

# LCF – 5759 BIOESTATÍSTICA

Hilton Thadeu Zarate do Couto

E-mail: [htzcouto@usp.br](mailto:htzcouto@usp.br)

# BIOESTATÍSTICA

- Desligar celular.
- Não ficar entrando e saindo da sala (trazer uma garrafinha de água). Somente em emergências.
- Trazer um PenDrive para armazenar os dados e programas.
- O aluno que quiser usar o próprio notebook será bem-vindo.
- PROGRAMA DA DISCIPLINA
- *AULAS (Segundas-feiras) das 14 às 17 h.: LAPI – Laboratório de Processamento de Imagens - LCF*
- **NOTA FINAL= (1ª PROVA T P + 2ª PROVA T P + EXERCÍCIOS)/3**

# FORMATO DOS EXERCÍCIOS:

- TIPO: RELATÓRIO CIENTÍFICO
- CAPA: TÍTULO, NOME(S), DISCIPLINA, DATA (DD/MM/AAAA)
- INTRODUÇÃO: COM OBJETIVOS BEM CLAROS
- REVISÃO BIBLIOGRÁFICA: MÁXIMO 3 PÁGINAS (DEVE ENFOCAR O ASSUNTO PRINCIPAL DO TRABALHO) MATERIAL E MÉTODOS
- RESULTADOS E DISCUSSÃO
- CONCLUSÕES
- BIBLIOGRAFIA (NORMAS DA ABNT)

# Exercícios

- Leiam alguma tese e/ou dissertação em sua área para servir de exemplo.
- Leiam também:  
<http://www.pdhonline.org/courses/g102f/g102f.htm>
- **Cuidado com plágio:**
  - 1. Assinar ou apresentar como seu (obra artística ou científica de outrem).
  - 2. Imitar (trabalho alheio) (Dicionário Aurélio)

# Definições e Origens da Estatística

- Biometria, Estatística Biológica, Bioestatística, Ecometria.
- Bioestatística: estatística aplicada a análise de dados biológicos.
- ***Estatística (definição)***: uma parte da matemática que trata da coleta, análise, interpretação e apresentação de uma massa de dados numéricos.
- **Outra definição**: Ciência e prática para o desenvolvimento do conhecimento humano através do uso de dados *empíricos* (originado ou baseado em experiências apenas) expressos na forma quantitativa.
- Empírico: Derivado de experimento ou de observação da realidade. (Dicionário Aurélio).
- Ainda outra definição: Tecnologia quantitativa para medir incertezas nas observações e experimentos e para avaliar as implicações destas incertezas. (Whitlock & Schluter, 2005)

# Definição de Estatística

- Estudo científico de dados numéricos extraídos de algum fenômeno natural
- Por que *estudo científico*?
- Por que *dados numéricos*?
- Por que *variação na natureza*?  
(animada e inanimada)

# Exemplo

- Título: Storks Deliver Babies ( $p = 0.008$ )
- Autor: Robert Matthews
- Local: Aston University, Birmingham, Inglaterra.
- E-mail: [rajm@compuserve.com](mailto:rajm@compuserve.com)
- Publicação: Teaching Statistics. Volume 22, N<sup>o</sup> 2, 2000, p. 36-38.

```

DATA A0;
INPUT PAIS $ CEGONHAS TX_NAS;*1000 POR ANO;
DATALINES;
Albânia      100      83
Áustria     300      87
Bélgica      1      118
Bulgária    5000     117
Dinamarca    9      59
França      140     774
Alemanha    3300     901
Grécia     2500     106
Holanda      4      188
Hungria     5000     124
Itália      5      551
Polônia     30000    610
Portugal    1500     120
Romênia     5000     367
Espanha     8000     439
Suíça       150      82
Turquia     25000    1576
;;;
ODS RTF;
TITLE1 'STORKS DELIVER BABIES - MATTHEWS, 2000';
PROC REG DATA=A0
PLOTS (ONLY)=PREDICTIONS (X=CEGONHAS) ;
MODEL TX_NAS = CEGONHAS;
RUN;
ODS RTF CLOSE;

```

**Programa SAS  
para análise dos  
dados  
apresentados  
por Matthews**

# Alta correlação entre taxa de nascimento e número de cegonhas em diversos países da Europa



# Análise feita com dados dos países europeus (Matthews, 2000).

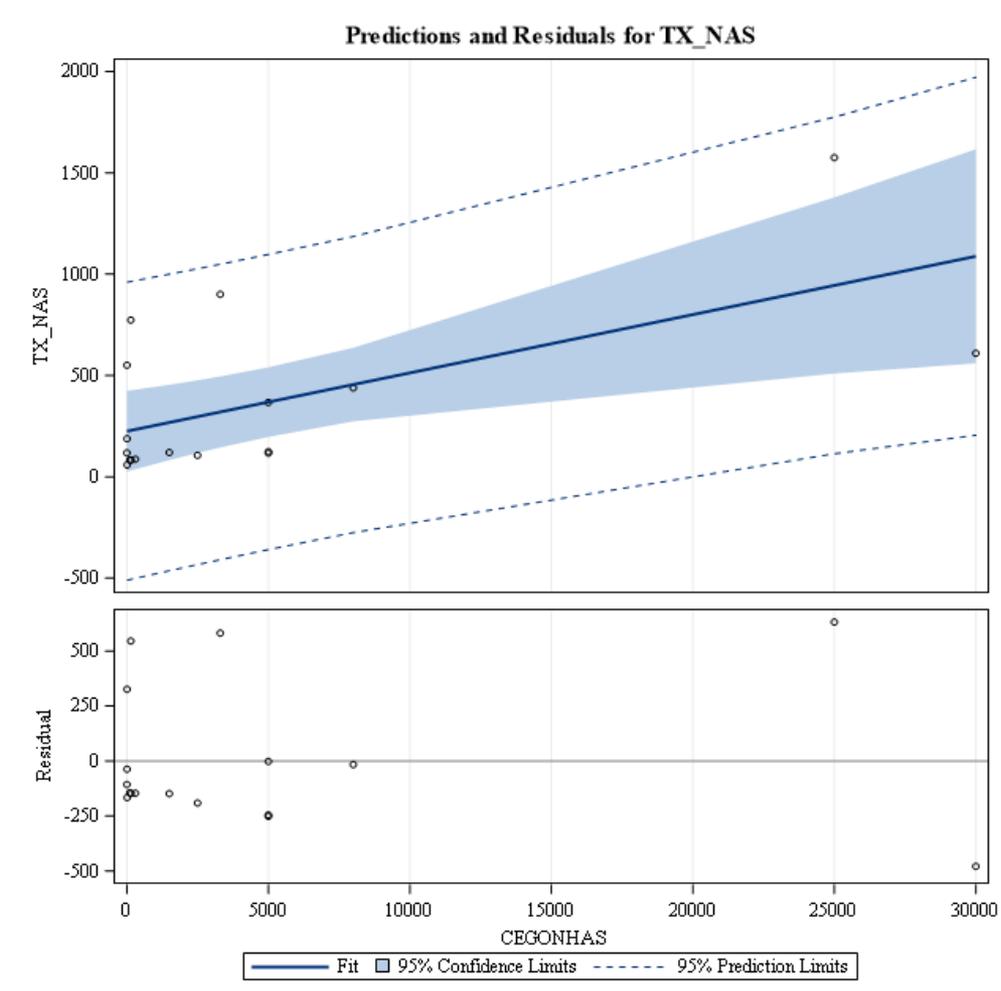
Análise feita no SAS com os dados publicados.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1035001	1035001	9.38	<b>0.0079</b>
Error	15	1655206	110347		
Corrected Total	16	2690208			

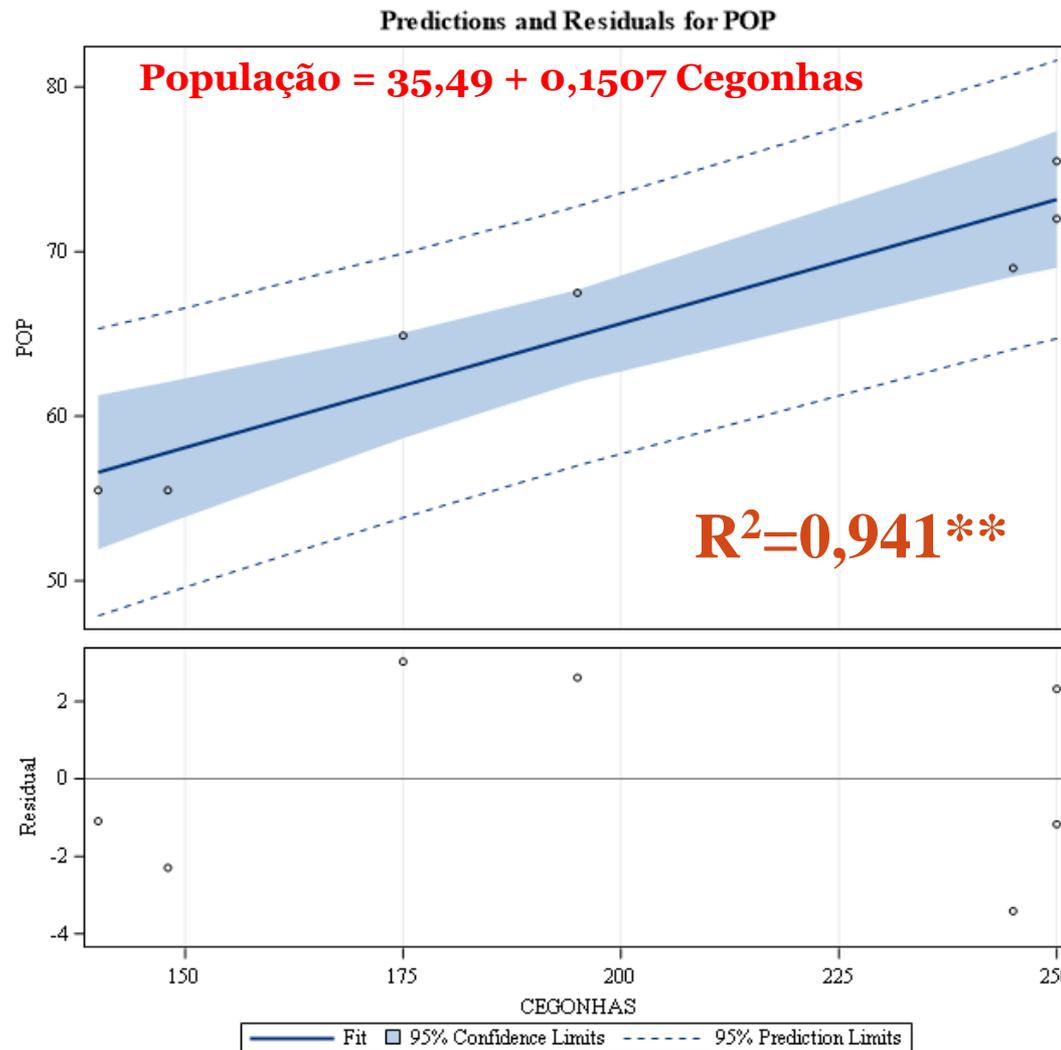
Root MSE	332.18533	R-Square	<b>0.3847</b>
Dependent Mean	370.70588	Adj R-Sq	0.3437
Coeff Var	89.60886		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	225.02869	93.56058	2.41	0.0295
cegonhas	1	0.02879	0.00940	3.06	0.0079

# Gráfico da análise

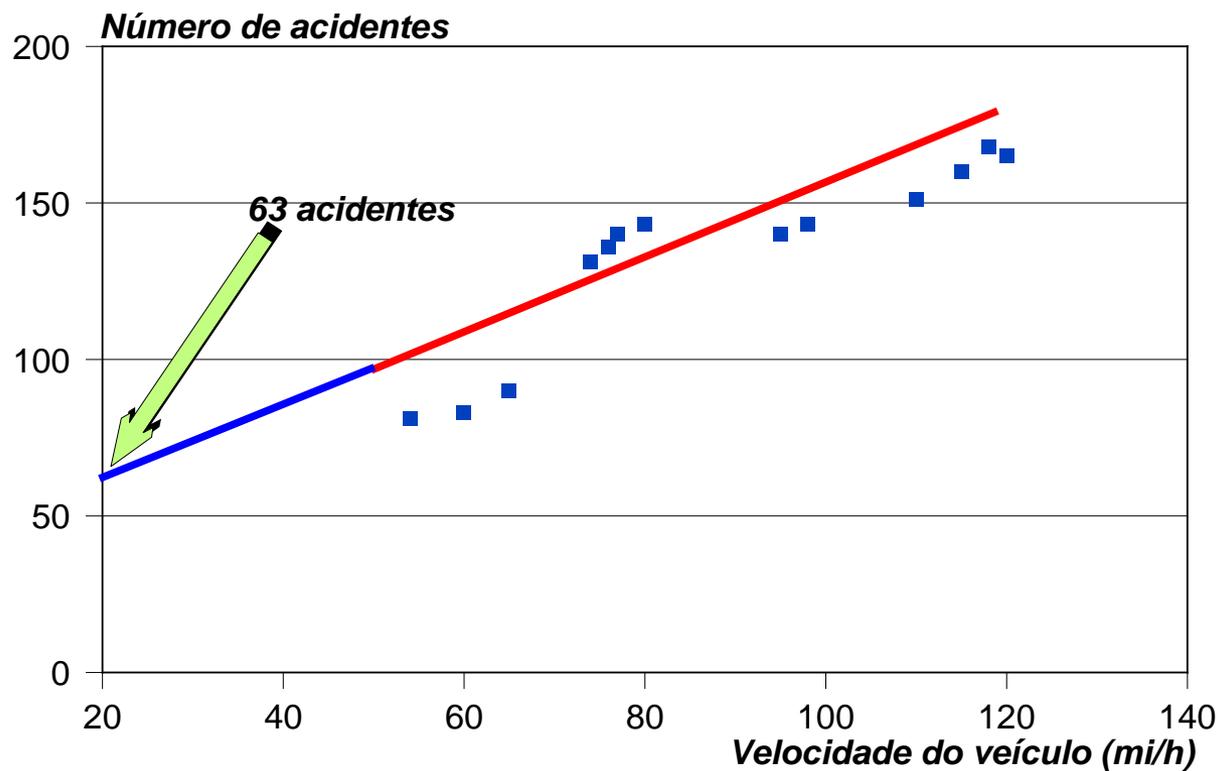


## Gráfico mostrando a população da cidade alemã de Oldenburg e o número de cegonhas, observados entre 1930 e 1936.



<u>Cegonhas</u>	<u>População</u> (mil)
140	55,5
148	55,5
175	64,9
195	67,5
245	69,0
250	72,0
250	75,5

Gráfico da relação entre a velocidade do veículo em milhas/h e o número de acidentes na cidade de Los Angeles-E.U., entre 1980 e 1985.



# Origens da Estatística

*Origem no século XVII: Ciência Política e Teoria da Probabilidade*

## Ciência Política

(Inglaterra) - estatística vital

John Graunt (1620-1674)

William Petty (1623-1687)

## Teoria da Probabilidade

(França) - jogos de azar

Blaise Pascal (1623-1662)

Pierre de Fermat (1601-1665)

Jacques Bernoulli (1654-1705)

Abraham de Moivre (1667-1754)

# Origens da Estatística

## Importância da Astronomia

Pierre Simon Laplace (1749-1827) -  
França- Métodos dos quadrados  
mínimos.

Karl Friedrich Gauss (1777- 1855)-  
Alemanha - Distribuição de dados  
(Gauss ou Normal)

## Aplicação na Biologia

Francis Galton (1822-1911) -  
Regressão e Correlação

Karl Pearson (1857-1936)-Continuou  
os estudos de Galton e iniciou os  
estudos da estatística descritiva.

Ronald Aylmer Fischer (1890-1962)-  
Estação Experimental de  
Rothamsted

# Teoria da Probabilidade

*Qual a chance de ganhar na Mega-Sena com apenas um jogo de 6 números?*

$$C(60/6) = 60! / ((60-6)! 6!) = 50.063.860$$

$$1 / 50.063.860 = 0,0000002\%$$

# DADOS

- **VARIÁVEL:** Característica medida em um indivíduo selecionado de um população em estudo. Ex.: altura, peso, mortalidade, etc.
- **OBSERVAÇÃO:** um dado obtido de uma variável  
Ex.: 13,3 kg, 1,67 m, etc.
- **DADOS:** conjunto de observações de uma ou mais variáveis realizadas em um conjunto de indivíduos.

# PARÂMETRO E ESTATÍSTICA

- **PARÂMETRO**: valor que descreve aspectos relacionados com a distribuição relativa da frequência de uma população.
- **ESTATÍSTICAS (ESTIMATIVAS DO PARÂMETRO)**: são os valores similares aos parâmetros, mas medidos na amostra.

# VARIÁVEIS

(Whitlock & Schluter, 2005)

## CATEGORIZADAS

**Dados categorizados são características dos indivíduos que não possuem magnitude na escala numérica.**

Ex.:

- Método de transmissão de doenças: ar, inseto, contato direto, água.
- Classe social: A, B, C, D, E.
- Tipo de sangue: A, B, O, AB.
- Severidade de ataque de pragas: nulo, baixo, médio alto e muito alto.

## NUMÉRICAS

**Dados numéricos são medidas quantitativas.**

Ex.:

- Temperatura corporal (graus centígrados)
- Tamanho do território (hectares)
- Remoção de carbono da atmosfera (ton de CO<sub>2</sub> /hectare).
- Idade das plantas em uma floresta nativa (anos)
- Número de habitantes por veículo na cidade de Piracicaba

# VARIÁVEIS CATEGORIZADAS

- São também chamadas de variáveis de atributo ou qualitativas.
- A medição de uma variável categorizada envolve a classificação de um indivíduo em um grupo.
- Uma variável categorizada é chamada de NOMINAL se os diferentes valores ou categorias não possuem uma ordem lógica. Ex.: método de transmissão de doenças, espécie de planta em uma floresta nativa (Amazônia). Nominal significa NOME ou seja, descreve um indivíduo de acordo com o grupo a que pertence.
- Por outro lado, uma variável categorizada ORDINAL pode ser ordenada mesmo sem ter a magnitude do valor numérico. Severidade do ataque de pragas florestais ou agrícolas, de nula a muito alta, é um exemplo de variável ordinal categorizada. Ordinal significa ter uma ordem.

# VARIÁVEIS NUMÉRICAS

- Como o nome diz, as observações das variáveis são números.
- As medições como contagens, dimensões, ângulos, taxas, e porcentagens são variáveis numéricas.
- As variáveis numéricas podem ser CONTÍNUAS ou DISCRETAS.
- **CONTÍNUAS**: Os valores das observações podem aumentar ou diminuir de quantidades infinitamente pequenas ou infinitamente grandes. Entre dois valores de uma variável contínua, um infinito número de outros valores podem ocorrer. Temperatura de uma estufa ou tamanho do território de um animal são exemplos.
- **DISCRETA**: Os valores numéricos são obtidos em unidades indivisíveis, como as contagens de animais em uma floresta. Número de animais vistos em um determinado dia é uma variável discreta, mas pode ser uma variável contínua se calculada como média diária em um conjunto de dias.
- **CUIDADO**: Muitas vezes o nome da espécie de árvore tomada em uma floresta é codificado em números para facilitar a coleta ou digitação. Embora sejam números, na realidade são nomes transformados em números e representam variáveis categorizadas e não numéricas.

# ESTUDOS CIENTÍFICOS

- Os estudos científicos podem ser divididos em dois tipos básicos:
- **Estudos observacionais:** coletam informações sobre os indivíduos como eles ocorrem na natureza, sem qualquer manipulação pelo pesquisador. Os estudos observacionais observam padrões e podem encontrar associação entre variáveis.
- Nos **estudos experimentais** um pesquisador poderá manipular uma ou mais variáveis do assunto estudado. Um estudo pode ser chamado de experimental se uma ou mais variáveis podem ser mudadas pelo pesquisador. Um estudo experimental devidamente delineado pode investigar causa e efeito.
- Por exemplo, a semente de *Tibouchina pulchra* germina a uma determinada temperatura e umidade. Se o pesquisador não manipula a temperatura, mas anota a temperatura e umidade e a taxa de germinação no campo, este estudo é de observação. Mas se as sementes são levadas ao laboratório e delineado um experimento para estudar causa (temperatura e umidade) e efeito (germinação) podemos estimar a temperatura e umidade adequadas para a germinação da *Tibouchina pulchra*.

# Tipos de Análises Estatísticas

- **Estatística Teórica (distribuição de dados).**
- **Estatística Descritiva e Gráficos.**
- **Análise de Variância e Teste de Comparação de Médias.**
- **Análise de Regressão.**
- **Análise Não-Paramétrica.**
- **Análise Multivariada.**
- **Outras Análises Estatísticas.**
- **Sem qualquer análise estatística.**

# PLATAFORMA DE COMPUTAÇÃO

- ANÁLISE ESTATÍSTICA NECESSITA DE FERRAMENTAS COMPUTACIONAIS
- MUITOS PACOTES ESTATÍSTICOS DISPONÍVEIS: R, SPSS, STATISTICA, JMP, SAS, EXCEL.
- TENTAÇÃO PARA PACOTES FÁCEIS DE USAR E APRENDER.
- ESCOLHER UM SISTEMA QUE NÃO TEM LIMITAÇÕES (LONGO TEMPO PARA FICAR OBSOLETO).

# O SISTEMA SAS

- STATISTICAL ANALYSIS SYSTEM
- COMEÇOU COMO UM SISTEMA DE ANÁLISE ESTATÍSTICA, MAS HOJE É MUITO MAIS.
- MUITOS ACHAM O SAS DIFÍCIL DE APRENDER (ORIENTADO À LINGUAGEM DE PROGRAMAÇÃO)
- PODE TAMBÉM SER USADO COM MENUS
- JMP É UM SAS COM MENU.
- VEJA: [www.sas.com](http://www.sas.com)

# ALGUMAS APLICAÇÕES DA ESTATÍSTICA NA ÁREA BIOLÓGICA

- **RESTAURAÇÃO DE ÁREAS DEGRADADAS:** espaçamento, adubação (mineral e organo-mineral), tipo de muda (tubete, tubetão, plântulas transplantadas), influência de ervas daninhas (aplicação de herbicidas, roçada, capina manual, uso de gel e irrigação, etc.), preparo de solo, espécies.
- **MICROBIOLOGIA:** estudo de produção de enzimas lignolíticas por fungos do solo e produção de etanol.
- **SAÚDE PÚBLICA:** desenvolvimento de vacinas e fármacos.
- **ECOLOGIA:** ciclagem e exportação de nutrientes pela retirada de árvores, impacto na conservação da fauna silvestre (ex. mineração, urbanização).
- **MUDANÇAS CLIMÁTICAS:** estudo de eventos climáticos extremos. Estimativa da quantidade de gases do efeito estufa sequestrados por árvores
- **PESQUISA DE OPINIÃO:** lançamento de produtos, opinião sobre políticos.

Quadro da análise de variância do crescimento do Ipê Roxo sob diferentes combinações de espécies pioneiras, com 6 anos de idade

Tratamento	Altura Média (m)	DAP Médio (cm)
A	4,8 <i>b</i>	5,4 <i>b</i>
B	4,7 <i>b</i>	5,2 <i>b</i>
C	6,1 <i>a</i>	8,1 <i>a</i>
D	6,2 <i>a</i>	7,6 <i>a</i>
E	2,1 <i>c</i>	3,4 <i>c</i>
Teste F	12,3**	9,6**
C.V. Exp. (%)	14,1	17,8
Média Geral	4,8	5,9

- A = 50% pioneira (*Trema sp*)  
 B = 75% pioneira (*Trema sp*)  
 C = 50% pioneira (*Tibouchina sp*)  
 D = 75% pioneira (*Tibouchina sp*)  
 E = Plantio puro de Ipê Roxo

Nota: médias seguidas da mesma letra não diferem entre si pelo teste Tukey a 5% de probabilidade

# Exercício para a próxima aula

- Um pesquisador deseja verificar quais os tipos de análise estatística que são mais usados na sua área de estudo e resolve tomar 15 trabalhos *científicos* de um periódico *científico* e analisar os métodos estatísticos utilizados. No final produza um relatório técnico para publicação interna de sua instituição de pesquisa.

# Periódicos (escolha um)

- Horticulture
- Limnology
- Aquatic Ecology
- Plant Pathology
- Journal of Applied Ecology
- Plant Ecology
- Forest Products Journal
- Journal of Animal Ecology
- Enzyme Research
- Epidemiology and Infection
- Animal Science Journal
- Journal of Mammalogy
- Forest Ecology and Management

*OBRIGADO !!!*

*A T É A*

*P R Ó X I M A*

*!!!*