PLANEJAMENTO E ANÁLISE DE EXPERIMENTOS
Regressão

9 a 14 de Setembro

Pós-Graduação em Produção Vegetal UFPR

Éder David Borges da Silva
Renato Gonçalves de Oliveira

Introdução

- Objetivo: Relacionar matematicamente duas ou mais variáveis, com uma função, por exemplo:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- Importante para entender os fenômenos físicos, químicos, biológicos, sociais, médicos...etc.
- Variáveis devem ser quantitativas
- Importante no estudo de otimização de processos

Classificação dos modelos quanto as parâmetros

- Modelo linear

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$\frac{\partial Y_i}{\partial \beta_0} = 0 \quad \frac{\partial Y_i}{\partial \beta_1} = X_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

- Modelo não-linear

$$Y_i = \beta_0 + \beta_1 X_i^{\beta_2} + \varepsilon_i$$

$$\frac{\partial Y_i}{\partial \beta_0} = 0 \quad \frac{\partial Y_i}{\partial \beta_1} = X_i^{\beta_2} \quad \frac{\partial Y_i}{\partial \beta_2} = \beta_1 X^{\beta_2} \ln(X_i)$$

Modelos estatísticos quanto ao # variáveis

- Modelo univariado:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- Modelo múltiplo:

$$Y_i = \beta_0 + \beta_1 X_i + \ldots + \beta_n X_n + \varepsilon_i$$

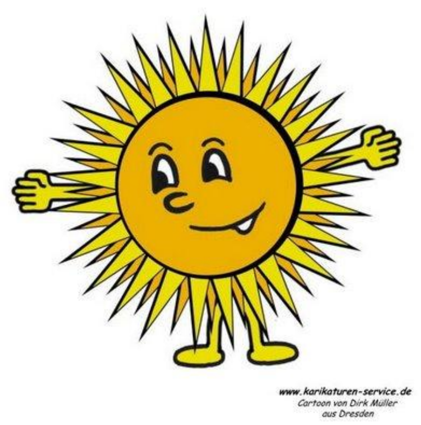
Análise de regressão linear

Experimento univariado

$$Y = \begin{bmatrix} 0.88 \\ 0.90 \\ 0.99 \\ 1.12 \\ 1.40 \\ 1.62 \\ 2.20 \\ 3.10 \end{bmatrix} \qquad X = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.5 \\ 0.8 \\ 1.0 \\ 1.5 \\ 2.0 \end{bmatrix}$$

Altura

Tempo de exposição



Experimento múltiplo

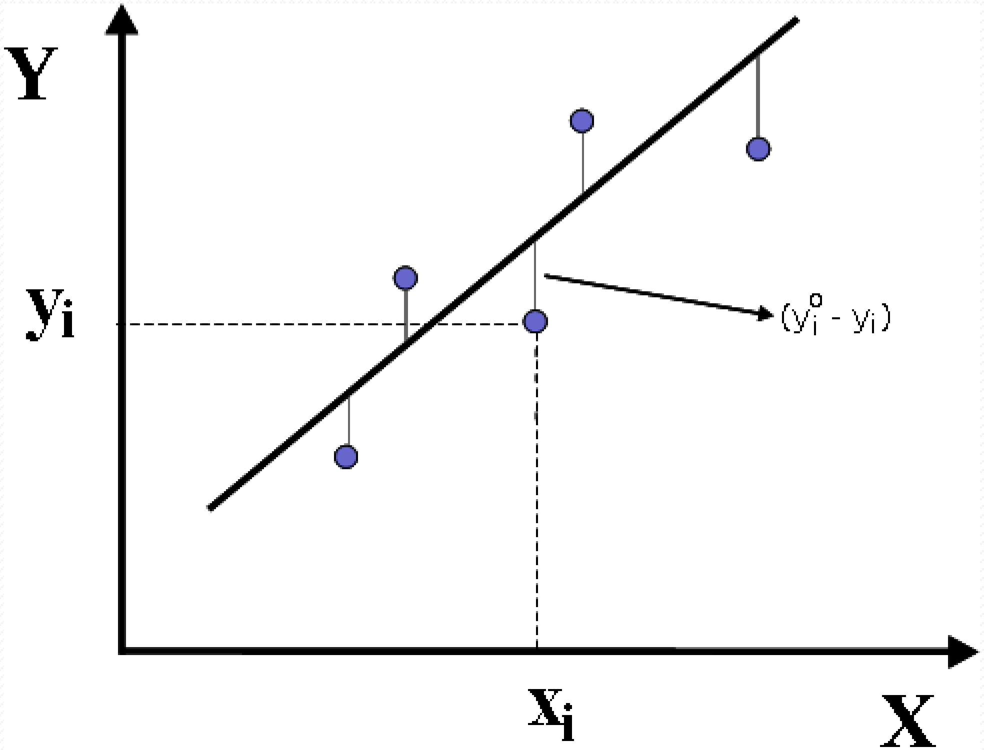
- Volume estimado de madeira

$$Y = \begin{bmatrix} 65 \\ 78 \\ 82 \\ 86 \\ 87 \\ 90 \\ 93 \\ 96 \\ 104 \\ 113 \end{bmatrix} \quad X_1 = \begin{bmatrix} 41 \\ 71 \\ 90 \\ 80 \\ 93 \\ 90 \\ 87 \\ 95 \\ 100 \\ 101 \end{bmatrix} \quad X_2 = \begin{bmatrix} 79 \\ 48 \\ 80 \\ 81 \\ 61 \\ 70 \\ 96 \\ 84 \\ 78 \\ 96 \end{bmatrix} \quad X_3 = \begin{bmatrix} 35 \\ 53 \\ 64 \\ 59 \\ 66 \\ 64 \\ 62 \\ 67 \\ 70 \\ 71 \end{bmatrix}$$

Volume/área   Área basal   Área basal relativa   Altura

Método dos mínimos quadrados

- Idéia do Método:



$$\min\left[S = \sum_{i=1}^{n} (y_i^0 - y_i)^2\right]$$

Estimador de Mínimo Quadrado

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \xrightarrow{\text{MMQ}} \underset{\sim}{\hat{\beta}} = (X'X)^{-1} X'Y$$

$$Y = \begin{bmatrix} 0.88 \\ 0.90 \\ 0.99 \\ 1.12 \\ 1.40 \\ 1.62 \\ 2.20 \\ 3.10 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0.1 & 0.01 \\ 1 & 0.2 & 0.04 \\ 1 & 0.3 & 0.09 \\ 1 & 0.5 & 0.25 \\ 1 & 0.8 & 0.64 \\ 1 & 1.0 & 1.00 \\ 1 & 1.5 & 2.25 \\ 1 & 2.0 & 4.00 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

X^0 X^1 X^2

Função lm

- **Fitting Linear Models**
- ajuste1 <- lm(Produção ~ altura)
- ajuste2 <- lm(Produção ~ altura+diametro)
- ajuste3 <- lm(Produção ~ I(altura)+I(diametro)+I(diametro)^2)

```
Summary(ajuste1)
Call:
lm(formula = Y ~ X1)

Residuals:
    Min       1Q     Median      3Q      Max
-10.7994  -2.6942  -0.1651   3.7156  13.0095

Coefficients:
             Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)  33.9634    11.4989     2.954     0.01832 *
X1            0.6537     0.1330     4.916     0.00117 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.106 on 8 degrees of freedom
Multiple R-squared: 0.7513,     Adjusted R-squared: 0.7202
F-statistic: 24.17 on 1 and 8 DF,  p-value: 0.001170
```

```
Anova(ajuste1)

Analysis of Variance Table

Response: Y
          df  Sum Sq    Mean Sq    F value   Pr(>F)
X1         1 1220.39    1220.39    24.165    0.001170 **
Residuals  8  404.01      50.50
```

ANOVA da Regressão

Tabela da análise de variância (ANOVA) do modelo de regressão.

F.V.	S.Q.	g.l.	Q.M.	F	p-valor
Regressão	$\hat{\beta}\,'X'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$	$p - 1$	SQreg/(p – 1)	QMreg/QMres	depende de F
Resíduos	$Y'Y - \hat{\beta}\,'X'Y$	$n - p$	SQres/(n – p)	---	---
Total	$Y'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$	$n - 1$	----		

n = tamanho amostral e p = número de parâmetros.

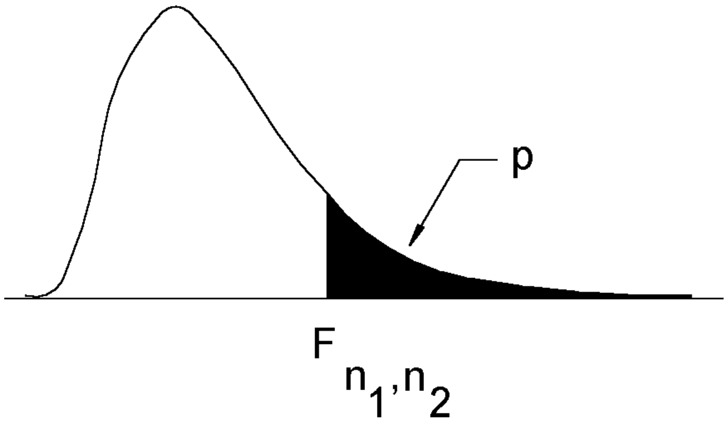
Fonte: Apostila Prof Suely Giolo UFPR

ANOVA da Regressão

FV	GL	QM	F_c	$Pr(F > F_c)$
Regressão	3	455,85296	10,65	0,0081
Erro	6	42,80685		
Total Corrigido	9			

$$H_0 : \beta_1 = \beta_2 = \beta_3 = ... = \beta m = 0$$

$H_1 : \beta_i \neq 0$ Para algum i=1,2,3,...,m

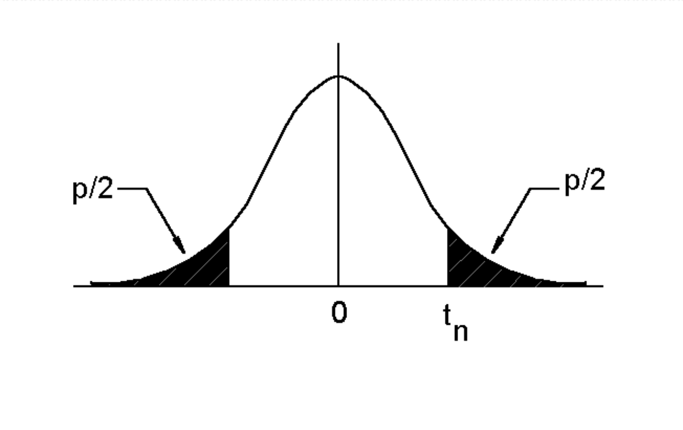


Teste t de Student

$$H_0 : \beta_i = \delta_0$$

$$H_0 : \beta_i \neq \delta_0$$

$$t_c = \frac{\hat{\beta}_i - \delta_0}{S_{(\hat{\beta}_i)}}$$



$$S_{(\hat{\beta}_i)} = \sqrt{(X'X)^{-1} QME}$$

Coeficiente de determinação R^2_{aju}

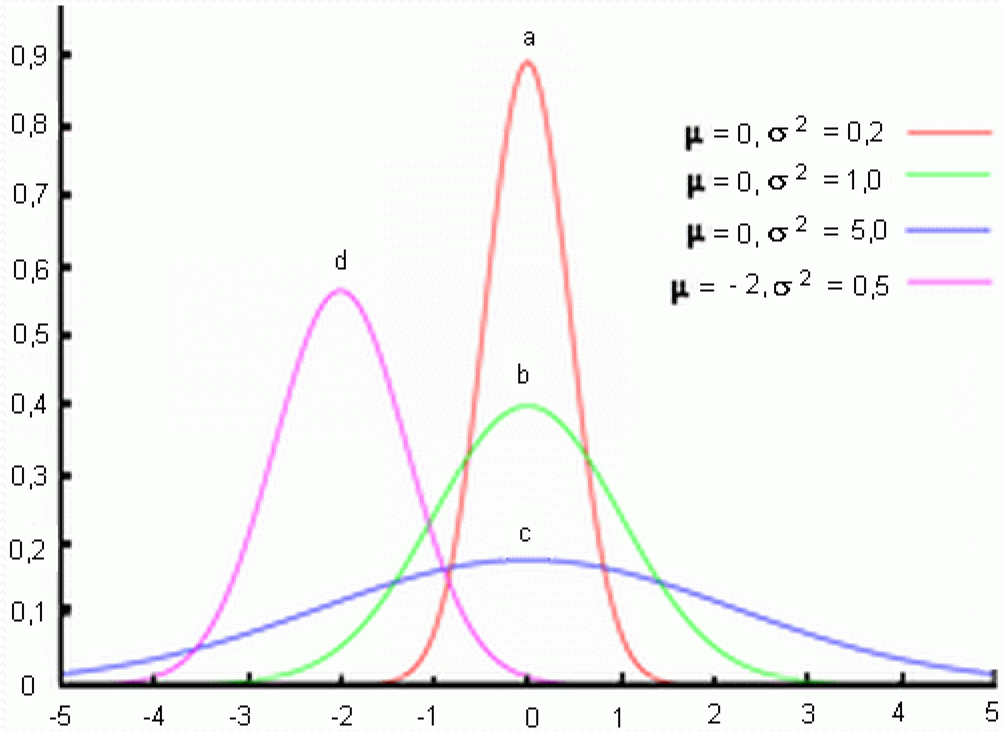
$$R^2 = 1 - \frac{SQE}{SQT}$$

$$R^2_{Aj} = 1 - \frac{n-i}{n-p}(1 - R^2)$$

- n tamanho da amostra
- p número de parâmetros
- i igual a 1, se inclui intercepto ou o sem intercepto

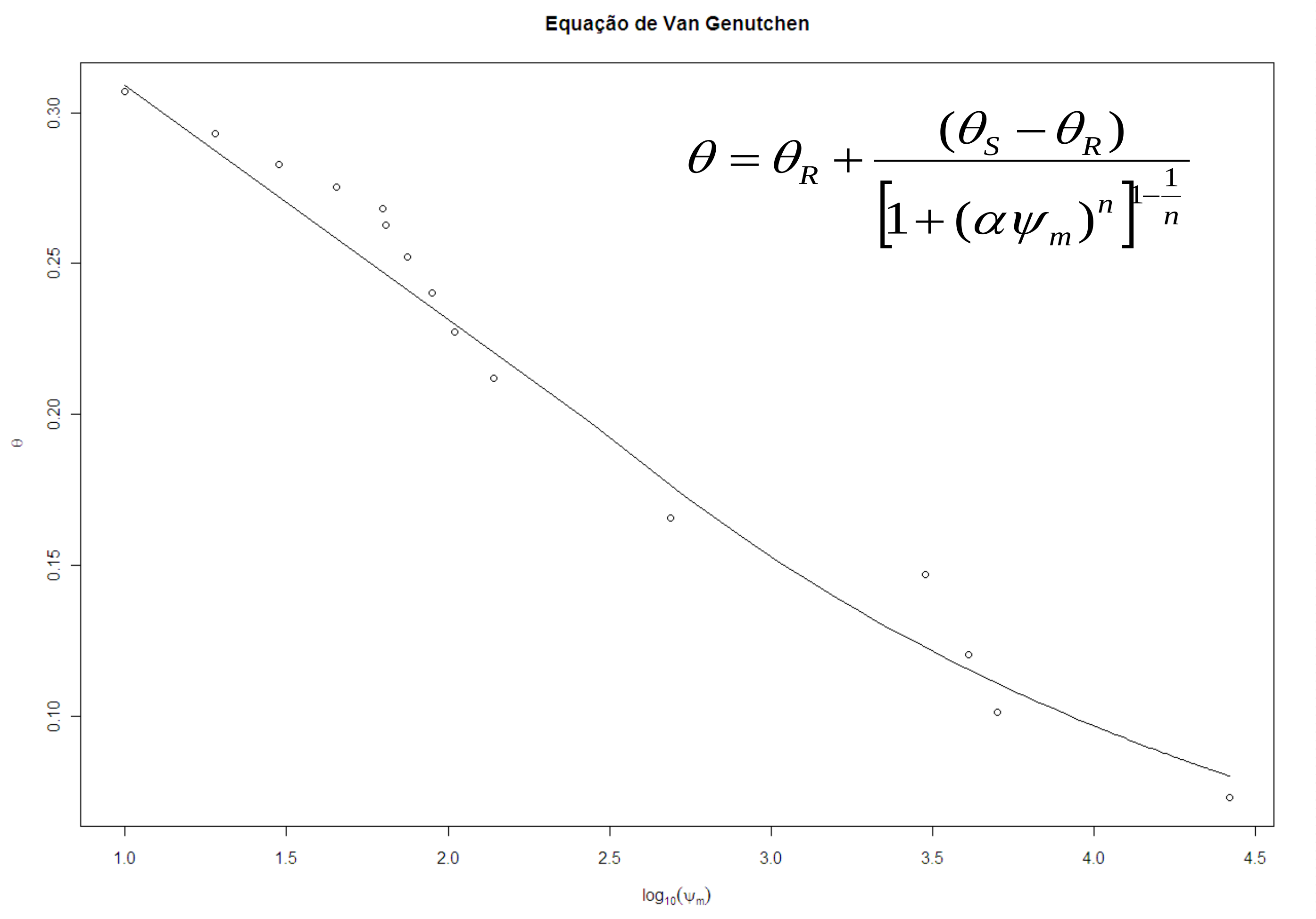
Análise do resíduo

- Com distribuição normal e μ=o e σ^2 conhecido



Análise de regressão não linear

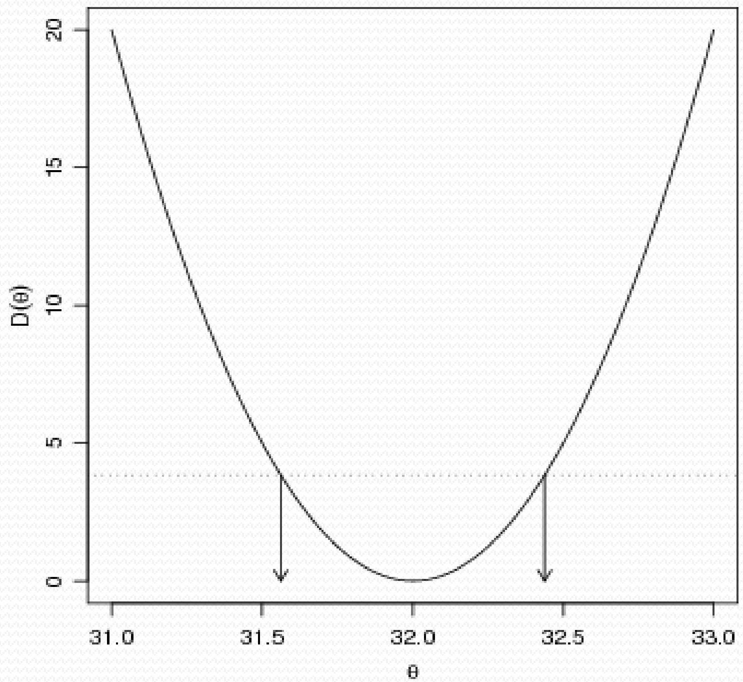
Introdução

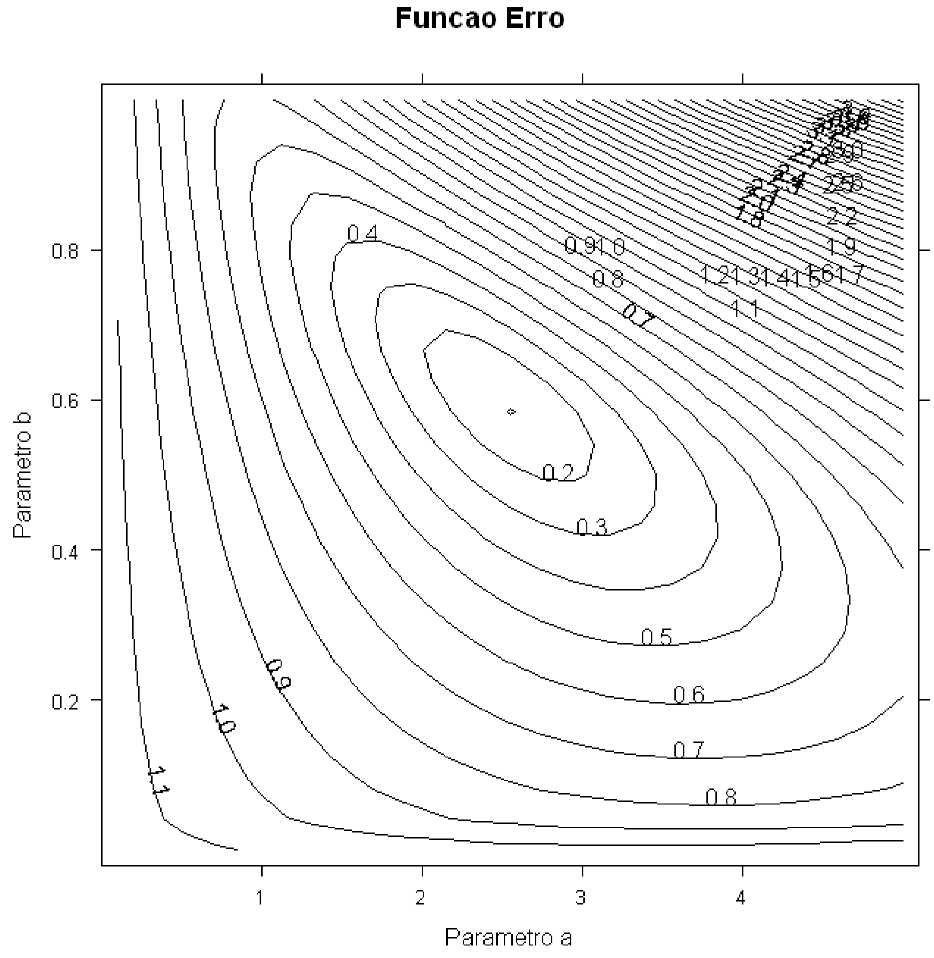


Equação de Van Genutchen
$$\theta = \theta_R + \frac{(\theta_S - \theta_R)}{\left[1 + (\alpha\psi_m)^n\right]^{1-\frac{1}{n}}}$$
0.30
0.25
0.20
0.15
0.10
θ
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
$\log_{10}(\psi_m)$

Métodos de Interativos

Ideia: SQE tem de ser minimizadas, em função dos parâmetros





Métodos de Interativos

$$X = \begin{bmatrix} \theta^{Z_1} & Z_1\alpha\theta^{(Z_1-1)} \\ \theta^{Z_2} & Z_2\alpha\theta^{(Z_2-1)} \\ \vdots & \vdots \\ \theta^{Z_n} & Z_n\alpha\theta^{(Z_n-1)} \end{bmatrix}$$

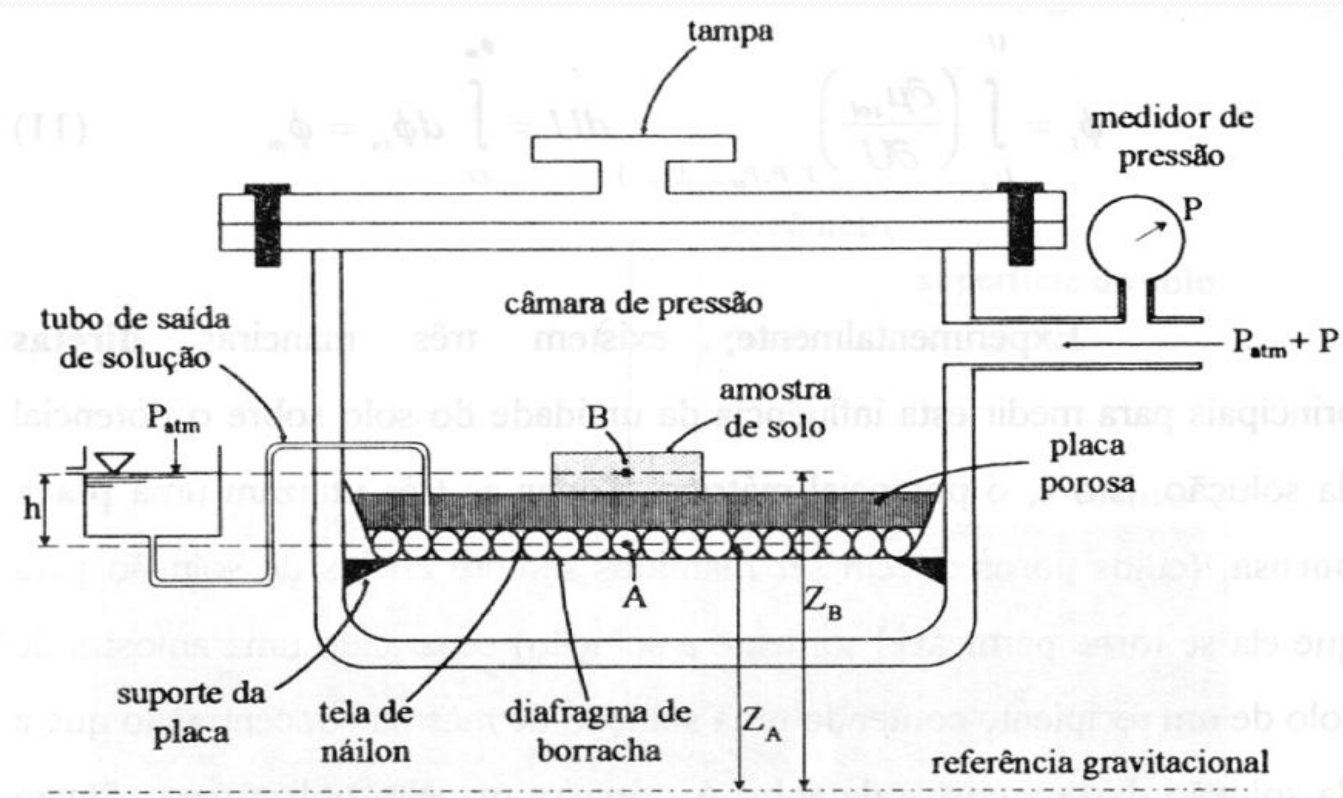
$$SQE\left(\underset{\sim}{\beta_k} + \lambda\underset{\sim}{\Delta}\right) < SQE\left(\underset{\sim}{\beta_k}\right)$$

$$\begin{cases} \text{Gradiente:} & \underset{\sim}{\Delta} = X'\underset{\sim}{e} \\ \text{Gauss-Newton:} & \underset{\sim}{\Delta} = (X'X)^{-}X'\underset{\sim}{e} \\ \text{Newton:} & \underset{\sim}{\Delta} = G^{-}X'\underset{\sim}{e} \\ \text{Marquardt:} & \underset{\sim}{\Delta} = [X'X + \delta\text{diag}(X'X)]^{-}X'\underset{\sim}{e} \end{cases}$$

Experimento

Determinação Laboratorial Câmara de Richards

$$pot = \begin{bmatrix} 10 \\ 19 \\ 30 \\ 45 \\ 63 \\ 64 \\ 75 \\ 89 \\ 105 \\ 138 \\ 490 \\ 3000 \\ 4100 \\ 5000 \\ 26300 \end{bmatrix} \quad u = \begin{bmatrix} 0.3071 \\ 0.2931 \\ 0.2828 \\ 0.2753 \\ 0.2681 \\ 0.2628 \\ 0.2522 \\ 0.2404 \\ 0.2272 \\ 0.2120 \\ 0.1655 \\ 0.1468 \\ 0.1205 \\ 0.1013 \\ 0.0730 \end{bmatrix}$$



Função nls

$$\theta = \theta_R + \frac{(\theta_S - \theta_R)}{\left[1 + (\alpha\psi_m)^n\right]^{1-\frac{1}{n}}}$$

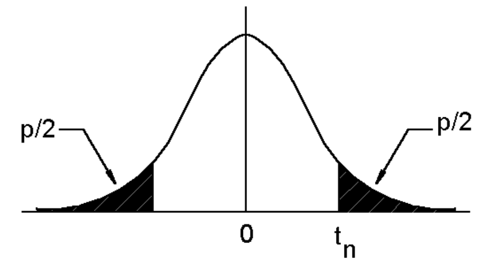
- ajuste<-nls(u~ur+(us-ur)/((1+(alpha*pot)^n)^(1-1/n)),start=list(us=0.2236,ur=0.0611,alpha=0.056,n=1.5351))

```
>summary(ajuste)
Formula: u ~ ur + (us - ur)/((1 + (alpha * pot)^n)^(1 - 1/n))

Parameters:
      Estimate    Std. Error  t value Pr(>|t|)
us     0.324120   0.017744   18.27   1.41e-09 ***
ur     0.007083   0.071083    0.10   0.922
alpha 0.038780   0.026202    1.48   0.167
n     1.211817   0.105207    11.52   1.77e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01104 on 11 degrees of freedom

Number of iterations to convergence: 10
Achieved convergence tolerance: 4.152e-06
```



Coeficiente de determinação R^2

- Método de calculo manual no R

```
# R^2
# soma de quadrados residual
SQE <- summary(ajuste)$sigma^2*summary(ajuste)$df[2]
SQE
# soma de quadrado total corrigida
SQT <- var(u)*(length(u)-1)
SQT
R2 <- 1 - SQE/SQT
R2
```

Enfim, o fim esta proximo....