

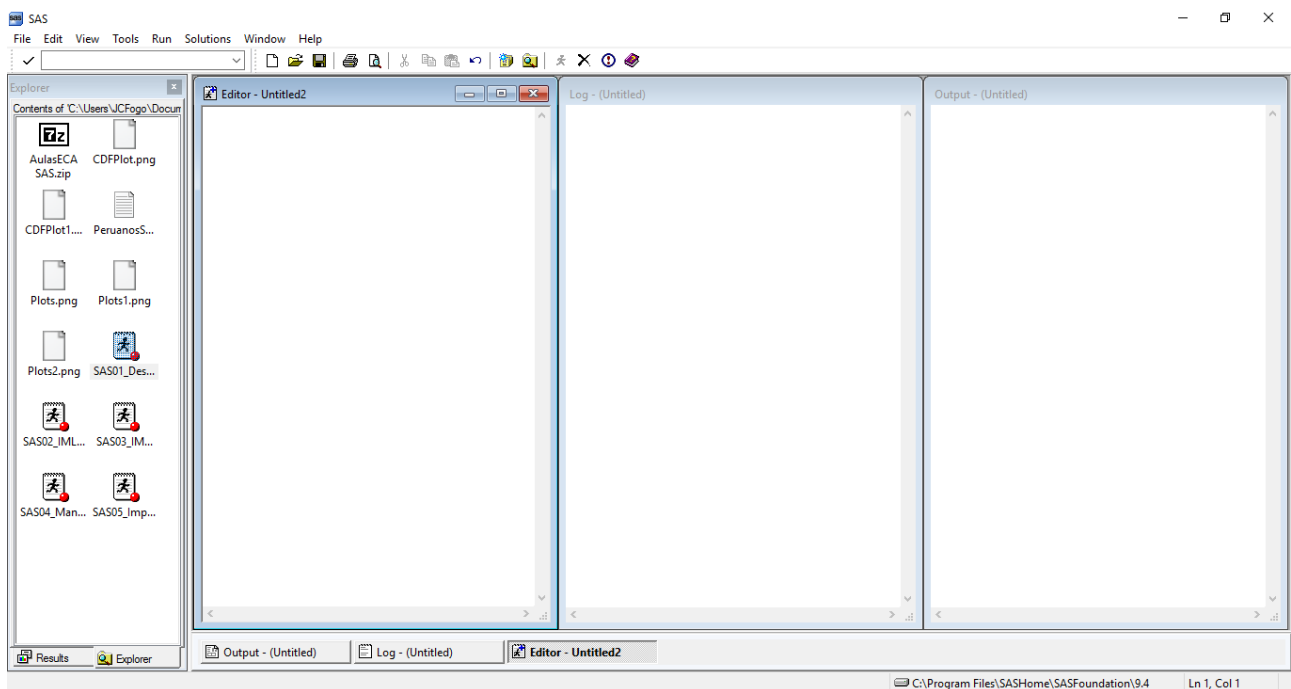
# O Sistema SAS<sup>1</sup>

## I - O ambiente SAS:

O Sistema SAS é um sistema integrado de software que permite o desenvolvimento de entradas, recuperação e manipulação de dados, análises estatísticas e matemáticas, relatórios, entre outros. Seu ambiente é formado por três janelas principais:

- a) **EDITOR**: editor de programas;
- b) **LOG**: janela de avisos e mensagens de erros durante a execução dos programas;
- c) **OUTPUT**: janela com saídas de programas;

mais uma janela auxiliar com o explorador **Explorer** e **Results** (ver figura).



**Figura 1:** Janela de abertura do SAS, versão 9.4.

Logo abaixo da linha de menu, existe uma *caixa de comando* para a execução direta de comandos específicos, válidos sempre para a janela ativa.



## II - Criando um programa SAS:

Os programas SAS podem ser criados através do editor SAS (janela **EDITOR**) ou de qualquer outro editor de texto para arquivos ASCII.

Um programa SAS é formado basicamente por dois módulos:

- i) **DATA**: destina-se à declaração e manipulação dos dados, aceitando comandos de programação.
- ii) **PROC**: com procedimentos para análise de dados, gráficos e relatórios.

<code>data Exemplo;</code>	→ identifica o conjunto de dados
<code>input var1 var2 ... vark ;</code>	→ declaração das variáveis
<code>cards;</code>	→ identifica o início do conjunto de dados
.....	
<code>{ dados</code>	→ entrada de dados
.....	
<code>;</code>	→ o sinal “;” indica o final dos dados.
.....	
<code>{ procedures ou proc's</code>	→ procedimentos para análise.
.....	

**Quadro 1:** Esquematização de um programa SAS.

Com a sequência acima é criado um conjunto de dados com o nome *Exemplo* que é identificado por SAS *dataset*, o qual será analisado com os procedimentos definidos a partir do final dos dados.

- Obs:**
- 1) Cada linha de comando deve ser encerrada por um ponto-e-vírgula “;”.
  - 2) Para os dados, nenhum sinal deve ser usado para indicar o fim da linha.
  - 3) Um único ponto-e-vírgula deve ser colocado após a última observação, indicando o final dos dados.

### Exemplo de um programa SAS:

```
/* Peso (kg) de coelhos híbridos Norfolk, abatidos após os 90 dias de vida.
   Variáveis: peso(kg) e idade(sem) */
```

```
data coelhos;
input peso idade;
cards;
```

```
2.50 14.9
2.58 16.2
2.60 15.4
2.62 15.9
2.65 16.8
2.66 16.1
2.58 14.7
2.70 17.3
2.55 14.5
2.57 14.3
2.70 17.7
2.62 16.0
2.59 16.3
2.54 15.0
2.53 15.5
2.20 11.8
;
```

```
/* Procedure para cálculo das estatísticas descritivas */
proc means data=coelhos;
var peso idade;
run;
```

### Regras para nomes no SAS:

i) Os nomes no SAS são usados para denominar conjuntos de dados, variáveis e outros itens. O 1º caracter deve ser uma letra ou um “*underscore*” ( \_ ). Brancos não podem ocorrer em nomes SAS.

ii) Variáveis alfanuméricas são identificadas pelo do símbolo " \$ " após a declaração do nome.

No exemplo:

```
INPUT nome $ sexo $ idade altura peso
```

as variáveis *nome* e *sexo* são alfanuméricas.

### III - Lendo um arquivo de dados externo:

Acrescentando o comando INFILE antes do INPUT, podemos usar um arquivo de dados externo.


<pre>DATA Exemplo;   INFILE 'C:\...\caminho...\nome.ext';   INPUT var1 var2; ..... {procedimentos .....</pre>
---

**Obs:** 1) Note que agora não precisamos do comando CARDS.

2) O arquivo *nome.ext* deve estar no formato ASCII, sem os nomes das variáveis, ou seja, deve conter apenas as observações.

3) Também podemos utilizar um arquivo de dados através do comando INCLUDE.

### IV - Executando um programa SAS:

Marcar as linhas, ou bloco, de comandos e clicar no botão *Submit*  (ou pressionar a tecla F8).

Ao resultados dos comandos executados são, automaticamente, apresentados na janela **OUTPUT**, desde que não haja nenhum erro.

**Obs:** 1) Pode-se ainda, executar o comando *Submit* na caixa de comandos, com a janela **EDITOR** ativa.

2) Pode-se executar o programa todo, ou partes deste, desde que sejam devidamente marcadas.

## V - Principais comandos do SAS:

No Sistema SAS existem alguns *comandos específicos* que podem ser utilizados diretamente na *caixa de comandos* ou mesmo através de teclas de atalho (*F-Keys*). Alguns comandos são relacionados abaixo:

Comando	Tecla de atalho (F-Key)	Descrição
HELP	F1	Aciona a janela <b>HELP</b> do SAS
WPGM	F5	Move o cursor para a janela <b>EDITOR</b>
LOG	F6	Move o cursor para a janela <b>LOG</b>
OUTPUT	F7	Move o cursor para a janela <b>OUTPUT</b>
ZOOM OFF; SUBMIT	F8	Restaura a janela ativa ao tamanho original; Executa os comandos selecionados quando a janela <b>EDITOR</b> estiver ativa
KEYS	F9	Edita as teclas de atalho <sup>2</sup>

<sup>2</sup> Podemos editar as teclas *F-keys* e reconfigurá-las.

Existem, ainda, outros comandos, dos quais destacamos:

- CLEAR: limpa a janela ativa (ou ctrl+E).
- TILE: organiza lado-a-lado as janelas **EDITOR**, **LOG** e **OUTPUT**.
- NEXT: move o cursor para a janela seguinte.
- BYE / ENDSAS: encerra o programa SAS.
- FILE: grava o conteúdo de uma janela ativa num arquivo.
- INCLUDE: abre um programa SAS na janela **EDITOR** ou inclui um conjunto de dados no corpo de um programa.
- LINESIZE =  $n$  (ou LS =  $n$ ):  $64 \leq n \leq 132$ , define o comprimento da linha\*.  
default LS = 132.
- PAGESIZE =  $n$  (ou PS =  $n$ ): especifica o número de linhas da página\*.  
default PS = 60.
- MISSING = 'caracter': especifica 'caracter' como valor perdido (*missing value*).

\* Os comandos PS e LS devem aparecer de preferência no início do programa.

## VI - Principais procedimentos do SAS:

Todo procedimento SAS vem identificado pela palavra PROC e pode apresentar diversas opções e subcomandos específicos. A estrutura dos PROC's são basicamente a mesma para todos, conforme descrição abaixo:

Sintaxe: **PROC NOME** opções;  
          <declarações>;

## VI.I - A seguir, apresentaremos os principais procedimentos do SAS:

### A) Procedimentos de estatísticas descritivas:

- PROC CORR** – fornece correlações e medidas de associação.
- PROC MEANS** – estatísticas descritivas.
- PROC UNIVARIATE** – estatísticas descritivas, incluindo quantis e plots descritivos.
- PROC CHART** – histogramas e gráficos.
- PROC FREQ** – faz distribuições de frequência para variáveis categóricas e tabelas cruzadas.
- PROC TABULATE** – tabelas com estatísticas descritivas.
- PROC TTEST** – compara médias de dois grupos de variáveis através do teste t.

### B) Procedimentos para impressão:

- PROC PRINT** – mostra o conteúdo de *SASdataset* na janela **OUTPUT**.
- PROC FORMS** – formas padrões, por exemplo mala direta.
- PROC PLOT** – para construir gráficos.
- PROC TIMEPLOT** – para gráficos de uma ou mais variáveis na ordem temporal.

### C) Procedimentos de regressão e análise de variância:

- PROC REG** – ajusta uma regressão linear com várias opções de diagnósticos e diversos métodos para seleção de modelos.
- PROC NLIN** – ajusta modelos de regressão não lineares.
- PROC ANOVA** – para análise de variância, análise de variância multivariada, análise de variância com medidas repetidas em experimentos balanceados, comparações múltiplas.
- PROC GLM** – ajusta modelos lineares gerais (regressão simples, múltipla, polinomial e ponderada, faz análise de variância para dados desbalanceados).
- PROC VARCOMP** – para componentes de variância.
- PROC LIFEREG** – ajusta modelos paramétricos para dados de tempos de vida.

### D) Procedimentos para dados multivariados:

- PROC PRINCOMP** – faz análise de componentes principais.
- PROC CANCORR** – faz análise de correlação canônica.
- PROC FACTOR** – faz análise de fatores ortogonais com rotações.
- PROC CLUSTER** – faz análise de cluster (ou agrupamentos).

### E) Procedimentos não paramétricos e para dados categóricos:

- PROC NPAR1WAY** – análise não paramétrica one-way para ranks.
- PROC CATMOD** – para tabelas de contingência.
- PROC LOGIST** – ajusta modelos logísticos.
- PROC PROBIT** – faz regressão probit, logística e logística ordinal.

### F) Procedimentos Utilitários:

- PROC APPEND** – anexa dados de um data set no fim de um outro data set.
- PROC COMPARE** – compara os conteúdos de dois data set's.
- PROC FORMAT** – define formatos de saída e de leitura de dados.
- PROC SORT** – ordena os dados de um *SASdataset*.

**VII - O procedimento PROC UNIVARIATE:** Estatísticas descritivas, incluindo quantis e plots descritivos para variáveis numéricas.

*Sintaxe:* PROC UNIVARIATE *opções*;

```

var    <variáveis>;
by     <variáveis>;
freq   <variáveis>;
weight <variáveis>;
id     <variáveis>;
output out=SASdataset Keyword=<nomes>;

```

onde: 1) Opções:

**DATA**=*SASdataset* – especifica o conjunto de dados.

**NOPRINT** – não mostra as saídas.

**PLOT** – opções de gráficos (ramos-folhas, box-plot, normal-plot).

**FREQ** – faz tabela de frequência.

**NORMAL** – faz teste de normalidade dos dados,  
                   se  $n < 51$  faz o teste de Shapiro-Wilks,  
                   se  $n \geq 51$  faz o teste de Kolmogorov-Sirnov.

**PCTLDEF**= $k$  – especifica o método para o cálculo dos quartis,  
                    $k = 1, 2, 3, 4$  ou  $5$  (default  $k = 5$ ).

**VARDEF**= $df/weight/n/wdf$  – especifica o divisor para o cálculo da variância (default = $df$ ),  
                   onde:

$df$  = graus de liberdade,  
 $weight$  = soma dos pesos,  
 $n$  = número de observações,  
 $wdf$  = soma dos pesos menos 1 ( $w-1$ ).

**ROUND**= $n$  – especifica as unidades para arredondamento

2) Descrição das declarações:

**var** – lista de variáveis para o PROC UNIVARIATE (default = *todas*).

\* usar sempre após uma declaração de output.

**by** – separa as análises em subgrupos definidos pela variável declarada.

\* os dados devem estar ordenados em ordem crescente pela variável em questão.

**freq** – considera os valores da variável como frequências.

\* se os dados não são inteiros, os valores são truncados.

**weight** – variável com pesos para o cálculo da média e variância ponderadas ( $\bar{x}_W$  e  $s_W^2$ ).

\* – valores negativos são considerados como zero.

– só para calcular  $\bar{x}_W$  e  $s_W^2$ .

**id** – usado para identificar as observações (ou casos).

\* Considera os 8 primeiros caracteres.

**output** – requisita um arquivo interno de saídas.

**out** = *SASdataset* – define o nome do arquivo interno de saídas.

**keyword** = *nomes* – identifica as estatísticas desejadas.

\* o comando OUTPUT grava o conteúdo da memória RAM num arquivo *SASdataset*.  
 Gravar as saídas num *SASdataset* não é o mesmo que salvar num arquivo definitivo,  
 o que pode ser obtido através do comando FILE. Os conteúdos dos *SASdataset* se  
 perdem quando ao sair do SAS.

Os nomes definidos através da opção **keyword** podem ser:

N – número de observações consideradas para os cálculos (menos MISSING).  
 NMISS – número de valores perdidos.  
 NOBS – número total de observações.  
 MEAN – média aritmética.  
 SUM – soma das observações.  
 STD – desvio padrão.  
 VAR – variância.  
 SKEWNESS – medida de simetria.  
 KURTOSIS – curtosi.  
 SUMWGT – soma dos pesos.  
 MAX – valor máximo.  
 MIN – valor mínimo.  
 RANGE – amplitude amostral (MAX-MIN).  
 Q3 – terceiro quartil.  
 MEDIAN – mediana.  
 Q1 – primeiro quartil.  
 QRANGE – amplitude interquartil (Q3-Q1).  
 P1 – 1<sup>o</sup> percentil (quantil 0.01).  
 P5 – 5<sup>o</sup> percentil (quantil 0.05).  
 P10 – 10<sup>o</sup> percentil (quantil 0.10).  
 P90 – 90<sup>o</sup> percentil (quantil 0.90).  
 P95 – 95<sup>o</sup> percentil (quantil 0.95).  
 P99 – 99<sup>o</sup> percentil (quantil 0.99).  
 MODE – moda (se houver mais do que uma moda, é usada a menor delas).

### VIII - O procedimento PROC MEANS: Estatísticas descritivas para variáveis numéricas.

Sintaxe: PROC MEANS *opções* **keyword**=<nomes>;

```

var    <variáveis>;
by     <variáveis>;
class  <variáveis>;
freq   <variáveis>;
weight <variáveis>;
id     <variáveis>;
output out=SASdataset;
  
```

#### Principais opções:

DATA=SASdataset – especifica o conjunto de dados.  
 NOPRINT – não mostra as saídas.  
 MAXDEC – número de casas decimais para os resultados (0 a 8).  
 FW – tamanho do campo para impressão.

#### Declarações:

**keyword** = *nomes* – identifica as estatísticas desejadas.  
**var** – lista de variáveis para o PROC MEANS (default = *todas*).  
**by** – especifica subgrupos.  
**class** – especifica as variáveis que definem subgrupos.  
**freq** – considera os valores da variável como frequências.  
**weight** – variável com pesos para o cálculo da média e variância ponderadas.  
**id** – usado para identificar as observações (casos).  
**output** – requisita um arquivo interno de saídas.  
**out** = SASdataset – define o nome do arquivo interno de saídas.

**Palavras chaves para estatísticas descritivas:**

N	número de observações
NMISS	número de valores perdidos
SUM	soma
SUMWGT	soma dos pesos (= $n$ para dados não ponderados)
CSS	soma de quadrados corrigida pela média
USS	soma de quadrados
MIN	valor mínimo da mostra
MAX	valor máximo da amostra
RANGE	amplitude dos dados
MEAN	média amostral
VAR	variância amostral
STDDEV STD	desvio padrão amostral
CV	coeficiente de variação
STDERR	erro padrão da média
SKEWNESS SKEW	simetria da amostra
KURTOSIS KURT	curtose da amostra
CLM	interval de confiança
LCLM	limite inferior do interval de confiança
UCLM	limite superior do interval de confiança

**Palavras chaves para quantis:**

MEDIAN P50	mediana
Q1 P25	1o. quartil
Q3 P75	3o. quartil
QRANGE	amplitude interquartil
P1	percentil 0.01 (1%)
P5	percentil 0.05 (5%)
P10	percentil 0.10 (10%)
P90	percentil 0.90 (90%)
P95	percentil 0.95 (95%)
P99	percentil 0.99 (99%)

**Palavras chaves para testar a hipótese  $H_0: \mu = 0$** 

T	valor $t$ observado
PROBT	$2P(T_{n-1} >  t_0 )$ .

**Default:** N, MEAN, STD, MIN, and MAX

**IX - O procedimento PROC PRINT:** Imprime os valores das variáveis de um arquivo de dados na janela **OUTPUT**.

*Sintaxe:* PROC PRINT data=nome;

<b>var</b>	<variáveis>;	→ seleciona variáveis.
<b>by</b>	<variáveis>;	→ especifica subgrupos.
<b>sum</b>	<variáveis>;	→ apresenta a soma das variáveis selecionadas.

\* A opção data=nome define o SASdataset a ser impresso. Se este não for definido, o último SASdataset utilizado pelo SAS será considerado.