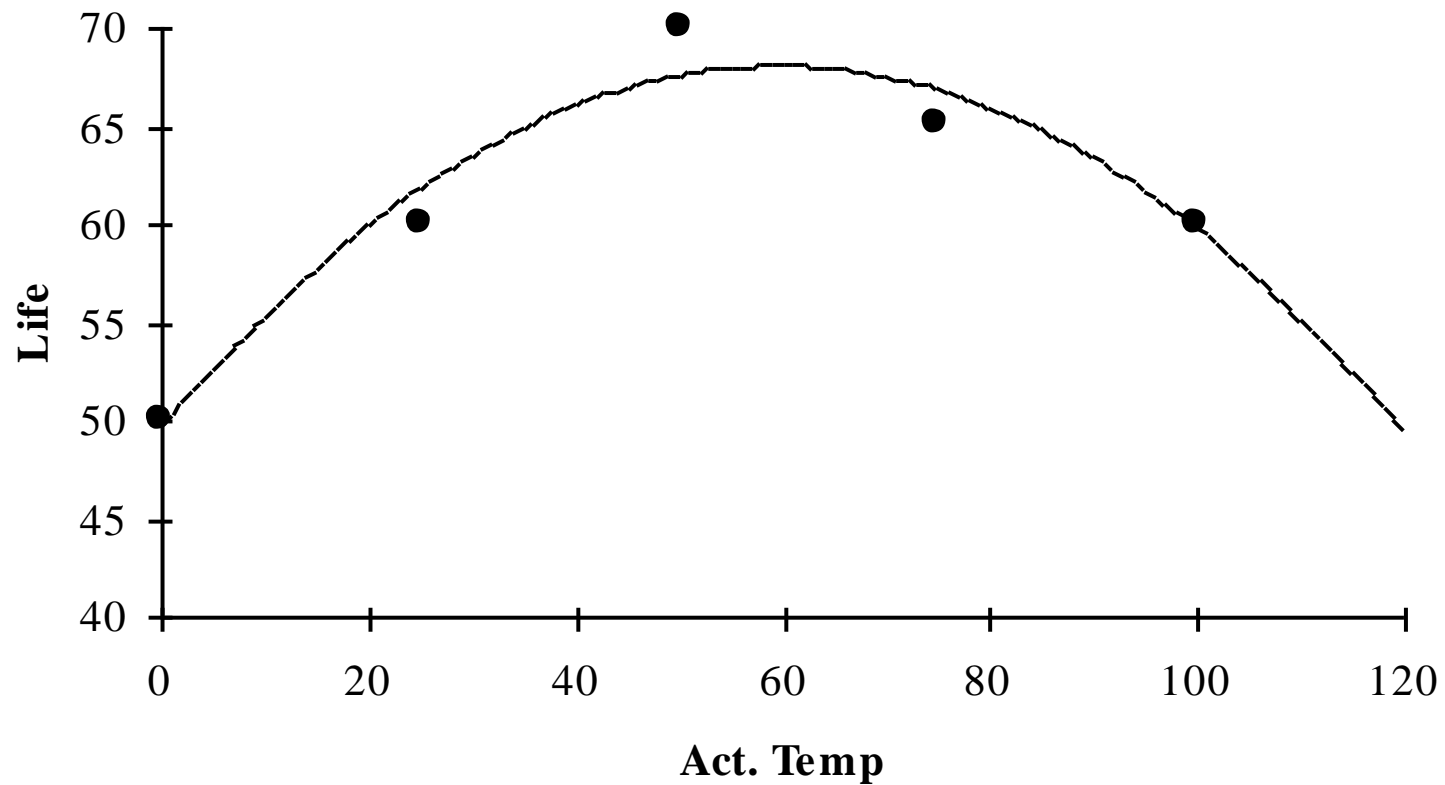
Taula Anova

$$L = 25.00 \quad Q_2 = -45.00$$
$$Q_4 = 30.00$$

$$C = 0.00$$

Source	SS	df	MS	F
Treat	660	4	165.0	23.57
Linear	187.50	1	187.50	26.79
Quadratic	433.93	1	433.93	61.99
Cubic	0.00	1	0.00	0.00
Quartic	38.57	1	38.57	5.51
Error	70	10	7.00	
Total	730	14		

$$y = 49.751 + 0.61429 x - 0.0051429 x^2$$



Comparacions Múltiples

El problema de les comparacions múltiples

- Suposem que es vol comparar l'eficàcia de 3 tractaments.
- Amb les eines que coneixem, acceptant normalitat, i considerant 3 contrasts t de Student **per parelles** consecutius:
 - fàrmac 1 contra fàrmac 2 ($\alpha=5\%$)
 - fàrmac 1 contra fàrmac 3 ($\alpha=5\%$)
 - fàrmac 2 contra fàrmac 3 ($\alpha=5\%$)
- aparentment es resol la qüestió plantejada. Però ...
- aquesta forma directa de plantejar el problema comporta un error metodològic: **el nivell de significació de las 3 proves juntes és superior a α**

-
- si plantejem els dos tests següents
 - fàrmac 1 contra fàrmac 2 ($\alpha=5\%$)
 - fàrmac 1 contra fàrmac 3 ($\alpha=5\%$)

els contrastos son independents, essent l'error de tipus I global:

$$\alpha_G = \text{nivell de significació global} = 1 - (1 - 0.05)^2 = 0.0975$$

- si plantejem les 3 a l'hora, en no ser independents, només podem afirmar que α_G està entre 0.0975 i 0.15.
- en general, quan més gran sigui el n° de grups, més gran serà α_G ¡acostant-se a 1!. Cal doncs una **tècnica alternativa**.

Contrasts conservadors vs lliberals

- Test **conservador**: per tal de mantenir el nivell de significació global proper (o igual) al nominal, disminueix el valor del nivell de comparació individual.

Conseqüència: tendeix a acceptar en excés H_0 (la igualtat de la parella de mitjanes), és **poc potent**.

- Test **lliberal**: para tal de mantenir el nivell de comparació individual proper al nominal, relaxa el control sobre el nivell de significació global.

Conseqüència: tendeix a rebutjar en excés H_0 .

Exemple

- Suposem que volem comparar la producció de 6 varietats de civada que han crescut en terrenys de les mateixes característiques (grandària, fertilitat, etc.).

Varietat	1	2	3	4	5	6
Rang	5	4	6	2	3	1
Mitjana	50.3	69.	24.	94.	75.	95.3

Font de variació	d.f.	Quadrats mitjans	F
Tract.	5	2976.44	24.8
Error	18	120.00	
Total	23		

Mètode de la diferencia mínima significativa

- dissenyat per Fisher, es realitza en 2 etapes:
 1. realitzar un Anova. Només si el factor és significatiu, passar a comparar les mitjanes (següent punt)
 2. comparar totes les possibles $a*(a-1)/2$ diferències de mitjanes contra un valor crític constant. Per a **dissenys no balancejats** aquest valor és

$$t_{\alpha(N-a)} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

FPLSD

- El mètode de les *mínimes diferències significatives protegit de Fisher* és abreujat per les seves sigles en anglès com FPLSD o simplement LSD.
- Equivalent de l'estadístic t quan tenim informació de a grups.
- Millor estima de σ^2 i més g.d.l (més potent que t senzilla) però nivell simultani real superior a α (en múltiples comparacions).

Mètode FPLSD: exemple

Varietat	6	4	5	2	1	3
Mitjana	95.3	94.	75.	69.	50.3	24.

$$FPLSD = t_{0.05}(18) \sqrt{2MS_E/n} = 2.101 \sqrt{2(120)/4} = 16.27$$

	94.0	75.0	69.0	50.3	24.0
Rang	2	3	4	5	6
95.3	1.3	20.3*	26.3*	45.0*	71.3*
94.0		19.0*	25.0*	43.7*	70.0*
75.0			6.0	24.7*	51.0*
69.0				18.7*	45.0*
50.3					26.3*

Mètode de Tukey

Suposem que tenim p mitjanes $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$

El test F ha detectat diferències significatives entre elles i volem determinar on estan aquestes diferències.

Ordenarem les mitjanes i calcularem les diferències 2 a 2

$$|\bar{x}_i - \bar{x}_j|$$

Si l'error estàndar de cada $\bar{x}_i = \frac{s}{\sqrt{n}} = \frac{\sqrt{MS_{Error}}}{\sqrt{n}}$

El valor crític de Tukey per a detectar diferències és

$$D_\alpha = q_\alpha \frac{s}{\sqrt{n}} = q_\alpha \frac{\sqrt{MS_{Error}}}{\sqrt{n}}$$

Direm que dues mitjanes són significativament diferents si la seva diferència de mitjanes supera aquest valor.

q_α = valor de les taules de Tukey del “studentized range”, amb p
=no. of mitjanes, ν = g.l. de l'Error

Si el disseny és no balancejat: $D_\alpha = \frac{q_\alpha}{\sqrt{2}} \sqrt{MS_{Error} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$

Exemple

$$HSD = Q_{0.05}(6,18) \sqrt{MS_E / n} = 4.49 \sqrt{120 / 4} = 24.59$$

	94.0	75.0	69.0	50.3	24.0
Rang	2	3	4	5	6
95.3	1.3	20.3	26.3*	45.0*	71.3*
94.0		19.0	25.0*	43.7*	70.0*
75.0			6.0	24.7*	51.0*
69.0				18.7	45.0*
50.3					26.3*

Varietat	6	4	5	2	1	3
Mitjana	95.3	94.	75.	69.	50.3	24.

Dunnett

	CONTROL	Met. 1	Met. 2	Met. 3
	55	55	55	50
	47	64	49	44
	48	64	52	41
Mean	50	61	52	45

Var.	SS	d.f.	MS	F
Treat.	402	3	134	7.05
Error	152	8	19	
Total	554	11		

$$Dun = d_{\alpha}(a, N - a) \sqrt{\frac{2MS_E}{n}} = 2.88 \sqrt{\frac{2 \times 19}{3}} = 10.25$$

Mètode de Scheffé

- Permet no només comparar mitjanes dos a dos, sinó també **testar múltiples contrastos lineals** (no cal que siguin ortogonals) tenint en compte les comparacions múltiples.

**Direm que un contrast L es significatiu (diferent de 0) quan:
El valor crític de Tukey per a detectar diferències és**

$$|\hat{L}| > \sqrt{(p-1)F_{\alpha; p-1, N-p}} \sqrt{MS_{Error} \sum_{j=1}^p c_j^2 / n_j}$$

$$L = c_1\mu_1 + c_2\mu_2 + \dots c_p\mu_p$$

$$\hat{L} = c_1\bar{x}_1 + c_2\bar{x}_2 + \dots c_p\bar{x}_p$$