

Introdução (RÁPIDA) ao SAS

O que é o SAS?

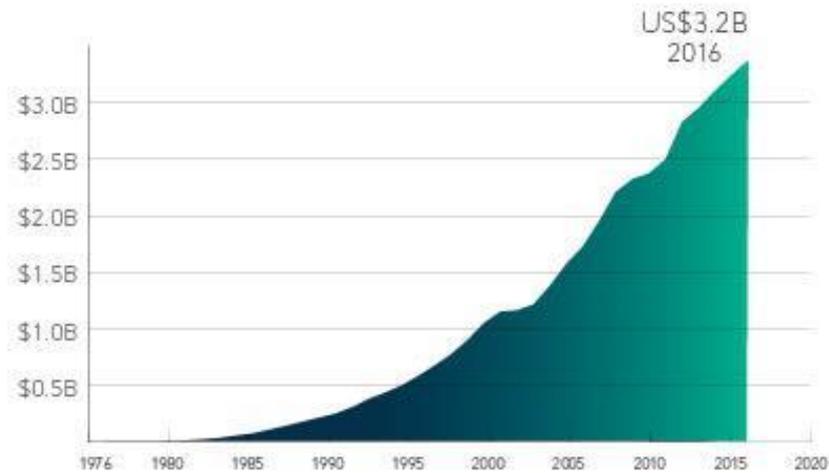
- Desenvolvido no início da década de 70 na North Carolina State University, Raleigh, NC
- No início, o objetivo era analisar estatisticamente experimentos agrícolas e para uso em investigação científica.
- É o pacote estatístico mais usado no mundo (mais de 70.000 organizações).
- SAS – Statistical Analysis System
- James Goodnight (137º mais rico do mundo: patrimônio de 10,4 bilhões de dólares, em janeiro de 2018, segundo a Revista Fortune).
- Mais de 14.000 empregados em 148 países.
- 94 das 100 maiores companhias no mundo usam SAS.
- Todas as 100 maiores universidades no mundo usam SAS (QS Top Universities).

SAS



James Goodnight

Annual Revenue 1976 – 2016



Continuous revenue growth since 1976

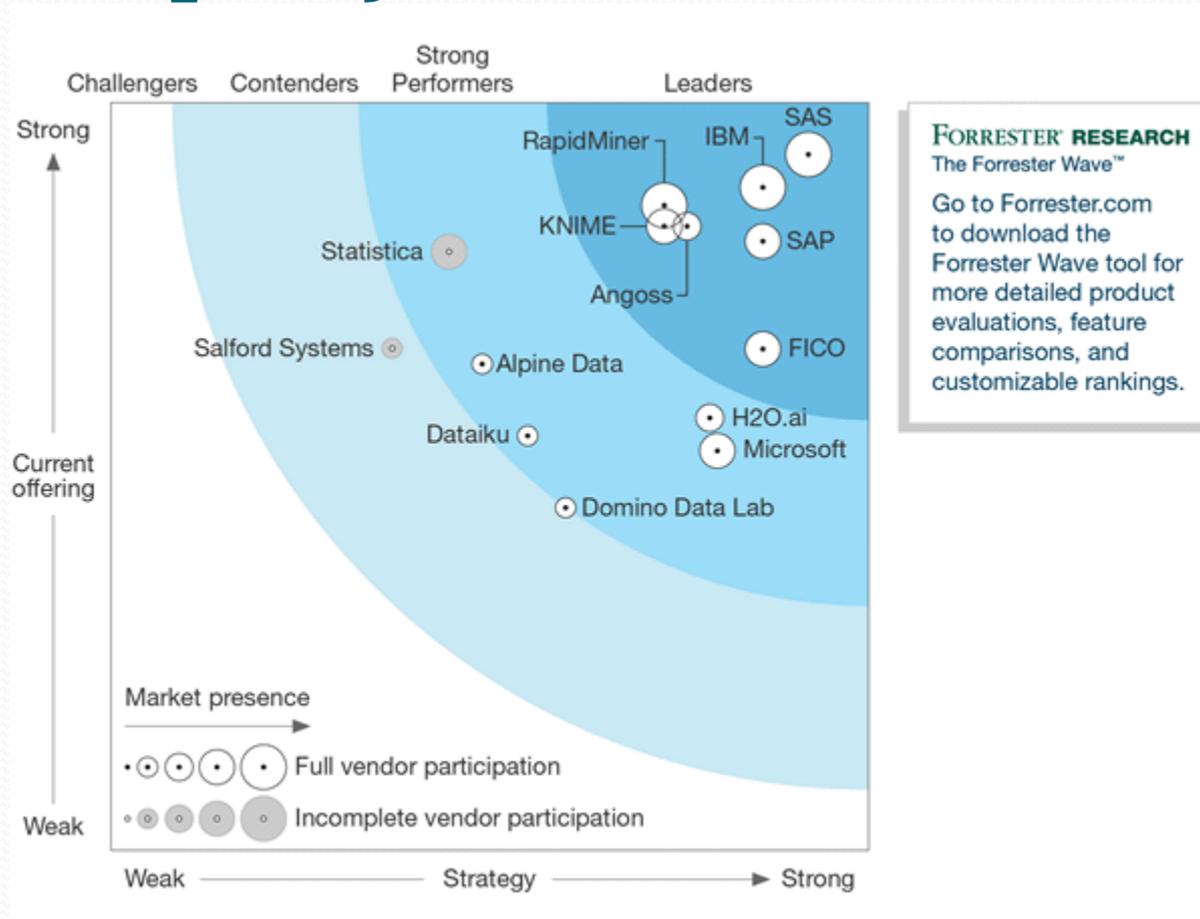
Record revenue of \$3.2 billion up 4 percent constant currency (1.3 percent USD) over 2015.



Liderança do SAS

- Especialista em análise de grande quantidade de dados (data mining).
- Dados meteorológicos: mais de 20.000 estações no mundo coletando dados a cada 15 minutos.
- Analítica: ciência da análise.
- Primeira empresa em analítica preditiva e mineração de dados, entre as 15 maiores empresas mundiais (Forrester, 2017).
- Em 2008 adquiriu a empresa TERAGRAM, especializada em **mineração de texto** (clientes: Yahoo!, Banco Mundial, CNN, Forbes, NY Times, Washington Post, etc.)
- Em 2010 adquiriu a empresa Assetlink, líder no gerenciamento integrado de mercado (marketing).

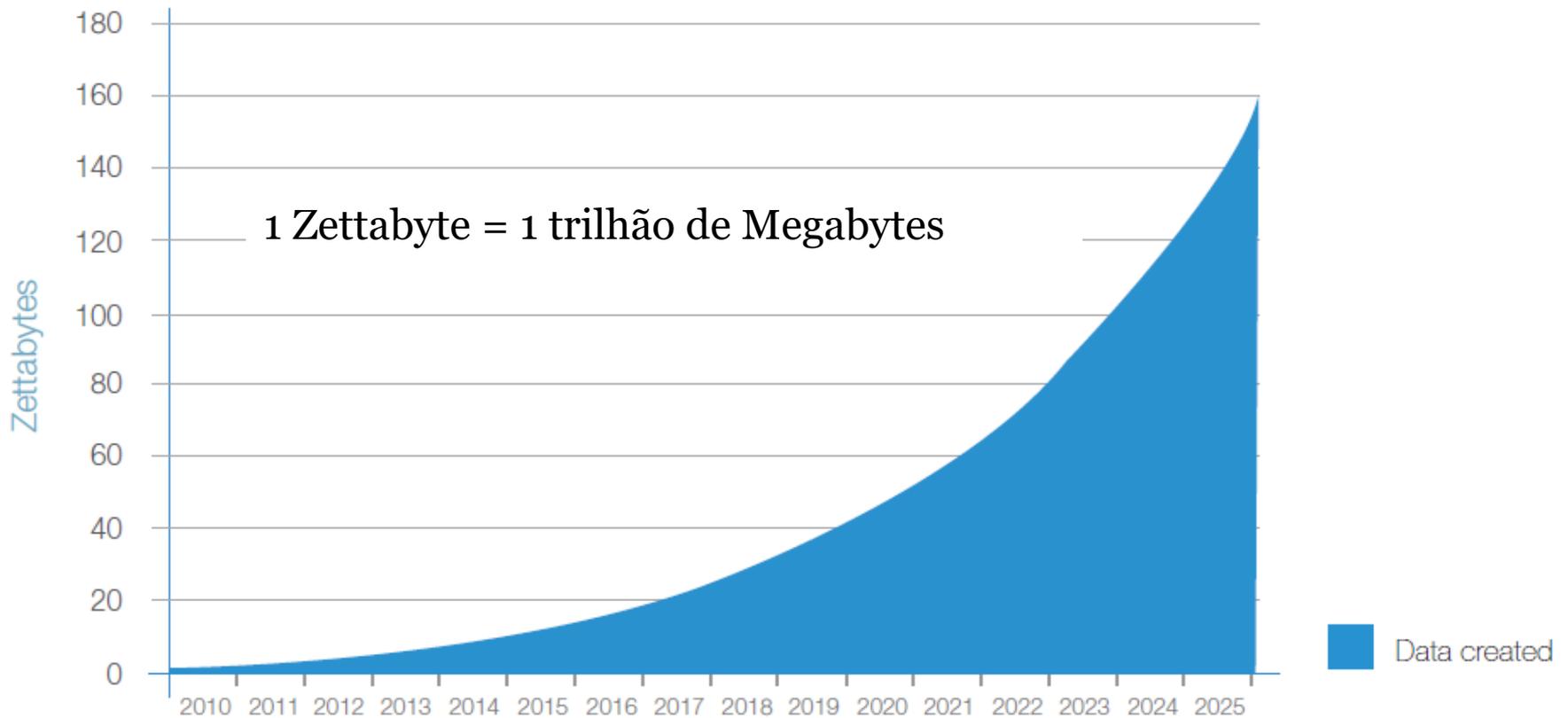
Analítica preditiva e Mineração de dados: posição dos sistemas (2017).



129452

Source: Forrester Research, Inc. Unauthorized reproduction, citation, or distribution prohibited.

Produção mundial de dados.



Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

Quantidade de dados analisados.



Rede de estações meteorológicas sob a supervisão da WMO.

- Mais de 11.000 na superfície terrestre.
- Mais de 1.300 estações aéreas fixas.
- Mais de 4.000 navios com estações meteorológicas.
- Mais de 1.200 bóias marítimas.
- Centenas de radares meteorológicos (976 em jan 18).
- Mais de 3.000 aeronaves comerciais equipadas com sistemas de monitoramento climático.

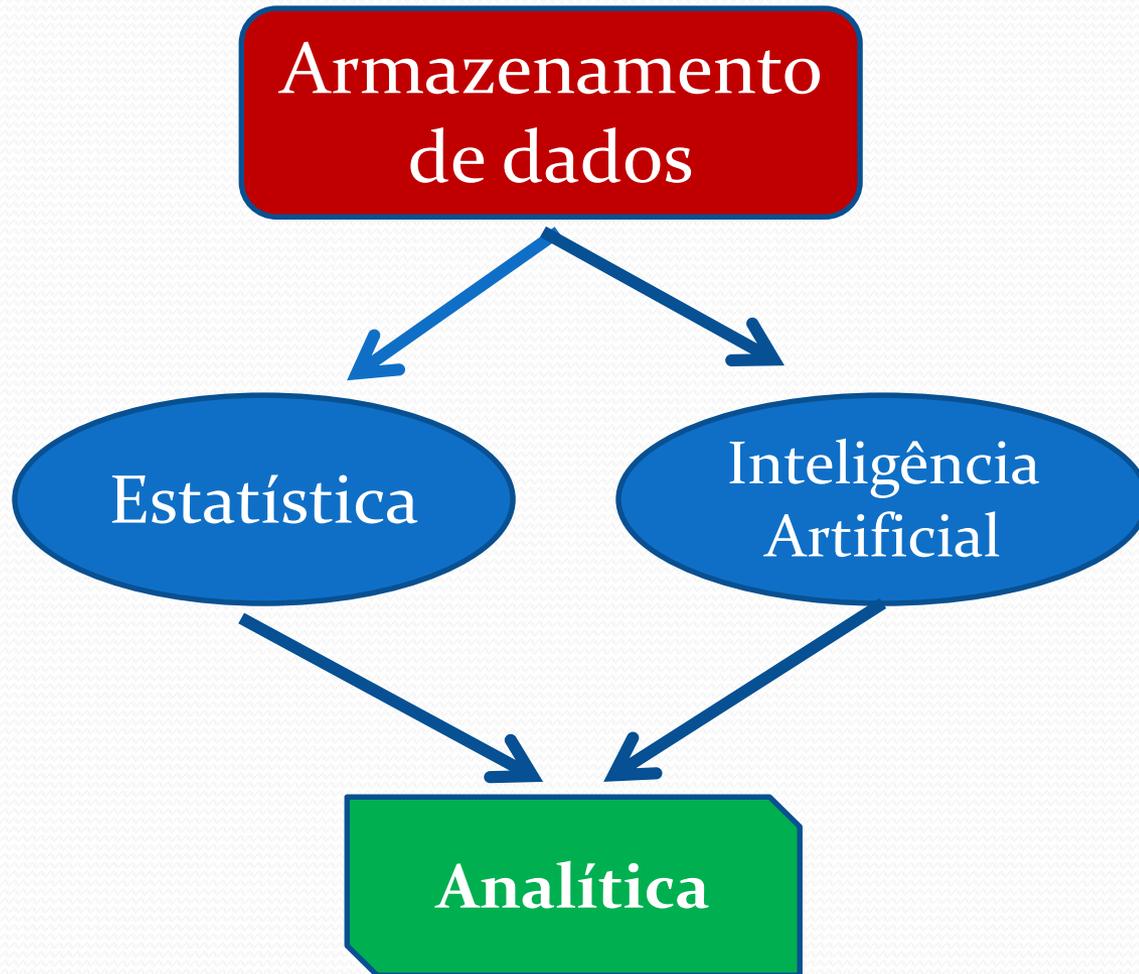
O que é analítica?

- Em 2007 a Universidade Estadual da Carolina do Norte fundou o Instituto de Analítica Avançada com o objetivo de estudar analítica. Analítica compreende desde a coleta e integração de dados, métodos estatísticos e processos complexos de otimização e modelagem para a tomada de decisões. É o primeiro mestrado em analítica nos Estados Unidos.
- Outra definição: parte da lógica que estuda análise.

O que é Analítica sob o ponto de vista prático?

- Processo de extração de conhecimento de grandes bases de dados convencionais ou não.
- O objetivo é encontrar automaticamente padrões, anomalias e regras, com o propósito de transformar dados em informações úteis para a tomada de decisões e/ou avaliação de resultados.

Áreas envolvidas



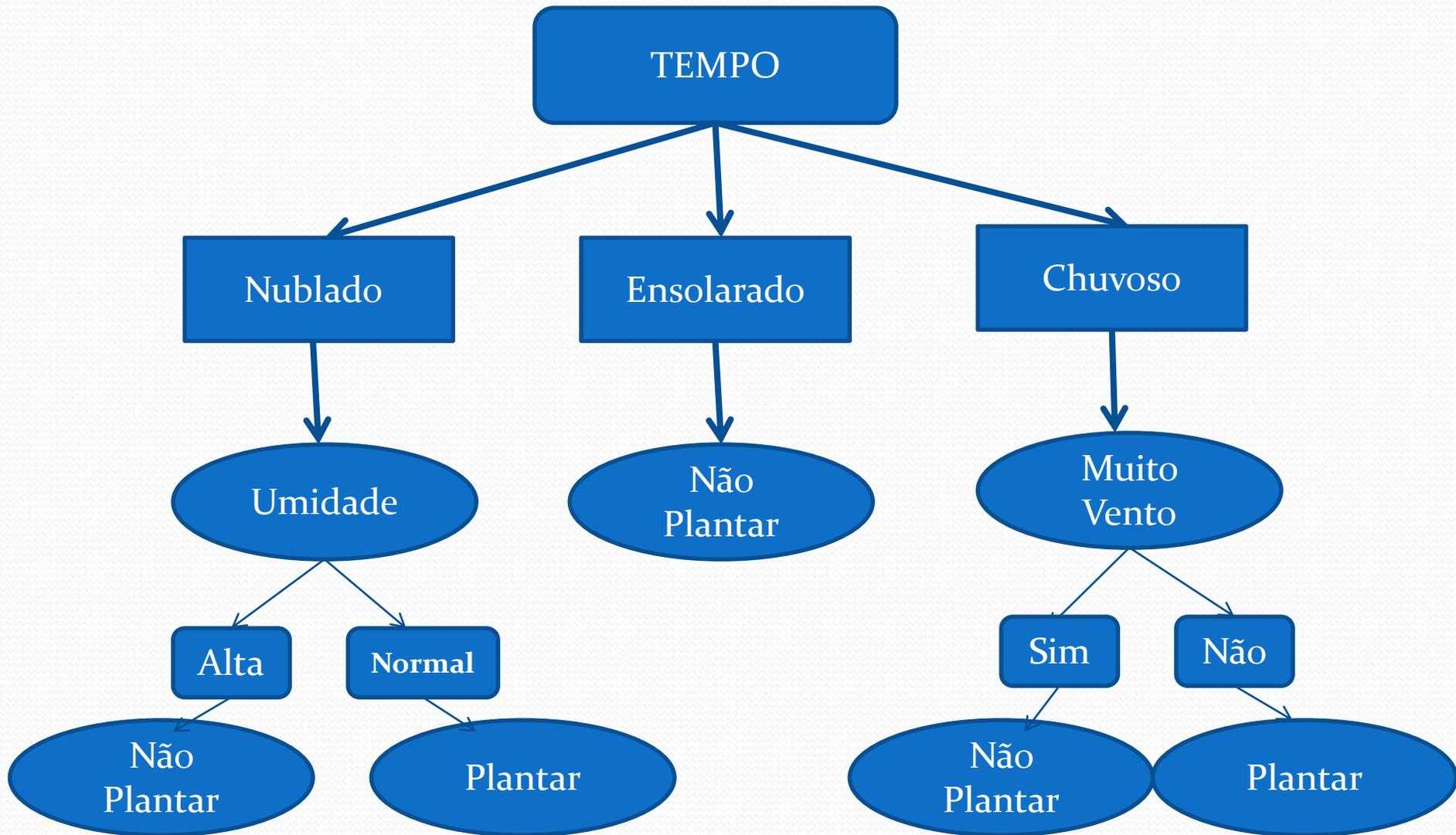
Inteligência Artificial

- Ciência e engenharia usada para fazer as máquinas inteligentes, usando programas computacionais.
- Ex.: fazer um programa de computador para uma máquina jogar xadrez.

Tarefas realizadas pela Analítica

- Modelagem preditiva (classificação e regressão).
- Segmentação (análise de agrupamento).
- Sumarização (média, variância, coeficiente de variação).
- Visualização (análise gráfica).

Árvore de Decisão de Plantio de Espécies Arbóreas Nativas



Exemplos de Aplicação

- **Bancos:** Selecionar clientes para estabelecer limites no cartão de crédito.
- **Medicina:** Um hospital nos Estados Unidos através da análise do banco de dados dos pacientes conseguiu economizar 300 dólares por paciente por ano direcionando o paciente para o atendimento.
- **Empresa produtora de cigarro:** Definir fornecedores que devem ser financiados através de estudos de índices de fidelidade, qualidade do produto, volume entregue.
- **Política:** a campanha do Presidente Obama usou os dados históricos de cada município para a alocação de recursos na campanha em 2012 e previu que o candidato teria uma votação de 56,5% e o resultado final foi de 56,6%.

Produtos SAS

- Base – SAS – procedimentos básicos para o manejo de dados.
- SAS/Stat: análise estatística
- SAS/Graph: apresentação gráfica de qualidade.
- SAS/OR : pesquisa operacional
- SAS/ETS: econometria e séries temporais.
- SAS/IML: linguagem matricial interativa
- SAS/AF: para desenvolvimento de menus e interfaces.
- SAS/QC: controle de qualidade

Documentação

- <http://www.sas.com>
- <http://support.sas.com/documentation/93/index.html> : manuais da versão 9.3 (em 2011 o SAS lançou a versão 9.3)
- <http://support.sas.com/documentation/>: documentação geral do SAS para todas as versões.

Estrutura básica do SAS

- 3 componentes principais na maioria dos programas:
 - Comandos DATA
 - DADOS
 - Comandos PROC (procedure).

Comandos DATA

- Lê os dados de fontes internas ou externas, manipulam os dados e combinam com outros dados.
- Se presta para preparar os dados para serem usados por um ou mais procedimentos.

- Exemplo:

- ```
DATA SOLO;
INPUT LOCAL $ PONTO PROF MO;
C=MO*0.58; /* Transforma matéria orgânica em
carbono*/
CO2 = C*44/12; /* Transforma carbono em CO2, gás do
efeito estufa */
DATALINES;
```

.....

# Comandos PROC

- Realiza as análises dos dados e produz a listagem na janela OUTPUT.
- Exemplos de Proc:
  - PROC PRINT
  - PROC ANOVA
  - PROC SORT
  - PROC MEANS
  - PROC UNIVARIATE
  - PROC GLM

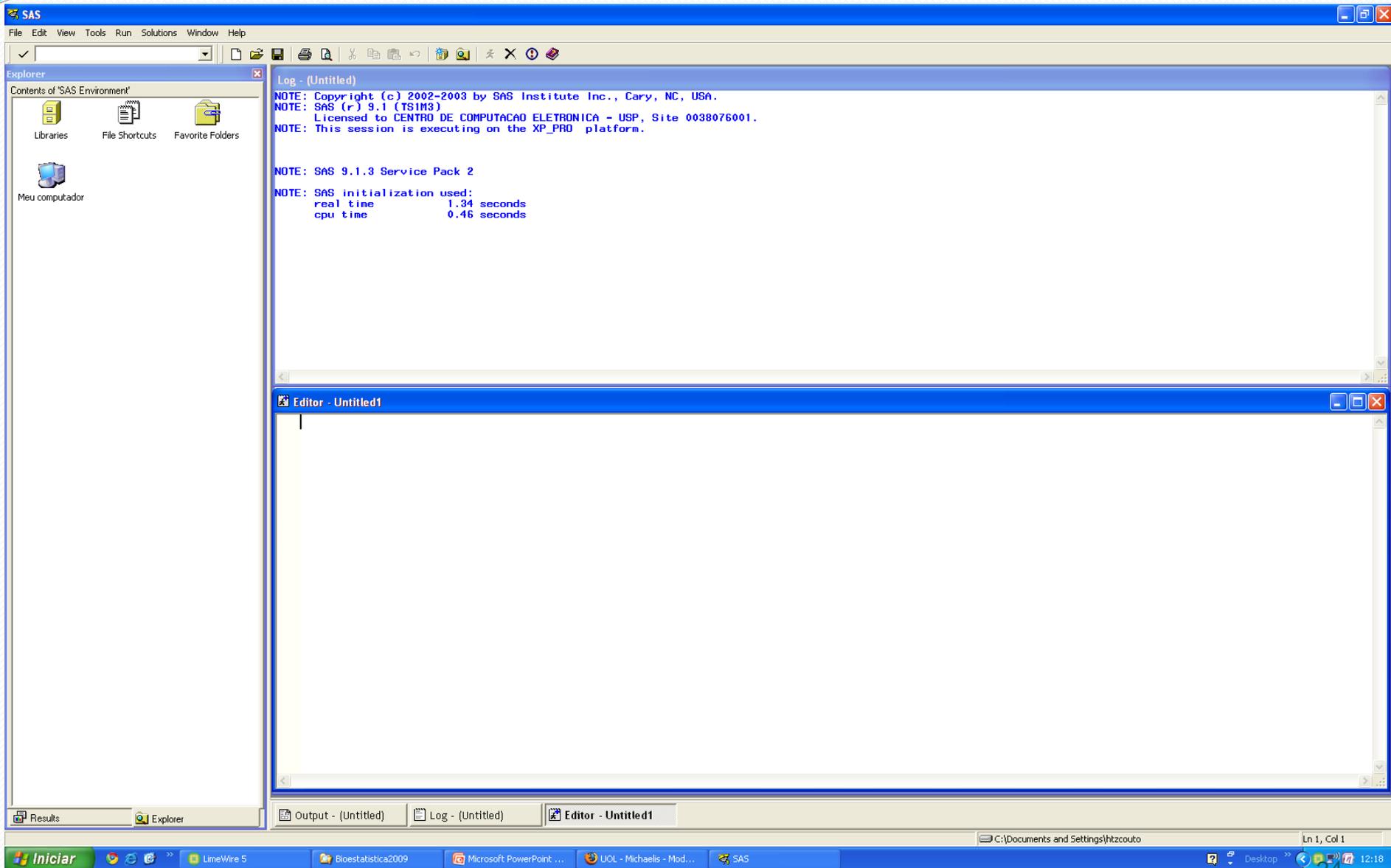
# Conceitos e Regras do SAS

- O nome SAS deve ter no máximo 32 caracteres, composto de letras, números e do caractere sublinhado (  )
- Não deve começar por número.
- Não deve ter espaços em branco.
- Cada sentença SAS deve terminar com o ponto e vírgula (;)
- Linhas que começam com \* e terminam com o ; são tratadas como comentário.
- Os comentários também podem ser colocados entre /\* e \*/.
- RUN; é o comando usado para executar o programa.

# Janelas

- 3 janelas principais
- EDITOR usada para criar, editar e executar um programa SAS.
- LOG relata o progresso do programa SAS, mostra os erros (vermelho) e os avisos (verde).
- OUTPUT mostra a saída de um programa SAS. Pode-se copiar, salvar ou imprimir a listagem do programa SAS.

# Principais janelas do SAS



# Entrada de dados para análise

- 3 modos de entrada de dados no SAS:
  - Ler dados digitados ou colados na janela EDITOR
  - Importar dados do Excel
  - Outros procedimentos, como importar arquivos .txt, Access, etc.

## Ler dados digitados ou colados no EDITOR

- DATA ALUNO;
- INPUT NOME \$ SEXO \$ ALTURA;
- DATALINES;
- JOÃO M 1.85
- MARIA F 1.65
- PEDRO M 1.72
- JOANA F 1.71
- ;
- PROC PRINT DATA=ALUNO; RUN;

# Importar dados do Excel

- PROC IMPORT OUT=WORK.Ao
- DATAFILE='D:\BIOESTATISTICA\DADOS.XLSX'
- DBMS=XLSX REPLACE;
- SHEET='PLAN1';
- RUN;

# Saída SAS

- A listagem (resultados das análises) pode ser gravada em HTML (**H**yper**T**ext**M**arkup**L**anguage); PS (**P**ost**S**cript); RTF (**R**ich**T**ext**F**ormat – pode salvar e editar no Microsoft WORD) e PDF (**P**ortable**D**ocument**F**ormat), entre outros formatos.
- Na janela OUTPUT do SAS: o conteúdo pode ser copiado e colado, por exemplo no Microsoft WORD, ou impresso diretamente.

# Programa SAS para gravar arquivo em PDF (Adobe Acrobat)

```
PROC IMPORT OUT=WORK.A0
DATAFILE='D:\BIOESTATISTICA\DADOS.XLSX'
DBMS=XLSX REPLACE;
SHEET='PLAN1';
RUN;
ODS PDF FILE='C:\RESULTADO.PDF';
PROC PRINT DATA=A0;
RUN;
ODS PDF CLOSE;
```

# Alguns comandos SAS

- LENGTH
- IF... THEN...;
- DELETE
- AND OR
- EQ, NE, GE, LE, GT, LT

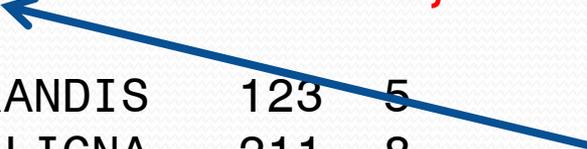
# Programa SAS:

## Uso do comando LENGTH

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Uso de comando condicional: IF ... THEN ...;

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF LOCAL ='RCLARO' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```



Caractere  
entre  
aspas

# Uso de comando condicional: IF ... THEN ...:

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF IDADE=7 THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

Número sem aspas

# Uso de comando condicional: AND (E)

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF IDADE=7 AND LOCAL='BAURÚ' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Uso de comando condicional: OR (OU)

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF LOCAL='LENÇÓIS' OR LOCAL='BAURÚ' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇÓIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Comandos de comparação

- EQ = IGUAL A
- NE = DIFERENTE DE
- GE = MAIOR OU IGUAL A
- LE = MENOR OU IGUAL A
- GT = MAIOR QUE
- LT = MENOR QUE

# Programa SAS: modificação dos dados.

```
DATA SOBREV;
 INPUT ESP $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: EQ (Igual a )

```
DATA SOBREV;
 INPUT ESP $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF TAXAMORT EQ 0 THEN DELETE;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: NE (Diferente de)

```
DATA SOBREV;
 INPUT ESP $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF ESP NE 'A02' THEN DELETE;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: GE (Maior ou igual a)

```
DATA SOBREV;
 INPUT ESP $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF IDADE GE 5 ;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: LE (Menor ou igual a)

```
DATA SOBREV;
 INPUT ESP $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF IDADE LE 2 THEN VIVOS=25 ;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Exercício

- A) Fazer um programa SAS para selecionar apenas as idades maiores que 4 (Use o comando GT).
- B) Fazer um programa SAS para selecionar apenas as idades cujos números de sobreviventes sejam menores que 15 (Use o comando LT).

Obrigado e até a  
próxima semana!!!