

EQUAÇÃO DE REGRESSÃO LINEAR MÚLTIPLA PARA ESTIMATIVA DO ERRO EXPERIMENTAL

MULTIPLE LINEAR REGRESSION EQUATION AS AN ESTIMATION OF EXPERIMENTAL ERROR

Alessandro Dal'Col Lúcio¹ David Ariovaldo Banzatto² Lindolfo Storck³
Thomas Newton Martin⁴ Leandro Homrich Lorentz⁵

RESUMO

Este trabalho teve por objetivo estimar equações de regressão linear múltipla tendo, como variáveis explicativas, as demais características avaliadas em experimento de milho e, como variáveis principais, a diferença mínima significativa em percentagem da média (DMS%) e quadrado médio do erro (QMe), para peso de grãos. Com 610 experimentos conduzidos na Rede de Ensaios Nacionais de Competição de Cultivares de Milho, realizados entre 1986 e 1996 (522 experimentos) e em 1997 (88 experimentos), estimaram-se duas equações de regressão, com os 522 experimentos, validando estas pela análise de regressão simples entre os valores reais e os estimados pelas equações, com os 88 restantes, observando que, para a DMS% a equação não estimava o mesmo valor que a fórmula original e, para o QMe, a equação poderia ser utilizada na estimação. Com o teste de Lilliefors, verificou-se que os valores do QMe aderiam à distribuição normal padrão e foi construída uma tabela de classificação dos valores do QMe, baseada nos valores observados na análise da variância dos experimentos e nos estimados pela equação de regressão.

Palavras-chave: *precisão experimental, análise de regressão, competição de cultivares, Zea mays.*

SUMMARY

The aims of this study were to estimate the multiple linear regression equation and to verify the possible relationship between dependent and independent variables. Dependent variables were the mean percentage of the least significant difference (LSD%) and the mean square of the error (MSe) for grain yield. Data from 522 experiments conducted from 1986 to 1996 and 88 experiments conducted in 1997 were used in a total of 610 experiments of the National Competition of Maize Cultivars. In the 522 experiments, two regression equations validated by the analysis of simple regression between the real

values and the foreseen for the equations were estimated, in the 88 experiments, it was observed that the regression equation was not a good estimation for the same original value for LSD%, but the equation can be used for the estimation of MSe. The application of Lilliefors test resulted in normal pattern distribution of MSe values. One classification table of MSe values was built based on observed values of variance analysis of the experiments and on the regression equation estimated value.

Key words: *experimental precision, regression analysis, cultivar competition, Zea mays.*

INTRODUÇÃO

Devido à importância da cultura do milho no Brasil, várias pesquisas são realizadas para exploração do potencial produtivo desta cultura, e o Centro Nacional de Pesquisa de Milho e Sorgo/EMBRAPA coordena o Ensaio Nacional de Cultivares de Milho, com o objetivo de avaliar cultivares de milho em comercialização ou em vias de lançamento, sendo seus resultados utilizados como suporte para elaboração de uma lista oficial de recomendação de cultivares de milho para as diferentes regiões do país.

Avaliando estes resultados, LÚCIO (1999) percebeu a existência de experimentos com precisão inadequada, que podem resultar em ineficientes discriminações entre os tratamentos (cultivares), induzindo a conclusões incorretas ao se fazerem as recomendações aos produtores. Esta falta

¹Engenheiro Agrônomo, Doutor, Professor Adjunto, Departamento de Fitotecnia (DF), Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS. E-mail: adlucio@ccr.ufsm.br. Autor para correspondência.

²Engenheiro Agrônomo, Doutor, Professor Adjunto, Departamento de Ciências Exatas, UNESP/FCAV, Jaboticabal, SP.

³Engenheiro Agrônomo, Doutor, Professor Titular, DF, CCR, UFSM, Bolsista CNPq.

⁴Aluno do Curso de Graduação em Agronomia, CCR, UFSM, bolsista CNPq.

⁵Aluno do Curso de Graduação em Agronomia, CCR, UFSM, bolsista FAPERGS.

de controle de qualidade dos experimentos de milho, resulta em prejuízos, muitas vezes, incalculáveis e significativos economicamente.

A interferência direta do erro experimental na análise dos dados se dá na estatística F e na variância da média estimada, usada nos procedimentos para comparações múltiplas. Nessas situações, quanto maior o erro, reduz-se a probabilidade de se obterem diferenças significativas entre as médias dos tratamentos, mesmo havendo tais diferenças.

Segundo autores como STEEL & TORRIE (1960), BANZATTO & KRONKA (1995) e STORCK *et al.* (2000), o uso dos princípios básicos da experimentação no planejamento, condução e análise de experimentos, é fundamental para se manter o erro experimental em níveis aceitáveis. Outras formas de minimizar o efeito do erro experimental na qualidade de experimentos é o uso de maior número de repetições, combinado com parcelas menores (GOMES, 1994 e CONAGIN *et al.*, 1995) e a realização o mais uniforme possível dos tratos culturais inerentes à cultura, no decorrer do experimento (LOPES & STORCK, 1995 e LÚCIO, 1997). Estes autores mostram que tratos culturais, como aplicação de inseticidas e adubações desuniformes, são capazes de aumentar o erro experimental. Já o desbaste, distribuição manual de sementes na linha de semeadura e controle de ervas daninhas são capazes de reduzir o erro para experimentos com a cultura do milho.

A avaliação da magnitude do erro experimental dá-se via estatísticas definidas como:

coeficiente de variação (CV) = $\frac{\sqrt{QMe}}{\hat{m}} \times 100$, onde,

QMe = quadrado médio do erro e \hat{m} = estimativa da média geral do experimento; Índice de variação (GOMES, 1991), IV = $\frac{CV}{\sqrt{J}}$, onde CV = coeficiente

de variação e J = número de repetições do experimento; diferença mínima significativa em percentagem da média do experimento (DMS%)

(LÚCIO, 1999), DMS% = $q_{\alpha(1;gle)} \frac{\sqrt{\frac{QMe}{J}}}{\hat{m}} \times 100$,

onde $q_{\alpha(1;gle)}$ = valor de tabela para o teste de Tukey em nível α de significância para I tratamentos e gle graus de liberdade do erro e QMe, J e \hat{m} possuem o mesmo significado já mencionado. A maior precisão será obtida com menores valores destas estatísticas e menores diferenças entre estimativas de médias aparecerão como significativas.

Tabelas foram apresentadas por GOMES (1990), SCAPIM *et al.* (1995) e LÚCIO

(1999), com critérios diferenciados de construção, mas com o mesmo propósito de classificação das estatísticas já descritas, em níveis muito alto, alto, médio, baixo e muito baixo, dando condições ao pesquisador de avaliar a precisão de seus experimentos, para diferentes culturas. A interpretação destas tabelas deverá ser realizada de forma criteriosa, pois as estatísticas utilizadas para suas construções não possuem boas propriedades e, se na escala utilizada existirem valores negativos, a estimativa do erro sofre restrições em seu uso. Outro fato que poderá “mascarar” a precisão de um experimento é a utilização de forma homogênea de um trato cultural capaz de aumentar a média final da variável em questão, fazendo com que a estimativa do erro seja reduzida sem que haja redução, ou maior controle, do erro experimental.

O presente trabalho teve por objetivo desenvolver um modelo de regressão linear múltipla, dando condições ao pesquisador de estimar o erro experimental em função da variabilidade entre os tratamentos e dentro das diferentes características morfológicas (altura, florescimento, etc.) observadas no decorrer do experimento, do ciclo dos cultivares estudadas e do número de tratamentos adotados no experimento, avaliando a precisão experimental obtida.

MATERIAL E MÉTODOS

O material utilizado refere-se aos resultados da Rede de Ensaios Nacionais de Competição de Cultivares de Milho, realizados entre os anos de 1986 e 1997, para materiais de ciclo normal, precoce e superprecoce. Estes resultados foram obtidos junto ao Centro Nacional de Pesquisa de Milho e Sorgo/EMBRAPA.

Em cada experimento, foi avaliado: o ciclo do material, a região de instalação do experimento (centro, sul e região não definida), o número de tratamentos (I), o coeficiente de variação (CV%) da variável rendimento e a média por tratamento para: peso de espigas, em kg/há, peso de grãos, em kg/ha dias até o florescimento, altura da planta, em cm, altura da espiga, em cm, percentagem de plantas acamadas, percentagem de plantas quebradas, estande final e número de espigas.

Cada ciclo dos materiais foi caracterizado pela soma térmica, por esta considerar a temperatura, apresentando potencialmente maior precisão para a determinação do ciclo (RECOMENDAÇÕES TÉCNICAS PARA CULTIVO DO MILHO, 1993). Assim, a classificação adotada para os materiais, conforme o

ciclo, foi: 800UC, para materiais de ciclo superprecoce; 830UC, para materiais de ciclo precoce e 880UC, para materiais de ciclo normal. Estes valores foram assumidos como a média do intervalo da soma térmica que caracteriza cada ciclo.

Os experimentos foram devidamente catalogados, e os resultados de cada um compuseram um arquivo. Destes, escolheram-se os arquivos referentes aos experimentos conduzidos entre os anos de 1986 e 1996 e, a partir deles calcularam médias e variâncias das médias dos tratamentos, para cada variável estudada no experimento, além do quadrado médio do erro (QMe) e diferença mínima significativa em porcentagem (DMS%) para a variável peso de grãos, criando um novo e único arquivo com estes resultados.

Da mesma forma, para os arquivos restantes, referentes aos experimentos conduzidos no ano de 1997, também foi aplicado esse programa, criando outro arquivo individual, para a validação das equações de regressão múltipla a serem ajustadas. Assim, o trabalho dividiu-se em duas partes, inicialmente a estimativa das equações de regressão linear múltipla e, a seguir, a validação destas equações.

O modelo de regressão linear múltipla utilizado é definido por DRAPER & SMITH (1966) e ELIAN (1988) como:

$Y_j = \alpha + \sum \beta_i X_{ij} + \epsilon_j$, onde Y_j = estimativas do QMe e/ou DMS%, α = coeficiente linear, β_i = coeficiente de regressão das variáveis independentes X_i , X_{ij} = variáveis independentes X_i na repetição j e ϵ_j = erro aleatório associado à variável Y na repetição j .

A estimação dos parâmetros das equações de regressão linear múltipla foi baseada no método "stepwise backward" apresentado por DRAPER & SMITH (1966) e ELIAN (1988). Na validação das equações, com os resultados dos experimentos realizados em 1997, foi realizado uma análise de regressão simples entre os valores estimados do QMe e DMS%, obtidos nas equações de regressão múltipla, e os valores observados nos experimentos e, quanto mais próximo da unidade se apresentou o valor do coeficiente de determinação (r^2), os valores observados e estimados se equívalem, ou seja, as duas formas de obtenção do QMe e DMS% são adequadas para estimativa dos valores, em nível de 5% de probabilidade de erro.

Para verificar a aderência à distribuição normal padrão dos valores estimados em cada equação (QMe e DMS%), nos experimentos conduzidos entre 1986 e 1996, foi realizado o teste

de Lilliefors, $\alpha = 5\%$ (CAMPOS, 1983), no qual quando não rejeitado $H_0 (D_{\text{máx}} < d_{\alpha(n)})$, os dados poderão ser estudados através da distribuição normal e uma classificação dos valores dos DMS% e QMe, em limites de classe, poderá ser realizada, com base na adaptação do critério de classificação proposto por LÚCIO (1999).

RESULTADOS E DISCUSSÃO

Foram utilizados 610 experimentos em rede de ensaios, dos quais 522 foram realizados entre os anos de 1986 e 1996 e 88 experimentos realizados no ano de 1997 e para a DMS%, a equação de regressão múltipla foi definida por: $DMS\% = 40,95 + 0,29I - 0,42M_{\text{pesp}}$, com um coeficiente de determinação, $R^2 = 0,7584$. A significância para o número de tratamentos (I) já era esperada, pois a própria equação para o cálculo desta, baseia-se nesta variável, ou seja, a medida que se avalia um número maior de tratamentos, aumenta-se o valor da DMS%. A média do peso de espigas apresentou-se com sinal negativo, mostrando que a medida que se aumenta o peso médio das espigas, há uma redução no valor da DMS%, mostrando que para se obter maior precisão experimental, deve-se planejar experimentos com um número menor de tratamentos e que estes possuam capacidade produtiva elevada.

Para a estatística QMe, a estimativa da equação de regressão foi função somente da média do peso de grãos e da variância do número de espigas: $QMe = 0,39 + 0,06V_{\text{pesp}}$, com $R^2 = 0,8102$. A variância do peso médio de espigas apresentou-se de tal forma que incrementos nesta proporcionam acréscimos no valor do QMe do peso de grãos, ou seja, quanto maiores as diferenças entre os tratamentos, no peso médio de espigas, maior será o QMe do peso de grãos, reduzindo assim a precisão experimental. Assim, verifica-se que, para se obter uma precisão experimental mais adequada, deve-se utilizar tratamentos (híbridos de milho) mais homogêneos entre si em relação à variância do peso de espigas, selecionando os tratamentos de tal forma que apresentem padrões próximos entre si, quando se trata de peso médio de espigas, favorecendo uma redução no valor do QMe do peso de grãos.

Na validação das equações, a estatística DMS%, apresentou o valor do coeficiente de determinação (r^2) entre os valores estimados e observados de 0,0034 e para o QMe, o valor do r^2 foi de 0,7820, mostrando que ao se estimar o valor do QMe, a equação de regressão ajustada pode ser utilizada, pois o poder de explicação foi próximo a

80%. Assim, os valores observados e estimados do QMe representam a mesma estatística, dando, juntamente com as tabelas de classificação da precisão uma indicação direta da magnitude do erro experimental inerente ao experimento. Já para a DMS%, os valores estimados e observados não se equivalem, pois o valor do r^2 foi muito baixo, mostrando uma total falta de relação entre eles.

A pressuposição a ser atendida na construção das tabelas de classificação da precisão, é a da aderência dos valores observados e estimados, à distribuição normal padrão, testada pelo teste de Lilliefors (CAMPOS, 1983). Como a equação de regressão para a DMS% não foi eficiente em estimar o verdadeiro valor da estatística, aplicou-se o teste apenas para a estatística original e para os valores observados e estimados do QMe, obtidos para os 522 experimentos realizados entre 1986 e 1996.

Para a DMS%, rejeitou-se H_0 , ou seja, o valor do $D_{m\acute{a}x} = 0,9724$ foi superior ao valor do $d_{5\% (522)} = 0,0388$, mostrando que estes dados não aderiram à distribuição normal, concordando com os resultados obtidos por LÚCIO (1997) em experimentos com a cultura do milho no estado do Rio Grande do Sul. Assim, não foi possível a aplicação do critério de classificação da precisão, através da estatística DMS%. Com os valores do QMe o $D_{m\acute{a}x}$ calculado foi de 0,0265 para o QMe estimado pela equação de regressão e de 0,0314 para o QMe observado nos experimentos, mostrando assim que os valores estimados e observados podem ser estudados através da distribuição normal padrão (não se rejeitou H_0), sendo possível a aplicação do critério de classificação já mencionado.

Com média e variância, respectivamente, de 0,5718t/ha e 0,5124, do QMe observado, e de 0,4902t/ha e 0,1444, do QMe estimado através da equação de regressão nos 522 experimentos, a classificação da precisão é a da tabela 1. Nesta tabela, a interpretação dos limites de classe é realizada no seguinte sentido. Limites muito baixo indicam excelente precisão, e muito alto, péssima precisão experimental. Assim, para o QMe_{obs} , não haverá condições de classificar um valor em muito baixo, pois o quadrado médio do erro não admite valores negativos, sendo desta forma a precisão classificada em, no máximo, boa (limite baixo do QMe).

Experimentos que apresentarem valores do QMe, observado no experimento ou estimado através da equação de regressão múltipla, classificados como muito alto, deverão ser descartados, pois apresentam uma péssima precisão,

Tabela 1 - Limites de classe para o quadrado médio do erro observado no experimento (QMe_{obs}) e estimado pela equação de regressão linear múltipla (QMe_{est}), em experimentos da Rede de Ensaios Nacionais de Competição de Cultivares de Milho. Santa Maria – RS, 2000.

	Limites de Classe				
	Muito baixo	Baixo	Médio	Alto	Muito alto
QMe_{obs}	$\leq 0,00$	0,00 – 0,23	0,23 – 0,86	0,86 – 1,41	$> 1,41$
QMe_{est}	$\leq 0,25$	0,25 – 0,39	0,39 – 0,59	0,59 – 0,73	$> 0,73$

não sofrendo assim nenhuma perda significativa de dados ao se realizar tal descarte. Aqueles experimentos nos quais os valores foram classificados como altos, deverão ser criteriosamente analisados, para que haja a identificação dos prováveis fatores que estão funcionando como fontes de erro.

CONCLUSÕES

O quadrado médio do erro, para peso de grãos, pode ser estimado por uma equação de regressão linear múltipla, tendo como variável independente a variância do peso de espiga, enquanto a diferença mínima significativa somente é estimada através de sua fórmula original.

O quadrado médio do erro, tanto o observado na análise da variância dos experimentos, quanto o estimado pela equação de regressão linear múltipla, aderiram à distribuição normal padrão, enquanto a diferença mínima significativa não aderiu.

REFERÊNCIAS BIBLIOGRÁFICAS

- BANZATTO, D.A., KRONKA, S.N. **Experimentação agrícola**. 3.ed. Jaboticabal : FUNEP, 1995. 247p.
- CAMPOS, H. de. **Estatística experimental não-paramétrica**. 4.ed. Piracicaba : FEALQ, 1983. 349p.
- CONAGIN, A., IGUE, T., NAGAI, V. **Tabelas para determinação do número de repetições no planejamento de experimentos**. Campinas : Instituto Agrônomo, Campinas. 1995. 17p. (Boletim Científico, 34).
- DRAPER, N.R., SMITH, R. **Applied regression analysis**. Nova York : John Wiley, 1966. 407p.
- ELIAN, S.N. **Análise de regressão**. São Paulo : IME/USP, 1988. 232p.
- EMPRESA Brasileira de Pesquisa Agropecuária. **Recomendações técnicas para o cultivo do milho**. Brasília : EMBRAPA - SPI, 1993. 204p.
- GOMES, F.P. A importância do número de repetições nos experimentos. **Revista de Agricultura**, Piracicaba, v.69, n.3, p.243-245, dez. 1994.

- GOMES, F.P. **Curso de estatística experimental**. 13.ed. Piracicaba : Nobel, 1990. 468p.
- GOMES, F.P. **O índice de variação, um substituto vantajoso do coeficiente de variação**. Piracicaba : Instituto de Pesquisas e Estudos Florestais. 1991. 4p. (Circular Técnica, 178).
- LOPES, S.J., STORCK, L. A precisão experimental para diferentes manejos na cultura do milho. **Ciência Rural**, Santa Maria, v.25, n.1, p.49-53, 1995.
- LÚCIO, A.D., STORCK, L., BANZATTO, D.A. Classificação dos experimentos de competição de cultivares quanto a sua precisão. **Pesquisa Agropecuária Gaúcha**, Porto Alegre, v.5, n.1, p.99-103, 1999.
- LÚCIO, A.D. **Parâmetros da precisão experimental das principais culturas anuais do Estado do Rio Grande do Sul**. Santa Maria-RS, 1997. 62p. Dissertação (Mestrado em Agronomia) – Curso de Pós-graduação em Agronomia. Centro de Ciências Rurais, Universidade Federal de Santa Maria.
- SCAPIM, C.A., CARVALHO, C.G.P., CRUZ, C.D. Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.5, p.683-686, 1995.
- STEEL, R.G.D., TORRIE, J.H. **Principles and procedures of statistics**. Nova York : McGraw-Hill, 1960. 481p.
- STORCK, L., GARCIA, D.C., LOPES, S.J., *et al.* **Experimentação vegetal**. Santa Maria : UFSM, 2000. 198p.

Ciência Rural, v. 31, n. 6, 2001.