

**PARTIÇÃO DA VARIÂNCIA DA INTERAÇÃO
GENÓTIPOS X AMBIENTES EM EFEITOS
EXPLICADOS PELOS MARCADORES E DESVIOS**

Heyder Diniz SILVA¹
Roland VENCOVSKY²
Glauce Cristina Ricardo RUMIN²

- **RESUMO:** Um dos problemas encontrado na análise de QTL's se refere à abordagem da interação QTL's x ambientes. Neste contexto, apresentou-se uma partição da variância da interação genótipos x ambientes em efeitos explicados pelos marcadores e desvios, a partir da qual obtiveram-se os estimadores da proporção da variância genética (p_m), e da variância da interação genótipos x ambientes (p_{ms}), explicados pelos marcadores moleculares. Estes estimadores independem de desvios das frequências alélicas dos marcadores em relação às esperadas (1:2:1 em uma geração F_2 , 1:1 em um retrocruzamento, etc.), porém apresentam uma alta probabilidade de obtenção de estimativas fora do intervalo paramétrico, principalmente para valores elevados destas proporções. Contudo, estas probabilidades podem ser reduzidas com o aumento do número de repetições e/ou ambientes nos quais os genótipos são avaliados.
- **PALAVRAS-CHAVE:** QTL, Interação QTL's x Ambientes, Seleção assistida, componentes de variância.

¹Faculdade de Matemática, Universidade Federal de Uberlândia, CEP 38408-902, Uberlândia - MG, Brasil

²Departamento de Genética, Escola Superior de Agricultura Luiz de Queiróz, Universidade de São Paulo, CEP 13418-900, Piracicaba, SP, Brasil

1 Introdução

As modernas técnicas de obtenção de marcadores moleculares permitem aos geneticistas construir mapas genéticos de alta densidade para qualquer espécie vegetal de interesse. Associados a técnicas estatísticas, esses mapas podem ser utilizados para detecção, localização e estimação dos efeitos de locos controladores de caracteres quantitativos (QTLs). Desse modo, os marcadores moleculares ligados a QTLs podem ser utilizados para aumentar a eficiência dos programas de seleção, por meio da técnica denominada seleção assistida por marcadores moleculares (SAM). Lande & Thompson (1990) propuseram a realização da SAM, utilizando um índice que combina as informações fenotípicas e moleculares, quantificadas por meio de um escore molecular, em proporções apropriadas. Neste procedimento, o escore molecular pode ser obtido mediante valores preditos por um modelo de regressão múltipla tendo como variáveis preditoras o número de “alelos” para cada loco marcador, sendo o peso atribuído a esse escore, no índice de seleção, uma função da proporção da variância genética explicada pelos marcadores moleculares.

O coeficiente de determinação (r^2) do modelo de regressão múltipla, utilizado na descrição genotípica do caráter, pode ser tomado como estimador da proporção da variância genética explicada pelos marcadores moleculares. Porém Bearzoti & Vencovsky (1998) argumentam que, devido ao fato de a esperança matemática da soma de quadrados de regressão envolver componentes de variâncias genéticas e ambientais, este estimador pode não ser muito adequado e apresentam um outro estimador, baseado na partição da variância genética em efeitos explicados pelos marcadores e desvios.

De acordo com Wang et al. (1999), a interação genótipos x ambientes tem representado um importante papel para os melhoristas e geneticistas quantitativos, e que, com a utilização de marcadores moleculares e delineamentos experimentais adequados, essa interação pode ser dissecada em seus componentes devido à interação QTLs x ambientes, que são de grande importância para os procedimentos de SAM. Desse modo, propõe-se, no presente trabalho, uma extensão do estimador da proporção da variância genética explicada pelos marcadores moleculares apresentado por Bearzoti & Vencovsky (1998), aplicável a experimentos conduzidos em um único ambiente, para o caso de experimentos conduzidos

em mais de um ambiente, bem como uma partição da variância da interação genótipos x ambientes em efeitos explicados pelos marcadores moleculares e os seus desvios.

2 Modelo matemático e partição das somas de quadrados

Considere-se um experimento no qual são avaliadas g progênie em r repetições em l ambientes. Então, um modelo matemático para a análise desses dados é:

$$y_{ij\ell} = \mu + s_{\ell} + t_i + (ts)_{i\ell} + e_{ij\ell} \quad (1)$$

em que

$y_{ij\ell}$ é o valor observado da progênie i , $i = 1, 2, \dots, g$, na repetição j , $j = 1, 2, \dots, r$ do ambiente ℓ , $\ell = 1, 2, \dots, l$;

μ é uma constante inerente a todas as observações;

s_{ℓ} é o efeito do ambiente ℓ ;

t_i é o efeito da progênie i ;

$(ts)_{i\ell}$ é o efeito da interação entre a progênie i e o ambiente ℓ ;

$e_{ij\ell}$ é o erro aleatório associado à observação $y_{ij\ell}$.

Na forma matricial, o modelo 1 é

$$\mathbf{y} = X\theta + \varepsilon \quad (2)$$

sendo:

\mathbf{y} o vetor de valores fenotípicos observados;

X a matriz do delineamento, e

ε o vetor de erros aleatórios.

- Somas de quadrados total, de parâmetros e resíduo

Sabe-se que:

$$SQ_{Total} = SQ_{Parâmetros} + SQ_{Resíduo};$$

ou seja:

$$\mathbf{y}'I\mathbf{y} = \mathbf{y}'P\mathbf{y} + \mathbf{y}'(I - P)\mathbf{y} \quad (3)$$

com $P = XX^+ = X(X'X)^G X'$, sendo X^+ a inversa de Moore-Penrose de X , $(X'X)^G$ qualquer inversa generalizada de $X'X$, e I uma matriz identidade de dimensões $n \times n$, ($n = grl$).

As somas de quadrados Total, de Parâmetros e Resíduo, estão associadas, respectivamente, a $r[I] = n$, $r[P] = l + g + lg + 1$ e $r[I - P] = lg(r - 1)$ graus de liberdade.

- Soma de quadrados de ambientes

A porção da SQParâmetros devida ao efeito de ambientes, denominada Soma de Quadrados de Ambientes, SQL, é dada por:

$$SQL = \mathbf{y}'(P_L - P_1)\mathbf{y}, \quad (4)$$

em que:

$P_L = X_L X_L^+$ e $P_1 = J J^+$, com X_L sendo uma submatriz de X , com dimensões $n \times (l + 1)$ referente aos efeitos de ambientes, e J é uma matriz de dimensões $n \times l$, de uns.

Esta soma de quadrados está associada a $r[P_L - P_1] = l - 1$ graus de liberdade.

- Soma de quadrados de progênes

A porção da SQParâmetros devida ao efeito de progênes, denominada Soma de Quadrados de Progênes, SQP, é dada por:

$$SQP = \mathbf{y}'(P_P - P_1)\mathbf{y}, \quad (5)$$

em que:

$P_P = X_P X_P^+$ com X_P sendo uma submatriz de X , com dimensões $n \times (g + 1)$ referente aos efeitos de progênes.

Esta associa-se a $r[P_P - P_1] = g - 1$ graus de liberdade.

- Soma de quadrados da interação

A porção da SQParâmetros devida aos efeitos de interação progênes x ambientes, denominada Soma de Quadrados da Interação, SQPxL, é dada por:

$$SQP \times L = \mathbf{y}'(P - P_L - P_P + P_1)\mathbf{y}, \quad (6)$$

estando associada a $r[P - P_L - P_P + P_1] = (g - 1)(l - 1)$ graus de liberdade.

• Partição das somas de quadrados de progênes e interação progênes x ambientes em efeitos explicados pelos marcadores e desvios.

Quando se possuem informações de marcadores moleculares para as progênes, o efeito de progênes (t_i) pode ser desdobrado,

por meio de regressão linear, em efeitos explicados pelos marcadores e desvios, o mesmo ocorrendo para a interação, do seguinte modo:

Adotando-se o modelo 1, e sendo t_i modelado por:

$$t_i = \sum_{m=1}^k \beta_m Z_{im} + h_i; \quad (7)$$

em que:

β_m é o coeficiente de regressão da marca m ;
 Z_{im} é uma variável dummy codificadora do genótipo da i -ésima progênie, quanto à marca t ;
 h_i é a porção de g_i não explicada pelas marcas (desvio).

Desse modo, a soma de quadrados de progênies pode ser desdobrada em soma de quadrados de regressão e desvios, sendo:

$$SQReg. = \mathbf{y}'(P_R - P_1)\mathbf{y}, \quad (8)$$

com $P_R = ZZ^+$, sendo Z uma matriz de dimensões $n \times k + 1$.

$$SQDesv. = \mathbf{y}'(P_P - P_R)\mathbf{y}. \quad (9)$$

Estas têm, respectivamente, $r[P_R - P_1] = k$ e $r[P_P - P_R] = g - k - 1$ graus de liberdade.

A soma de quadrados da interação progênies x ambientes também pode ser particionada em soma de quadrados de interação regressão x ambientes e soma de quadrados de interação desvios x ambientes, por meio das seguintes relações:

Sabe-se que:

$$SQP \times L = \sum_{\ell} SQP_{\ell} - SQP, \quad (10)$$

em que:

SQP_{ℓ} é a soma de quadrados de progênies no ambiente ℓ ;
 SQP é a soma de quadrados de progênies na análise conjunta, considerando os l ambientes.

Aplicando-se a relação 10 às somas de quadrados de regressão e desvios, tem-se:

$$SQReg. \times Amb. = \sum_{\ell} SQReg.\ell - SQReg.; \quad (11)$$

$$SQDesv. \times Amb. = \sum_{\ell} SQDesv.\ell - SQDesv. \quad (12)$$

3 Esperanças matemáticas das somas de quadrados e dos quadrados médios.

Assumindo-se que:

$$s_{\ell} \sim N(0, \sigma_s^2);$$

$$t_i \sim N(0, \sigma_g^2);$$

$$(ts)_{i\ell} \sim N(0, \sigma_{gs}^2);$$

$$e_{ij\ell} \sim N(0, \sigma^2);$$

s_{ℓ} , t_i , $(ts)_{i\ell}$ e $e_{ij\ell}$ são independentes;

$cov(t_i, t_{i'}) = 0$; o mesmo sendo válido para s e ts ;

tem-se que as esperanças matemáticas dos quadrados médios de locais, progênes, interação progênes por locais e resíduo são (Searle *et al.*, 1992; Vencovsky & Barriga, 1992; Cruz & Regazzi, 1994):

$$E[QML] = \sigma^2 + r\sigma_{gs}^2 + rg\sigma_t^2; \quad (13)$$

$$E[QMP] = \sigma^2 + r\sigma_{gs}^2 + rl\sigma_g^2; \quad (14)$$

$$E[QMP \times L] = \sigma^2 + r\sigma_{gs}^2; \quad (15)$$

$$E[QMResíduo] = \sigma^2. \quad (16)$$

Para obtenção das esperanças matemáticas dos quadrados médios de regressão, desvios, interação regressão x ambientes e desvios x ambientes assume-se ainda que:

$$\sigma_g^2 = \sigma_m^2 + \sigma_h^2;$$

$$\sigma_{gs}^2 = \sigma_{ms}^2 + \sigma_{hs}^2$$

$$Cov(y_{ij\ell}, y_{ij'\ell} / m_i) = \sigma^2 + \sigma_h^2 + \sigma_{hs}^2$$

$$Cov(y_{ij\ell}, y_{ij'\ell} / m_i) = \sigma_h^2 + \sigma_{hs}^2$$

$$Cov(y_{ij\ell}, y_{ij'\ell'} / m_i) = \sigma_h^2$$

Deste modo, a matriz de variâncias e covariâncias, dados os genótipos das progênies, quanto aos marcadores moleculares é:

$$V^* = \begin{bmatrix} \sigma_h^2 + \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 & \dots & 0 & 0 & 0 & 0 \\ \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & \sigma_b^2 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & \dots & \sigma_h^2 & \sigma_h^2 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \sigma_b^2 & \sigma_b^2 \\ 0 & 0 & \dots & 0 & 0 & \sigma_h^2 & \sigma_h^2 \\ \sigma_b^2 & \sigma_b^2 & \dots & 0 & 0 & 0 & 0 \\ \sigma_h^2 & \sigma_h^2 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 & 0 & 0 \\ 0 & 0 & \dots & \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 \\ 0 & 0 & \dots & 0 & 0 & \sigma_h^2 + \sigma_{hs}^2 & \sigma_h^2 + \sigma_{hs}^2 \end{bmatrix}$$

Sabendo-se que a esperança matemática de uma forma quadrática $\mathbf{y}'A\mathbf{y}$ é dada por:

$$E[\mathbf{y}'A\mathbf{y}] = tr[AV] + E[\mathbf{y}]'AE[\mathbf{y}] \quad \text{Searle (1971),}$$

com V sendo a matriz de variâncias e covariâncias de \mathbf{y} .

As expressões das esperanças matemáticas dos quadrados médios de regressão, desvios, interação regressão x ambientes e interação desvios x ambientes são:

$$\begin{aligned} \bullet E[SQDesv.] &= E[\mathbf{y}'(P_P - P_R)\mathbf{y}] = tr[(P_P - P_R)V^*] + \\ &E[\mathbf{y}]'(P_P - P_R)E[\mathbf{y}] \\ &= tr[(P_P - P_R)V^*] + \beta'Z'(P_P - P_R)Z\beta \\ &= (g - k - 1)(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2) \end{aligned}$$

conseqüentemente,

$$E[QMDesv.] = E\left[\frac{1}{g - k - 1}SQDesv.\right] = \sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2$$

$$\begin{aligned} \bullet E[SQReg.] &= E[\mathbf{y}'(P_R - P_1)\mathbf{y}] = tr[(P_R - P_1)V^*] + \\ &E[\mathbf{y}]'(P_R - P_1)E[\mathbf{y}] \\ &= tr[(P_R - P_1)V^*] + \beta'Z'(P_R - P_1)Z\beta \\ &= k(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2) + \beta'Z'(P_R - P_1)Z\beta \end{aligned}$$

Mas SQP = SQReg. + SQDesv.

$$\mathbf{y}'(P_P - P_1)\mathbf{y} = \mathbf{y}'(P_R - P_1)\mathbf{y} + \mathbf{y}'(P_P - P_R)\mathbf{y}$$

$$\begin{aligned}
E[\mathbf{y}'(P_P - P_1)\mathbf{y}] &= E[\mathbf{y}'(P_R - P_1)\mathbf{y}] + E[\mathbf{y}'(P_P - P_R)\mathbf{y}] \\
(g-1)\sigma^2 + (g-1)(r\sigma_{g_s}^2 + rl\sigma_g^2) &= k(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) + \\
\beta'Z'(P_P - P_1)Z\beta + (g-k-1)(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) & \\
(g-1)(r\sigma_{g_s}^2 + rl\sigma_g^2) &= \beta'Z'(P_P - P_1)Z\beta + (g-1)(r\sigma_{h_s}^2 + rl\sigma_h^2) \\
\text{Lembrando que: } \sigma_g^2 &= \sigma_h^2 + \sigma_m^2; \sigma_{g_s}^2 = \sigma_{h_s}^2 + \sigma_{m_s}^2 \\
(g-1)[r(\sigma_{h_s}^2 + \sigma_{m_s}^2) + rl(\sigma_h^2 + \sigma_m^2)] &= \beta'Z'(P_P - P_1)Z\beta + \\
(g-1)(r\sigma_{h_s}^2 + rl\sigma_h^2) & \\
(g-1)(r\sigma_{m_s}^2 + rl\sigma_m^2) &= \beta'Z'(P_P - P_1)Z\beta
\end{aligned}$$

portanto,

$$\begin{aligned}
E[SQReg.] &= k(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) + \beta'Z'(P_P - P_1)Z\beta \\
&= k(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) + (g-1)(r\sigma_{m_s}^2 + rl\sigma_m^2)
\end{aligned}$$

conseqüentemente,

$$E[QMReg.] = E\left[\frac{1}{k}SQReg.\right] = \sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2 + \frac{(g-1)}{k}(r\sigma_{m_s}^2 + rl\sigma_m^2) \quad (17)$$

$$\begin{aligned}
\bullet E[SQReg. \times Amb.] &= E[SQReg._1] + E[SQReg._2] + \dots + \\
E[SQReg._l] - E[SQReg.] &= \\
\sum_{s=1}^l [k(\sigma_s^2 + r\sigma_{h_s}^2) + (g-1)r\sigma_{m_s}^2] - [k(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) + \\
(g-1)(r\sigma_{m_s}^2 + rl\sigma_m^2)] &= \\
k\sum_{s=1}^l \sigma_s^2 + rk\sum_{s=1}^l \sigma_{h_s}^2 + (g-1)r\sum_{s=1}^l \sigma_{m_s}^2 - k(\sigma^2 + r\sigma_{h_s}^2 + \\
rl\sigma_h^2) - (g-1)(r\sigma_{m_s}^2 + rl\sigma_m^2) &= \\
\text{mas } \sum_{s=1}^l \sigma_s^2 = l\sigma^2; \sum_{s=1}^l \sigma_{h_s}^2 = l(\sigma_{h_s}^2 + \sigma_h^2); \sum_{s=1}^l \sigma_{m_s}^2 = \\
l(\sigma_{m_s}^2 + \sigma_m^2) & \\
kl\sigma^2 + rkl(\sigma_{h_s}^2 + \sigma_h^2) + (g-1)rl(\sigma_{m_s}^2 + \sigma_m^2) - k(\sigma^2 + r\sigma_{h_s}^2 + \\
rl\sigma_h^2) - (g-1)(r\sigma_{m_s}^2 + rl\sigma_m^2) &= \\
k(l-1)\sigma^2 + rk(l-1)\sigma_{h_s}^2 + (g-1)(l-1)r\sigma_{m_s}^2 &
\end{aligned}$$

conseqüentemente,

$$E[QMReg. \times Amb.] = E\left[\frac{1}{k(l-1)}SQReg. \times Amb.\right] = \sigma^2 + r\sigma_{h_s}^2 + \frac{(g-1)}{k}r\sigma_{m_s}^2 \quad (18)$$

$$\begin{aligned}
\bullet E[SQDesv. \times Amb.] &= E[SQDesv._1] + E[SQDesv._2] + \dots + \\
E[SQDesv._l] - E[SQDesv.] &= \\
\sum_{s=1}^l (g-k-1)(\sigma_s^2 + r\sigma_{h_s}^2) - (g-k-1)(\sigma^2 + r\sigma_{h_s}^2 + rl\sigma_h^2) &= \\
(g-k-1)(l-1)\sigma^2 + r(g-k-1)(l-1)\sigma_{h_s}^2 &
\end{aligned}$$

conseqüentemente,

$$E[QMDesv. \times Amb.] = E \left[\frac{1}{(g-k-1)(l-1)} SQDesv. \times Amb. \right] = \sigma^2 + r\sigma_{hs}^2. \quad (19)$$

Assim, pode-se esquematizar o quadro de análise de variância (ANAVA) apresentado na tabela 1.

Tabela - 1: Esquema da análise de variância conjunta com o desdobramento da soma de quadrados de tratamentos e interação, em efeitos explicados pelos marcadores moleculares e desvios, com as respectivas esperanças matemáticas dos quadrados médios.

FV	GL	SQ	E(QM)
Ambientes	$l-1$	$\mathbf{y}'(P_L - P_1)\mathbf{y}$	$\sigma^2 + r\sigma_g^2 + rg\sigma_s^2$
Progênes	$g-1$	$\mathbf{y}'(P_P - P_1)\mathbf{y}$	$\sigma^2 + r\sigma_g^2 + rl\sigma_g^2$
Prog. x Amb.	$(g-1)(l-1)$	$\mathbf{y}'(P - P_L - P_P + P_1)\mathbf{y}$	$\sigma^2 + r\sigma_g^2$
Resíduo	$lg(r-1)$	$\mathbf{y}'(I - P)\mathbf{y}$	σ^2
Reg.	k	$\mathbf{y}'(P_R - P_1)\mathbf{y}$	E^*
Desv.	$g-k-1$	$\mathbf{y}'(P_P - P_R)\mathbf{y}$	$\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2$
Reg. x Amb.	$k(l-1)$	$\sum_{s=1}^l SQR_s - \mathbf{y}'(P_R - P_1)\mathbf{y}$	$\sigma^2 + r\sigma_{hs}^2 + \frac{(g-1)}{k} r\sigma_{ms}^2$
Desv. x Amb.	$(g-k-1)(l-1)$	$\sum_{s=1}^l SQDR_s - \mathbf{y}'(P_P - P_R)\mathbf{y}$	$\sigma^2 + r\sigma_{hs}^2$
Total	$lrg-1$		

onde $E^* = \sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2 + \frac{(g-1)r}{k}(\sigma_{ms}^2 + l\sigma_m^2)$;
 $P_R = ZZ'$; SQR_s = Soma de quadrados de regressão no ambiente s ;
 $SQDR_s$ = Soma de quadrados de desvios no ambiente s ;

Vale a pena ressaltar que essas expressões das esperanças matemáticas dos quadrados médios independem das proporções genotípicas da população segregante utilizada, quanto ao loco marcador, bem como de desvios dessas proporções em relação às proporções esperadas (1:2:1 em uma geração F_2 , 1:1 em RC, etc.).

4 Estimadores dos componentes de variância

Com base nas expressões das esperanças matemáticas dos quadrados médios, apresentadas na tabela 1, obtiveram-se os estimadores das componentes de variância residual (σ^2), genética (σ_g^2), de ambientes (σ_s^2), interação progênes x ambientes (σ_{gs}^2), devida ao efeito de interação desvios x ambientes (σ_{hs}^2), interação marcas x ambientes (σ_{ms}^2), desvios (σ_h^2) e marcas (σ_m^2), dados por:

$$\hat{\sigma}^2 = QMResiduo; \tag{20}$$

$$\hat{\sigma}_{hs}^2 = \frac{QMDesv. \times Amb. - QMResiduo}{r}; \tag{21}$$

$$\hat{\sigma}_{ms}^2 = \frac{k(QMReg. \times Amb. - QMDesv. \times Amb.)}{r(g-1)}; \tag{22}$$

$$\hat{\sigma}_h^2 = \frac{QMDesv. - QMDesv. \times Amb.}{rl}; \tag{23}$$

$$\hat{\sigma}_m^2 = \frac{k[(QMReg - QMDesv) - (QMReg \times Amb - QMDesv \times Amb)]}{rl(g-1)}; \tag{24}$$

$$\hat{\sigma}_{gs}^2 = \frac{QMProg. \times Amb. - QMResiduo}{r}; \tag{25}$$

$$\hat{\sigma}_g^2 = \frac{QMProg. - QMProg. \times Amb.}{rl}; \tag{26}$$

$$\hat{\sigma}_s^2 = \frac{QMAmb. - QMProg. \times Amb.}{rg}; \tag{27}$$

5 Testes de hipóteses

Sob normalidade tem-se que:

$$\begin{aligned} \frac{SQResiduo}{\sigma^2} &\sim \chi_{gl(r-1)}^2, \text{ central;} \\ \frac{SQDesv. \times Amb.}{\sigma^2 + r\sigma_{hs}^2} &\sim \chi_{(g-k-1)(l-1)}^2, \text{ central;} \\ \frac{SQReg. \times Amb.}{\sigma^2 + r\sigma_{hs}^2 + \frac{(g-1)}{r}k\sigma_{ms}^2} &\sim \chi_{k(l-1)}^2, \text{ central;} \\ \frac{SQDesv.}{\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2} &\sim \chi_{(g-k-1)}^2, \text{ central;} \\ \frac{SQReg.}{\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2 + \frac{g-1}{k}(r\sigma_{ms}^2 + rl\sigma_m^2)} &\sim \chi_{(k)}^2, \text{ central;} \end{aligned}$$

e são independentes duas a duas. Nestas condições, as hipóteses de nulidade dos componentes σ_{hs}^2 , σ_{ms}^2 , σ_h^2 e σ_m^2 , podem ser testadas via as estatísticas F apresentadas na tabela 2.

Tabela - 2: Estatísticas F para o teste das hipóteses de nulidade dos componentes de variância, com suas respectivas distribuições associadas

Hipótese	Estatística F	Distribuição associada
$\left\{ \begin{array}{l} H_0: \sigma_{h^s}^2 = 0 \text{ vs} \\ H_a: \sigma_{h^s}^2 > 0 \end{array} \right.$	$\frac{QMDesv. \times Amb.}{QMResiduo}$	$\mathcal{F}_{(g-k-1)(l-1), lg(r-1)}$
$\left\{ \begin{array}{l} H_0: \sigma_{m^s}^2 = 0 \text{ vs} \\ H_a: \sigma_{m^s}^2 > 0 \end{array} \right.$	$\frac{QMReg. \times Amb.}{QMDesv. \times Amb.}$	$\mathcal{F}_{k(l-1), (g-k-1)(l-1)}$
$\left\{ \begin{array}{l} H_0: \sigma_h^2 = 0 \text{ vs} \\ H_a: \sigma_h^2 > 0 \end{array} \right.$	$\frac{QMDesv.}{QMDesv. \times Amb.}$	$\mathcal{F}_{g-k-1, (g-k-1)(l-1)}$
$\left\{ \begin{array}{l} H_0: \sigma_m^2 = 0 \text{ vs} \\ H_a: \sigma_m^2 > 0 \end{array} \right.$	$\frac{QMReg. + QMDesv. \times Amb.}{QMDesv. + QMReg. \times Amb.}$	$\mathcal{F}_{v1, v2}$

$$v1 \approx \frac{[QMReg + QMDesv. \times Ambientes]^2}{\frac{[QMReg]^2}{GLReg} + \frac{[QMDesv. \times Ambientes]^2}{GLDesv. \times Ambientes}} e$$

$$v2 \approx \frac{[QMDesv + QMReg. \times Ambientes]^2}{\frac{[QMDesv]^2}{GLDesv} + \frac{[QMReg \times Ambientes]^2}{GLReg \times Ambientes}}$$

6 Estimadores das proporções da variância genética e da variância da interação genótipos x ambientes explicadas pelos marcadores moleculares

Uma extensão natural do estimador \hat{p} apresentado por Bearzoti & Vencovsky (1998) para a situação em que as progênies são avaliadas em vários ambientes é:

$$\hat{p}_m = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2}, \quad (28)$$

em que o estimador $\hat{\sigma}_m^2$ é dado por 24, e o estimador $\hat{\sigma}_h^2$, por 23.

Um estimador da proporção da variância da interação genótipos x ambientes, devida à interação marcadores moleculares x ambientes, é dado por:

$$\hat{p}_{ms} = \frac{\hat{\sigma}_{ms}^2}{\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2}, \quad (29)$$

em que o estimador $\hat{\sigma}_{ms}^2$ é dado por 22, e o estimador $\hat{\sigma}_{hs}^2$, por 21.

6.1 Esperança e Variância de $\hat{\rho}_m$ e $\hat{\rho}_{ms}$

Em virtude da inviabilidade de expressar as esperanças e variâncias de $\hat{\rho}_m$ e $\hat{\rho}_{ms}$ explicitamente, derivaram-se expressões aproximadas para esses parâmetros, utilizando-se a expansão em série de Taylor da razão entre duas variáveis aleatórias (X e Y), conforme apresentado por Mood *et al.* (1974), cujas expressões são:

$$E\left[\frac{X}{Y}\right] = \frac{E[X]}{E[Y]} - \frac{Cov[X, Y]}{E^2[Y]} + \frac{E[X]}{E^3[Y]}V[Y] \quad (30)$$

$$V\left[\frac{X}{Y}\right] = \frac{E^2[Y]V[X] + E^2[X]V[Y] - 2E[X]E[Y]Cov[X, Y]}{E^4[Y]} \quad (31)$$

Dado que as variáveis aleatórias em questão são combinações lineares de quadrados médios, suas variâncias foram obtidas por:

$$V[QM] = \frac{2[E[QM]]^2}{v} \quad (\text{Searle et al., 1992}) \quad (32)$$

em que v são os graus de liberdade associados ao quadrado médio.

Um estimador não tendencioso, desta variância (32), apresentado por Searle *et al.* (1992) é:

$$\hat{V}[QM] = \frac{2(QM)^2}{v + 2} \quad (33)$$

Esperança de $\hat{\rho}_m$

Sendo $\hat{\rho}_m = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2}$ (28), e fazendo $x = \hat{\sigma}_m^2$ e $y = \hat{\sigma}_m^2 + \hat{\sigma}_h^2$ em 30 decorre que:

$$E[\hat{\rho}_m] = E\left[\frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2}\right] = \frac{E[\hat{\sigma}_m^2]}{E[\hat{\sigma}_m^2 + \hat{\sigma}_h^2]} - \frac{Cov[\hat{\sigma}_m^2, \hat{\sigma}_m^2 + \hat{\sigma}_h^2]}{E^2[\hat{\sigma}_m^2 + \hat{\sigma}_h^2]} + \frac{E[\hat{\sigma}_m^2]}{E^3[\hat{\sigma}_m^2 + \hat{\sigma}_h^2]}V[\hat{\sigma}_m^2 + \hat{\sigma}_h^2]$$

$$= \frac{\sigma_m^2}{\sigma_m^2 + \sigma_h^2} - \frac{V(\hat{\sigma}_m^2)Cov[\hat{\sigma}_m^2, \hat{\sigma}_h^2]}{(\sigma_m^2 + \sigma_h^2)^2} + \frac{\sigma_m^2}{(\sigma_m^2 + \sigma_h^2)^3}[V(\hat{\sigma}_m^2) + V(\hat{\sigma}_h^2) + 2Cov(\hat{\sigma}_m^2, \hat{\sigma}_h^2)]$$

∴

$$E(\hat{\rho}_m) = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_h^2} - \frac{V(\hat{\sigma}_m^2)Cov[\hat{\sigma}_m^2, \hat{\sigma}_h^2]}{(\sigma_m^2 + \sigma_h^2)^2} + \frac{\sigma_m^2}{(\sigma_m^2 + \sigma_h^2)^3}[V(\hat{\sigma}_m^2) + V(\hat{\sigma}_h^2) + 2Cov(\hat{\sigma}_m^2, \hat{\sigma}_h^2)], \quad (34)$$

com,

$$\begin{aligned}
 V(\hat{\sigma}_h^2) &= V\left[\frac{1}{rl}(QMDesv. - QMDesv. \times Amb.)\right] \\
 &= \frac{1}{(rl)^2} [V(QMDesv.) + V(QMDesv. \times Amb.) - \\
 &\quad - 2Cov(QMDesv., QMDesv. \times Amb.)] \\
 &= \frac{1}{(rl)^2} \left[\frac{2[E(QMDesv.)]^2}{g-k-1} + \frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)} \right] \\
 &= \frac{1}{(rl)^2} \left[\frac{2(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2)^2}{g-k-1} + \frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)} \right]; \quad (35)
 \end{aligned}$$

$$\begin{aligned}
 V(\hat{\sigma}_m^2) &= V\left[\frac{k}{rl(g-1)}[(QMReg. - QMDesv.) - (QMReg. \times Amb. - QMDesv. \times Amb.)]\right] \\
 &= \left(\frac{k}{rl(g-1)}\right)^2 [V(QMReg.) + V(QMDesv.) + V(QMReg. \times Amb.) \\
 &\quad + V(QMDesv. \times Amb.) - 2Cov(QMReg., QMDesv.) \\
 &\quad - 2Cov(QMReg., QMDesv. \times Amb.) + 2Cov(QMReg., QMReg. \times Amb.) \\
 &\quad - 2Cov(QMReg., QMDesv. \times Amb.) + 2Cov(QMDesv., QMReg. \times Amb.) \\
 &\quad - 2Cov(QMDesv., QMDesv. \times Amb.) - 2Cov(QMDesv., QMReg. \times Amb.)] \\
 &= \left(\frac{k}{rl(g-1)}\right)^2 \left[\frac{2[E(QMReg.)]^2}{k} + \frac{2[E(QMDesv.)]^2}{g-k-1} + \frac{2[E(QMReg. \times Amb.)]^2}{k(l-1)} \right. \\
 &\quad \left. + \frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)} \right] \\
 &= \left(\frac{k}{rl(g-1)}\right)^2 \left[\frac{2(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2 + \frac{r(g-1)}{k}(\sigma_{ms}^2 + l\sigma_m^2))^2}{k} + \frac{2(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2)^2}{g-k-1} \right. \\
 &\quad \left. + \frac{2(\sigma^2 + r\sigma_{hs}^2 + \frac{r(g-1)}{k}\sigma_{ms}^2)^2}{k(l-1)} + \frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)} \right]; \quad (36)
 \end{aligned}$$

$$\begin{aligned}
 Cov(\hat{\sigma}_m^2, \hat{\sigma}_h^2) &= \\
 &= Cov\left[\frac{k}{(g-1)(rl)}(QMReg. - QMDesv. - QMReg. \times Amb. + QMDesv. \times Amb.), \right. \\
 &\quad \left. \frac{1}{rl}(QMDesv. - QMDesv. \times Amb.)\right] \\
 &= \frac{k}{(rl)^2(g-1)} [Cov(QMReg., QMDesv.) - Cov(QMReg., QMDesv. \times Amb.) \\
 &\quad - Cov(QMDesv., QMDesv.) + Cov(QMDesv., QMDesv. \times Amb.) \\
 &\quad - Cov(QMReg. \times Amb., QMDesv.) + Cov(QMReg. \times Amb., QMDesv. \times Amb.) \\
 &\quad + Cov(QMDesv. \times Amb., QMDesv.) - Cov(QMDesv. \times Amb., QMDesv. \times Amb.)] \\
 &= \frac{k}{(rl)^2(g-1)} [V(QMDesv.) - V(QMDesv. \times Amb.)] \\
 &= \frac{k}{(rl)^2(g-1)} \left[\frac{2[E(QMDesv.)]^2}{g-k-1} - \frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)} \right] \\
 &= \frac{k}{(rl)^2(g-1)} \left[\frac{2(\sigma^2 + r\sigma_{hs}^2 + rl\sigma_h^2)^2}{g-k-1} - \frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)} \right]. \quad (37)
 \end{aligned}$$

Variância de $\hat{\rho}_m$

Fazendo $x = \hat{\sigma}_m^2$ e $y = \hat{\sigma}_m^2 + \hat{\sigma}_h^2$ em 31 tem-se:

$$\begin{aligned}
 \text{Var}[\hat{p}_m] &= \text{Var} \left[\frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2} \right] = \\
 &= \frac{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^2 V(\hat{\sigma}_m^2) + (\hat{\sigma}_m^2)^2 V(\hat{\sigma}_h^2) - 2\hat{\sigma}_m^2 (\hat{\sigma}_m^2 + \hat{\sigma}_h^2) \text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)}{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^4} \\
 &= \frac{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^2 V(\hat{\sigma}_m^2) + \hat{\sigma}_m^4 [V(\hat{\sigma}_m^2) + V(\hat{\sigma}_h^2) + 2\text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)] - 2\hat{\sigma}_m^2 (\hat{\sigma}_m^2 + \hat{\sigma}_h^2) [V(\hat{\sigma}_m^2) + \text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)]}{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^4} \\
 &= \frac{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^2 V(\hat{\sigma}_m^2) + \hat{\sigma}_m^4 V(\hat{\sigma}_m^2) + \hat{\sigma}_m^4 V(\hat{\sigma}_h^2) - 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 V(\hat{\sigma}_m^2) - 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 \text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)}{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^4} \\
 &= \frac{(\hat{\sigma}_m^4 + 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 + \hat{\sigma}_h^4) V(\hat{\sigma}_m^2) + \hat{\sigma}_m^4 V(\hat{\sigma}_h^2) + \hat{\sigma}_m^2 V(\hat{\sigma}_h^2) - 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 V(\hat{\sigma}_m^2) - 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 \text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)}{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^4} \\
 &\quad \therefore \\
 \text{Var}(\hat{p}_m) &= \text{Var} \left[\frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2} \right] = \frac{\hat{\sigma}_h^4 V(\hat{\sigma}_m^2) + \hat{\sigma}_m^4 V(\hat{\sigma}_h^2) - 2\hat{\sigma}_m^2 \hat{\sigma}_h^2 \text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)}{(\hat{\sigma}_m^2 + \hat{\sigma}_h^2)^4}, \quad (38)
 \end{aligned}$$

com,

$V(\hat{\sigma}_m^2)$ dado por 36;
 $V(\hat{\sigma}_h^2)$ dado por 35 e
 $\text{Cov}(\hat{\sigma}_m^2, \hat{\sigma}_h^2)$ dado por 37.

Esperança de \hat{p}_{ms}

Dado que $\hat{p}_{ms} = \frac{\hat{\sigma}_{ms}^2}{\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2}$ (29), então fazendo $x = \hat{\sigma}_{ms}^2$ e $y = \hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2$ em 30 decorre:

$$\begin{aligned}
 E[\hat{p}_{ms}] &= E \left[\frac{\hat{\sigma}_{ms}^2}{\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2} \right] = \frac{E[\hat{\sigma}_{ms}^2]}{E[\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2]} - \frac{\text{Cov}[\hat{\sigma}_{ms}^2, \hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2]}{E^2[\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2]} + \frac{E[\hat{\sigma}_{ms}^2]}{E^3[\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2]} V[\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2] \\
 &= \frac{\sigma_{ms}^2}{\sigma_{ms}^2 + \sigma_{hs}^2} - \frac{V(\hat{\sigma}_{ms}^2) + \text{Cov}[\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2]}{(\sigma_{ms}^2 + \sigma_{hs}^2)^2} + \frac{\sigma_{ms}^2}{(\sigma_{ms}^2 + \sigma_{hs}^2)^3} [V(\hat{\sigma}_{ms}^2) + V(\hat{\sigma}_{hs}^2) + 2\text{Cov}(\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2)] \\
 &\quad \therefore \\
 E(\hat{p}_{ms}) &= \frac{\sigma_{ms}^2}{\sigma_{ms}^2 + \sigma_{hs}^2} - \frac{V(\hat{\sigma}_{ms}^2) + \text{Cov}[\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2]}{(\sigma_{ms}^2 + \sigma_{hs}^2)^2} + \frac{\sigma_{ms}^2}{(\sigma_{ms}^2 + \sigma_{hs}^2)^3} [V(\hat{\sigma}_{ms}^2) + V(\hat{\sigma}_{hs}^2) \\
 &\quad + 2\text{Cov}(\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2)], \quad (39)
 \end{aligned}$$

com

$$\begin{aligned}
 V(\hat{\sigma}_{ms}^2) &= V \left[\frac{k}{r(g-1)} (QMReg. \times Amb. - QMDesv. \times Amb.) \right] \\
 &= \left(\frac{k}{r(g-1)} \right)^2 [V(QMReg. \times Amb.) + V(QMDesv. \times Amb.) \\
 &\quad - 2\text{Cov}(QMReg. \times Amb., QMDesv. \times Amb.)] \\
 &= \left(\frac{k}{r(g-1)} \right)^2 \left[\frac{2[E(QMReg. \times Amb.)]^2}{k(l-1)} + \frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)} \right] \\
 &= \left(\frac{k}{r(g-1)} \right)^2 \left[\frac{2(\sigma^2 + r\sigma_{hs}^2 + \frac{r(g-1)}{k}\sigma_{ms}^2)^2}{k(l-1)} + \frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)} \right]. \quad (40)
 \end{aligned}$$

$$\begin{aligned}
V(\hat{\sigma}_{hs}^2) &= V\left[\frac{1}{r}(QMDesv. \times Amb. - QMRes.)\right] \\
&= \frac{1}{r^2}[V(QMDesv \times Amb.) + V(QMRes.) - 2cov(QMDesv \times Amb., QMRes.)] \\
&= \frac{1}{r^2}\left[\frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)} + \frac{2[E(QMRes.)]^2}{lg(r-1)}\right] \\
&= \frac{1}{r^2}\left[\frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)} + \frac{2(\sigma^2)^2}{lg(r-1)}\right]. \tag{41}
\end{aligned}$$

$$\begin{aligned}
Cov(\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2) &= \\
&= Cov\left[\frac{k}{r(g-1)}(QMReg. \times Amb. - QMDesv. \times Amb.), \frac{1}{r}(QMDesv. \times Amb. - QMRes.)\right] \\
&= \frac{k}{r^2(g-1)}[Cov(QMReg. \times Amb., QMDesv. \times Amb.) - Cov(QMReg. \times Amb., QMRes.) \\
&\quad - Cov(QMDesv. \times Amb., QMDesv. \times Amb.) - Cov(QMDesv. \times Amb., QMRes.)] \\
&= \frac{k}{r^2(g-1)}V(QMDesv. \times Amb.) = \frac{k}{r^2(g-1)}\left[\frac{2[E(QMDesv. \times Amb.)]^2}{(g-k-1)(l-1)}\right] \\
&= \frac{k}{r^2(g-1)}\left[\frac{2(\sigma^2 + r\sigma_{hs}^2)^2}{(g-k-1)(l-1)}\right]. \tag{42}
\end{aligned}$$

Variância de \hat{p}_{ms}

De 29 e 31 tem-se:

$$\begin{aligned}
Var[\hat{p}_{ms}] &= Var\left[\frac{\hat{\sigma}_{ms}^2}{\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2}\right] \\
&= \frac{\hat{\sigma}_{hs}^4 V(\hat{\sigma}_{ms}^2) + \hat{\sigma}_{ms}^4 V(\hat{\sigma}_{hs}^2) - 2\hat{\sigma}_{ms}^2 \hat{\sigma}_{hs}^2 Cov(\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2)}{(\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2)^4} \tag{43}
\end{aligned}$$

com

$$\begin{aligned}
&V(\hat{\sigma}_{ms}^2) \text{ dada por 40;} \\
&V(\hat{\sigma}_{hs}^2) \text{ dada por 41 e} \\
&Cov(\hat{\sigma}_{ms}^2, \hat{\sigma}_{hs}^2) \text{ dada por 42;}
\end{aligned}$$

Estas expressões das variâncias de \hat{p}_m e de \hat{p}_{ms} podem ser empregadas na construção de intervalos de confiança, utilizando a aproximação normal. Porém, devido ao fato de não serem expressões triviais, necessitando da utilização de um grande número de outras estimativas ($\hat{\sigma}^2$, $\hat{\sigma}_m^2$, $\hat{\sigma}_h^2$, $\hat{\sigma}_{ms}^2$ e $\hat{\sigma}_{hs}^2$), as quais já estão sujeitas a erros, os valores obtidos por estas expressões podem ser um tanto quanto viesados. Adicionando-se o fato de Bearzoti & Vencovsky terem verificado, via simulação computacional, que

a aproximação normal para a distribuição de \hat{p} não é satisfatória para valores elevados de p ($p = 0,9$), mesmo ao utilizar um grande número de progênies ($g=100$), sugere-se lançar mão de métodos de reamostragem, tais como o “bootstrap”, para obtenção dos intervalos de confiança para \hat{p}_m e de \hat{p}_{ms} .

6.1.1 Probabilidade de obtenção de estimativas de \hat{p}_m e \hat{p}_{ms} maiores que um

Conforme mostrado por Bearzoti & Vencovsky (1998), para o caso de experimentos conduzidos em um único ambiente, a obtenção de estimativas de p fora do intervalo paramétrico ($0 \leq p \leq 1$) não é tão incomum, principalmente, para valores elevados de p ($p > 0,9$). Sendo o estimador \hat{p}_m dado por:

$\hat{p}_m = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2}$, tem-se que:

$$\begin{aligned}
 P(\hat{p}_m > 1) &= P\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_m^2 + \hat{\sigma}_h^2} > 1\right) \\
 &= P(\hat{\sigma}_h^2 < 0) \\
 &= P(QMDesv. < QMDesv. \times Amb.), \text{ de 35} \\
 &= P\left(\frac{QMDesv.}{QMDesv. \times Amb.} < 1\right) \\
 &= P\left(\frac{\sigma^2 + r\sigma_{hs}^2 + r\sigma_h^2}{\sigma^2 + r\sigma_{hs}^2} \mathcal{F}_{(g-k-1), (g-k-1)(l-1)} < 1\right) \\
 &= P\left(\mathcal{F}_{(g-k-1), (g-k-1)(l-1)} < \frac{\sigma^2 + r\sigma_{hs}^2}{\sigma^2 + r\sigma_{hs}^2 + r\sigma_h^2}\right) \quad (44)
 \end{aligned}$$

Verifica-se na eq. 44 que o aumento do número de repetições por local não altera os parâmetros da distribuição, mas diminui o valor f , abaixo do qual resultam estimativas de p_m maiores que um, e, conseqüentemente, a probabilidade de ocorrência desse evento. Aumentando-se o número de ambientes, também diminui-se o valor f , porém essa alteração provoca simultaneamente uma alteração na distribuição \mathcal{F} , que, mantido o mesmo ponto de corte, aumenta a probabilidade de ocorrência de valores inferiores a este ponto. Os parâmetros g (número de progênies) e k (número de marcas) alteram apenas o formato da distribuição, sendo que, para um mesmo ponto de corte, o aumento de g , mantendo k fixo, diminui a área sob a curva abaixo do ponto de corte. Por outro lado, o aumento de k , mantendo g fixo aumenta essa área, aumentando assim, a probabilidade de obtenção de estimativas de p_m maiores que um.

Para o estimador \hat{p}_{ms} , tem-se:

$$\begin{aligned}
 P(\hat{p}_{ms} > 1) &= P\left(\frac{\hat{\sigma}_{ms}^2}{\hat{\sigma}_{ms}^2 + \hat{\sigma}_{hs}^2} > 1\right) \\
 &= P(\hat{\sigma}_{hs}^2 < 0) \\
 &= P(QMDesv. \times Amb. < QMResíduo), \text{ de 21} \\
 &= P\left(\frac{QMDesv. \times Amb.}{QMResíduo} < 1\right) \\
 &= P\left(\frac{\sigma^2 + r\sigma_{hs}^2}{\sigma^2} \mathcal{F}_{(g-k-1)(l-1), gl(r-1)} < 1\right) \\
 &= P\left(\mathcal{F}_{(g-k-1)(l-1), gl(r-1)} < \frac{\sigma^2}{\sigma^2 + r\sigma_{hs}^2}\right) \quad (45)
 \end{aligned}$$

No caso da proporção da interação genótipos x ambientes, devida à interação marcadores x ambientes, tem-se que os parâmetros experimentais número de progênies avaliadas (g) e número de ambientes (l) atuam somente sobre o formato da distribuição, sendo que, para um mesmo ponto de corte, o aumento desses valores diminui a probabilidade de ocorrência de um valor inferior ao ponto de corte. Ou seja, tanto o aumento do número de progênies, quanto o de ambientes nas quais as progênies são avaliadas diminuem a probabilidade de obtenção de estimativas de p_{ms} maiores que um, sendo que a maior redução é obtida aumentando-se o número de ambientes. Assim como no caso de \hat{p}_m , o incremento do número de marcas utilizadas aumenta a probabilidade de obtenção de estimativas de p_{ms} maiores que um.

A título de exemplo, encontram-se apresentadas na Tabela 3 as probabilidades de obtenção de estimativas de p_m e p_{ms} maiores que um, em função do número de marcas utilizadas no modelo (k) e dos valores paramétricos de p_m e p_{ms} , considerando um experimento, no qual foram avaliadas 100 progênies em quatro ambientes, com duas repetições por ambiente. Considerou-se que o caráter apresentou uma herdabilidade ao nível de média de progênies igual a 0,40, e que $\sigma_g^2 = \sigma_{gs}^2$. Verifica-se nesta tabela que, para valores elevados, tanto de p_m , quanto de p_{ms} , são grandes as probabilidades de obtenção de estimativas desses parâmetros maiores que um. Isso ocorre mesmo se trabalhando com 200 indivíduos, sendo que essas probabilidades crescem como o aumento no número de marcas utilizadas (k). Como anteriormente comentado, um modo de reduzir a probabilidade de obtenção de

estimativas maiores que um consiste em aumentar o número de repetições neste exemplo o aumento para 3 repetições diminuiu o probabilidade de obtenção de estimativas de p_m maiores que um de 0,37, quando se avaliaram 100 progênies com $p = 0,9$, para 0,31 e $P(p_{ms} > 1)$ de 0,43 para 0,39 na mesma situação.

Tabela - 3: Probabilidade de obtenção de estimativas de p_m e p_{ms} maiores que um, em um experimento no qual foram avaliadas 100 e 200 progênies em 4 ambientes com 2 repetições por ambiente. Considerou-se $h^2 = 0,40$, e $\sigma_g^2 = \sigma_{gs}^2$.

$P(\hat{p}_m > 1)$						
p_m	$g = 100$			$g = 200$		
	$k = 10$	$k = 15$	$k = 20$	$k = 10$	$k = 15$	$k = 20$
0,9	0,3641	0,3685	0,3730	0,2984	0,3010	0,3037
0,7	0,1533	0,1607	0,1686	0,0651	0,0677	0,0705
0,5	0,0533	0,0587	0,0647	0,0087	0,0095	0,0104
0,3	0,0163	0,0190	0,0222	0,0008	0,0009	0,0013
0,1	0,0045	0,0057	0,0071	0,0000	0,0000	0,0001

$P(\hat{p}_{ms} > 1)$						
p_{ms}	$g = 100$			$g = 200$		
	$k = 10$	$k = 15$	$k = 20$	$k = 10$	$k = 15$	$k = 20$
0,9	0,4327	0,4343	0,4361	0,4011	0,4020	0,4030
0,7	0,3041	0,3077	0,3116	0,2282	0,2302	0,2321
0,5	0,1999	0,2045	0,2095	0,1113	0,1133	0,1153
0,3	0,1233	0,1278	0,1328	0,0468	0,0482	0,0497
0,1	0,0716	0,0755	0,0798	0,0171	0,0179	0,0187

7 Conclusões

- i. Obteve-se um estimador da proporção da variância genética, explicada pelos marcadores moleculares, aplicável ao índice de seleção de Lande & Thompson (1992), no caso de experimentos conduzidos em mais de um ambiente, bem como um estimador da proporção da variância da interação genótipos x ambientes explicada pela interação QTLs x ambientes.
- ii. Os estimadores apresentados independem de desvios das frequências alélicas dos marcadores em relação às esperadas,

porém possuem alta probabilidade de geração de estimativas fora do intervalo paramétrico, que pode ser reduzida com o aumento do número de ambientes e/ou repetições nos quais os genótipos são avaliados.

- iii. As expressões da esperança e da variância dos estimadores apresentados, obtidas via expansão em série de Taylor, não são triviais; em vista disto sugere-se obter tais estimativas via processos de reamostragem.

Agradecimentos

Os autores agradecem à Coordenação para Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

SILVA, H.D.; VENCOVSKY, R.; RUMIN, G.C.R.- Partition of the genotype x environment interaction variance in effects explained by molecular markers and deviations. *Rev. Mat. Estat.* (São Paulo), v. 20, p.207-226, 2002.

- *ABSTRACT: One of the problems found in the QTL analysis refer to the QTL x environment interaction. We consider this aspect by partitioning the genotype x environment interaction variance in components explained by the molecular markers and deviations. This allowed estimating the proportion of the genetic variance (p_m), and genotype x environment variance (p_{ms}), explained by the markers. These estimators are not affected by deviations of allelic frequencies of the markers in relation to the expected values (1:2:1 in a F2 generation, 1:1 in a backcross, etc). However, there is a high probability of obtaining estimates out of the parametric range, specially for high values of this proportion. Nevertheless, these probabilities can be reduced by increasing the number of replications and/or environments where the progenies are evaluated.*

- **KEYWORDS:** *QTL, QTL's x environment, Marker assisted selection, variance components.*

Referências

- BEARZOTI, E.; VENCOSKY, R. Estimation of the proportion of genetic variance explained by molecular markers. *Genet. Mol. Biol.*, v.21, n.4, p.557-66, 1998.
- CRUZ, C. D.; REGAZZI, A. J. *Modelos biométricos aplicados ao melhoramento genético* Viçosa, MG: Imprensa Universitária, 1994. 390p.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, v.126, p.743-56, 1990.
- MOOD, A. L.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. 3.ed. Tokio: McGraw-Hill, Kogakusha, 1974. 564p.
- SEARLE, S. R. *Linear models*. New York: John-Wiley, 1971. 501p.
- SEARLE, S. R.; CASELA, J.; McCULLOCH, C. E. *Variance components*. New York: John-Wiley, 1992. 532p.
- VENCOSKY, R.; BARRIGA, P. *Genética biométrica no fitomelhoramento* Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.
- WANG, D. L.; ZHU, J.; LI, Z. K.; PATERSON, A.H. Mapping QTLs with epistatic effects and QTL x environment interaction by mixed linear model approaches. *Theor. Appl. Genet.*, v.99, p.1255-1264, 1999.

Recebido em 15.10.2001.