5. ANÁLISE DA VARIÂNCIA

5.1 INTRODUÇÃO

Consideremos primeiro um exemplo simples. Suponhamos que pretendemos comparar os valores médios de duas variáveis aleatórias com distribuição gaussiana com valores médios μ_1 e μ_2 e variâncias iguais $\sigma_1^2 = \sigma_2^2 = \sigma^2$ com base em duas amostras $\left(X_{11},\ldots,X_{1n_1}\right)$ e $\left(X_{21},\ldots,X_{2n_2}\right)$ independentes de dimensões n_1 e n_2 respectivamente ($n_1 + n_2 = n$). Isto é, queremos testar $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ ou equivalentemente

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 0 \tag{1}$$

A estatística $\overline{X}_1 - \overline{X}_2$ é o estimador de $\mu_1 - \mu_2$ e tem distribuição

$\overline{X}_1 - \overline{X}_2 \cap Gau\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$ (note-se que as variâncias são iguais). Assim sendo, rejeitaremos H_0 quando $\left|\overline{X}_1 - \overline{X}_2\right| > k$, onde k é tal que para um nível de significância α

$$P\left(\left|\overline{X}_1 - \overline{X}_2\right| > k \middle| H_0 \; verdadeira\right) = \alpha$$

Sob a validade da hipótese nula

$$\overline{X}_1 - \overline{X}_2 \cap Gau\left(0, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \tag{2}$$

ou ainda

$$\frac{\overline{X}_1 - \overline{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \cap Gau(0,1) \tag{3}$$

Usualmente a variância σ^2 é desconhecida, logo é necessário estimá-la a partir das duas amostras observadas. O estimador é então

$$S_p^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \tag{4}$$

onde $S_i^2 \; i = 1,2$ é a variância empírica da iésima amostra, a S_p^2 chama-se "pooled variance". A distribuição deste estimador é

$$\frac{(n_1+n_2-2)S_p^2}{\sigma^2} = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \cap \chi^2_{n_1+n_2-2} \tag{5}$$

dado que esta v.a. é a soma de dois qui-quadrados independentes com $(n_1 - 1)$ e $(n_2 - 1)$ graus de liberdade respectivamente.

Finalmente, sob a validade da hipóse nula, a variável aleatória

$$T = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \cap t_{n_1+n_2-2} \tag{6}$$

e agora podemos obter a região de rejeição do teste de nível α

$$\text{Rejeitar } H_0 \text{ se } |T| > t_{1-\frac{\alpha}{2};n_1+n_2-2} \tag{7}$$

Exercício: Para obter mais eficiência numa linha de montagem de determinada fábrica os novos empregados necessitam de um mês de treino. Foi sugerido um novo método de estágio e foi efectuado um teste para comparar o método novo com o método tradicional. Dois grupos de nove empregados cada foram treinados durante 3 semanas, um grupo usando o novo método e o outro o método tradicional. No fim do período de treino de 3 semanas foi registado o tempo (em minutos) que cada operário levava a montar um determinado equipamento. Os resultados obtidos encontram-se na tabela seguinte.

Método	**Observações**								
Standard	32	37	35	28	41	44	35	31	34
Novo	35	31	29	25	34	40	27	32	31

Supondo que os tempos de montagem X_i , (i=1, 2) seguem aproximadamente uma distribuição gaussiana e têm variâncias aproximadamente iguais, teste ao nível de significância de 5% a hipótese $H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 0$.

5.2 ANÁLISE DE VARIÂNCIA SIMPLES (A UM FACTOR)

Suponhamos agora que temos k populações e que queremos comparar os seus valores médios. No caso do exemplo, poderíamos querer comparar o efeito médio de vários métodos de treino. Várias experiências envolvem estudos sobre o efeito de um ou mais *factores* na variável resposta (observações). Estes factores (variáveis controladas) podem ter vários *níveis* cada. Nesta secção vamos dedicar-nos ao estudo da análise de variância a um factor (one-way). Como exemplo, suponhamos que queríamos comparar colheitas em campos que recebem diferentes fertilizantes, aqui temos um factor (fertilizante) com vários níveis (variedade do fertilizante utilizado), ou que queríamos comparar o efeito de diferentes doses (níveis) de um antibiótico (factor) na eficácia da cura de uma determinada patologia. Neste caso não tem sentido comparar os valores médios dois a dois, mas sim compará-los globalmente. Fisher (1925) desenvolveu uma metodologia que permite realizar estas comparações simultaneamente.

Consideremos agora k amostras *independentes*

$$\begin{array}{l} \left(X_{11},\ldots X_{1n_1}\right) \\ \left(X_{21},\ldots X_{2n_2}\right) \\ . \\ . \\ . \\ \left(X_{k1},\ldots X_{1n_k}\right) \end{array}$$

extraídas de k populações Gaussianas $\left(\mu_i,\sigma\right) i = 1,\ldots,k$, onde cada amostra é um vector com n_i observações do efeito do iésimo nível do factor que estamos a estudar. Admitamos a hipótese dos diferentes níveis do factor em estudo terem o mesmo efeito médio, isto é,

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu \quad versus \quad H_1 : \exists i, j : \mu_i \neq \mu_j \tag{1}$$

No caso de a hipótese nula ser verdadeira todas as nossas observações podiam ser consideradas uma amostra de dimensão $n = \sum_{i=1}^{k} n_i$ de uma população gaussiana com o mesmo valor médio μ e desvio-padrão σ. E então o efeito médio global do factor em estudo pode ser avaliado por

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} \tag{2}$$

Por outro lado, o efeito médio μ_i de cada nível (tratamento) do factor **A** em estudo pode ser avaliado pela média de cada subamostra, i.e.,

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \tag{3}$$

A soma dos quadrados que nos permite estimar a variância da amostra pode ser particionada numa soma de duas parcelas do seguinte modo:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i.}+\overline{X}_{i.}-\overline{X}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i.}\right)^2 + \sum_{i=1}^{k} n_i\left(\overline{X}_{i.}-\overline{X}\right)^2 \tag{4}$$

Dado que o termo relativo ao produto cruzado se anula. Esta partição da soma de quadrados pode escrever-se

$$SST = SSE + SSA \tag{5}$$

e chama-se *Partição da Soma dos Quadrados*, sendo o primeiro membro a *Soma total de quadrados*, a primeira parcela do 2º membro a *Soma dos Quadrados Residual* (mede a variabilidade *dentro de cada amostra*) e a segunda parcela a *Soma dos quadrados Entre Amostras* (ou devida ao factor A).

Note-se que $\overline{X}_{i.} - \overline{X} = \hat{\mu}_i - \hat{\mu}$ é estimador da diferença $\mu_i - \mu$ e avalia o efeito do iésimo nível do factor em estudo no valor médio comum μ (caso H_0 verdadeira), isto é, podemos escrever $\mu_i = \mu + (\mu_i - \mu) = \mu + \alpha_i$ e $\hat{\alpha}_i = \overline{X}_{i.} - \overline{X}$. Então SSA pode escrever-se $SSA = \sum_{i=1}^{k} n_i \hat{\alpha}^2$ e é de esperar que esta variável assuma grandes valores se a hipótese nula não for verdadeira. Assim, devemos rejeitar H_0 quando *SSA*> C sendo este ponto crítico calculado a partir do nível de significância do teste, mas para isso precisamos de saber a distribuição da estatítica de teste sob a validade da hipótese nula.

Quando H_0 verdadeira

$$\frac{SSA}{\sigma^2} \cap \chi^2_{k-1} \tag{6}$$

Mas usualmente σ^2 é desconhecido e é preciso obter um seu estimador.

A variável

$$\frac{SSE}{\sigma^2}=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i.}\right)^2}{\sigma^2}=\frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{\sigma^2}\cap\chi^2_{\sum_{i=1}^{k}(n_i-1)=n-k} \tag{7}$$

(s_i^2 é a variância empírica da iésima amostra (observações do iésimo nível do factor A e portanto, $\frac{(n_i-1)s_i^2}{\sigma^2}\cap\chi^2_{n_i-1}$, i =1,…, k independentes) e $E\left(\frac{SSE}{n-k}\right)=E(MSE)=\sigma^2$,

isto é, *MSE* é um estimador centrado da variância. σ^2. As parcelas do 2º membro da partição (5) são v.a.s independentes, logo a variável

$$F=\frac{SSA/\sigma^2(k-1)}{SSE/\sigma^2(n-k)}=\frac{MSA}{MSE}\cap F_{k-1;n-k} \tag{8}$$

que já não depende de σ^2 e será agora a estatística de teste. Tenderá a assumir valores “grandes”quando a hipótese nula for falsa (uma vez que o numerador aumenta se H_0 falsa).

Finalmente rejeitamos H_0 quando

$$F>F_{1-\alpha;k-1,n-k} \tag{9}$$

Para um problema específico é costume dispor os resultados numa tabela de Análise de Variância ou ANOVA (abreviatura de Analisys of Variance).

Tabela de **ANOVA** -1

Origem da Variação	**Soma de Quadrados**	**Graus de liberdade**	**Média Soma dos Quadrados**	**Razão F**
Entre Amostras	$SSA=\sum_{i=1}^{k}n_i\left(\overline{X}_{i.}-\overline{X}\right)^2$	*k-1*	$MSA=SSA/(k-1)$	$F=\frac{MSA}{MSE}$
Residual	$SS_e=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i.}\right)^2$	*n-k*	*MSE = SSE/(n-k)*	
Total	$SS_T=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}\right)^2$	*n-1*		

Exemplo 1: Suponhamos que pretendemos comparar três marcas distintas de baterias

Marcas		
A	**B**	**C**
40	60	60
30	40	50
50	55	70
50	65	65
30		75
		40

Suponhamos que as três amostras provêem de populações Gaussianas com valores médios μ_i, i=1,2,3 e variância σ^2 desconhecida. Vamos testar a hipótese de que as 3 marcas não diferem no tempo médio de vida X, isto é, $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$ *versus* $H_1 : \exists i, j : \mu_i \neq \mu_j \quad i, j = 1,2,3$, ao nível de significância de 5%.

Quando temos de efectuar os cálculos manualmente é preferível utilizar as expressões simplificadas das somas de quadrados.

$$SSA = \sum_{i=1}^{k} n_i \left(\overline{X}_{i.} - \overline{X}\right)^2 = \sum_{i=1}^{k} n_i \overline{X}_{i.}^2 - n\overline{X}^2 \;:$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}\right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}^2 - n\overline{X}^2$$

E obter por diferença

$SSE = SST\text{-}SSA$

Para os dados da tabela anterior:

SST = 43300-15x 52^2 = 2740 $\qquad \overline{x}_{1.} = 40, \overline{x}_{2.} = 55, \overline{x}_{3.} = 60$

SSA = 41700-15x 52^2=1140 e SSE = 2740-1140=1600

Podemos apresentar os resultados na tabela seguinte

ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	1140	2	570	4,275	***0,04***	**3,89**
Within Groups	1600	12	133,33			
Total	2740	14				

A penúltima coluna da tabela dá-nos o valor do p-value < 0,05 e a última coluna o ponto crítico $F_{0.95;2,12} = 3{,}89 < F_{observ} = 4{,}275$, o que nos leva a decidir pela rejeição da hipótese nula ao nível de significância de 5%. Isto é, as marcas das baterias diferem no tempo médio de vida.

5.3 ANÁLISE DE VARIÂNCIA DUPLA (A DOIS FACTORES)

Em muitas situações estamos interessados em investigar os possíveis efeitos de dois factores no resultado de uma experiência. Por exemplo, a qualidade do grão e o tipo de fertilizante utilizado influenciam ambos a produção agrícola, a classificação obtida por alunos de uma turma pode ser influenciada pelo tamanho da classe e pelo professor que lecciona a disciplina, etc.

Consideremos então dois factores A e B a influenciar o resultado de uma experiência e suponhamos que dispomos de uma observação para cada combinação do iésimo nível do factor A e do jésimo nível do factor B representada por X_{ij}, i=1,…,I e j=1,…,J . Esquematicamente,

	1	2	…	J	Média Linha
1	X_{11}	X_{12}	…	X_{1J}	$\overline{X}_{1.}$
2	X_{21}	X_{22}	…	X_{2J}	$\overline{X}_{2.}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
I	X_{I1}	X_{I2}	…	X_{IJ}	$\overline{X}_{I.}$
Média Coluna	$\overline{X}_{.1}$	$\overline{X}_{.2}$		$\overline{X}_{.J}$	$\overline{X}$

Onde as médias de linha, coluna e global são iguais a

$$\overline{X}_{.j} = \frac{1}{I}\sum_{i=1}^{I} X_{ij}\ ,\ j = 1,\ldots J\ ;\ \overline{X}_{i.} = \frac{1}{J}\sum_{j=1}^{J} X_{ij}\ \ ,i = 1,\ldots,I\ ;\ \overline{X} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij} \qquad n = IJ \quad (1)$$

O modelo probabilístico subjacente a esta experiência é:

$$X_{ij} \cap Gau(\mu_{ij}, \sigma) \quad independentes$$

ou ainda (2)

$$X_{ij} = \mu_{ij} + \varepsilon_{ij}\ , i = 1,...,I;\ j = 1,...,J \quad com \quad \varepsilon_{ij} \cap Gau(0,\sigma) \quad independentes$$

onde

$$\mu_{ij} = \mu + \alpha_i + \beta_j \tag{3}$$

Sendo α_i o *efeito do iésimo nível* do factor A e β_j *o efeito do jésimo nível* do factor B.

Assume-se ainda que

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0 \tag{4}$$

Desta condição resulta que:

$$\begin{aligned} \mu_{i.} &= \frac{1}{J} \sum_{j=1}^{J} \mu_{ij} = \mu + \alpha_i \quad i.e. \quad \alpha_i = \mu - \mu_{i.}\ , i = 1,...,n \\ \mu_{.j} &= \frac{1}{I} \sum_{i=1}^{I} \mu_{ij} = \mu + \beta_j \quad i.e. \quad \beta_j = \mu - \mu_{.j}\ , j = 1,...,J \end{aligned} \tag{5}$$

Atendendo a (2) e (3) podemos ainda escrever o modelo na forma

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}\ , X_{ij} \cap Gau(\mu + \alpha_i + \beta_j, \sigma)\, independentes, i = 1,...,I;\ j = 1,...,J$$

O nosso objectivo é averiguar a possível influência dos factores A e B no resultado da experiência, o que em termos das observações se reflectirá nos seus valores médios. As hipóteses de interesse são pois

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_I = 0 \quad versus \quad H_1 : \exists_{i:}\ \alpha_i \neq 0 \tag{6}$$

e

$$H_0' : \beta_1 = \beta_2 = ... = \beta_J = 0 \quad versus \quad H_1' : \exists_{j:}\ \beta_j \neq 0 \tag{7}$$

A primeira hipótese nula diz que o factor A não produz efeito no resultado da experiência, e a segunda hipótese nula analogamente para o factor B.

Tal como na análise de variância simples podemos escrever uma partição da soma de quadrados

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(X_{ij}-\overline{X}\right)^2=\sum_{i=1}^{I}\sum_{j=1}^{J}\left(X_{ij}-\overline{X}_{i.}+\overline{X}_{i.}+\overline{X}_{.j}-\overline{X}_{.j}-\overline{X}+\overline{X}-\overline{X}\right)^2$$

ou seja (8)

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(X_{ij}-\overline{X}\right)^2=J\sum_{i=1}^{I}\left(\overline{X}_{i.}-\overline{X}\right)^2+I\sum_{j=1}^{J}\left(\overline{X}_{.j}-\overline{X}\right)^2+\sum_{i=1}^{I}\sum_{j=1}^{J}\left(X_{ij}-\overline{X}_{i.}-\overline{X}_{.j}+\overline{X}\right)^2$$

Ou ainda $SST = SSA + SSB + SSE$ (9)

A *Soma total de quadrados* foi subdividida na *soma dos quadrados devida ao factor A*, *soma dos quadrados devida ao factor B e soma dos quadrados residual.*

Vejamos como estimar a partir das observações o efeito dos níveis de cada factor. Atendendo a (5) tem-se

$$\hat{\alpha}_i = \hat{\mu}_{i.} - \hat{\mu} = \overline{X}_{i.} - \overline{X} \;, i = 1,\dots,I$$
$$\hat{\beta}_j = \hat{\mu}_{.j} - \hat{\mu} = \overline{X}_{.j} - \overline{X} \;, j = 1,\dots,J \qquad (10)$$

e as somas de quadradros devidas aos factores A e B podem escrever-se

$$SSA = J\sum_{i=1}^{I}\hat{\alpha}_i^2$$
$$SSB = I\sum_{j=1}^{J}\hat{\beta}_j^2 \qquad (11)$$

o que nos permite concluir que devemos rejeitar a hipótese H_0 para “grandes” valores de *SSA*, e rejeitar H_0' para “grandes” valores de *SSB*.

Isto é,

Rejeitar H_0 quando $SSA > k_1$

Rejeitar H_0' quando $SSB > k_2$

Como calcular as constantes k_1 e k_2? Primeiro temos de saber qual a distribuição das estatísticas de teste. Pode provar-se que sob a validade da hipótese H_0 a v.a.

$$\frac{SSA}{\sigma^2} \cap \chi^2_{I-1} \qquad (12)$$

e que sob a validade da hipótese H_0' a v.a.

$$\frac{SSB}{\sigma^2} \cap \chi^2_{J-1} \qquad (13)$$

Por outro lado o estimador da variância $\hat{\sigma}^2$ *é* MSE e em qualquer circunstância tem-se:

$$\frac{SSE}{\sigma^2} \cap \chi^2_{(I-1)(J-1)} \tag{14}$$

Prova-se também que as v.a.'s *SSA, SSB* e *SSE* são independentes e então

$$\frac{SSA/\sigma^2(I-1)}{SSE/\sigma^2(J-1)} = \frac{MSA}{MSE} \cap F_{I-1;(I-1)(J-1)} \tag{15}$$

$$\frac{SSB/\sigma^2(J-1)}{SSE/\sigma^2(J-1)} = \frac{MSB}{MSE} \cap F_{J-1;(I-1)(J-1)} \tag{16}$$

Se fixarmos o nível de significância α podemos finalmente escrever as regiões de rejeição das hipóteses H_0 e H_0'

$$\begin{array}{l} Re\ jeitar\ H_0\ se\ \dfrac{MSA}{MSE} > F_{1-\alpha;I-1,(I-1)(J-1)} \\ Re\ jeitar\ H_0'\ se\ \dfrac{MSB}{MSE} > F_{1-\alpha;J-1,(I-1)(J-1)} \end{array} \tag{17}$$

A tabela de ANOVA para este problema é

Tabela de **ANOVA** -2

Origem da Variação	**Soma de Quadrados**	**Graus de liberdade**	**Média Soma dos Quadrados**	**Razão F**
Linhas (A)	$SSA = J\sum_{i=1}^{I}\left(\overline{X}_{i.} - \overline{X}\right)^2$	*I-1*	$MSA = SSA/(I-1)$	$F = \dfrac{MSA}{MSE}$
Colunas (B)	$SSB = I\sum_{j=1}^{J}\left(\overline{X}_{.j} - \overline{X}\right)^2$	*J-1*	$MSB = \dfrac{SSB}{J-1}$	$F = \dfrac{MSB}{MSE}$
Residual	$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}\right)^2$	*(I-1)(J-1)*	$MSE = \dfrac{SSE}{(I-1)(J-1)}$	
Total	$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2$	*IJ-1*		

Exemplo 2: Na tabela seguinte estão registadas as produções (kg/lote) de 3 variedades de trigo, obtidas com 4 fertilizantes distintos.

	Variedade	*de*	*trigo*
Fertilizante	**A**	**B**	**C**
α	8	3	7
β	10	4	8
γ	6	5	6
δ	8	4	7

Serão as três variedades de trigo idênticas quanto à produção média? E os fertilizantes serão igualmente eficazes?

Assumido o modelo $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}\ , i = 1,2,3,4;\ j = 1,2,3$, onde α_i é o efeito do iésimo fertilizante e β_j o efeito da jésima variedade de trigo. As hipóteses a testar são

$$H_0 : \alpha_1 = ... = \alpha_4 = 0 \quad vs \quad H_1 : \exists i : \alpha_i \neq 0$$

e

$$H_0' : \beta_1 = \beta_2 = \beta_3 = 0 \quad vs \quad H_1' : \exists j : \beta_j \neq 0$$

Usando expressões simplificadas para as somas de quadrados como na secção anterior obtem-se:

$$SSA = 3\sum_{i=1}^{4} \bar{x}_{i.}^2 - 12\bar{x}^2 = 4.67 \quad SSB = 4\sum_{j=1}^{3} \bar{x}_{.j}^2 - 12\bar{x}^2 = 34.67$$

$$SST = \sum_{i=1}^{4}\sum_{j=1}^{3} x_{ij}^2 - 12\bar{x}^2 = 46.67 \quad SSE = SST - (SSA + SSB) = 7.33$$

ANOVA

Origem da Variação	**Soma de Quadrados**	**Graus de liberdade**	**Média Soma dos Quadrados**	**Razão F**
Fertilizante (A)	*4.67*	*3*	*1.56*	***1.28***
Variedade de trigo (B)	*34.67*	*2*	*17.33*	***14.2***
Residual	*7.33*	*6*	*1.22*	
Total	*46.67*	*11*		

Por outro lado, $F_{0.95;3,6} = 4.76$ *e* $F_{0.95;2,6} = 5.14$. Uma vez que 1.28 < 4.76, não se rejeita a hipótese H_0 ao nível e significância 5% e face à amostra observada, i.e., os diferentes fertilizantes parecem ser igualmente eficazes. Pelo contrário, 14.2>5.14 e rejeita-se a hipótese H_0' ao nível e significância 5%, logo as três variedades de trigo não são idênticas quanto à produção média.