UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

ESTATÍSTICA USANDO MINITAB

Renzo Joel Flores Ortiz

Ela Mercedes Medrano de Toscano

BELO HORIZONTE
MINAS GERAIS - BRASIL
SETEMBRO DE 2010

Estatística Usando Minitab

Renzo Joel Flores Ortiz
Graduando em Estatística
Universidade Federal de Minas Gerais

Ela Mercedes Medrano de Toscano
Professora Associada do Departamento de Estatística
Universidade Federal de Minas Gerais

Belo Horizonte
Minas Gerais - Brasil
Setembro de 2010

PREFÁCIO

O presente relatório busca apresentar os recursos mais utilizados do Minitab em cursos de Estatística Básica por meio de exercícios resolvidos.

Organizado em seis capítulos, o texto aborda Estatística Descritiva, Probabilidade e Testes de Hipóteses.

O trabalho foi desenvolvido durante o período de abril a dezembro de 2008, como atividade integrante do Programa de Monitoria de Graduação - PMG/UFMG, contando com o apoio de bolsa desse programa. Os autores agradecem o apoio recebido para o desenvolvimento deste trabalho.

Sugestões e correções serão muito bem recebidas e ajudarão no aprimoramento e desenvolvimento de próximas versões.

Renzo Joel Flores Ortiz
(renzojfo@gmail.com)

Ela Mercedes Medrano de Toscano
(mercedes@est.ufmg.br)

Belo Horizonte, Setembro de 2010

Sumário

Capítulo 1 – Introdução

Trabalhando em qualquer campo do conhecimento estamos sujeitos a analisar e a entender conjuntos de dados. Seja em universidades, bancos, indústrias ou hospitais, o estatístico propõe soluções para profissionais das mais diversas áreas.

Com a demanda por resultados mais rápidos e eficientes, hoje o profissional de Estatística é amparado de *softwares* (pacotes) especificamente voltados para a análise estatística.

Desenvolvido em 1972 na Universidade Estadual da Pensilvânia, o Minitab é um exemplo de pacote estatístico largamente utilizado por empresas e universidades ao redor do mundo.

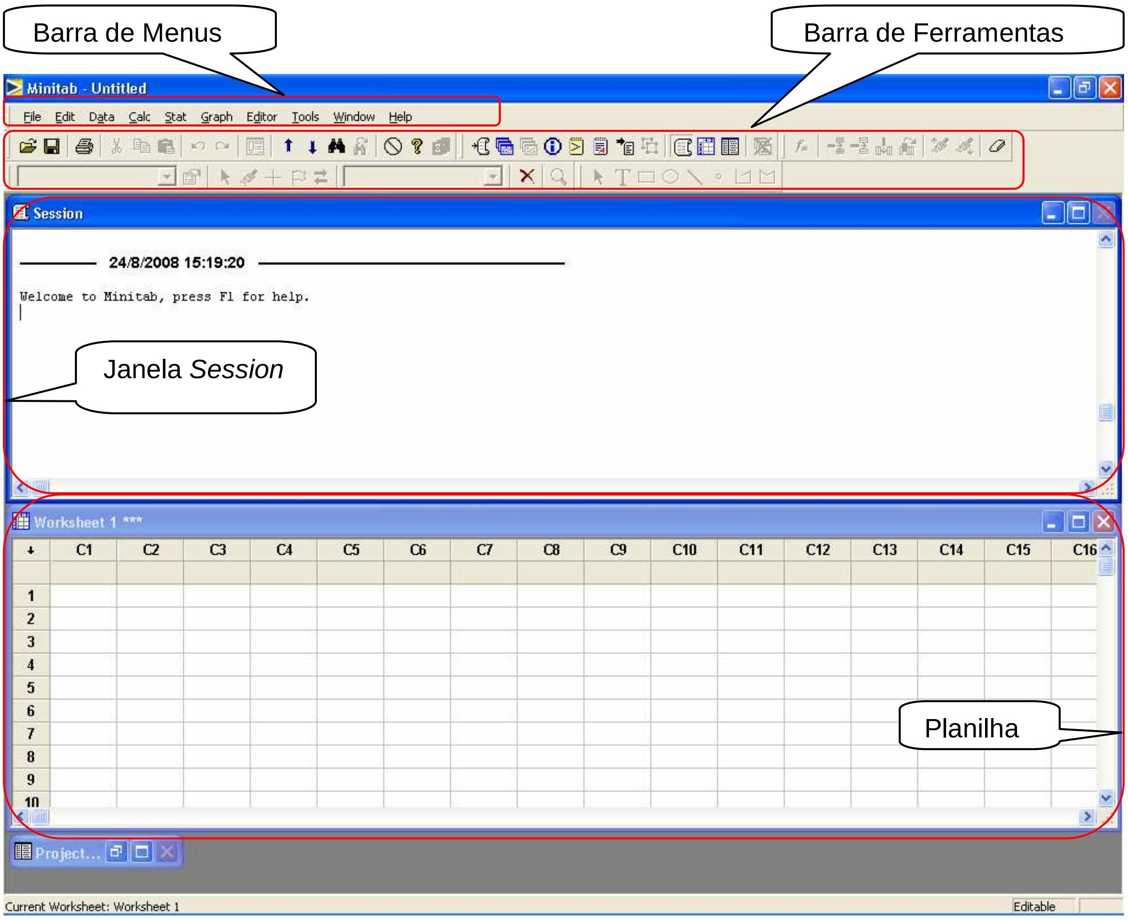
No mercado o Minitab é aplicado a empresas dos mais variados segmentos, como por exemplo: General Electric, LG, Nokia, AT&T, Rolls Royce, Samsung, Xerox Corporation, dentre outras.

Nas universidades, disciplinas de Estatística integram o currículo de muitos cursos de graduação: desde as engenharias, passando pelas ciências humanas até a área da saúde, é possível ver como a Estatística é essencial para a formação de profissionais qualificados, independente das áreas de atuação.

Considerando a importância do computador como ferramenta de apoio ao ensino, *Estatística Usando Minitab* visa apresentar, de forma simples e objetiva, os recursos mais utilizados do Minitab em cursos de estatística básica. Trabalharemos em referência a versão 15, entretanto não descartamos o uso de sua versão precedente, a versão 14, que possui uma interface semelhante.

Capítulo 2 – Descrição e Características do Minitab

Uma das características que faz do Minitab um dos pacotes de maior preferência é a facilidade para a utilização dos seus recursos. A seguir apresentamos uma breve descrição desse ambiente.



2.1 – Barra de Menus

As operações e comandos possíveis de serem executados no Minitab estão disponíveis na Barra de Menus.

- **File** – Menu relacionado a operações com arquivos (abrir, salvar,...).
- **Edit** – Menu relacionado a operações para editar arquivos (copiar, colar,...).
- **Data** – Menu relacionado a manipulação de dados na planilha (transpor colunas, exibir postos...).

- **Calc** – Menu relacionado à cálculos (estatísticas, probabilidades,...).

- **Stat** – Menu relacionado à análise estatística (estatística descritiva, teste de hipóteses,...).

- **Graph** – Menu relacionado a construção de gráficos (histograma, gráfico de pontos,...).

- **Editor** – Menu relacionado a comandos na janela *Session*.

- **Tools** – Menu que contém utilidades diversas como bloco de notas, calculadora, etc.

- **Window** – Menu que gerencia as janelas do Minitab (minimizar, restaurar,...).

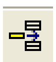
- **Help** – Menu que disponibiliza ajuda sobre o ambiente do Minitab.

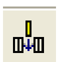
2.2 – Barra de Ferramentas

A Barra de Ferramentas nos fornece maior facilidade para manipulação de algo que se esteja trabalhando, como uma planilha ou um gráfico. Muitas das opções da Barra de Ferramentas podem ser habilitadas no menu ***Tools*** → ***Toolbars***.

A seguir descrevemos algumas dessas ferramentas:

Assinalar fórmula a uma coluna.


Inserir linha.


Inserir coluna.

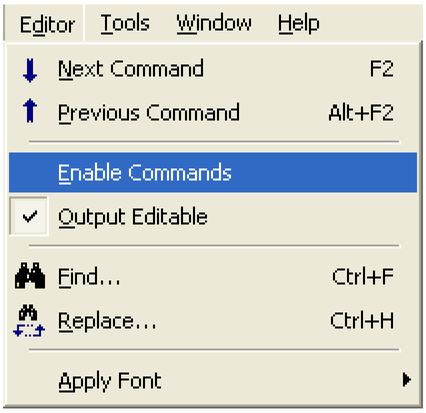

Mover colunas.

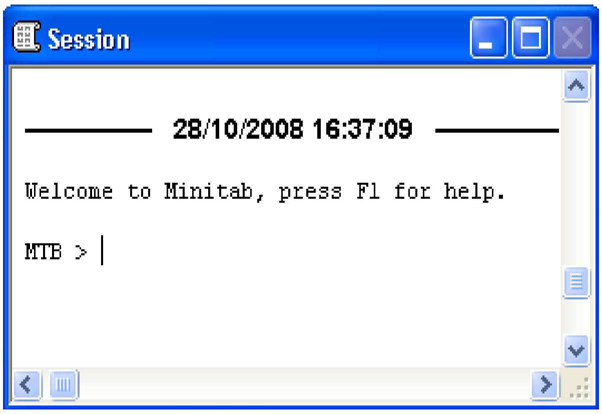
2.3 – Janela Session

A janela ***Session*** é o “lugar” onde aparecem as saídas dos procedimentos estatísticos disponíveis no Minitab. Esses procedimentos podem ser acessados através de menus ou por comandos apropriados chamados de *session commands*.

Para habilitar o uso de *session commands*, acesse na Barra de Menus:

- ***Editor → Enable Commands***





A janela ***Session*** estará habilitada para o uso de *session commands* quando aparecer um cursor piscando e o procedimento para executar comandos consiste, basicamente, em escrevê-los e teclar **Enter** ao final de cada linha.

2.4 – Planilha (Worksheet)

É um tipo de tabela utilizada para cálculos ou apresentação de dados.
Ela é composta por linhas e colunas, sendo que em cada coluna aparece toda a informação de uma variável e em cada linha aparece toda a informação de todas as variáveis de um individuo ou objeto. Ao par linha e coluna existe uma área denominada célula (*cell*).

Capítulo 3 – Estatística Descritiva

Estatística Descritiva é a área da Estatística que apresenta métodos para organizar, resumir, comparar e descrever dados através de tabelas, gráficos e medidas resumo. Apresentamos neste capítulo os principais recursos gráficos e numéricos para uma análise descritiva no Minitab.

Como objeto de estudo foi utilizado um banco de dados referente aos candidatos ao curso de Estatística no Vestibular 2008 da UFMG[1].
As variáveis analisadas são: *Sexo* (em que F designa feminino e M masculino), *Nota 1ª Etapa Matemática* (que pode assumir valores entre 0 e 8) e *Nota Matemática 2ª Etapa* (que pode assumir valores entre 0 e 100).

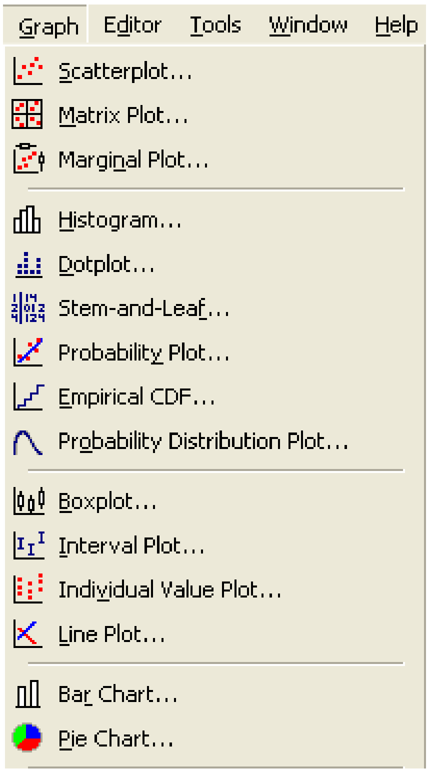
Worksheet 1 ***

↓	C1-T	C2	C3
	Sexo	Nota Matemática 1ª Etapa	Nota Matemática 2ª Etapa
1	F	7	0
2	M	6	24
3	F	6	44
4	F	3	53
5	F	2	*
6	M	2	14
7	F	5	35
8	F	1	*
9	M	4	*
10	F	2	*
11	F	0	*

[1] Os autores agradecem a Profa. Glaura Franco do Departamento de Estatística da UFMG pelos dados cedidos.

3.1 – Gráficos

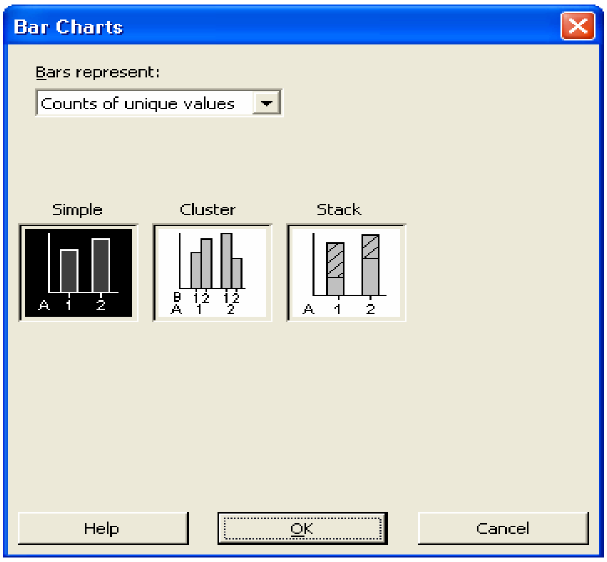
Um gráfico é uma forma de apresentação de dados que permite uma leitura rápida e global de um fenômeno estudado. Os tipos de gráficos disponíveis no Minitab se encontram no menu ***Graph***.



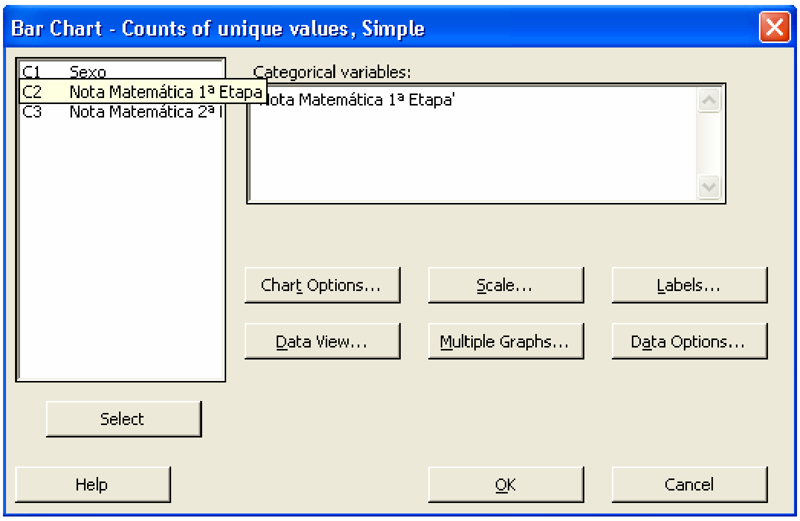
3.1.1 – Gráfico de Barras

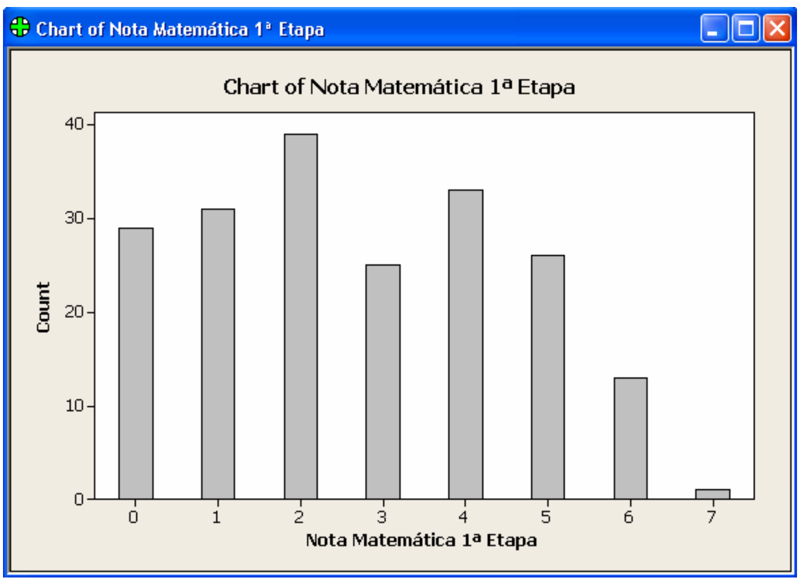
O Gráfico de Barras é representado no plano cartesiano com os valores da variável no eixo das abscissas e as frequências ou porcentagens no eixo das ordenadas. A cada valor da variável temos uma barra com altura correspondendo à sua frequência ou porcentagem. É um gráfico que se adapta bem às variáveis discretas ou qualitativas ordinais.

- **Exemplo 3.1.1** – Construa o Gráfico de Barras para a variável *Nota Matemática 1ª Etapa*.
  - Selecione ***Graph → Bar Chart***
  - <u>Janela *Bar Charts*</u>: Escolha a opção ***Counts of unique values*** em ***Bars represent*** e a seguir selecione ***Simple → OK***



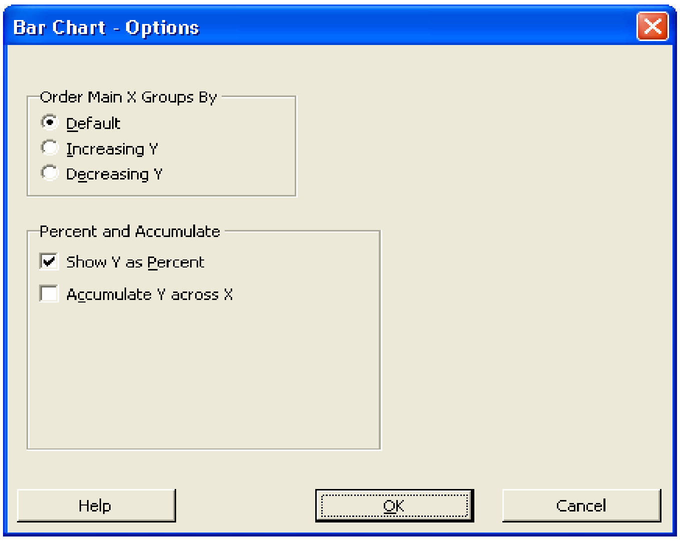
- Janela *Bar Chart - Counts of unique values, Simple*: Selecione a coluna ***C2 Nota Matemática 1ª Etapa*** para o campo ***Categorical variables*** → ***OK***



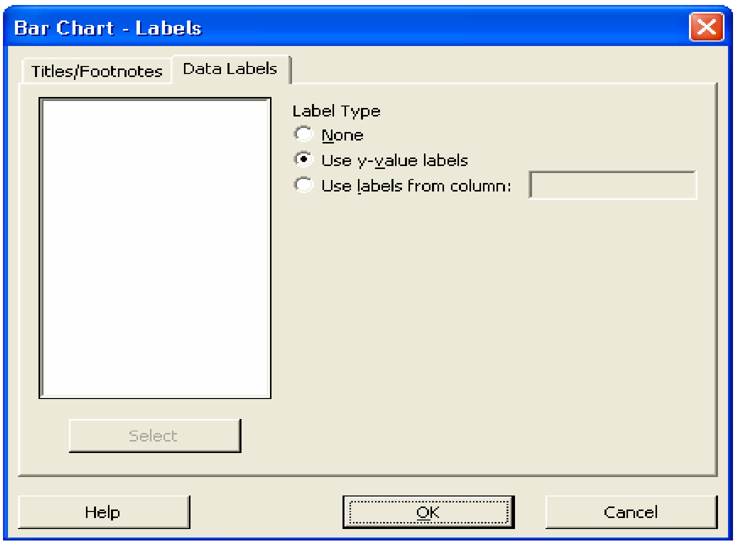


Os gráficos possuem opções com que podemos acrescentar mais ou menos informações. Um incremento que poderíamos incluir ao gráfico de barras seria exibir as porcentagens das respectivas barras. Para isso, siga os próximos passos a partir da janela ***Bar Chart - Counts of unique values, Simple***:

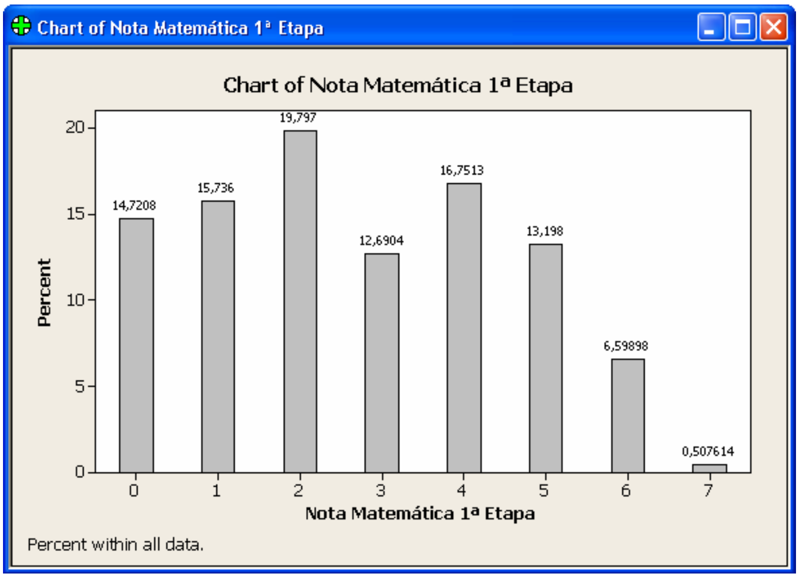
- Janela *Bar Chart - Counts of unique values, Simple*: Clique em ***Chart options*** e marque a opção ***Show Y as Percent*** → ***OK***



- Janela *Bar Chart - Counts of unique values, Simple*: Clique em ***Labels*** → ***Data Labels*** e marque a opção ***Use y-value labels*** → ***OK***



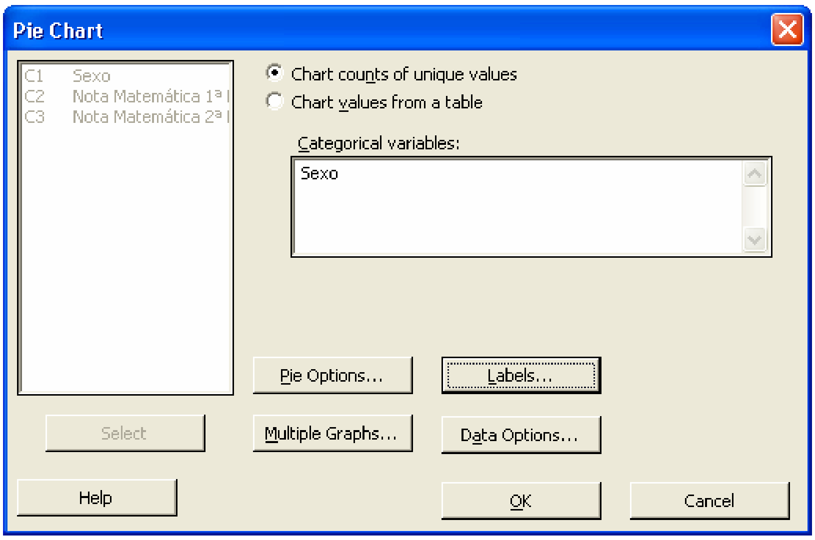
- Janela *Bar Chart - Counts of unique values, Simple*: → ***OK***

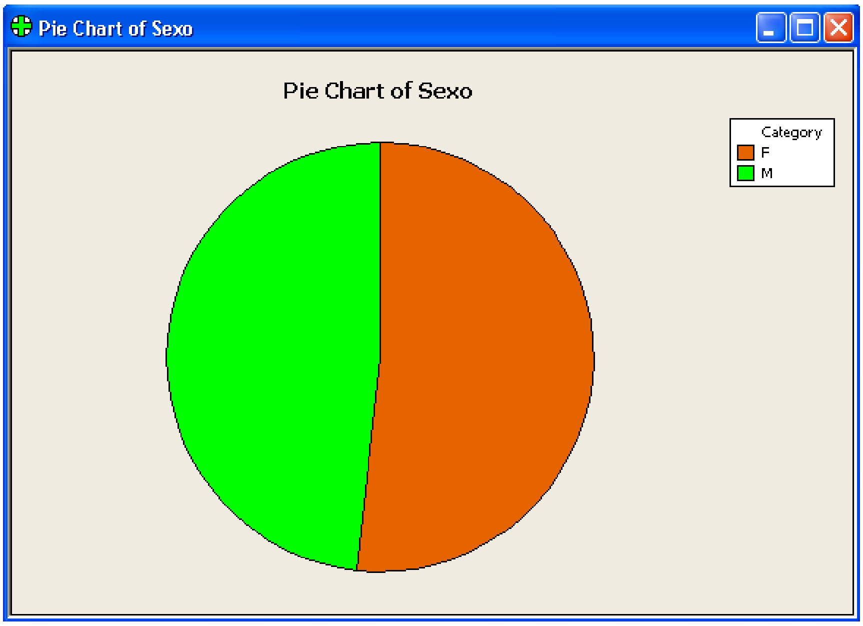


3.1.2 – Gráfico de Setores (Gráfico de Pizza)

O Gráfico de Setores ou de "Pizza" consiste num círculo dividido em setores correspondentes às porcentagens ou frequências das variáveis representadas. É indicado para representar variáveis qualitativas.

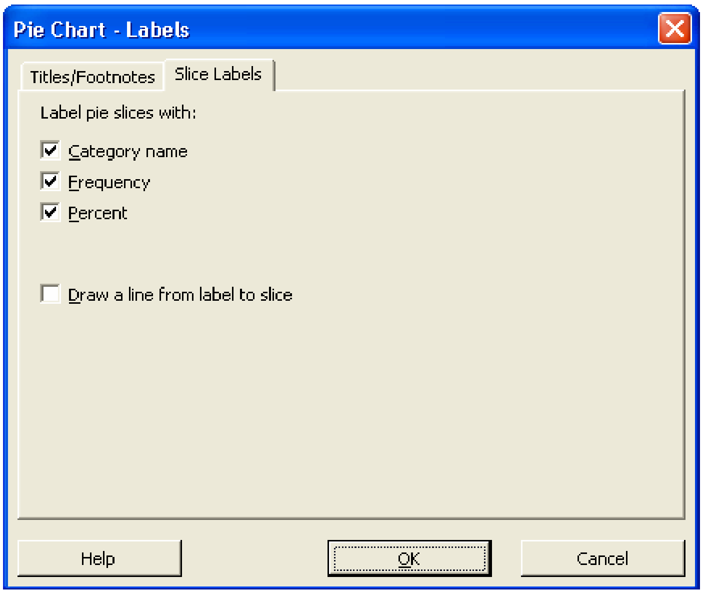
- **Exemplo 3.1.2** – Construa o Gráfico de Setores para a variável *Sexo*.
  - Selecione ***Graph*** → ***Pie Chart***
  - Janela *Pie Chart*: Marque a opção ***Chart counts of unique values*** e selecione a coluna ***C1 Sexo*** para o campo ***Categorical variables*** → ***OK***



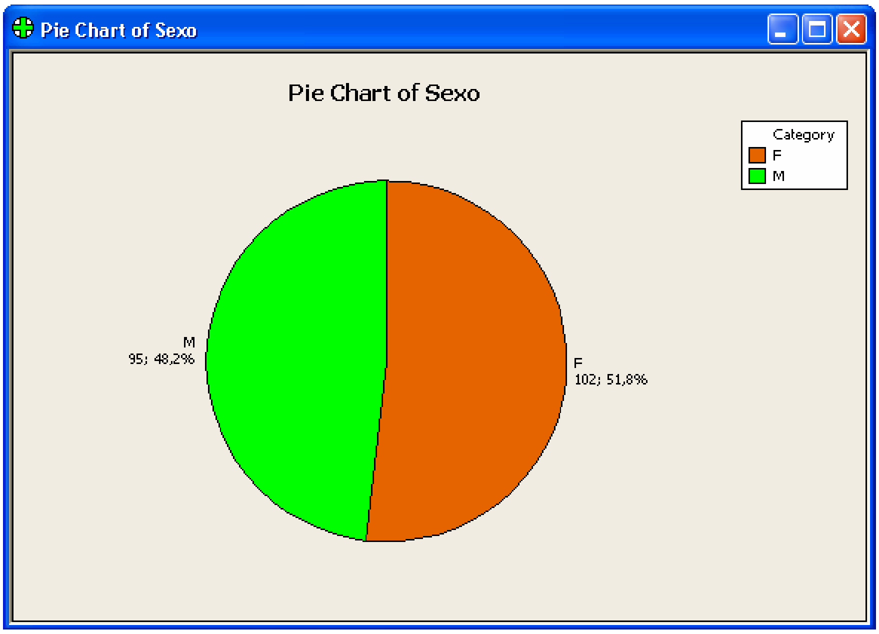


Pode-se incluir ao gráfico de setores os nomes e as frequências dos respectivos setores, além das porcentagens, como feito no exemplo do gráfico de barras. Para isso, siga os próximos passos a partir da janela ***Pie Chart***:

- Janela *Pie Chart*: Clique em ***Labels*** → ***Slice Labels*** e marque as opções ***Category name***, ***Frequency*** e ***Percent*** → ***OK***



- Janela *Pie Chart*: → ***OK***



3.1.3 – Histograma

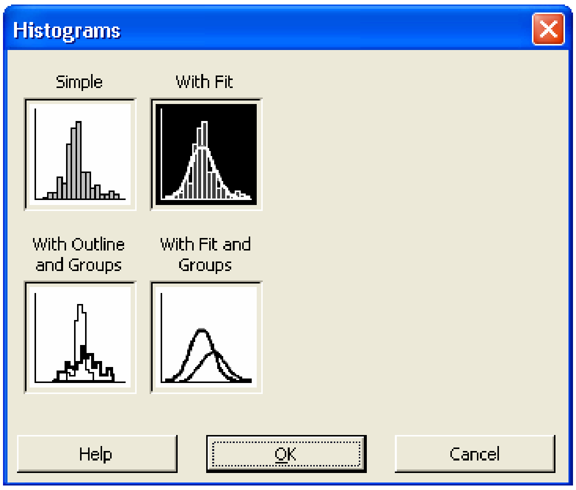
O histograma é um gráfico de distribuição de frequências em barras justapostas com bases que são faixas de valores da variável e cuja área é igual a frequência relativa da respectiva faixa.

A construção de histogramas tem caráter preliminar em qualquer estudo e é um importante indicador da distribuição dos dados. Indica, por exemplo, se os dados seguem uma distribuição normal. A suposição de que os dados seguem uma distribuição normal é assumida para a maioria dos métodos estatísticos mais utilizados, como o teste *t* de Student, ANOVA, regressão linear e intervalos de confiança. Este fato somado a resultados teóricos fundamentais (Teorema Central do Limite) faz com que a distribuição normal seja a distribuição teórica mais importante em Estatística.

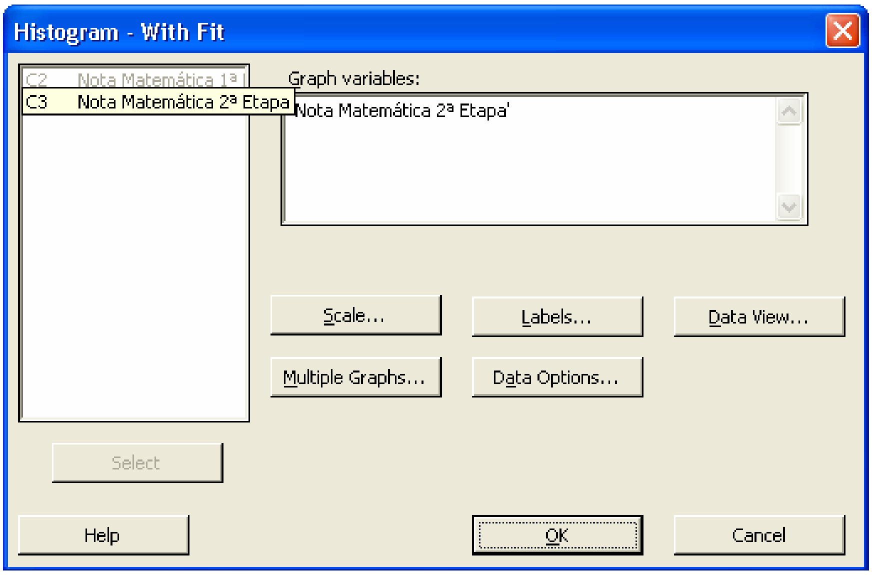
- **Exemplo 3.1.3** – Construa o Histograma para a variável *Nota Matemática 2ª Etapa* ajustando a esse gráfico uma curva normal com sua respectiva média e desvio padrão.

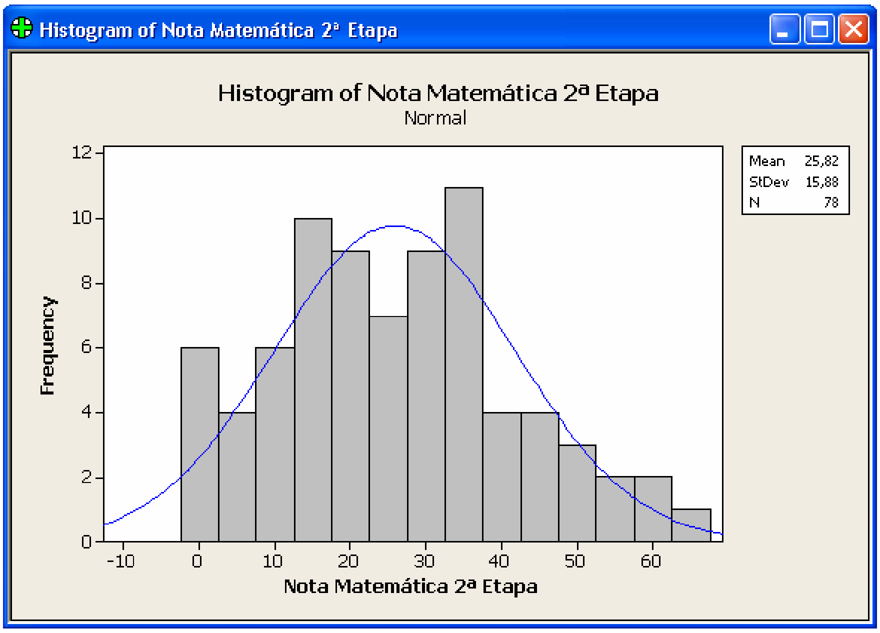
  - Selecione ***Graph*** → ***Histogram***

  - Janela *Histograms*: Selecione ***With Fit*** → ***OK***



- Janela *Histogram - With Fit*: Selecione a coluna ***C3 Nota Matemática 2ª Etapa*** para o campo ***Graph variables*** → ***OK***

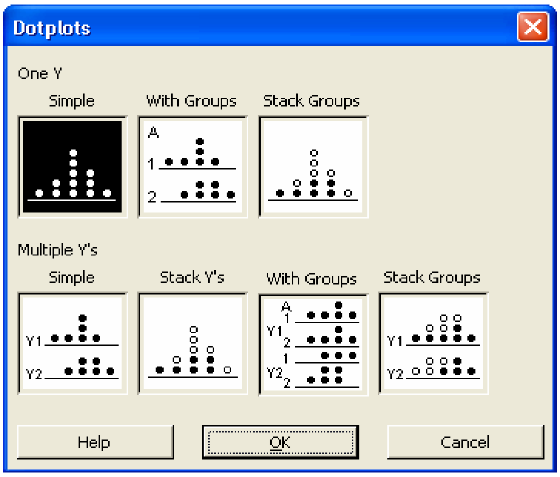




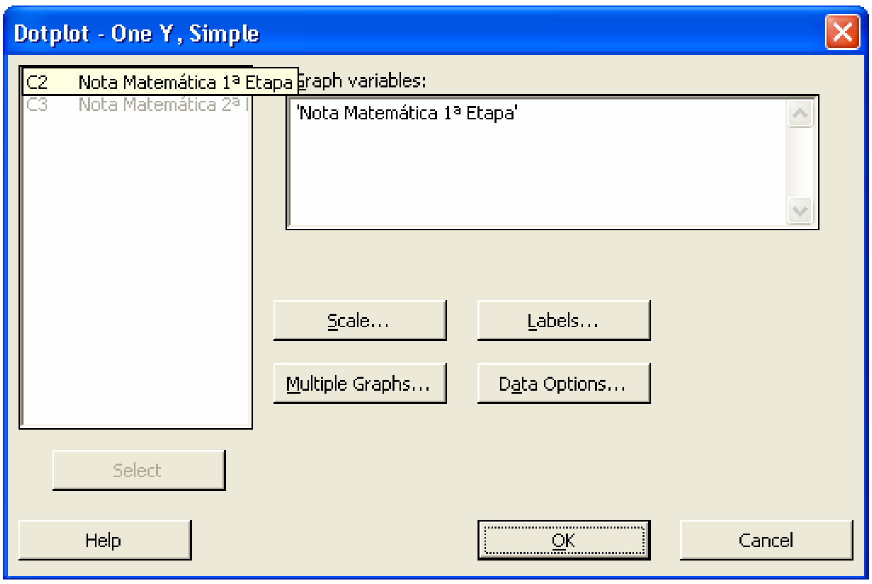
3.1.4 – Gráfico de Pontos

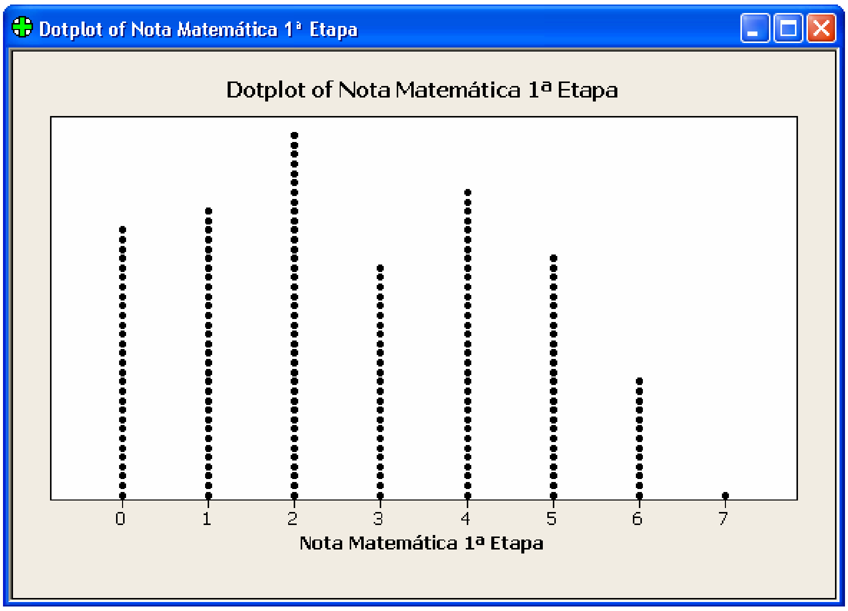
O Gráfico de Pontos consiste em representar dados por pontos ao longo de uma reta provida de escala. Valores repetidos são empilhados um em cima do outro.

- **Exemplo 3.1.4** – Construa o gráfico de pontos para a variável *Nota Matemática 1ª Etapa*.
  - Selecione ***Graph* → *Dotplot***
  - Janela *Dotplots*: Na linha ***One Y*** selecione ***Simple* → *OK***



  - Janela *Dotplot - One Y, Simple*: Selecione a coluna ***C2 Nota Matemática 1ª Etapa*** para o campo ***Graph variables* → *OK***

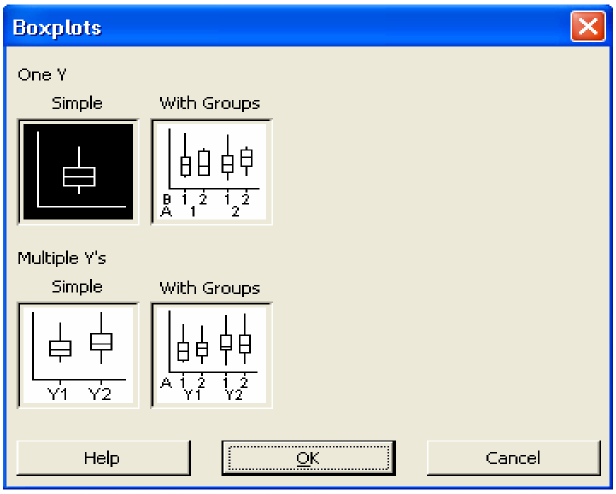




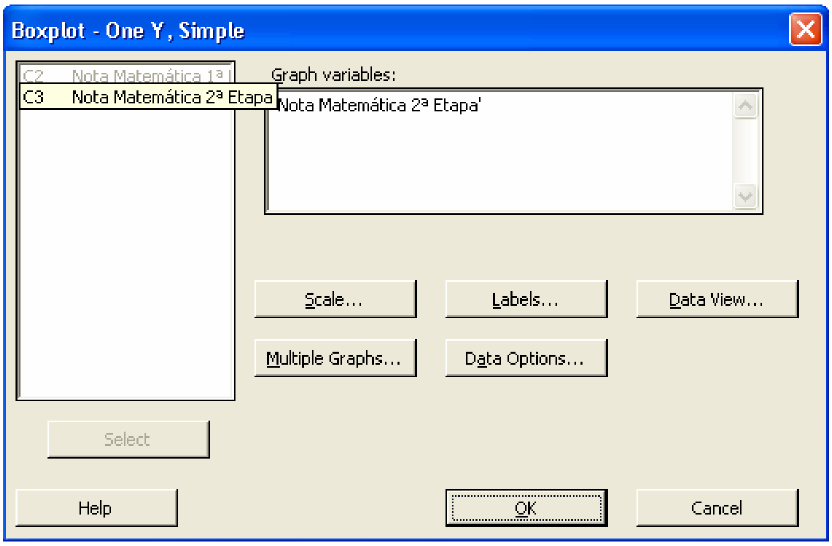
3.1.5 – Boxplot

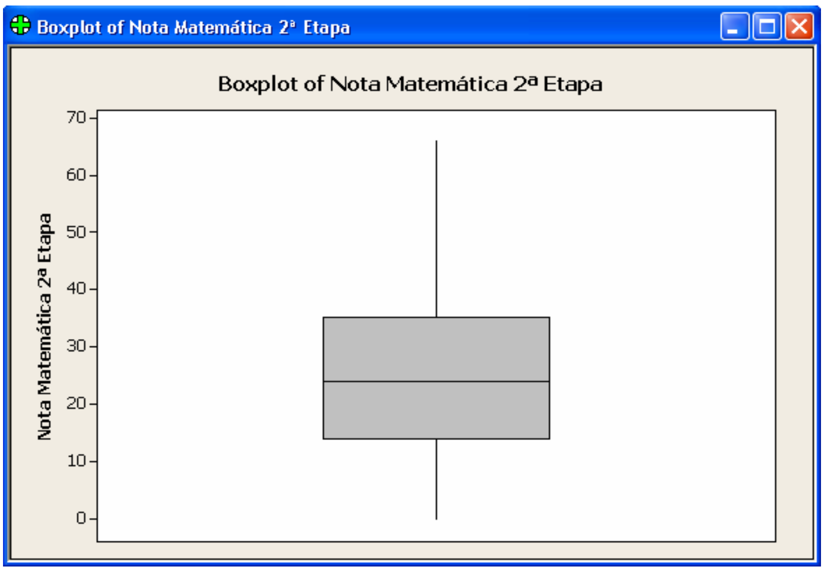
O *Boxplot* é um gráfico no formato de "caixa", cujos limites são o primeiro quartil e o terceiro quartil. A mediana é representada por um traço no interior da caixa e segmentos de reta são colocados da caixa até os valores máximo e mínimo, que não sejam observações discrepantes. Este gráfico fornece informações sobre posição, dispersão, assimetria, caudas e a presença de dados discrepantes, sejam atípicos ou *outliers*.

- **Exemplo 3.1.5** – Construa o *Boxplot* para a variável *Nota Matemática 2ª Etapa*.
  - Selecione ***Graph*** → ***Boxplot***
  - Janela *Boxplots*: Na linha ***One Y*** selecione ***Simple*** → ***OK***



- Janela *Boxplot - One Y, Simple*: Selecione a coluna ***C3 Nota Matemática 2ª Etapa*** para ***Graph variables*** → ***OK***

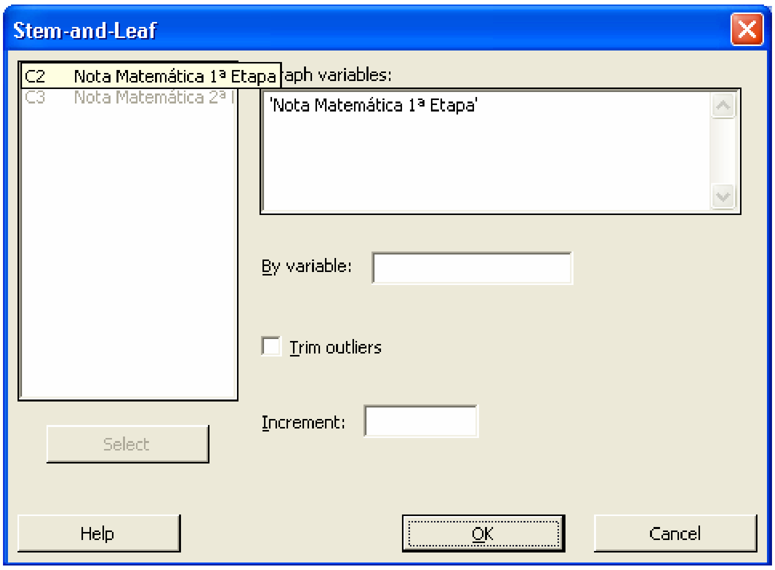


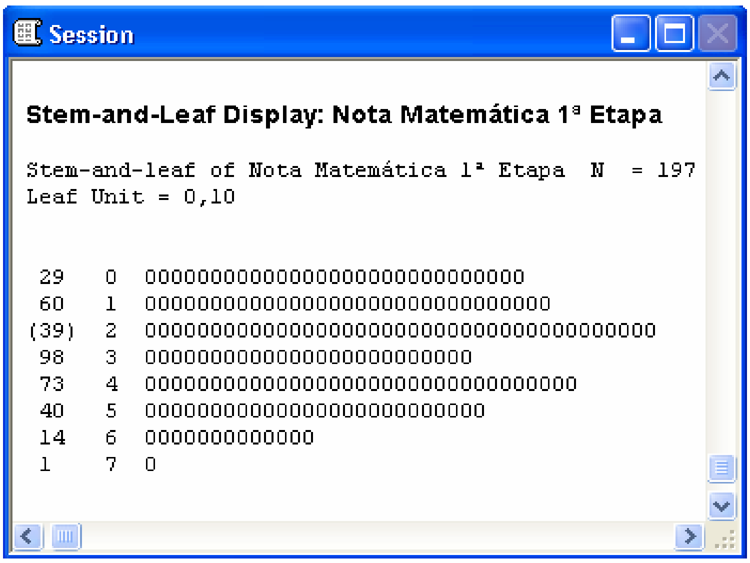


3.1.6 – Ramo-e-Folhas

O diagrama de Ramo-e-Folhas consiste em representar dados separando cada valor em duas partes: o *ramo* e a *folha* (está última, localizada à direita do *ramo*). As *folhas* representam as unidades e o *ramo* representa as dezenas. No Minitab há ainda uma fila à esquerda do *ramo* que funciona como contador de suas respectivas *folhas*. Além de mostrar como um conjunto de valores está distribuído, esse diagrama é útil, pois não há perda de informação dos dados.

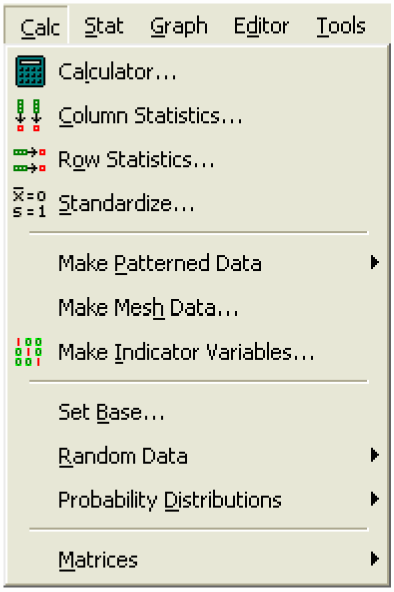
- **Exemplo 3.1.6** – Construa o diagrama de Ramo-e-Folhas para a variável *Nota Matemática 1ª Etapa*.
  - Selecione ***Graph*** → ***Steam-and-Leaf***
  - Janela *Steam-and-Leaf*: Selecione a coluna ***C2 Nota Matemática 1ª Etapa*** para ***Graph variables*** → ***OK***



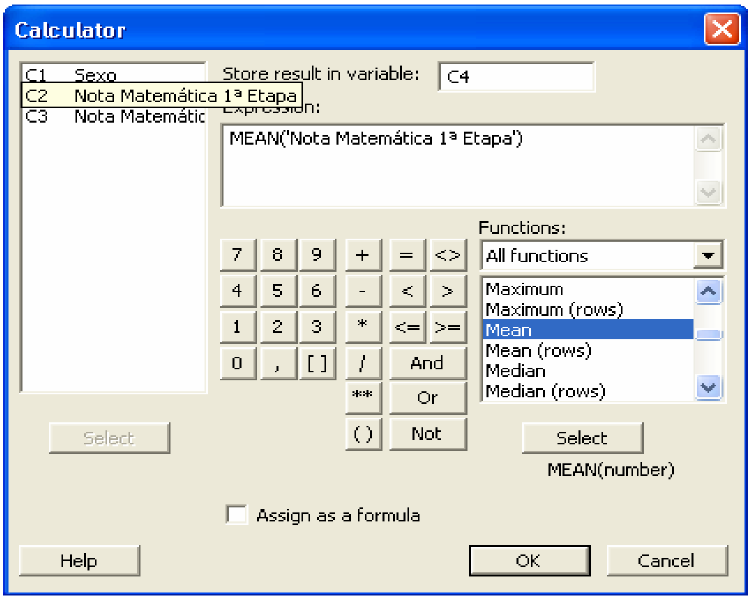


3.2 – Medidas Descritivas

Além do uso de tabelas e gráficos, outra maneira de analisar um conjunto de dados quantitativos é através de medidas resumo ou descritivas, que nada mais são do que números. Tais medidas se classificam como de posição (moda, média, mediana) e dispersão (amplitude, variância, desvio padrão). Mostraremos neste capítulo como obter medidas descritivas usando os menus ***Calc*** e ***Stat*** do Minitab.



- **Exemplo 3.2.1** – Calcule a média da variável *Nota Matemática 1ª Etapa*.
  - Selecione ***Calc*** → ***Calculator***
  - <u>Janela *Calculator*</u>: Selecione ***Mean*** → ***Select***. Depois escolha a coluna ***C2 Nota Matemática 1ª Etapa*** → ***Select***. No campo ***Store result in variable*** indique a coluna onde deseja armazenar o resultado. → ***OK***

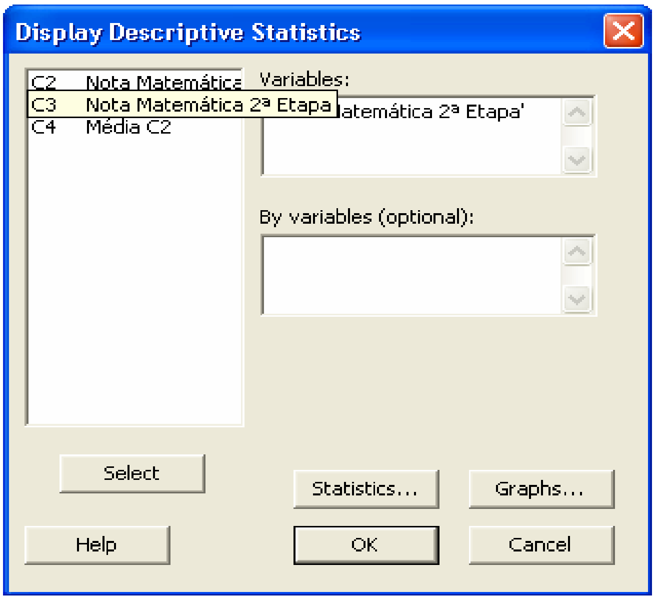


<u>Nota</u>: Na janela ***Calculator*** é possível calcular valores provindo de diversas operações como, por exemplo, aritméticas, estatísticas, trigonométricas, dentre outras.

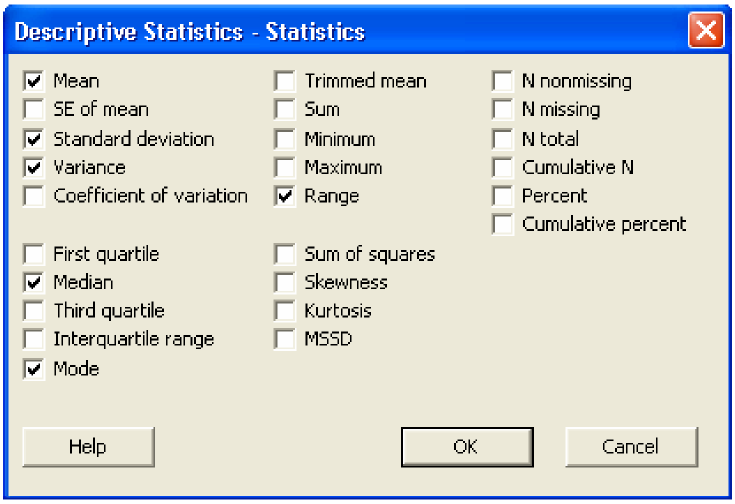
- **Exemplo 3.2.2** – Calcule a moda, mediana, média, amplitude, variância e desvio padrão para a variável *Nota Matemática 2ª Etapa*.

  Ao invés de calcularmos as medidas separadamente na janela ***Calculator*** (como visto no exemplo 3.2.1), optemos por outro caminho:

  - Selecione ***Stat*** → ***Basic Statistics*** → ***Display Descriptive Statistics***
  - <u>Janela *Display Descriptive Statistics*</u>: Selecione a coluna ***C3 Nota Matemática 2ª Etapa*** para o campo ***Variables*** e a seguir clique em ***Statistics***.



- Janela *Descriptive Statistics - Statistics*: Selecione as medidas de interesse; ***Mode*** (moda), ***Median*** (mediana), ***Mean*** (média), ***Range*** (amplitude), ***Variance*** (variância) e ***Standard deviation*** (desvio padrão). → ***OK***



Nota: Observe que na janela ***Descriptive Statistics - Statistics*** é possível encontrar outros valores de interesse como, por exemplo, ***First quartile*** (primeiro quartil), ***Third quartile*** (terceiro quartil), ***Interquartile range*** (intervalo interquartil), etc.

- Janela *Display Descriptive Statistics*: → ***OK***

Session

Descriptive Statistics: Nota Matemática 2ª Etapa

```
                                                                         N for
Variable                 Mean  StDev  Variance  Median  Range  Mode   Mode
Nota Matemática 2ª Etapa  25,82  15,88    252,15   24,00  66,00     0      6
```

Nota: No Minitab também podemos fazer a análise descritiva de um conjunto de dados através de comandos na janela ***Session***.

- Na janela ***Session*** digite o seguinte comando:

  `MTB > Describe C3.` Ou `MTB > Desc C3.`

Session

```
MTB > Describe C3.
```

Descriptive Statistics: Nota Matemática 2ª Etapa

```
Variable                   N   N*   Mean  SE Mean  StDev  Minimum     Q1
Nota Matemática 2ª Etapa  78  119  25,82     1,80  15,88     0,00  14,00

Variable                  Median     Q3  Maximum
Nota Matemática 2ª Etapa   24,00  35,25    66,00
```

```
MTB > Desc C3.
```

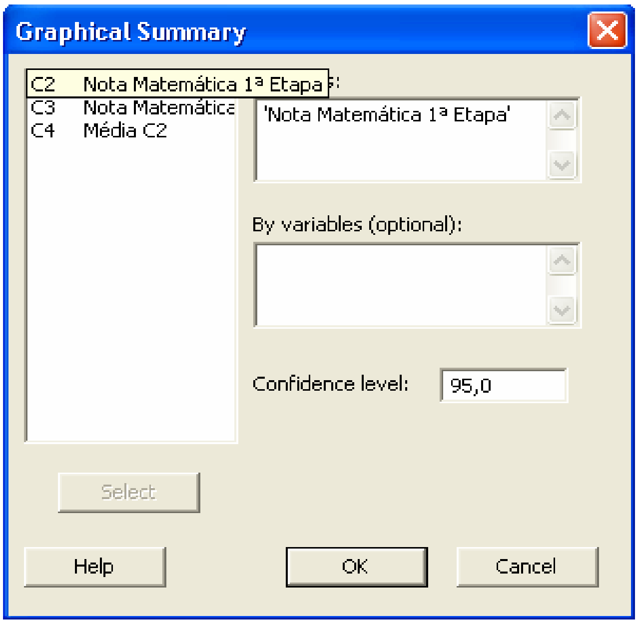
Descriptive Statistics: Nota Matemática 2ª Etapa

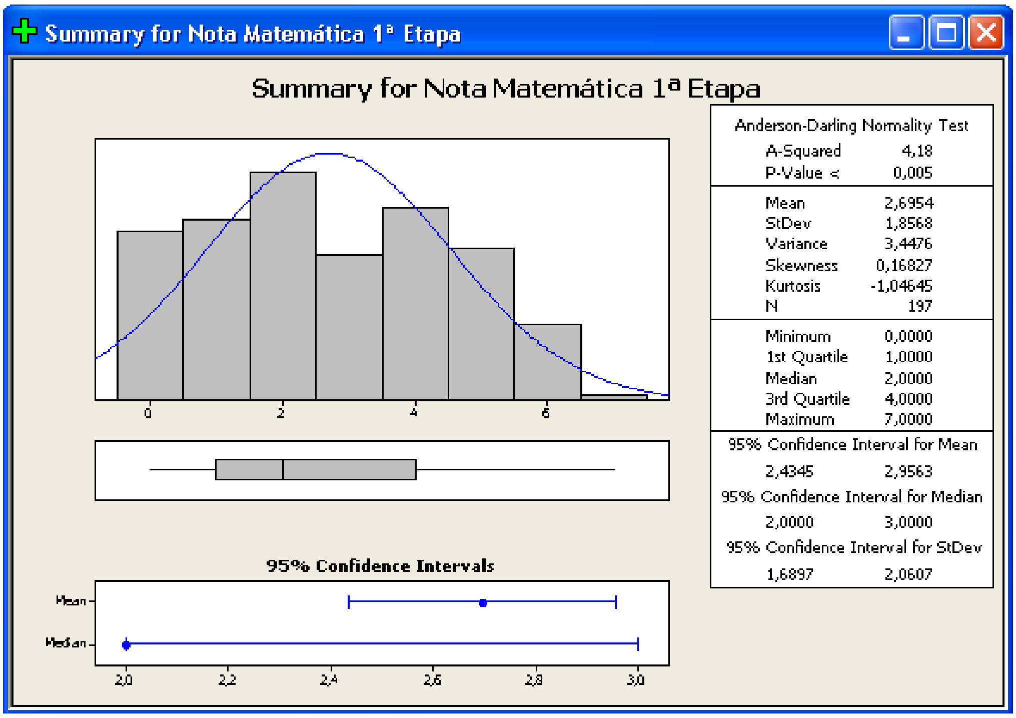
```
Variable                   N   N*   Mean  SE Mean  StDev  Minimum     Q1
Nota Matemática 2ª Etapa  78  119  25,82     1,80  15,88     0,00  14,00

Variable                  Median     Q3  Maximum
Nota Matemática 2ª Etapa   24,00  35,25    66,00
```

- **Exemplo 3.2.3** – Obtenha uma saída que permita a análise exploratória da variável *Nota Matemática 1ª Etapa*.

  - Selecione ***Stat → Basic Statistics → Graphical Summary***

  - Janela *Graphical Summary*: Para o campo ***Variables*** selecione a coluna ***C2 Nota Matemática 1ª Etapa***. Verifique o nível de confiança em ***Confidence level → OK***


Graphical Summary
C2 Nota Matemática 1ª Etapa
C3 Nota Matemática
C4 Média C2
'Nota Matemática 1ª Etapa'
By variables (optional):
Confidence level:
95,0
Select
Help
OK
Cancel


Summary for Nota Matemática 1ª Etapa
Anderson-Darling Normality Test
A-Squared 4,18
P-Value < 0,005
Mean 2,6954
StDev 1,8568
Variance 3,4476
Skewness 0,16827
Kurtosis -1,04645
N 197
Minimum 0,0000
1st Quartile 1,0000
Median 2,0000
3rd Quartile 4,0000
Maximum 7,0000
95% Confidence Interval for Mean
2,4345 2,9563
95% Confidence Interval for Median
2,0000 3,0000
95% Confidence Interval for StDev
1,6897 2,0607
95% Confidence Intervals
Mean
Median

Capítulo 4 – Probabilidade

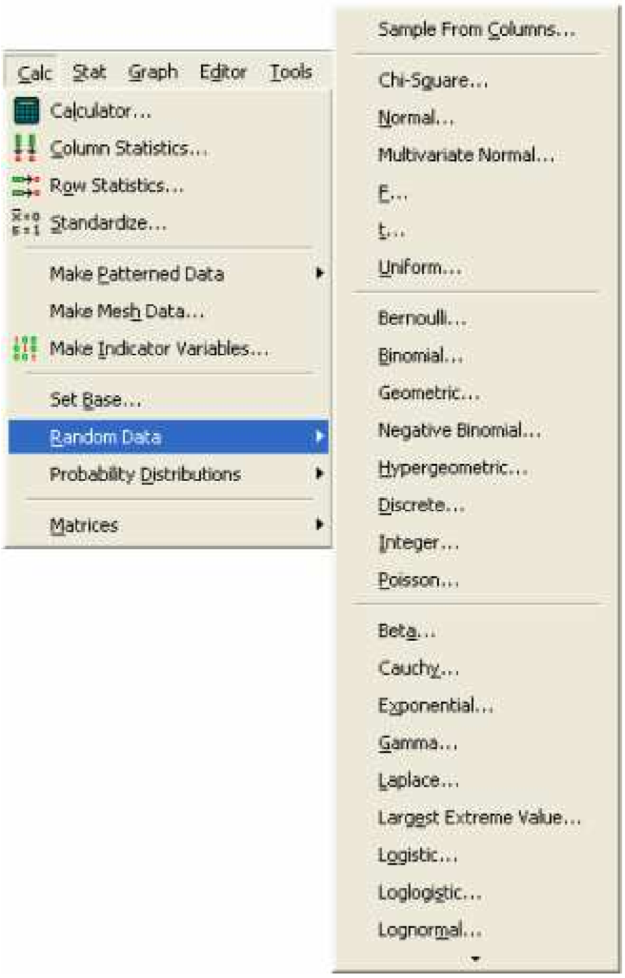
O Minitab atende a certos tópicos comumente abordados em cursos introdutórios de Probabilidade. São estes:

- Geração de Números Aleatórios
- Cálculo de Valores de Distribuições de Probabilidade
- Ajuste de Distribuições

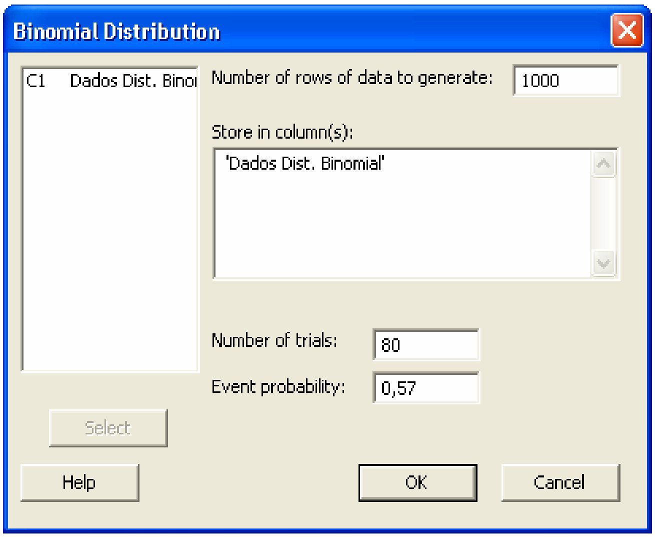
4.1 – Números Aleatórios

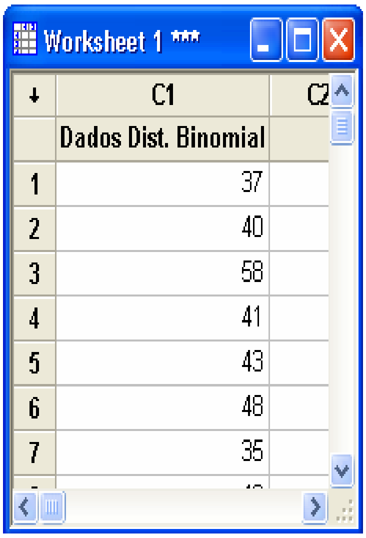
4.1.1 – Geração de Números Aleatórios

Podemos gerar números aleatórios de acordo com um tipo de distribuição de probabilidade. No Minitab essas distribuições encontram-se no menu ***Calc*** → ***Random Data***.



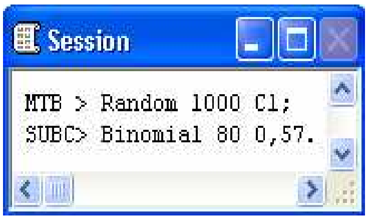
- **Exemplo 4.1.1** – Gere 1000 números aleatórios segundo uma distribuição Binomial com parâmetros $n = 80$ e $p = 0,57$.
    - Selecione ***Calc*** → ***Random Data*** → ***Binomial***
    - Janela *Binomial Distribution*: No campo ***Number of rows of data to generate*** digite a quantidade de números a gerar; no campo ***Store in column(s)*** escolha a coluna em que esses números serão gerados e nos campos ***Number of Trials*** e ***Event probability*** informe o valor dos parâmetros n e p, respectivamente, da distribuição Binomial. → ***OK***





- De forma equivalente, na janela ***Session***, digite os seguintes comandos:

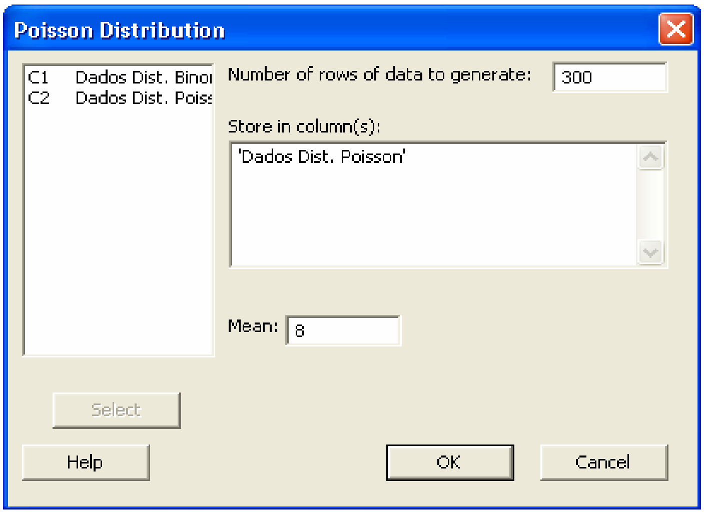
```
MTB > Random 1000 C1;
SUBC> Binomial 80 0,57.
```

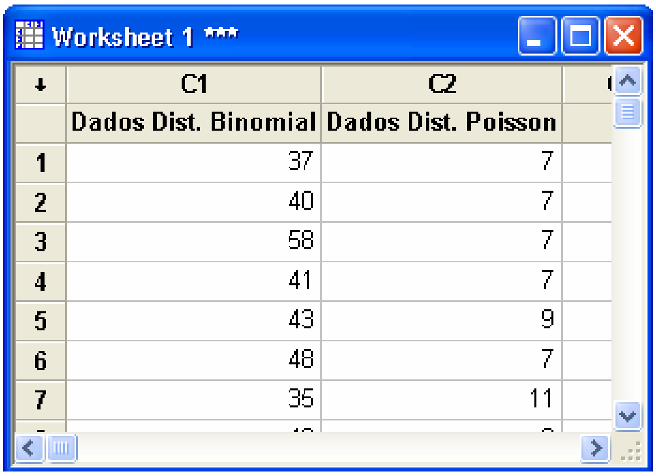


- **Exemplo 4.1.2** – Gere 300 números aleatórios segundo uma distribuição de Poisson com parâmetro λ = 8.

  - Selecione ***Calc*** → ***Random Data*** → ***Poisson***

  - Janela *Poisson Distribution*: No campo ***Number of rows of data to generate*** digite a quantidade de números a gerar; no campo ***Store in column(s)*** escolha a coluna em que esses números serão gerados e no campo ***Mean*** informe o valor da média da distribuição de Poisson (para uma Poisson a média é igual ao valor de seu parâmetro λ). → ***OK***





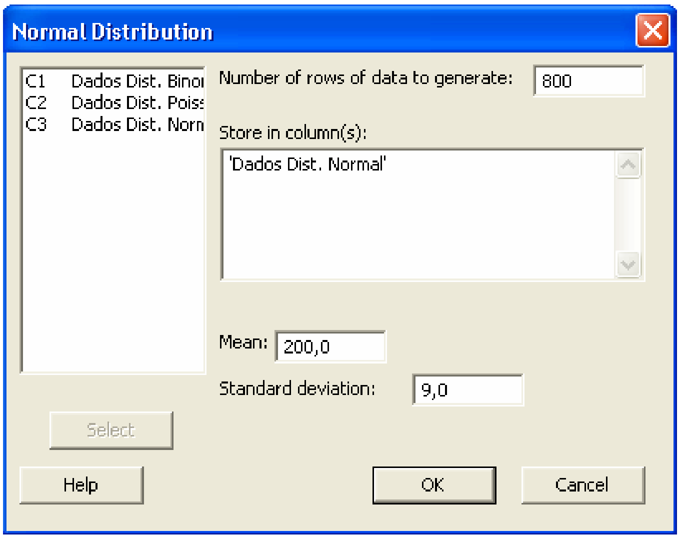
- De modo equivalente, na janela ***Session***, digite os seguintes comandos:

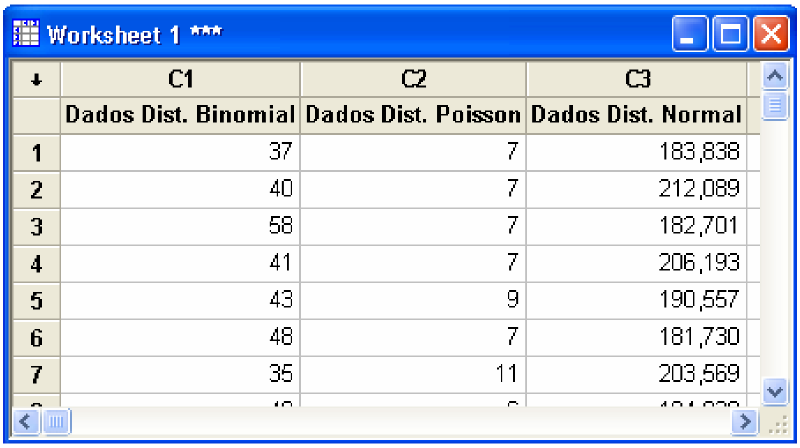
```
MTB > Random 300 C2;
SUBC> Poisson 8.
```

- **Exemplo 4.1.3** – Gere 800 números aleatórios segundo uma distribuição Normal com parâmetros $\mu = 200$ e $\sigma = 9$.

  - Selecione ***Calc*** → ***Random Data*** → ***Normal***

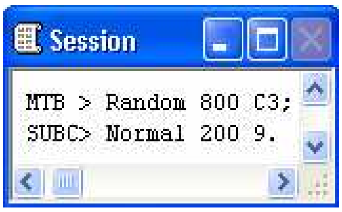
  - <u>Janela *Normal Distribution*</u>: No campo ***Number of rows of data to generate*** digite a quantidade de números a gerar; no campo ***Store in column(s)*** escolha a coluna em que esses números serão gerados e nos campos ***Mean*** e ***Standard deviation*** informe o valor da média e do desvio padrão, respectivamente, da distribuição Normal. → ***OK***





- De forma equivalente, na janela ***Session***, digite os seguintes comandos:

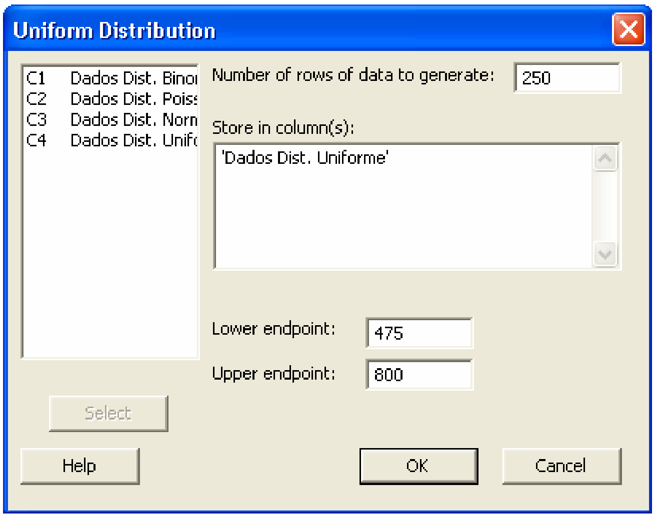
```
MTB > Random 800 C3;
SUBC> Normal 200 9.
```

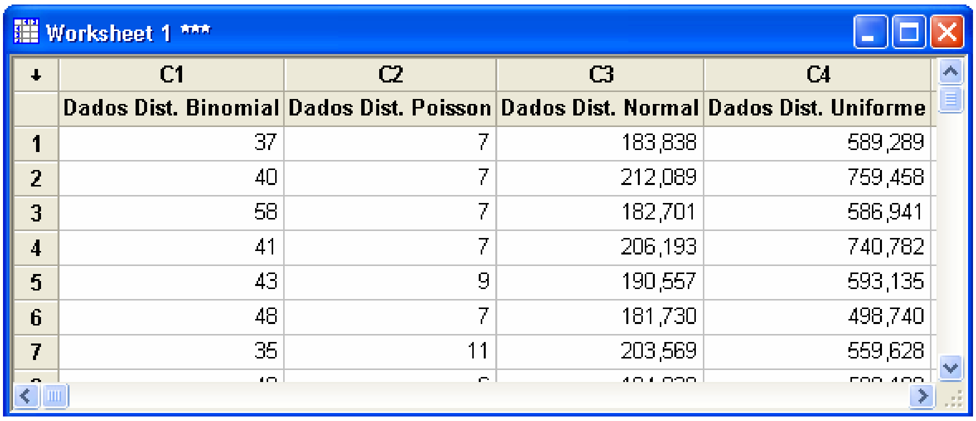


- **Exemplo 4.1.4** – Gere 250 números aleatórios segundo uma distribuição Uniforme com parâmetros $\alpha = 475$ e $\beta = 800$.

  - Selecione ***Calc*** → ***Random Data*** → ***Uniform***

- Janela *Uniform Distribution*: No campo ***Number of rows of data to generate*** digite a quantidade de números a gerar; no campo ***Store in column(s)*** escolha a coluna em que esses números serão gerados e nos campos ***Lower endpoint*** e ***Upper endpoint*** informe o valor dos parâmetros α e β, respectivamente, da distribuição Uniforme. → ***OK***





↓	C1	C2	C3	C4
	Dados Dist. Binomial	Dados Dist. Poisson	Dados Dist. Normal	Dados Dist. Uniforme
1	37	7	183,838	589,289
2	40	7	212,089	759,458
3	58	7	182,701	586,941
4	41	7	206,193	740,782
5	43	9	190,557	593,135
6	48	7	181,730	498,740
7	35	11	203,569	559,628

- De modo equivalente, na janela ***Session***, digite os seguintes comandos:

```
MTB > Random 250 C4;
SUBC> Uniform 475 800.
```

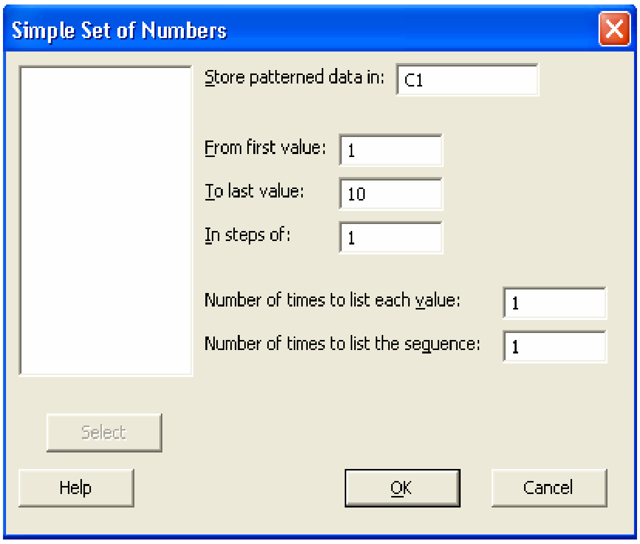
4.1.2 – Amostragem Aleatória Simples

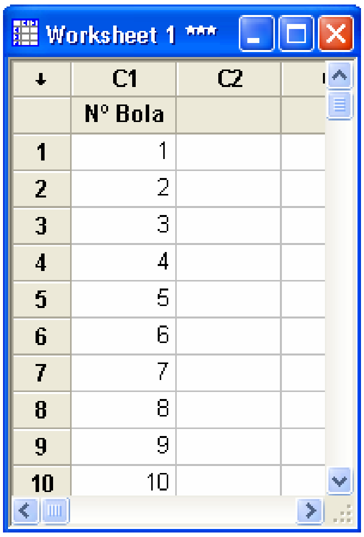
No próximo exemplo mostraremos como o Minitab pode ser usado para obter amostras aleatórias simples, com e sem reposição, de um conjunto de dados.

- **Exemplo 4.1.5** – De uma urna contendo bolas numeradas de 1 a 10, obtenha duas amostras aleatórias de tamanho 5: uma com e outra sem reposição.

  - Selecione ***Calc*** → ***Make Patterned Data*** → ***Simple Set of Numbers*** para primeiro gerarmos números de 1 a 10 correspondentes às bolas da urna.

  - <u>Janela *Simple Set of Numbers*</u>: Em ***Store patterned data in*** escolha a coluna em que os números serão armazenados. Em ***From first value*** e em ***To last value***, digite qual deve ser o primeiro e o último valor a serem gerados, respectivamente. E no campo ***In Steps of*** informe o intervalo de cada número em relação ao seguinte. → ***OK***





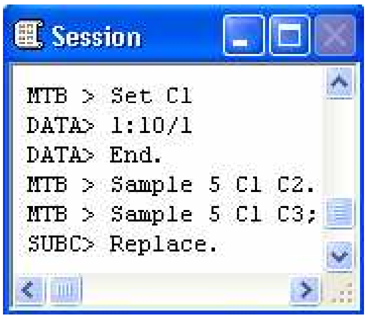
  - Para gerar números de 1 a 10 na janela ***Session*** digite os comandos:

```
MTB > Set C1
DATA> 1:10/1
DATA> End.
```

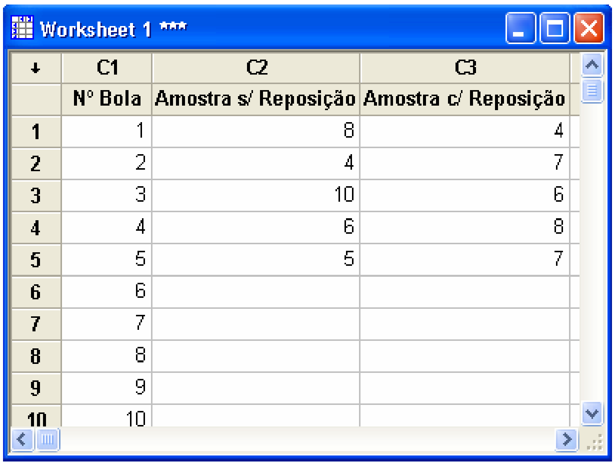
  - Para obter as duas amostras aleatórias de tamanho 5 digite os comandos:

`MTB > Sample 5 C1 C2.` **(Amostra Sem Reposição)**

`MTB > Sample 5 C1 C3;` **(Amostra Com Reposição)**
`SUBC> Replace.`



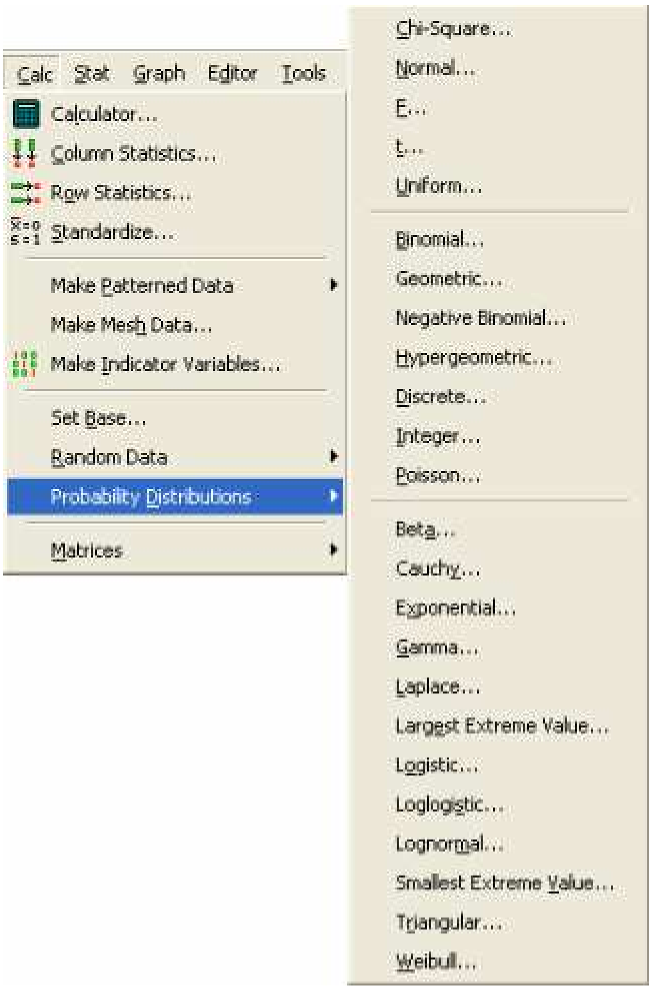
```
MTB > Set C1
DATA> 1:10/1
DATA> End.
MTB > Sample 5 C1 C2.
MTB > Sample 5 C1 C3;
SUBC> Replace.
```

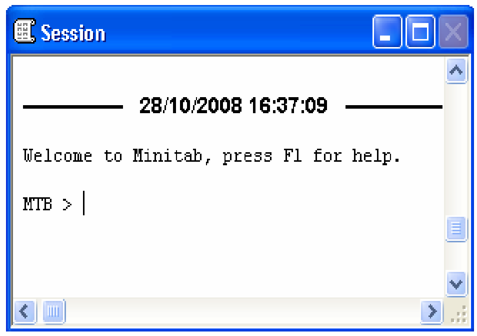


↓	C1	C2	C3
	N° Bola	Amostra s/ Reposição	Amostra c/ Reposição
1	1	8	4
2	2	4	7
3	3	10	6
4	4	6	8
5	5	5	7
6	6		
7	7		
8	8		
9	9		
10	10		

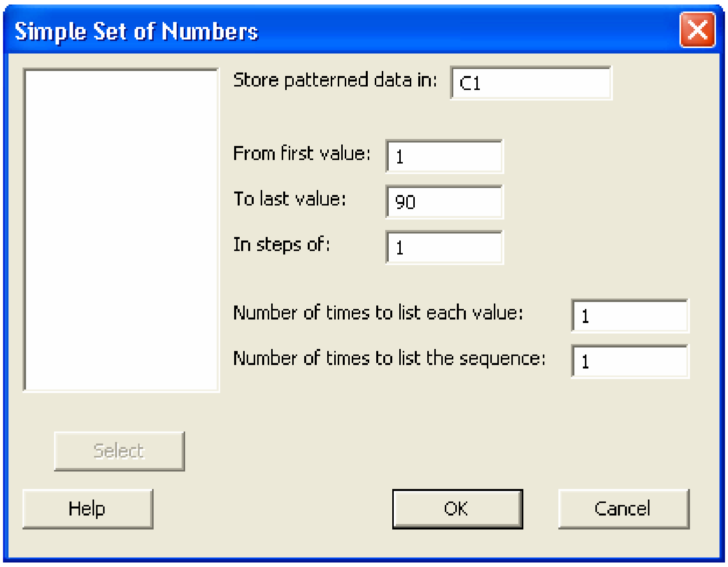
4.2 – Cálculo de Valores de Distribuições de Probabilidade

No Minitab podemos calcular valores de funções de densidade de probabilidade, probabilidade acumulada e probabilidade acumulada inversa para certas distribuições de uma variável aleatória usando o menu ***Calc*** → ***Probability Distributions*** e comandos na janela ***Session***.



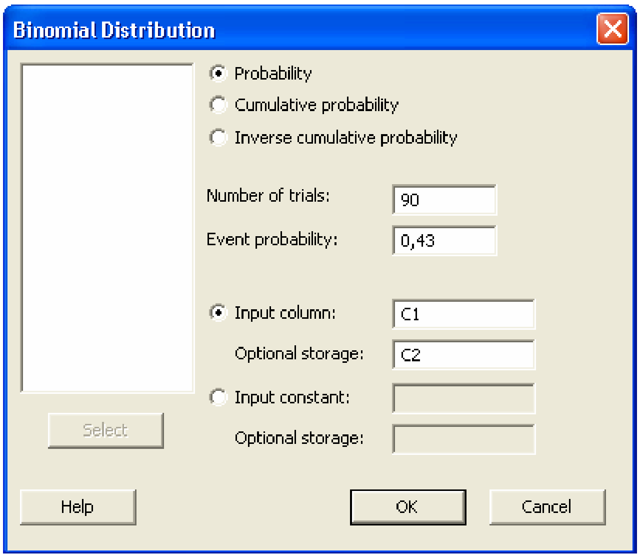


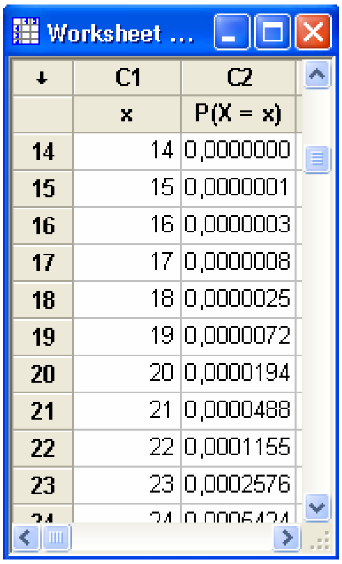
- **Exemplo 4.2.1** – Obtenha os valores da função densidade de probabilidade de uma distribuição Binomial com parâmetros n = 90 e p = 0,43.
  - Selecione ***Calc*** → ***Make Patterned Data*** → ***Simple Set of Numbers*** para primeiro gerarmos números de 1 a 90 correspondentes aos ensaios da distribuição Binomial.
  - Janela *Simple Set of Numbers*: Em ***Store patterned data in*** escolha a coluna em que os números serão armazenados. Em ***From first value*** e em ***To last value***, digite qual deve ser o primeiro e o último valor a serem gerados, respectivamente. E no campo ***In Steps of*** informe o intervalo de cada número em relação ao seguinte. → ***OK***



- Selecione ***Calc*** → ***Probability Distributions*** → ***Binomial***

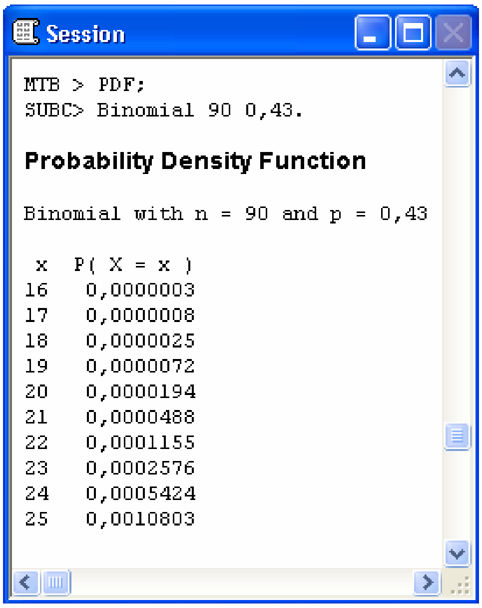
- <u>Janela *Binomial Distribution*</u>: Marque a opção ***Probability*** e em ***Number of trials*** e ***Event Probability*** informe o valor dos parâmetros *n* e *p*, respectivamente, da distribuição Binomial. Em ***Input column*** e em ***Optional storage*** informe, respectivamente, a coluna de entrada e saída da função de probabilidade. → ***OK***





- De forma equivalente, na janela ***Session***, digite os seguintes comandos:

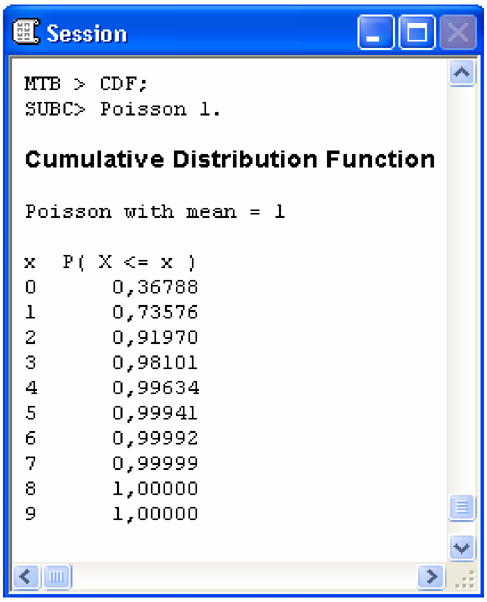
```
MTB > PDF;
SUBC> Binomial 90 0,43.
```



- **Exemplo 4.2.2** – Encontre os valores da função de probabilidade acumulada de uma variável aleatória que tem distribuição de Poisson com parâmetro λ = 1.

  - Na janela ***Session*** digite os seguintes comandos:

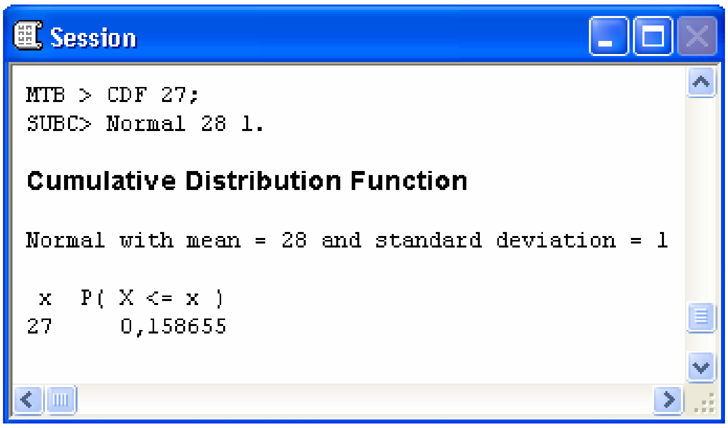
```
MTB > CDF;
SUBC> Poisson 1.
```



- **Exemplo 4.2.3** – Calcule P(X ≤ 27) de uma distribuição Normal com parâmetros $\mu = 28$ e $\sigma = 1$.

  - Na janela ***Session*** digite os seguintes comandos:

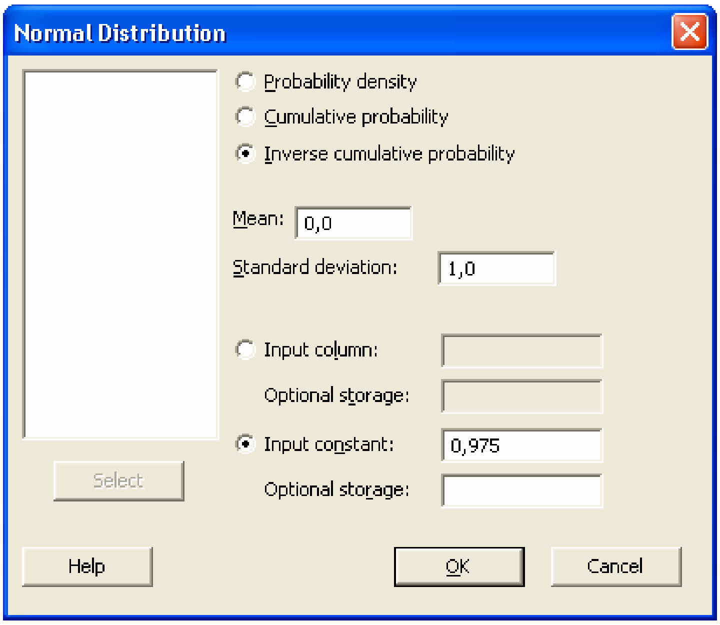
```
MTB > CDF 27;
SUBC> Normal 28 1.
```



- **Exemplo 4.2.4** – Seja *X ~ Normal(0, 1)*. Encontre x tal que P(X ≤ x) = 0,975.

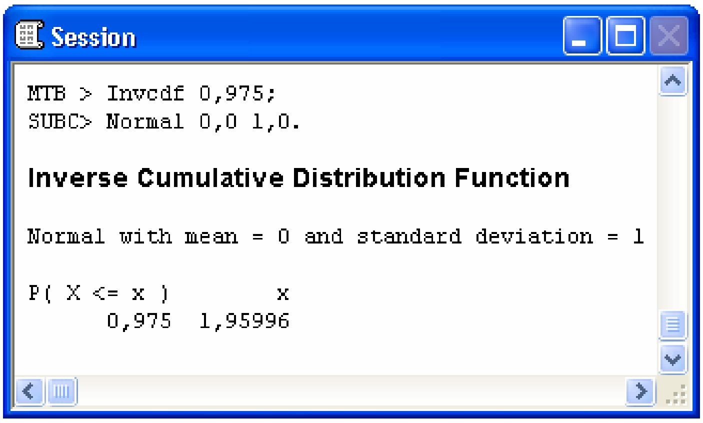
  - Selecione ***Calc*** → ***Probability Distributions*** → ***Normal***

  - Janela Normal Distribution: Marque a opção ***Inverse cumulative probability*** e em ***Mean*** e ***Standard Deviation*** informe o valor da média e do desvio padrão, respectivamente, da distribuição Normal. Marque a opção ***Input constant*** e nesse campo informe o valor da função de probabilidade para o qual estamos procurando a inversa. → ***OK***



- De forma equivalente, na janela ***Session***, digite os seguintes comandos:

```
MTB > Invcdf 0,975;
SUBC> Normal 0,0 1,0.
```

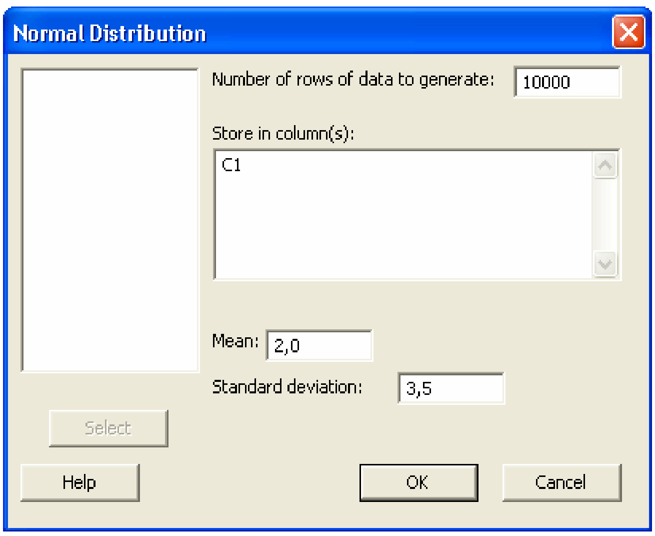


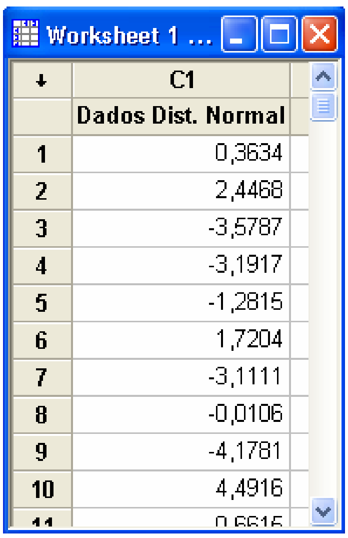
<u>Nota</u>: Este procedimento é útil, por exemplo, em Testes de Hipóteses para encontrar os valores críticos de uma distribuição de referência. Esses mesmos valores também são encontrados em tabelas de distribuições: Normal, Fisher, Qui-quadrado, etc.

- **Exemplo 4.2.5** – Gere 10000 números aleatórios segundo uma distribuição Normal com parâmetros μ = 2 e σ = 3,5. Calcule sua função densidade de probabilidade e faça seu gráfico.

  - Selecione ***Calc*** → ***Random Data*** → ***Normal***

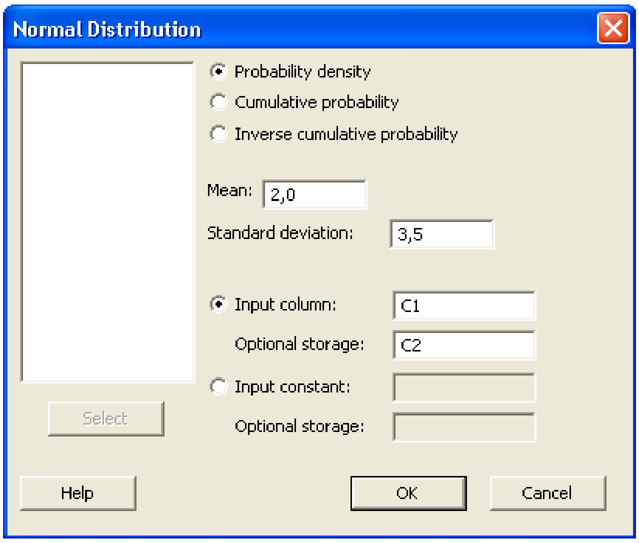
  - <u>Janela *Normal Distribution*</u>: No campo ***Number of rows of data to generate*** digite a quantidade de números a gerar; no campo ***Store in column(s)*** escolha a coluna em que esses números serão gerados e nos campos ***Mean*** e ***Standard deviation*** informe o valor da média e do desvio padrão, respectivamente, da distribuição Normal. → ***OK***

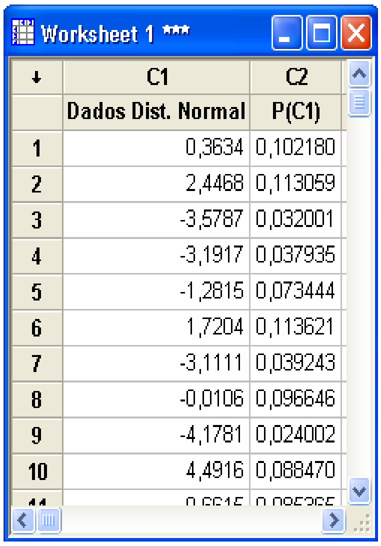




- Selecione ***Calc*** → ***Probability Distributions*** → ***Normal***

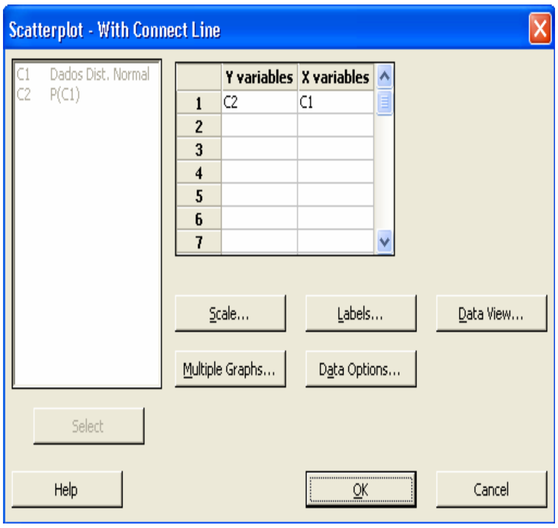
- <u>Janela *Normal Distribution*</u>: Marque a opção ***Probability density*** e nos campos ***Mean*** e ***Standard deviation*** informe o valor dos parâmetros μ e σ, respectivamente, da distribuição Normal. Em ***Input column*** e em ***Optional storage*** informe a coluna de entrada e saída, respectivamente, da função de probabilidade. → ***OK***

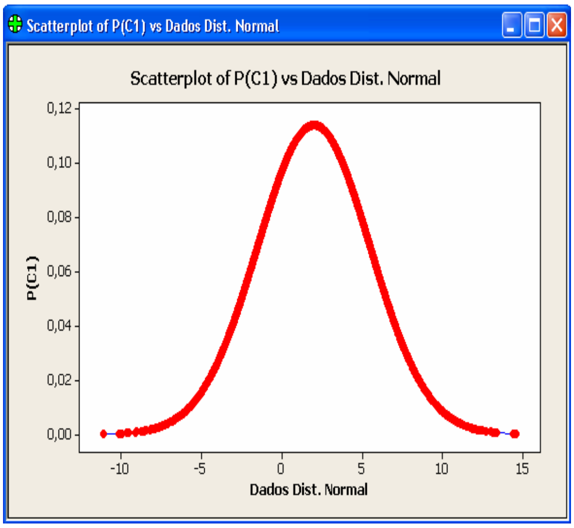




- Selecione ***Graph*** → ***Scatterplot*** → ***With Connect Line*** → ***OK***

- <u>Janela *Scatterplot - With Connect Line*</u>: Selecione as ordenadas e abscissas correspondentes à distribuição Normal. → ***OK***

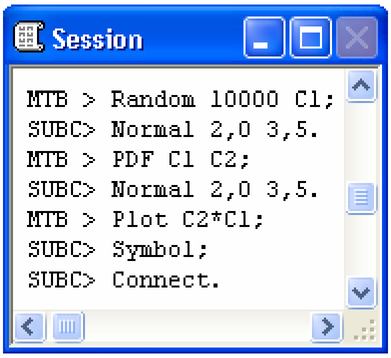




- De forma equivalente, na janela ***Session*** digite os seguintes comandos:

```
MTB > Random 10000 C1;
SUBC> Normal 2,0 3,5.

MTB > PDF C1 C2;
SUBC> Normal 2,0 3,5.
```

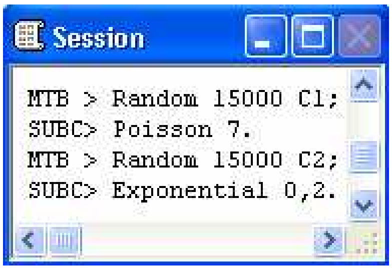
```
MTB > Plot C2*C1;
SUBC> Symbol;
SUBC> Connect.
```

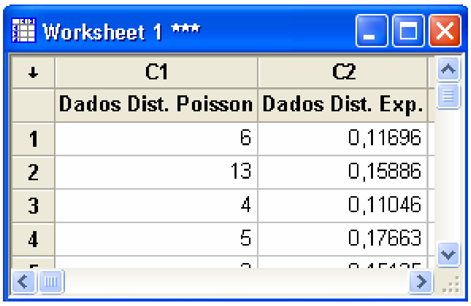


- **Exemplo 4.2.6** – Gere duas amostras de tamanho 15000: uma de Poisson com parâmetro λ = 7 e outra Exponencial com parâmetro ϴ = 5. Em seguida faça o que se pede.

  - Podemos gerar as duas amostras na janela ***Session*** digitando os seguintes comandos:

```
MTB > Random 15000 C1;
SUBC> Poisson 7.

MTB > Random 15000 C2;
SUBC> Exponential 0,2.
```



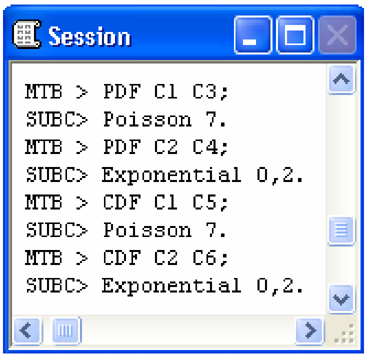


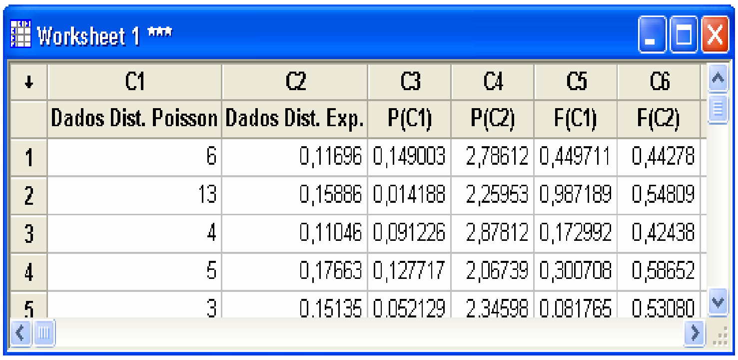
a) Calcule as funções densidade e acumulada das duas amostras.

  - Na janela ***Session*** digite os seguintes comandos:

```
MTB > PDF C1 C3;
SUBC> Poisson 7.

MTB > PDF C2 C4;
SUBC> Exponential 0,2.

MTB > CDF C1 C5;
SUBC> Poisson 7.
```

```
MTB > CDF C2 C6;
SUBC> Exponential 0,2.
```



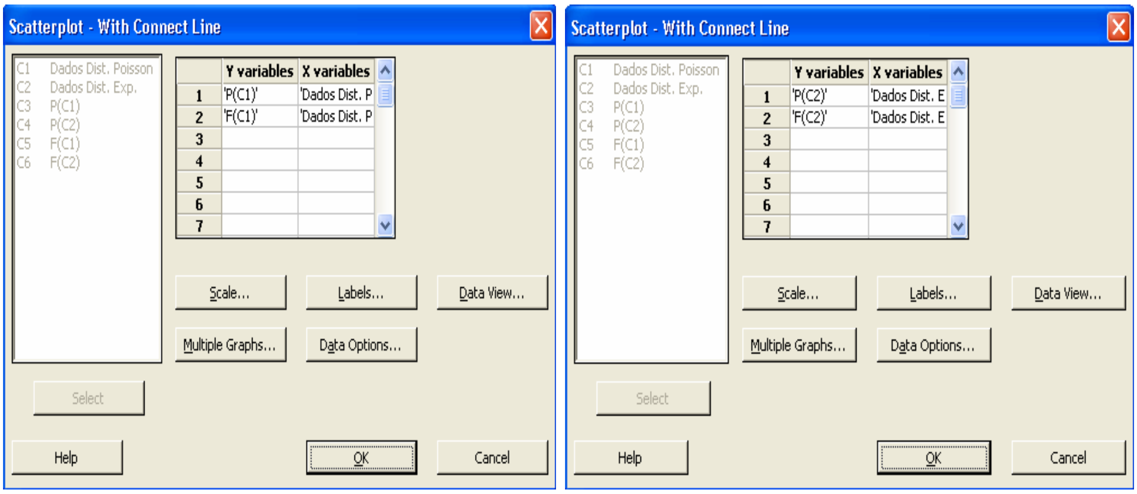


↓	C1	C2	C3	C4	C5	C6
	Dados Dist. Poisson	Dados Dist. Exp.	P(C1)	P(C2)	F(C1)	F(C2)
1	6	0,11696	0,149003	2,78612	0,449711	0,44278
2	13	0,15886	0,014188	2,25953	0,987189	0,54809
3	4	0,11046	0,091226	2,87812	0,172992	0,42438
4	5	0,17663	0,127717	2,06739	0,300708	0,58652
5	3	0.15135	0.052129	2.34598	0.081765	0.53080

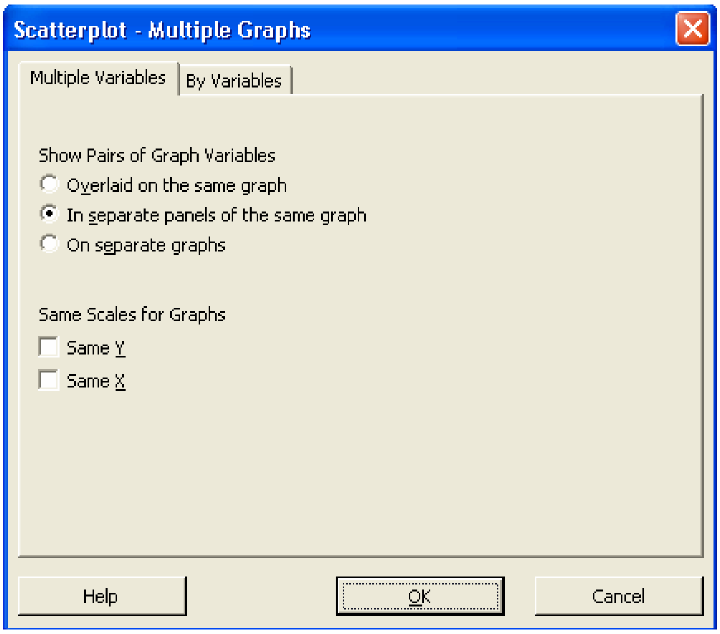
b) Represente num mesmo gráfico as funções densidade e acumulada da amostra de Poisson encontrada no item a. Faça o mesmo para a amostra Exponencial.

- Selecione ***Graph*** → ***Scatterplot*** → ***With Connect Line*** → ***OK***

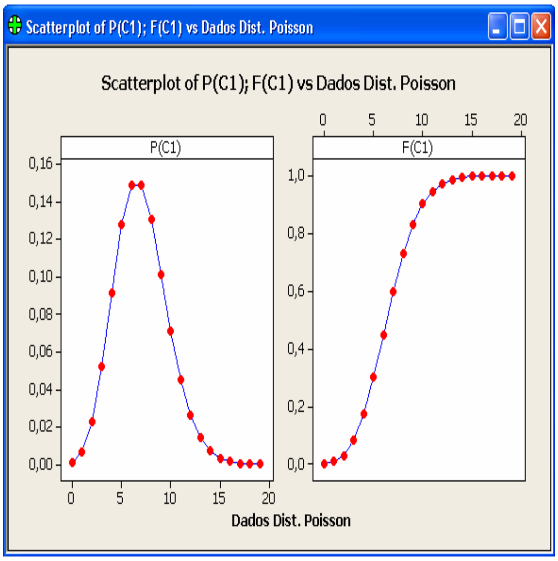
- Janela *Scatterplot - With Connect Line*: Selecione as ordenadas e abscissas correspondentes a cada uma das distribuições e a seguir clique em ***Multiple Graphs***.

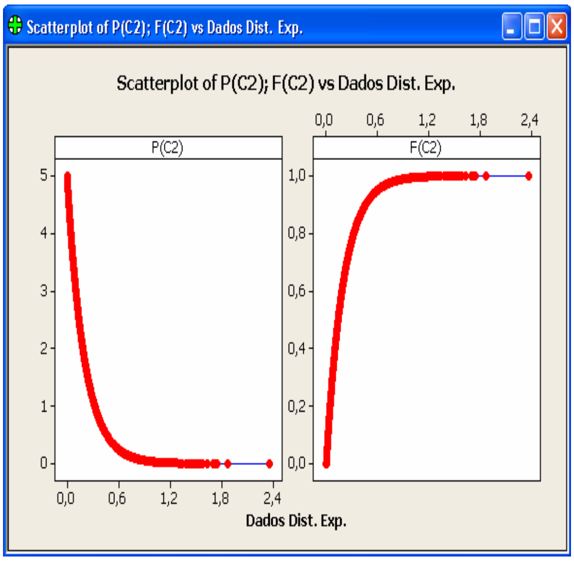


- Janela *Scatterplot - Multiple Graphs*: Selecione a opção ***In separate panels of the same graph*** → ***OK***



- <u>Janela *Scatterplot - With Connect Line*</u>: → ***OK***

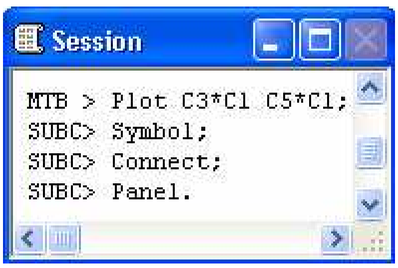




- Ou da mesma forma, na janela ***Session*** digite os seguintes comandos:

```
MTB > Plot C3*C1 C5*C1;
SUBC> Symbol;
SUBC> Connect;
SUBC> Panel.
```

```
MTB > Plot C4*C2 C6*C2;
SUBC> Symbol;
SUBC> Connect;
SUBC> Panel.
```



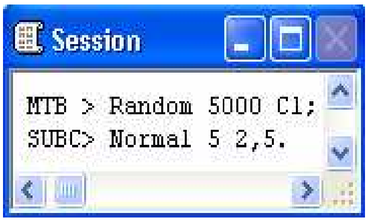
4.3 – Ajuste de Distribuições

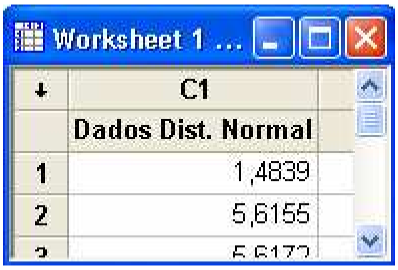
Em determinada fase de um estudo estatístico surge o interesse em saber se um conjunto de dados segue uma distribuição ou modelo probabilístico. A seguir é introduzido como o Minitab auxilia a escolha de tais modelos.

- **Exemplo 4.3.1** – Gere uma amostra de tamanho 5000 segundo uma distribuição Normal com parâmetros μ = 5 e σ = 2,5. A seguir teste se a amostra ajusta uma distribuição Normal.

  - Optemos por gerar os números na janela ***Session***. Para isso, digite os seguintes comandos:

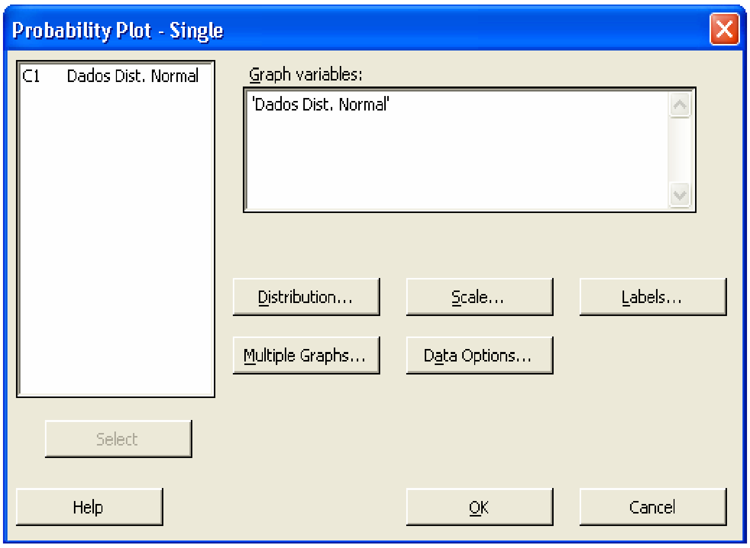
```
MTB > Random 5000 C1;
SUBC> Normal 5 2,5.
```



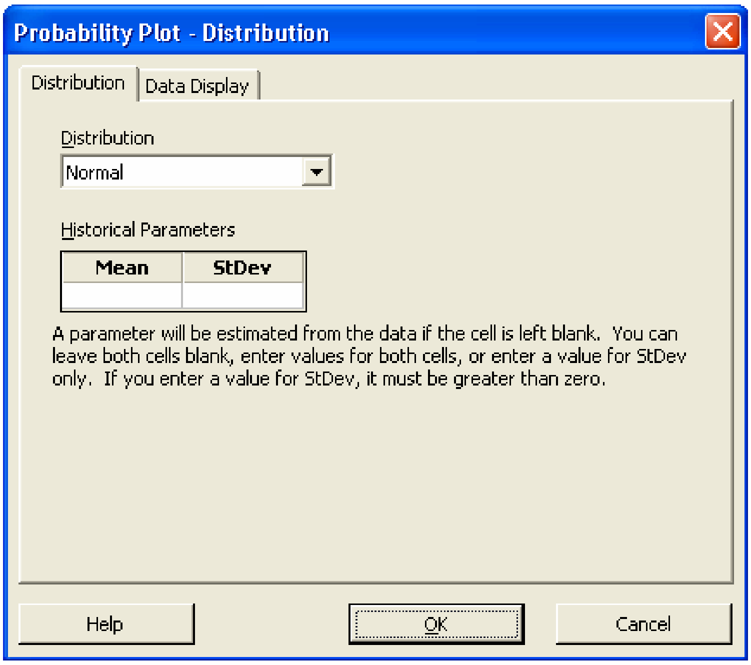


  - Selecione ***Graph*** → ***Probability Distribution Plot*** → ***Single*** → ***OK***

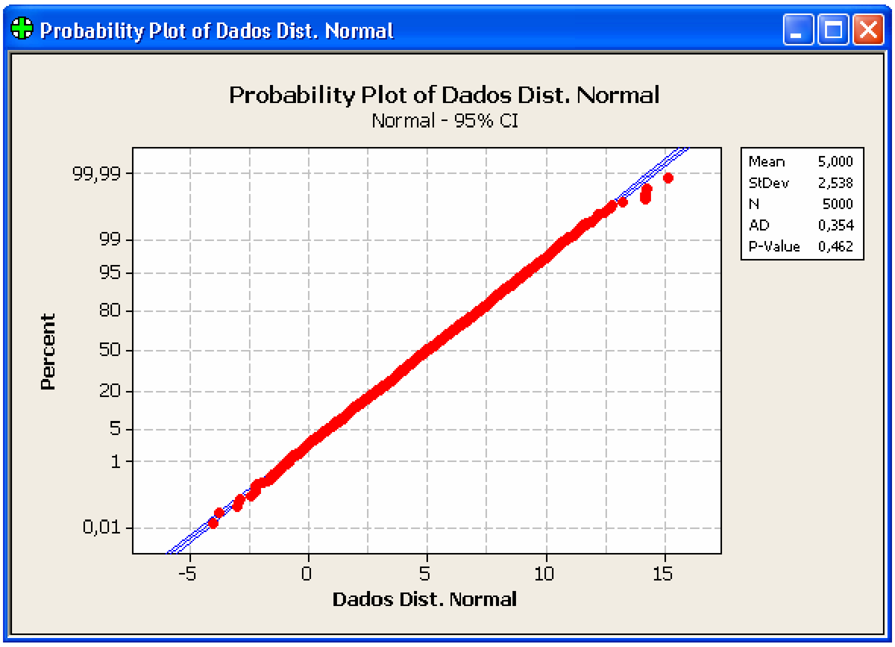
  - <u>Janela *Probability Plot - Single*</u>: No campo ***Graph variables*** escolha a coluna onde estão os dados a serem testados e a seguir clique em ***Distribution***.



  - <u>Janela *Probability Plot - Distribution*</u>: Escolha a distribuição a ser testada para o conjunto de dados em ***Distribution*** → ***OK***



- Janela *Probability Plot - Single*: → ***OK***

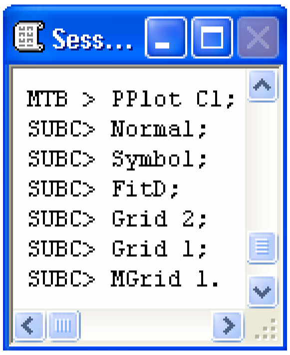


- Na janela ***Session***, para verificar se o conjunto de dados se ajusta bem a uma distribuição Normal, digite os seguintes comandos:

```
MTB > PPlot C1;
SUBC> Normal;
SUBC> Symbol;
SUBC> FitD;
SUBC> Grid 2;
SUBC> Grid 1;
SUBC> MGrid 1.
```

Ou simplesmente

```
MTB > PPlot C1;
SUBC> Normal.
```

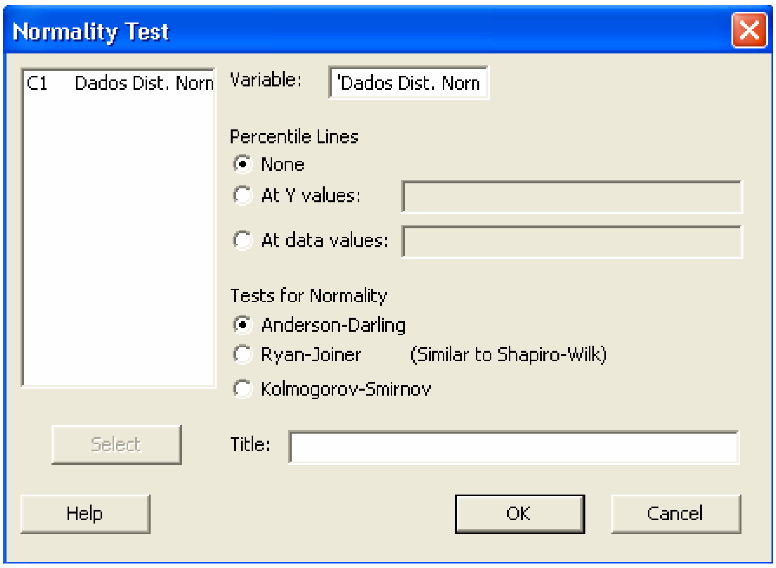


<u>Interpretação dos Resultados</u>: Obtemos que o p-valor = 0,462 > α = 0,05 (nível de significância) e, portanto podemos afirmar que o conjunto de dados se ajusta bem a uma distribuição Normal. Vale notar também, do gráfico, que os pontos estão situados em torno de uma reta.

<u>Nota</u>: A seguir outra forma de testar a normalidade dos dados.

- ***Stat → Basic Statistics → Normality Test***

- <u>Janela *Normality Test*</u>: Em ***Variable*** selecione a coluna onde estão os dados a serem testados e escolha um dos testes de normalidade: ***Anderson-Darling***, ***Ryan-Joiner*** ou ***Kolmogorov-Smirnov***. → ***OK***



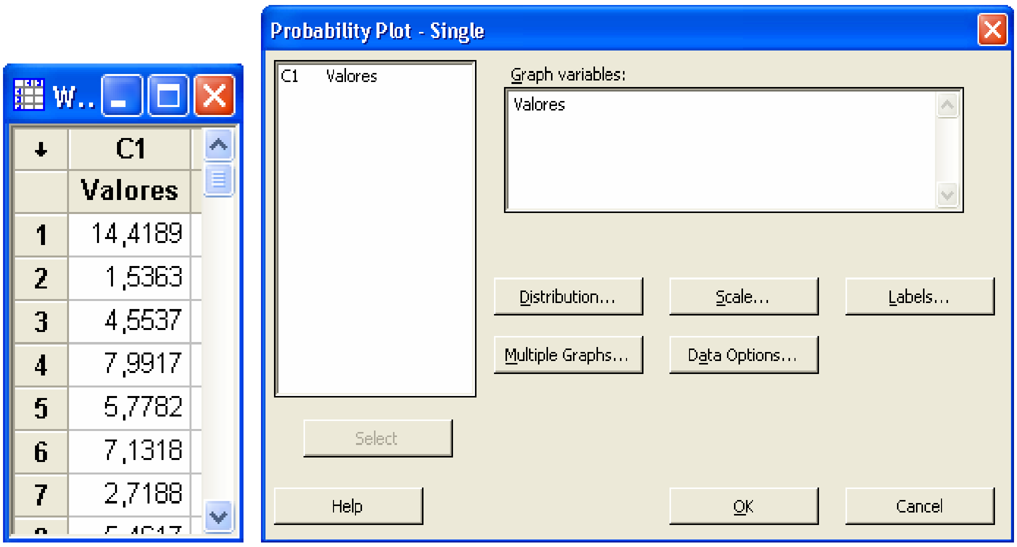
Verifique que chegaremos à mesma conclusão: o conjunto de dados se ajusta bem a uma distribuição Normal.

- **Exemplo 4.3.2** – Teste se os valores abaixo se ajustam bem a uma distribuição Weibull.

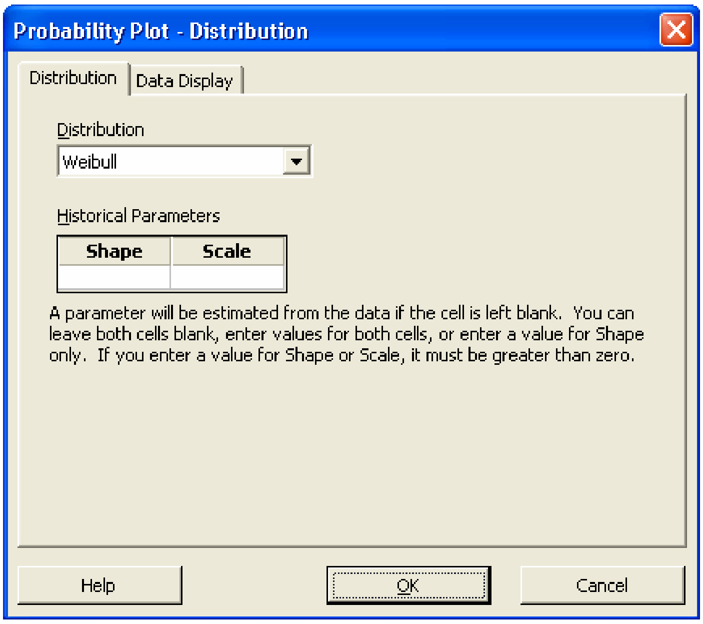
Valores	
14,4189	10,9302
1,53627	4,07240
4,55370	8,55290
7,99175	13,9009
5,77815	1,72682
7,13175	24,1585
2,71879	8,93994
5,46169	5,88014
2,94740	2,85988
16,4731	1,69562
3,15996	14,3234

- Selecione ***Graph*** → ***Probability Distribution Plot*** → ***Single*** → ***OK***

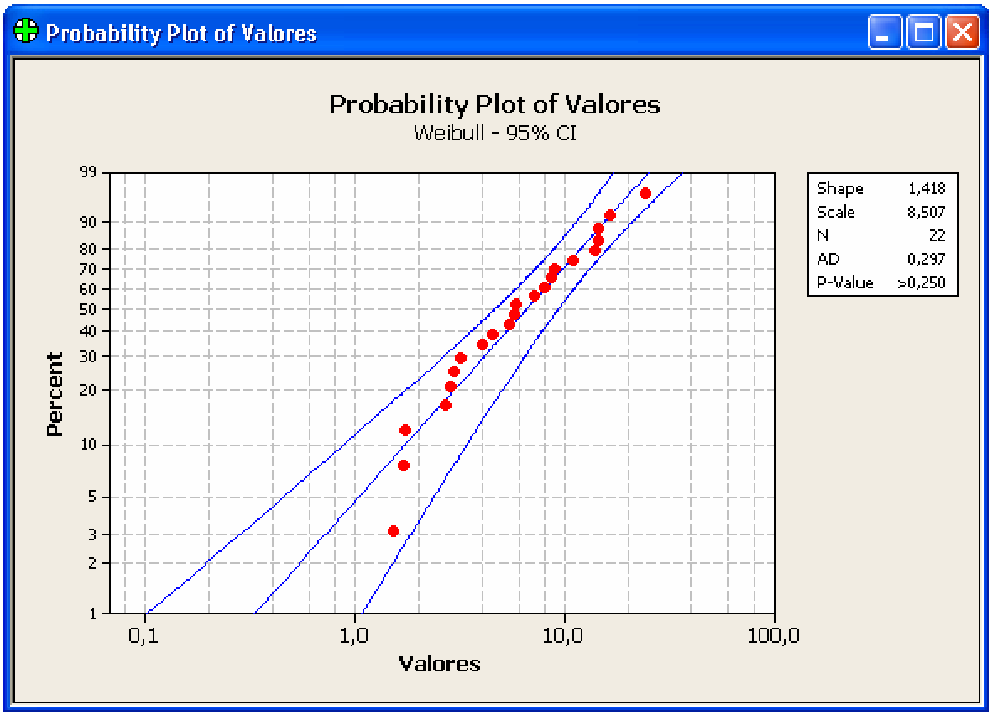
- Janela *Probability Plot - Single*: No campo ***Graph variables*** escolha a coluna onde estão os dados a serem testados e a seguir clique em ***Distribution***.



- Janela *Probability Plot - Distribution*: Escolha a distribuição a ser testada para o conjunto de dados em ***Distribution*** → ***OK***

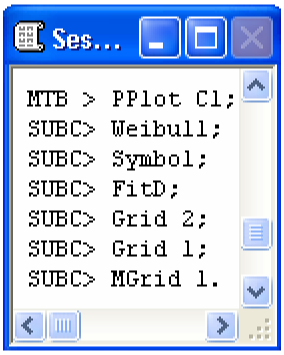


- Janela *Probability Plot - Single*: → ***OK***



- Na janela ***Session***, para verificar se o conjunto de dados se ajusta bem a uma distribuição Weibull, digite os seguintes comandos:

```
MTB > PPlot C1;
SUBC> Weibull;
SUBC> Symbol;
SUBC> FitD;
SUBC> Grid 2;
SUBC> Grid 1;
SUBC> MGrid 1.
```



Ou simplesmente

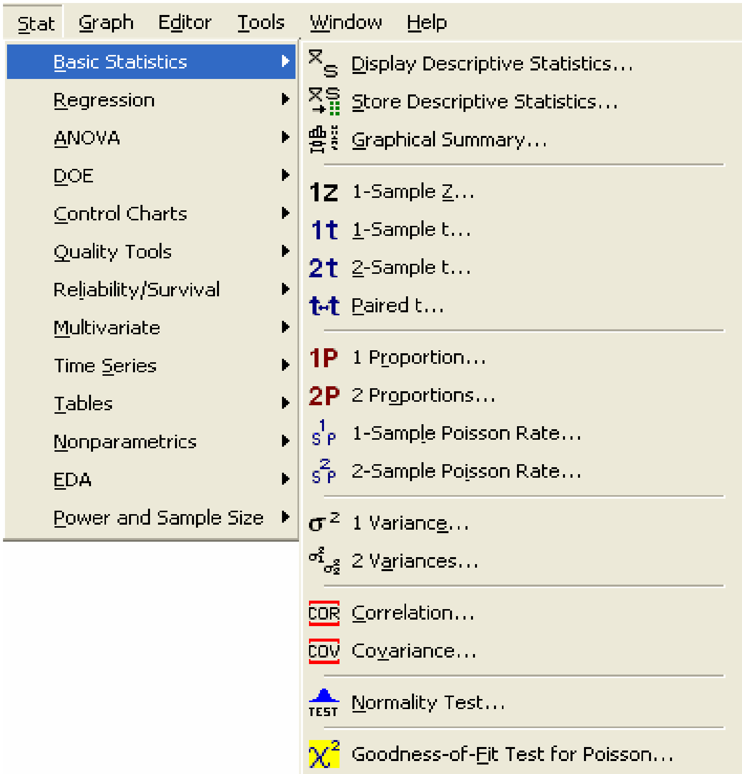
```
MTB > PPlot C1;
SUBC> Weibull.
```

<u>Interpretação dos Resultados</u>: Obtemos que o p-valor > 0,250 > α = 0,05 (nível de significância) e, portanto podemos afirmar que o conjunto de dados se ajusta bem a uma distribuição Weibull. Vale notar também, do gráfico, que os pontos estão situados em torno de uma reta.

Capítulo 5 – Teste de Hipóteses

Neste capítulo iremos introduzir os procedimentos e recursos básicos do Minitab para fazer inferências sobre um parâmetro de uma população.

As opções de testes a serem utilizados encontram-se no menu ***Stat*** → ***Basic Statistics***.



5.1 – Teste para a Média Populacional μ com σ conhecido

- **Exemplo 5.1.1** – A seguir temos nove medidas de um produto. É de conhecimento que essas medidas seguem uma distribuição Normal com σ = 0,2. Testar se a média das medidas é igual a 5.

Medidas do Produto								
4,9	5,1	4,6	5	5,1	4,7	4,4	4,7	4,6

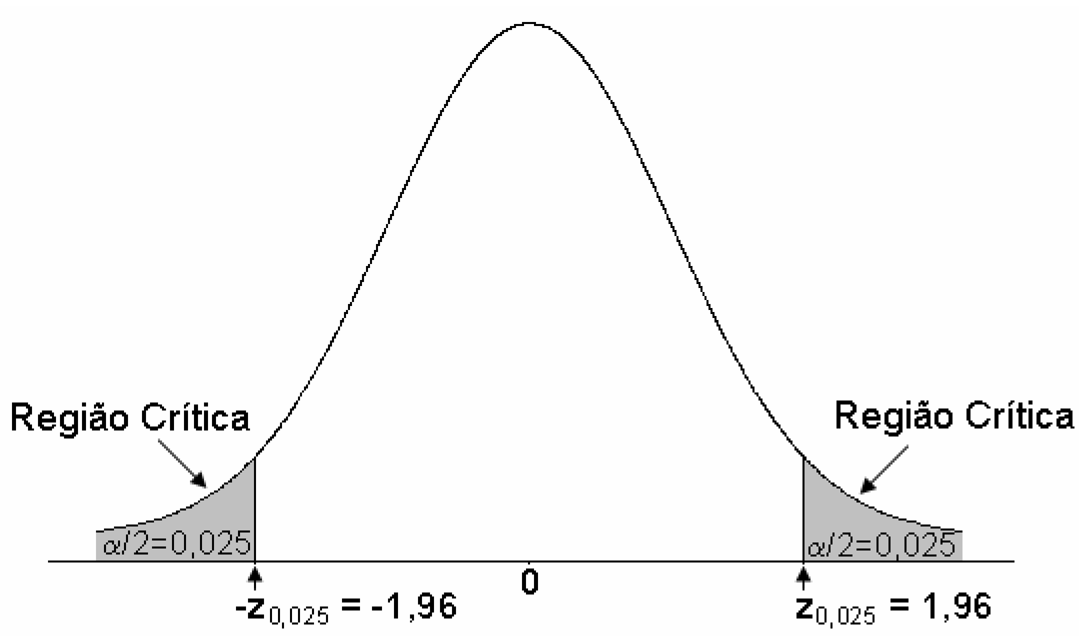
Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu = 5 \\ H_1: \mu \neq 5 \end{cases}$ Nível de Significância: α = 0,05

Estatística de Teste:

$$z = \frac{\overline{x} - \mu}{(\sigma/\sqrt{n})}$$

Valores Críticos[2]: $z_{0,025}$ = 1,96 e $-z_{0,025}$ = -1,96 (Valores Tabelados)



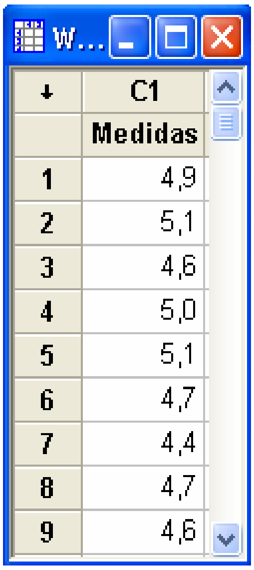
Usando o Minitab:

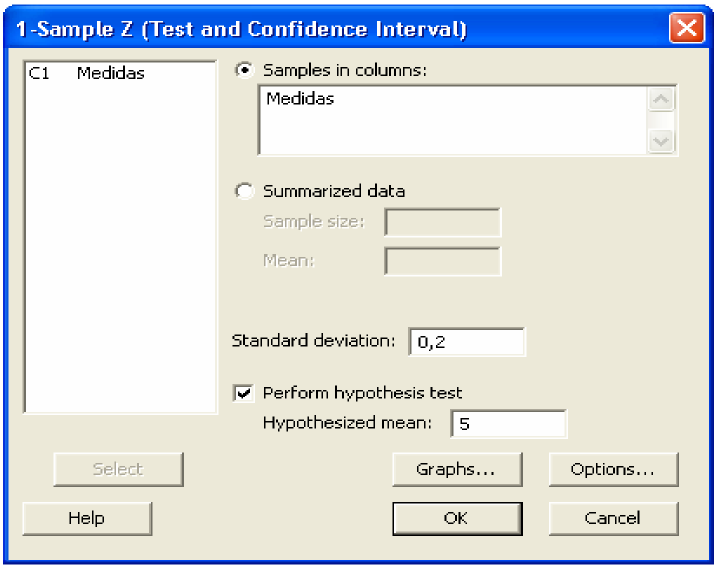
- Selecione ***Stat*** → ***Basic Statistics*** → ***1-Sample Z***

- Janela *1-Sample Z*: Selecione ***Samples in Columns*** (amostras em colunas) e escolha a coluna da amostra a ser testada. Introduza o valor

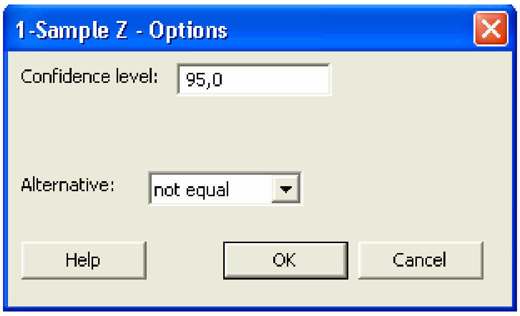
[2] No Minitab, valores críticos de uma distribuição de referência podem ser calculados usando a função de probabilidade acumulada inversa como visto no exemplo 4.2.4 do Capítulo 4.

de σ em ***Standard deviation*** (desvio padrão); marque a opção ***Perform hypothesis test*** (fazer teste de hipóteses) e informe o valor da média populacional μ em ***Hypothesized mean***. A seguir clique em ***Options***.

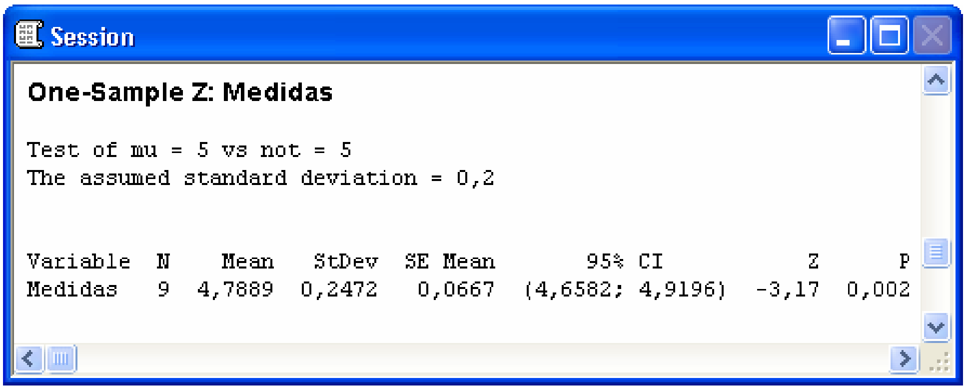




- <u>Janela *1-Sample Z - Options*</u>: Introduza o nível de confiança em ***Confidence Level*** e para ***Alternative*** (hipótese alternativa), selecione ***not equal*** (teste bilateral: $\mu \neq \mu_0$). → ***OK***



- <u>Janela *1-Sample Z*</u>: → ***OK***



<u>Resultados</u>:

- Como o valor de z_{obs} = -3,17 pertence à região crítica: {z < -1,96 ou z > 1,96}, rejeitamos H_0 ao nível de 5% de significância.

- Como o valor-p = 0,002 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância[3].
- Como o intervalo de 95% de confiança não contém o valor de μ sob H_0 (μ = 5), rejeitamos H_0 ao nível de 5% de significância.

Conclusão: Há evidência estatística suficiente para rejeitarmos a afirmativa de que a média populacional seja igual a 5.

- **Exemplo 5.1.2** – A tensão de ruptura de cabos produzidos por um fabricante apresenta média (μ) de 1800 kg e desvio padrão (σ) de 100 kg. Mediante nova técnica de produção, proclamou-se que a tensão de ruptura pode ter aumentado. Para testar essa declaração, selecionou-se uma amostra de 50 cabos, chegando-se a uma média amostral de 1830 kg. Pode-se confirmar a declaração ao nível de 5% de significância?

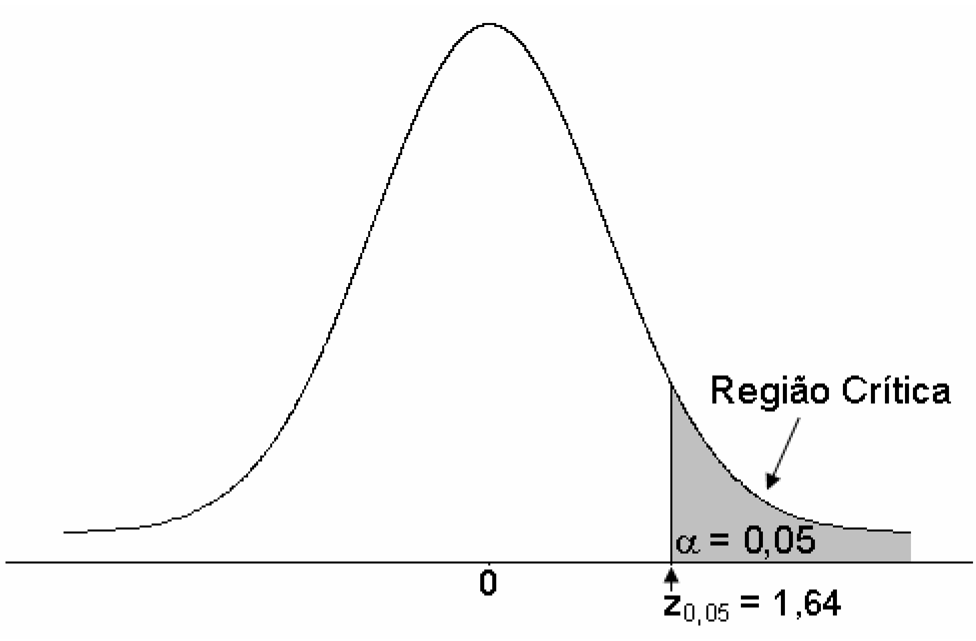
Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu = 1800 \\ H_1: \mu > 1800 \end{cases}$ Nível de Significância: α = 0,05

Estatística de Teste:

$$z = \frac{\overline{x} - \mu}{(\sigma/\sqrt{n})}$$

Valor Crítico: $z_{0,05}$ = 1,64 (Valor Tabelado)

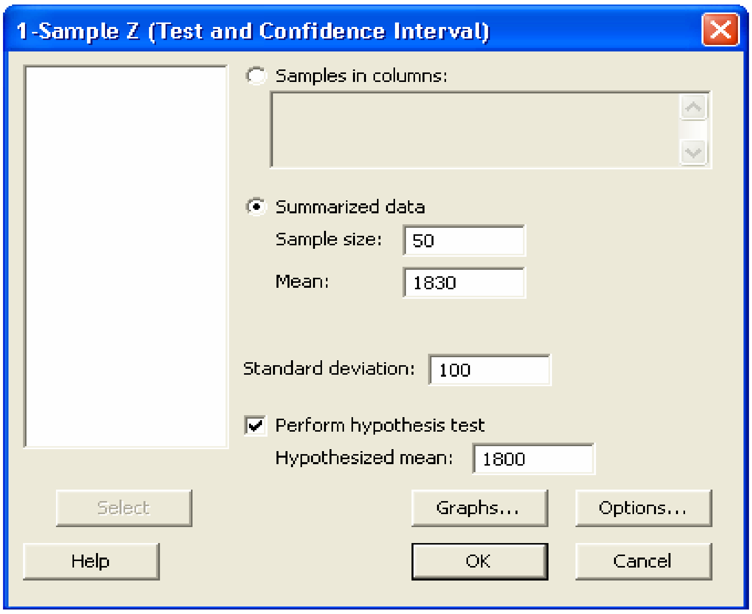


Usando o Minitab:

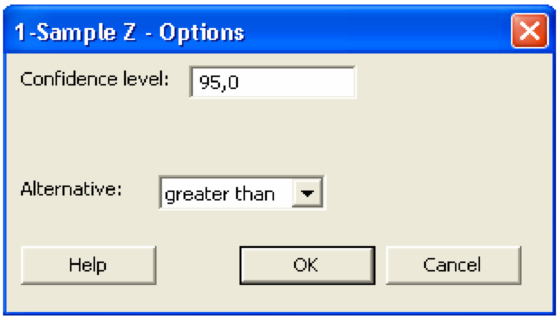
[3] Em linhas gerais rejeitamos H_0 a um nível de significância maior que o valor-p encontrado.

- Selecione ***Stat*** → ***Basic Statistics*** → ***1-Sample Z***

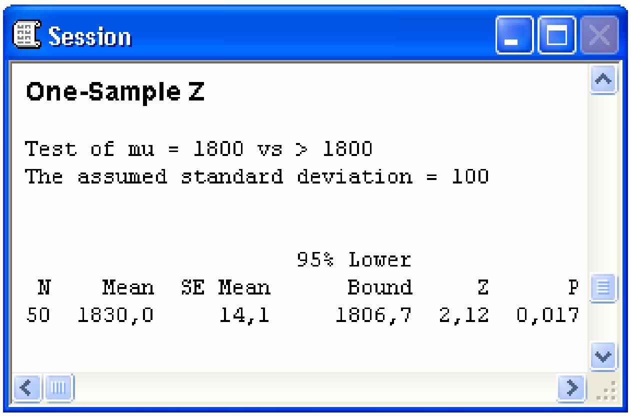
- Janela *1-Sample Z*: Selecione ***Summarized data*** e preencha os campos ***Sample size*** (tamanho da amostra), ***Mean*** (média amostral) e ***Standard deviation*** (desvio padrão). Marque a opção **Perform hypothesis test** (fazer teste de hipóteses) e informe o valor da média populacional μ em ***Hypothesized mean***. A seguir clique em ***Options***.



- Janela *1-Sample Z - Options*: Introduza o nível de confiança em ***Confidence Level*** e para ***Alternative*** (hipótese alternativa), selecione ***greater than*** (teste unilateral: $\mu > \mu_0$). → ***OK***



- Janela *1-Sample Z*: → ***OK***



N	Mean	SE Mean	95% Lower Bound	Z	P
50	1830,0	14,1	1806,7	2,12	0,017

<u>Resultados</u>:

- Como o valor de z_{obs} = 2,12 pertence à região crítica: {z > 1,96}, rejeitamos H_0 ao nível de 5% de significância.
- Como o valor-p = 0,017 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.
- Como o limite inferior do intervalo de 95% de confiança é maior que o valor de μ sob H_0 (μ = 1800), rejeitamos H_0 ao nível de 5% de significância.

<u>Conclusão</u>: A nova técnica produz cabos com uma tensão de ruptura média (μ) significativamente diferente de 1800 kg.

5.2 – Teste para a Média Populacional μ com σ desconhecido

- **Exemplo 5.2.1** – Doze latas de lubrificante de certa marca acusam os seguintes conteúdos médios (em decilitros):

Conteúdos Médios											
10,2	9,7	10,1	10,3	10,1	9,8	9,9	10,4	10,3	9,8	10,4	10,2

Ao nível de 1%, testar a hipótese de que o conteúdo médio das latas do lubrificante é igual a 10 dl. Assumir que os dados seguem uma distribuição normal.

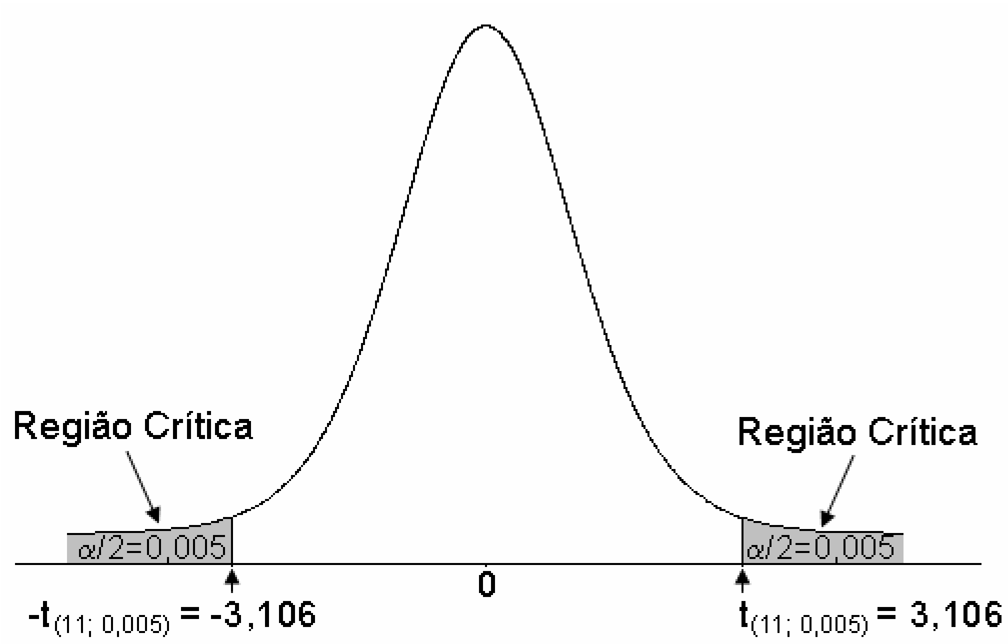
Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu = 10 \\ H_1: \mu \neq 10 \end{cases}$ Nível de Significância: $\alpha = 0{,}01$

Estatística de Teste: t com 11 graus de liberdade

$$t = \frac{\overline{x} - \mu}{(s/\sqrt{n})}$$

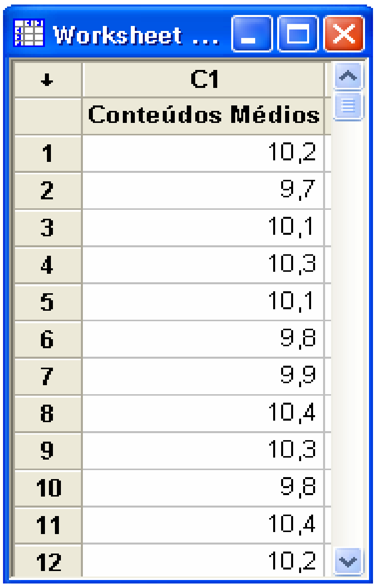
Valores Críticos: $t_{(11;\,0{,}005)} = 3{,}106$ e $-t_{(11;\,0{,}005)} = -3{,}106$ (Valores Tabelados)

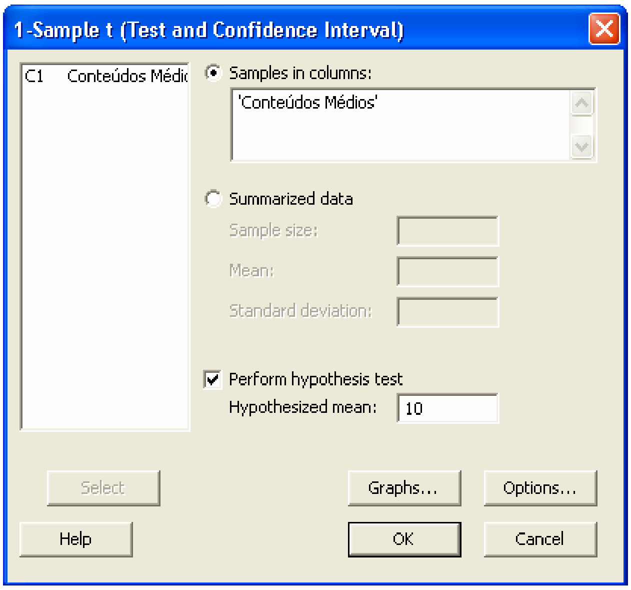


Usando o Minitab:

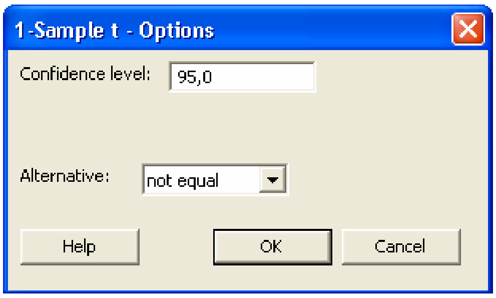
- Selecione ***Stat*** → ***Basic Statistics*** → ***1-Sample t***

- Janela *1-Sample t*: Selecione ***Samples in Columns*** (amostras em colunas) e escolha a coluna da amostra a ser testada. Marque a opção ***Perform hypothesis test*** (fazer teste de hipóteses) e informe o valor da média populacional μ em ***Hypothesized mean***. A seguir clique em ***Options***.

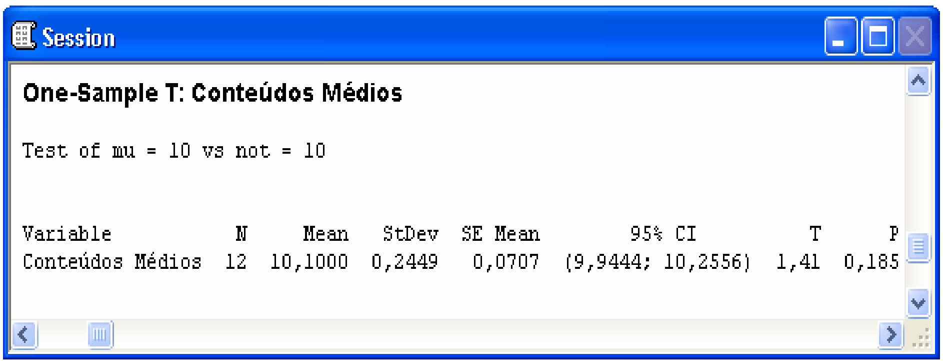




- Janela *1-Sample t - Options*: Introduza o nível de confiança em ***Confidence Level*** e para ***Alternative*** (hipótese alternativa), selecione ***not equal*** (teste bilateral: $\mu \neq \mu_0$). → ***OK***



- Janela *1-Sample t*: → ***OK***



Resultados:

- Como o valor de t_{obs} = 1,41 não pertence à região crítica: {t < -3,106 ou t > 3,106}, deixamos de rejeitar H_0 ao nível de 1% de significância.
- Como o valor-p = 0,185 > α = 0,01, deixamos de rejeitar H_0 ao nível de 1% de significância.
- Como o intervalo de 99% de confiança contém o valor de μ sob H_0 (μ = 10), deixamos de rejeitar H_0 ao nível de 1% de significância.

Conclusão: Não há evidência suficiente para rejeitar a afirmativa que o conteúdo médio das latas de lubrificante seja igual a 10 dl.

- **Exemplo 5.2.2** – Um artigo do *New York Times* observou que a vida média de 35 regentes homens de orquestras sinfônicas era de 73,4 anos, em contraste com a vida média de 69,5 anos para os homens da população em geral. Supondo que os 35 regentes tenham tempo médio de vida com um desvio padrão de 8,7 anos, use o nível de significância 0,05 para testar a afirmativa de que regentes homens de orquestras sinfônicas têm tempo médio de vida maior do que 69,5 anos. Assumir que a variável de interesse segue uma distribuição normal.

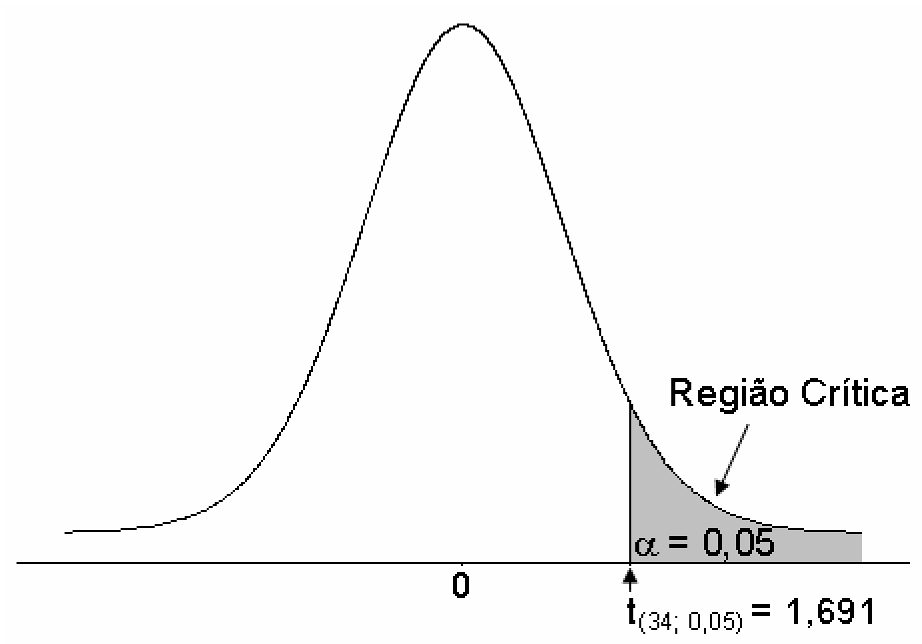
Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu = 69,5 \\ H_1: \mu > 69,5 \end{cases}$ Nível de Significância: α = 0,05

Estatística de Teste: t com 34 graus de liberdade

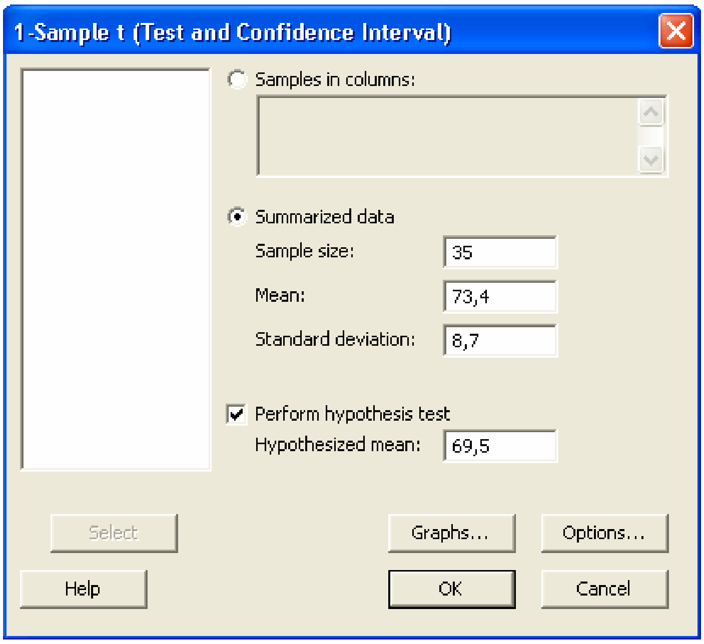
$$t = \frac{\overline{x} - \mu}{(s/\sqrt{n})}$$

Valor Crítico: $t_{(34;\ 0,05)}$ = 1,691 (Valor Tabelado)

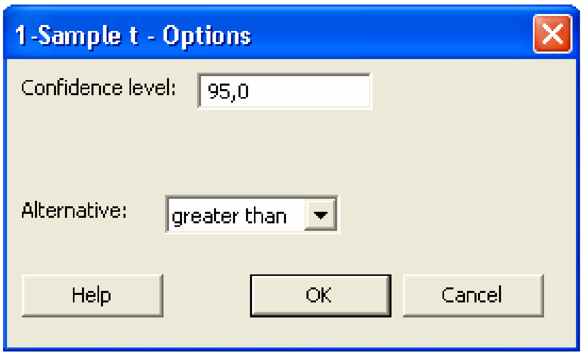


Usando o Minitab:

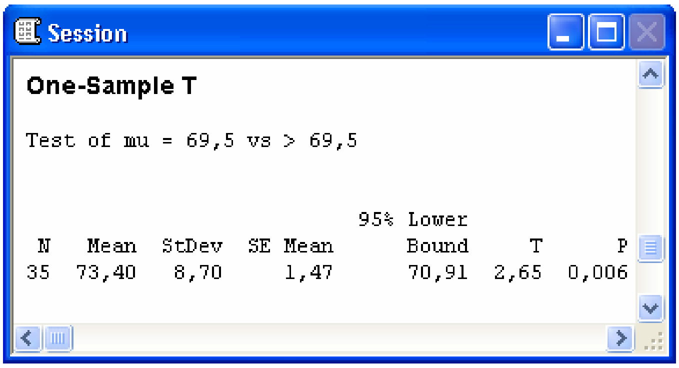
- Selecione ***Stat*** → ***Basic Statistics*** → ***1-Sample t***
- Janela *1-Sample t*: Selecione ***Summarized data*** e preencha os campos ***Sample size*** (tamanho da amostra), ***Mean*** (média amostral) e ***Standard deviation*** (desvio padrão). Marque a opção ***Perform hypothesis test*** (fazer teste de hipóteses) e informe o valor da média populacional μ em ***Hypothesized mean***. A seguir clique em ***Options***.



- Janela *1-Sample t - Options*: Introduza o nível de confiança em ***Confidence Level*** e para ***Alternative*** (hipótese alternativa), selecione ***greater than*** (teste unilateral: $\mu > \mu_0$). → ***OK***



- <u>Janela *1-Sample t*</u>: → ***OK***



<u>Resultados</u>:

- Como o valor de t_{obs} = 2,65 pertence à região crítica: {t > 1,691}, rejeitamos H_0 ao nível de 5% de significância.
- Como o valor-p = 0,006 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.
- Como o limite inferior do intervalo de 95% de confiança é maior que o valor de μ sob H_0 (μ = 69,5), rejeitamos H_0 ao nível de 5% de significância.

<u>Conclusão</u>: O tempo médio de vida de regentes homens é significativamente diferente de 69,5 anos.

5.3 – Teste de Hipóteses para Uma Proporção

- **Exemplo 5.3** – Um jornal alega que 25% dos seus leitores pertencem à classe A. Que regra de decisão adotaríamos para testar essa hipótese, contra a alternativa de que a porcentagem verdadeira não é 25% no nível de 5% de significância? Se em uma amostra de 740 leitores encontramos 156 da classe A, qual a sua decisão a respeito da alegação veiculada pelo jornal?

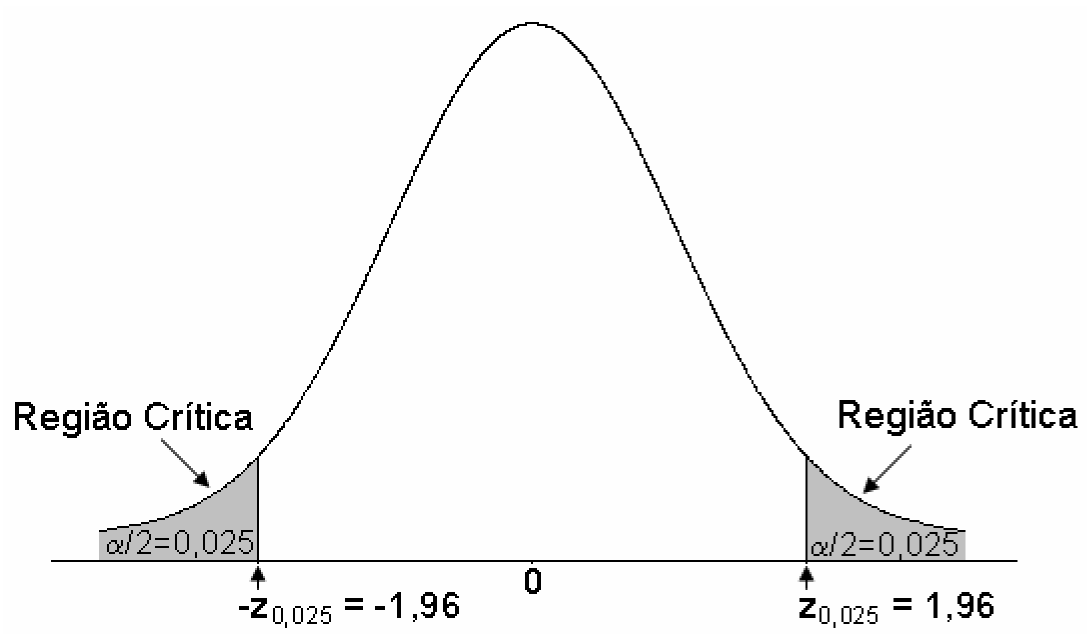
  Para este exemplo temos:

  Hipóteses: $\begin{cases} H_0: p = 0{,}25 \\ H_1: p \neq 0{,}25 \end{cases}$ Nível de Significância: $\alpha = 0{,}05$

  Estatística de Teste (Aproximando pela Normal):

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

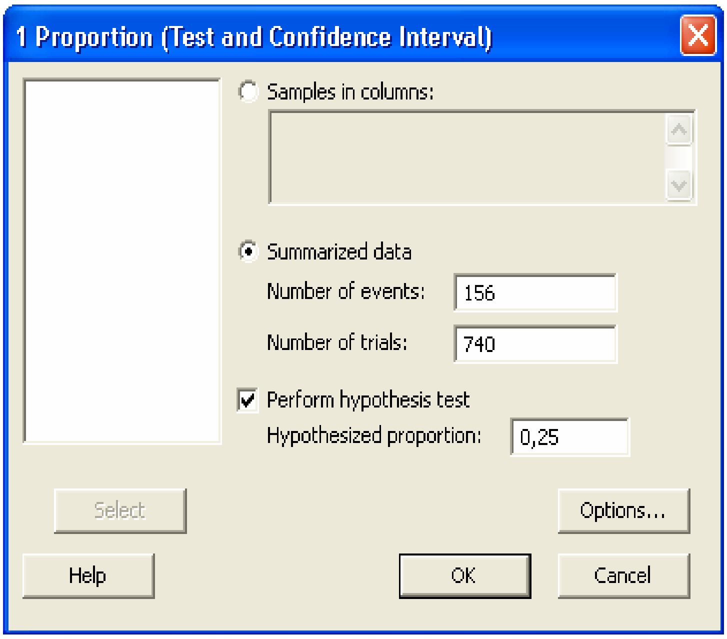
  Valores Críticos: $z_{0,025} = 1{,}96$ e $-z_{0,025} = -1{,}96$ (Valores Tabelados)



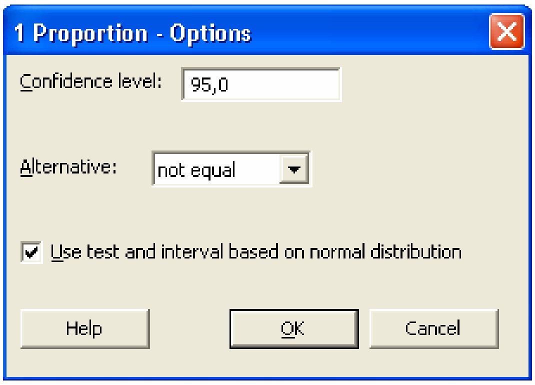
  Usando o Minitab:

- Selecione ***Stat*** **→** ***Basic Statistics*** **→** ***1 Proportion***

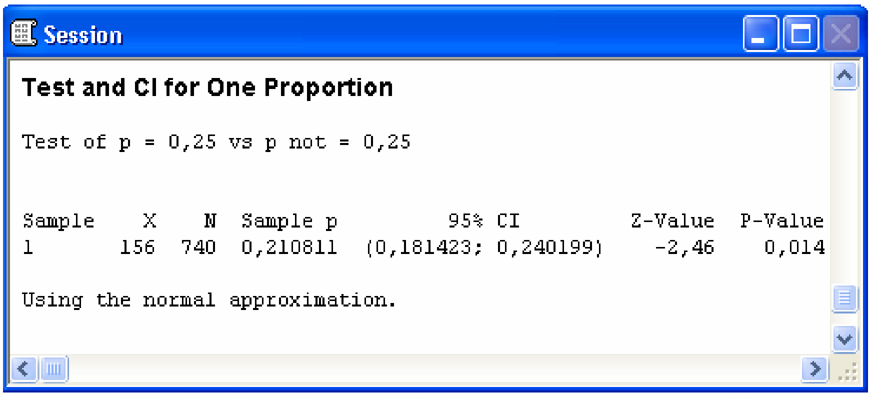
- <u>Janela *1 Proportion*</u>: Selecione ***Summarized data*** e preencha os campos ***Number of events*** (número de eventos) e ***Number of trials*** (número de ensaios). Marque a opção ***Perform hypothesis test*** (fazer teste de hipóteses) e informe o valor da proporção populacional *p* a ser testada em ***Hypothesized proportion***. A seguir clique em ***Options***.



- <u>Janela *1 Proportion - Options*</u>: Introduza o nível de confiança em ***Confidence Level*** e para ***Alternative*** (hipótese alternativa), selecione ***not equal*** (teste bilateral: $p \neq p_0$). Marque a opção ***Use test and interval based on normal distribution*** (o Minitab calcula a estatística de teste e o intervalo de confiança fazendo a aproximação da distribuição Binomial pela Normal). → ***OK***



- <u>Janela *1 Proportion*</u>: → ***OK***



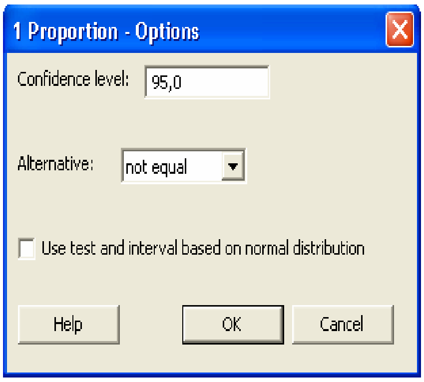
Resultados:

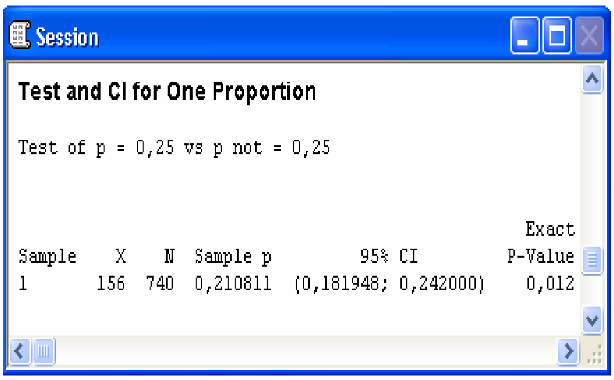
- Como o valor de $z_{obs} = -2{,}46$ pertence à região crítica: $\{z < -1{,}96$ ou $z > 1{,}96\}$, rejeitamos H_0 ao nível de 5% de significância.
- Como o valor-p = 0,014 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.
- Como o intervalo de 95% de confiança não contém o valor de *p* sob H_0 (*p* = 0,25), rejeitamos H_0 ao nível de 5% de significância.

Conclusão: A porcentagem de leitores pertencentes à classe A é significativamente diferente de 25%.

Nota 1: No teste de hipótese para proporções, a distribuição de probabilidade que se usa é uma aproximação da distribuição Binomial a Normal se, e somente se, $np > 5$ e $np(1-p) > 5$, onde *n* é o número de ensaios (ou eventos) e *p* é a proporção em teste. Neste exemplo a distribuição Normal pode ser usada satisfatoriamente.

Nota 2: Caso não tivéssemos marcado a opção ***Use test and interval based on normal distribution*** na janela ***1 Proportion - Options***, o Minitab iria usar o Teste da Razão de Máxima Verossimilhança para testar a proporção. A conclusão seria a mesma do processo anterior.





Resultados:

- Como o valor-p = 0,012 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.

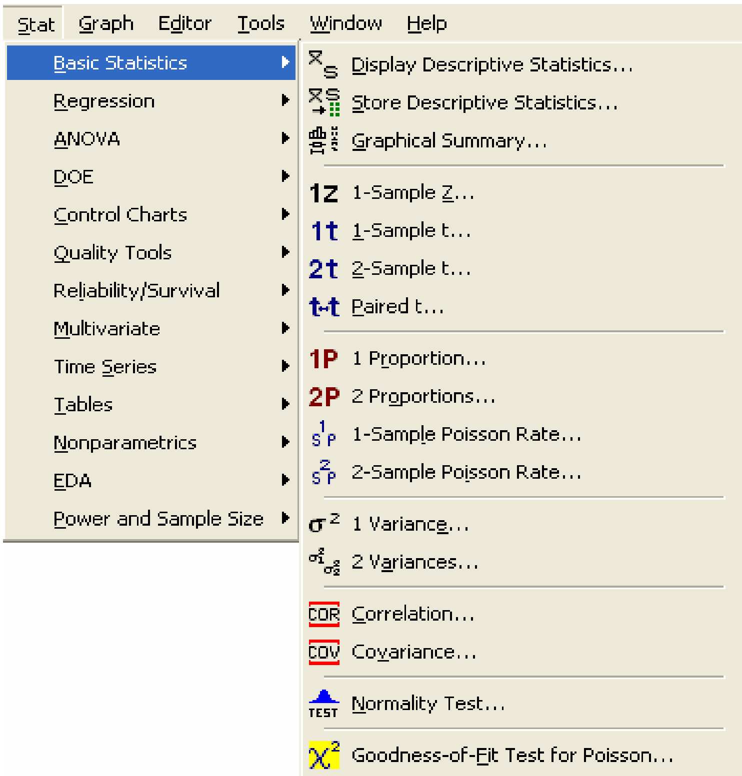
- Como o intervalo de 95% de confiança não contém o valor de p sob H_0 (p = 0,25), rejeitamos H_0 ao nível de 5% de significância.

<u>Conclusão</u>: A porcentagem de leitores pertencentes à classe A é significativamente diferente de 25%.

Capítulo 6 – Inferência a Partir de Duas Amostras

Neste capítulo apresentamos os procedimentos e recursos básicos do Minitab para fazer inferência a partir de duas amostras.

As opções de testes a serem utilizados encontram-se no menu ***Stat* → *Basic Statistics***.



Nota: Nos exemplos deste capítulo, as entradas dos testes são conjuntos de dados, mas também podem ser os dados sumarizados isto é, os valores dos parâmetros: média, desvio padrão e tamanho de amostra.

6.1 – Teste para Comparar Médias Populacionais: Duas Amostras Dependentes (Teste t - Pareado)

- **Exemplo 6.1** – *Captropil* é um medicamento desenvolvido para baixar a pressão sangüínea sistólica. Na tabela a seguir estão as leituras de pressão sangüínea sistólica (em mm de mercúrio) de pacientes antes e depois da ingestão do medicamento. Use um nível de significância de 0,05 para testar a afirmativa de que as medidas de pressão sangüínea sistólica são mais baixas depois do uso de *Captropil*.

Paciente	A	B	C	D	E	F	G	H	I	J	K	L
Antes	200	174	198	170	179	182	193	209	185	155	169	210
Depois	191	170	177	167	159	151	176	183	159	145	146	177

Para este exemplo temos:

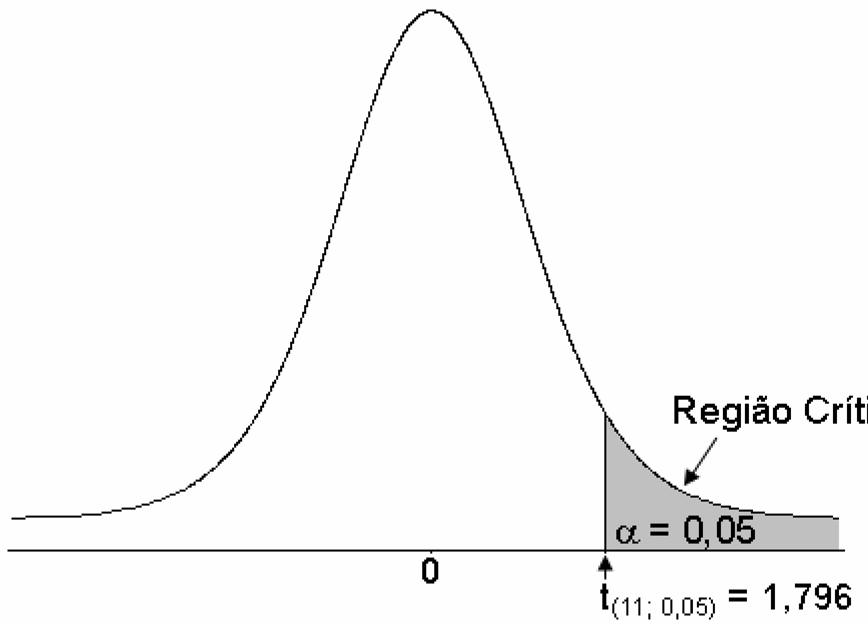
Hipóteses: $\begin{cases} H_0: \mu_d = 0 \\ H_1: \mu_d > 0 \end{cases}$ d = Antes - Depois

Nível de Significância: $\alpha = 0{,}05$

Estatística de Teste: t com 11 graus de liberdade

$$t = \frac{\overline{d} - \mu_d}{(s_d/\sqrt{n})}$$

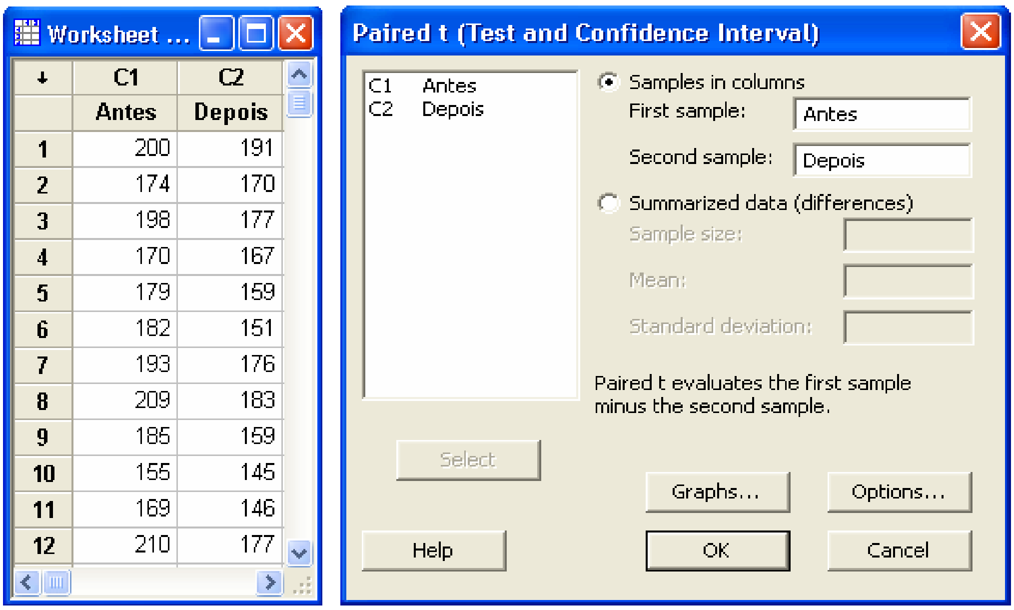
Valor Crítico: $t_{(11;\ 0,05)} = 1{,}796$ (Valor Tabelado)



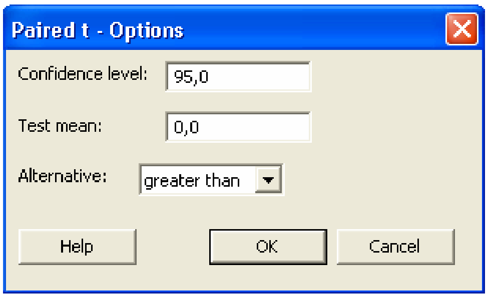
Usando o Minitab:

- Selecione ***Stat*** → ***Basic Statistics*** → ***Paired t***

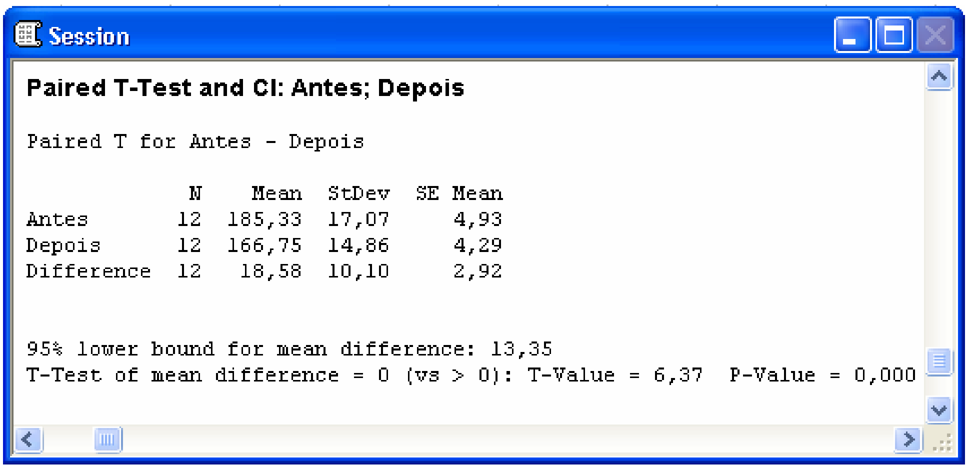
- <u>Janela *Paired t*</u>: Selecione ***Samples in Columns*** (amostras em colunas) e escolha as colunas para ***First sample*** (primeira amostra) e ***Second Sample*** (segunda amostra). A seguir clique em ***Options***.



- <u>Janela *Paired t - Options*</u>: Introduza o nível de confiança em ***Confidence Level*** e o valor da média μ_d a ser testada em ***Test mean***. Para ***Alternative*** (hipótese alternativa), selecione ***greater than*** (teste unilateral: $\mu_d > \mu_0$). → ***OK***



- Janela *Paired t*: → ***OK***



Resultados:

- Como o valor de t_{obs} = 6,37 pertence à região crítica: {t > 1,796}, rejeitamos H_0 ao nível de 5% de significância.
- Como o valor-p = 0,000 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.
- Como o limite inferior do intervalo de 95% de confiança é maior que o valor de μ sob H_0 (μ = 0), rejeitamos H_0 ao nível de 5% de significância.

Conclusão: A medida de pressão sanguínea sistólica é significativamente diferente de zero depois do uso de *Captropil*.

6.2 – Teste para Comparar Médias Populacionais: Duas Amostras Independentes com $\sigma_1 = \sigma_2$

- **Exemplo 6.2** – Dois grupos de ratos foram alimentados apresentando alto e baixo conteúdo de proteína. A tabela abaixo dá o peso, em gramas, para cada rato ao final do experimento. Apesar de não conhecidas, as variâncias populacionais para os dois grupos de ratos são consideradas iguais com base em estudos anteriores. Testar se as dietas são equivalentes ou não.

Alto	123	134	146	104	119	124	161	107	83	129	97	113
Baixo	70	118	101	85	107	132	94					

Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu_{Alto} = \mu_{Baixo} \\ H_1: \mu_{Alto} \neq \mu_{Baixo} \end{cases} \rightarrow \begin{cases} H_0: \mu_{Alto} - \mu_{Baixo} = 0 \\ H_1: \mu_{Alto} - \mu_{Baixo} \neq 0 \end{cases}$

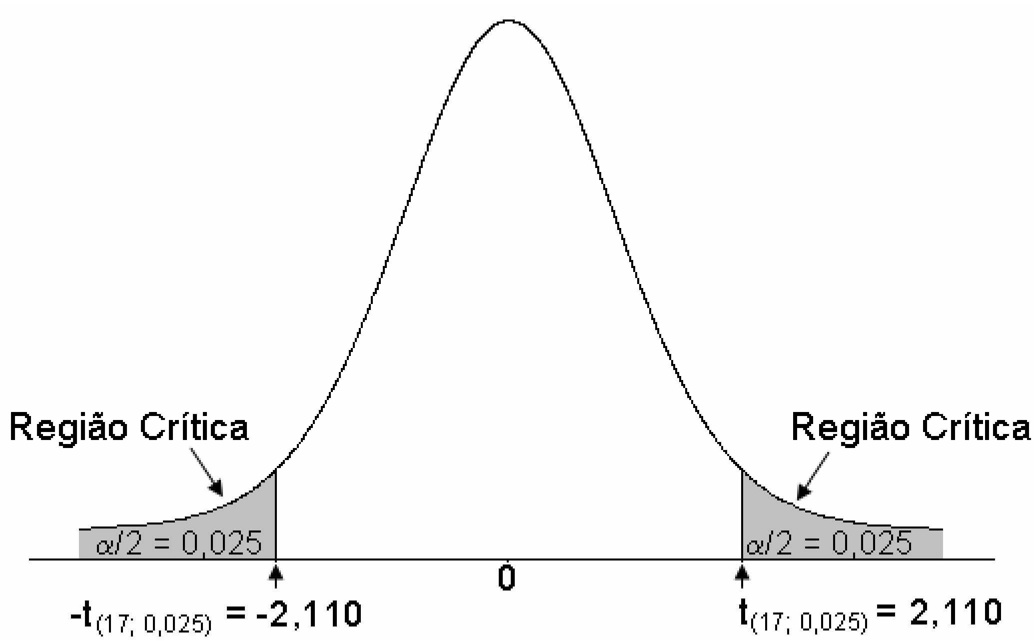
Nível de Significância: $\alpha = 0,05$

Estatística de Teste: t com 17 graus de liberdade

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

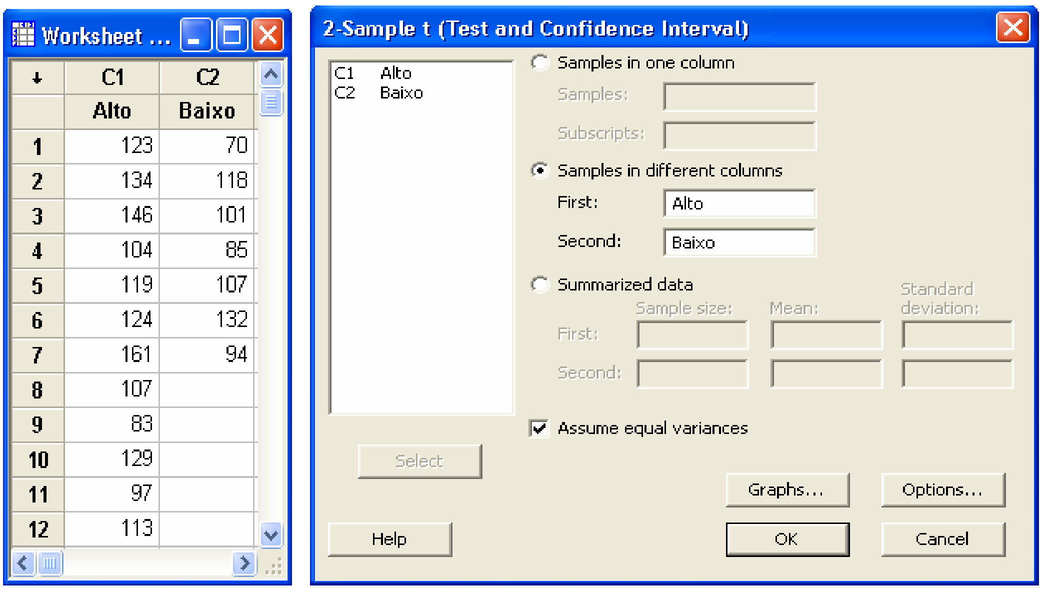
Valores Críticos: $t_{(17;\,0,025)} = 2,110$ e $-t_{(17;\,0,025)} = -2,110$ (Valores Tabelados)



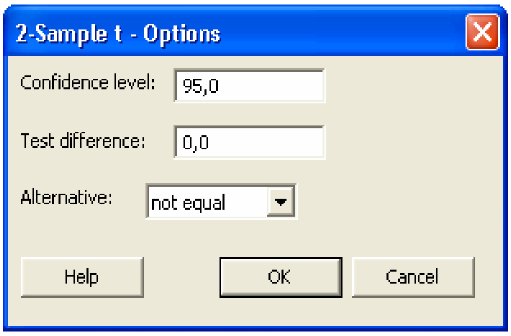
Usando o Minitab:

- Selecione ***Stat*** → ***Basic Statistics*** → ***2-Sample t***

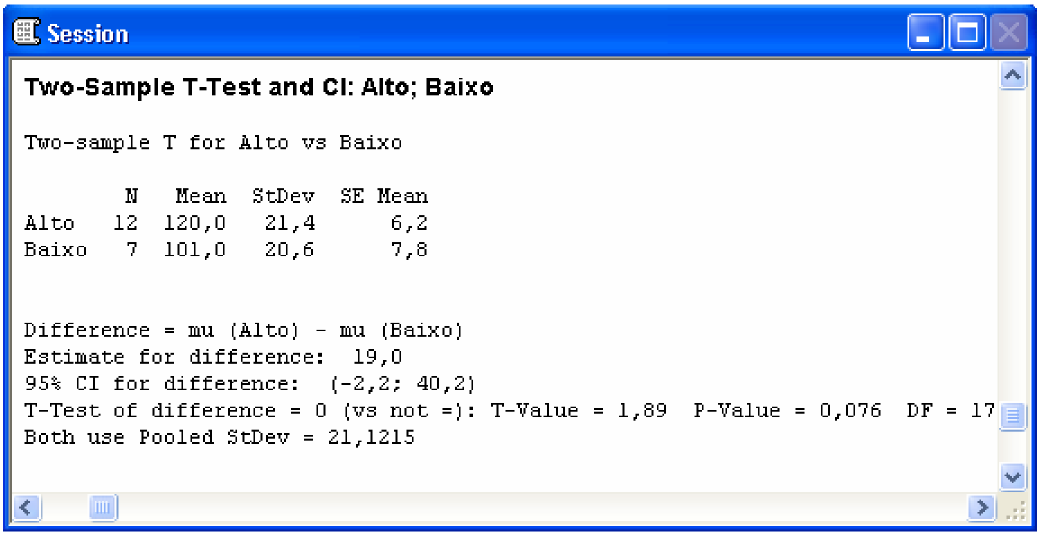
- Janela *2-Sample t*: Selecione ***Samples in different columns*** (amostras em colunas diferentes) e escolha as colunas para ***First*** (primeira amostra) e ***Second*** (segunda amostra). Marque a opção ***Assume equal variances*** (assumir variâncias iguais) e a seguir clique em ***Options***.



- Janela *2-Sample - Options*: Introduza o nível de confiança em ***Confidence Level*** e o valor da diferença entre as médias populacionais testadas: μ_1 - μ_2 em ***Test difference***. Para ***Alternative*** (hipótese alternativa), selecione ***not equal*** (teste bilateral: $\mu_1 - \mu_2 \neq 0$). → ***OK***



- Janela *2-Sample t*: → ***OK***



Resultados:

- Como o valor de t_{obs} = 1,89 não pertence à região crítica: {t < -2,110 ou t > 2,110}, deixamos de rejeitar H_0 ao nível de 5% de significância.
- Como o valor-p = 0,076 > α = 0,05, deixamos de rejeitar H_0 ao nível de 5% de significância.
- Como o intervalo de 95% de confiança contém o valor de $\mu_{Alto} - \mu_{Baixo}$ sob H_0 ($\mu_{Alto} - \mu_{Baixo} = 0$), deixamos de rejeitar H_0 ao nível de 5% de significância.

Conclusão: Não há evidência estatística suficiente para rejeitar a afirmativa de que as dietas com alto e baixo conteúdo de proteínas sejam equivalentes.

6.3 – Teste para Comparar Médias Populacionais: Duas Amostras Independentes com $\sigma_1 \neq \sigma_2$

- **Exemplo 6.3** – Consulte os dados amostrais listados abaixo e use um nível de significância de 0,05 para testar a afirmativa de que a quantidade média de alcatrão em cigarros tamanho *king-size* com filtro é menor do que a quantidade média de alcatrão em cigarros tamanho *king-size* sem filtro. Sabe-se que as variâncias populacionais são diferentes, sendo seus valores desconhecidos. Todas as medidas estão em miligramas e os dados são da Comissão Federal de Comércio dos Estados Unidos.

Com Filtro	16	15	16	14	16	1	16	18	10	14	12
	11	14	13	13	13	16	16	8	16	11	
Sem Filtro	23	23	24	26	25	26	21	24			

Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \mu_{Com\ Filtro} = \mu_{Sem\ Filtro} \\ H_1: \mu_{Com\ Filtro} < \mu_{Sem\ Filtro} \end{cases} \rightarrow \begin{cases} H_0: \mu_{Com\ Filtro} - \mu_{Sem\ Filtro} = 0 \\ H_1: \mu_{Com\ Filtro} - \mu_{Sem\ Filtro} < 0 \end{cases}$

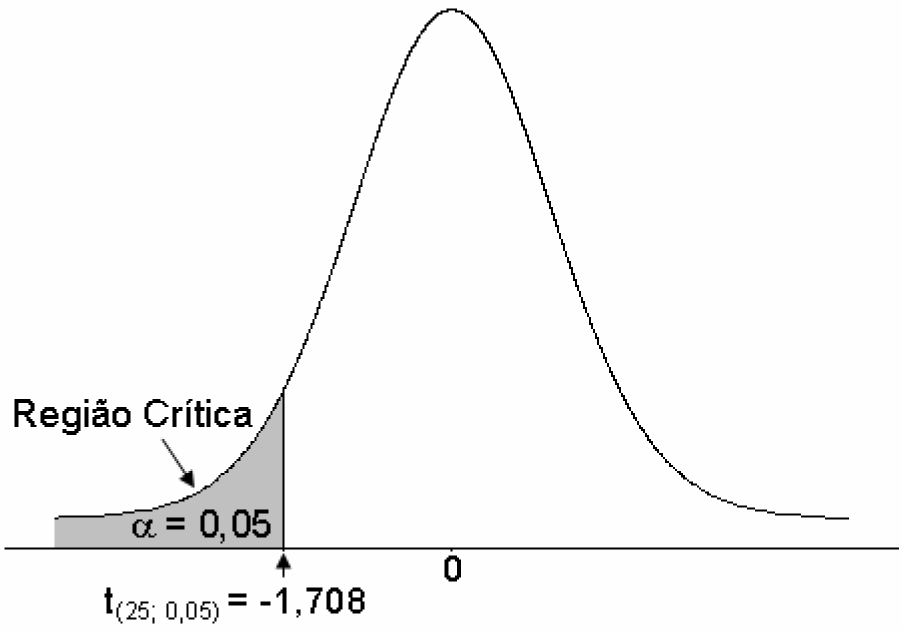
Nível de Significância: $\alpha = 0,05$

Estatística de Teste: t com v graus de liberdade

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$v = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

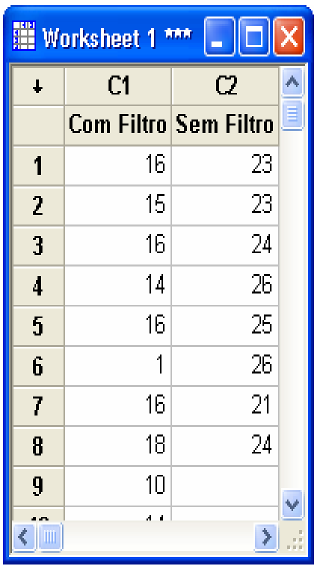
Valor Crítico: $t_{(25;\ 0,05)} = -1,708$ (Valor Tabelado)

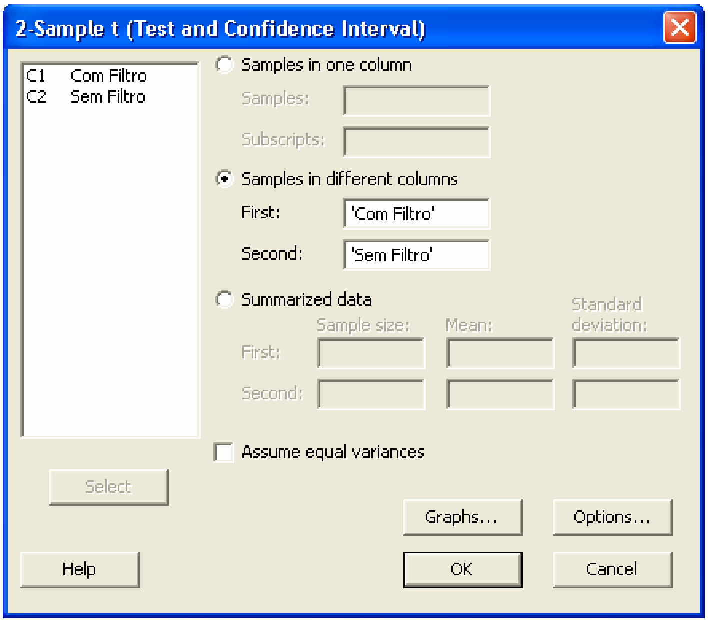


Usando o Minitab:

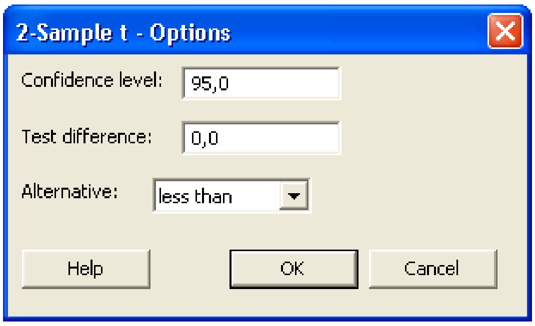
- Selecione ***Stat*** → ***Basic Statistics*** → ***2-Sample t***

- Janela *2-Sample t*: Selecione ***Samples in different columns*** (amostras em colunas diferentes) e escolha as colunas para ***First*** (primeira amostra) e ***Second*** (segunda amostra). Como foi mencionado que as variâncias populacionais são desconhecidas e diferentes, deixamos de marcar a opção ***Assume equal variances*** (assumir variâncias iguais). A seguir clique em ***Options***.

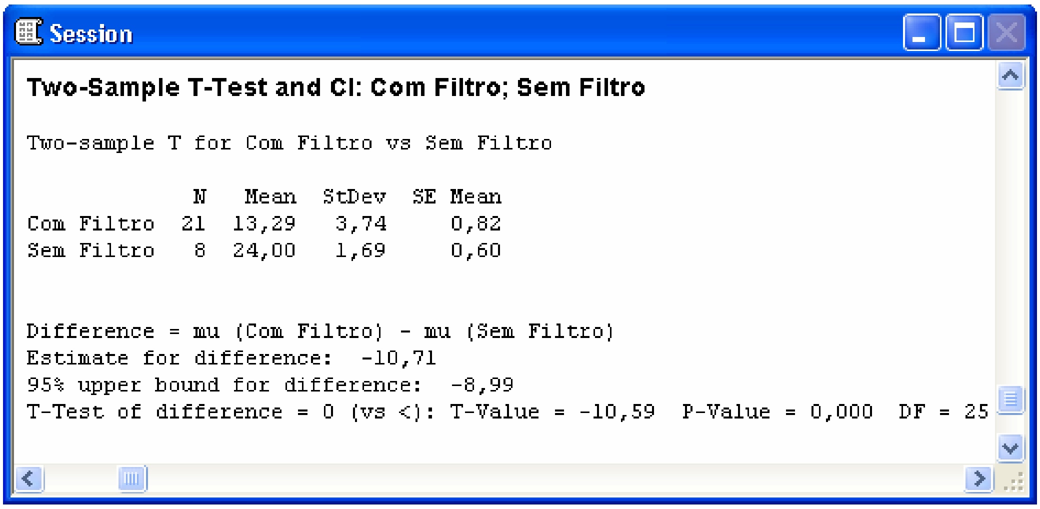




- Janela *2-Sample t - Options*: Introduza o nível de confiança em ***Confidence Level*** e o valor da diferença entre as médias populacionais testadas: μ_1 - μ_2 em ***Test difference***. Para ***Alternative*** (hipótese alternativa), selecione ***Less than*** (teste unilateral: $\mu_1 - \mu_2 < 0$). → ***OK***



- <u>Janela *2-Sample t*</u>: → ***OK***



<u>Resultados</u>:

- Como o valor de $t_{obs} = -10,59$ pertence à região crítica: $\{t < -1,708\}$, rejeitamos H_0 ao nível de 5% de significância.
- Como o valor-p = 0,000 < α = 0,05, rejeitamos H_0 ao nível de 5% de significância.
- Como o limite superior do intervalo de 95% de confiança é menor que o valor de $\mu_{Com\ Filtro} - \mu_{Sem\ Filtro}$ sob H_0 ($\mu_{Com\ Filtro} - \mu_{Sem\ Filtro} = 0$), rejeitamos H_0 ao nível de 5% de significância.

<u>Conclusão</u>: A quantidade média de alcatrão em cigarros tamanho *king-size* com filtro é significativamente diferente da quantidade média de alcatrão em cigarros tamanho *king-size* sem filtro.

6.4 – Teste para Comparar Variâncias Populacionais

É importante ressaltar que o teste para comparação de variâncias é também útil como um procedimento preliminar em testes de comparação de médias, auxiliando a escolha das técnicas adequadas.

- **Exemplo 6.4** – Coletaram-se dados amostrais em um estudo de suplementos de cálcio e seus efeitos sobre a pressão sangüínea. Os valores da amostra estão listados a seguir. No nível de significância de 0,10, teste a afirmativa de que os dois grupos amostrais originam-se de populações com a mesma variância.

Placebo	124,6	104,8	96,5	116,3	106,1	128,8	107,2	123,1
	118,1	108,5	120,4	122,5	113,6			
Cálcio	129,1	123,4	102,7	118,1	114,7	120,9	104,4	116,3
	109,6	127,7	108,0	124,3	106,6	121,4	113,2	

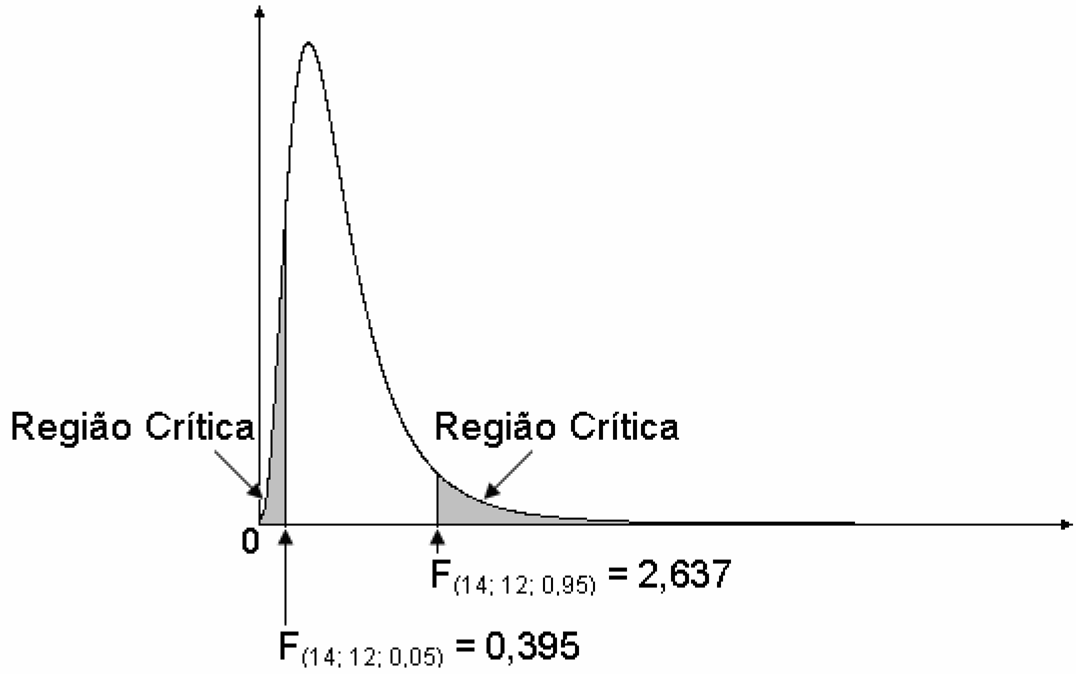
Para este exemplo temos:

Hipóteses: $\begin{cases} H_0: \sigma^2_{Placebo} = \sigma^2_{Cálcio} \\ H_1: \sigma^2_{Placebo} \neq \sigma^2_{Cálcio} \end{cases}$ Nível de Significância: $\alpha = 0{,}10$

Estatística de Teste: F com 14 e 12 graus de liberdade.
(S_1^2 representa o maior valor das duas variâncias amostrais)

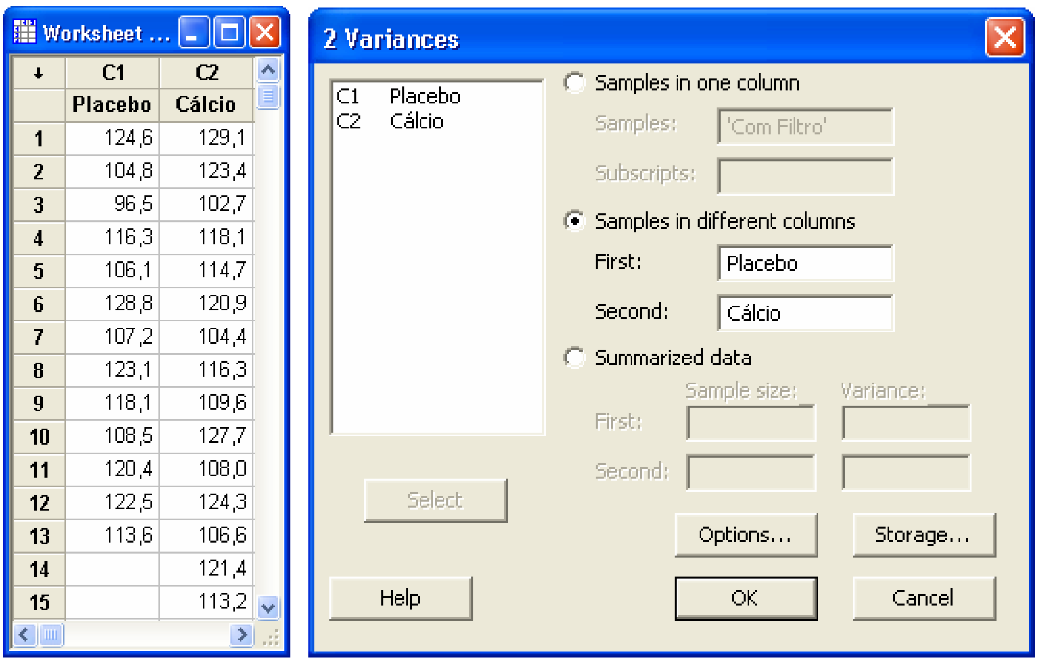
$$F = \frac{S_1^2}{S_2^2}$$

Valores Críticos: $F_{(14;\ 12;\ 0,95)} = 0{,}395$ e $F_{(14;\ 12;\ 0,05)} = 2{,}637$ (Valores Tabelados)

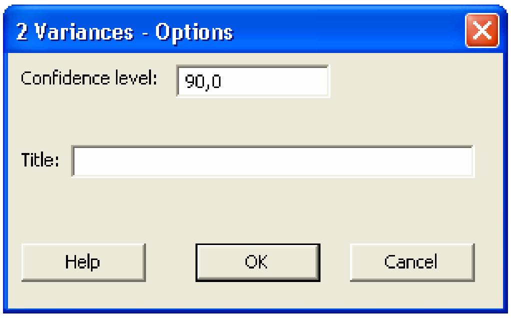


Usando o Minitab:

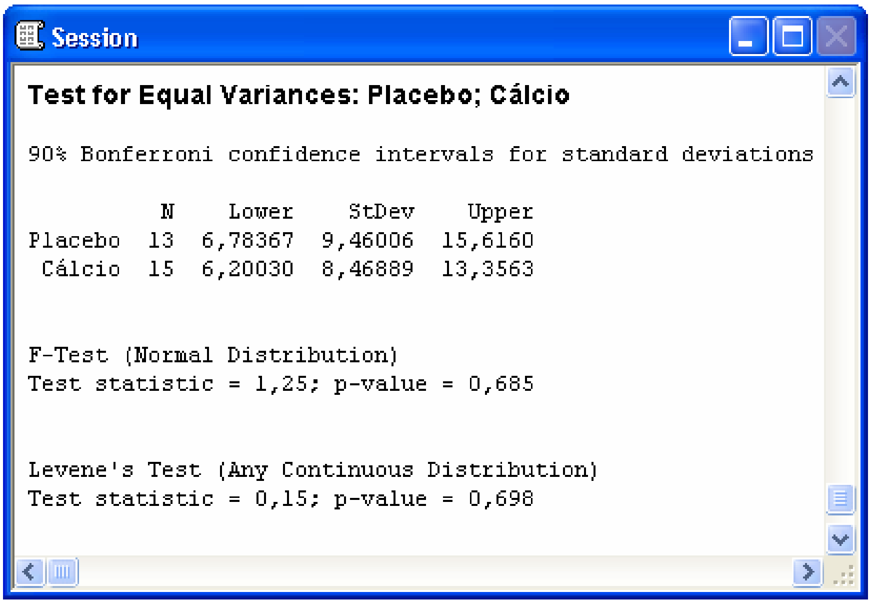
- Selecione ***Stat*** → ***Basic Statistics***→ ***2 Variances***
- <u>Janela *2 Variances*</u>: Selecione ***Samples in different columns*** (amostras em colunas diferentes) e escolha as colunas para ***First*** (primeira amostra) e ***Second*** (segunda amostra). A seguir clique em ***Options***.

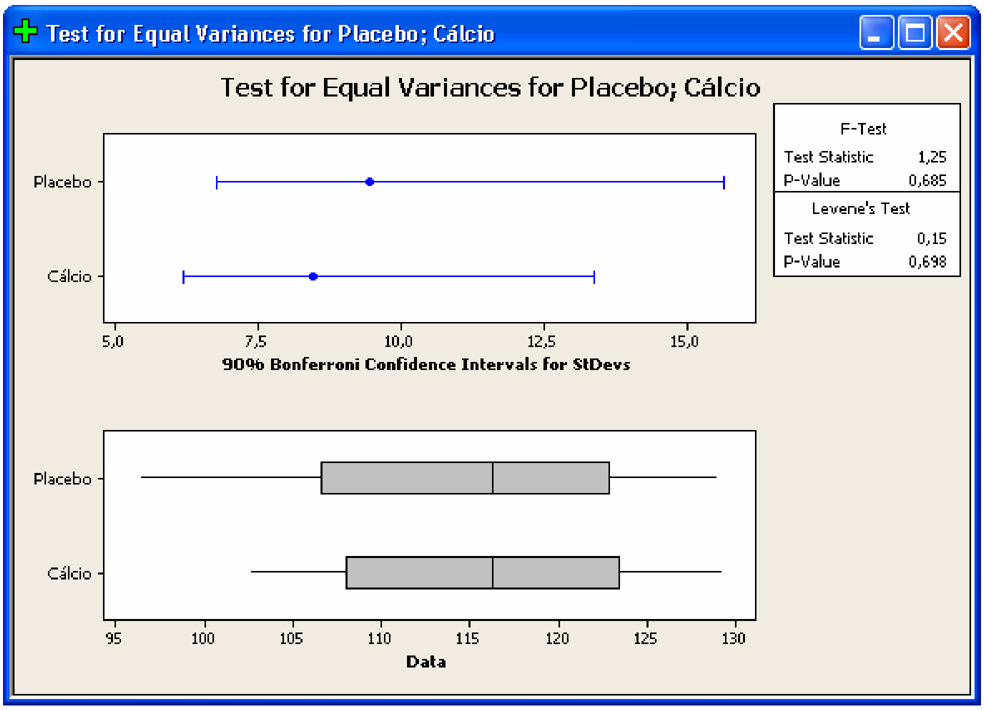


- <u>Janela *2 Variances - Options*</u>: Preencha os campos ***Confidence Level*** (nível de confiança) e ***Title*** (caso queira atribuir um título ao gráfico do teste de variâncias). → ***OK***



- <u>Janela *2 Variances*</u>: → ***OK***





Resultados:

- Como o valor de $F_{obs} = 1,25$ não pertence à região crítica: $\{F < 0,395$ ou $F > 2,637\}$, deixamos de rejeitar H_0 ao nível de 10% de significância.
- Como o valor-p = 0,685 > α = 0,10, deixamos de rejeitar H_0 ao nível de 10% de significância.

Conclusão: Há evidência estatística suficiente para afirmar que os dois grupos provenham de populações com variâncias iguais.

Nota 1: Da perspectiva de estatísticas de teste é importante esclarecer que escolhendo as colunas para ***First*** (primeira amostra) e ***Second*** (segunda amostra) estamos definindo, também, que elas correspondem às amostras do numerador e denominador, respectivamente, da estatística de teste F. Essa informação é necessária para determinar os valores críticos e, consequentemente, a região crítica.

Nota 2: Usualmente encontramos tabelas da distribuição F correspondentes à cauda direita. O valor crítico da cauda esquerda pode ser encontrado tomando o recíproco do valor crítico da cauda direita com os graus de liberdade invertidos. No exemplo temos:

$$F_{(14;\ 12;\ 0,95)} = 1 / F_{(12;\ 14;\ 0,05)} = 1 / 2,534 = 0,395$$

Referências Bibliográficas

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 5. ed. São Paulo: Saraiva, 2002.

FARIAS, A. A.; CÉSAR, C. C.; SOARES, J. F. *Introdução à Estatística*. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 2003.

MAGALHÃES, M. N.; Pedroso de Lima, A. C. *Noções de Probabilidade e Estatística*. 6. ed. São Paulo: Edusp, 2008.

MEYER, P. L. *Probabilidade: Aplicações à Estatística*. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1995.

ROSS, S. M. *A First Course in Probability*. 7. ed. Upper Saddle River: Prentice Hall, 2005.

TRIOLA, M. F. *Introdução à Estatística*. 9. ed. Rio de Janeiro: Livros Técnicos e Científicos, 2005.