<u>Comments:</u> For $R(\alpha|\mu)$,

(i) $\sum_{j=1}^{b} \dfrac{n_{ij}}{n_{i.}} \mu_{ij}$ may not be equal for all $i = 1, \ldots, a$, when

$\dfrac{1}{b} \sum_{j=1}^{b} \mu_{ij}$ are equal for all $i = 1, \ldots, a$.

(ii) $\sum_{j=1}^{b} \dfrac{n_{ij}}{n_{i.}} \mu_{ij}$ may be equal for all $i = 1, \ldots, a$, when

$\dfrac{1}{b} \sum_{j=1}^{b} \mu_{ij}$ are not equal for some $i = 1, \ldots, a$.

Consider

$$R(\beta|\mu,\alpha) = \mathrm{Y}^T (P_{\mu,\alpha,\beta} - P_{\mu,\alpha})\mathrm{Y}$$

and the corresponding F-statistic

$$F = \frac{R(\beta|\mu,\alpha)/(b-1)}{MSE} \sim F_{(b-1, n_{..}-ab)}(\delta^2)$$

Here,

$$\frac{1}{\sigma^2} R(\beta|\mu,\alpha) \;\sim\; \chi^2_{\mathrm{rank}(\mathrm{X}_{\mu,\alpha,\beta}) - \mathrm{rank}(\mathrm{X}_{\mu,\alpha})}(\delta^2)$$

$$[1+(a-1)+(b-1)] - [1+(a-1)]$$
$$= b-1 \text{ degrees of freedom}$$

and

$$\delta^2 = \frac{1}{2\sigma^2}[(P_{\mu,\alpha,\beta} - P_{\mu,\alpha})X\beta]^T[(P_{\mu,\alpha,\beta} - P_{\mu,\alpha})X\beta]$$

$$P_{\mu,\alpha,\beta}X = X_{\mu,\alpha,\beta}[X^T_{\mu,\alpha,\beta}X_{\mu,\alpha,\beta}]^{-}X^T_{\mu,\alpha,\beta}X$$

$$= X_{\mu,\alpha,\beta}\left[\begin{array}{c|cc|ccc} n_{..} & n_{1.} & n_{2.} & n_{.1} & n_{.2} & n_{.3} \\ \hline n_{1.} & n_{1.} & 0 & n_{11} & n_{12} & n_{13} \\ n_{2.} & 0 & n_{2.} & n_{21} & n_{22} & n_{23} \\ \hline n_{.1} & n_{11} & n_{21} & n_{.1} & 0 & 0 \\ n_{.2} & n_{12} & n_{22} & 0 & n_{.2} & 0 \\ n_{.3} & n_{13} & n_{23} & 0 & 0 & n_{.3} \end{array}\right]^{-} X^T_{\mu,\alpha,\beta}X$$

call this $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} A^{-1}B \\ I \end{bmatrix} [C - B^T A^{-1} B]^{-1} [-B^T A^{-1} \,|\, I]$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & C^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -C^{-1}B^T \end{bmatrix} [A - BC^{-1}B^T]^{-1} [I \,|\, -BC^{-1}]$$

$$= \begin{bmatrix} W & -WBC^{-1} \\ -C^{-1}B^T W & C^{-1} + C^{-1}B^T W B C^{-1} \end{bmatrix}$$

where $W = [A - BC^{-1}B^T]^{-1}$

The null hypothesis is

$$H_0: \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} (\beta_j + \gamma_{ij}) - \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} \left(\sum_{k=1}^{b} \frac{n_{ik}}{n_{i.}} (\beta_k + \gamma_{ik}) \right) = 0$$

for all $j = 1, \ldots, b$

With respect to the cell means,

$$E(Y_{ijk}) = \mu_{ij},$$

this null hypothesis is

$$H_0: \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} \mu_{ij} - \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} \left(\sum_{k=1}^{b} \frac{n_{ik}}{n_{i.}} \mu_{ik} \right) = 0$$

for all $j = 1, 2, \ldots, b$.

Consider

$$R(\gamma|\mu,\alpha,\beta) = \mathrm{Y}^T[P_X - P_{\mu,\alpha,\beta}]\mathrm{Y}$$

and the associated F-statistic

$$F = \frac{R(\gamma|\mu,\alpha,\beta)/[(a-1)(b-1)]}{MSE} \sim F_{(a-1)(b-1),n_{..}-ab}(\delta^2)$$

The null hypothesis is:

$$H_0: \ (\mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell}) = (\gamma_{ij} - \gamma_{i\ell} - \gamma_{kj} + \gamma_{k\ell}) = 0$$

for all (i,j) and (k,ℓ) .

Type I sums of squares

Source of variation.	d.f.	sums of squares	Mean square	F	p-val
Soil types	$a-1=1$	$R(\alpha\|\mu) = 52.5$	52.5	3.94	.0785
Var.	$b-1=2$	$R(\beta\|\mu,\alpha) = 124.73$	62.4	4.68	.0405
Interaction	(a-1)(b-1) =2	$R(\gamma\|\mu,\alpha,\beta) = 222.76$	111.38	8.35	.0089
Resid.	$n_{\bullet\bullet} - ab = 9$	$\mathrm{Y}^T(I-P_X)\mathrm{Y} = 120$	13.33		
Corr. total	$n_{\bullet\bullet} - 1 = 14$	$\mathrm{Y}^T(I-P_1)\mathrm{Y} = 520$			
Corr. for the mean	1	$R(\mu) = 3375$			

Summary:

Sums of Squares	Associated null hypothesis
$R(\mu)$	$H_0: \mu + \sum_{i=1}^{a}\frac{n_{i.}}{n_{..}}\alpha_i + \sum_{j=1}^{b}\frac{n_{.j}}{n_{..}}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{n_{..}}\gamma_{ij} = 0$ $\left(\text{or } H_0: \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{n_{..}}\mu_{ij} = 0\right)$
$R(\alpha\|\mu)$	$H_0: \alpha_i + \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}(\beta_j + \gamma_{ij})$ are equal $\left(\text{or } H_0: \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}\mu_{ij}\right.$ are equal $\left.\right)$
$R(\beta\|\mu,\alpha)$	$H_0: \beta_j + \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\gamma_{ij} = \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\sum_{k=1}^{b}\frac{n_{ik}}{n_{k.}}(\beta_k + \gamma_{ik})$ for all $j = 1,\ldots,b$ $\left(\text{or } H_0: \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\mu_{ij} = \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\sum_{k=1}^{b}\frac{n_{ik}}{n_{i.}}\mu_{ik} \text{ for all } j = 1,\ldots,b\right)$
$R(\gamma\|\mu,\alpha,\beta)$	$H_0: \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$ for all (i,j) and (k,ℓ) (or $H_0: \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0$ for all $(i,j)$and (k,ℓ)

Sums of Squares	Associated null hypothesis
$R(\mu)$	$H_0: \mu + \sum_{i=1}^{a}\frac{n_{i.}}{n_{..}}\alpha_i + \sum_{j=1}^{b}\frac{n_{.j}}{n_{..}}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{n_{..}}\gamma_{ij} = 0$ $\left(\text{or } H_0: \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{n_{..}}\mu_{ij} = 0\right)$
$R(\beta\|\mu)$	$H_0: \beta_j + \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}(\alpha_j + \gamma_{ij})$ are equal for all $j = 1,\ldots,b$ $\left(\text{or } H_0: \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\mu_{ij} \text{ are equal for all } j = 1,\ldots,b\right)$
$R(\alpha\|\mu,\beta)$	$H_0: \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}(\alpha_{ij} + \gamma_{ij}) = \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}\sum_{k=1}^{a}\frac{n_{kj}}{n_{.j}}(\alpha_k + \gamma_{kj})$ for all $i = 1,\ldots,a$ $\left(\text{or } H_0: \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}\mu_{ij} = \sum_{j=1}^{b}\frac{n_{ij}}{n_{i.}}\left[\sum_{k=1}^{a}\frac{n_{kj}}{n_{.j}}\mu_{kj}\right] \text{ for all } i = 1,\ldots,a\right)$
$R(\gamma\|\mu,\alpha,\beta)$	$H_0: \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$ for all (i,j) and (k,ℓ) $\left(\text{or } H_0: \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0 \text{ for all } (i,j) \text{ and } (k,\ell)\right)$

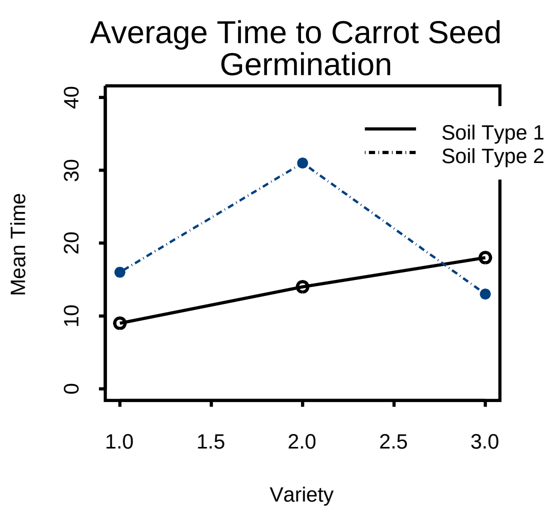
Type I sums of squares

Source of variat.	d.f.	sums of squares	Mean square	F	p-val
"Soils"	$a-1=1$	$R(\alpha\|\mu)=52.50$	52.5	3.94	.0785
"Var."	$b-1=2$	$R(\beta\|\mu,\alpha)=124.73$	62.4	4.68	.0405
Inter-action	(a-1)(b-1) =2	$R(\gamma\|\mu,\alpha,\beta)$ =222.76	111.38	8.35	.0089
"Res."	$\Sigma\Sigma(n_{ij}-1)$ =9	$Y^T(I-P_X)Y$ =120.00	13.33		
Corr. total	$n_{..}-1=14$	$Y^T(I-P_1)Y$ =520.00			

Source of variat.	d.f.	sums of squares	Mean square	F	p-val
"Var."	$b-1=2$	$R(\beta\|\mu)=93.33$	46.67	3.50	.0751
"Soils"	$a-1=1$	$R(\alpha\|\mu,\beta)=83.90$	83.90	6.29	.0334
Inter-action	(a-1)(b-1) =2	$R(\gamma\|\mu,\alpha,\beta)$ = 222.76	111.38	8.35	.0089
"Res."	$\Sigma\Sigma(n_{ij}-1)$ =9	$Y^T(I-P_X)Y$ =120.00	13.33		
Corr. total	$n_{..}-1=14$	$Y^T(I-P_1)Y$ =520.00			

Type II sums of squares:

Source of variat.	d.f.	sums of squares	Mean square	F	p-val
"Soils"	$a-1=1$	$R(\alpha\|\mu,\beta)=83.90$	83.90	6.3	.0339
"Var."	$b-1=2$	$R(\beta\|\mu,\alpha)=124.73$	62.37	4.7	.0405
Inter-action	(a-1)(b-1) =2	$R(\gamma\|\mu,\alpha,\beta)=222.76$	111.38	8.4	.0089
"Res."	$n_{\bullet\bullet}-ab$ =9	$Y^T(I-P_X)Y=120$	13.33		
Corr. total	$n_{\bullet\bullet}-1$	$Y^T(I-P_1)Y=520$			



Examine the soil type effect on time to germination for each variety:

Time to Germination

	Soil Type 1		Soil Type 2			
Variety	$\bar{Y}_{1j.}$	$S_{\bar{Y}_{1j.}}$	$\bar{Y}_{2j.}$	$S_{\bar{Y}_{2j.}}$	t	p-value
$j=1$	9.0	2.11	16.0	1.83	-2.51	.0333
$j=2$	14.0	2.58	31.0	3.65	-3.80	.0042
$j=3$	18.0	2.58	13.0	2.11	1.50	.1679

- Time to germination for variety 2 is shorter in soil type 1.
- Time to germination for variety 1 may also be shorter in soil type 1.
- For variety 3 there is no significant difference in average germination times for the two soil types.

In the previous analysis:

$$\begin{aligned}\bar{Y}_{ij.} &= \hat{\mu}_{ij} \\ &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}\end{aligned}$$

is the OLS estimator (b.l.u.e.) for

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

Also,

$$S_{\bar{Y}_{ij.}} = \sqrt{\frac{MSE}{n_{ij}}} \quad \text{for } i = 1, \ldots, a \text{ and } j = 1, \ldots, b$$

Since $\bar{Y}_{1j.}$ is independent of $\bar{Y}_{2j.}$

$$t = \frac{\bar{Y}_{1j.} - \bar{Y}_{2j.}}{\sqrt{MSE(\frac{1}{n_{1j}} + \frac{1}{n_{2j}})}} \quad \text{for } j = 1, \ldots, b$$

Method of Unweighted Means

(Type III sums of squares in SAS when $n_{ij} > 0$ for all (i, j)).

Consider the cell means reparameterization of the model:

$$\begin{aligned}Y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \\ &= \mu_{ij} + \epsilon_{ijk}\end{aligned}$$

$$\underset{\mathbf{Y}}{\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{214} \\ Y_{221} \\ Y_{231} \\ Y_{232} \\ Y_{233} \end{bmatrix}} = \underset{D}{\begin{bmatrix} 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \end{bmatrix}} \underset{\mu}{\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{113} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{213} \\ \epsilon_{214} \\ \epsilon_{221} \\ \epsilon_{231} \\ \epsilon_{232} \\ \epsilon_{233} \end{bmatrix}$$

The model is

$$\mathbf{Y} = D\mu + \epsilon$$

The least squares estimator (b.l.u.e.) for μ is

$$\begin{aligned}\hat{\mu} &= (D^T D)^{-1} D^T \mathbf{Y} \\ &= \begin{bmatrix} n_{11}^{-1} &&&&& \\ & n_{12}^{-1} &&&& \\ && n_{13}^{-1} &&& \\ &&& n_{21}^{-1} && \\ &&&& n_{22}^{-1} & \\ &&&&& n_{23}^{-1} \end{bmatrix} \begin{bmatrix} Y_{11.} \\ Y_{12.} \\ Y_{12.} \\ Y_{21.} \\ Y_{22.} \\ Y_{23.} \end{bmatrix} \\ &= \begin{bmatrix} \bar{Y}_{11.} \\ \bar{Y}_{12.} \\ \bar{Y}_{13.} \\ \bar{Y}_{21.} \\ \bar{Y}_{22.} \\ \bar{Y}_{23.} \end{bmatrix}\end{aligned}$$

Test the null hypothesis

$$H_0: \frac{1}{b}\sum_{j=1}^{b}\mu_{1j} = \frac{1}{b}\sum_{j=1}^{b}\mu_{2j} = \cdots = \frac{1}{b}\sum_{j=1}^{b}\mu_{aj}$$

vs.

$$H_A: \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} \neq \frac{1}{b}\sum_{j=1}^{b}\mu_{kj} \text{ for some } i \neq k$$

The OLS estimator (b.l.u.e.) for $\frac{1}{b}\sum_{j=1}^{b}\mu_{ij}$ is

$$\tilde{Y}_{i..} = \frac{1}{b}\sum_{j=1}^{b}\bar{Y}_{ij.}$$

with

$$\begin{aligned} Var(\tilde{Y}_{i..}) &= \frac{1}{b^2}\sum_{j=1}^{b}\frac{\sigma^2}{n_{ij}} \\ &= \frac{\sigma^2}{b^2}\sum_{j=1}^{b}\frac{1}{n_{ij}} \end{aligned}$$

Express the null hypothesis in matrix form:

$$H_0: C_1\mu = 0$$

where

$$\begin{aligned} C_1\mu &= ([I_{a-1} | -1_{a-1}] \otimes 1_b^T)\mu \\ &= \begin{bmatrix} 1_b^T & & & -1_b^T \\ & 1_b^T & & -1_b^T \\ & & \ddots & \vdots \\ & & 1_b^T & -1_b^T \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1b} \\ \mu_{21} \\ \vdots \\ \mu_{2b} \\ \vdots \\ \mu_{ab} \end{bmatrix} \\ &= \begin{bmatrix} \sum_j \mu_{1j} - \sum_j \mu_{aj} \\ \vdots \\ \sum_j \mu_{a-1,j} - \sum_j \mu_{aj} \end{bmatrix} \end{aligned}$$

Then

$$\begin{aligned} C_1\mathrm{b} &= C_1(D^TD)^{-1}D^T\mathrm{Y} \\ &= \begin{bmatrix} \sum_j \bar{Y}_{1j.} - \sum_j \bar{Y}_{aj.} \\ \vdots \\ \vdots \\ \sum_j \bar{Y}_{a-1,j.} - \sum_j \bar{Y}_{aj.} \end{bmatrix} \end{aligned}$$

is the OLS estimator (b.l.u.e.) of $C_1\mu$, and

$$\begin{aligned} Var(C_1\mathrm{b}) &= Var(C_1(D^TD)^{-1}D^T\mathrm{Y}) \\ &= C_1(D^TD)^{-1}D^T(\sigma^2 I)D(D^TD)^{-1}C_1^T \\ &= \sigma^2 C_1(D^TD)^{-1}D^TD(D^TD)^{-1}C_1^T \\ &= \sigma^2 C_1(D^TD)^{-1}C_1^T \end{aligned}$$

Compute

$$\begin{aligned} SS_{H_0} &= (C_1\mathrm{b} - 0)^T[C_1(D^TD)^{-1}C_1^T]^{-1}(C_1\mathrm{b} - 0) \\ &= \mathrm{Y}^TD(D^TD)^{-1}C_1^T[C_1(D^TD)^{-1}C_1^T]^{-1} \\ &\qquad C_1(D^TD)^{-1}D^T\mathrm{Y} \end{aligned}$$

Use result 4.7 to show

$$\frac{1}{\sigma^2}SS_{H_0} \sim \chi^2_{(a-1)}(\delta^2)$$

Check that

$$\begin{aligned} A\Sigma &= \frac{1}{\sigma^2}D(D^TD)^{-1}C_1^T[C_1(D^TD)^{-1}C_1^T]^{-1} \\ &\qquad C_1(D^TD)^{-1}D^T(\sigma^2 I) \end{aligned}$$

is idempotent and that

$$a - 1 = \text{rank}(C_1(D^TD)^{-1}C_1^T)$$

Compute:

$$SSE = \mathrm{Y}^T(I - P_D)\mathrm{Y}$$

where

$$P_D = D(D^TD)^{-1}D^T$$

Use result 4.7 to show

$$\frac{1}{\sigma^2}SSE \sim \chi^2_{\Sigma\Sigma(n_{ij}-1)}$$

Use result 4.8 to show that

$$SSE = \mathrm{Y}^T\underline{(I - P_D)}\mathrm{Y}$$

↖ **call this A_1**

is distributed independently of

$$SS_{H_o} = \mathrm{Y}^T\underline{D(D^TD)^{-1}C_1^T[C_1(D^TD)^{-1}C_1^T]^{-1}C_1(D^TD)^{-1}D^T}\mathrm{Y}$$

↖ **call this A_2**

Check that

$$\begin{aligned} A_1\Sigma A_2 &= A_1(\sigma^2 I)A_2 \\ &= \sigma^2 A_1A_2 \\ &= \sigma^2(I - P_D)(D(D^TD)^{-1}C_1^T \\ &\quad (C_1(D^TD)^{-1}C_1^T)^{-1}C_1(D^TD)^{-1}D^T \\ &= 0 \end{aligned}$$

This is true because $(I - P_D)D = 0$.

Then

$$\begin{aligned} F &= \frac{SS_{H_0}/(a-1)}{SSE/_{(\Sigma\Sigma(n_{ij}-1))}} \\ &\sim F_{(a-1,\Sigma\Sigma(n_{ij}-1))}(\delta^2) \end{aligned}$$

where

$$\delta^2 = \frac{1}{\sigma^2}\mu^T C_1^T\,[C_1(D^TD)^{-1}C_1^T]^{-1}C_1\mu$$

Reject

$$H_0 : \frac{1}{b}\sum_{j=1}^{b}\mu_{1j} = \frac{1}{b}\sum_{j=1}^{b}\mu_{2j} = \cdots = \frac{1}{b}\sum_{j=1}^{b}\mu_{aj}$$

if

$$F = \frac{SS_{H_0}/(a-1)}{SSE/_{(\Sigma\Sigma(n_{ij}-1))}} \;>\; F_{(a-1,\Sigma\Sigma(n_{ij}-1))\,,\,\alpha}$$

or, if

$$\begin{aligned} p\text{-value} &= Pr\{F_{(a-1,\Sigma\Sigma(n_{ij}-1))} > F\} \\ &< \alpha \end{aligned}$$

Test

$$H_0: \frac{1}{a}\sum_{i=1}^{a} \mu_{i1} = \frac{1}{a}\sum_{i=1}^{a} \mu_{i2} = \cdots = \frac{1}{a}\sum_{i=1}^{a} \mu_{ib}$$

vs.

$H_A: \frac{1}{a}\sum_{i=1}^{a} \mu_{ij} \neq \frac{1}{a}\sum_{i=1}^{a} \mu_{ik}$ for some $j \neq k$

Write the null hypothesis in matrix form as

$$H_0: C_2\mu = 0$$

where

$$\begin{aligned} C_2 &= 1_a^T \otimes [I_{b-1} | -1_{b-1}] \\ &= \left[\begin{array}{ccccc|cc|ccccc} 1 & & & & -1 & \cdots & \cdots & 1 & & & & -1 \\ & 1 & & & -1 & & & & 1 & & & -1 \\ & & \ddots & & \vdots & & & & & \ddots & & \vdots \\ & & & 1 & -1 & & & & & & 1 & -1 \end{array}\right] \end{aligned}$$

then

$$C_2\mu = C_2 \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1b} \\ \mu_{21} \\ \vdots \\ \mu_{ab} \end{bmatrix} = \begin{bmatrix} \frac{1}{a}\sum_{i=1}^{a} \mu_{i1} - \frac{1}{a}\sum_{i=1}^{a} \mu_{ib} \\ \vdots \\ \frac{1}{a}\sum_{i=1}^{a} \mu_{i,b-1} - \frac{1}{a}\sum_{i=1}^{a} \mu_{ib} \end{bmatrix}$$

Compute

$$\begin{aligned} SS_{H_{0,2}} = \; & Y^T D(D^TD)^{-1}C_2^T \\ & [C_2(D^TD)^{-1}C_2^T]^{-1} \\ & C_2(D^TD)^{-1}D^TY \end{aligned}$$

and reject H_0 if

$$F = \frac{SS_{H_{0,2}}/(b-1)}{SSE/_{(\Sigma\Sigma(n_{ij}-1))}} > F_{(b-1,\Sigma\Sigma(n_{ij}-1)),\alpha}$$

Test for Interaction:

Test

$$H_0: \mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell} = 0$$

for all (i,j) and (k,ℓ)

vs.

$$H_A: \mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell} \neq 0$$

for some $(i \neq k)$ and $(j \neq \ell)$.

Write the null hypothesis in matrix form as

$$H_0: C_3\,\mu = 0$$

where

$$C_3 = [I_{a-1} | -1_{a-1}] \otimes [I_{b-1} | -1_{b-1}]$$

Compute

$$b = (D^TD)^{-1}D^TY = \begin{bmatrix} \bar{Y}_{11.} \\ \vdots \\ \bar{Y}_{ab.} \end{bmatrix}$$

$$\begin{aligned} SS_{H_{0,3}} &= (C_3 b - 0)^T[C_3(D^TD)^{-1}C_3^T]^{-1} \\ & \quad (C_3 b - 0) \\ &= Y^T D(D^TD)^{-1}C_3^T[C_3(D^TD)^{-1}C_3^T]^{-1} \\ & \quad C_3(D^TD)^{-1}D^TY \end{aligned}$$

and reject H_0 if

$$F = \frac{SS_{H_{0,3}}/((a-1)(b-1))}{SSE/_{(\Sigma\Sigma(n_{ij}-1))}} > F_{((a-1)(b-1),\Sigma\Sigma(n_{ij}-1)),\alpha}$$

PROC GLM is SAS reports this as Type III sums of squares.

Source of variation	d.f.	Sum of Squares	Mean Square	F	p-val
Soils	a-1=1	$SS_{H_0} =$ 123.77	123.77	9.28	.0139
Var.	b-1=2	$SS_{H_{0,2}} =$ 192.13	96.06	7.20	.0135
Inter.	(a-1)(b-1)=2	$SS_{H_{0,3}} =$ 222.76	111.38	8.35	.0089

Note that

$$\begin{aligned} \mathrm{Y}^T P_1 \mathrm{Y} \;&+\; \mathrm{Y}^T D(D^TD)^{-1}[C_1(D^TD)^{-1}C_1^T]^{-1} \\ &\qquad C_1(D^TD)^{-1}D^T\mathrm{Y} \\ &+\; \mathrm{Y}^T D(D^TD)^{-1}C_2^T[C_2(D^TD)^{-1}C_2^T]^{-1} \\ &\qquad C_2(D^TD)^{-1}D^T\mathrm{Y} \\ &+\; \mathrm{Y}^T D(D^TD)^{-1}C_3^T[C_3(D^TD)^{-1}C_3^T]^{-1} \\ &\qquad C_3(D^TD)^{-1}D^T\mathrm{Y} \\ &+\; \mathrm{Y}^T(I-P_D)\mathrm{Y} \end{aligned}$$

do not necessarily sum to $\mathrm{Y}^T\mathrm{Y}$, nor do the middle three terms (SS_{H_0}, $SS_{H_0,2}$, $SS_{H_0,3}$) necessarily sum to

$$SS_{\text{model,corrected}} = \mathrm{Y}^T(P_D - P_1)\mathrm{Y} \ ,$$

nor are (SS_{H_0}, $SS_{H_0,2}$, $SS_{H_0,3}$) necessarily independent of each other.

Note that

$$SS_{H_0} = \sum_{i=1}^{a} w_i \left[\tilde{Y}_{i.} - \frac{\sum_{k=1}^{a} w_k \tilde{Y}_{k.}}{\sum_{k=1}^{a} w_k}\right]^2$$

where

$$\begin{aligned} \tilde{Y}_{i.} &= \frac{1}{b}\sum_{j=1}^{b}\bar{Y}_{ij.} \\ w_i &= \left[\frac{1}{b^2}\sum_{j=1}^{b}\frac{\sigma^2}{n_{ij}}\right]^{-1} = \sigma^2\left[Var(\tilde{Y}_{i.})\right]^{-1} \end{aligned}$$

and $\tilde{Y}_{i.}$ is not necessarily equal to

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} Y_{ijk}}{\sum_{j=1}^{b} n_{ij}} = \frac{\sum_{j=1}^{b} n_{ij}\bar{Y}_{ij.}}{\sum_{j=1}^{b} n_{ij}}$$

Furthermore,

$$SS_{H_0,2} = \sum_{j=1}^{b} w_j \left[\tilde{Y}_{.j} - \frac{\sum_{\ell=1}^{a} w_\ell \tilde{Y}_{.\ell}}{\sum_{\ell=1}^{a} w_\ell}\right]^2$$

where

$$\begin{aligned} \tilde{Y}_{.j} &= \frac{1}{a}\sum_{i=1}^{a}\bar{Y}_{ij.} \\ w_j &= \left[\frac{1}{a^2}\sum_{i=1}^{a}\frac{\sigma^2}{n_{ij}}\right]^{-1} = \sigma^2\left[Var(\tilde{Y}_{.j})\right]^{-1} \end{aligned}$$

and $\tilde{Y}_{.j}$ is not necessarily equal to

$$\bar{Y}_{.j} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{n_{ij}} Y_{ijk}}{\sum_{i=1}^{a} n_{ij}} = \frac{\sum_{i=1}^{a} n_{ij}\bar{Y}_{ij.}}{\sum_{i=1}^{a} n_{ij}}$$

Balanced factorial experiments

$$n_{ij} = n \quad \text{for} \quad i = 1, \ldots, a$$
$$j = 1, \ldots, b$$

Example 8.2: Sugar Cane Yields (from Snedecor and Cochran)

	Nitrogem Level		
	150 lb/acre	210 lb/acre	270 lb/acre
Variety 1	$Y_{111} = 70.5$	$Y_{121} = 67.3$	$Y_{131} = 79.9$
	$Y_{112} = 67.5$	$Y_{122} = 75.9$	$Y_{132} = 72.8$
	$Y_{113} = 63.9$	$Y_{123} = 72.2$	$Y_{133} = 64.8$
	$Y_{114} = 64.2$	$Y_{124} = 60.5$	$Y_{134} = 86.3$
Variety 2	$Y_{211} = 58.6$	$Y_{221} = 64.3$	$Y_{231} = 64.4$
	$Y_{212} = 65.2$	$Y_{222} = 48.3$	$Y_{232} = 67.3$
	$Y_{213} = 70.2$	$Y_{223} = 74.0$	$Y_{233} = 78.0$
	$Y_{214} = 51.8$	$Y_{224} = 63.6$	$Y_{234} = 72.0$
Variety 3	$Y_{311} = 65.8$	$Y_{321} = 64.1$	$Y_{331} = 56.3$
	$Y_{312} = 68.3$	$Y_{322} = 64.8$	$Y_{332} = 54.7$
	$Y_{313} = 72.7$	$Y_{323} = 70.9$	$Y_{331} = 66.2$
	$Y_{314} = 67.6$	$Y_{324} = 58.3$	$Y_{334} = 54.4$

For a balanced experiment, Type I, Type II, and Type III sums of squares are the same:

$$R(\alpha|\mu) = R(\alpha|\mu,\beta) = SS_{H_0} = n\,b \sum_{i=1}^{a} (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$R(\beta|\mu) = R(\beta|\mu,\alpha) = SS_{H_{0,2}} = n\,a \sum_{j=1}^{b} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$R(\gamma|\mu,\alpha,\beta) = SS_{H_{0,3}} = n \sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

Summary

Sum of Squares	Associated null hypothesis
$R(\mu) = \mathbf{Y}^T P_1 \mathbf{Y} = a\,b\,n\,\bar{Y}_{...}^2$	$H_0: \mu + \frac{1}{a}\sum_{i=1}^{a}\alpha_i + \frac{1}{b}\sum_{j=1}^{b}\beta_j + \frac{1}{a\,b}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij} = 0$ $\left(H_0: \frac{1}{a\,b}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} = 0\right)$
$R(\alpha\|\mu) = R(\alpha\|\mu,\beta) = n\,b\sum_{i=1}^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$H_0: \alpha_i + \frac{1}{b}\sum_{j=1}^{b}(\beta_j + \gamma_{ij})$ are equal $\left(H_0: \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} \text{ are equal}\right)$
$R(\beta\|\mu) = R(\beta\|\mu,\alpha) = n\,a\sum_{j=1}^{b}(\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$H_0: \beta_j + \frac{1}{a}\sum_{i=1}^{a}(\alpha_i + \gamma_{ij})$ are equal $\left(H_0: \frac{1}{a}\sum_{i=1}^{a}\mu_{ij} \text{ are equal}\right)$

$$R(\gamma|\mu,\alpha,\beta) = n \sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$H_0: \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$ for all (i,j) and (k,ℓ)

$(H_0: \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0$ for all (i,j) and (k,ℓ) $)$

```
# A file with the S-PLUS commands is
# posted as  cane.ssc

# Enter the data.  Note that the first
# line of this file is a line of data,
# not a line of variable names.

> cane <- read.table("cane.dat",
    col.names=c("Variety","Nitrogen",
                "Yield"))

#  Create factors

> cane$V <- as.factor(cane$Variety)
> cane$N <- as.factor(cane$Nitrogen)

# Print the data frame

> cane
```



```
   Variety Nitrogen Yield   N V
 1       1      150  70.5 150 1
 2       1      150  67.5 150 1
 3       1      150  63.9 150 1
 4       1      150  64.2 150 1
 5       1      210  67.3 210 1
 6       1      210  75.9 210 1
 7       1      210  72.2 210 1
 8       1      210  60.5 210 1
 9       1      270  79.9 270 1
10       1      270  72.8 270 1
11       1      270  64.8 270 1
12       1      270  86.3 270 1
13       2      150  58.6 150 2
14       2      150  65.2 150 2
15       2      150  70.2 150 2
16       2      150  51.8 150 2
17       2      210  64.3 210 2
18       2      210  48.3 210 2
```

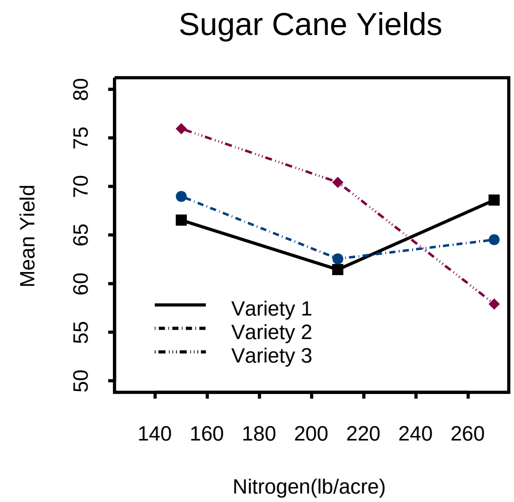


```
19       2      210  74.0 210 2
20       2      210  63.6 210 2
21       2      270  64.4 270 2
22       2      270  67.3 270 2
23       2      270  78.0 270 2
24       2      270  72.0 270 2
25       3      150  65.8 150 3
26       3      150  68.3 150 3
27       3      150  72.7 150 3
28       3      150  67.6 150 3
29       3      210  64.1 210 3
30       3      210  64.8 210 3
31       3      210  70.9 210 3
32       3      210  58.3 210 3
33       3      270  56.3 270 3
34       3      270  54.7 270 3
35       3      270  66.2 270 3
36       3      270  54.4 270 3
```



```
# Compute mean yields for all combinations
# of nitrogen levels and varieties and
# Make a profile plot. At this point
# UNIX users should open a graphics
# window with the  motif( ) function.

> means <- tapply(cane$Yield,
           list(cane$Variety,cane$Nitrogen),
              mean)
> means


     150    210    270
1 66.525 68.975 75.950
2 61.450 62.550 70.425
3 68.600 64.525 57.900
```

```
#  Set up the profile plot
> par(fin=c(7,7),cex=1.2,lwd=3,mex=1.5,mkh=.20)
> x.axis <- unique (cane$Nitrogen)
> matplot(c(130,270), c(50,80),
     type="n", xlab="Nitrogen(lb/acre)",
     ylab="Mean Yield",
     main= "Sugar Cane Yields")

#  Add a profile for each soil type
> matlines(x.axis,means,type='l',
               lty=c(1,3,5),lwd=3)

# Plot symbols for the sample means
> matpoints(x.axis,means, pch=c(15,16,18))

# Add a legend to the plot
> legend(130,60, legend=c('Variety 1',
          'Variety 2','Variety 3'),
          lty=c(1,3,5),bty='n')
```


```
# Fit a model with main effects and
# interaction effects. Compute both
# sets of Type I sums of squares.

options(contrasts=c('contr.sum','contr.ploy'))

> lm.out1 <- lm(Yield~N*V, data=cane)
> anova(lm.out1)

Analysis of Variance Table

Response: Yield

Terms added sequentially (first to last)
          Df Sum of Sq  Mean Sq F Value    Pr(F)
        N  2    56.541  28.2703 0.60847 0.551478
        V  2   319.374 159.6869 3.43698 0.046797
      N:V  4   559.788 139.9469 3.01211 0.035471
Residuals 27  1254.460  46.4615
```



```
> lm.out2 <- lm(Yield~V*N, data=cane)
> anova(lm.out2)

Analysis of Variance Table

Response: Yield

Terms added sequentially (first to last)
          Df Sum of Sq  Mean Sq F Value    Pr(F)
        V  2   319.374 159.6869 3.43698 0.046797
        N  2    56.541  28.2703 0.60847 0.551478
      V:N  4   559.788 139.9469 3.01211 0.035471
Residuals 27  1254.460  46.4615
```

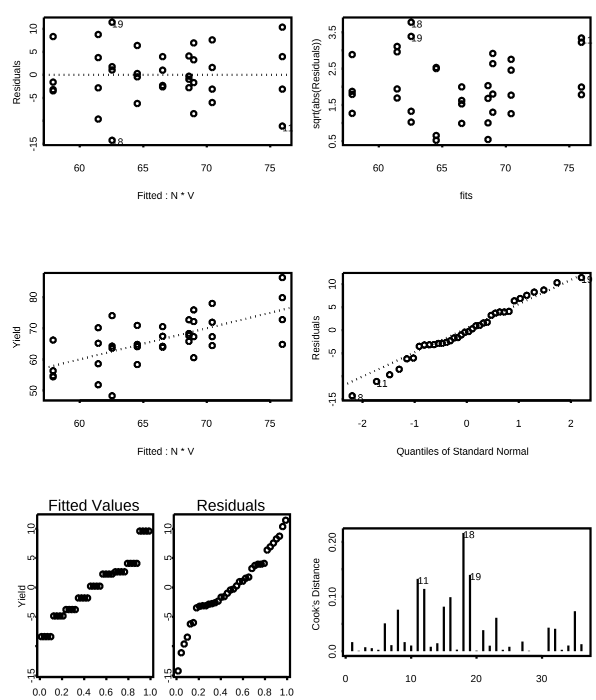
```
> summary(lm.out2, correlation=F)

Call: lm(formula = Yield ~ V * N, data = cane)
Residuals:
    Min     1Q  Median    3Q   Max
 -14.25 -3.131 -0.3625 3.956 11.45

Coefficients:           Std.
               Value   Error  t value  Pr(>|t|)
(Intercept)  66.3222  1.1360  58.3800   0.0000
         V1   4.1611  1.6066   2.5900   0.0153
         V2  -1.5139  1.6066  -0.9423   0.3544
         N1  -0.7972  1.6066  -0.4962   0.6238
         N2  -0.9722  1.6066  -0.6051   0.5501
       V1N1  -3.1611  2.2721  -1.3913   0.1755
       V2N1  -2.5611  2.2721  -1.1272   0.2696
       V1N2  -0.5361  2.2721  -0.2360   0.8152
       V2N2  -1.2861  2.2721  -0.5660   0.5760

Residual standard error: 6.816 on 27 df
Multiple R-Squared: 0.4272
F-statistic: 2.517 on 8 and 27 df,
             the p-value is 0.03462
```



```
> model.matrix(lm.out2)
   (Intercept) V1 V2 N1 N2 V1N1 V2N1 V1N2 V2N2
 1           1  1  0  1  0    1    0    0    0
 2           1  1  0  1  0    1    0    0    0
 3           1  1  0  1  0    1    0    0    0
 4           1  1  0  1  0    1    0    0    0
 5           1  1  0  0  1    0    0    1    0
 6           1  1  0  0  1    0    0    1    0
 7           1  1  0  0  1    0    0    1    0
 8           1  1  0  0  1    0    0    1    0
 9           1  1  0 -1 -1   -1    0   -1    0
10           1  1  0 -1 -1   -1    0   -1    0
11           1  1  0 -1 -1   -1    0   -1    0
12           1  1  0 -1 -1   -1    0   -1    0
13           1  0  1  1  0    0    1    0    0
14           1  0  1  1  0    0    1    0    0
15           1  0  1  1  0    0    1    0    0
16           1  0  1  1  0    0    1    0    0
17           1  0  1  0  1    0    0    0    1
18           1  0  1  0  1    0    0    0    1
```



```
19           1  0  1  0  1    0    0    0    1
20           1  0  1  0  1    0    0    0    1
21           1  0  1 -1 -1    0   -1    0   -1
22           1  0  1 -1 -1    0   -1    0   -1
23           1  0  1 -1 -1    0   -1    0   -1
24           1  0  1 -1 -1    0   -1    0   -1
25           1 -1 -1  1  0   -1   -1    0    0
26           1 -1 -1  1  0   -1   -1    0    0
27           1 -1 -1  1  0   -1   -1    0    0
28           1 -1 -1  1  0   -1   -1    0    0
29           1 -1 -1  0  1    0    0   -1   -1
30           1 -1 -1  0  1    0    0   -1   -1
31           1 -1 -1  0  1    0    0   -1   -1
32           1 -1 -1  0  1    0    0   -1   -1
33           1 -1 -1 -1 -1    1    1    1    1
34           1 -1 -1 -1 -1    1    1    1    1
35           1 -1 -1 -1 -1    1    1    1    1
36           1 -1 -1 -1 -1    1    1    1    1

# Create diagnostic plots
> par(mfrow=c(3,2))
> plot(lm.out1)
```

```
# Create a data frame containing the original
# data and the residuals and estimated means

> data.frame(cane$Nitrogen,cane$Variety,
        cane$Yield,Pred=lm.out1$fitted,
        Resid=round(lm.out1$resid,3))

    X1 X2   X3   Pred   Resid
 1 150  1 70.5 66.525   3.975
 2 150  1 67.5 66.525   0.975
 3 150  1 63.9 66.525  -2.625
 4 150  1 64.2 66.525  -2.325
 5 210  1 67.3 68.975  -1.675
 6 210  1 75.9 68.975   6.925
 7 210  1 72.2 68.975   3.225
 8 210  1 60.5 68.975  -8.475
 9 270  1 79.9 75.950   3.950
10 270  1 72.8 75.950  -3.150
11 270  1 64.8 75.950 -11.150
12 270  1 86.3 75.950  10.350
```



```
13 150  2 58.6 61.450  -2.850
14 150  2 65.2 61.450   3.750
15 150  2 70.2 61.450   8.750
16 150  2 51.8 61.450  -9.650
17 210  2 64.3 62.550   1.750
18 210  2 48.3 62.550 -14.250
19 210  2 74.0 62.550  11.450
20 210  2 63.6 62.550   1.050
21 270  2 64.4 70.425  -6.025
22 270  2 67.3 70.425  -3.125
23 270  2 78.0 70.425   7.575
24 270  2 72.0 70.425   1.575
25 150  3 65.8 68.600  -2.800
26 150  3 68.3 68.600  -0.300
27 150  3 72.7 68.600   4.100
28 150  3 67.6 68.600  -1.000
29 210  3 64.1 64.525  -0.425
30 210  3 64.8 64.525   0.275
31 210  3 70.9 64.525   6.375
32 210  3 58.3 64.525  -6.225
33 270  3 56.3 57.900  -1.600
34 270  3 54.7 57.900  -3.200
35 270  3 66.2 57.900   8.300
36 270  3 54.4 57.900  -3.500
```



```
# Compute Type III sums of squares and
# corresponding F-tests.

# Generate an identity matrix and a
# vector of ones

> Iden <- function(n) diag(rep(1,n))
> one  <- function(n) matrix(rep(1,n),ncol=1)

# Compute the transpose of the model
# matrix for the cell means model

> s  <- length(unique(cane$Nitrogen))
> t  <- length(unique(cane$Variety))
> st <- s*t
> r  <- length(cane$Yield)/(st)
> D <- t(kronecker(Iden(st), t(one(r))))
```



```
# Least squares estimation
> y <- matrix(cane$Yield,ncol=1)
> b <- solve(crossprod(D)) %*% crossprod(D,y)
> yhat <- D %*% b
> sse  <- crossprod(y-yhat)
> df2  <- nrow(y) - st

>c1 <- kronecker( cbind(Iden(s-1),-one(s-1)),
                 t(one(t)) )
> q1 <- t(b) %*% t(c1)%*% solve( c1 %*%
         solve(crossprod(D)) %*% t(c1))%*%
           c1 %*% b
> df1<- s-1
> f  <- (q1/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c1
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    1    1    0    0    0   -1   -1   -1
[2,]    0    0    0    1    1    1   -1   -1   -1

> data.frame(SS=q1,df=df1,F.stat=f,p.value=p)

        SS df   F.stat    p.value
1 319.3739  2 3.436975 0.04679743
```

```
> c2 <- kronecker( t(one(s)),
         cbind(Iden(t-1),-one(t-1)) )
> q2 <- t(b) %*% t(c2)%*%solve( c2 %*%
        solve(crossprod(D)) %*% t(c2))%*%
           c2 %*% b
> df1<- t-1
> f  <- (q2/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c2

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    0   -1    1    0   -1    1    0   -1
[2,]    0    1   -1    0    1   -1    0    1   -1


> data.frame(SS=q2,df=df1,F.stat=f,p.value=p)

        SS df   F.stat  p.value
1 56.54056  2 0.608467 0.551478
```



```
> c3 <- kronecker( cbind(Iden(s-1),-one(s-1)),
                cbind(Iden(t-1),-one(t-1)) )
> q3 <- t(b) %*% t(c3)%*% solve( c3 %*%
        solve(crossprod(D)) %*% t(c3))%*%
           c3 %*% b
> df1<- (s-1)*(t-1)
> f  <- (q3/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c3

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    0   -1    0    0    0   -1    0    1
[2,]    0    1   -1    0    0    0    0   -1    1
[3,]    0    0    0    1    0   -1   -1    0    1
[4,]    0    0    0    0    1   -1    0   -1    1

> data.frame(SS=q3,df=df1,F.stat=f,p.value=p)

        SS df   F.stat    p.value
1 559.7878  4 3.012107 0.03547072
```


Conclusions:

- **Variety 3 exhibits a "linear" decrease in yield as nitrogen increases from 150 lb/acre to 270 lb/acre.**
- **Varieties 1 and 2 exhibit parallel "linear" increasing trends in yield as nitrogen increases from 150 lb/acre to 270 lb/acre.**

- **Variety 1 appears to provide a consistently higher yield than Variety 2, but the difference in these two varieties is not "significant" at the .05 level.**
- **Variety 3 seems to do as well as Variety 1 at 150 lb/acre of nitrogen.**

```
/* Analysis of completely randomized
   factorial experiements with an
   application to the sugar cane data
   from Snedecor and Cochran. This
   program is posted as  cane.sas  */

data set1;
  infile 'cane.dat';
  input variety nitrogen yield;
  run;

/* Print the data */

proc print data=set1;
   var yield;
   run;

/* Compute an ANOVA table */
```



```
proc glm data=set1;
  class variety nitrogen;
  model yield = variety|nitrogen /
          p clm alpha=.05  ss1 ss2
          ss3 ss4 e e1 e2 e3 e4;
  output out=setr r=resid p=yhat;
  lsmeans variety*nitrogen / stderr pdiff;
  means variety nitrogen / tukey;
  contrast 'n-linear' nitrogen -1 0 1;
  contrast 'n-quad' nitrogen -1 2 -1;
  contrast 'v1-v2' variety 1 -1 0;
  contrast '(v1+v2)-v3' variety .5 .5 -1;
  contrast '(v1-v2)*(n-lin)' variety*nitrogen
                -1 0 1 1 0 -1 0 0 0;
  contrast '(v1-v2)*(n-quad)' variety*nitrogen
                -1 2 -1 1 -2 1 0 0 0;
  contrast '(.5(v1+v2)-v3)*(n-lin)'
       variety*nitrogen
             -.5 0 .5 -.5 0 .5 1 0 -1;
  contrast '(.5(v1+v2)-v3)*(n-quad)'
       variety*nitrogen
            -.5 1 -.5 -.5 1 -.5 1 -2 1;
```



```
  estimate 'n-linear' nitrogen -1 0 1;
  estimate 'n-quad' nitrogen -1 2 -1;
  estimate 'v1-v2' variety 1 -1 0;
  estimate '(v1+v2)-v3' variety .5 .5 -1;
  estimate '(v1-v2)*(n-lin)' variety*nitrogen
                    -1 0 1 1 0 -1 0 0 0;
  estimate '(v1-v2)*(n-quad)' variety*nitrogen
                    -1 2 -1 1 -2 1 0 0 0;
  estimate '(.5(v1+v2)-v3)*(n-lin)'
           variety*nitrogen
              -.5 0 .5 -.5 0 .5 1 0 -1;
  estimate '(.5(v1+v2)-v3)*(n-quad)'
           variety*nitrogen
              -.5 1 -.5 -.5 1 -.5 1 -2 1;
  run;
```



```
/* Make a profile plots for the interaction
   between varieties and nitrogen levels  */

/* UNIX users can use the following options */
/*  goptions cback=white colors=(black)
    targetdevice=ps300 rotate=landscape; */

/* Windows users can use the following */
goptions  cback=white colors=black
          device=WIN target=WINPRTC;


proc sort data=set1; by variety nitrogen;
proc means data=set1 noprint;
  by variety nitrogen;
  var yield;
  output out=means mean=my;
  run;
```

```
axis1 label=(f=swiss h=2.5)
      ORDER = 120 to 300 by 30
      value=(f=swiss h=2.0)  w=3.0
      length= 5.5 in;

axis2 label=(f=swiss h=2.0)
      order = 50 to 80 by 10
      value=(f=swiss h=2.0) w= 3.0
      length = 5.5 in;

 SYMBOL1 V=CIRCLE H=2.0 w=3 l=1 i=join ;
 SYMBOL2 V=DIAMOND H=2.0 w=3 l=3 i=join ;
 SYMBOL3 V=square H=2.0 w=3 l=9 i=join ;

 PROC GPLOT DATA=means;
   PLOT my*nitrogen=variety /
             vaxis=axis2 haxis=axis1;
   TITLE1  H=3.0 F=swiss "Sugar Cane Yields";
       LABEL my='Mean Yield';
       LABEL nitrogen = 'Nitrogen (lb/acre)';
   RUN;
```


General Form of Estimable Functions

Effect			Coefficients
Intercept			L1
variety	1		L2
variety	2		L3
variety	3		L1-L2-L3
nitrogen		150	L5
nitrogen		210	L6
nitrogen		270	L1-L5-L6
variety*nitrogen	1	150	L8
variety*nitrogen	1	210	L9
variety*nitrogen	1	270	L2-L8-L9
variety*nitrogen	2	150	L11
variety*nitrogen	2	210	L12
variety*nitrogen	2	270	L3-L11-L12
variety*nitrogen	3	150	L5-L8-L11
variety*nitrogen	3	210	L6-L9-L12
variety*nitrogen	3	270	L1-L2-L3-L5-L6+L8 +L9+L11+L12

Type I Estimable Functions

Effect			variety
Intercept			0
variety	1		L2
variety	2		L3
variety	3		-L2-L3
nitrogen		150	0
nitrogen		210	0
nitrogen		270	0
variety*nitrogen	1	150	0.3333*L2
variety*nitrogen	1	210	0.3333*L2
variety*nitrogen	1	270	0.3333*L2
variety*nitrogen	2	150	0.3333*L3
variety*nitrogen	2	210	0.3333*L3
variety*nitrogen	2	270	0.3333*L3
variety*nitrogen	3	150	-0.3333*L2-0.3333*L3
variety*nitrogen	3	210	-0.3333*L2-0.3333*L3
variety*nitrogen	3	270	-0.3333*L2-0.3333*L3

Type I Estimable Functions

			------------Coefficients-------	
Effect			nitrogen	variety*nitrogen
Intercept			0	0
variety	1		0	0
variety	2		0	0
variety	3		0	0
nitrogen		150	L5	0
nitrogen		210	L6	0
nitrogen		270	-L5-L6	0
variety*nitrogen	1	150	0.3333*L5	L8
variety*nitrogen	1	210	0.3333*L6	L9
variety*nitrogen	1	270	-0.3333*L5-0.3333*L6	-L8-L9
variety*nitrogen	2	150	0.3333*L5	L11
variety*nitrogen	2	210	0.3333*L6	L12
variety*nitrogen	2	270	-0.3333*L5-0.3333*L6	-L11-L12
variety*nitrogen	3	150	0.3333*L5	-L8-L11
variety*nitrogen	3	210	0.3333*L6	-L9-L12
variety*nitrogen	3	270	-0.3333*L5-0.3333*L6	L8+L9+L11+L12

Type II Estimable Functions

Effect		----Coefficients---- variety
Intercept		0
variety	1	L2
variety	2	L3
variety	3	-L2-L3
nitrogen	150	0
nitrogen	210	0
nitrogen	270	0
variety*nitrogen	1 150	0.3333*L2
variety*nitrogen	1 210	0.3333*L2
variety*nitrogen	1 270	0.3333*L2
variety*nitrogen	2 150	0.3333*L3
variety*nitrogen	2 210	0.3333*L3
variety*nitrogen	2 270	0.3333*L3
variety*nitrogen	3 150	-0.3333*L2-0.3333*L3
variety*nitrogen	3 210	-0.3333*L2-0.3333*L3
variety*nitrogen	3 270	-0.3333*L2-0.3333*L3

Type II Estimable Functions

Effect		-------------Coefficients---------	
		nitrogen	variety*nitrogen
Intercept		0	0
variety	1	0	0
variety	2	0	0
variety	3	0	0
nitrogen	150	L5	0
nitrogen	210	L6	0
nitrogen	270	-L5-L6	0
variety*nitrogen	1 150	0.3333*L5	L8
variety*nitrogen	1 210	0.3333*L6	L9
variety*nitrogen	1 270	-0.3333*L5-0.3333*L6	-L8-L9
variety*nitrogen	2 150	0.3333*L5	L11
variety*nitrogen	2 210	0.3333*L6	L12
variety*nitrogen	2 270	-0.3333*L5-0.3333*L6	-L11-L12
variety*nitrogen	3 150	0.3333*L5	-L8-L11
variety*nitrogen	3 210	0.3333*L6	-L9-L12
variety*nitrogen	3 270	-0.3333*L5-0.3333*L6	L8+L9+L11+L12

Type III Estimable Functions

Effect		----Coefficients---- variety
Intercept		0
variety	1	L2
variety	2	L3
variety	3	-L2-L3
nitrogen	150	0
nitrogen	210	0
nitrogen	270	0
variety*nitrogen	1 150	0.3333*L2
variety*nitrogen	1 210	0.3333*L2
variety*nitrogen	1 270	0.3333*L2
variety*nitrogen	2 150	0.3333*L3
variety*nitrogen	2 210	0.3333*L3
variety*nitrogen	2 270	0.3333*L3
variety*nitrogen	3 150	-0.3333*L2-0.3333*L3
variety*nitrogen	3 210	-0.3333*L2-0.3333*L3
variety*nitrogen	3 270	-0.3333*L2-0.3333*L3

Type III Estimable Functions

Effect		-------------Coefficients--------	
		nitrogen	variety*nitrogen
Intercept		0	0
variety	1	0	0
variety	2	0	0
variety	3	0	0
nitrogen	150	L5	0
nitrogen	210	L6	0
nitrogen	270	-L5-L6	0
variety*nitrogen	1 150	0.3333*L5	L8
variety*nitrogen	1 210	0.3333*L6	L9
variety*nitrogen	1 270	-0.3333*L5-0.3333*L6	-L8-L9
variety*nitrogen	2 150	0.3333*L5	L11
variety*nitrogen	2 210	0.3333*L6	L12
variety*nitrogen	2 270	-0.3333*L5-0.3333*L6	-L11-L12
variety*nitrogen	3 150	0.3333*L5	-L8-L11
variety*nitrogen	3 210	0.3333*L6	-L9-L12
variety*nitrogen	3 270	-0.3333*L5-0.3333*L6	L8+L9+L11+L12

Type IV Estimable Functions

Effect			----Coefficients---- variety
Intercept			0
variety	1		L2
variety	2		L3
variety	3		-L2-L3
nitrogen		150	0
nitrogen		210	0
nitrogen		270	0
variety*nitrogen	1	150	0.3333*L2
variety*nitrogen	1	210	0.3333*L2
variety*nitrogen	1	270	0.3333*L2
variety*nitrogen	2	150	0.3333*L3
variety*nitrogen	2	210	0.3333*L3
variety*nitrogen	2	270	0.3333*L3
variety*nitrogen	3	150	-0.3333*L2-0.3333*L3
variety*nitrogen	3	210	-0.3333*L2-0.3333*L3
variety*nitrogen	3	270	-0.3333*L2-0.3333*L3

Type IV Estimable Functions

Effect			-------------Coefficients------- nitrogen	variety*nitrogen
Intercept			0	0
variety	1		0	0
variety	2		0	0
variety	3		0	0
nitrogen		150	L5	0
nitrogen		210	L6	0
nitrogen		270	-L5-L6	0
variety*nitrogen	1	150	0.3333*L5	L8
variety*nitrogen	1	210	0.3333*L6	L9
variety*nitrogen	1	270	-0.3333*L5-0.3333*L6	-L8-L9
variety*nitrogen	2	150	0.3333*L5	L11
variety*nitrogen	2	210	0.3333*L6	L12
variety*nitrogen	2	270	-0.3333*L5-0.3333*L6	-L11-L12
variety*nitrogen	3	150	0.3333*L5	-L8-L11
variety*nitrogen	3	210	0.3333*L6	-L9-L12
variety*nitrogen	3	270	-0.3333*L5-0.3333*L6	L8+L9+L11+L12

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F	Pr > F
Model	8	935.7022	116.9628	2.52	0.0346
Error	27	1254.4600	46.4615		
C. Total	35	2190.1622			

Source	DF	Type I SS	Mean Square	F	Pr > F
variety	2	319.3739	159.6869	3.44	0.0468
nitrogen	2	56.5406	28.2703	0.61	0.5515
var*nit	4	559.7878	139.9469	3.01	0.0355

Source	DF	Type II SS	Mean Square	F	Pr > F
variety	2	319.3739	159.6869	3.44	0.0468
nitrogen	2	56.5406	28.2703	0.61	0.5515
var*nit	4	559.7878	139.9469	3.01	0.0355

Source	DF	Type III SS	Mean Square	F	Pr > F
variety	2	319.3739	159.6869	3.44	0.0468
nitrogen	2	56.5406	28.2703	0.61	0.5515
var*nit	4	559.7878	139.9469	3.01	0.0355

Source	DF	Type IV SS	Mean Square	F	Pr > F
variety	2	319.3739	159.6869	3.44	0.0468
nitrogen	2	56.5406	28.2703	0.61	0.5515
var*nit	4	559.7878	139.9469	3.01	0.0355

Least Squares Means

variety	nitrogen	LSMEAN yield	Standard Error	Pr > |t|
1	150	66.525	3.408133	<.0001
1	210	68.975	3.408133	<.0001
1	270	75.950	3.408133	<.0001
2	150	61.450	3.408133	<.0001
2	210	62.550	3.408133	<.0001
2	270	70.425	3.408133	<.0001
3	150	68.600	3.408133	<.0001
3	210	64.525	3.408133	<.0001
3	270	57.900	3.408133	<.0001

Least Squares Means for effect variety*nitrogen
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: yield

i/j	1	2	3	4	5
1		0.6154	0.0610	0.3017	0.4168
2	0.6154		0.1594	0.1301	0.1937
3	0.0610	0.1594		0.0056	0.0098
4	0.3017	0.1301	0.0056		0.8212
5	0.4168	0.1937	0.0098	0.8212	
6	0.4255	0.7658	0.2617	0.0735	0.1139
7	0.6702	0.9386	0.1389	0.1495	0.2202
8	0.6815	0.3640	0.0252	0.5289	0.6852
9	0.0848	0.0296	0.0009	0.4678	0.3432

i/j	6	7	8	9
1	0.4255	0.6702	0.6815	0.0848
2	0.7658	0.9386	0.3640	0.0296
3	0.2617	0.1389	0.0252	0.0009
4	0.0735	0.1495	0.5289	0.4678
5	0.1139	0.2202	0.6852	0.3432
6		0.7079	0.2315	0.0150
7	0.7079		0.4053	0.0350
8	0.2315	0.4053		0.1806
9	0.0150	0.0350	0.1806	

Contrast	DF	Contrast SS	Mean Square
n-linear	1	39.5266667	39.5266667
n-quad	1	17.0138889	17.0138889
v1-v2	1	193.2337500	193.2337500
(v1+v2)-v3	1	126.1401389	126.1401389
(v1-v2)*(n-lin)	1	0.2025000	0.2025000
(v1-v2)*(n-quad)	1	1.6875000	1.6875000
(.5(v1+v2)-v3)*(n-lin)	1	528.0133333	528.0133333
(.5(v1+v2)-v3)*(n-quad)	1	29.8844444	29.8844444

Contrast	F Value	Pr > F
n-linear	0.85	0.3645
n-quad	0.37	0.5501
v1-v2	4.16	0.0513
(v1+v2)-v3	2.71	0.1110
(v1-v2)*(n-lin)	0.00	0.9478
(v1-v2)*(n-quad)	0.04	0.8503
(.5(v1+v2)-v3)*(n-lin)	11.36	0.0023
(.5(v1+v2)-v3)*(n-quad)	0.64	0.4296

Parameter	Estimate	Standard Error	t Value
n-linear	2.5666667	2.7827289	0.92
n-quad	-2.9166667	4.8198279	-0.61
v1-v2	5.6750000	2.7827289	2.04
(v1+v2)-v3	3.9708333	2.4099139	1.65
(v1-v2)*(n-lin)	0.4500000	6.8162659	0.07
(v1-v2)*(n-quad)	2.2500000	11.8061189	0.19
(.5(v1+v2)-v3)*(n-lin)	19.9000000	5.9030595	3.37
(.5(v1+v2)-v3)*(n-quad)	-8.2000000	10.2243989	-0.80

Parameter	Pr > |t|
n-linear	0.3645
n-quad	0.5501
v1-v2	0.0513
(v1+v2)-v3	0.1110
(v1-v2)*(n-lin)	0.9478
(v1-v2)*(n-quad)	0.8503
(.5(v1+v2)-v3)*(n-lin)	0.0023
(.5(v1+v2)-v3)*(n-quad)	0.4296

Two factor experiments with empty cells

Data from Littell, Freund, and Spector, 1991, ***SAS System for Linear Models,*** **3rd edition, SAS Institute, Cary, N.C.**

Factor A	Factor B $j=1$	$j=2$	$j=3$
$i=1$	$Y_{111}=5$ $Y_{112}=6$	$Y_{121}=2$ $Y_{122}=3$ $Y_{123}=5$ $Y_{124}=6$ $Y_{125}=7$	–
$i=2$	$Y_{211}=2$ $Y_{212}=3$	$Y_{221}=8$ $Y_{222}=8$ $Y_{223}=9$	$Y_{231}=4$ $Y_{232}=4$ $Y_{233}=6$ $Y_{234}=6$ $Y_{235}=7$

Sample sizes:

Factor A	Factor B $j=1$	$j=2$	$j=3$
$i=1$	$n_{11}=2$	$n_{12}=5$	–
$i=2$	$n_{21}=2$	$n_{22}=3$	$n_{23}=5$

Effects model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

for $(i,j) \neq (1,3)$
and $k = 1, \ldots, n_{ij}$

$$\begin{aligned} \mu_{ij} &= E(\bar{Y}_{ij.}) \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij} \end{aligned}$$

is estimable for all $(i,j) \neq (1,3)$.

Functions of parameters that are <u>not</u> estimable include:

$$\mu_{13} = \mu + \alpha_1 + \beta_3 + \gamma_{13}$$

$$\begin{aligned} \bar{\mu}_{..} &= \frac{1}{6}\sum_{i=1}^{2}\sum_{j=1}^{3}\mu_{ij} \\ &= \mu + \frac{1}{2}(\alpha_1+\alpha_2) + \frac{1}{3}(\beta_1+\beta_2+\beta_3) \\ &\quad + \frac{1}{6}(\gamma_{11}+\gamma_{12}+\gamma_{13}+\gamma_{21}+\gamma_{22}+\gamma_{23}). \end{aligned}$$

$$\bar{\mu}_{1.} = \frac{1}{3}\sum_{j=1}^{3}\mu_{1j}$$

$$\bar{\mu}_{.3} = \frac{1}{2}(\mu_{13}+\mu_{23})$$

Two factor classifications with empty cells:

- No single "best" or "correct" analysis.
- Analysis of variance
  - Test for interaction is useful
  - Use SSE to estimate the error variance σ^2.
  - Tests for "main effects" may not be meaningful, especially in the presence of interaction.

- Compute F-tests and sums of squares for meaningful contrasts.
- Compare estimated means for different combinations of factor levels.
- Consider the combinations of factor levels as levels of a single "combined" factor.
  - one-way ANOVA
  - contrasts
  - compare means


```
/* SAS code for analyzing data
   from the two factor experiment
   with no data for one combination
   of factors>  This code is posted
   as  littell.sas   */

data set1;
  input A B y;
  cards;
 1  1   5
 1  1   6
 1  2   2
 1  2   3
 1  2   5
 1  2   6
 1  2   7
```

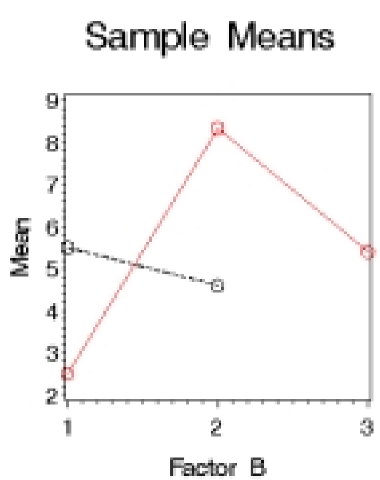


```
 2  1   2
 2  1   3
 2  2   8
 2  2   8
 2  2   9
 2  3   4
 2  3   4
 2  3   6
 2  3   6
 2  3   7
run;

/*  Print the data  */

proc print data=set1;
run;

/* Compute sample means for all
   factor combinations with data.
   Make a profile plot. */
```

```
proc sort data=set1; by a b;
proc means data=set1 noprint; by a b;
  var Y;
  output out=means mean=my;
  run;

goptions  cback=white colors=black
          device=WIN target=WINPRTC;

/*
 goptions cback=white colors=(black)
    targetdevice=ps300 rotate=landscape;
*/

axis1 label=(f=swiss h=2.0)
      value=(f=swiss h=1.8)
      w=3.0  length= 5.0 in;

axis2 label=(f=swiss h=2.0 a=90 r=0)
      value=(f=swiss h=1.8)
      w= 3.0  length = 5.0 in;
```



```
SYMBOL1 V=circle H=2.0 w=3 l=1 i=join;
SYMBOL2 V=diamond H=2.0 w=3 l=3 i=join;

proc gplot data=means;
  plot my*b=a / vaxis=axis2 haxis=axis1;
  title ls=0.8in H=3.0 F=swiss "Sample Means";
       label my='Mean';
       label b = 'Factor B';
  footnote ls=0.4in '  ';
 run;


/* Perform analysis of variance where
   facror A is entered into the model
   before factor B.  Use the LSMEANS
   statement to compare means for
   different combinations of factor A
   and factor B.  */
```


```
proc glm data=set1;
  class A B;
  model y = A B A*B / solution ss1 ss2
            ss3 ss4 e e1 e2 e3 e4 p;
  means A B A*B;
  lsmeans A*B / pdiff tdiff stderr;
    estimate 'A1-A2' A 1 -1 / e;
    contrast 'A1-A2' A 1 -1 / e;
    estimate 'A1-A2 within B1'  A 1 -1
              A*B 1 0  -1 0 0 / e;
    estimate 'A1-A2 within B2'  A 1 -1
              A*B 0 1 0 -1 0 / e;
    estimate 'A1-A2 over B'  A 1 -1
              A*B .5 .5 -.5 -.5 0 / e;
    estimate 'B1-B2 over A'  B 1 -1 0
              A*B .5 -.5 .5 -.5 0 / e;
    estimate 'B3-.5(B1+B2) in A2' B -.5 -.5 1
              A*B 0 0 -.5 -.5 1 / e;
    estimate 'interaction' A*B 1 -1 -1 1 0 / e;
    run;
```

```
/*  Do everything with a one-factor ANOVA by
    combining the two factors into a single
    factor with 5 categories.  */

data set1; set set1;
  C=10*A+B;
  run;

proc glm data=set1;
  class C;
  model y = C / solution e e2;
  estimate 'C11-C21' C 1 0 -1 0 0;
  estimate 'C12-C22' C 0 1 0 -1 0;
  estimate '.5(C11+C12-C21+C22)'
               C .5 .5 -.5 -.5 0;
  estimate '.5(C11-C12+C21-C22)'
               C .5 -.5 .5 -.5 0;
  estimate 'C23-.5(C21+C22)' C 0 0 -.5 -.5 1;
  estimate 'C11-C12-C21+C22' C 1 -1 -1 1 0;
  lsmeans C / stderr tdiff pdiff;
  run;
```


General Form of Estimable Functions

Effect		Coefficients
Intercept		L1
A	1	L2
A	2	L1-L2
B	1	L4
B	2	L5
B	3	L1-L4-L5
A*B	1 1	L7
A*B	1 2	L2-L7
A*B	2 1	L4-L7
A*B	2 2	-L2+L5+L7
A*B	2 3	L1-L4-L5

Type III Estimable Functions

		-----------Coefficients-----------		
Effect		A	B	A*B
Intercept		0	0	0
A	1	L2	0	0
A	2	-L2	0	0
B	1	0	L4	0
B	2	0	L5	0
B	3	0	-L4-L5	0
A*B	1 1	0.5*L2	0.25*L4-0.25*L5	L7
A*B	1 2	0.5*L2	-0.25*L4+0.25*L5	-L7
A*B	2 1	-0.5*L2	0.75*L4+0.25*L5	-L7
A*B	2 2	-0.5*L2	0.25*L4+0.75*L5	L7
A*B	2 3	0	-L4-L5	0

Type IV Estimable Functions

		------Coefficients------		
Effect		A	B	A*B
Intercept		0	0	0
A	1	L2	0	0
A	2	-L2	0	0
B	1	0	L4	0
B	2	0	L5	0
B	3	0	-L4-L5	0
A*B	1 1	0.5*L2	0	L7
A*B	1 2	0.5*L2	0	-L7
A*B	2 1	-0.5*L2	L4	-L7
A*B	2 2	-0.5*L2	L5	L7
A*B	2 3	0	-L4-L5	0

NOTE: Other Type IV estimable functions exist.

General Form of Estimable Functions

Effect		Coefficients
Intercept		L1
C	11	L2
C	12	L3
C	21	L4
C	22	L5
C	23	L1-L2-L3-L4-L5

Type II Estimable Functions

Effect		-Coefficients- C
Intercept		0
C	11	L2
C	12	L3
C	21	L4
C	22	L5
C	23	-L2-L3-L4-L5

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	45.8157	11.4539	5.27	0.0110
Error	12	26.0667	2.1722		
C. Total	16	71.8824			

Parameter	Estimate	Standard Error	t	Pr > |t|
C11-C21	3.0000	1.4738	2.04	0.0645
C12-C22	-3.7333	1.0763	-3.47	0.0046
.5(C11+C12-C21+C22)	-0.3667	0.9125	-0.40	0.6949
.5(C11-C12+C21-C22)	-2.4667	0.9125	-2.70	0.0192
C23-.5(C21+C22)	0.0167	0.9418	-0.02	0.9862
C11-C12-C21+C22	6.7333	1.8250	3.69	0.0031

Least Squares Means

C	LSMEAN y	Standard Error	Pr > |t|	LSMEAN Number
11	5.5000	1.0421	0.0002	1
12	4.6000	0.6591	<.0001	2
21	2.5000	1.0422	0.0336	3
22	8.3333	0.8509	<.0001	4
23	5.4000	0.6591	<.0001	5

Least Squares Means for Effect C
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: y

i/j	1	2	3	4	5
1		0.7299	2.0355	-2.1059	0.0811
		0.4795	0.0645	0.0569	0.9367
2	-0.7299		1.7030	-3.4685	-0.8582
	0.4795		0.1143	0.0046	0.4076
3	-2.0355	-1.70301		-4.3357	-2.3518
	0.0645	0.1143		0.0010	0.0366
4	2.1059	3.46853	4.3357		2.7253
	0.0569	0.0046	0.0010		0.0184
5	-0.0811	0.85824	2.3518	-2.7253	
	0.9367	0.4076	0.0366	0.0184	

Estimable functions for Type IV sums of squares may depend on

- location of empty cells
- ordering of the levels for the row and column factors

<u>Example:</u> Exchange columns 1 and 3 in the previous example.

	Factor 2		
Factor 1	A (old j=3)	B	C (old j=1)
$i = 1$	$-$	$\bar{Y}_{12.} = 4.6$ $n_{12} = 5$	$\bar{Y}_{13} = 5.5$ $n_{13} = 2$
$i = 2$	$\bar{Y}_{21.} = 5.4$ $n_{21} = 5$	$\bar{Y}_{22} = 8.33$ $n_{22} = 3$	$\bar{Y}_{23.} = 2.5$ $n_{23} = 2$

Type IV estimable functions for Factor B:

	Main Effects				Interaction		
	A	B	C		A	B	C
$i=1$	–	0	0	$i=1$	–	.5	-.5
$i=2$	1	0	-1	$i=2$	0	.5	-.5

$$\mu_{2A} - \mu_{2C} \qquad \frac{1}{2}(\mu_{1B} + \mu_{2B}) - \frac{1}{2}(\mu_{1C} + \mu_{2C})$$

In either case, Type IV sums of squares and testable functions are not the same as Type III sums of squares and testable functions.

Additive model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

$$i = 1, \ldots, a$$

$$j = 1, \ldots, b$$

$$k + 1, \ldots, n_{ij}$$

For this model

$$E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j$$

may be estimable when $n_{ij} = 0$.

For example 8.1, $n_{13} = 0$, but

$$\begin{aligned} \mu_{13} &= \mu + \alpha_1 + \beta_3 \\ &= (\mu + \alpha_2 + \beta_3) - (\mu + \alpha_2 + \beta_2) \\ &\quad + (\mu + \alpha_1 + \beta_2) \\ &= \mu_{23} + (\mu_{12} - \mu_{22}) \\ &= E(\bar{Y}_{23.} - \bar{Y}_{22.} + \bar{Y}_{12.}) \end{aligned}$$

Summary

Sum of Squares	Associated null hypothesis
$R(\mu)$	$H_0: \mu + \sum_{i=1}^{a} \frac{n_{i.}}{n_{..}} \alpha_i + \sum_{j=1}^{b} \frac{n_{.j}}{n_{..}} \beta_j = 0$
	(or $H_0: \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{n_{ij}}{n_{..}} \mu_{ij} = 0$)
$R(\alpha \mid \mu)$	$H_0: \alpha_i + \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}} \beta_j$ are equal for all $i = 1, \ldots, a$
	(or $H_0: \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}} \mu_{ij}$ are equal for all $i = 1, \ldots, a$)
$R(\beta \mid \mu, \alpha)$	$H_0: \beta_j$ are equal for all $j = 1, \ldots, b$

Sum of Squares	Associated null hypothesis
$R(\mu)$	$H_0 : \mu + \sum_{i=1}^{a} \frac{n_{i.}}{n_{..}} \alpha_i + \sum_{j=1}^{b} \frac{n_{.j}}{n_{..}} \beta_j = 0$
	(or $H_0 : \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{n_{ij}}{n_{..}} \mu_{ij} = 0$)
$R(\beta \mid \mu)$	$H_0 : \beta_j + \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} \alpha_i$ are equal for all $j = 1, \ldots, b$
	(or $H_0 : \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}} \mu_{ij} areequal$ for all $j = 1, \ldots, b$)
$R(\alpha \mid \mu, \beta)$	$H_0 : \alpha_i$ are equal for all $i = 1, \ldots, a$