

Sumário

PROC FREQ	3
PROC UNIVARIATE;	5
PROC MEANS	6
PROC PLOT.....	8
PROC CHART	10
PROC STANDARD	12

PROC FREQ

O comando PROC FREQ gera tabelas de frequências de uma ou mais variáveis. Ele é utilizado também para gerar tabelas de contingência ou cruzamentos.

Exemplo:

```
DATA GRUPO ;  
INPUT NOME $  
SEXO $ IDADE ALTURA PESO RELIGIAO ;  
CARDS ;  
ALFREDO M 14 69 112 2  
ALICE F 13 56 84 2  
BERNADETE F 14 62 102 2  
MARTA F 13 65 98 4  
HENRIQUE M 14 63 102 3  
MARCOS M 12 57 83 3  
HELENA F 12 59 84 2  
CLARA F 15 62 102 2  
MARCELO M 13 62 84 4  
FREDERICO M 12 59 99 3  
MATEUS M 12 59 99 2  
ELIANE F 11 51 50 2  
ALINE F 14 64 90 2  
HELOISA F 12 56 77 2  
TANIA F 15 66 112 3  
FELIPE M 16 72 150 3  
ROBERTO M 12 64 128 4  
REINALDO M 15 67 113 1  
MAURICIO M 11 57 85 2  
ANACELIA F 13 62 108 3  
TOMAS M 99 99 85 9  
EDUARDO M 15 66 112 2
```

```
PROC FREQ ;
```

Resultado: O comando PROC FREQ gera tabelas de frequências simples para todas as variáveis do DATA SET.

Diagrama resumido do comando: PROC FREQ opções ;
TABLES especificações / opções ;
BY variáveis ;

PROC FREQ opções ;

As "opções" são:

PAGE

Informe para imprimir uma tabela por página

DATA = data set

Nome do DATA SET sobre o qual o procedimento deve atuar. Se esta opção for omitida, o procedimento atua sobre o último DATA SET criado ou referenciado em PROC's anteriores.

Exemplo: PROC FREQ DATA = GRUPO ;

Serão feitas tabelas de frequências dos dados do DATA SET chamado GRUPO, definido em algum DATA step anterior a esta PROC.

ORDER = FREQ

Esta opção faz com que os valores das frequências das variáveis sejam classificados em ordem decrescente.

Exemplo: PROC FREQ ORDER = FREQ ;

ORDER = DATA

Esta opção faz com que os resultados apareçam conforme a ordem de ocorrência no arquivo de dados.

Exemplo: PROC FREQ DATA = XYZ ORDER = DATA ;

TABLES especificações / opções ;

O TABLES serve para selecionar variáveis e definir o tipo de tabela a ser gerada. Uma PROC FREQ pode conter mais de um comando TABLES.

Exemplos: TABLES IDADE ;

Será gerada uma tabela de frequências simples da variável IDADE.

TABLES SEXO * RELIGIAO ;

Será gerada uma tabela de contingências, com o cruzamento das variáveis SEXO e RELIGIAO.

As tabelas podem ser:

- unidimensionais: TABLE região;
- bi dimensionais: TABLE região*queima;
- tridimensionais: TABLE região*queima*produção;

As "especificações" do TABLES podem ser dos tipos a seguir:

TABLES A * (B C) ; equivale a TABLES A * B A * C ;

TABLES (A -- C) ; equivale a TABLES A B C ;

TABLES (A -- C) * D ; equivale a TABLES A * D B * D C * D ;

As "opções" do TABLES podem ser:

MISSING

Inclui os "missing values" nos cálculos.

OUT = data set

Gera arquivo SAS contendo os valores das variáveis e a frequência. Se for pedido mais de um TABLES, será gerado arquivo com o último deles.

CHISQ

Faz teste Qui-quadrado e outras medidas de associação baseadas no Qui-quadrado.

MEASURES

Gera um conjunto de medidas de associação e seus erros-padrão.

ALL

Gera todas os testes e medidas das outras opções.

EXACT

Realiza o teste de FISHER

EXPECTED

Gera frequências esperadas das células.

DEVIATION

Imprime a diferença entre o valor observado e o esperado na célula.

NOFREQ

Não imprime as frequências das células num cruzamento.

NOPERCENT

Não imprime as porcentagens das células num cruzamento.

NOROW

Não imprime as porcentagens das linhas nas células de um cruzamento.

NOCOL

Não imprime as porcentagens das colunas nas células de um cruzamento.

NOPRINT

Não imprime as tabelas, mas somente as medidas estatísticas especificadas nas "opções".

```
Exemplo: DATA GRUPO ;  
          INPUT NOME $ SEXO $  
             IDADE ALTURA PESO RELIGIAO $ ;  
          IF IDADE LE 13 THEN IDADE = 1 ;
```

```

IF IDADE GT 13 THEN IDADE = 2 ;
IF IDADE EQ 99 THEN IDADE = . ;
IF PESO EQ 99 THEN PESO = . ;
IF ALTURA EQ 999 THEN ALTURA = . ;
IF RELIGIAO EQ 9 THEN RELIGIAO = . ;
CARDS ;
linhas de dados
DATA G2 ;
SET GRUPO ;
IF ALTURA LT 100 ;
PROC FREQ DATA = GRUPO ;
TABLES IDADE ;
LABEL IDADE 'GRUPOS: ATE 13 E MAIS DE 13 ANOS' ;
TITLE 'TABELA DE IDADE PARA O ARQUIVO - GRUPO' ;
PROC SORT DATA = GRUPO ;
BY SEXO ;
PROC FREQ ORDER = FREQ ;
TABLES ALTURA PESO ;
BY SEXO ;
TITLE 'TABELA DE ALTURA E PESO PARA ARQUIVO-GRUPO, POR SEXO' ;
PROC FREQ DATA = G2 ;
TITLE 'DATA SET G2 - INCLUI REGISTROS COM ALTURA < 100' ;
TITLE2 CRUZAMENTOS COM OPCOES - PRIMEIRA TABELA' ;
TABLES SEXO * RELIGIAO / CHISQ EXPECTED ;
PROC FREQ ;
TITLE2 CRUZAMENTOS COM OPCOES - SEGUNDA TABELA' ;
TABLES SEXO * RELIGIAO * IDADE / NOPERCENT ;

```

PROC UNIVARIATE;

```
PROC UNIVARIATE NORMAL PLOT; VAR ALTURA;
```

```
HISTOGRAM AGE/NORMAL (COLOR=RED W=5);
```

```
TITLE 'PROC UNIVARIATE EXAMPLE';
```

```
FOOTNOTE 'Evaluate distribution of variables';
```

Opções comuns para PROC UNIVARIATE incluem:

DATA= - Specifies data set to use

NORMAL - Produces a test of normality

FREQ – Produces a frequency table

PLOT – Produces stem-and-leaf plot

PROC MEANS

Este comando gera estatísticas descritivas de variáveis numéricas.

```
Exemplo: DATA A ;  
          INPUT TESTE1 TESTE2 LUGAR $ DIA @@ ;  
          CARDS ;  
          04 54 S 1 07 70 N 1 10 69 N 2 04 52 S 1  
          07 70 S 2 08 74 N 1 04 60 S 1 07 62 S 2  
          07 80 N 1 06 61 S 2 06 77 N 2 08 75 N 2  
          PROC MEANS ;
```

Resultado: Este programa gera estatísticas descritivas das variáveis TESTE1, TESTE2 e DIA, que são numéricas.

Diagrama resumido do comando: PROC MEANS opções ;
VAR lista de variáveis ;
BY lista de variáveis ;
OUTPUT opções ;

PROC MEANS opções ;

As "opções" podem ser:

DATA = data set

Nome do DATA SET sobre o qual o procedimento deve atuar. Se esta opção for omitida, o procedimento atua sobre o último DATA SET criado ou referenciado em PROC's anteriores.

Exemplo: PROC MEANS DATA = A ;

Serão geradas estatísticas descritivas para os dados do DATA SET chamado A, definido em algum DATA step anterior a esta PROC.

MAXDEC = n

Define o número de casas decimais (0 a 8) a ser impresso nos resultados.

N

Imprime o número de observações incluídas nos cálculos.

NMISS

Imprime o número de missing values.

MEAN

Imprime a média.

STD

Imprime o desvio padrão.

MIN

Imprime o menor valor.

MAX

Imprime o maior valor.

RANGE

Imprime a amplitude total: diferença entre o maior e o menor valor.

SUM

Imprime a soma (total).

VAR

Imprime a variância.

USS

Imprime a soma de quadrados não corrigida.

CSS

Imprime a soma de quadrados corrigida.

STDERR

Imprime o erro padrão da média.

CV

Imprime o coeficiente de variação.

SKEWNESS

Imprime a medida de assimetria.

KURTOSIS

Imprime a medida de curtose (achatamento).

T

Imprime a estatística T-Student para testar a hipótese de que a média é zero.

PRT

Imprime a probabilidade determinada pela estatística T-Student.

OUTPUT opções ;

Este comando gera arquivo SAS contendo resultados gerados pela PROC MEANS. As opções são:

`OUT = dataset estatística = nomevars ;`

Gera arquivo SAS com o nome especificado em OUT e especifica quais as estatísticas a serem incluídas no data set gerado, e o nome da variável pelo qual esta estatística será reconhecida.

As estatísticas são:

N, NMISS, MEAN, STD, MIN, MAX, RANGE, SUM, VAR, USS, CSS, STDERR, CV, SKEWNESS, KURTOSIS, T, PRT, SUMWGT.

```
Exemplo: PROC MEANS ;
          VAR X1 X2 ;
          BY GRUPO ;
          OUTPUT OUT = ESTATS MEAN = MX1 MX2 STD = SX1 ;
```

Se a variável GRUPO possui dois valores, A e B, o arquivo ESTATS terá duas observações, com as seguintes variáveis: GRUPO, MX1, MX2 e SX1, sendo MX1 e MX2 as médias de X1 e X2, respectivamente, e SX1 o desvio padrão de X1.

```
Exemplo: DATA A ;
          INPUT TESTE1 TESTE2 LUGAR $ DIA @@ ;
          CARDS ;
          04 54 S 1 07 70 N 1 10 69 N 2 04 52 S 1
          07 70 S 2 08 74 N 1 04 60 S 1 07 62 S 2
          07 80 N 1 06 61 S 2 06 77 N 2 08 75 N 2
          PROC MEANS ;
          VAR TESTE1 TESTE2 ;
          PROC MEANS DATA = A MEXDEC = 3 NMISS RANGE USS CSS
                   SKEWNESS KURTOSIS T PRT ;
          VAR TESTE1 TESTE2 ;
          TITLE 'ESTATISTICAS SELECCIONADAS ;
          PROC SORT ;
          BY LUGAR DIA ;
          PROC MEANS MAXDEC = 3 ;
          BY LUGAR DIA ;
          VAR TESTE1 TESTE2 ;
          OUTPUT OUT = NOVO MEAN = T1MED T2MED STDERR = T1EM T2EM ;
          TITLE 'ESTATISTICAS POR LUGAR E DIA' ;
          PROC PRINT DATA = NOVO ;
          TITLE 'NOVO DATA SET' ;
```

PROC PLOT

O procedimento PROC PLOT produz gráficos bidimensionais.

```
Exemplo: DATA TESTE ;
          INPUT X Y TRAT $ @@ ;
          CARDS ;
          125 3.1 P 135 9.0 P 144 14.9 P 153 20.2 P
          136 2.0 Q 152 8.5 Q 160 12.6 Q 165 17.1 Q
          154 3.2 R 164 10.5 R 173 15.4 R 183 20.7 R
```

```
PROC PLOT ;  
PLOT Y * X ;
```

Diagrama resumido do comando: PROC PLOT opções ;
PLOT especificações / opções ;
BY lista de variáveis ;

PROC PLOT opções ;

As "opções" são:

DATA = data set

Nome do DATA SET sobre o qual o procedimento deve atuar. Se esta opção for omitida, o procedimento atua sobre o último DATA SET criado ou referenciado em PROC's anteriores.

Exemplo: PROC PLOT DATA = B ;

Serão feitos gráficos com os dados do DATA SET chamado B, definido em algum DATA step anterior a esta PROC.

UNIFORM

Faz com que todos os gráficos produzidos para as categorias das variáveis especificadas no comando BY sejam impressos com a mesma escala para serem comparadas.

NOLEGEND

Não imprime legendas no cabeçalho dos gráficos.

PLOT especificações / opções ;

As "especificações" podem ser do tipo:

vertical * horizontal

Exemplo: PLOT Y * X ;

No eixo vertical aparece a variável Y e X no horizontal. Os pontos de intersecção serão mostrados pelo carácter "A".

vertical * horizontal = carácter'

Exemplo: PLOT Y * X = '+' ;

Os pontos do gráfico serão indicados pelo carácter "+" ao invés de "A".

vertical * horizontal = variável

Exemplo: PLOT Y * X = SEXO ;

Os pontos serão indicados por "F" ou "M", por exemplo, de acordo com o resultado observado para a variável SEXO em cada registro.

As "opções" podem ser:

VAXIS = valores

Especifica os intervalos a serem marcados no eixo vertical.

Exemplo: PROC PLOT ;
 PLOT Y * X / VAXIS = 1 TO 50 BY 5 ;

O eixo vertical conterá marcas a cada 5 unidades, sendo a primeira com valor 1, até 50.

HAXIS = valores

Especifica os intervalos a serem marcados no eixo horizontal.

Exemplo: PROC PLOT ;
 PLOT Y * X / HAXIS = 0 TO 20 BY 2 ;

O eixo horizontal conterá marcas a cada 2 unidades, sendo a primeira com valor 0, até 20.

VZERO e HZERO

O gráfico começa de zero nos eixos vertical e horizontal, respectivamente.

OVERLAY

Esta opção , utilizada para sobrepor gráficos.

Exemplo: PROC PLOT ;
 PLOT Y * X A * B / OVERLAY ;

Os gráficos de Y por X e A por B serão impressos no mesmo sistema de eixos.

Exemplo: DATA TESTE ;
 INPUT X Y TRAT \$ V1 V2 ;
 CARDS ;
 125 3.1 P 75 85
 135 9.0 P 110 112
 144 14.9 P 75 81
 153 20.2 P 105 108
 136 2.0 Q 112 115
 152 8.5 Q 75 77
 160 12.6 Q 90 99
 165 17.1 Q 70 73
 154 3.2 R 118 121
 164 10.5 R 103 100
 173 15.4 R 118 121
 183 20.7 R 102 110
 PROC PLOT ;
 PLOT Y * X = TRAT / VAXIS = 125 TO 185 BY 5 ;
 TITLE 'GRAFICO DE Y POR X' ;
 PROC PLOT ;
 PLOT Y * X = TRAT V1 * V2 / OVERLAY ;
 TITLE 'GRAFICOS SOBREPOSTOS' ;

PROC CHART

O comando PROC CHART produz gráficos de barras, colunas, setores e histogramas usando estatísticas como frequências, porcentagens, somas e médias. Os valores das estatísticas são representados pelo tamanho das barras.

Exemplos: DATA GRAF ;

```
INPUT SEXO $ IDADE @@ ;
```

```
CARDS ;
```

```
F 30 M 33 F 31 M 40 F 25 M 19 F 52 M 51 F 42 M 28
```

```
F 21 F 34 M 25 M 18 M 60 F 37 M 27 F 48 F 39 F 50
```

```
PROC CHART ;
```

```
VBAR SEXO ;
```

Diagrama resumido do comando: PROC CHART opção ;

VBAR variáveis / opções ;

HBAR variáveis / opções ;

BLOCK variáveis / opções ;

PIE variáveis / opções ;

BY variáveis ;

PROC CHART opção ;

A "opção" ,:

DATA = data set

Nome do DATA SET sobre o qual o procedimento deve atuar. Se esta opção for omitida, o procedimento atua sobre o último DATA SET criado ou referenciado em PROC's anteriores.

Exemplo: PROC CHART DATA = GRAF ;

Serão feitos gráficos utilizando dados do DATA SET chamado GRAF, definido em algum DATA step anterior a esta PROC.

VBAR variáveis / opções ;

Faz gráfico de barras verticais das variáveis especificadas.

HBAR variáveis / opções ;

Faz gráfico de barras horizontais das variáveis especificadas.

BLOCK variáveis / opções ;

Faz gráfico de blocos ou colunas tridimensionais.

PIE variáveis / opções ;

Faz gráfico de setores das variáveis especificadas.

Algumas "opções" para todos os tipos de gráficos são:

TYPE = FREQ (frequência)
TYPE = PERCENT ou PCT (porcentagens)
TYPE = CFREQ (frequências acumuladas)
TYPE = CPCT (porcentagens acumuladas)
TYPE = SUM (totais)
TYPE = MEAN (médias)

O TYPE especifica qual o tipo de medida deve ser representado pelas barras ou secções.
GROUP = variável
Repete o gráfico, de acordo com os valores da variável especificada, no mesmo sistema de eixos.

Exemplo: PROC CHART ;
 VBAR DEPT / GROUP = SEXO ;
Será gerado um gráfico para SEXO masculino e outro para SEXO feminino no mesmo eixo.

SUBGROUP = variável Divide as barras ou secções de acordo com os valores da variável especificada.

Exemplo: PROC CHART ;
 VBAR DEPT / SUBGROUP = SEXO ;
Será produzido um gráfico com barras compostas de duas partes, sendo uma para representar SEXO masculino e a outra SEXO feminino.

SUMVAR = variável
A "variável" será utilizada no cálculo de médias, somas ou frequências.

Exemplo: PROC CHART ;
 VBAR LOCAL / TYPE = MEAN SUMVAR = VENDAS ;
Produz um gráfico mostrando a média de VENDAS por LOCAL.

Exemplos: DATA GRAF2 ;
 INPUT DEPT \$ SEXO \$ WT P75 P76 P77 ANO VENDAS ;
 CARDS ;
 P F 17 3500 2500 800 1975 3500
 P F 17 3500 2500 800 1976 2500
 P F 17 3500 2500 800 1977 800
 P M 21 3651 5391 4500 1975 3651
 P M 21 3651 5391 4500 1976 5391
 P M 21 3651 5391 4500 1977 4500
 P M 21 2644 3500 3000 1975 2644
 P M 21 2644 3500 3000 1976 3500
 P M 21 2644 3500 3000 1977 3000
 T F 12 5672 6100 7400 1975 5672
 T F 12 5672 6100 7400 1976 6100
 T F 12 5672 6100 7400 1977 7400

```

T M 45 1253 4698 9345 1975 1253
T M 45 1253 4698 9345 1976 4698
T M 45 1253 4698 9345 1977 9345
R F 02 9050 12062 15931 1975 9050
R F 02 9050 12062 15931 1976 12062
R F 02 9050 12062 15931 1977 15931
R M 34 . . . 1975 .
PROC CHART ;
  VBAR DEPT / TYPE = PCT ;
  BY SEXO ;
PROC CHART ;
  VBAR DEPT / TYPE = CPCT GROUP = ANO ;
PROC CHART ;
  VBAR ANO / SUBGROUP = DEPT ;
PROC CHART ;
  PIE DEPT / SUMVAR = VENDAS ;
PROC CHART ;
  BLOCK DEPT / TYPE = MEAN SUMVAR = VENDAS GROUP = SEXO ;

```

PROC STANDARD

```

data dados;
input x y @@;
cards;

5 80 4 100 3 40 2 60 1 20

proc corr;
  var x;
  with y; run;

proc standard data=Sol mean=0 std=1 vardef=n out=zscores;
  var x y; run;

proc print; run;

data corr;
  set zscores;
  xy=x*y;

proc means mean;
  var x y xy; run;

```