

Diagnóstico exploratório em análise de regressão para dados de nematologia

GODOI, L.S.O.¹; OLIVEIRA, M.C.N. de²

¹Universidade Pitágoras Unopar; bolsista da FAPED, Londrina, PR; ²Pesquisadora, Embrapa Soja

Introdução

A soja [*Glycine max* (L.) Merr.] é uma cultura de grande importância econômica no Brasil e a principal do agronegócio do país, segundo maior produtor mundial com a área plantada correspondente a 35.802 mil hectares e a produção estimada no ano agrícola 2018/2019 de 114,3 milhões de toneladas (Conab, 2019). O primeiro lugar na produção mundial de soja atualmente é ocupado pelos Estados Unidos e outros países de produção expressiva são a Argentina, a China e a Índia. O Brasil vem apresentando crescentes aumentos na produção de soja, não apenas pelo incremento na área plantada, mas também pelos ganhos em produtividade com perspectivas de atingir a liderança na produção dessa commodity. Para tal o país deve investir em esforços na solução dos problemas que limitam o potencial da cultura como os relacionados aos aspectos fitossanitários.

Dentre os patógenos que podem reduzir o potencial produtivo da cultura, destaca-se o nematoide do cisto da soja (NCS), *Heterodera glycines* Ichinohe (Silva et al., 2006). O gênero *Heterodera*, além de apresentar grande variabilidade genética, caracteriza-se pela formação de cistos, que são estruturas constituídas pelo corpo da fêmea adulta morta que contém os ovos. Essas estruturas são de cor marrom e altamente resistentes às condições adversas do ambiente. Sua ocorrência, cujo principal sintoma visual é a clorose das folhas das plantas afetadas, se dá em reboleiras, o que dificulta a obtenção de amostra representativa. O método tradicional utilizado na amostragem no campo em áreas extensas, como é o caso da soja, requer coleta de grande número de amostras de solo que não refletem com precisão a distribuição espacial do nematoide. O conhecimento da distribuição espacial desse patógeno, assim como a forma de seu desenvolvimento na lavoura é fundamental para a elaboração de estratégias de manejo e controle.

Os cuidados para um diagnóstico seguro a fim de evitar a infestação do nematoide devem ocorrer antes da sementeira da soja e o agricultor deve ficar em alerta para verificar se realmente as reboleiras que apresentam a clorose das plantas se deve à infestação de nematoides e não a outros problemas como deficiência de algum macro e/ou micronutriente, compactação dos solos, excesso ou falta de calagem ou interação com outros organismos patogênicos (Silva et al., 2006). Tais cuidados devem ser estendidos também ao tratamento estatístico dos dados, levando em conta o problema de obtenção de amostras representativas.

Em diversos trabalhos que abordam o problema dos nematoides, como Dias et al. (1998), Farias (1999), Pinheiro et al. (2008), Santana et al. (2009), Debiasi et al. (2016) e Zhou et al. (2017), a avaliação estatística se restringiu ao uso de valores médios das variáveis respostas e erro padrão ou análise de regressão e correlação, ou estatísticas gráficas como o boxplot e análise geoestatística o que pode levar a equívocos na interpretação dos resultados.

A proposta deste trabalho foi avaliar o coeficiente de correlação de Pearson, a análise de regressão linear com diagnóstico de regressão linear (DRL) para as variáveis peso de massa fresca (PMF), número de cistos viáveis (NCV), número de vagens por planta (NVP) e Índice de Vegetação da Diferença Normalizada (NDVI) em 21 pontos amostrais (Draper; Smith, 1981). Na análise de diagnóstico foi possível detectar a existência de pontos influentes que interferiram nas estimativas dos coeficientes do modelo de regressão, pois esses pontos induzem a resultados equivocados.

Material e Métodos

Os dados utilizados nesse estudo foram obtidos de experimento conduzido no município de Florínea (lat. 22° 51'S, long. 50°, 39'W), no Vale do Paranapanema, tradicional região de cultivo de soja do Estado de São Paulo (Santos Junior, 2002). O tipo de solo na área é latossolo vermelho escuro e o solo antes do preparo para plantio apresentou pH 5,5. Na adubação de pré-plantio, foram utilizados 285 kg/ha da formulação 0-20-20 e a lavoura foi plantada no dia 15 de outubro de 2000 com a cultivar BRS133. A precipitação pluviométrica registrada do plantio até a tomada das leituras foi de 632 mm, considerada normal em relação às médias anuais. As 21 áreas infestadas

por *H. glycines* foram avaliadas pelo aparelho portátil espectrorradiômetro, selecionadas por meio de avaliação visual, considerando níveis diferentes de sintomas de ataque do nematoide refletidos na lavoura. Os sintomas variaram de severo até ausência. Cada área amostral foi georreferenciada utilizando-se receptor de GPS (Global Positioning System) de navegação, modelo Garmin GPS II. Em cada parcela onde foram tomadas leituras, coletou-se uma amostra composta de solo, a partir de três subamostras, retiradas numa profundidade de até 30 cm. Essas amostras foram encaminhadas ao Laboratório de Nematologia do Departamento de Fitossanidade da UNESP/FCAV, Campus de Jaboticabal, SP, onde foram processadas segundo Jones (1945). Para cada amostra da área infestada por *H. glycines*, foi determinado o número de cistos extraídos por 50 cm³ de solo.

Foram coletadas também cinco plantas de soja em cada área amostral, para determinação do peso da matéria fresca (PMF) e contagem do número de vagens por planta (NVP). Esses dados foram utilizados de forma complementar, para serem relacionados aos valores de NDVI e à infestação de nematoides em função do número de cistos viáveis (NCV). Foram avaliados os pressupostos de normalidade e aleatoriedade dos resíduos (Shapiro; Wilk, 1965; Parente, 1984), bem como os coeficientes de assimetria e curtose. As análises estatísticas utilizadas para avaliar a presença de pontos influentes foram: distância de Cook, taxa de covariância, DFBetas, DFFits e resíduos estudantizados.

Considerando que duas variáveis quantitativas X e Y, observadas em cada indivíduo sob estudo e, no caso de existir uma relação entre Y e X, ela poderá ser escrita por uma função matemática para qualquer uma das variáveis em estudo. Para investigar essa relação podem ser usados gráficos de dispersão e ou correlação que podem indicar uma relação linear de dependência entre as duas variáveis. Essa relação pode ser expressa por:

$$Y = \beta_0 + \beta_1 X \quad (1)$$

Se as mensurações realizadas para as variáveis não caem sobre uma linha reta, ou seja, os valores observados são diferentes, essas discrepâncias, os quais são identificados como erros estatísticos, assim o modelo de regressão linear, matematicamente é indicado como:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2)$$

em que:

Y indica a variável resposta,

X é a variável regressora e β_j , $j = 0, 1$ são chamados de coeficientes de regressão e ε indica o erro aleatório com distribuição Normal com média zero e variância constante .

Quanto mais distantes estiverem os pontos observados da reta, maiores serão os erros, aumentando assim a falta de ajuste dos dados ao modelo de regressão linear. Para estimar os parâmetros do modelo de regressão linear foi utilizado o método de mínimos quadrados. Esse método é denominado de mínimos quadrados porque o seu objetivo é encontrar a reta Y que torne mínima a soma dos quadrados dos erros. Não será apresentada a demonstração deste método. Para uma dada equação $y = x$, para minimizar a soma dos quadrados dos erros fica como segue:

$$SQE = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (3)$$

Diferenciando a equação 3 em relação aos parâmetros β_0 e β_1 e igualando a expressão (3) a zero obtém-se as estimativas de $\hat{\beta}_0$ e $\hat{\beta}_1$ por mínimos quadrados, mostrados nas equações:

$$a = \frac{(\sum XY)(\sum Y^2) - (\sum Y)(\sum XY)}{N \sum Y^2 - (\sum Y)^2} \quad (4)$$

$$b = \frac{N \sum XY - (\sum X)(\sum Y)}{N \sum Y^2 - (\sum Y)^2} \quad (5)$$

O coeficiente de correlação de Pearson (r) é muito utilizado quando se quer avaliar o grau de associação linear entre duas variáveis quantitativas X e Y. Embora não implique em causalidade, o coeficiente de correlação exprime em números a relação entre as variáveis, assumindo valores entre -1 e 1; sua interpretação está associado com o sinal deste coeficiente. Valores negativos do coeficiente r indicam que o crescimento linear de uma variável está associado ao decréscimo linear da outra. Por outro lado, valores positivos de r informam um crescimento linear diretamente proporcional de ambas va-

riáveis. A ausência de associação linear é indicada por valores de r próximos à zero. O cálculo de r é dado pela Equação 6:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (6)$$

O coeficiente que é utilizado para avaliar a qualidade do ajustamento de uma linha de regressão linear, ou seja, o qual mede a proporção da variação do total dos dados que é explicada é o coeficiente de determinação R^2 . Este coeficiente é sempre positivo e seus valores encontram-se no intervalo entre 0 e 1. Estas estatísticas são expressas pelas equações (7) e (8).

$$R^2 = \frac{SQRegressão}{SQTotal} \quad (7)$$

$$R_{Aj}^2 = 1 - \frac{n-1}{n-p} (1 - R^2) \quad (8)$$

em que:

n é o tamanho da amostra,

p é o número de parâmetros (incluindo o intercepto) e,

i é igual a 1, se o modelo inclui o intercepto ou 0, se o modelo não incluir 0 (zero)

O coeficiente de determinação indica a percentagem da variação total de Y explicada pela variável X . Esta estatística deve ser usada com cautela, pois um valor alto denão é um indicativo de que seja um bom preditor linear. Além desta estatística pode-se também utilizar o coeficiente de determinação ajustado (R_{Aj}^2), que considera o número de parâmetros no modelo, assim como o tamanho da amostra, fornecendo uma medida mais adequada para comparar modelos com diferentes quantidades de parâmetros. Em geral o diagnóstico de valores influentes na análise de regressão linear é utilizado para que desvios entre os valores observados e ajustados do modelo linear que permitam avaliar e verificar o seu grau de influência sobre a análise. Várias são as estatísticas utilizadas para medir a influência na análise de regressão. A influência nas observações tem um forte impacto sobre as estimativas dos parâmetros e em consequência pode produzir mudança na inclinação da reta, na linha de regressão (afetando o intercepto) e influência na inclinação apenas por um ponto. Dependendo da localização de um ponto a linha de regressão muda ligeiramente em direção a essas observações. As estatísticas distância de

Cook, DFBETAS, DFFITS e a Taxa de Covariância são apresentadas nas equações a seguir:

$$\text{Distância de Cook: } D_i = \frac{(\mathbf{b}_{(i)} - \mathbf{b})' (\mathbf{X}' \mathbf{X}) (\mathbf{b}_{(i)} - \mathbf{b})}{p \text{MSE}} = r_i^2 \left(\frac{h_{ii}}{1 - h_{ii}} \right) \quad (9)$$

em que:

\mathbf{b}_i é o vetor das estimativas dos parâmetros sem incluir i-ésima observação,

\mathbf{b} é o vetor das estimativas dos parâmetros a partir da regressão que inclui todas as observações,

MSE é o quadrado médio do erro a partir da regressão usando todas as observações,

r_i é o i-ésimo resíduo estudentizado e,

h_{ii} é o i-ésimo valor de alavancagem.

$$\text{DFBETAS}_{j(i)} = \frac{b_j - b_{j(i)}}{\hat{\sigma}(b_j)} \quad (10)$$

em que:

b_j é a estimativa do parâmetro para a j-ésima variável independente;

$b_{j(i)}$ é a estimativa do parâmetro para j-ésima variável independente com a i-ésima observação deletada a partir da análise;

$\hat{\sigma}(b_j)$ é o erro padrão da j-ésima estimativa do parâmetro.

$$\text{DFFITS}_i = \frac{\hat{Y}_i - \hat{Y}_{(i)}}{S(\hat{Y}_i)} \quad (11)$$

em que:

\hat{Y} é o i-ésimo valor predito;

$\hat{Y}_{(i)}$ é o i-ésimo valor predito quando a i-ésima observação é deletada a partir da análise;

$S(\hat{Y}_i)$ é o erro padrão do i-valor predito.

$$\text{COVRATIO}_i = \frac{|s_i^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}|}{|s^2 (\mathbf{X}' \mathbf{X})^{-1}|} \quad (12)$$

As taxas de covariância têm algumas indicações: 1) taxa maior que 1 indica que a i -ésima observação aumenta a precisão das estimativas, 2) taxa menor que 1 indica que a i -ésima observação diminui a precisão das estimativas e, 3) taxa próxima a 1 indica que a i -ésima observação tem pouco efeito sobre a precisão das estimativas. Todas as estatísticas anteriores que medem a influência sobre o modelo de regressão linear, bem como as análises de regressão foram obtidas utilizando o software estatístico Statistical Analysis System – SAS versão 9.4 (SAS Institute, 2012).

Resultados e Discussão

As plantas coletadas em Florínea nos 21 pontos amostrais para avaliar a infestação de nematoides, apresentaram alta variabilidade (Tabela 1). Os coeficientes de assimetria (G1) e curtose (G2) foram todos maiores do que zero antes e após a retirada de pontos influentes. Os coeficientes de variação (CV) obtidos estão no intervalo de 19,57 a 150,17 (%). Somente para o número de cistos viáveis com 21 observações não apresentou distribuição normal. O NCV foi o que apresentou maiores coeficientes de assimetria e curtose para $N=21$ (Tabela 1). Baseados nos resultados das estatísticas descritivas com $N=13$, 16 e 21 e pelas estatísticas: distância de Cook, resíduos estudentizados, taxa de covariância, DFFITS e DFBETAS, foi dado continuidade na análise de diagnóstico, com o objetivo de detectar pontos que interferem nas estimativas dos coeficientes de regressão - intercepto e coeficiente angular (Tabelas 1 e 2, Figura 1).

Tabela 1. Estatísticas descritivas para as variáveis PMF, NCV, NVP e NDVI com N=13, N=16 e N= 21

Variáveis	Amostra	G1	G2	CV (%)	W	Pr(W)	Variância
PMF	13	0,29	-1,39	36,99	0,92	0,31	1316,17
	16	0,58	-0,38	42,9	0,94	0,45	1063,22
	21	0,34	-1,03	50,5**	0,93	0,14	1848,06
NCV	13	0,08	-1,49	66,3	0,92	0,27	11,0
	16	0,46	-1,09	74,5**	0,89	0,051	7,8
	21	3,40*	12,98*	150,17**	0,57	0,0001	147,79
NVP	13	-0,005	-1,28	38,45	0,95	0,57	400,91
	16	0,33	-1,34*	43,56	0,92	0,17	274,86
	21	0,11	-1,37*	45,80**	0,93	0,17	453,26
NDVI	13	-0,46	-0,93	19,57	0,92	0,27	0,019
	16	-0,61	0,08	32,90	0,95	0,53	0,040
	21	-0,73	-0,11	32,09	0,92	0,110	0,043

*Estatísticas G1 e G2 com valores altos - **Coeficientes de Variação (%) muito alto – G1= coeficiente de assimetria – G2= coeficiente de curtose – CV= coeficiente de variação – W= teste de normalidade de Shapiro & Wilk – Pr(W)= valor da Probabilidade da estatística W – PMF = peso de massa fresca em (g/ha) – NCV=número de cistos viáveis – NVP= número de vagens por planta – NDVI= Índice vegetativo com diferença normalizada

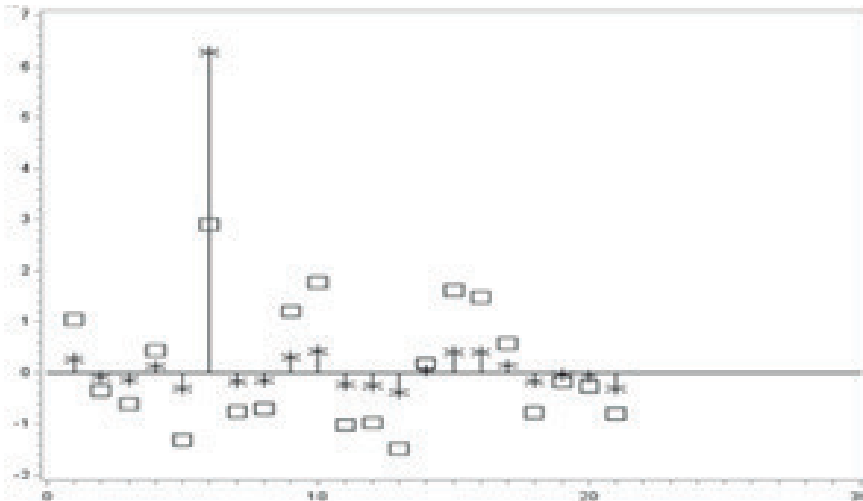


Figura 1. Estatísticas DFFITS e resíduos estudentizados usando a Regressão entre (PMF) em g/ha e (NCV).

Tabela 2. Estatísticas Distância de Cook, resíduos estudentizados, taxa de covariância, DFFITS e DFBETAS, que avaliam as influências no modelo de regressão entre PMF X Ncistos. N=21.

DCook's	Rstudent	Cov Ratio	DFFITS	DFBETAS	
				intercepto	Ncistos
0,0076	0,44117	1,1723	0,12084	0,1208	-0,0681
0,0778	1,48504	0,95059	0,40677	0,4068	-0,2293
0,0375	1,04341	1,05921	0,27432	0,2738	-0,1408
0,0308	-0,97933	1,06864	-0,24784	-0,2459	0,1133
0,0666	-1,48756	0,94018	-0,37645	-0,3736	0,172
0,0459	1,21154	1,01352	0,3066	0,3042	-0,1401
0,0104	0,57944	1,13795	0,14166	0,139	-0,0559
0,0837	1,76412	0,85833	0,4313	0,4233	-0,1702
0,0731	1,63066	0,8964	0,39867	0,3913	-0,1573
0,0023	-0,27297	1,17108	-0,06674	-0,0655	0,0263
0,001	0,18039	1,17268	0,0428	0,0413	-0,014
0,0009	-0,17884	1,16758	-0,04064	-0,037	0,0071
0,0128	-0,70215	1,10883	-0,15771	-0,1379	0,0145
0,0095	-0,60668	1,1235	-0,13566	-0,1127	0,0011
0,0142	-0,74469	1,10102	-0,16652	-0,1383	0,0013
0,0255	-1,01086	1,04759	-0,22604	-0,1877	0,0018
0,0031	-0,34495	1,15493	-0,07737	-0,0604	-0,0059
0,0153	-0,76406	1,09898	-0,17315	-0,1257	-0,0274
0,0461	-1,31081	0,98021	-0,30941	-0,1883	-0,0967
0,0453	-0,79678	1,18507	-0,29809	-0,0219	-0,2332
14,1862	2,90328	2,9367	6,2822	-2,1895	6,098

Foram obtidos pelo DRL as estatísticas DFFITS, DFBETAS, Taxa de covariância (Covratio), Distância de Cook (Dcook's) e Resíduos estudentizados (Rstudentized), tomando por base o modelo de regressão linear entre (PMF) e (NCV). Foi observado que existem pontos influentes nas variáveis em estudo. Na Figura 1 e na Tabela 2 a estatística DFFITS indicou um grande valor (6,28220) e está denotado por um * e um quadrado indicando os resíduos estudentizados (2,90328) obtidos pelo SAS. Estes mesmos resultados e outros podem ser confirmados na Tabela 2 para todas as estatísticas.

Os coeficientes de variação, a correlação e a regressão para todas as variáveis com 16 e 21 observações estão na Tabela 3. Os maiores valores de CV estão nas análises com N=21 indicando que existem valores anômalos e podem ser confirmados com as estatísticas assinaladas em vermelho na Tabela 2 e Figura 1. Em face desses resultados foi continuada a investigação para não ocorrer interpretações equivocadas para os resultados de pesquisa sobre os nematoides. Observar que as estatísticas DCooks, resíduos estudentizados, DFFITS e DFBETAS estão todos alterados com a presença de um valor influente, considerando sempre que os valores residuais foram além do intervalo de (-2, +2) ou nas proximidades deste intervalo (Tabela 2).

Tabela 3. Coeficientes de Variação, de correlação de Pearson e coeficientes de determinação entre as variáveis. N=21.

	CV ₁₆	CV ₂₁	r ₁₆	r ₂₁	R ²	R ²
NCV x NVP	52,70	125,66	-0,71	-0,58	0,49	0,33
NCV x PMF	32,12	42,40	-0,69	-0,58	0,79	0,34
NCV x NDVI	55,24	104,47	-0,67	-0,73	0,45	0,54
NVP x PMF	16,49	21,76	0,93	0,91	0,86	0,83
PMF x NDVI	19,79	23,83	0,90	0,89	0,80	0,79
NDVI x NVP	17,79	18,82	0,92	0,92	0,84	0,84

CV (%) = Coeficiente de variação, r = coeficiente de correlação,

Pela análise de regressão da Tabela 4, o valor do R² foi muito baixo devido aos valores influentes observados na Figura 2, os quais estão concentrados entre 0 e 10 para o NCV e outros entre 20 e 60. Na Tabela 2 percebe-se a necessidade de novas análises sem estes valores. Uma vez detectado a variabilidade nos dados e vários pontos influentes distante da reta de regressão entre PMF e NCV com N=21 foram realizados os mesmos procedimentos anteriores para N=16.

Tabela 4. Análise de Regressão Linear entre Peso de massa fresca (PMF) e Número de cistos viáveis (NCV) N=21.

Causas de variação	Grau de liberdade	Somas de quadrados	Quadrados médios	Valor de F	Pr(F)
NCV	1	13807,43	13807,43	9,81	0,0055
RESÍDUO	19	26750,60	1407,93		
CORREÇÃO TOTAL	20	40558,03			

CV(%)=42,40 Média=88,44

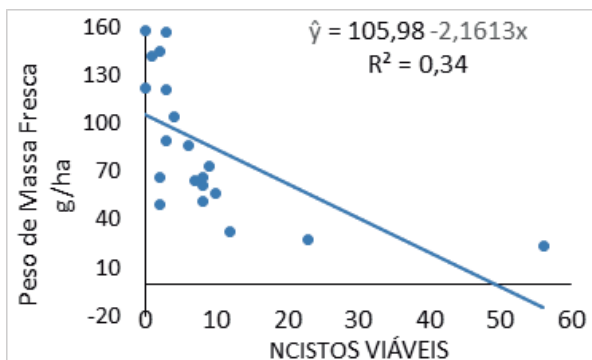


Figura 2. Regressão entre o peso de matéria fresca (PMF) em g/ha e número de cistos viáveis (NCV). N=21.

Os dados médios foram insuficientes para este estudo de caso, na aplicação de diagnóstico exploratório em regressão linear com dados de nematoides, mas foi essencial para elucidar que em análise de regressão ocorrem *outliers* e pontos influentes e deve-se ter cautela usando estatísticas apropriadas para avaliá-los. O ideal é que nesse levantamento o tamanho da amostra fosse N=100. O objetivo original do trabalho (Santos Junior et al, 2002) foi explorar o uso de espectrorradiometria no visível e infravermelho próximo para avaliar o nível populacional do nematoide de cistos *Heterodera glycines* 112 dias após o plantio. Portanto, esse estudo foi dado continuidade até que reduzisse o coeficiente de variação, baixasse o número de *outliers*, resíduos estudentizados e DFFITS, permitindo o estudo coerente para uma análise de regressão linear (Figuras 3 e 4, Tabelas 5 e 6).

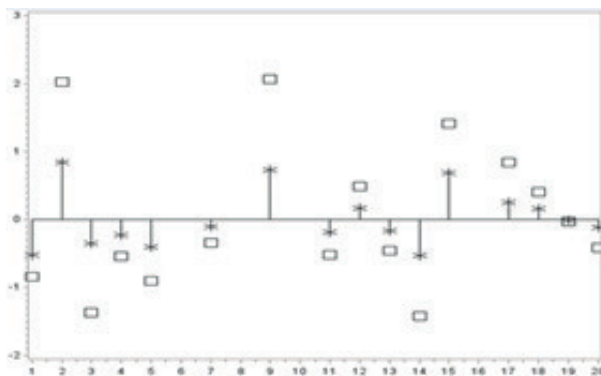


Figura 3. Estatísticas DFFITS e resíduos estudentizados usando o modelo de Regressão entre (PMF) em g/ha e (NCV). N=16.

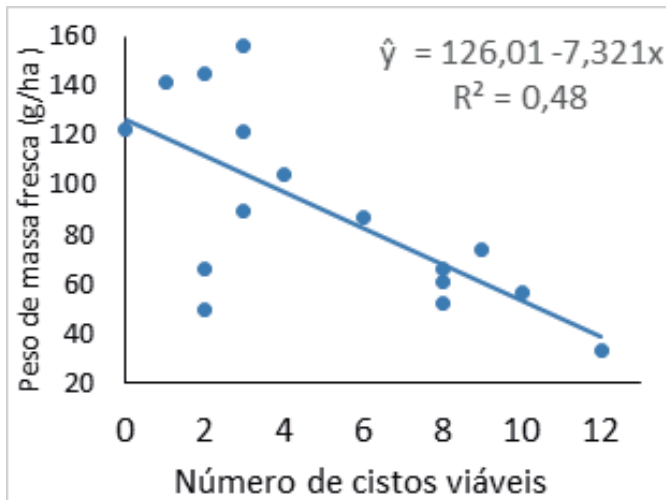


Figura 4. Regressão entre o (PMF) em g/ha e (NCV), N=16.

Tabela 5. Estatísticas distância de Cook, resíduos estudentizados, taxa de covariância, DFFITS e DFBETAS, que avaliam as influências no modelo de regressão entre PMF X Ncistos. N= 16.

DCook's	Rstudent	Cov ratio	DFFITS	----- DFBETAS-----	
				Intercepto	Ncistos
0.002	-0,1392	1,4299	-0,0677	-0,0677	0,0555
0.064	0.8587	1.2154	0.3543	0.3523	-0.2676
0.174	-1.8141	0.8295	-0.6366	-0.6173	0.4175
0.327	-2.8188	0.5018	-0.9892	-0.9591	0.6486
0.097	1.2827	1.0264	0.4501	0.4364	-0.2952
0.018	0.6053	1.1975	0.1832	0.1663	-0.0925
0.168	2.1508	0.6887	0.6508	0.5907	-0.3286
0.013	-0.5248	1.2140	-0.1588	-0.1441	0.0802
0.003	0.2581	1.2320	0.0698	0.0547	-0.0201
0.001	0.1472	1.2388	0.0394	0.0135	0.0101
0.000	-0.0478	1.2966	-0.0165	0.0014	-0.0105
0.003	-0.2265	1.2869	-0.0779	0.0067	-0.0499
0.020	-0.5680	1.2350	-0.1955	0.0168	-0.1251
0.021	0.4946	1.2999	0.1999	-0.0460	0.1489
0.002	0.1233	1.4198	0.0587	-0.0197	0.0478
0.010	-0.2064	1.6565	-0.1365	0.0645	-0.1217

Tabela 6. Análise de Regressão Linear entre Peso de massa fresca (PMF) e Número de cistos viáveis (NCV). N = 16.

Causas de Variação	Grau de liberdade	Somas de Quadrados	Quadrados Médios	Valor de F	Valor de p
NCV	1	10662,55011	10662,55011	13,04	0,0028
RESÍDUO	14	11443,13939	817,36710		
CORREÇÃO TOTAL	15	22105,68950			
CV(%)=32,14	R ² =0,48	Média=88,95			

Na análise de regressão com N=13 os valores observados estão mais próximos da reta estimada. Inicialmente com N=21 a variância residual foi de 1407,93 (Tabela 4) e com N=13 o quadrado médio do resíduo foi 354,04 (Tabela 7). Os mesmos procedimentos estatísticos foram aplicados para o conjunto de dados com N=13 e com esta amostra foram obtidos resultados com baixa variabilidade reduzindo os outliers e o valor do coeficiente de determinação foi de $R^2 = 0,75$ e $CV(\%) = 19,19$ (Tabelas 7 e 8, Figuras 5 e 6). Na Figura 5 as estatísticas DFFITS foram bem menores apresentando somente um valor de resíduo estudentizado (2,8455) alto. No entanto, vale ressaltar que este estudo focou apenas em ajustes de regressão linear, o que não necessariamente se aplica à todas as relações entre duas variáveis.

Tabela 7. Análise de Regressão Linear entre Peso de massa fresca (PMF) e Número de cistos viáveis (NCV). N=13.

Causas de variação	Graus de liberdade	Somas de Quadrados	Quadrado Médio	Valor de F	Pr(F)
NCV	1	11899,70371	11899,70371	33,61	0,0001
RESÍDUO	11	3894,44398	354,04036		
CORREÇÃO TOTAL	12	15794,14769			
CV(%)=19,19	R ² = 0,75	Média =98,06			

Tabela 8. Estatísticas Distância de Cook, resíduos estudentizados, Taxa de covariância, DFFITS e DFBETAS, que avaliam as influências no modelo de regressão entre PMF X Ncistos com N=13.

DCook's	Rstudent	Cov ratio	DFFITS	-----DFBETAS-----	
				Intercepto	NCistos
0,377	-1,5269	1.0845	-0.9199	-0.9199	0.7758
0,013	0,3157	1.4793	0.1569	0.1561	-0.1228
0,097	1,0799	1.1352	0.4449	0.4313	-0.3050
0,003	0,2118	1.3432	0.0734	0,0663	-0,0390
0,296	2,8455	0.4138	0.9861	0.8910	-0.5242
0,46	-1.687	0.8212	-0.5848	-0.5284	0.3109
0,002	-0,1838	1.3128	-0.0558	-0.0427	0.0167
0,001	-0,1205	1.3178	-0.0366	-0.0095	-0.0110
0,003	-0,1914	1.4051	-0.0788	0.0147	-0.0540
0,020	-0,4701	1.3548	-0.1937	0.0362	-0.1328
0,090	-1,0310	1.1565	-0.4247	0.0793	-0.2912
0,079	0,7879	1.3378	0.3916	-0.1272	0.3063
0,021	0,3274	1,6144	0,1972	-0.0833	0.1663

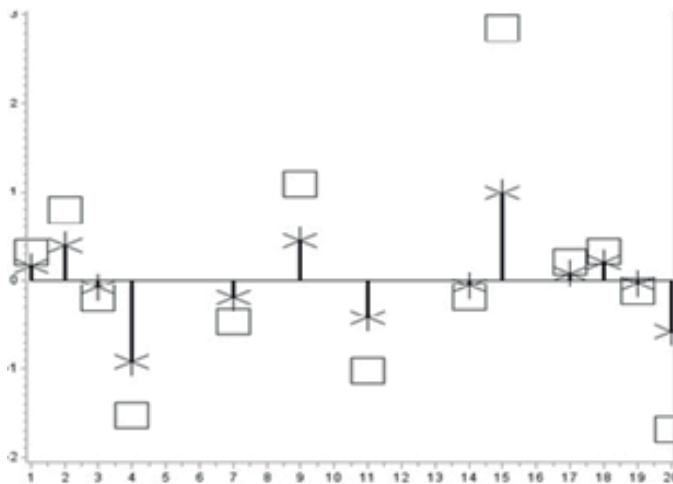


Figura 5. Estatísticas DFFITS e resíduos estudentizados para o modelo de Regressão entre PMF em g/ha e NCV. N=13.

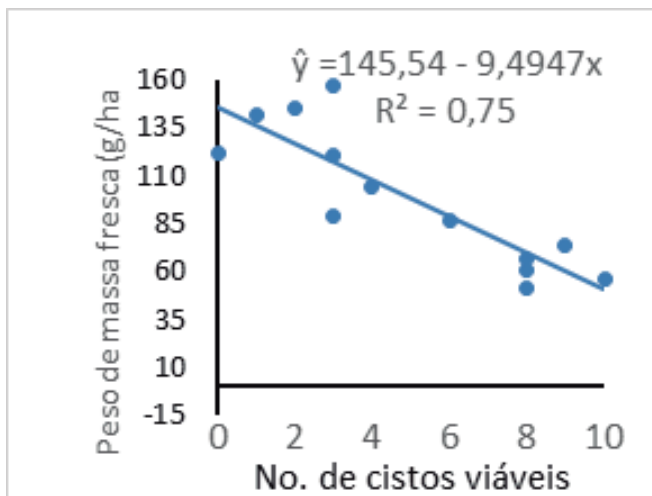


Figura 6. Regressão entre o peso de matéria fresca (PMF) em g/ha e (NCV). N=13

Conclusão

O método de diagnóstico para dados influentes foi eficiente para identificar pontos que interferem na análise de regressão linear com garantia de se obter um bom preditor linear. As estatísticas dos resíduos estudentizados e DFFITS são eficientes para detectar pontos influentes. Os pontos influentes interferem nos resultados de pesquisa.

Referências

CONAB. **Acompanhamento de safra brasileira:** grãos, oitavo levantamento, maio 2019, safra 2018/19. 69 p. Disponível em: <<https://www.conab.gov.br/info-agro/safra/safra/graos/boletim-da-safra-de-graos>>. Acesso em: 10 mai. 2019.

DEBIASI, H.; FRANCHINI, J. C.; DIAS, W. P.; RAMOS JUNIOR, E. U.; BALBINOT JUNIOR, A. A. Práticas culturais na entressafra da soja para o controle de *Pratylenchus brachyurus*. **Pesquisa Agropecuária Brasileira**, Brasília, v.51, n.10, p.1720-1728, 2016.

DIAS, W. P.; SILVA, J. F. V.; KIIHL, R. A. de S.; HIROMOTO, D. M.; ABDELNOOR, R. V. Quebra de resistência da CV. Hartwig por população de campo do nematoide de cisto da soja (*Heterodera glycines*). **Pesquisa Agropecuária Brasileira**, Brasília, v.33, n.6, p.971-974, jun.1998.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd. ed. New York: John Wiley & Sons, 1998. 706 p.

- FARIAS, P. R. S. **Distribuição espacial do nematoide reniforme com o uso de geoestatística, em um solo com rotação de culturas**. 1999. 109 f. Tese (Doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal.
- JONES, F. G. W. Soil populations of beet eelworm (*Heterodera schachtii* Schm.) in relation to cropping. **Annals of Applied Biology**, v. 32, p. 351-380, 1945.
- PARENTE, R. C. P. **Aspectos da análise de resíduos**. 1984. 118 f. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiróz", Piracicaba.
- PINHEIRO, J. B.; POZZA, E. A.; POZZA, A. A. A.; MOREIRA, A. S.; ALVES, M. C.; CAMPOS, V.P. Influência da nutrição na distribuição espacial do nematoide de cisto da soja. **Nematologia Brasileira**, v. 32, n. 4, p. 270-278, 2008.
- SANTANA, H.; PIRES, E.; COMERLATO, A. P.; NASU, E. G. C.; FURLANETTO, C. Variabilidade genética em populações de campo do nematoide de cisto da soja provenientes dos estados do Paraná e Rio Grande do Sul. **Tropical Plant Pathology**, v. 34, n. 4, p. 261-264, 2009.
- SANTOS JUNIOR, R. F. dos; SANTOS, J. M. dos; RUDORFF, B. F. T.; MARCHIORATO, I. A. Detecção de *Heterodera glycines* em plantio de soja mediante espectrorradiometria no visível e infravermelho próximo. **Fitopatologia Brasileira**, Brasília, v. 27, n. 4, p. 355-360, 2002.
- SAS INSTITUTE. **SAS/STAT: user's Guide**. Version 9.4. Cary: SAS Institute, 2009. 7869 p.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality. **Biometrika**, v. 52, p. 591-611, 1965.
- SILVA, J. F. V. da; DIAS, W. P.; GARCIA, A.; CARNEIRO, G. E. de S. Perdas por nematoides chegam a 10,6% da soja mundial. **Visão Agrícola**, Piracicaba, v. 3, n. 5, p. 103-107, 2006.
- ZHOU, Y.; WANG, Y.; ZHU, X.; LIU, R.; XIANG, P.; CHEN J.; LIU, X.; DUAN, Y.; CHEN, L. Management of the soybean cyst Nematode *Heterodera glycines* with combinations of different rhizobacterial strains on soybean. **PLoS One**, v. 13, n. 3, e0194287, 2017. DOI: 10.1371/journal.pone.0194287.