

# Ponto 13 (Teórica) - Delineamentos Experimentais

Ben Dêivide

20 de Setembro de 2017

## 1 Conceitos iniciais

Quando se realiza uma pesquisa científica, estamos interessados em saber quais as características que influenciam no objetivo da pesquisa. Logicamente, estamos tentando entender as características de uma população, por meio de uma amostra, que influenciam os seus elementos dentro do objetivo da pesquisa. Para isso, realizamos um experimento. Para nossa proposta, definimos

**Definição 1** (Experimento). *O procedimento para realização de inferências sobre as características que supostamente influenciam no desempenho das unidades experimentais, é chamado de Experimento.*

As características que estão presentes nas unidades em um experimento podem ser ilustradas na Figura 1.

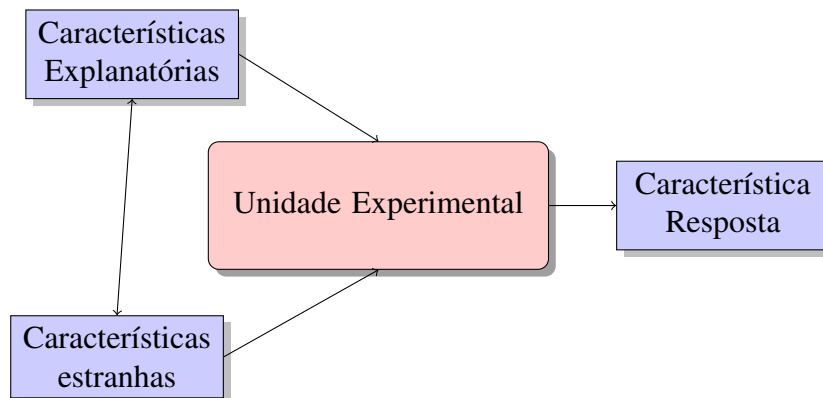


Figura 1: Variáveis que envolvem a unidade experimental em um experimento.

As características que exprimem o desempenho das unidades experimentais, é chamada de característica explanatória. Esse desempenho é visto por meio da variável resposta, coleta ao final ou ao longo do experimento. Para a análise de experimentos, as características explanatórias serão as características controladas de interesse da pesquisa pelo pesquisador, impondo níveis dessas características as unidades experimentais, para saber o quanto que níveis influenciam no desempenho da característica resposta. Para o nosso estudo, chamaremos essa característica de fator experimental.

**Definição 2** (Fator Experimental de tratamento). *A característica explanatória cujos níveis são atribuídos às unidades experimentais para verificar a influência destas no desempenho da variável resposta, é chamada de fator experimental de tratamento ou simplesmente fator.*

Experimentos que apresentam apenas um fator experimental de tratamento, dizemos que temos um experimento unifatorial. Quando se apresenta mais de um fator, temos experimentos fatoriais.

**Definição 3** (Tratamento). *Considere um experimento  $k$  fatores cada um com  $n$  níveis. Se  $k = 1$ , definimos os  $n$  níveis desse fator de tratamento. Se  $k > 1$ , a combinação dos  $n$  níveis de todos os fatores também será definida como tratamento.*

As características respostas ou variáveis respostas, representam o desempenho da unidade experimental. Quando este é devido ao tratamento, dizemos que esse desempenho representa o efeito do tratamento exercido na unidade experimental proporcional uma variação na variável resposta. As características respostas e explanatórias representam o objetivo da pesquisa, de modo que estas são escolhidas com bases científicas ou empíricas. Apesar de podermos ter diversas variáveis respostas, apenas uma será analisada de cada vez com as variáveis explanatórias.

**Definição 4** (Unidade Experimental). *A unidade experimental é a maior fração no experimento que recebe aleatoriamente um tratamento dos fatores envolvidos.*

Na área agropecuária também é costume chamar a unidade experimental de parcela. Contudo, o pesquisador nem sempre tem total controle das características (ou variáveis) envolvidas nas unidades experimentais. Estas são chamadas de características estranhas. São determinadas por exclusão, sendo características que não refletem no desempenho da unidade experimental, do qual não faz parte do objetivo da pesquisa.

**Definição 5** (Fator experimental intrínseco). *O fator experimental cuja a manifestação é inerente a unidade experimental devido as características estranhas, é chamado de fator experimental intrínseco.*

As três classes de características envolvidas na unidade experimental são definidas a seguir.

**Definição 6** (Material Experimental). *A característica explanatória, característica estranha e característica resposta representam o material experimental.*

O reconhecimento do material experimental é fundamental para a consecução do objetivo da pesquisa. Para isso, é fundamental o planejamento da resposta, das condições experimental<sup>1</sup>, do controle dos fatores intrínsecos e o delineamento experimental. Este último é o objetivo dessa dissertação que será detalhada mais a frente.

Como observado na Figura 1, na unidade experimental ocorre um confundimento entre o fator (característica explanatória) e a variável resposta (característica estranha). Esse confundimento pode ser entendido como erro experimental.

**Definição 7** (Erro Experimental). *Em um experimento, o desempenho da variável resposta que é atribuível aos fatores intrínsecos é chamado de erro experimental.*

O pesquisador poderá até controlar esse confundimento, porém, por mais que se exerça o controle nos fatores intrínsecos, é inevitável o confundimento.

**Definição 8** (Controle Experimental). *Em um experimento, todas as ações realizadas sobre as unidades experimentais exercidas pelo pesquisador para evitar o erro experimental, denominamos de controle experimental.*

---

<sup>1</sup>Os tratamentos para o experimento unifatorial e fatorial representam as condições experimentais.

O que se deseja é que esse confundimento seja irrelevante para o desempenho da variável resposta, e assim, garantir que o desempenho das variáveis respostas sejam devidas ao efeito dos tratamentos quando estes diferirem entre si.

O controle experimental pode ser implantado no experimento usando os seguintes processos: controle de técnicas experimentais, controle estatístico, controle local e casualização. Claro que nem todos os fatores intrínsecos são identificados para a execução do controle local. Por isso, deve-se fazer a identificação das possíveis características estranhas no experimento e assim, realizar o procedimento do controle experimental.

**Definição 9** (Controle de técnicas experimentais - CTE). *O controle físico exercido na unidade experimental com o propósito de diminuir o erro experimental é chamado de Controle de técnicas estatísticas.*

Exemplos de CTE são apresentados a seguir.

**Exemplo 1.** *São exemplos de CTE: uso de sementes sadias com vigor homogêneo; condução de experimento no campo com terreno plano e uniforme quanto as características de solo; técnicas de cultivo uniforme; controle homogêneo de incidências de pragas e doenças.*

Esse tipo de controle pode ter implicações na amostra, uma vez que as características estranhas são eliminadas da amostra, e isso contribui em prejuízos para representação da população em estudo. Observe que esse controle afeta a manifestação das características estranhas.

Um outro tipo de controle experimental é o controle estatístico. Esse tipo de controle não afeta a manifestação das características estranhas.

**Definição 10** (Controle Estatístico). *A correção feita no desempenho da variável resposta devido a valores observados de uma ou mais fatores experimentais intrínsecos nas unidades experimentais, que comumente chamamos estes fatores de covariáveis, denominamos de Controle estatístico.*

Esse controle resulta em um tipo de análise estatística chamada de análise de covariância, sendo utilizado quando os níveis dos fatores intrínsecos são heterogêneos e quantitativos. Quando realizamos o procedimento do controle estatístico a análise covariância resulta na junção de dois tipos de análise: análise de regressão e análise de variância (ANAVA).

Geralmente, esse tipo de controle pode ser realizado juntamente com o controle local. O Controle local bem como a casualização farão parte do que chamamos dos princípios básicos da experimentação. Estes princípios foram desenvolvidos por Fisher em 1930, com o intuito do experimento apresentar conclusões válidas para a tomada de decisão sobre as hipóteses estatísticas.

## 2 Delineamentos Experimentais

### 2.1 Delineamento Inteiramente Casualizado (DIC)

#### 2.1.1 Análise de Variância

Considere  $I$  tratamentos de um experimento unifatorial com  $n_i$  repetições, para  $i = 1, 2, \dots, I$ . A variável resposta  $Y_{ij}$  representa a  $j$ -ésima variável aleatória no  $i$ -ésimo tratamento. A Tabela 1 apresenta as realizações de  $Y_{ij}$  em um determinado experimento unifatorial. Pela notação con-

Tabela 1: Observações em um experimento unifatorial no desenho (DIC).

<b>Tratamento</b>					<b>Total</b>	<b>Médias</b>
<b>(Nível)</b>	<b>Realizações (Observações)</b>					
1	$y_{11}$	$y_{12}$	$\dots$	$y_{1n_1}$	$y_{1.}$	$\bar{y}_1.$
2	$y_{21}$	$y_{22}$	$\dots$	$y_{2n_2}$	$y_{2.}$	$\bar{y}_2.$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\vdots$
$I$	$y_{I1}$	$y_{I2}$	$\dots$	$y_{In_I}$	$y_{I.}$	$\bar{y}_I.$
<b>Total</b>					$y_{..}$	$\bar{y}_{..}$

sideraremos  $n_i$  o número de repetições das observações no  $i$ -ésimo tratamento e  $N = \sum_{i=1}^I n_i$ , em que segue as demais notações:

$$y_i = \sum_{j=1}^{n_i} y_{ij} \quad (1)$$

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \quad (2)$$

$$y_{..} = \sum_{i=1}^I y_i = \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} \quad (3)$$

$$\bar{y}_{..} = \frac{y_{..}}{N} \quad (4)$$

O modelo de médias de caselas para esses dados pode ser descrito da seguinte forma:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad (5)$$

tal que  $\mu_i$  é a média populacional desconhecida do  $i$ -ésimo tratamento e  $\epsilon_{ij}$  é o erro aleatório (variável aleatória) na  $j$ -ésima observação no  $i$ -ésimo tratamento que incorpora todas as outras variabilidades do experimento atribuído aos fatores intrínsecos. Assumimos também que

$$E[\epsilon_{ij}] = 0, \quad Var[\epsilon_{ij}] = \sigma^2 \text{ e } Cov[\epsilon_{ij}, \epsilon_{hk}] = 0, \quad \forall i \neq h, \forall j \neq k, \quad \epsilon_{ij} \sim N(0, \sigma^2) \quad (6)$$

Como consequência de (6), temos

$$E[y_{ij}] = \mu_i, \quad Var[\epsilon_{ij}] = \sigma^2, \text{ e } Cov[y_{ij}, y_{hk}] = 0, \quad \forall i \neq h, \forall j \neq k, \quad Y_{ij} \sim N(\mu_i, \sigma^2) \quad (7)$$

Para estimarmos  $\mu_i$  do modelo (5), geralmente utilizamos o método de mínimos quadrados, que consiste em minimizar

$$S = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \quad (8)$$

com respeito aos  $\mu_i$ s, derivamos  $S$  em relação a  $\mu_i$  e igualamos a 0. Assim,

$$\begin{aligned} \frac{\partial}{\partial \mu_i} S &= \frac{\partial}{\partial \mu_i} \left[ \sum_{j=1}^{n_i} (y_{1j} - \mu_1)^2 + \sum_{j=1}^{n_i} (y_{2j} - \mu_2)^2 + \dots + \sum_{j=1}^{n_i} (y_{Ij} - \mu_I)^2 \right] \\ &= \left( -2 \sum_{j=1}^{n_1} (y_{1j} - \mu_1) \right) + \dots + \left( -2 \sum_{j=1}^{n_I} (y_{Ij} - \mu_I) \right). \end{aligned} \quad (9)$$

Igualando a zero os  $i$  termos, temos

$$-2 \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i) = 0 \Rightarrow \hat{\mu}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} = \bar{y}_i. \quad (10)$$

Como a derivada segunda de  $S$  com relação a  $\mu_i$  é positiva, isto é,

$$\frac{\partial^2}{\partial^2 \mu_i} S = 2n_i \mu_i,$$

logo  $\hat{\mu}_i$  é um ponto de mínimo.

Uma outra forma de obtermos (10) sem o uso de cálculos, por meio de (5), segue

$$\begin{aligned} S &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} [y_{ij} - \bar{y}_i - (\mu_i - \bar{y}_i)]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (\mu_i - \bar{y}_i)^2 - 2 \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(\mu_i - \bar{y}_i) \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (\mu_i - \bar{y}_i)^2 - 2 \sum_{i=1}^I (\mu_i - \bar{y}_i) \left[ \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) \right], \end{aligned}$$

uma vez que  $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = \sum_{j=1}^{n_i} y_{ij} - n_i \bar{y}_i = \sum_{j=1}^{n_i} y_{ij} - n_i \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} = 0$ , logo

$$S = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (\mu_i - \bar{y}_i)^2. \quad (11)$$

Intuitivamente, a única forma de minimizar (11) com relação a  $\mu_i$ , é considerar no segundo termo  $\hat{\mu}_i = \bar{y}_i$ , uma vez que os termos são somas de quadrado, que por sua vez são termos não negativos.

Podemos reescrever o modelo (5) considerando  $\mu_i = \mu + \tau_i$ , isto é,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad (12)$$

em que  $\mu$  é uma constante (parâmetro) desconhecida e  $\tau_i$  é o parâmetro que representa o efeito do tratamento  $i$  na unidade experimental. Esse modelo é chamado modelo de efeitos. Observe que em (12) há  $1+I$  parâmetros a serem estimados e somente  $I$  média de caselas para estimá-los, é o que chamamos de modelo superparametrizado. Dessa forma, devemos determinar alguma restrição (condições marginais) para os  $\tau_i$ 's, uma vez que não temos solução para minimizar (8) com relação a  $\mu_i = \mu + \tau_i$ . Dessa forma, não temos valores únicos para  $\mu$  e os  $\tau_i$ 's, e sim, infinitas soluções, dependendo da restrição imposta ao modelo.

**Exemplo 2.** Vejamos um experimento com  $i = 2$  e  $j = 3$ , isto é,

$$y_{11} = \mu + \tau_1 + \epsilon_{11}$$

$$y_{12} = \mu + \tau_1 + \epsilon_{12}$$

$$y_{13} = \mu + \tau_1 + \epsilon_{13}$$

$$y_{21} = \mu + \tau_2 + \epsilon_{21}$$

$$y_{22} = \mu + \tau_2 + \epsilon_{22}$$

$$y_{23} = \mu + \tau_2 + \epsilon_{23}$$

Vamos considerar por exemplo que  $\mu = 15$  possa representar uma constante atribuída a todas as unidades experimentais, e que  $\tau_1 = 1$  e que  $\tau_2 = 3$ . O modelo fica:

$$\begin{aligned} y_{1j} &= 15 + 1 + \epsilon_{1j} = 16 + \epsilon_{1j}, & j = 1, 2, 3, \\ y_{2j} &= 15 + 3 + \epsilon_{2j} = 18 + \epsilon_{2j}, & j = 1, 2, 3. \end{aligned}$$

Observe que simplesmente com o último resultado não temos como determinar que  $\mu = 15$ ,  $\tau_1 = 1$  e  $\tau_2 = 3$ , porque

$$\begin{aligned} y_{1j} &= 10 + 6 + \epsilon_{1j} = 16 + \epsilon_{1j}, & j = 1, 2, 3, \\ y_{2j} &= 10 + 8 + \epsilon_{2j} = 18 + \epsilon_{2j}, & j = 1, 2, 3, \end{aligned}$$

ou

$$\begin{aligned} y_{1j} &= 25 - 9 + \epsilon_{1j} = 16 + \epsilon_{1j}, & j = 1, 2, 3, \\ y_{2j} &= 25 - 7 + \epsilon_{2j} = 18 + \epsilon_{2j}, & j = 1, 2, 3. \end{aligned}$$

Nesse caso, podemos fazer uma restrição, tal que  $\tau_1 + \tau_2 = 0$ , e com isso para os modelos

$$\begin{aligned} y_{1j} &= 16 + \epsilon_{1j}, & j = 1, 2, 3, \\ y_{2j} &= 18 + \epsilon_{2j}, & j = 1, 2, 3, \end{aligned}$$

a única condição para  $\mu$ ,  $\tau_1$  e  $\tau_2$  são:

$$\begin{aligned} y_{1j} &= 16 + \epsilon_{1j} = 17 - 1 + \epsilon_{1j}, & j = 1, 2, 3, \\ y_{2j} &= 18 + \epsilon_{2j} = 17 + 1 + \epsilon_{2j}, & j = 1, 2, 3. \end{aligned}$$

Dessa forma, percebemos que  $\mu_i = \mu + \tau_i$  são únicos e podem ser estimados.

Algumas restrições são bem utilizadas. Uma das restrições mais utilizada para (12) é assumir  $\sum_{i=1}^I \tau_i = 0$ . Essa restrição é bem utilizada pelas facilidades algébricas no sistema de equações normais e a consequência de  $\tilde{\mu}$  representar a média geral do experimento, isto é,  $\tilde{\mu} = y_{..}/N$ . Uma outra restrição utilizada pelo programa SAS é  $\tau_I = 0$ , do qual a constante  $\tilde{\mu}$  será a média para o  $I$ -ésimo nível ao invés da média geral do experimento como na primeira restrição. Por fim, poderíamos apresentar outras restrições, mas apresentaremos esta última utilizada pelo programa R, do qual assume  $\tau_1 = 0$ , em que a constante  $\tilde{\mu}$  será a média para o primeiro nível ao invés da média geral do experimento como na primeira restrição.

A seguir, apresentaremos as soluções de mínimos quadrados para essas três restrições. Inicialmente, devemos minimizar (8) com relação a  $\mu_i = \mu + \tau_i$ , isto é,

$$S = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i)^2. \quad (13)$$

Derivando  $S$  com relação a  $\mu$  e  $\tau_i$ , temos

$$\begin{aligned} \frac{\partial}{\partial \mu} S &= \frac{\partial}{\partial \mu} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i)^2 \\ &= -2 \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i) \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} - N\mu - \sum_{i=1}^I n_i \tau_i, \end{aligned} \quad (14)$$

e

$$\frac{\partial}{\partial \tau_i} S = \frac{\partial}{\partial \tau_i} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i)^2 \quad (15)$$

$$= \frac{\partial}{\partial \tau_i} \left[ \sum_{j=1}^{n_1} (y_{1j} - \mu - \tau_1)^2 + \dots + \sum_{j=1}^{n_I} (y_{Ij} - \mu - \tau_I)^2 \right] \quad (16)$$

$$= -2 \sum_{j=1}^{n_1} (y_{1j} - \mu - \tau_1) - \dots - 2 \sum_{j=1}^{n_I} (y_{Ij} - \mu - \tau_I) \quad (17)$$

$$= \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i). \quad (18)$$

Igualando a 0 as expressões (14) e (15), temos o que chamamos de sistema de equações normais (SEN), isto é,

$$\left. \begin{aligned} N\tilde{\mu} + \sum_{i=1}^I n_i \tilde{\tau}_i &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} & (I) \\ n_i \tilde{\mu} + n_i \tilde{\tau}_i &= \sum_{j=1}^{n_i} y_{ij} & (II) \end{aligned} \right\} SEN \quad (19)$$

Para a restrição  $\tau = \sum_{i=1}^I \tau_i = 0$ . No modelo (12), poderia ter considerado o seguinte modelo sob a restrição  $\tau$ ,

$$Y_{ij} = (\mu + \tau) + (\tau_i - \tau) + \epsilon_{ij} \quad (20)$$

$$= \tilde{\mu} + \tilde{\tau}_i + \epsilon_{ij}, \quad (21)$$

em que  $\tilde{\mu} = \mu + \tau$  e  $\tilde{\tau}_i = \tau_i - \tau$ , que nada afetaria. Podemos ainda afirmar que  $\sum_{i=1}^I \tilde{\tau}_i = \sum_{i=1}^I (\tau_i - \tau) = \sum_{i=1}^I \tau_i - I\tau = 0$ .

Dessa forma, podemos expressar as soluções do SEN. Usando (I) em (19), temos

$$\begin{aligned} N\tilde{\mu} + \sum_{i=1}^I n_i (\tau_i - \tau) &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}, & (\tilde{\tau}_i = \tau_i - \tau) \\ N\tilde{\mu} + \sum_{i=1}^I n_i \left( \tau_i - \sum_{i=1}^I \tau_i \right) &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}, & \left( \tau = \sum_{i=1}^I \tau_i \right) \\ N\tilde{\mu} + \sum_{i=1}^I n_i \left( \underbrace{\sum_{i=1}^I (\tau_i - \tau_i)}_{=0} \right) &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}, & \left( \tau = \sum_{i=1}^I \tau_i \right) \\ N\tilde{\mu} &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}, \end{aligned}$$

logo,

$$\tilde{\mu} = \frac{\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}}{N} = \bar{y}_{..} \quad (22)$$

Usando (II) em (19), temos

$$\tilde{\tau}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} - \tilde{\mu} = \bar{y}_i. - \bar{y}_{..} \quad (23)$$

Quando o número de repetições  $n_i = n$  é igual para os tratamentos, o SEN em (19) pode ser reescrito como

$$\left. \begin{aligned} N\tilde{\mu} + n \sum_{i=1}^I \tilde{\tau}_i &= \sum_{i=1}^I \sum_{j=1}^n y_{ij} & (I) \\ n\tilde{\mu} + n\tilde{\tau}_i &= \sum_{j=1}^n y_{ij} & (II) \end{aligned} \right\} SEN \quad (24)$$

e usando a restrição  $\sum_{i=1}^I \tau_i = 0$ , as soluções para  $\tilde{\mu}$  e  $\tilde{\tau}$  podem ser facilmente obtidas. Na expressão (I) em (24), dado a restrição, temos

$$\tilde{\mu} = \frac{\sum_{i=1}^I \sum_{j=1}^n y_{ij}}{N} = \bar{y}_{..}$$

Usando (II) em (24), temos

$$\tilde{\tau}_i = \frac{\sum_{j=1}^n y_{ij}}{n} - \tilde{\mu} = \bar{y}_i. - \bar{y}_{..}$$

As restrições utilizadas no SAS e no programa R são respectivamente,  $\tau_I = 0$  e  $\tau_1 = 0$ . Essas restrições são feitas por esses programas devido a facilidade computacional da restrição. Para a primeira restrição,  $\tau_I = 0$ , temos para (II) de (19), temos a seguinte solução

$$\tilde{\tau}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} - \tilde{\mu} = \bar{y}_i. - \tilde{\mu}, \forall i = 1, 2, \dots, I - 1. \quad (25)$$

Para (I) temos a segunda solução, sob  $\tau_I = 0$ ,

$$\begin{aligned} N\tilde{\mu} + (n_1\tilde{\tau}_1 + \dots + n_{I-1}\tilde{\tau}_{I-1} + n_I\tilde{\tau}_I) &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} \\ N\tilde{\mu} + (n_1\tilde{\tau}_1 + \dots + n_{I-1}\tilde{\tau}_{I-1}) &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij} \\ N\tilde{\mu} + \sum_{i=1}^{I-1} n_i\tilde{\tau}_i &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij} \\ N\tilde{\mu} + \sum_{i=1}^{I-1} n_i(\bar{y}_i. - \tilde{\mu}) &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij}, \quad (\tilde{\tau}_i = \bar{y}_i. - \tilde{\mu}) \\ N\tilde{\mu} + \sum_{i=1}^{I-1} n_i\bar{y}_i. - \tilde{\mu} \sum_{i=1}^{I-1} n_i &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij} \\ N\tilde{\mu} + \sum_{i=1}^{I-1} n_i\bar{y}_i. - \tilde{\mu}(N - n_I) &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij}, \quad \left( N = \sum_{i=1}^{I-1} n_i + n_I \right) \\ \sum_{i=1}^{I-1} n_i\bar{y}_i. + n_I\tilde{\mu} &= \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} y_{ij}. \end{aligned} \quad (26)$$



Uma vez que  $\sum_{i=1}^{I-1} n_i \bar{y}_i = \sum_{i=1}^{I-1} n_i \frac{y_i}{n_i} = \sum_{i=1}^{I-1} y_i$  e que  $\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^I y_i$ , então  $\sum_{i=1}^I y_i - \sum_{i=1}^{I-1} y_i = y_I$ . Logo, a solução para (26) é

$$\tilde{\mu} = \frac{y_I}{n_I} = \bar{y}_I. \quad (27)$$

Para a restrição imposta pelo programa R segue no mesmo raciocínio apresentado acima.

Infinitas soluções temos para  $\tilde{\mu}$  e  $\tilde{\tau}_i$  dependendo da restrição imposta. Entretanto, o que é estimável é  $\hat{\mu}_i = \tilde{\mu} + \tilde{\tau}_i = \bar{y}_i$ .

Um outro parâmetro desconhecido a ser estimado é a variância do erro do modelo, isto é  $Var[\epsilon_{ij}] = \sigma^2$ . Podemos definir o erro do modelo por:

$$\epsilon_{ij} = y_{ij} - \mu_i, \quad (28)$$

sendo,

$$\hat{\epsilon}_{ij} = y_{ij} - \hat{\mu}_i, \quad (29)$$

o preditor do erro, também chamado de resíduo.

Podemos obter um estimador para  $\sigma^2$  por meio da soma de quadrados residual (ou do erro) (SQE), dada por

$$\begin{aligned} SQE &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2 \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad (\hat{\mu}_i = \bar{y}_i) \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij}^2 + \bar{y}_i^2 - 2y_{ij}\bar{y}_i) \\ &= \sum_{i=1}^I \left( \sum_{j=1}^{n_i} y_{ij}^2 + \sum_{j=1}^{n_i} \bar{y}_i^2 - 2 \sum_{j=1}^{n_i} y_{ij}\bar{y}_i \right) \\ &= \sum_{i=1}^I \left( \sum_{j=1}^{n_i} y_{ij}^2 + n_i \bar{y}_i^2 - 2 \sum_{j=1}^{n_i} y_{ij}\bar{y}_i \right) \\ &= \sum_{i=1}^I \left( \sum_{j=1}^{n_i} y_{ij}^2 + n_i \bar{y}_i^2 - 2n_i \bar{y}_i^2 \right), \quad \left( 2 \sum_{j=1}^{n_i} y_{ij}\bar{y}_i = 2n_i \bar{y}_i^2 \right) \\ &= \sum_{i=1}^I \left( \sum_{j=1}^{n_i} y_{ij}^2 - n_i \bar{y}_i^2 \right) \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^I n_i \bar{y}_i^2 \end{aligned} \quad (30)$$

Agora, observe que podemos expressar  $\bar{y}_i$  da seguinte forma:

$$\begin{aligned}
 \bar{y}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \\
 &= \frac{1}{n_i} \sum_{j=1}^{n_i} (\mu_i + \epsilon_{ij}), \quad (\text{Sem perda de generalidade}) \\
 &= \mu_i + \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij} \\
 &= \mu_i + \bar{\epsilon}_i,
 \end{aligned}$$

sendo  $\bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij}$ . Logo,

$$y_{ij} - \bar{y}_i = \epsilon_{ij} - \bar{\epsilon}_i. \quad (31)$$

Usando (31) em (32), temos

$$SQE = \sum_{i=1}^I \sum_{j=1}^{n_i} \epsilon_{ij}^2 - \sum_{i=1}^I n_i \bar{\epsilon}_i^2. \quad (32)$$

A esperança de de (32) pode ser expressa por:

$$\begin{aligned}
 E[SQE] &= \sum_{i=1}^I \sum_{j=1}^{n_i} E[\epsilon_{ij}^2] - \sum_{i=1}^I n_i E[\bar{\epsilon}_i^2] \\
 &= \sum_{i=1}^I \sum_{j=1}^{n_i} \sigma^2 - \sum_{i=1}^I n_i \frac{\sigma^2}{n_i} \\
 &= N\sigma^2 - I\sigma^2 \\
 &= (N - I)\sigma^2,
 \end{aligned}$$

mostrando ser um estimador viesado de  $\sigma^2$ .