Pós-Graduação em Agronomia - CPGA-Solos
Análise Multivariada Aplicada as Ciências Agrárias

Regressão linear múltipla

Carlos Alberto Alves Varella

Objetivo da disciplina

- Ensinar modelagem estatística de fenômenos naturais aos alunos de pós-graduação utilizando técnicas da estatística multivariada.

Ementa da disciplina

- Regressão linear múltipla
- Regressão linear múltipla para dados repetidos
- Validação da predição
- Correlação múltipla
- Análise de componentes principais
- Análise discriminante de Fisher
- Análise de variância multivariada - MANOVA
- Análise de variáveis canônicas

Avaliações

- Uma Prova
- Trabalhos semanais
- Trabalho final: Cada aluno deverá apresentar um seminário e um trabalho escrito sobre aplicações de técnicas da estatística multivariada em sua tese.

Recursos computacionais

- SAS: recomendado para análises estatísticas multivariadas por Revistas de nível internacional.

Local para baixar arquivos da disciplina pela Internet

- http://www.ufrrj.br/institutos/it/deng/varella/multivariada.htm

Modelos Lineares (revisão)

Modelos lineares

- Seja Y a variável que queremos predizer a partir de um conjunto de variáveis preditoras X_1, X_2, ..., X_p. Então podemos escrever:

$$\boxed{Y = f\left(X_1, X_2, \cdots, X_p, \varepsilon\right)}$$

- Y representa a resposta;
- X_1,X_2,..., X_p são as variáveis estudadas;
- ε representa outro conjunto de variáveis não consideradas no estudo;

Requisitos da função $f(\cdot)$

- Deve prestar-se ao tratamento matemático;
- Deve ser adequada para o conjunto de dados em estudo;
- Deve ser simples ou pelo menos mais simples dentre as concorrentes.

Condição para que um modelo seja linear

- Um modelo para as observações Y será linear se:

$$E(Y) = X\beta$$

- Vamos estudar o caso em que os erros são normalmente distribuídos, independentes e homocedásticos.

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N\left(\mu, \sigma^2\right)$$

- Este modelo é definido como Modelo Linear de Gauss-Markov-Normal.

A superfície de resposta

- É a superfície gerada pelos valores da variável de resposta. O modelo linear para uma única variável de resposta ‘Y’ com ‘p’ variáveis preditoras é:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + e_i$$

$$i = 1,2,\cdots,n.$$

Y_i = superfície de resposta
n = número de observações;
p = número de variáveis preditoras.

- O modelo linear é a chave do negócio, isto é, tem inúmeras aplicações na estatística multivariada.

Duas situações são encontradas na modelagem

1. A matriz X'X de variáveis preditoras 'X' é de posto coluna completo. Neste caso o modelo é chamado de posto completo ou modelo de regressão. É o modelo que estamos estudando;
2. A matriz X'X de variáveis preditoras 'X' é de posto coluna incompleto. Neste caso o modelo é chamado de posto incompleto é o modelo da ANOVA (ANalysis Of VAriance)

Posto ou Rank de matrizes

- Número de linhas ou colunas linearmente independentes de uma matriz.
- Em nosso caso, o posto é o número de colunas linearmente independentes da matriz X'X, sendo X a matriz dos valores das variáveis preditoras ou "independentes"
- No programa computacional MATLAB o comando *rank* faz uma estimativa do posto de matrizes.

Condições para que a matriz X'X seja de posto coluna completo

- O posto ou rank da matriz X'X deve ser igual a 'p+1', ou seja:

$$posto(X'X) = p + 1$$

- *p* é o número de variáveis preditoras estudas no modelo.

Condições para que a matriz X'X tenha inversa $(X'X)^{-1}$

- As matrizes que possuem inversa são chamadas NÃO SINGULARES.
- Somente matrizes quadradas podem ser não singulares. Contudo, nem toda matriz quadrada é não singular;

Quando uma matriz quadrada é singular?

- Seu determinante é nulo; det(X'X)
- Ao menos uma de suas raízes características é nula. As raízes características são os autovalores da matriz; eig(X'X)
- Seu posto é menor que p; rank(X'X)
- Não é definida positiva ou negativa.

Matriz definida positiva (negativa)

- Quando todos os autovalores são positivos (negativos).

Regressão Linear Múltipla

Introdução

- É uma técnica da estatística multivariada utilizada para a predição de valores de uma ou mais variáveis de resposta (dependentes) a partir de diversas variáveis preditoras ou independentes.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 5th ed. Upper Saddle River, New Jersey: Prentice-Hall, 2002, 767 p.

Introdução (Cont.)

- Pode também ser utilizada para estudar o efeito dos preditores sobre as variáveis de resposta.
- Primeiro trabalho sobre o assunto: Regression Towards Mediocrity in Heredity Stature. Journal of the Anthropological Institute, 15 (1885). 246-263.
- Mediocridade em função da estatura hereditária
- Estatística UNIVARIADA. Segundo JOHNSON & WICHERN (2002) nesse artigo o autor não percebeu a importância da técnica para análises multivariadas.

Modelagem da Regressão Linear

Pressuposições da modelagem

- O modelo utilizado é o de Gauss-Markov-Normal
- Pressupõe que a resposta apresenta uma média. Pressupõe ainda que essa média contem erros provenientes de medições aleatórias e de outras fontes não explicitadas pelo modelo.
- O erro, e conseqüentemente a resposta, são tratados como variáveis aleatórias, que o comportamento é caracterizado assumindo-se uma distribuição NORMAL para os dados experimentais.

Estimadores dos parâmetros pelo método dos mínimos quadrados

- Este método consiste em se determinar o estimador que minimiza a soma do quadrado das diferenças entre valores observados e valores preditos pelo modelo.

$$Y = X\beta + \varepsilon \quad \textit{é o modelo linear}$$

$$\textit{Queremos determinar } \hat{\beta} \textit{ o estimador de } \beta$$

O erro da modelagem

- O erro do modelo na forma matricial é:

$$\varepsilon = Y - X\beta$$

$$\varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{21} \cdots X_{p1} \\ 1 & X_{12} & X_{22} \cdots X_{p2} \\ \cdots & \cdots & \cdots \quad \cdots \quad \cdots \\ 1 & X_{1n} & X_{2n} \cdots X_{pn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{bmatrix}$$

- O problema consiste em se ajustar um modelo de regressão.

Estimadores dos parâmetros pelo método dos mínimos quadrados

Modelo de regressão

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi},\ i = 1,2,\cdots,n.$$

- O estimador de beta é chamado de beta chapéu e pode ser determinado por outros métodos de minimização do erro, como por exemplo o método da máxima verossimilhança.

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdots \\ \hat{\beta}_p \end{bmatrix}$$

O método dos mínimos quadrados

- Sabendo que o erro do modelo é:

$$\varepsilon = Y - X\beta$$

- Então o somatório ao quadrado das diferenças dos erros pode ser representado na forma matricial por:

$$Z = \left\| Y - X\beta \right\|^2$$

- De acordo com o método temos que minimizar Z

Minimização da função Z

$$Z = \left\| Y - X\beta \right\|^2$$

$$Z = \left(Y - X\beta\right)' \left(Y - X\beta\right)$$
$$Z = \left(Y' - \beta' X'\right)\left(Y - X\beta\right)$$
$$Z = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

- As matrizes Y’Xβ e β’X’Y uma é a transposta da outra e são de dimensão 1x1, então as matrizes são iguais.

Diferenciando a função Z

$$Z = Y'Y - 2\beta' X' Y + \beta' X' X\beta$$

$$dZ = -2(d\beta')X'Y + (d\beta')X'X\beta + \beta' X' X(d\beta)$$

- As matrizes (dβ')X'Xβ e β'X'X(dβ) uma é a transposta da outra e são de dimensão 1x1, então as matrizes são iguais.

$$dZ = -2(d\beta')X'Y + 2(d\beta')X'X\beta$$
$$dZ = 2(d\beta')\cdot(X'X\beta - X'Y)$$

Fazendo com que a diferencial de Z seja igual a zero

$$dZ \equiv 0$$

- Para que a diferencial de Z seja zero

$$2(d\beta')\cdot(X'X\beta - X'Y) = 0$$

- Para que dZ seja zero, (X’Xβ-X’Y) deve ser igual a zero.

$$X'X\hat{\beta} - X'Y = 0$$

O beta chapéu

- Assim é chamado o vetor estimador dos parâmetros de beta.
- O vetor beta chapéu é determinado resolvendo-se o sistema de equações normais:

$$X'X\hat{\beta} = X'Y$$

Solução do sistema de equações normais

$$X'X\hat{\beta} = X'Y$$

- Multiplicando-se ambos os membros do sistema de equações por

$$(X'X)^{-1}$$

- Temos: $$(X'X)^{-1}X'X\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

- O modelo de regressão pressupõe um beta chapéu único não tendencioso (blue). Mas isso precisa de ser testado.

Conseqüências da estimação

- O modelo que estamos estudando é o Linear de Gauss-Markov-Normal.

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N\left(\mu, \sigma^2\right)$$

$$\varepsilon = Y - X\beta \; \textit{este é o erro do modelo}$$

A média do modelo linear

$$E(Y) = X\beta \text{ é a esperança matemática da população, também conhecido como média } '\mu'.$$

- Quando trabalhos com dados experimentais assumimos que o estimador da média ‘x barra’ pode representar a média ‘μ’ da população. Mas depois precisamos testar se isso é verdadeiro.

Os valores preditos pelo modelo

$\hat{Y} = X\hat{\beta}$ *são os valores preditos pelo modelo, isto é, valores obtidos para* Y *em função de uma combinação linear de valores de variáveis preditoras* X *e do estimador de* β*, o* $\hat{\beta}$*.*

- Quando trabalhos com dados experimentais determinamos o beta chapéu a partir de amostras da população. Por isso é que precisamos testar se esse beta é mesmo estimador não tendencioso.

O erro do modelo de regressão

$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta}$ *é o erro do modelo ajustado, também chamado de resíduo ou desvio.*

- Este é o erro que calculamos quando trabalhamos com dados experimentais.
- É um vetor que descreve a distribuição dos dados experimentais. Muitas inferências sobre nossos dados podem ser feitas analisando-se esse vetor.

O que queremos modelar

$$Y = \hat{Y} + \hat{\varepsilon}$$

$$Y:\ \textit{é o fenômeno que queremos modelar;}$$

$$\hat{Y}:\textit{é a modelagem do fenômeno estudado;}$$

$$\hat{\varepsilon}:\textit{é o erro na modelagem do fenômeno.}$$

- Quando trabalhos com dados experimentais assumimos que nossas observações são capazes de modelar o fenômeno, e depois testamos.

Prática 1

- Na tabela abaixo apresentamos os valores de uma amostra de 6 observações das variáveis Y_i, X_{1i} e X_{2i}.

Y_i	X_{1i}	X_{2i}
1,5	0	0
6,5	1	2
10,0	1	4
11,0	2	2
11,5	2	4
16,5	3	6

Fonte: Apostila de INF 664 Modelos Lineares. Adair José Regazzi,UFV, Viçosa, 2002.

Montar do sistema de equações normais

- Quando a regressão é com intercepto adicionados uma coluna de uns na matriz de dados.

X com intercepto

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \end{bmatrix}$$

X sem intercepto

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 2 \\ 1 & 4 \\ 2 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$$

Resposta Y

$$Y = \begin{bmatrix} 1,5 \\ 6,5 \\ 10,0 \\ 11,0 \\ 11,5 \\ 16,5 \end{bmatrix}$$

Obtenção da matriz X'X

- Esta matriz é obtida multiplicando-se a transposta da matriz X por ela mesma.

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 & 3 \\ 0 & 2 & 4 & 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 9 & 18 \\ 9 & 19 & 36 \\ 18 & 36 & 76 \end{bmatrix}$$

Obtenção da matriz X'Y

- Esta matriz é obtida multiplicando-se a transposta da matriz X pelo vetor Y.

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 & 3 \\ 0 & 2 & 4 & 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 6,5 \\ 10,0 \\ 11,0 \\ 11,5 \\ 16,5 \end{bmatrix} = \begin{bmatrix} 57 \\ 111 \\ 220 \end{bmatrix}$$

Sistema de equações normais

- Estimativa de beta pelos método dos mínimos quadrados

$$\begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} = \begin{bmatrix} 6 & 9 & 18 \\ 9 & 19 & 36 \\ 18 & 36 & 76 \end{bmatrix}^{-1} \begin{bmatrix} 57 \\ 11 \\ 220 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

$\hat{\beta}_0$: é o intercepto da equação de regressão;

β_1 e β_2 : são os regressores.

$Y_i = 2 + 3X_{1i} + 1X_{2i}$: é a equação de regressão

C:\disciplinas\Análise Multivariada Aplicada as Ciências Agrárias\pratica1.m

File Edit View Text Debug Breakpoints Web Window Help

Stack: Base

```
%Exemplo para montagem do sistema de equações normais
%Metodo dos minimos quadrados
%matriz X : variaveis perditoras
x=[1 0 0
   1 1 2
   1 1 4
   1 2 2
   1 2 4
   1 3 6];
%matriz Y: resposta
y=[1.5
   6.5
   10
   11
   11.5
   16.5];
%matriz x'x do sistema de equaçoes normais
xlinhax=x'*x;
%matriz x'y do sistema de equaçoes normais
xlinhay=x'*y;
%beta chapeu o estimador de beta
betahat=inv(xlinhax)*xlinhay
```

Programa na linguagem MATLAB

script Ln 1 Col 16

Exemplos de comandos do Programa computacional MATLAB

```
%posto da matriz X'X
rankxlinhax=rank(xlinhax)
%determinante da matriz X'X
detxlinhax=det(xlinhax)
%autovalores da matriz X'X
eigxlinha=eig(xlinhax)
```

MATLAB

File Edit View Web Window Help

Current Directory: C:\disciplinas\Análise Multivariada Aplicada as C

Resultados obtidos no Programa computacional MATLAB

```
betahat =
```

Vetor de parâmetros

```
    2.0000
    3.0000
    1.0000


rankxlinhax =
```

Posto da matriz

```
     3


detxlinhax =
```

Determinante da matriz

```
   240


eigxlinha =
```

Autovalores da matriz

```
    1.3303
    1.8442
   97.8255
```

Start

Análise de Variância da Regressão Linear

Análise de variância da regressão linear

- A análise de variância da regressão é a estatística utilizada para testar os regressores. A hipótese nula é que todos os regressores são iguais e zero. Caso isso não ocorra o resultado da análise é significativo, isto é, rejeita-se a hipótese nula.
- A análise de variância não testa o intercepto.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

Algumas Pressuposições do Modelo

- Beta chapéu é um estimador não tendencioso:

$$E\left(\hat{\beta}\right) = \beta$$

- A esperança do erro do modelo é zero e a esperança da variância dos erros é constante:

$$E\left(\varepsilon\right) = \phi \quad e \quad V\left(\varepsilon\right) = I\sigma^2$$

Variâncias e Covariâncias do Vetor Estimador dos Parâmetros

- O vetor estimador dos parâmetros é beta chapéu:
- A covariância deste vetor é:

$$\mathrm{Cov}\,(\hat{\beta}) = \mathrm{E}[(\hat{\beta}-\beta)\,(\hat{\beta}-\beta)^{'}] = (\mathrm{X'X})^{-1}\sigma^2$$

$$Cov(\hat{\beta}) = (X'X)^{-1}\hat{\sigma}^2$$

$$Cov(\hat{\beta}) = (X'X)^{-1}s^2$$

- s^2 é o Quadrado médio do resíduo.

Soma de Quadrado do Resíduo

- Soma dos quadrados dos desvios entre os valores observados e os estimados pela equação de regressão.

$$SQ\,Re\,s = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2$$

- Escrito na forma matricial é:

$$SQ\,Re\,s = Y'Y - \hat{\beta}'X'Y$$

Soma de Quadrado Total

$$SQTotal = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

- Matricialmente podemos escrever:

$$SQTotal = Y'Y - c$$

$$c = \frac{1}{n} Y'uu'Y$$

- u é um vetor de 1's de dimensão n x 1.

Soma de Quadrado da Regressão

$$SQ\,Re\,g = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$$

- Na forma matricial escrevemos:

$$SQ\,Re\,g = \hat{\beta}'\,X'\,Y - \frac{1}{n}\,Y'\,u\,u'\,Y$$

Esquema da análise de variância da regressão

Causa de variação	GL	SQ	QM	F
Regressão	p	$\hat{\beta}'X'Y - c$	SQReg/p	$\frac{QM\,Reg}{QM\,Res}$
Resíduo	n-p-1	$Y'Y - \hat{\beta}'X'Y$	SQRes/n-p-1	
Total	n-1	$Y'Y - c$		

- n =número de observações;
- p =número de variáveis
- Análise para dados não repetidos

Teste F dos parâmetros

- F é utilizado para testar a hipótese:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

- É o mesmo que testar se:

$$\beta_1 = \beta_2 = \cdots = \beta_p = 0$$

- Se os erros e_i têm distribuição normal e se o quociente

$$F = \frac{QM\,Reg}{QM\,Res}$$

- tem distribuição F (central) com p e n-p-1 graus de liberdade.

Quando o teste F é significativo?

- Quando F é maior que o tabelado;
- Quando rejeitamos a hipótese nula;
- Contudo não é possível concluir quais parâmetros são significativos;
- Exceto para o caso particular de $p=1$.

Teste *t* dos parâmetros

- *Utilizado para testar hipótese a respeito dos parâmetros da regressão .*
- *A estatística utilizada é:*

$$t = \frac{\hat{\beta}_i - \beta_i}{s(\hat{\beta}_i)}, \textit{associado a } (n - p - 1) \textit{ gl.}$$

- *O teste é significativo quando t é maior que o valor tabelado.*

Hipóteses a Respeito dos Parâmetros no Modelo Linear

- A hipótese de nulidade pode ser construída a partir de ***m*** combinações lineares independentes

$$H_0 : c'\beta = \theta$$

- c' é uma matriz com ***m*** linhas e ***p+1*** colunas

$$c' = [c_0 \; c_1 \; c_2 \; \cdots \; c_p]$$

- θ é um vetor ***m***-dimensional de constantes conhecidas.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$$

Estatística F usada para testar a hipótese H_0:c'β=θ

- Estatística de Wald
  Para teste F simultâneo dos parâmetros

$$F(H_0) = \frac{(C'\hat{\beta} - \theta)'[C'(X'X)^{-1}C]^{-1}(C'\hat{\beta} - \theta)}{m\hat{\sigma}^2}$$

- Sendo verdadeira a hipótese de nulidade a estatística **F(H_0)** tem distribuição **F** com ***m*** e ***n-posto[X]=n-p-1*** graus de liberdade.

Exemplo: testar a hipótese $H_0:\beta_1=\beta_2=0$

$$H_0: c'\beta = \theta \quad \Rightarrow \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow H_0: \beta_1 = 0 \text{ e } \beta_2 = 0$$

- Posto [c']=***m***=2

$$c'\hat{\beta} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$c'\hat{\beta} - \theta = \begin{bmatrix} 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Exemplo: testar a hipótese $H_0:\beta_1=\beta_2=0$

$$c'(x'x)^{-1}c = \frac{1}{240}\begin{bmatrix} 132 & -54 \\ -54 & 33 \end{bmatrix}$$

$$\left[c'(x'x)^{-1}c\right]^{-1} = \begin{bmatrix} 33/6 & 54/6 \\ 54/6 & 132/6 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 33/6 & 54/6 \\ 54/6 & 132/6 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 125{,}50$$

Exemplo: testar a hipótese $H_0:\beta_1=\beta_2=0$

$$\hat{\sigma}^2 = s^2 = QMR = \frac{y'y - \hat{\beta}'x'y}{n-p-1} = \frac{3,00}{6-2-1} = 1,00$$

$$F(H_0) = \frac{125,50}{2\cdot(1,00)} = 62,75^{**}$$

$$F_{1\%}(2;3) = 30,82$$

- **Rejeita-se a hipótese $H_0:\beta_1=\beta_2=0$**

Estatística *t* usada para testar a hipótese H_0:c'β=θ

- Podemos usar *t* para testar hipóteses a respeito de combinações lineares dos parâmetros

$$t = \frac{c'\hat{\beta} - c'\beta}{\sqrt{\hat{V}(c'\hat{\beta})}}, \textit{associado} \text{ a (n - p - 1) gl.}$$

$$\mathrm{n - p - 1 = n - posto(X) = GLR}$$

Teste Simultâneo dos Parâmetros

- Testa uma única hipótese;
- Testa um vetor de betas;
- Não é o mesmo que testar os betas separadamente.
- Isto é, ***testar***

$$H_0 : \beta_1 = 0 \ \text{ e } \ H_1 : \beta_2 = 0$$

- ***Não é o mesmo que testar***

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{ou} \quad H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Programa SAS (reg_cap1.sas)

```
proc reg data=sas.ind_v9;
/*ndvi rnir gnir arvi savi gndvi*/
model N = gndvi;
output out=p  p=yhat r=resid;
print p;
run;
quit;
proc reg;
model yhat=N;
test N=1, intercept=0;
run;
plot yhat*N;
run;
quit;
```

Output do SAS – Análise de variância do modelo de regressão

The SAS System 23:15 Thursday, October 7, 2009 5

The REG Procedure
Model: MODEL1
Dependent Variable: N N

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	20710	3451.59735	4.39	0.0293
Error	8	6290.41589	786.30199		
Corrected Total	14	27000			

Root MSE	28.04108	R-Square	0.7670
Dependent Mean	60.00000	Adj R-Sq	0.5923
Coeff Var	46.73513		

Teste t dos beta-chapéu do modelo de regressão

Parameter Estimates

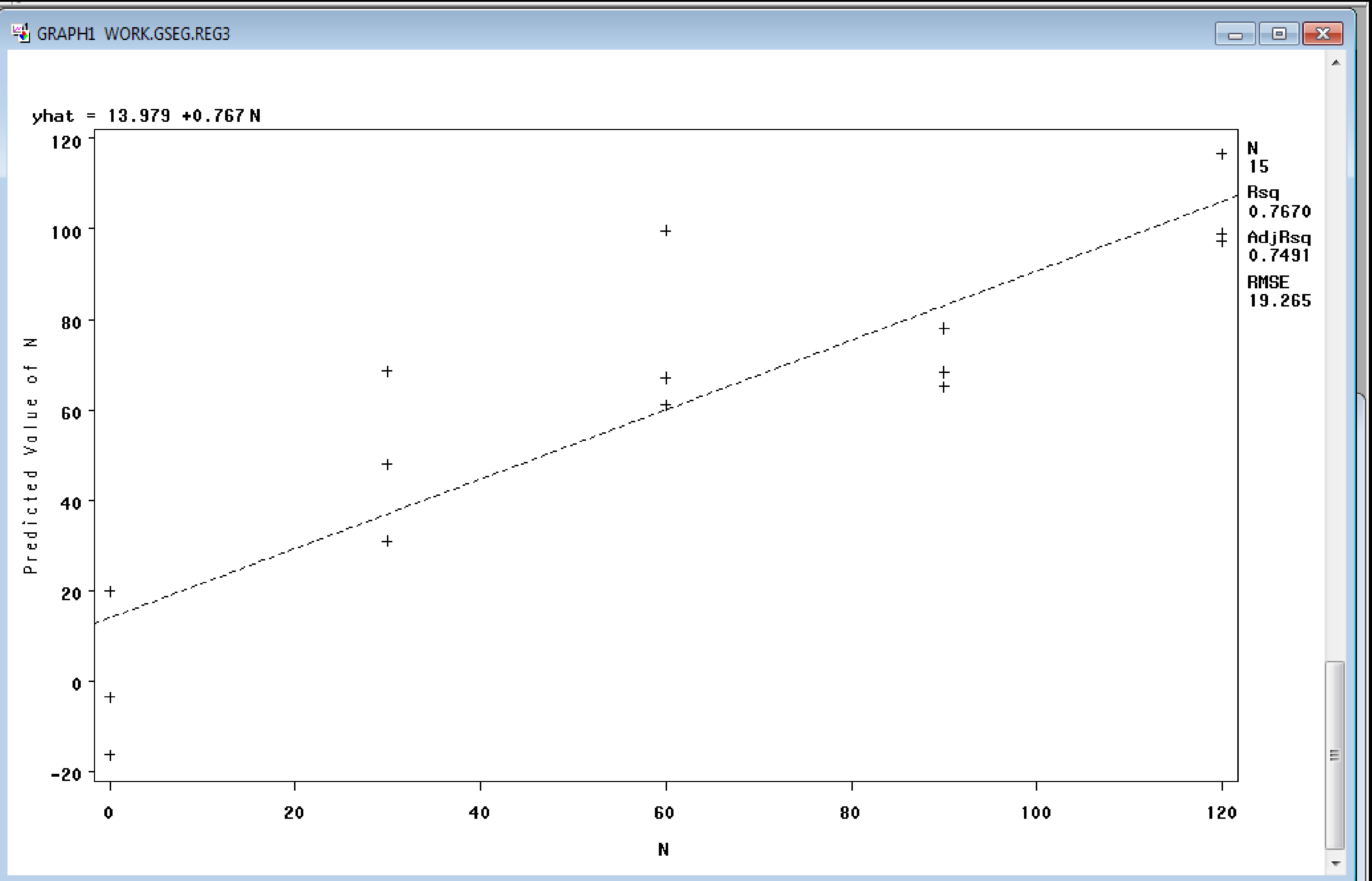
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1835.59747	1483.61562	1.24	0.2511
NDVI	NDVI	1	-15182	19298	-0.79	0.4541
RNIR	RNIR	1	-1698.66240	3814.27214	-0.45	0.6679
GNIR	GNIR	1	-413.90081	2665.47402	-0.16	0.8804
ARVI	ARVI	1	546.46984	283.26026	1.93	0.0898
SAVI	SAVI	1	8350.10834	13196	0.63	0.5445
GNDVI	GNDVI	1	594.04446	2908.94995	0.20	0.8433

Níveis de N preditos pelo modelo

Obs	Dependent Variable	Predicted Value	Residual
1	0	-16.4019	16.4019
2	0	-3.4152	3.4152
3	0	19.8021	-19.8021
4	30.0000	30.9970	-0.9970
5	30.0000	68.5033	-38.5033
6	30.0000	47.8805	-17.8805
7	60.0000	67.1267	-7.1267
8	60.0000	99.6748	-39.6748
9	60.0000	61.1820	-1.1820
10	90.0000	68.4044	21.5956
11	90.0000	65.1605	24.8395
12	90.0000	78.0660	11.9340
13	120.0000	97.4010	22.5990
14	120.0000	116.5953	3.4047
15	120.0000	99.0235	20.9765

Sum of Residuals	-3.6067E-11
Sum of Squared Residuals	6290.41589
Predicted Residual SS (PRESS)	28335

Gráfico: Predito x Observado



Conclusão

- O modelo de regressão multivariado proposto não pode ser utilizado para predizer níveis de N aplicados no solo.

Exemplo de regressão linear múltipla com duas vaiáveis independentes

Y	X1	X2
1,5	0	0
6,5	1	2
10	1	4
11	2	2
11,5	2	4
16,5	3	6

Programa SAS

```
proc reg data=reg.inf664p26;
/*Seleção por teste t dos parâmetros da regressão*/
model Y = X1 X2 / selection=stepwise;
output out=p  p=yhat r=resid;
print p;
run;
quit;
proc reg;
model yhat=Y;
test Y=1, intercept=0;
run;
plot yhat*Y;
run;
quit;
```

Resumo do Stepwise

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	X1		X1	1	0.9201	0.9201	8.2727	46.04	0.0025
2	X2		X2	2	0.0566	0.9767	3.0000	7.27	0.0740

Valores preditos

Model: MODEL1
Dependent Variable: Y Y

Output Statistics			
Obs	Dependent Variable	Predicted Value	Residual
1	1.5000	2.0000	-0.5000
2	6.5000	7.0000	-0.5000
3	10.0000	9.0000	1.0000
4	11.0000	10.0000	1.0000
5	11.5000	12.0000	-0.5000
6	16.5000	17.0000	-0.5000

Sum of Residuals	0
Sum of Squared Residuals	3.00000
Predicted Residual SS (PRESS)	22.21749

Regressão entre predito e observado

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: yhat Predicted Value of Y

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	122.57004	122.57004	167.33	0.0002
Error	4	2.92996	0.73249		
Corrected Total	5	125.50000			

Root MSE	0.85586	R-Square	0.9767
Dependent Mean	9.50000	Adj R-Sq	0.9708
Coeff Var	9.00902		

Validação da predição

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	0.22179	0.79783	0.28	0.7948
Y	Y	1	0.97665	0.07550	12.94	0.0002

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable yhat				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	2	0.03502	0.05	0.9539
Denominator	4	0.73249		