

Rogério L. Craveiro
Edcarlos M. de Souza





ANÁLISE EXPLORATÓRIA DE DADOS BIOLÓGICOS

Com o Software Action

“É fácil mentir usando a estatística, difícil é dizer a verdade sem ela”.

Andrejs Dunkels (Maindonald, 1992)



SUMÁRIO

1. INTRODUÇÃO	5
2. VISÃO GERAL DA ANÁLISE ESTATÍSTICA	6
2.1 Estatística Descritiva	6
2.2 Estatísticas Inferenciais	6
3. VOCABULÁRIO BÁSICO ESTATÍSTICO	7
3.1 População	7
3.2 Amostra	8
3.3 Variável	10
3.4 Tipos de Variáveis	10
Exercício Proposto.....	11
4. PLANEJAMENTO DA COLETA DE DADOS.....	12
4.1 Técnicas de Amostragem.....	12
4.2 Obtenção de Amostras.....	12
4.2.1 Amostra Aleatória ou Probabilística.....	13
4.2.2 Amostra Aleatória Simples.....	13
4.2.3 Amostra Aleatória Estratificada	15
5. ORGANIZAÇÃO DOS DADOS.....	17
5.1 Dados Brutos	17
5.2 Rol	17
6. TABELAS DE DISTRIBUIÇÃO DE FREQUÊNCIA.....	20
6.1 Tabela de Distribuição de Frequência simples	20
Exercício proposto.....	25
6.2 Interpretação da tabela de distribuição de frequência gerada pelo <i>Action</i>	25
6.3 Tabelas de Distribuição de Frequência com Intervalos de Classe.....	27
6.4 Interpretação da tabela de distribuição de frequência com intervalos de classe gerada pelo <i>Action</i>	31
7. GRÁFICOS	34
7.1 Gráficos para variáveis qualitativas.....	34
7.1.1 Gráfico em setores (Pizza).....	34
Exercício Proposto.....	38
7.2 Gráficos para variáveis Quantitativas.....	38
7.2.1 Gráfico em Barras ou Colunas	38



Exercício Proposto.....	41
7.2.2 Histograma	41
Exercício Proposto.....	45
8. MEDIDAS ESTATÍSTICAS	46
8.1 Medidas de Tendência Central	46
8.1.1 Média Aritmética.....	46
8.1.2 Mediana	48
8.1.3 Moda.....	48
8.2 Medidas de Dispersão.....	49
8.2.1 Amplitude	49
8.2.2 Variância.....	50
8.2.3 Desvio Padrão.....	51
8.2.4 Coeficiente de Variação.....	52
8.3 Cálculo de medidas estatísticas utilizando o <i>software Action</i>	52
8.3.1 Análise sem fator	52
8.3.2 Análise com um único fator e diversos níveis.....	55
Exercício Proposto.....	59
8.4 Quartil.....	60
Exercício Proposto.....	61
8.5 <i>Box plot</i> (Gráfico de Caixa).....	61
8.5.1 Gerando <i>Box plot</i> com o <i>software Action</i>	63
8.5.2 Gerando <i>Box plot</i> para a comparação entre dois ou mais grupos.....	66
Exercício proposto.....	69
9. NOÇÕES DE CORRELAÇÃO	70
9.1 Diagrama de dispersão.....	70
9.2 Coeficiente de Correlação	72
9.3 Cálculo do coeficiente de correlação com o <i>Software Action</i>	75
Exercício Proposto.....	77
Referências	84



1. INTRODUÇÃO

A análise exploratória de dados também é conhecida como estatística descritiva. A palavra “Estatística” vem de *status*, que significa em latim Estado. Com essa palavra faziam-se as descrições e dados relacionados aos Estados, tornando a Estatística um meio de administração para os governantes.

Atualmente as inúmeras aplicações de metodologias de análise nos estudos dos fenômenos biológicos e a integração da estatística como ciência na essência do método científico são algumas das razões que justificam a importância do conhecimento das técnicas estatísticas na investigação científica.

A estatística reúne técnicas de coleta de dados que visam de informações e interpretação de resultados. Neste sentido, existem diversos aspectos estudados com suas peculiaridades, sendo que diversos ramos da análise estatística são hoje objeto de especialidades, dentre estes podemos destacar: Análise de Séries Temporais, Geoestatística, Análise Multivariada, Análise de Regressão, Análise de Sobrevivência, Estatística Espacial e outros. Existem também situações em que se pretende realizar estudos estatísticos com especificidade em áreas, como é o caso da Estatística Matemática, Econometria, Bioestatística e outros. O nosso objeto de estudo neste material terá como foco a Análise Exploratória de Dados relacionados a Bioestatística.

A Bioestatística é a Estatística aplicada às áreas Biológicas. O investigador na área biológica muitas vezes se depara com a necessidade de coletar dados, esses dados tem uma característica comum que é a variabilidade, ou seja, os dados coletados não são iguais, eles variam. Essa característica é a principal razão para a existência da estatística. A escolha pelo tema surge da experiência prática dos autores ao trabalharem com disciplinas ligadas a esta área na Universidade Federal do Acre.

Esse material pretende contribuir no apoio a iniciação de investigadores na área biológica no campo das técnicas e métodos estatísticos. Para realização do mesmo, optamos por realizar os procedimentos com o uso de softwares estatísticos específicos, no caso em questão com o auxílio do *Software Action*. Por tanto, as demonstrações matemáticas mais complexas foram ocultadas para dar lugar as análises e interpretações dos resultados. No entanto, buscamos dar uma visão básica sobre os cálculos realizados pelo software *Action*. O motivo pelo qual escolhemos o *Action* como ferramenta de apoio nas análises estatísticas foi pela sua possibilidade de utilização com o *Microsoft Office Excel*, que é um *software* bastante conhecido, mesmo que a maioria de seus usuários não explorem recursos avançados de cálculo. A coleta, organização, análise e apresentações dos dados, provenientes das investigações científicas, serão estudadas neste material, que apontará algumas técnicas para que esses dados sejam validados.

Não esperamos e nem temos a intenção de esgotar o tema com este material, nossa pequena contribuição está relacionada a aplicação dos conceitos estatísticos usuais na análise exploratória de dados e a utilização do *software Action* como facilitador da aprendizagem, como também do tratamento da informação. Buscamos utilizar, sempre



que possível exemplo de conjunto de dados reais, utilizados em pesquisas realizadas, em grande maioria por colegas de profissão ligados as áreas biológicas. Desde já, agradecemos os mesmos pela valiosa contribuição.

2. VISÃO GERAL DA ANÁLISE ESTATÍSTICA

Quando realizamos uma análise estatística, podemos se deparar em geral com dois tipos de situação: A primeira é aquela que realiza apenas uma discussão descritiva sobre os dados, a segunda aquela que faz inferências sobre informações da população a partir da amostra. Estas recebem o nome de Estatística Descritiva e Estatística Inferencial respectivamente.

2.1 Estatística Descritiva

A estatística descritiva interessa-se pela organização e apresentação dos dados. Para isso, utiliza-se de ferramentas como os gráficos, tabelas e medidas descritivas. É utilizada, principalmente, na etapa inicial da pesquisa, buscando modelos que poderão ser utilizados numa etapa final, o que chamamos de inferência estatística.

2.2 Estatísticas Inferenciais

A estatística inferencial busca reunir um conjunto de técnicas que possibilitem utilizar dados de amostras para se chegar a conclusões sobre populações. Estas inferências são possíveis pela utilização de teoria da probabilidade. Em outras palavras, o pesquisador tem a necessidade de realizar afirmações sobre determinada característica do objeto de pesquisa. Para isso, ele busca um conjunto menor de observações, realiza alguns procedimentos matemáticos e probabilísticos e passa a concluir ou afirmar algo sobre o fenômeno estudado. A estatística inferencial abrange uma vasta teoria e técnicas sofisticadas para chegar a tais conclusões. Podemos dizer, que em grande parte dos trabalhos científicos esta é a situação mais utilizada. Entretanto, muitas vezes, para preceder a um bom procedimento inferencial, necessita-se observar os dados com mais cautela, procurando algum padrão ou informação importante dentro do conjunto de dados observável.

De certo então, que a análise exploratória dos dados, está mais ligada ou até mesmo pode ser considerada a própria análise descritiva ou Estatística Descritiva. Neste sentido, a Estatística descritiva pode ser por si só uma análise que já justifica toda uma pesquisa, com é o caso em estudos de casos, ou pode ser um método que antecede uma boa análise inferencial.

3. VOCABULÁRIO BÁSICO ESTATÍSTICO

3.1 População

Uma população consiste em todos os itens ou indivíduos ao qual o pesquisador deseja tirar conclusões.

ATENÇÃO

O conceito de População em Estatística é diferente do conceito de população em Biologia.

Segundo Odum (2004), **População** é o conjunto de organismos da mesma espécie (ou outros grupos no seio dos quais os indivíduos podem trocar informações genéticas) ocupando um dado espaço. Ou seja, população em biologia refere-se exclusivamente a serem vivos.

Em Estatística uma população pode ser constituída de animais ou pessoas, mas também pode ser constituída por certidões de óbito, certidões de nascimento, lote de produção de um remédio, entre outros.

A seguir alguns exemplos de populações em Estatística.

O sistema vascular do corpo humano pode ser considerado uma população em Estatística.

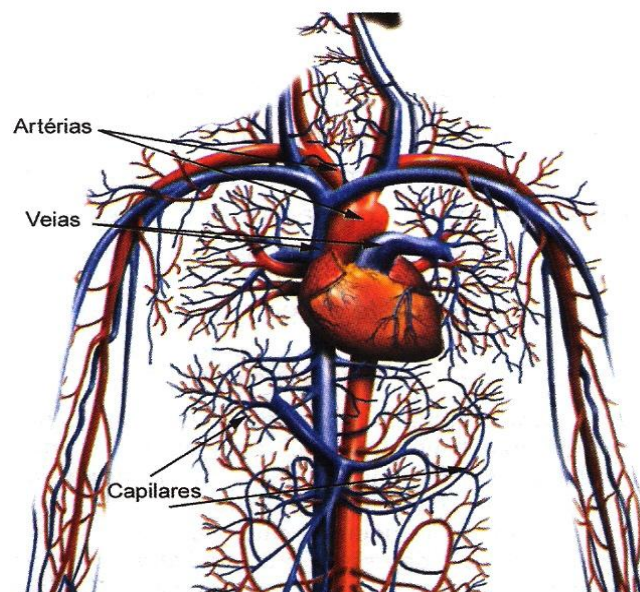


Figura-1 - Vasos sanguíneos do corpo humano.
Forte: www.colegiocardeal.com.br

As hemácias e leucócitos sanguíneos são exemplos de população em Estatística.

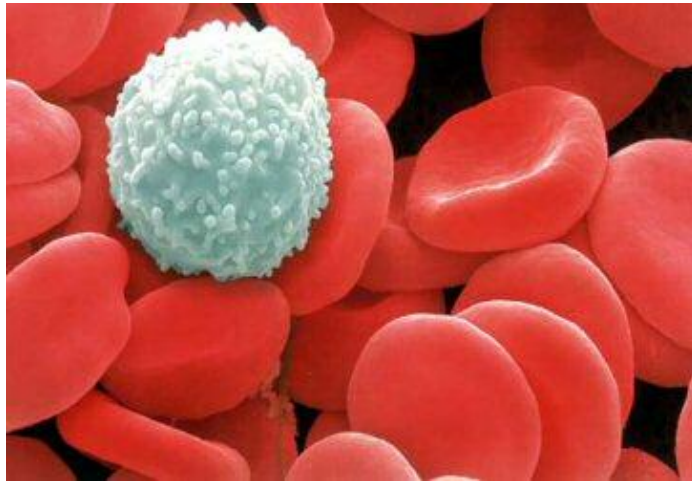


Figura-2 - Microscopia eletrônica mostrando as hemácias (em vermelho) e um glóbulo branco (em branco). Fonte: www.sobiologia.com.br

População de peixes, algas, plantas e seres microscópicos do mar.



Figura-3 - Peixes, algas, plantas e seres microscópicos do mar. Fonte: www.sobiologia.com.br

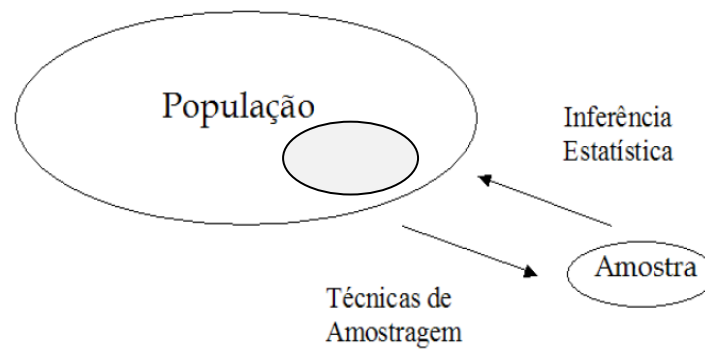
3.2 Amostra

Uma amostra corresponde a parcela da população selecionada para análise.

A escolha de uma amostra requer cuidados específicos para que os resultados não sejam distorcidos.



DIAGRAMA DA ESTATÍSTICA



Exemplos de amostras em biologia

Amostra de sangue para exames laboratoriais.



Figura -4 Tubo de ensaio com amostra sanguínea
Fonte: <http://www.mundodastribos.com>

Amostra de cobras



Figura -5 Biólogo coletando espécies de cobras para análise.
Fonte: <http://gambarini.com.br/blog/2009/01/29/herpetologia-sapos-cobras-pererecas-lagartos-144705/>

Amostra de nadadeira de peixes



Figura-6 Coleta de amostra de nadadeira de peixes
Fonte: www.embrapa.br

3.3 Variável

Uma variável corresponde a uma característica de um item ou de um indivíduo.

Ex: Peso, altura, gênero, espécie de um animal, quantidade de proteína total no sangue, temperatura, cor, tipo de vocalização, etc.

3.4 Tipos de Variáveis

3.4.1 Qualitativas

São aquelas variáveis que indicam qualidades, atributos, características não numéricas de forma geral.

Ex: Gênero, espécie de um animal, cor, tipo de vocalização, etc.

(a) Qualitativas nominais:

São aquelas que não permitem uma ordenação natural.

Ex: Cor de uma flor, raça de um animal ou o sexo de um paciente, etc.

(b) Qualitativas ordinais:

Por sua vez, são aquelas que admitem uma ordenação natural.

Ex: Estágio de uma doença, ciclo de vida de uma espécie, mês de observação de um fenômeno, etc.

3.4.2 Quantitativas

Resumem-se a medidas, pesagens ou contagens.

(a) Quantitativas discretas:

São representadas pelas contagens.

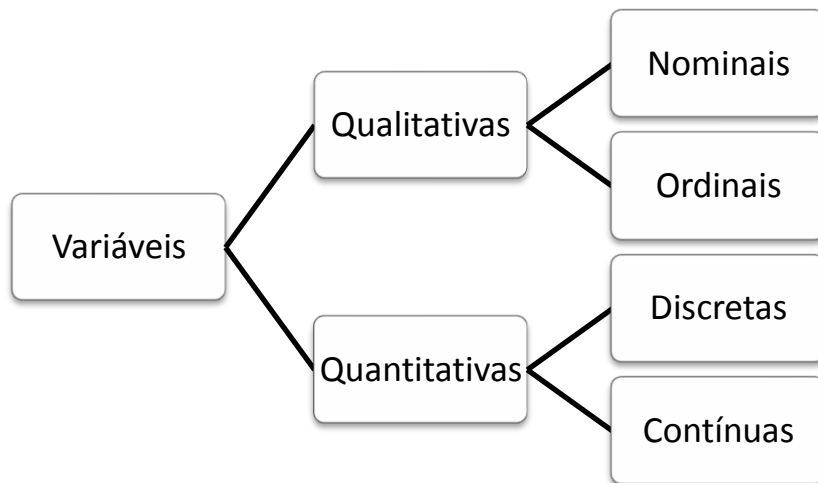
Ex: Número de filhos nascidos vivos, número de células aneuplóides por antera, etc.

(b) Quantitativas contínuas:

São representadas pelas medições ou pesagens

Ex: Peso de um animal, altura de um indivíduo, etc.

A seguir, visualizaremos um diagrama com o resumo dos tipos de variáveis estatísticas, para facilitar nossa familiarização com os termos utilizados nesta área. Pois o conhecimento correto desses termos é fundamental para o entendimento da disciplina.



Exercício Proposto

1. Denomine o tipo de variável que está sendo coletada nos casos abaixo:

- a) Espécies de abelhas
- b) Unidade relativa do ar (%)
- c) Temperatura (°C)
- d) Precipitação (mm)
- e) Número de óbitos por acidente ofídico
- f) Tipo de serpente
- g) Abundância de lagartos
- h) Ciclo de vida de uma espécie

2. Cite três exemplos além dos que foram citados acima das seguintes variáveis:

- a) Variável qualitativa ordinal;
- b) Variável quantitativa discreta.

4. PLANEJAMENTO DA COLETA DE DADOS

Nos trabalhos de investigação científica, existe uma sequência a ser seguida. Primeiro temos que identificar o assunto que será estudado, em seguida devemos realizar o planejamento da pesquisa das variáveis relevantes a serem incluídas, questionando-se, por exemplo, quais são as hipóteses a comprovar e qual o procedimento utilizado para refutar ou não estas hipóteses. Em muitos casos, não necessariamente a pesquisa envolve uma hipótese inicial, pois o meio do qual ela se desenvolve poderá propiciar os questionamentos durante o percurso, mesmo assim, um planejamento das ações deve ser realizado com cautela. Logo após, passamos a identificar qual a forma que os dados serão coletados, nesse momento devemos descrever o método ou procedimento utilizado na obtenção dos mesmos. A partir daí, organizar o cronograma das atividades, os custos que serão envolvidos, o exame das informações disponíveis sobre o tema escolhido, que pode ser feito através de uma pesquisa bibliográfica do que já se sabe sobre o assunto. O próximo passo é a coleta dos dados, que consiste na obtenção dos dados das variáveis que compõem o fenômeno a ser estudado.

4.1 Técnicas de Amostragem

Após a definição da população a ser estudada, é preciso estabelecer a técnica de amostragem, ou seja, o procedimento que será utilizado para escolher os elementos que irão compor a amostra. Dependendo da técnica utilizada, tem-se um tipo de amostra.

Muitas são as razões pelas quais os pesquisadores utilizam as amostras em vez da população. Uma delas é a questão do custo e da demora de se obter dados de toda uma população. Outra razão é a existência de populações tão grandes que estudá-las por completo seria praticamente impossível. Outras vezes é impossível estudar toda a população porque o estudo destrói as unidades. Por exemplo, se quisermos saber a tipagem sanguínea de um indivíduo não podemos retirar-lhe todo o sangue, por isso, coletamos uma amostra sanguínea.

4.2 Obtenção de Amostras

Para obtermos uma amostra, primeiro é preciso definir os critérios que serão usados para a seleção das unidades que irão compô-la. De acordo com a técnica usada, tem-se um tipo de amostra. Estudaremos neste material apenas a amostra aleatória ou



probabilística por ser a mais simples e uma das mais utilizada no meio biológico, no entanto, isso não impede que o leitor busque conhecer as outras técnicas de amostragem, na verdade recomendamos o estudo das mesmas por serem de extrema importância para as pesquisas posteriores.

4.2.1 Amostra Aleatória ou Probabilística

Para obter uma amostra aleatória, o pesquisador retira ao acaso n unidades da população. A escolha dessas n unidades pode ser obtida através de sorteio. Dessa forma, toda unidade da população tem a mesma chance de pertencer à amostra.

Após conhecer a população o pesquisador deve identificar por nome ou por número cada unidade e, em seguida, sortear os elementos que constituirão à amostra.

Dois casos de amostra aleatória serão aqui estudados: Simples e estratificada.

4.2.2 Amostra Aleatória Simples

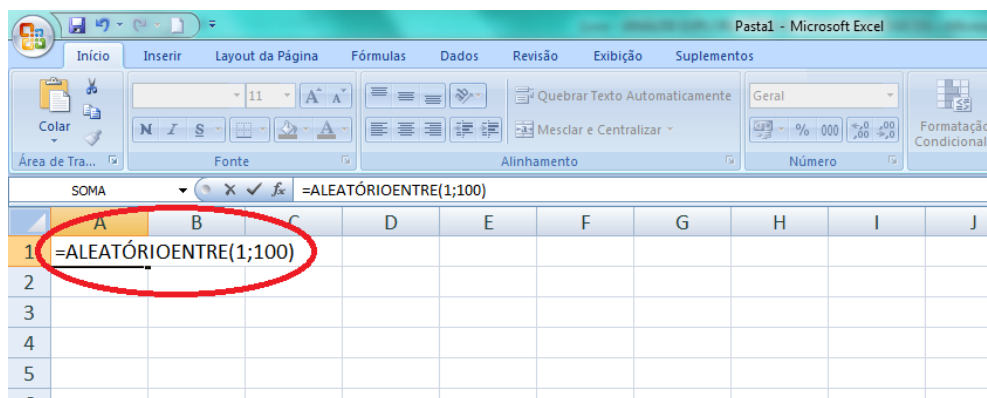
Quando temos uma população homogênea para a variável em estudo, utilizamos a **amostra aleatória simples** que é obtida através de sorteio das unidades da população. Neste cenário, cada ponto da população tem igual probabilidade de ser sorteado. Existem alguns procedimentos que podem ser utilizados para este sorteio, a ideia é um sorteio simples, como se cada elemento da população pudesse ser representado por um número colocado em um globo similar aquele utilizado em jogos da megasena, tendo a mesma chance de ser sorteado. Alguns livros apresentam algumas técnicas e tabelas de números aleatórios (que ao nosso entender já estão um pouco ultrapassados pelos recursos tecnológicos). Apontamos então para o uso de um programa simples¹ e bastante conhecido: O Microsoft *Excel*. Nesta planilha eletrônica existem algoritmos já inseridos para este fim. O exemplo abaixo discorre sobre estes procedimentos.

Imagine que um pesquisador queira escolher uma amostra aleatória de 10 indivíduos em uma população homogênea de 100 unidades observáveis. Ele pode numerar cada unidade e gerar os números aleatórios no Excel.

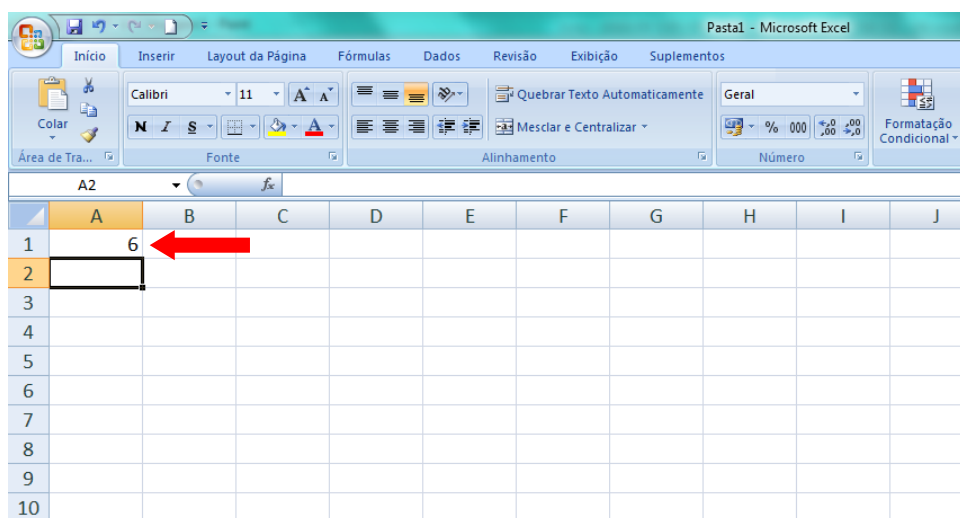
Para gerarmos os números aleatórios no *Excel* procedemos da seguinte forma:

1. Digite na célula A1 o comando: **=ALEATÓRIOENTRE(1;100)**

¹ Existem diversos programas com esta finalidade, alguns com uma vasta aplicação de recursos na área estatística como o caso do Programa R, SAS, SISVAR e outros.

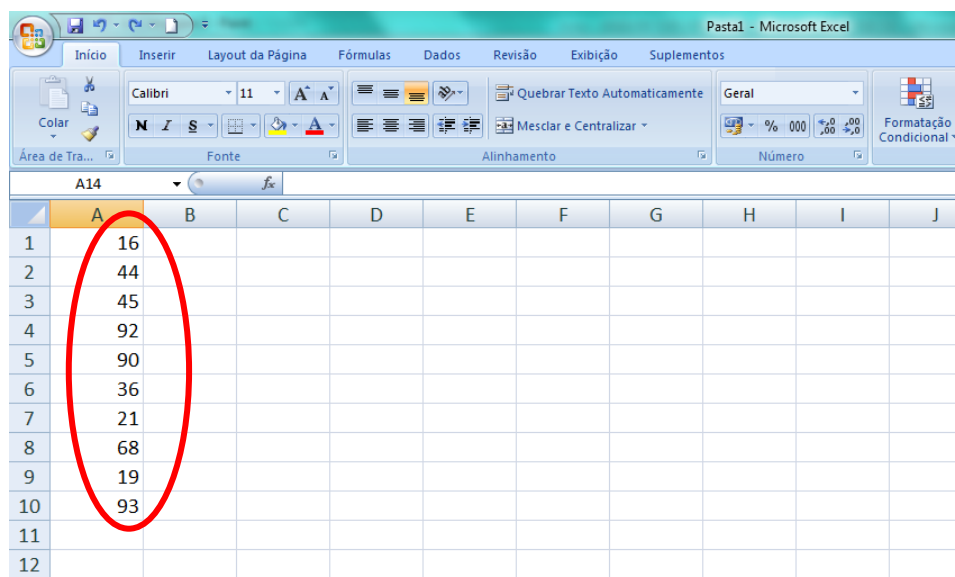


2. Tecle **ENTER**.



Neste exemplo o Excel gerou o número 6, no entanto, poderão aparecer em seu exemplo números entre 1 e 100.

3. Em seguida, copiamos o comando da célula A1 na célula A2 até a célula A10.



Observe que ao copiar o comando para as demais células o valor da primeira célula também muda. Isso garante que os números serão gerados aleatoriamente a cada nova célula copiada.

[Clique aqui](#) e assista ao vídeo tutorial sobre como sortear amostras aleatórias simples utilizando o Excel.

4.2.3 Amostra Aleatória Estratificada

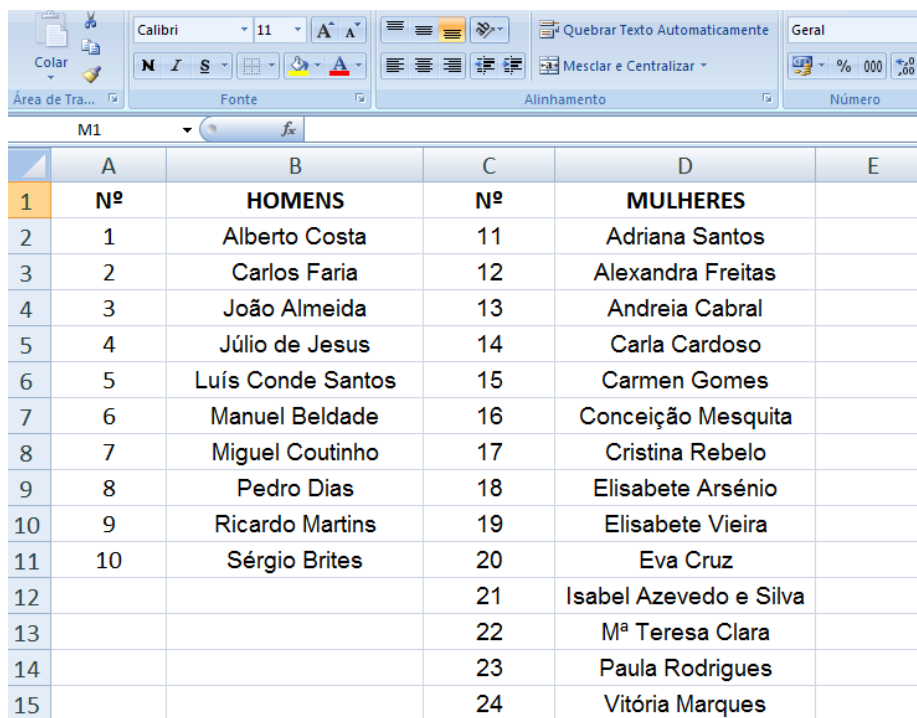
Usamos a *amostra aleatória estratificada* quando a população é constituída por *unidades heterogêneas* para a variável que se quer estudar. Nesse caso, as unidades da população devem ser identificadas; depois, as unidades similares devem ser reunidas em subgrupos chamados *estratos*. O sorteio é feito dentro de cada estrato.

Como exemplo, imagine que um pesquisador queira sortear indivíduos para uma entrevista num grupo de 24 pessoas, das quais 10 são homens e 14 são mulheres, ele optou por uma amostra aleatória estratificada e deseja sortear metade dos homens e metade das mulheres.

Para realizar o sorteio primeiramente dividi-se os indivíduos em dois grupos ou estratos, no nosso caso, homens e mulheres, em seguida, sorteia-se metade dos homens e metade das mulheres.

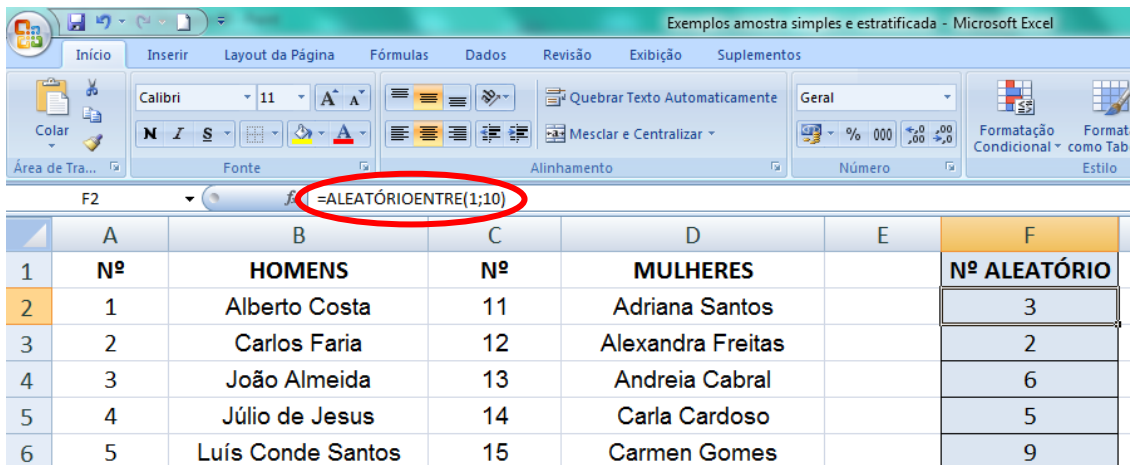
Para realizar o sorteio utilizando o Microsoft Excel procedemos da seguinte forma:

1. Construímos uma tabela contendo uma coluna com números ordenados e outra com os nomes dos indivíduos.



	A	B	C	D	E
1	Nº	HOMENS	Nº	MULHERES	
2	1	Alberto Costa	11	Adriana Santos	
3	2	Carlos Faria	12	Alexandra Freitas	
4	3	João Almeida	13	Andreia Cabral	
5	4	Júlio de Jesus	14	Carla Cardoso	
6	5	Luís Conde Santos	15	Carmen Gomes	
7	6	Manuel Beldade	16	Conceição Mesquita	
8	7	Miguel Coutinho	17	Cristina Rebelo	
9	8	Pedro Dias	18	Elisabete Arsénio	
10	9	Ricardo Martins	19	Elisabete Vieira	
11	10	Sérgio Brites	20	Eva Cruz	
12			21	Isabel Azevedo e Silva	
13			22	Mª Teresa Clara	
14			23	Paula Rodrigues	
15			24	Vitória Marques	

2. Digitamos em uma célula em branco do *Excel* o seguinte comando: =ALEATÓRIOENTRE(1;10). Com este comando o *Excel* gera um número aleatório entre 1 e 10, este número corresponderá ao indivíduo sorteado no grupo dos homens. Copiamos a fórmula nas outras células até que o número de indivíduos sorteados seja completado. Repetimos o mesmo procedimento para o grupo das mulheres, alterando apenas o intervalo de números aleatórios, que no nosso caso será de 11 a 24. Digitamos então o seguinte comando: =ALEATÓRIOENTRE(11;24). De forma similar o *Excel* gera um número aleatório entre 11 e 24 que corresponderá ao indivíduo sorteado do grupo das mulheres.



	A	B	C	D	E	F
1	Nº	HOMENS	Nº	MULHERES		Nº ALEATÓRIO
2	1	Alberto Costa	11	Adriana Santos		3
3	2	Carlos Faria	12	Alexandra Freitas		2
4	3	João Almeida	13	Andreia Cabral		6
5	4	Júlio de Jesus	14	Carla Cardoso		5
6	5	Luís Conde Santos	15	Carmen Gomes		9

3. Para que o *Excel* procure o nome correspondente ao número sorteado digitamos o seguinte comando:

=PROCV(valor_procurado;matriz_tabela;num_índice_coluna;[procurar_intervalo])



	A	B	C	D	E	F	G
1	Nº	HOMENS	Nº	MULHERES		Nº ALEATÓRIO	NOME
2	1	Alberto Costa	11	Adriana Santos		3	João Almeida
3	2	Carlos Faria	12	Alexandra Freitas		1	Alberto Costa
4	3	João Almeida	13	Andreia Cabral		9	Ricardo Martins
5	4	Júlio de Jesus	14	Carla Cardoso		8	Pedro Dias
6	5	Luís Conde Santos	15	Carmen Gomes		8	Pedro Dias
7	6	Manuel Beldade	16	Conceição Mesquita			
8	7	Miguel Coutinho	17	Cristina Rebelo		Nº ALEATÓRIO	NOME
9	8	Pedro Dias	18	Elisabete Arsénio		21	Isabel Azevedo e Silva
10	9	Ricardo Martins	19	Elisabete Vieira		24	Vitória Marques
11	10	Sérgio Brites	20	Eva Cruz		21	Isabel Azevedo e Silva
12			21	Isabel Azevedo e Silva		24	Vitória Marques
13			22	Mª Teresa Clara		20	Eva Cruz
14			23	Paula Rodrigues		15	Carmen Gomes
15			24	Vitória Marques		20	Eva Cruz

[Clique aqui](#) e assista ao vídeo tutorial sobre como sortear amostras aleatórias estratificadas utilizando o Excel.



5. ORGANIZAÇÃO DOS DADOS

Após a coleta das informações obtidas das variáveis que caracterizam os elementos da população ou da amostra em estudo, os dados devem ser analisados e interpretados com o auxílio de métodos estatísticos. A primeira etapa é fazer uma análise descritiva, ou seja, deve-se organizar e descrever os dados, para a identificação de valores que representem um elemento comum entre os dados e para a qualificação da variabilidade presente nos mesmos.

5.1 Dados Brutos

O que chamamos de *dados brutos* são os dados da forma que o pesquisador os obtém, ou seja, antes de qualquer organização ou tratamento estatístico. Por exemplo, na tabela a seguir estão apresentados os índices pluviométricos registrados num período de 50 dias na cidade de Cruzeiro do Sul – Ac.

Tabela 1 – Índice pluviométrico registrado num período de 50 dias na cidade de Cruzeiro do Sul – AC, no ano de 2016.

163	152	177	156	157	175	164	175	156	158
166	175	173	144	175	174	154	163	152	155
142	157	157	174	162	142	169	174	174	169
144	177	167	158	149	173	172	178	157	149
145	174	164	151	169	155	151	162	160	164

Fonte: Estação meteorológica da UFAC – Campus Floresta.

Perceba que os dados da forma que estão representados na tabela acima nos dão algumas informações, como por exemplo: os valores variam diariamente; o valor máximo do índice pluviométrico registrado foi 178 e o valor mínimo foi 142. No entanto, se a amostra fosse maior, precisaríamos de mais tempo e atenção para obter estas informações, pois os dados não estão organizados.

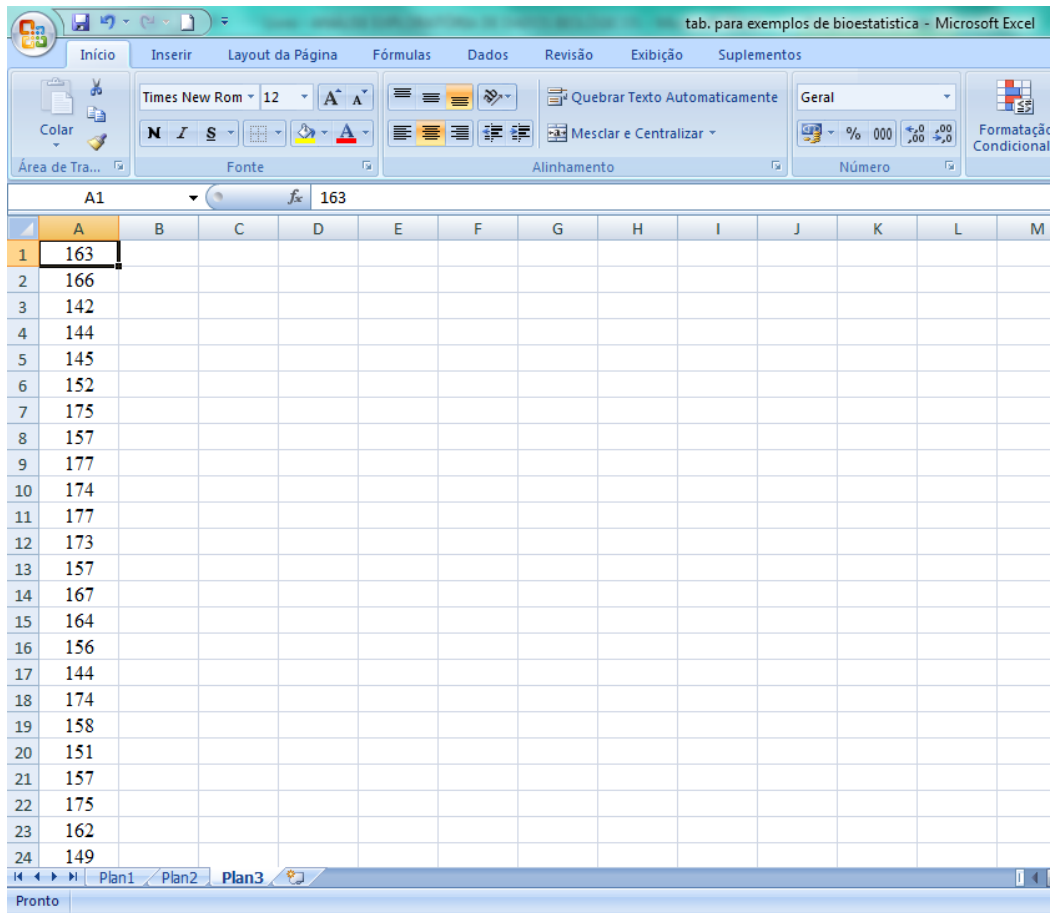
5.2 Rol

O que chamamos de *Rol* é a organização dos dados brutos em ordem crescente, essa organização serve para verificarmos quais os maiores e menores valores, além de verificarmos se alguns elementos se repetem.

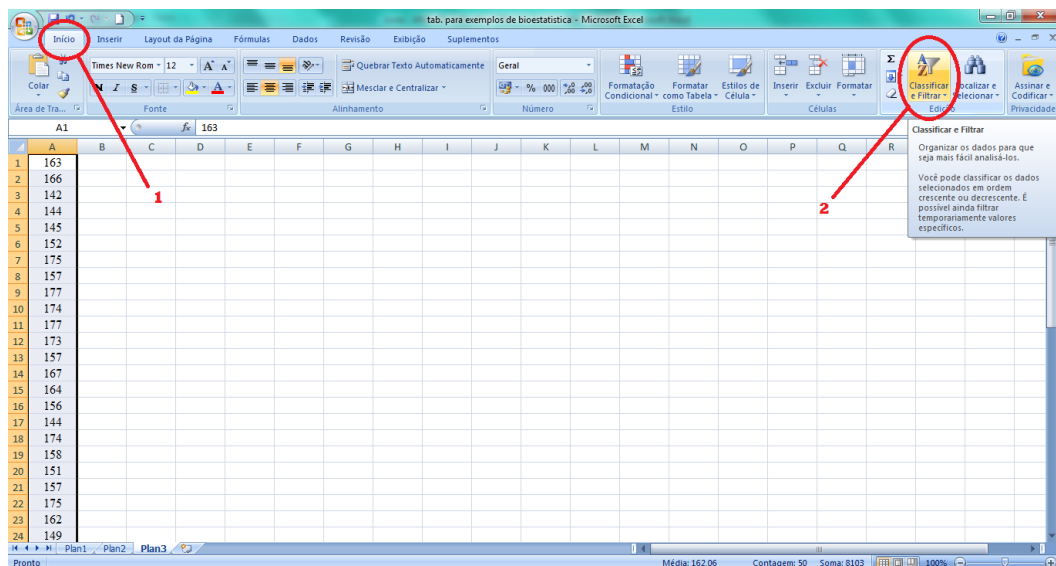
5.2.1 Organizando dados numéricos no Excel

Para organizar os dados em ordem crescente com o *Microsoft Excel*, realize os seguintes passos:

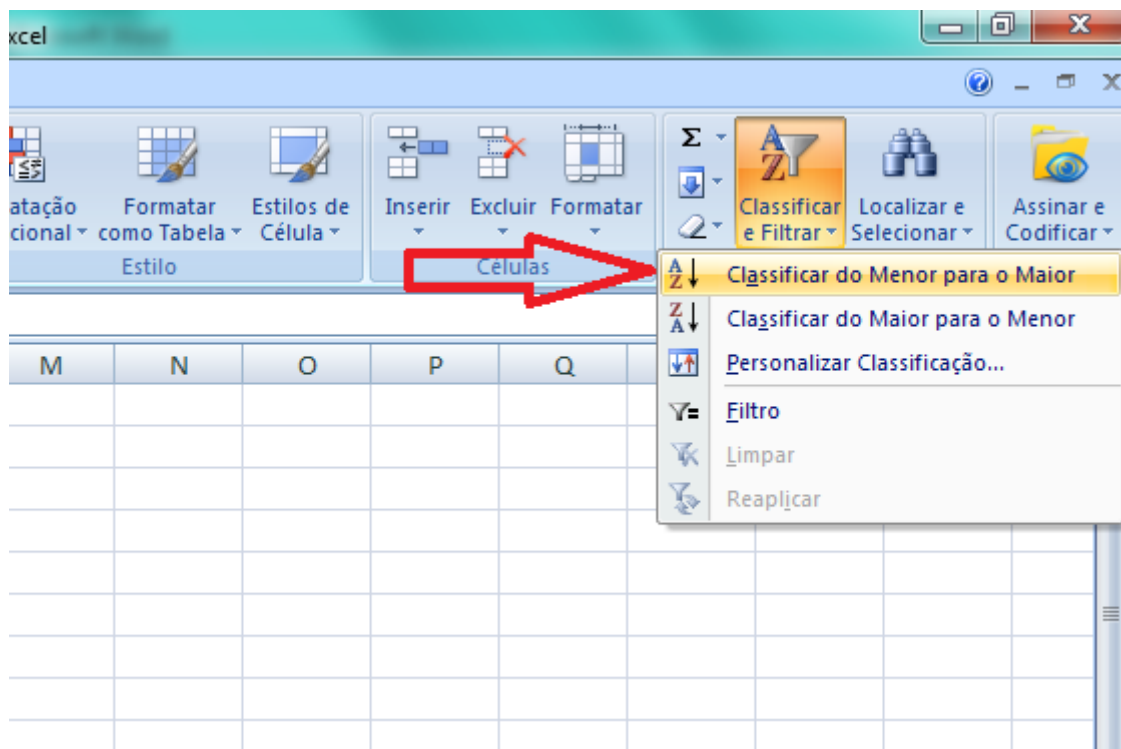
1. Digite os dados da Tabela 1 na coluna A, como mostra a figura abaixo.



2. Selecione os dados da coluna A, clique em **Início** na barra de ferramentas e em seguida no ícone **Classificar e Filtrar**.



3. Aparecerão as opções de classificação, clique em **Classificar do Menor para o Maior**.



Os dados serão classificados em ordem crescente. Logo, o Rol da tabela anterior ficaria assim representado:

Tabela 2 – Valores em ordem crescente do índice pluviométrico registrado num período de 50 dias na cidade de Cruzeiro do Sul – Ac.

142	149	152	156	158	163	166	172	174	175
142	149	154	157	158	163	167	173	174	175
144	151	155	157	160	164	169	173	174	177
144	151	155	157	162	164	169	174	175	177
145	152	156	157	162	164	169	174	175	178

Com os dados organizados em ordem crescente as informações se tornam mais fáceis de serem coletadas minimizando o tempo utilizado nas análises e diminuindo a ocorrência de erros.

[Clique aqui](#) para assistir ao vídeo tutorial sobre como organizar dados numéricos em um Rol utilizando o Excel.

6. TABELAS DE DISTRIBUIÇÃO DE FREQUÊNCIA

6.1 Tabela de Distribuição de Frequência simples

As tabelas de distribuição de frequência organizam os dados de uma maneira mais eficiente, facilitando a compreensão e interpretação dos mesmos. As tabelas de distribuição de frequência simples mostram os valores obtidos e o número de vezes que cada valor foi observado.

6.1.2 Gerando tabelas de distribuição de frequência simples com o Action

Para exemplificar, consideremos os dados obtidos em uma pesquisa sobre a idade de 50 alunos da Universidade Federal do Acre – UFAC, Campus Floresta, na cidade de Cruzeiro do Sul – AC, no ano de 2015.

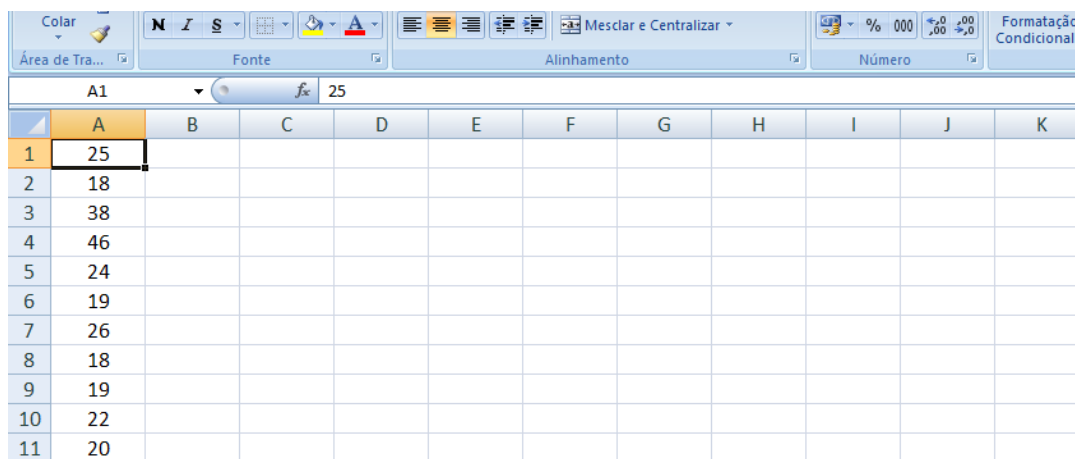
Tabela 3 – Idade em anos de 50 alunos da Universidade Federal do Acre – UFAC, Campus Floresta, na cidade de Cruzeiro do Sul – AC, no ano de 2015.

25	20	23	19	24
18	19	25	19	26
38	19	23	18	29
46	25	30	23	19
24	22	20	28	18
19	22	21	22	29
26	19	22	26	26
18	21	18	24	20
19	26	23	20	35
22	21	18	19	26

Fonte: Alunos do 3º período do curso de Bacharelado em Ciências Biológicas da UFAC – Campus Floresta

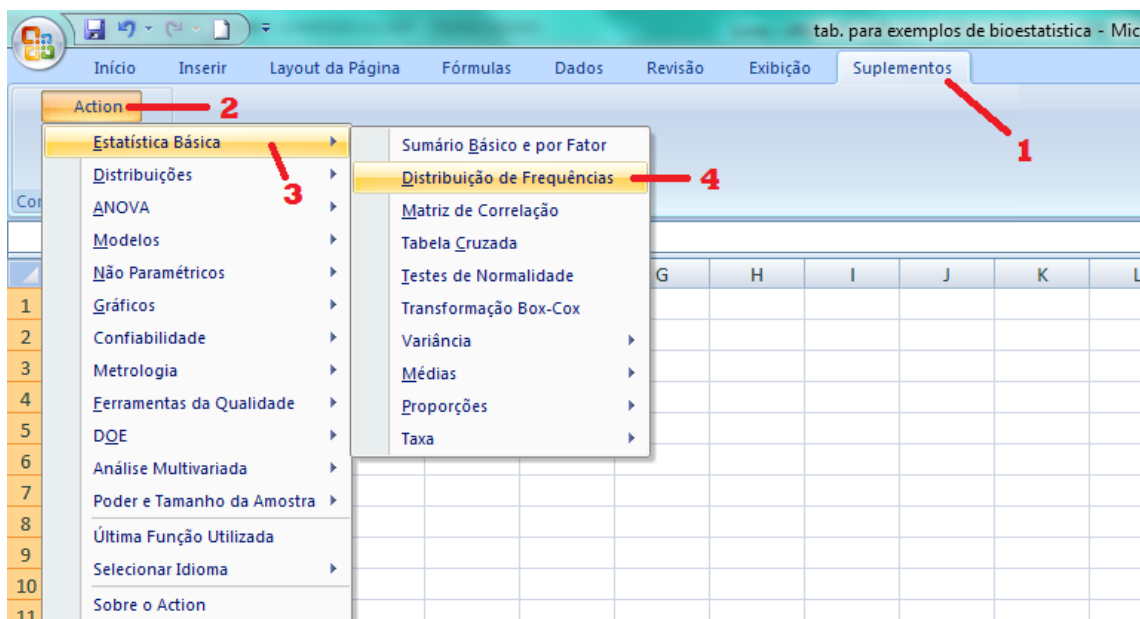
Para construir a tabela de distribuição de frequência dos dados da Tabela 3, com o auxílio de Suplemento *Action* do *Excel* siga os seguintes passos:

1. Digite os dados da Tabela 3, na Coluna A, conforma a figura abaixo.

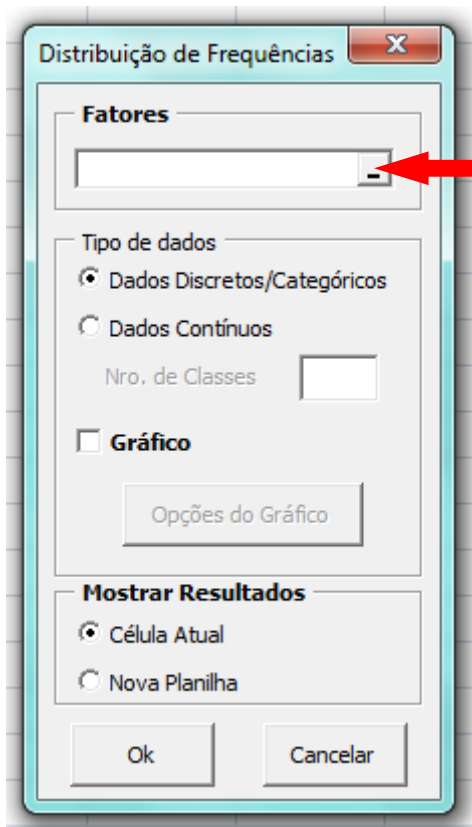


	A	B	C	D	E	F	G	H	I	J	K
1	25										
2	18										
3	38										
4	46										
5	24										
6	19										
7	26										
8	18										
9	19										
10	22										
11	20										

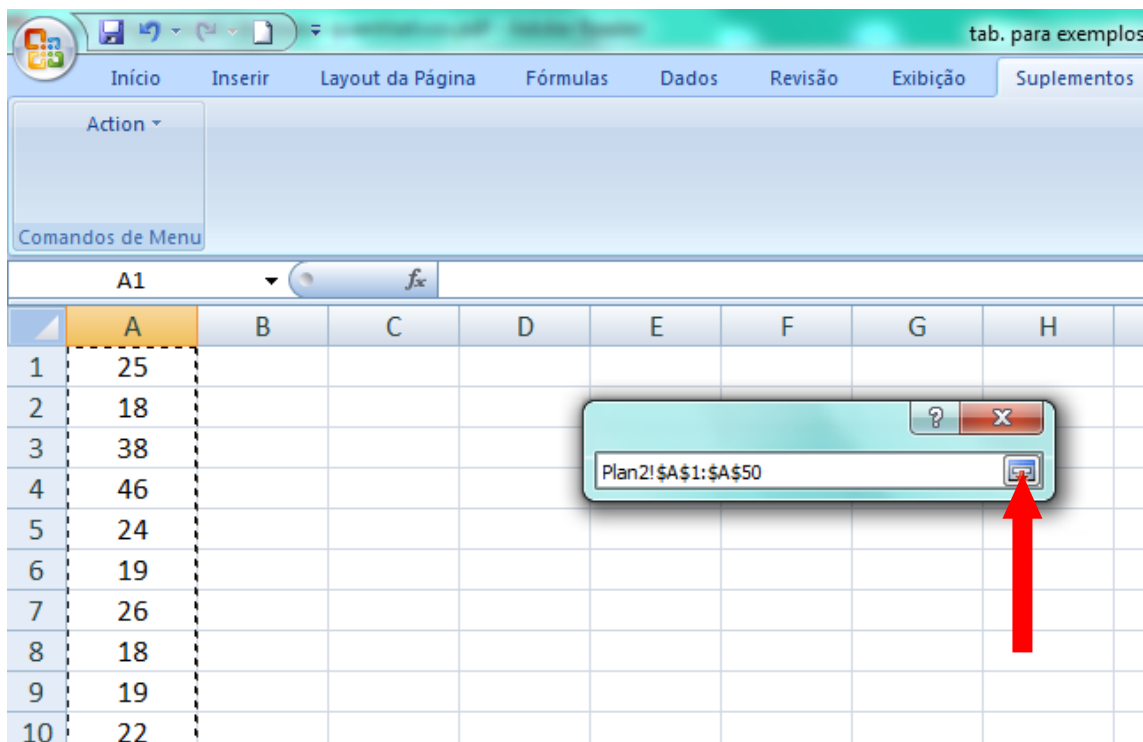
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.



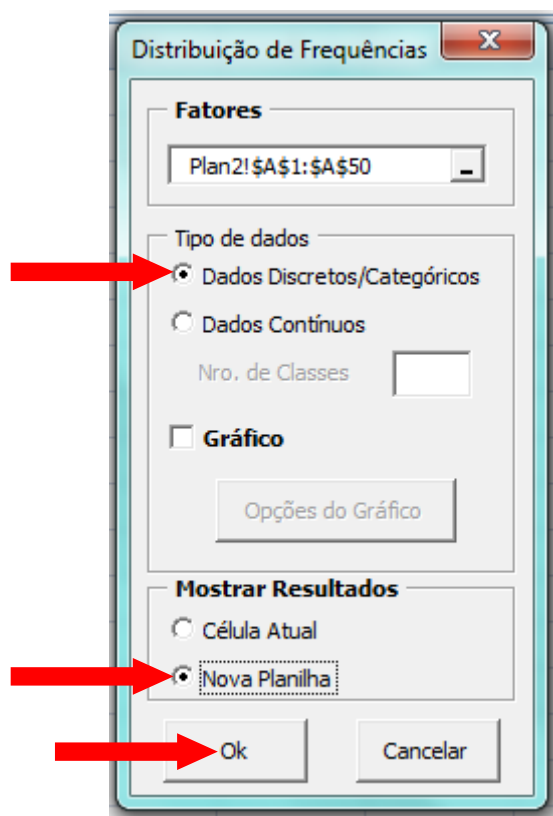
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique no canto direito da caixa de **Fatores** conforme indicado na figura abaixo, em seguida selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A50.



4. Clique novamente no canto direito da caixa de **Fatores** conforme indicado na figura abaixo.



5. Clique em Dados **Discretos/Categóricos**, depois em **Nova Planilha** e clique **Ok**.





Com estes passos o *Excel* construirá a tabela de distribuição de frequência simples dos dados selecionados, com os fatores, a frequência absoluta, a frequência relativa, a frequência percentual e a frequência percentual acumulada, conforme figura abaixo.

Comandos de Menu						
A5		f _x Fatores				
	A	B	C	D	E	F
1	DISTRIBUIÇÃO DE FREQUÊNCIAS					
2						
3	DADOS DO PROCESSO					
4						
5	Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.	
6	18	6	0,12	12	12	
7	19	9	0,18	18	30	
8	20	4	0,08	8	38	
9	21	3	0,06	6	44	
10	22	5	0,1	10	54	
11	23	4	0,08	8	62	
12	24	3	0,06	6	68	
13	25	3	0,06	6	74	
14	26	6	0,12	12	86	
15	28	1	0,02	2	88	
16	29	2	0,04	4	92	
17	30	1	0,02	2	94	
18	35	1	0,02	2	96	
19	38	1	0,02	2	98	
20	46	1	0,02	2	100	
21	Total	50				

[Clique aqui](#) para assistir ao vídeo tutorial sobre como gerar tabelas de distribuição de frequência simples utilizando o suplemento *Action* de *Excel*.

O exemplo a seguir utiliza dados qualitativos sobre gênero, espécie e tipo de captura de abelhas coletados pelo Prof^o Dr. Rogério Oliveira Souza na região de Santa Luzia, localizada no município de Cruzeiro do Sul, no período de agosto de 2010.

Tabela 4 – Gênero, espécie e tipo de captura de abelhas na Região de Santa Luzia, Cruzeiro do Sul – Acre.

Gênero	Espécie	Captura	Gênero	Espécie	Captura
Eulaema	Elmeriana	Cineol	Euglossa	Egbidentata	Salicilato
Eulaema	Elmeriana	Cineol	Euglossa	Egmixta	Salicilato
Eulaema	Elmeriana	Cineol	Euglossa	Egmixta	Salicilato
Eulaema	Elmeriana	Cineol	Euglossa	Egignita	Salicilato
Eulaema	Elmeriana	Cineol	Euglossa	Egimperialis	Salicilato
Eulaema	Elmeriana	Salicilato	Euglossa	Egimperialis	Salicilato
Eulaema	Elmeriana	Salicilato	Euglossa	Egimperialis	Salicilato
Eufriesea	Efpulchra	Salicilato	Euglossa	Egignita	Salicilato
Euglossa	Egmixta	Salicilato	Euglossa	Egignita	Salicilato

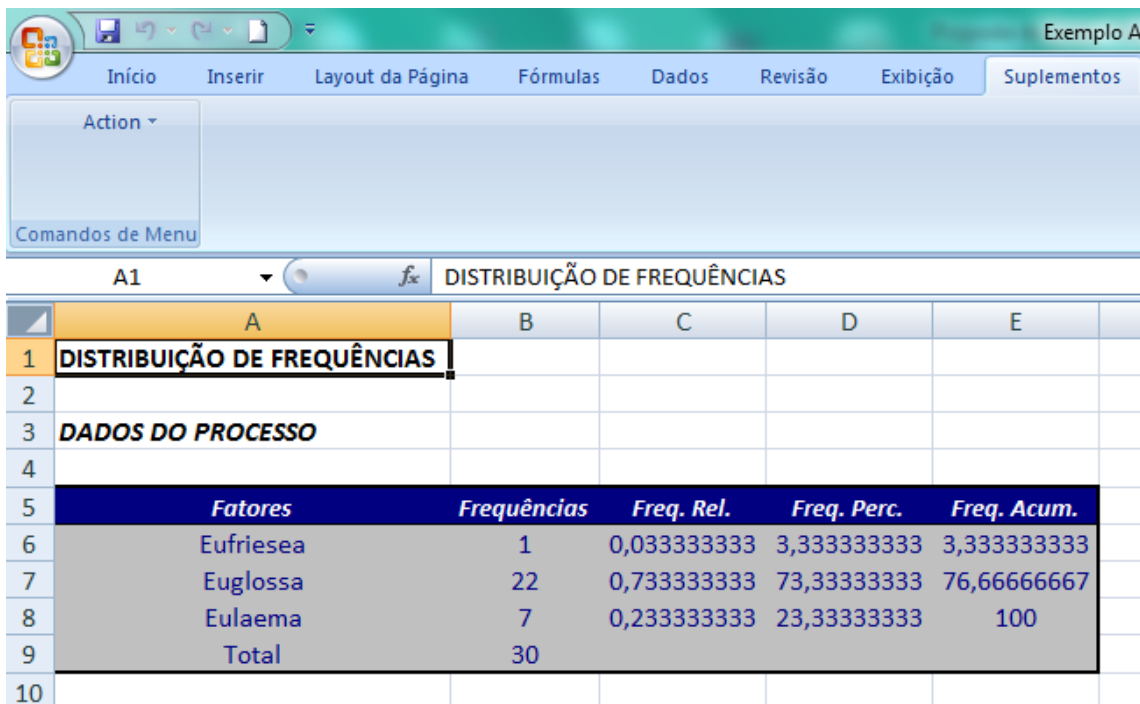
Euglossa	Egmixta	Salicilato	Euglossa	Egignita	Salicilato
Euglossa	Egmixta	Salicilato	Euglossa	Egignita	Salicilato
Euglossa	Egmixta	Salicilato	Euglossa	Egoccidentalis	Cineol
Euglossa	Egcognata	Salicilato	Euglossa	Egignita	Cineol
Euglossa	Egmixta	Salicilato	Euglossa	Egoccidentalis	Cineol
Euglossa	Egcognata	Salicilato	Euglossa	Egimperialis	Cineol

Fonte: Prof^o Dr. Rogério Oliveira Souza

Para construirmos a tabela de distribuição de frequência da variável *Gênero* da Tabela 4, com o auxílio de Suplemento *Action* do *Excel* procedemos da seguinte forma:

1. Digite os dados da variável *Gênero* da Tabela 4, na Coluna A do Excel.
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique na caixa de **Fatores**, em seguida selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A30.
4. Clique em Dados **Discretos/Catégoricos**, depois em **Nova Planilha** e clique **Ok**.

Com estes passos o *Action* construirá a tabela de distribuição de frequência simples dos dados da variável *Gênero*, com os fatores, a frequência absoluta, a frequência relativa, a frequência percentual e a frequência percentual acumulada, conforme figura abaixo.



The screenshot shows the Excel interface with the 'Action' dialog box open. The 'Action' menu is set to 'Distribuição de Frequência'. The resulting table is as follows:

DISTRIBUIÇÃO DE FREQUÊNCIAS					
DADOS DO PROCESSO					
	Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
6	Eufriesea	1	0,033333333	3,333333333	3,333333333
7	Euglossa	22	0,733333333	73,33333333	76,66666667
8	Eulaema	7	0,233333333	23,33333333	100
9	Total	30			

[Clique aqui](#) e assista ao vídeo tutorial.

Exercício proposto

Para exercitar o que aprendeu até agora sobre a construção de tabelas de frequência simples, construa uma tabela de distribuição de frequência simples para as variáveis: Espécie e Captura da Tabela 4, utilizando o *Action*.

6.2 Interpretação da tabela de distribuição de frequência gerada pelo *Action*

Passaremos agora a interpretar cada coluna da tabela de distribuição de frequência gerada pelo *Action*.

6.2.1 Fatores

Os fatores (x) são todos os valores observados no conjunto de dados selecionados.

Em nossos exemplos os fatores são a *Idade* no primeiro caso e o *Gênero* no segundo caso.

6.2.2 Frequências (Frequência Absoluta)

A coluna Frequências (f) nos dá a frequência absoluta, que é o número de vezes que cada valor da coluna *Fatores* foi observado. Por exemplo, o número 18 foi observado 6 vezes no conjunto de dados sobre a idades dos alunos.

25	20	23	19	24
18	19	25	19	26
38	19	23	18	29
46	25	30	23	19
24	22	20	28	18
19	22	21	22	29
26	19	22	26	26
18	21	18	24	20
19	26	23	20	35
22	21	18	19	26

5	Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
6	18	6	0,12	12	12
7	19	9	0,18	18	30
8	20	4	0,08	8	38
9	21	3	0,06	6	44
10	22	5	0,1	10	54
11	23	4	0,08	8	62
12	24	3	0,06	6	68
13	25	3	0,06	6	74



6.2.3 Freq. Rel. (Frequência Relativa)

A Frequência Relativa (fr) é a proporção com que cada valor ocorre. É obtida dividindo-se cada valor da frequência absoluta pela sua soma (Σ).

$$fr = \frac{f}{\Sigma f}$$

Onde,

fr : Frequência relativa

f : Frequência absoluta

Σf : Somatório da frequência absoluta

Na primeira linha e terceira coluna da tabela de distribuição de frequência gerada pelo *Action* dos dados sobre a idade dos alunos, observamos o número 0,12 que foi obtido da seguinte forma:

$$fr = \frac{6}{50} = 0,12$$

Os demais números da coluna **Freq. Rel.** foram obtidos de forma semelhante. Sempre que possível é recomendável utilizarmos a frequência relativa em detrimento da absoluta, tendo em vista que com essa podemos comparar dados de pesquisas semelhantes mais com tamanhos amostrais diferentes. Por exemplo, em termos absolutos poderíamos dizer que o número 20 é maior do que o número 10, porém, se os tamanhos amostrais para este caso fossem respectivamente 1000 e 50, teríamos o número 20 correspondendo a 2% do total, enquanto que o número 10 corresponderia a 20%, ou seja, neste caso, o número 10 é dez vezes mais representativo do que o número 20.

6.2.4 Freq. Perc. (Frequência Percentual %)

A frequência percentual apresenta o valor da frequência relativa na forma percentual (%). Com isso, podemos observar mais claramente a proporção com que cada valor ocorre. Em algumas situações, esta frequência poderá ser interpretada como uma estimativa da probabilidade de encontramos valores iguais aos dos dados coletados se realizamos a mesma pesquisa com grupos semelhantes.

6.2.5 Freq. Acum. (Frequência Percentual Acumulada)

A frequência percentual acumulada é obtida somando-se a frequência (f) da linha desejada (x) com as frequências dos valores de x menores do que o considerado. Por isso, sua soma não precisa ser realizada. A frequência Percentual Acumulada indica o valor percentual acumulado igual ou menor do que um determinado elemento do Fator. Por exemplo, na linha cinco da coluna **Freq. Acum.** da tabela do exemplo anterior, está o número 54, isso significa que 54% dos alunos entrevistados tem idade igual ou menor que 22 anos.

Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
18	6	0,12	12	12
19	9	0,18	18	30
20	4	0,08	8	38
21	3	0,06	6	44
22	5	0,1	10	54
23	4	0,08	8	62
24	3	0,06	6	68
25	3	0,06	6	74

6.3 Tabelas de Distribuição de Frequência com Intervalos de Classe

Quando os valores de uma característica variam muito, como é o caso da estatura e o peso das pessoas, uma tabela de distribuição de frequência simples tenderia a ser muito extensa, perdendo sua característica de facilitar a interpretação e a compreensão dos valores. A solução é agrupar os dados por intervalos de classe. Isso é recomendável quando a variável estuda é do tipo contínua, podendo assumir qualquer valor num determinado intervalo.

Para exemplificar, observemos uma pesquisa realizada pelos alunos do Curso de Bacharelado em Ciências Biológicas sobre a estatura de 50 alunos da Universidade Federal do Acre – UFAC, Campus Floresta, na cidade de Cruzeiro do Sul – AC, no ano de 2015.

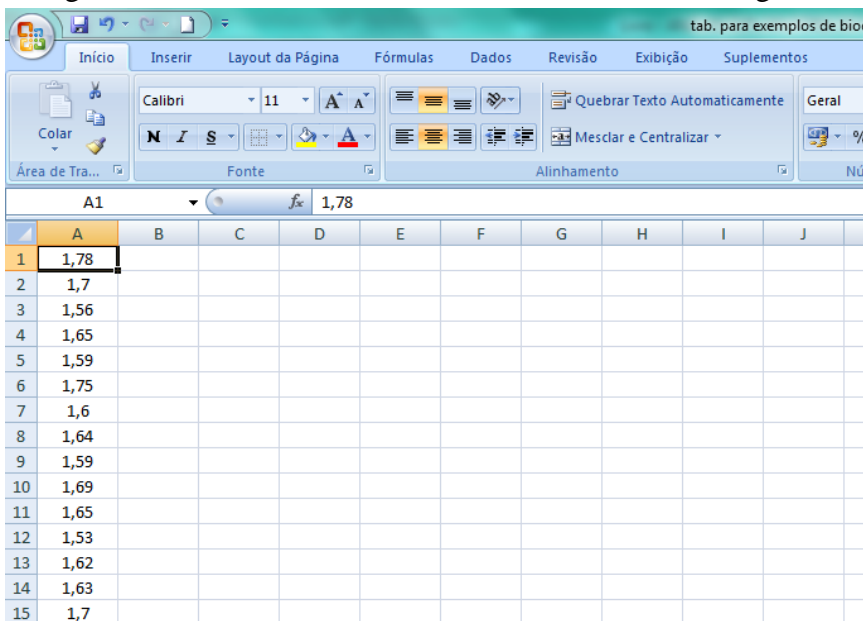
Tabela 5 – Estatura em metros de 50 alunos da Universidade Federal do Acre – UFAC, Campus Floresta, na cidade de Cruzeiro do Sul – AC, no ano de 2015.

1,78	1,65	1,58	1,58	1,59
1,7	1,53	1,69	1,8	1,69
1,56	1,62	1,64	1,73	1,76
1,65	1,63	1,82	1,7	1,77
1,59	1,7	1,69	1,58	1,67
1,75	1,65	1,72	1,55	1,6
1,6	1,67	1,54	1,58	1,73
1,64	1,61	1,78	1,67	1,83
1,59	1,7	1,6	1,7	1,68
1,69	1,55	1,63	1,77	1,58

Fonte: Alunos do 3º período do curso de Bacharelado em Ciências Biológicas da UFAC – Campus Floresta.

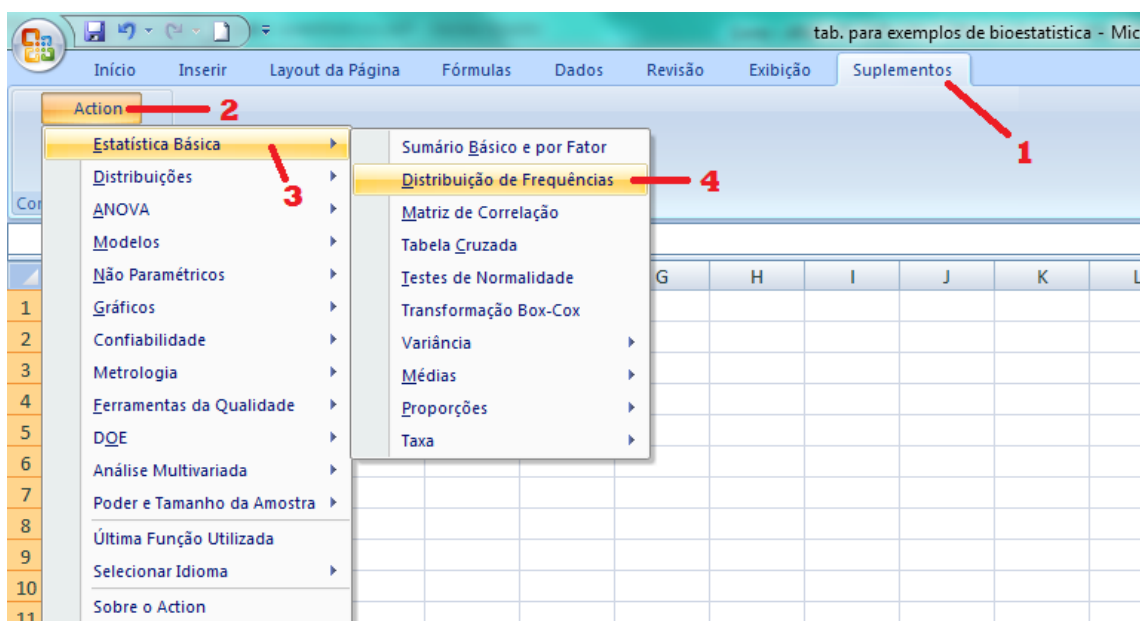
Para construirmos a tabela de distribuição de frequência com intervalos de classe dos dados da Tabela 5, utilizando o *Action* siga os seguintes passos:

1. Digite os dados da Tabela 5, na Coluna A, conforma a figura abaixo.

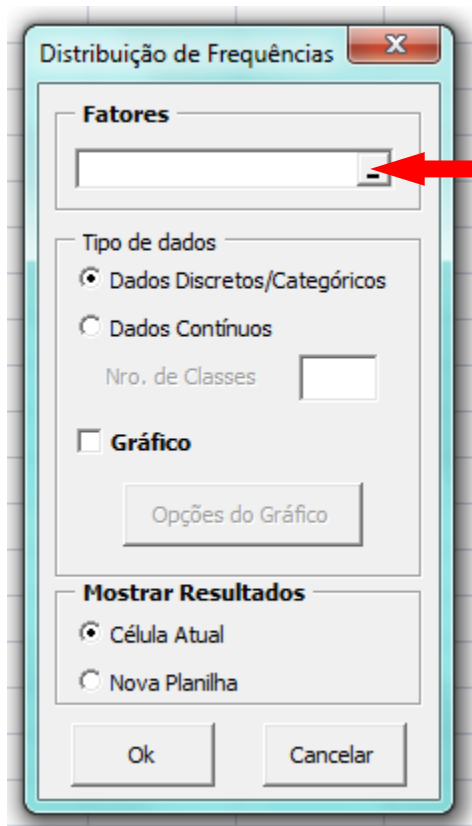


	A	B	C	D	E	F	G	H	I	J
1	1,78									
2	1,7									
3	1,56									
4	1,65									
5	1,59									
6	1,75									
7	1,6									
8	1,64									
9	1,59									
10	1,69									
11	1,65									
12	1,53									
13	1,62									
14	1,63									
15	1,7									

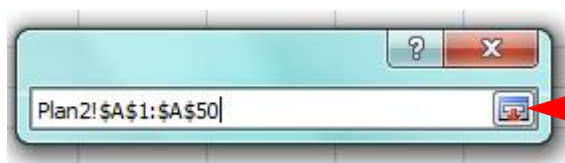
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.



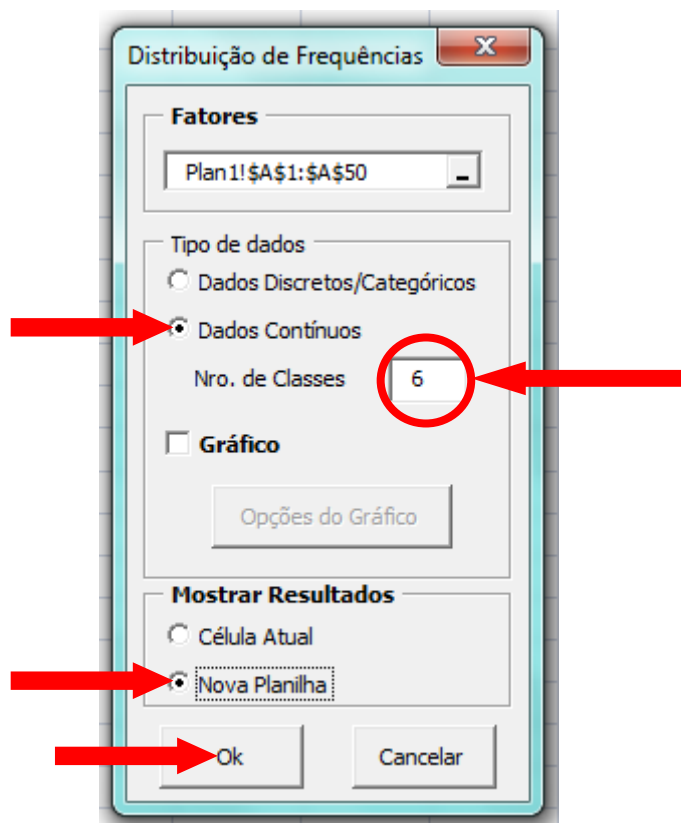
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique no canto direito da caixa de **Fatores** conforme indicado na figura a seguir, após selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A50.



4. Clique novamente no canto direito da caixa de **Fatores** conforme indicado na figura abaixo.



5. Clique em **Dados Contínuos**, e em seguida, escolha quantas classes terá sua tabela. Neste exemplo, decidimos que nossa tabela terá seis linhas com intervalos de classe, então digitamos o número 6 na caixa **Nro. De Classe**, depois é só clicar em **Nova Planilha** e então **Ok**.



Seguindo estes passos, o *Action* construirá a tabela de distribuição de frequência com intervalos de classe dos dados selecionados, com as classes, a frequência absoluta, a frequência relativa, a frequência percentual, a frequência percentual acumulada, a densidade e o ponto médio do intervalo.

Tabela 6 – Tabela de distribuição de frequência com intervalo de classe da estatura em metros de 50 alunos da Universidade Federal do Acre – UFAC, gerada pelo *Action*.

	A	B	C	D	E	F	G
1	DISTRIBUIÇÃO DE FREQUÊNCIAS						
2							
3	<i>Classe</i>	<i>Frequência</i>	<i>Freq. Relativa</i>	<i>Porcentagem</i>	<i>Porc. Acumulada</i>	<i>Densidades</i>	<i>Ponto médio</i>
4	[1,53 ; 1,58)	5	0,1	10	10	2	1,555
5	[1,58 ; 1,63)	13	0,26	26	36	5,2	1,605
6	[1,63 ; 1,68)	10	0,2	20	56	4	1,655
7	[1,68 ; 1,73)	11	0,22	22	78	4,4	1,705
8	[1,73 ; 1,78)	6	0,12	12	90	2,4	1,755
9	[1,78 ; 1,83)	5	0,1	10	100	2	1,805
10							

[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar tabelas de distribuição de frequência com intervalo de classe utilizando o *Action*.



6.4 Interpretação da tabela de distribuição de frequência com intervalos de classe gerada pelo *Action*

6.4.1 Classes

As *classes* são subgrupos dos valores observados no conjunto de dados selecionados. No nosso exemplo, dividimos os dados em seis grupos (classes). Na literatura existem algoritmos para se determinarem esses números. Dessa forma, o *Action* subdividiu-os em intervalos de 0,05 cm dentro de cada classe. O *Action* simboliza esse intervalo com o ponto e vírgula, no entanto, também são utilizados os símbolos (┌), que significa, inclui o limite inferior e exclui o limite superior.

O cálculo matemático realizado para determinar o intervalo de classes é o seguinte:

1) Encontramos o menor e maior valor da variável:

$$X_{\text{máximo}} = 1,83 \text{ m}$$

$$X_{\text{mínimo}} = 1,53 \text{ m}$$

2) Calculamos a *Amplitude* subtraindo o maior pelo menor valor:

$$A = X_{\text{máximo}} - X_{\text{mínimo}} = 1,83 - 1,53 = \mathbf{0,30}$$

3) Para determinar o intervalo de cada classe, basta dividir a amplitude pelo número de classes em que desejamos agrupar os dados da variável. No nosso exemplo escolhemos o número 6. Esse valor pode ser calculado por meio da fórmula de Sturges da seguinte forma:

$$K \cong 1 + 3,22 \times \log(n)$$

onde,

n = tamanho da amostra

$$K \cong 1 + 3,22 \times \log(50) \qquad K = 6,47$$

Usando o arredondamento encontramos o número 6, que é o número de linhas que terá nossa tabela de frequência com intervalos de classe.

$$IC = \frac{A}{K} = \frac{0,30}{6} = \mathbf{0,05}$$

Onde:

IC: intervalo de classe

A: amplitude

K: número de classes (linhas)



6.4.2 Frequências (Frequência Absoluta)

A coluna Frequências nos dá a frequência com que cada valor foi observado dentro do intervalo definido. Por exemplo, o número 5 na primeira linha da coluna frequência indica que existem cinco alunos com estatura entre 1,53 e 1,58 m, ou seja, só são contados os alunos com estatura até 1,57 m. Se existir algum alunos com estatura de 1,58 m, este será contado no intervalo seguinte, pois o intervalo inclui o limite inferior e exclui o limite superior.

O cálculo da frequência relativa, da porcentagem e da porcentagem acumulada é semelhante aos das tabelas de frequência simples, já explicados anteriormente.

6.4.3 Densidade

O cálculo da densidade é obtido da seguinte forma:

$$d = \frac{fr}{IC}$$

Na tabela do nosso exemplo o valor da densidade do primeiro intervalo é 2.

$$d = \frac{0,1}{0,05} = 2$$

Os demais valores foram obtidos da mesma forma.

6.4.4 Ponto Médio

O ponto médio é a média entre o limite interior e o limite superior de cada classe. Assim, representando por \bar{X}_i o ponto médio do intervalo da classe i , temos:

$$\bar{X}_i = \frac{l_i + L_i}{2}$$

em que l_i representa o limite inferior da classe e L_i o limite superior da mesma classe.

Para exemplificar, observe na quarta linha da tabela de frequência gerada pelo *Action*, encontramos o número 1,705 referente ao ponto médio do intervalo de classe. Este valor foi obtido da seguinte forma:

$$\bar{X}_i = \frac{1,68 + 1,73}{2}$$

$$\bar{X}_i = 1,705$$



	A	B	C	D	E	F	G
1	DISTRIBUIÇÃO DE FREQUÊNCIAS						
2							
3	<i>Classe</i>	<i>Frequência</i>	<i>Freq. Relativa</i>	<i>Porcentagem</i>	<i>Porc. Acumulada</i>	<i>Densidades</i>	<i>Ponto médio</i>
4	[1,53 ; 1,58)	5	0,1	10	10	2	1,555
5	[1,58 ; 1,63)	13	0,26	26	36	5,2	1,605
6	[1,63 ; 1,68)	10	0,2	20	56	4	1,655
7	[1,68 ; 1,73)	11	0,22	22	78	4,4	1,705
8	[1,73 ; 1,78)	6	0,12	12	90	2,4	1,755
9	[1,78 ; 1,83)	5	0,1	10	100	2	1,805
10							

6.4.5 Interpretação da Densidade e do ponto médio

A densidade e o ponto médio são valores gerados pelo *Action* para a construção dos histogramas que são gráficos de variáveis quantitativas contínuas, assunto que estudaremos a seguir. Podemos interpretar o valor da densidade como sendo a altura das barras do gráfico. O ponto médio é a escolha de um número que represente toda a classe para efeito do cálculo de algumas medidas matemáticas. Sabe-se que matematicamente num intervalo de número reais, existem infinitos pontos, então, para realização dos cálculos é necessário utilizar um número que possa representar toda a classe.



7. GRÁFICOS

Os gráficos são uma maneira de representar os dados de uma pesquisa com mais clareza e simplicidade, nos dando uma visão mais imediata do comportamento dos dados, permitindo chegar-se a conclusões sobre a evolução de fenômenos ou como relacionam-se os valores de uma série, por isso devem ser autoexplicativos e de fácil compreensão.

7.1 Gráficos para variáveis qualitativas

7.1.1 Gráfico em setores (Pizza)

O gráfico em setores ou gráfico de pizza é um gráfico mais usado para a apresentação de variáveis qualitativas nominais, mas também pode ser usado em variáveis quantitativas discretas. O gráfico em setores nos dá uma noção da proporção de cada setor em relação ao todo.

Para construirmos o gráfico em setores para a variável “Curso” da tabela 7 abaixo, utilizando o *Action*, procedemos da seguinte forma:

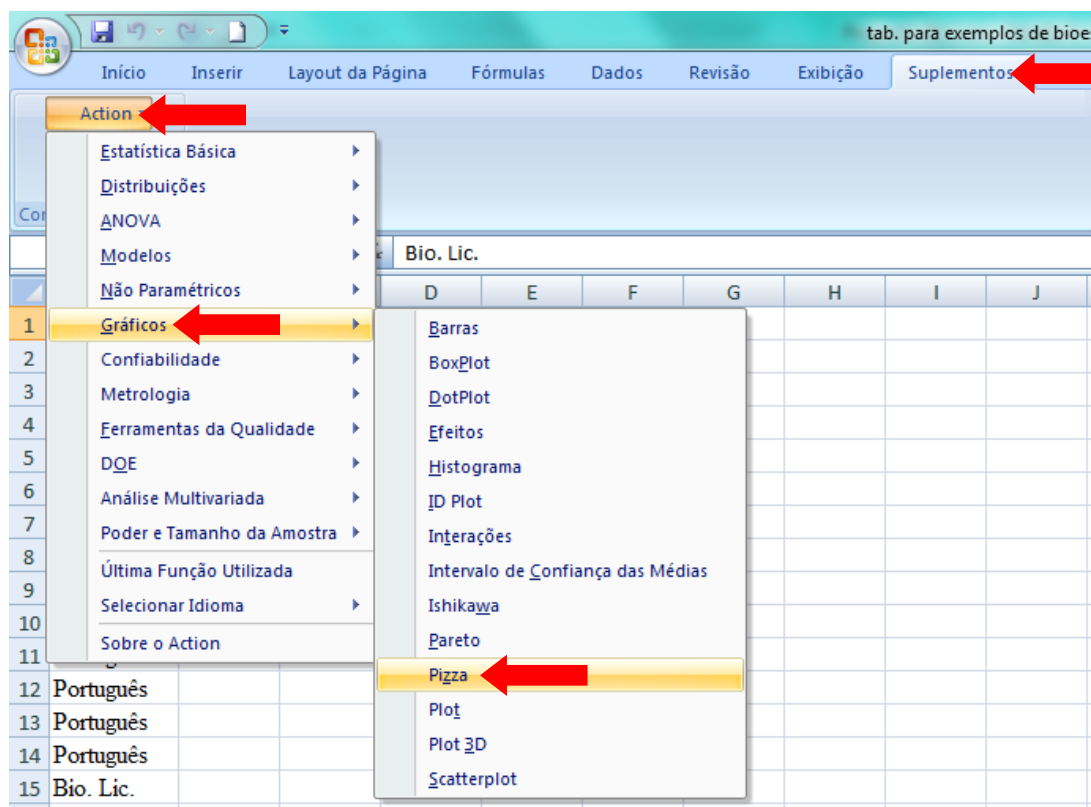
Tabela 7 – Pesquisa realizada com 50 alunos da Universidade Federal do Acre – UFAC – Campus Floresta, na cidade de Cruzeiro do Sul – Acre.

Curso	Sexo	Cor/Raça	Estado Civil	Curso	Sexo	Cor/Raça	Estado Civil
Bio. Lic.	M	Preto	Solteiro	Enfermagem	M	Parda	Casado
Matemática	F	Parda	Solteiro	Bio. Lic.	F	Parda	Casado
Matemática	F	Parda	Casado	Eng. Agro.	M	Branca	Solteiro
Matemática	M	Branco	Casado	Pedagogia	F	Parda	Solteiro
Bio. Lic.	F	Branco	Solteiro	Letras Inglês	F	Branca	Solteiro
Matemática	M	Parda	Casado	Letras Inglês	F	Parda	Solteiro
Matemática	F	Parda	Solteiro	Letras Inglês	M	Parda	Solteiro
Matemática	F	Parda	Solteiro	Letras Inglês	F	Branca	Solteiro
Matemática	F	Parda	Solteiro	Bio. Lic.	F	Parda	Solteiro
Português	M	Branco	Solteiro	Bio. Lic.	F	Branca	Solteiro
Português	F	Parda	Solteiro	Letras Inglês	F	Parda	Solteiro
Português	M	Parda	Solteiro	Bio. Lic.	F	Parda	Solteiro
Português	F	Preto	Solteiro	Bio. Lic.	F	Parda	Solteiro
Português	M	Parda	Casado	Bio. Lic.	M	Parda	Solteiro
Bio. Lic.	M	Parda	Solteiro	Bio. Bach.	F	Parda	Solteiro
Enfermagem	M	Parda	Solteiro	Bio. Lic.	F	Parda	Casado
Matemática	F	Parda	Solteiro	Enfermagem	F	Parda	Casado
Bio. Lic.	F	Branco	Solteiro	Bio. Bach.	F	Parda	Casado

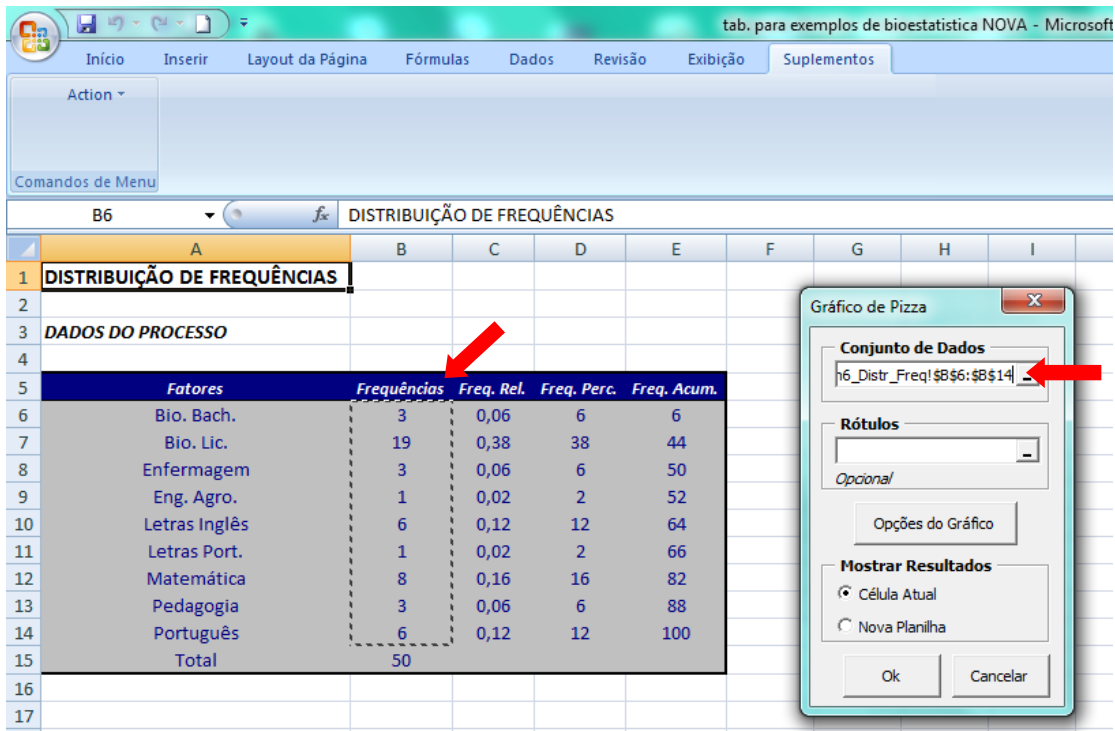
Bio. Lic.	F	Parda	Casado	Bio. Bach.	M	Parda	Solteiro
Bio. Lic.	F	Parda	Solteiro	Bio. Lic.	F	Branca	Solteiro
Bio. Lic.	F	Branco	Solteiro	Bio. Lic.	F	Parda	Solteiro
Bio. Lic.	F	Parda	Casado	Letras Port.	F	Parda	Solteiro
Português	F	Parda	Casado	Letras Inglês	M	Parda	Solteiro
Pedagogia	M	Parda	Casado	Pedagogia	F	Parda	Casado
Bio. Lic.	M	Parda	Solteiro	Bio. Lic.	F	Parda	Solteiro

Fonte: Alunos do 3º período do curso de Licenciatura em Ciências Biológicas da Universidade Federal do Acre – UFAC, 2014.

1. Digitamos os dados da variável *Curso* em uma coluna do *Excel*;
2. Construimos uma tabela de distribuição de frequência simples dos dados da variável;
3. Após a construção tabela de frequência, clicamos em **Suplementos, Action, Gráficos e Pizza**, conforme figura abaixo;



4. A janela **Gráfico de Pizza** será aberta. Clicamos na caixa **Conjunto de Dados** e selecionamos os dados da coluna *frequências* sem o Total;



tab. para exemplos de bioestatística NOVA - Microsoft

Início Inserir Layout da Página Fórmulas Dados Revisão Exibição Suplementos

Comandos de Menu

B6 DISTRIBUIÇÃO DE FREQUÊNCIAS

Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
Bio. Bach.	3	0,06	6	6
Bio. Lic.	19	0,38	38	44
Enfermagem	3	0,06	6	50
Eng. Agro.	1	0,02	2	52
Letras Inglês	6	0,12	12	64
Letras Port.	1	0,02	2	66
Matemática	8	0,16	16	82
Pedagogia	3	0,06	6	88
Português	6	0,12	12	100
Total	50			

Gráfico de Pizza

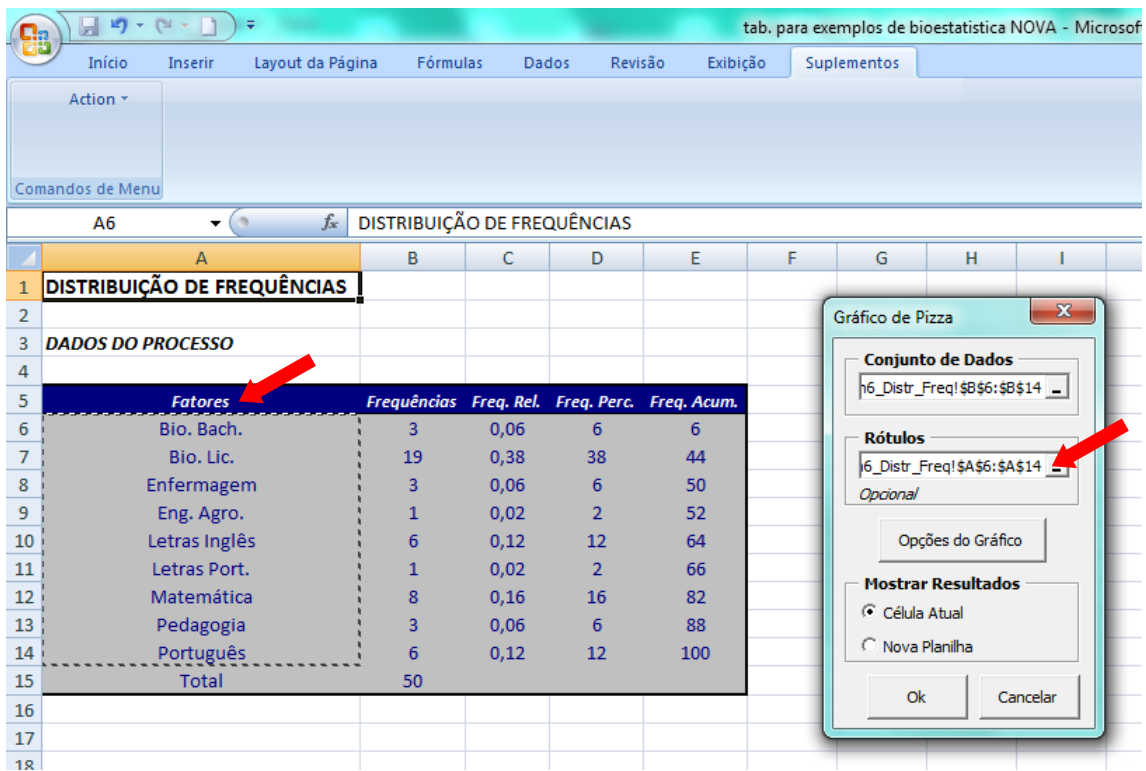
Conjunto de Dados
=D6:D14

Rótulos

Mostrar Resultados
 Célula Atual
 Nova Planilha

Ok Cancelar

5. Em seguida, clicamos na caixa **Rótulos** e selecionamos os dados da coluna *Fatores* sem o Total;



tab. para exemplos de bioestatística NOVA - Microsoft

Início Inserir Layout da Página Fórmulas Dados Revisão Exibição Suplementos

Comandos de Menu

A6 DISTRIBUIÇÃO DE FREQUÊNCIAS

Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
Bio. Bach.	3	0,06	6	6
Bio. Lic.	19	0,38	38	44
Enfermagem	3	0,06	6	50
Eng. Agro.	1	0,02	2	52
Letras Inglês	6	0,12	12	64
Letras Port.	1	0,02	2	66
Matemática	8	0,16	16	82
Pedagogia	3	0,06	6	88
Português	6	0,12	12	100
Total	50			

Gráfico de Pizza

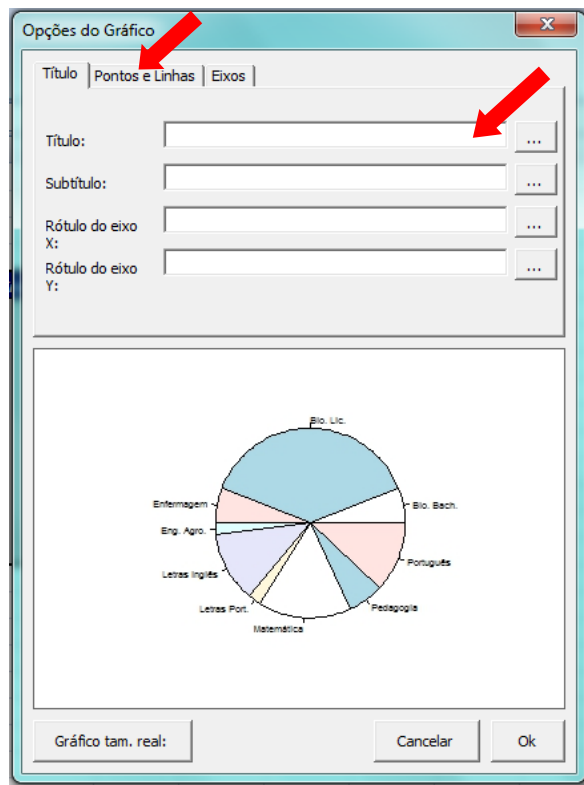
Conjunto de Dados
=D6:D14

Rótulos
=A6:A14

Mostrar Resultados
 Célula Atual
 Nova Planilha

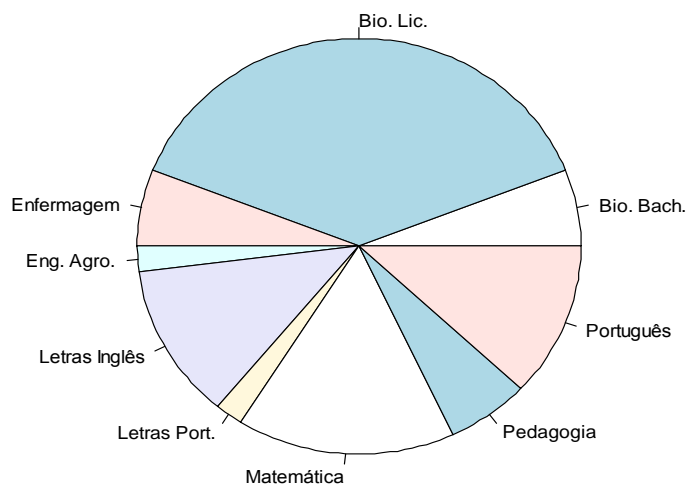
Ok Cancelar

6. Ao clicarmos em **Opções de Gráfico** da janela **Gráfico de Pizza** uma nova janela será aberta, onde podemos editar o gráfico em setores, colocando o título, os rótulos dos eixos e outras edições que acharmos convenientes, como por exemplo, alterar a cor, o tipo de fonte, o tamanho, etc.



7. Após a edição do gráfico podemos ainda visualiza-lo em tamanho real clicando em **Gráfico tam. real:** no canto inferior esquerdo. Após a visualização, fechamos a janela **Chart Preview**, clicamos em nova planilha e clicamos em Ok na janela de Opções de Gráfico.

Curso dos Alunos Entrevistados





Acima podemos visualizar o gráfico em setores gerado pelo *Action* para a variável *curso* da tabela 7. O gráfico nos mostra de maneira simples que a maioria dos alunos entrevistados foi dos cursos de Licenciatura em Ciências Biológicas, seguidos pelos alunos do curso de Matemática, Português e Letras Inglês. Cursos do turno noturno, indicando que a pesquisa foi realizada principalmente neste turno. O gráfico de pizza é um gráfico muito usado quando o pesquisador quer indicar um contraste visível entre os dados observados.

[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar gráfico em setores (pizza) utilizando o suplemento *Action* do *Excel*.

Exercício Proposto

1. Construa com o auxílio do Suplemento *Action*, um gráfico em setores para as variáveis: *sexo*, *cor/raça* e *estado civil* da tabela 7.
2. Após a análise dos gráficos descreva quais as principais características constates no conjunto de dados observados.

7.2 Gráficos para Variáveis Quantitativas

A depender do tipo de variável quantitativa (discreta ou contínua), podemos utilizar tipos de gráficos diferentes, sendo que alguns deles são mais recomendados para cada tipo de variável. Para variáveis discretas recomendamos o gráfico em barras ou colunas.

7.2.1 Gráfico em Barras ou Colunas

O gráfico em barras ou gráfico em colunas também pode ser usado para representar dados qualitativos, no entanto, em algumas situações também é utilizado para representar dados de variáveis quantitativas discretas, ou seja, quando a série é composta por números inteiros..

Para construirmos um gráfico em Barras para a variável “Idade” da tabela 8 abaixo, utilizando o *Action*, procedemos da seguinte forma:

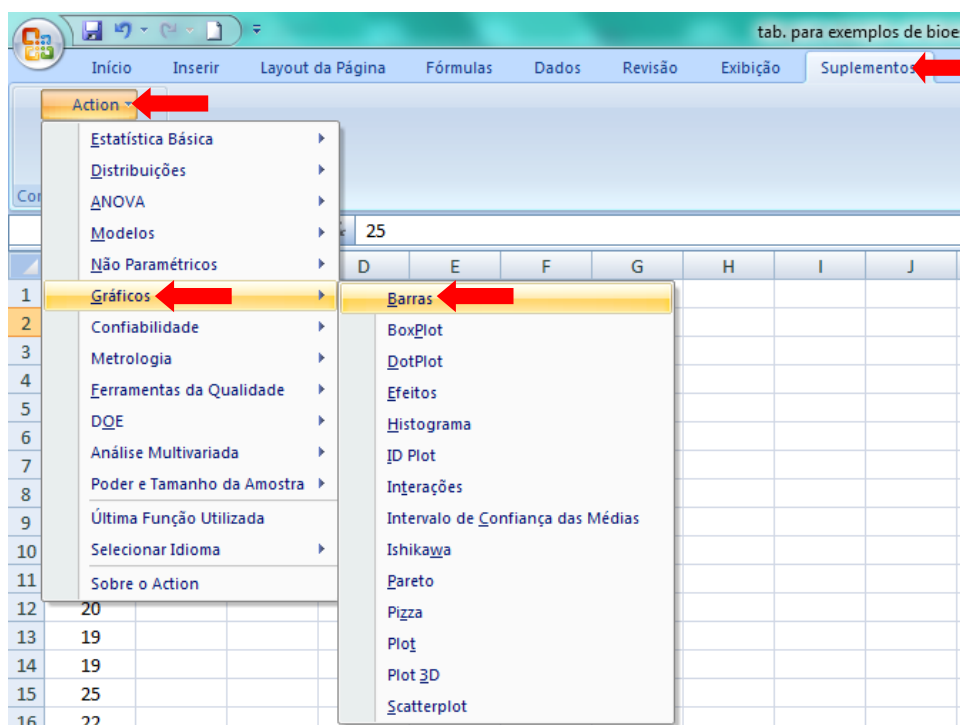
Tabela 8 – Pesquisa realizada com 50 alunos da Universidade Federal do Acre – UFAC – Campus Floresta, na cidade de Cruzeiro do Sul – Acre.

Idade (anos)	Nº de Filhos	Nº de Irmãos	Idade (anos)	Nº de Filhos	Nº de Irmãos
25	0	4	21	1	3
18	0	0	22	2	5
38	2	8	18	0	6

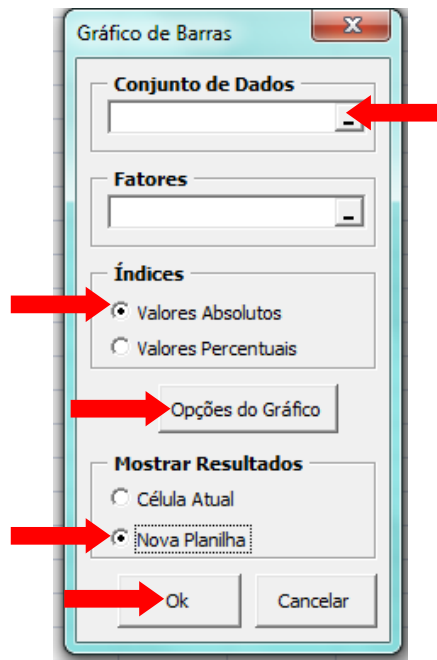
46	2	1	23	0	6
24	0	4	18	0	2
19	1	2	19	0	2
26	1	3	19	0	2
18	0	2	18	0	1
19	0	5	23	1	1
22	1	2	28	1	3
20	0	2	22	2	6
19	0	8	26	2	0
19	0	8	24	2	4
25	1	5	20	0	2
22	0	4	19	0	2
22	0	6	24	2	3
19	0	16	26	0	4
21	0	2	29	0	2
26	1	3	19	0	2
21	0	2	18	0	1
23	0	7	29	3	10
25	0	4	26	0	2
23	1	4	20	0	2
30	2	2	35	3	6
20	0	4	26	0	12

Fonte: Alunos do 3º período do curso de Licenciatura em Ciências Biológicas da Universidade Federal do Acre – UFAC, 2014.

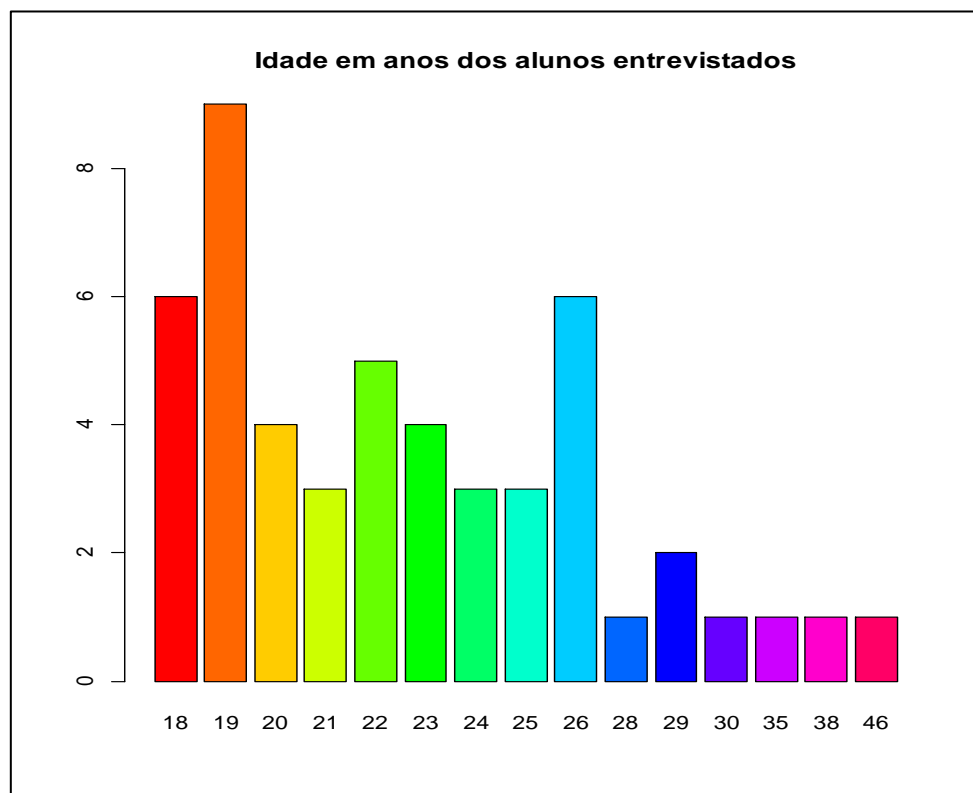
1. Digitamos os dados em uma coluna do Excel;
2. Clicamos em **Suplementos, Action, Gráficos e Barras**, conforme figura abaixo;



3. A janela *Gráfico de Barras* será aberta. Clicamos na caixa **Conjunto de Dados** e selecionamos os dados sobre a idade dos alunos, em seguida, selecionamos *Valores Absolutos* em **Índices**. Também podemos editar o gráfico clicando em *Opções de Gráfico* como visto no exemplo anterior. Em seguida, clicamos em *Nova Planilha* e Ok.



4. Com estes passos o *Action* gera o gráfico em barras da variável Idade dos dados observados.





[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar gráficos em barras utilizando do suplemento *Action* do *Excel*.

Através do gráfico em barras gerado pelo *Action* podemos realizar algumas análises, como por exemplo:

- a maioria dos alunos entrevistados tem 19 anos de idade;
- quase todos os alunos entrevistados tem idade entre 18 e 26 anos;
- houve a ocorrência de dados discrepantes, observados na presença de um aluno de 46 anos de idade.

Estas são algumas análises básicas que o pesquisador pode realizar apenas observando o gráfico acima.

Exercício Proposto

1. Construa com o auxílio do Suplemento *Action*, um gráfico em barras para as variáveis: *Números de Filhos e Números de Irmãos* da tabela 8.
2. Após a análise dos gráficos descreva quais as principais características constates no conjunto de dados observados.

7.2.2 Histograma

O histograma é um gráfico de barras verticais da tabela de distribuição de frequência de dados quantitativos contínuos. A base do histograma é o intervalo de classe (*IC*) e a altura é a densidade (*d*).

Para exemplificar, construiremos um histograma no *Excel*, utilizando o *software Action* dos dados coletados sobre o pH de lagos de meandro abandonado da região do Rio Purus – Acre.

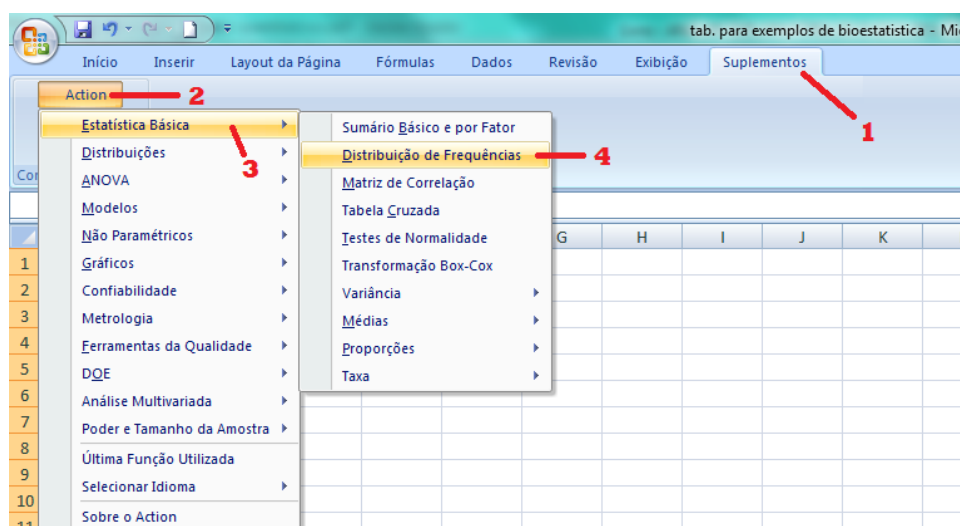
Tabela 9 – Índice de pH de lagos de meandro abandonado da região do Rio Purus – Acre.

6,21	6,48	6,50	6,77	7,34	6,40
6,10	6,48	6,50	6,66	7,31	6,90
6,80	6,68	6,50	6,66	7,24	7,64
6,90	6,70	6,50	6,70	7,60	7,60
6,70	7,10	6,50	6,50	7,60	6,81
5,90	6,11	6,58	6,87	8,23	7,74
6,06	5,99	6,63	7,25	8,12	7,92
5,80	6,99	6,60	6,60	8,10	7,76
6,20	6,89	6,70	6,83	8,20	7,76
6,33	6,71	6,93	6,50	7,80	7,76

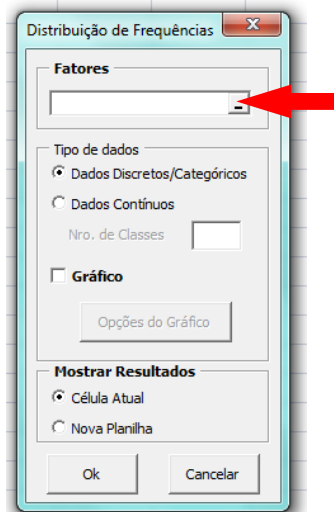
6,13	6,10	6,10	6,23	7,95	6,23
5,42	6,17	7,37	6,72	7,88	6,72
5,86	6,80	6,75	6,72	6,19	6,72
5,60	6,90	6,20	6,84	7,82	6,84
6,90	6,30	6,70	6,40	7,41	6,40
6,11	6,48	6,83	6,40	7,77	6,25
6,35	6,28	6,82	5,87	7,81	6,45
6,32	6,43	6,82	5,98	6,74	6,22
6,35	6,38	6,82	5,84	6,74	6,22
6,01	6,80	6,82	5,93	6,74	6,22

Fonte: Prof^o Lucena Rocha

1. Digitamos os dados na coluna A do Excel.
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.



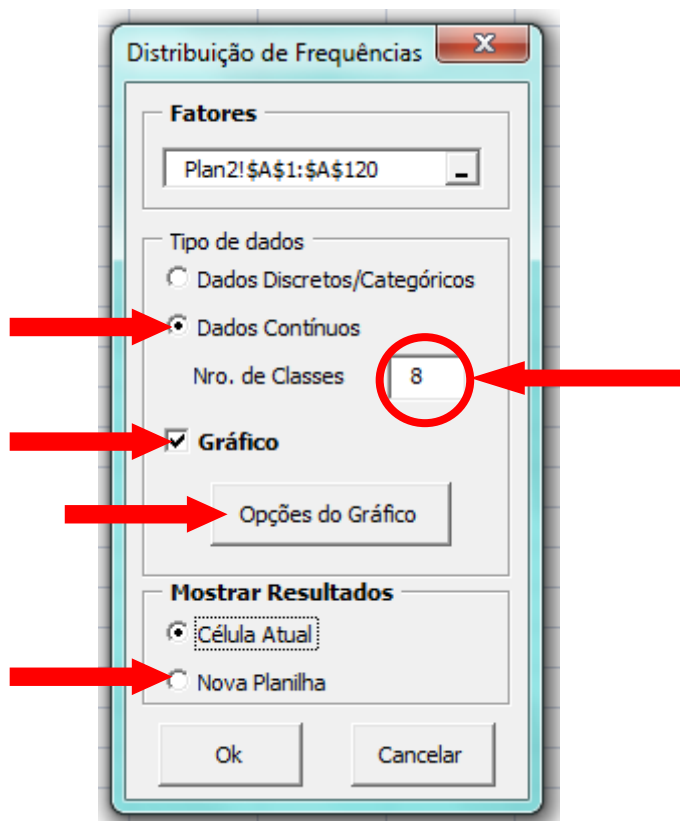
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique no canto direito da caixa de Fatores conforme indicado na figura abaixo, em seguida selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A120.



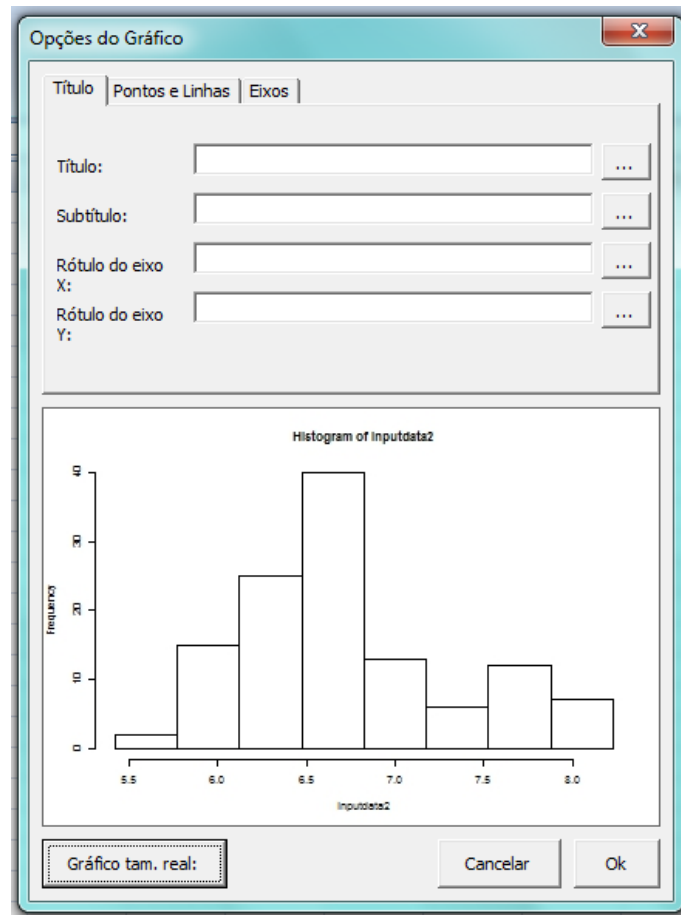
4. Clique novamente no canto direito da caixa de **Fatores** conforme indicado na figura abaixo.



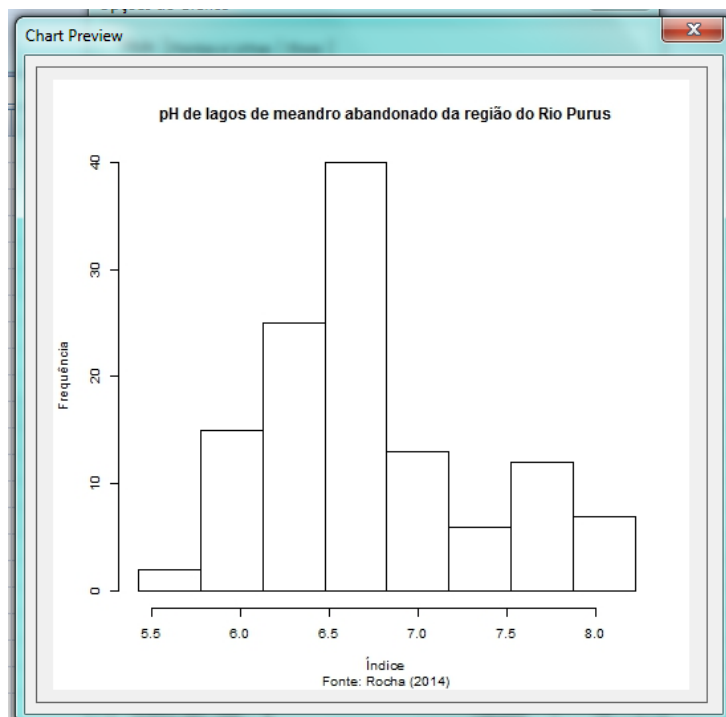
5. Clique em **Dados Contínuos** e escolha quantas classes terá a tabela. Neste exemplo, decidimos que nossa tabela terá oito linhas com intervalos de classe, então digitamos o número 8 na caixa **Nro. De Classe**, depois clicamos em **Gráfico** e em seguida em **Opções do Gráfico**.



6. Ao clicarmos em **Opções de Gráfico**, uma nova janela será aberta, onde podemos editar o histograma, colocando o título, os rótulos dos eixos e outras edições que acharmos convenientes, como por exemplo, alterar a cor das barras do gráfico, o tipo de fonte, o tamanho, etc.



7. Após a edição do histograma ainda podemos visualiza-lo em tamanho real clicando na caixa **Gráfico tam. real.**:





8. Agora é só fechar a janela *Chart Preview*, clicar em **Ok** na janela Opções do gráfico, clicarmos em **Nova Planilha** na janela de Distribuição de Frequência e Ok.

Além do histograma o *Action* gera a tabela de distribuição de frequência em uma planilha conjunta para facilitar a visualização dos resultados.

[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar histogramas utilizando o *Action*.

Exercício Proposto

1. Construa com o auxílio do Suplemento *Action*, um histograma para a variável altura da tabela abaixo.

Altura em metros de 50 alunos da UFAC – Campus Floresta – Cruzeiro do Sul – Acre, 2014.				
1,78	1,65	1,58	1,58	1,59
1,7	1,53	1,69	1,8	1,69
1,56	1,62	1,64	1,73	1,76
1,65	1,63	1,82	1,7	1,77
1,59	1,7	1,69	1,58	1,67
1,75	1,65	1,72	1,55	1,6
1,6	1,67	1,54	1,58	1,73
1,64	1,61	1,78	1,67	1,83
1,59	1,7	1,6	1,7	1,68
1,69	1,55	1,63	1,77	1,58

Fonte: Alunos do 3º período do curso de Licenciatura em Ciências Biológicas da Universidade Federal do Acre – UFAC, 2014.

2. Após a análise do gráfico, descreva quais as principais características constates no conjunto de dados observados.



8. MEDIDAS ESTATÍSTICAS

Após a organização dos dados com a construção de tabelas e gráficos, a análise exploratória de dados também utiliza os cálculos de medidas estatísticas, que resumem as informações coletadas, nos dando uma visão mais ampla dos dados. Essas medidas, que também são conhecidas como medidas descritivas, recebem o nome de *estatísticas* quando calculada com os dados da amostra, e de *parâmetros* quando calculadas com dados populacionais.

Dentre as medidas estatísticas as mais utilizadas são as *medidas de tendência central* (ou de *posição*) e as *medidas de dispersão* (ou de *variabilidade*).

8.1 Medidas de Tendência Central

As medidas de tendência central são valores calculados para representar um conjunto de dados de uma forma ainda mais resumida que as tabelas e os gráficos. Quando um pesquisador necessita representar por meio de um número um conjunto de informações que variam, então ele precisa escolher um valor central, ou seja, um número em torno do qual os dados observados se distribuam, esse número representará todo o conjunto de informações. As medidas de tendência central são: média aritmética, mediana e moda.

8.1.1 Média Aritmética

A média aritmética é uma das medidas estatísticas mais utilizadas na representação de um conjunto de dados. Para calcularmos a média aritmética de um conjunto de valores, somamos os valores da série de dados e dividimos pelo número de observações.

$$\text{média} = \frac{\text{Soma de todos os dados}}{\text{número de observações}}$$

Símbolos matemáticos

Para representarmos esta fórmula usamos os seguintes símbolos matemáticos:

Média da população: μ (letra grega *mi* minúsculo)

Média da amostra: \bar{x} (lê-se xis barra)

Soma dos dados: $\sum x$ (lê-se somatório de x)

Número de observações: n



Em termos dos símbolos matemáticos podemos então escrever a média da seguinte forma:

$$\text{Média da população: } \mu = \frac{\sum x}{n}$$

$$\text{Média da amostra: } \bar{x} = \frac{\sum x}{n}$$

No exemplo a seguir, observamos a situação em que um biólogo coletou dados sobre a cobertura total de vegetação aquática no lago de Itapira em Rio Branco - AC, obtendo os seguintes resultados:

4,2 3,1 4,2 6 8,6

Para calcular a média amostral dos dados coletados procedemos da seguinte forma:

Chamamos os dados coletados de x e a posição que ele ocupar de i , que ficará subscrito ao lado de x , logo representamos por x_i .

$$x_1 = 4,2 ; x_2 = 3,1 ; x_3 = 4,2 ; x_4 = 6 ; x_5 = 8,6$$

Como foram coletadas cinco amostras, $n = 5$.

O cálculo da média será:

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} \\ \bar{x} &= \frac{4,2 + 3,1 + 4,2 + 6 + 8,6}{5} \\ \bar{x} &= 5,22\end{aligned}$$

A média aritmética é uma das medidas estatísticas mais simples de ser calculada e interpretada, no entanto, em alguns casos, esta medida não é a mais indicada para representar um conjunto de dados, por sofrer influência de valores extremos.

Por exemplo, imagine que a tabela abaixo representa a idade de 5 alunos de uma turma.

19	17	17	20	18
----	----	----	----	----

A média de idade dos alunos desta turma seria $\bar{x} = 18,2$ anos.

Se após um determinado tempo fosse incluído nesta turma um aluno com 60 anos de idade, a média de idade dos alunos desta turma passaria a ser $\bar{x} = 25,2$ anos, um número que não representa a maioria das idades dos alunos desta turma. Por tanto, a média aritmética é mais indicada quando temos um conjunto de valores aproximadamente simétrico, ou seja, sem valores extremos.

Uma medida estatística que não sofre influência dos valores extremos é a mediana.



8.1.2 Mediana

A mediana (Me) é o valor que ocupa a posição central em um conjunto de dados ordenados, dividindo a série em duas partes. Quando os dados apresentam uma quantidade par de valores, há dois valores centrais, nestes casos, o cálculo da mediana é feito encontrando-se a média dos dois valores centrais do conjunto de dados ordenados.

Com os dados do exemplo anterior, observe como poderíamos calcular a mediana.

17	17	18	19	20
----	----	-----------	----	----

Como nesta série temos um número ímpar de valores a mediana é o valor central.

$$Me = 18$$

Se usarmos a mesma situação do exemplo anterior e incluirmos um aluno com 60 anos de idade os dados seriam:

17	17	18	19	20	60
----	----	-----------	-----------	----	----

Nesta nova série, o número de valores é par, então, procedemos da seguinte forma:

$$Me = \frac{18 + 19}{2} = \mathbf{18,5}$$

Perceba que o valor extremo não influenciou o valor da mediana de forma significativa. Dessa forma, a mediana é mais indicada quando temos um conjunto de dados que possuem valores extremos, enquanto que a média é um parâmetro indicado quando o conjunto de dados é simétrico.

8.1.3 Moda

A moda (Mo) é o valor que ocorre com mais frequência em um conjunto de dados. Se todos os valores forem diferentes não há moda. Um conjunto de dados pode ser bimodal, quando tem duas modas, trimodal, quando tem três modas ou multimodal quando tem mais de três modas.

O valor da moda é encontrado apenas pela observação da série de dados.

Veja no conjunto de dados abaixo como encontrar a moda.

0	6	4	7	3
5	3	5	1	2
5	2	3	9	7
6	6	9	8	9

$$Mo = 5$$



8.2 Medidas de Dispersão

As medidas de dispersão ou medidas de variabilidade são usadas para descrever a variação de um conjunto de dados entre si. Se todos os valores de um conjunto de dados forem iguais, então não existe dispersão, mas se os dados não são iguais, existe uma dispersão entre eles. O motivo pelo qual um pesquisador precisa calcular essas medidas, é que, ao escolher uma medida de tendência central para representar um conjunto de dados, esta, não informa se há ou não variabilidade entre os dados.

Para exemplificar, observe a tabela abaixo sobre a transparência da água em três locais da região do Vale do Juruá - AC.

Tabela 10 – Dados sobre a transparência da água em três locais da região do vale do Juruá - AC.

Repetição	Igarapé Preto (A)	Anurí (B)	Cameta (C)
1	45	41	25
2	45	42	30
3	45	43	35
4	45	44	40
5	45	45	45
6	45	46	50
7	45	47	55
8	45	48	60
9	45	49	65
Média (\bar{x})	45	45	45
Mediana (Me)	45	45	45

Fonte: Dados fictícios

Observe que os dados do local “A” não apresentam dispersão, já os valores do local “B” apresentam certa dispersão em torno da média $\bar{x} = 45$, e os valores do local “C” apresentam uma dispersão em torno da média maior do que a do local B. No entanto, se o pesquisador observa-se os dados apenas através da média ou da mediana, não poderia dizer se os dados variaram ou não, pois nos três locais tanto a média quanto a mediana são iguais.

As medidas descritivas mais comuns para quantificar a dispersão são: *amplitude*, *variância*, *desvio-padrão* e *coeficiente de variação*.

8.2.1 Amplitude

A medida de dispersão mais simples é a *amplitude de variação* (a) que é a diferença entre o maior e o menor valor em um conjunto de dados amostrais.

$$a = \text{maior valor} - \text{menor valor}$$

Para calcularmos a amplitude de variação dos dados da tabela 10, procedemos da seguinte forma:

$$\text{Local A: } a = 45 - 45 \quad a = 0$$

$$\text{Local B: } a = 49 - 41 \quad a = 8$$

$$\text{Local C: } a = 65 - 25 \quad a = 40$$

Uma das vantagens de se usar a amplitude como medida de dispersão é a sua simplicidade, no entanto, é uma medida muito limitada, por depender somente dos valores extremos.

8.2.2 Variância

Para entendermos a variância temos que consideramos que os dados se distribuem em torno da média, logo o grau de dispersão de um conjunto de dados pode ser medido pelo afastamento dos mesmos em relação à média. Para medirmos esta dispersão usamos então a variância que é definida como a soma dos quadrados dos desvios dividida pelo tamanho da amostra menos 1 ($n - 1$). Chamamos o valor de ($n - 1$) de *grau de liberdade*. Dessa forma, a variância de uma amostra que é indicada por s^2 , é dada pela fórmula:

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$$

A primeira vista esta fórmula pode parecer difícil, mas na verdade ela facilita o trabalho se realizarmos os cálculos intermediários em uma tabela. Vejamos o exemplo a seguir:

Imagine que um biólogo precisa verificar a variação dos dados sobre a riqueza de peixes no lago Bom Lugar na região do Rio Purus – Ac. Os dados estão apresentados na tabela abaixo.

Repetição	Riqueza (x)
1	5
2	3
3	2
4	1
5	7

Cálculos intermediários para a obtenção da variância.



x	x^2
5	25
3	9
2	4
1	1
7	49
$\sum x = 18$	$\sum x^2 = 88$

Substituindo os valores de $\sum x^2$ e $\sum x$ na fórmula da variância temos:

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$$

$$s^2 = \frac{88 - \frac{(18)^2}{5}}{5 - 1}$$

$$s^2 = 5,8$$

A variância apresenta o resultado da variação dos dados elevada ao quadrado, por isso, sua interpretação se torna um pouco difícil. Para resolver esse problema, precisamos de outra medida de variabilidade chamada desvio padrão.

8.2.3 Desvio Padrão

O desvio padrão (s) é uma medida de variabilidade que apresenta o valor da variação na mesma unidade de medida do conjunto de dados amostrais. Para obtermos o valor do desvio padrão basta calcularmos a raiz quadrada da variância.

$$s = \sqrt{VAR}$$

Para a variação da riqueza de peixes no lago Bom Lugar na região do Rio Purus – AC, cuja variância já foi calculada, tem-se o desvio padrão:

$$s = \sqrt{5,8}$$

$$s = 2,4$$



8.2.4 Coeficiente de Variação

O coeficiente de variação é uma medida relativa de variabilidade, utilizada para comparar o grau de concentração dos dados em torno da média. É calculado dividindo-se o desvio padrão pela média do conjunto de dados. Seu resultado pode ser expresso em termos decimais ou percentuais.

$$CV = \frac{s}{\bar{x}}$$

Quanto menor for o valor do coeficiente de variação mais homogênea será considerada a série e quanto maior for o coeficiente de variação mais heterogeneia será considerada a série.

Se precisarmos calcular o coeficiente de variação dos dados sobre a transparência da água nos três locais da região do vale do Juruá apresentados na tabela 8, procedemos da seguinte forma:

$$\text{Local A: } CV = \frac{0}{45} = 0 \quad \text{ou } 0\%$$

$$\text{Local B: } CV = \frac{2,74}{45} = 0,06 \quad \text{ou } 6\%$$

$$\text{Local C: } CV = \frac{13,69}{45} = 0,30 \quad \text{ou } 30\%$$

8.3 Cálculo de medidas estatísticas utilizando o *software Action*.

O *Software Action* possibilita que o cálculo das principais medidas de tendência central e de dispersão seja realizado de maneira simples e rápida, com apenas alguns cliques. Veja a seguir, os procedimentos para a construção da tabela “Sumário básico e por fator”, que gera o valor das medidas estatísticas desejadas.

8.3.1 Análise sem fator

Quando nos referimos a análise sem fator, estamos dizendo que as medidas estatísticas serão apresentadas para uma única variável, não havendo um fator de comparação entre elas.

Suponha que um pesquisador precisa calcular a média, a variância e o desvio padrão dos dados climáticos de agosto a dezembro de 2010, coletados na captura de abelhas na região de Santa Luzia município de Cruzeiro do Sul – AC.

Tabela 11 – Dados de precipitação em *mm*, coletados entre os meses de agosto a dezembro de 2010, na região de Cruzeiro do Sul – AC.

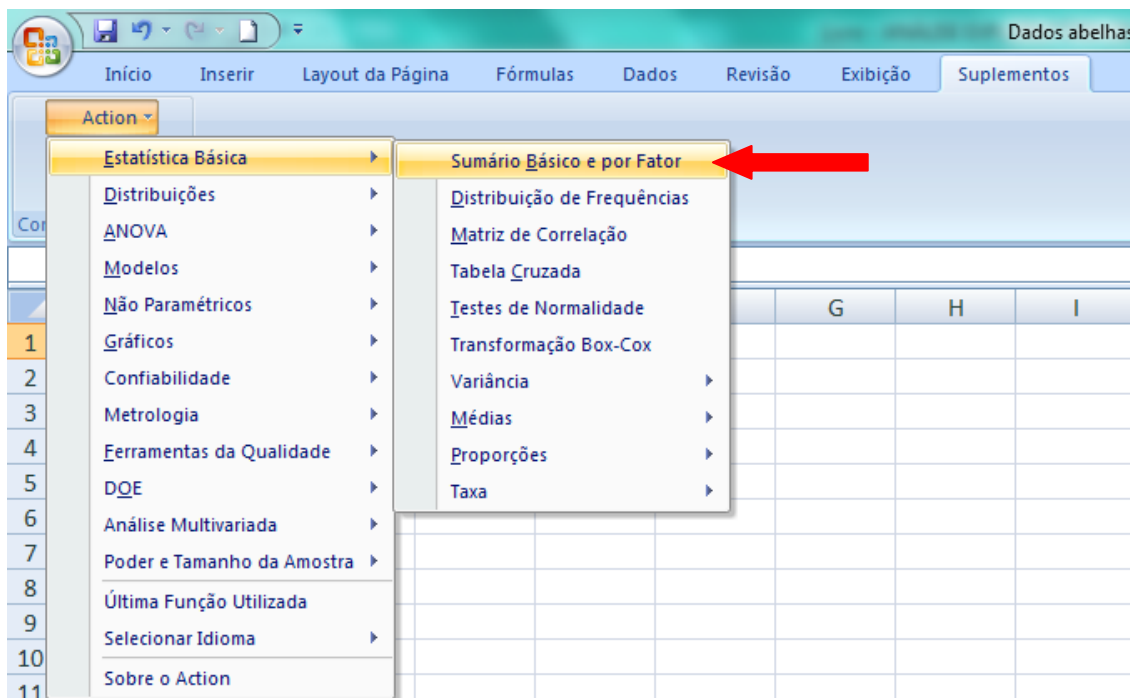
8,64	4,67	4,67	0,76	0,76	8,57	8,57	0,76	8,64	0,57
8,64	4,67	4,67	0,76	0,57	8,57	8,57	0,76	8,64	0,57
8,64	4,67	4,67	0,76	0,57	8,57	8,57	0,76	8,57	0,57
8,64	4,67	4,67	0,76	8,57	8,57	8,57	0,76	0,76	0,57
4,67	4,67	4,67	0,76	0,57	8,57	8,57	0,76	0,76	0,57
4,67	4,67	4,67	0,76	0,57	8,57	8,57	8,64	0,76	0,57
4,67	4,67	4,67	0,76	0,57	8,57	8,57	8,64	0,57	0,57
4,67	4,67	4,67	0,76	0,57	8,57	8,57	8,64	4,67	0,57
4,67	4,67	4,67	0,76	0,57	8,57	8,57	8,64	4,67	0,57
4,67	4,67	4,67	0,76	8,57	8,57	0,76	8,64	0,57	0,57

Fonte: Prof^o Rogério Oliveira Souza

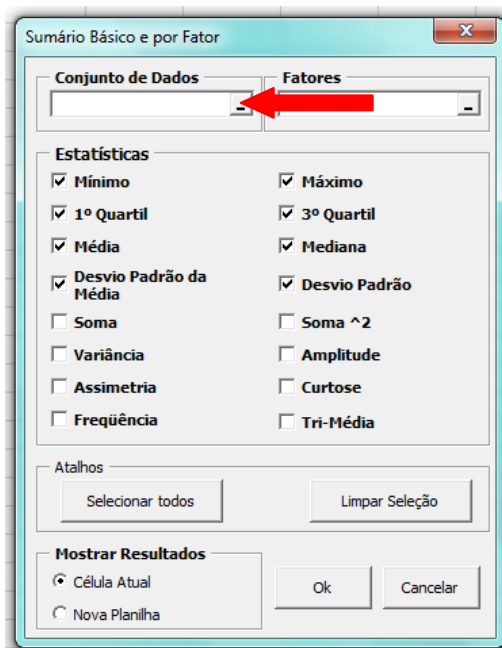
Para realizarmos os cálculos das medidas estatísticas de apenas uma variável utilizando o *software Action*, procedemos da seguinte forma:

1. Digitamos os dados na coluna A do Excel.

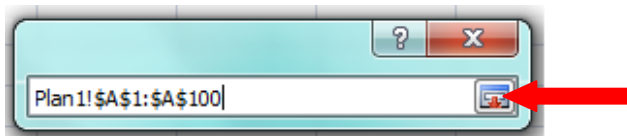
2. Clique em **Suplementos** na barra de ferramentas, em seguida em **Action**, **Estatística Básica** e **Sumário básico e por fator**.



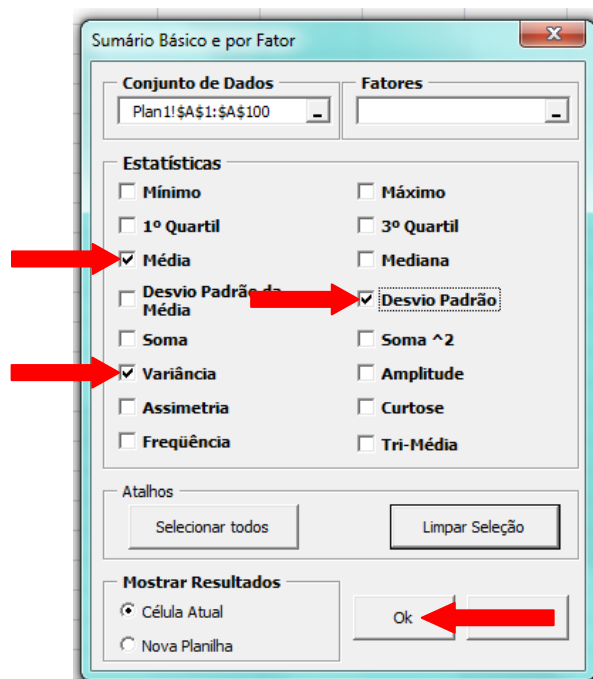
3. A janela de Sumário Básico e por Fator do *Action* será aberta. Clique no canto direito da caixa de **Conjunto de Dados**, conforme indicado na figura abaixo. Em seguida, selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A100.



4. Clique novamente no canto direito da caixa “Conjunto de Dados” conforme indicado na figura abaixo.



5. Clique em **limpar seleção** e em seguida selecione as medidas estatísticas que deseja calcular, no nosso exemplo selecionaremos a **média**, a **variância** e o **desvio padrão**, em seguida, clique em **Ok**.





6. O *Action* gera uma tabela com os valores das medidas estatísticas selecionadas, conforme figura abaixo.

	A	B	C
1	SUMÁRIO BÁSICO E POR FATOR		
2			
3	DADOS DO PROCESSO		
4			
5	Média	Desvio Padrão	Variância
6	4,405007	3,371748	11,368686

[Clique aqui](#) e assista ao vídeo tutorial sobre como calcular medidas estatísticas sem fatores utilizando o *Action*.

8.3.2 Análise com um único fator e diversos níveis

Também podemos construir um sumário básico com as medidas estatísticas, e comparar essas medidas com outra variável utilizando a ferramenta “fatores”, disponível no *software Action*.

Vejam os a seguir um exemplo da aplicação de fatores na análise das medidas estatísticas.

Um biólogo precisa calcular a média, a variância e o desvio padrão dos dados sobre a temperatura de lagos de meandros abandonados na região do Rio Purus – Ac. Para isso dispõe dos dados da tabela abaixo.

Tabela 12 – Dados sobre a temperatura em graus Celsius (°C) da água em lagos de meandros abandonados da região do Rio Purus – Ac.

Lago	Temp.	Lago	Temp.	Lago	Temp.	Lago	Temp.
Itapira	26,12	Flor_ouro	26,41	Lago_Verde	31	Santana	36,5
Itapira	30,5	Flor_ouro	29,72	Lago_Verde	29,66	Santana	31,5
Itapira	29,2	Flor_ouro	30,1	Lago_Verde	30	Santana	27,8
Itapira	28,9	Flor_ouro	28,8	Lago_Verde	28,5	Santana	24
Itapira	25,34	Flor_ouro	27,7	Lago_Verde	28,8	Santana	29
Cameta	28,6	Anuri	26,8	Ig_Preto	26,91	Bom_Lugar	33,4
Cameta	29	Anuri	25,9	Ig_Preto	27,74	Bom_Lugar	35,1
Cameta	28,35	Anuri	25,11	Ig_Preto	26,82	Bom_Lugar	29,6
Cameta	27	Anuri	25,11	Ig_Preto	27,7	Bom_Lugar	29,6
Cameta	27,89	Anuri	25,11	Ig_Preto	29,9	Bom_Lugar	29,6
Santana	31	Itapira	30	Flor_ouro	28,5	Lago_Verde	32,2
Santana	39,8	Itapira	30	Flor_ouro	26,29	Lago_Verde	28,5
Santana	28	Itapira	30	Flor_ouro	28,9	Lago_Verde	34
Santana	29,22	Itapira	30	Flor_ouro	26,36	Lago_Verde	34,4

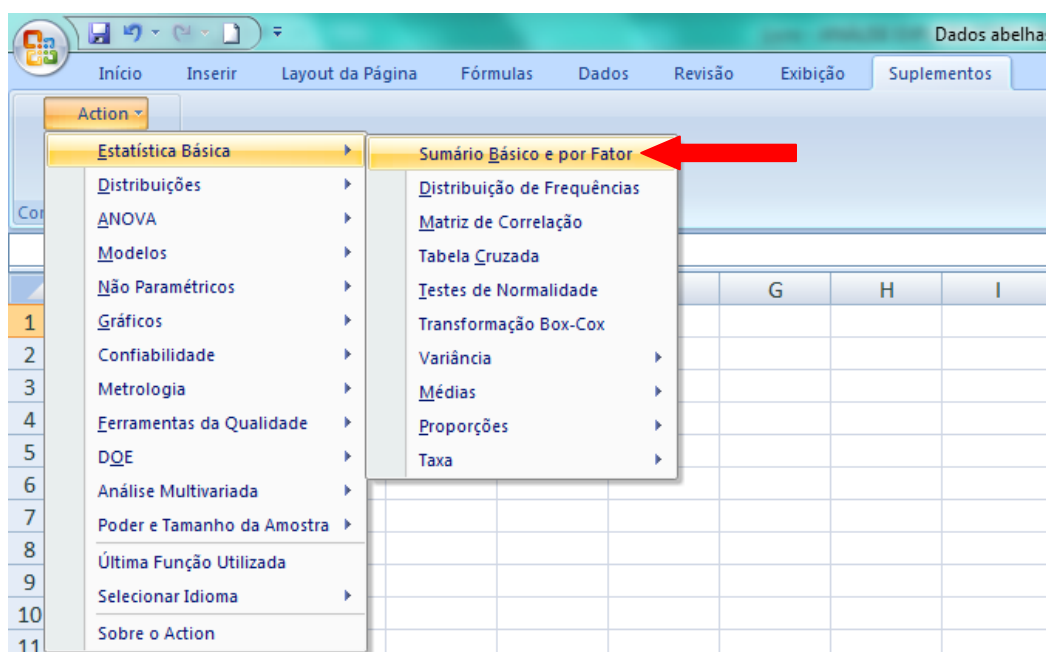
Santana	30,6	Itapira	30	Flor_ouro	29,72	Lago_Verde	26,4
Bom_Lugar	28,7	Cameta	28,9	Anuri	26,6	Ig_Preto	31,4
Bom_Lugar	27	Cameta	34	Anuri	25,11	Ig_Preto	36,9
Bom_Lugar	26,66	Cameta	32	Anuri	27	Ig_Preto	33,1
Bom_Lugar	32	Cameta	31	Anuri	24,9	Ig_Preto	33,1
Bom_Lugar	33,02	Cameta	30	Anuri	5,31	Ig_Preto	33,1
Lago_Verde	25	Santana	31,9	Itapira	30,01	Flor_ouro	26,41
Lago_Verde	25	Santana	31	Itapira	32	Flor_ouro	29,72
Lago_Verde	25,9	Santana	31,5	Itapira	30,8	Flor_ouro	30,1
Lago_Verde	25,8	Santana	32,1	Itapira	35	Flor_ouro	28,8
Lago_Verde	25,6	Santana	29,33	Itapira	35	Flor_ouro	28,8
Ig_Preto	28,9	Bom_Lugar	29,5	Cameta	31,1	Anuri	29
Ig_Preto	24,4	Bom_Lugar	29,4	Cameta	29,1	Anuri	27
Ig_Preto	26,9	Bom_Lugar	28,5	Cameta	30,2	Anuri	30
Ig_Preto	26,9	Bom_Lugar	28,5	Cameta	37,4	Anuri	30
Ig_Preto	26,9	Bom_Lugar	28,5	Cameta	33,5	Anuri	30

Fonte: Prof^ª. Lucena Rocha

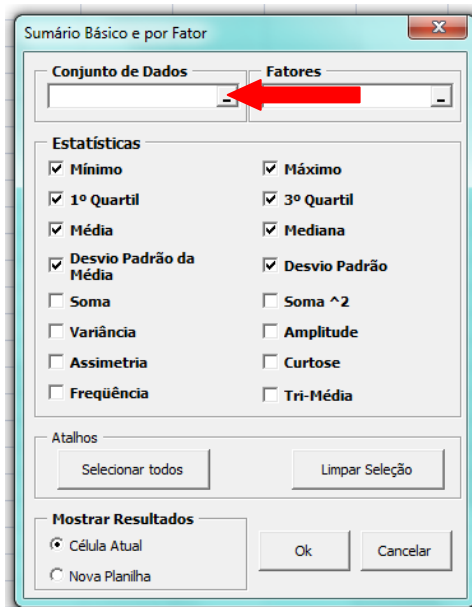
Para comparar a temperatura média e a variação da série de dados entre os lagos, utilizando o *Action*, procedemos do seguinte modo:

1. Digitamos os dados da variável lago na coluna A e os dados da variável temperatura na coluna B do Excel.

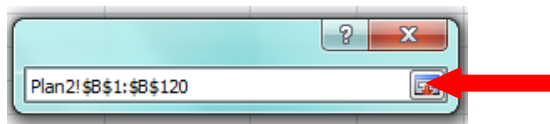
2. Clique em **Suplementos** na barra de ferramentas, em seguida em **Action**, **Estatística Básica** e **Sumário básico e por fator**.



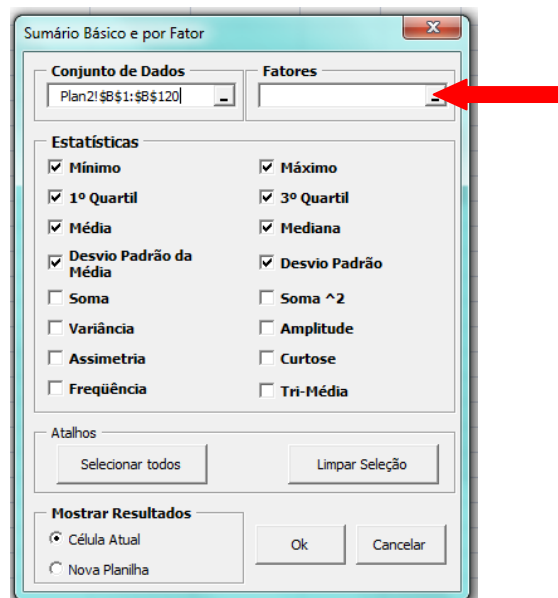
3. A janela de Sumário Básico e por Fator do *Action* será aberta. Clique no canto direito da caixa **Conjunto de Dados**, conforme indicado na figura abaixo. Em seguida, selecione os dados da coluna B, clicando na célula B1 e arrastando o ponteiro do mouse até a célula B120.



4. Clique novamente no canto direito da caixa **Conjunto de Dados**, conforme indicado na figura abaixo.



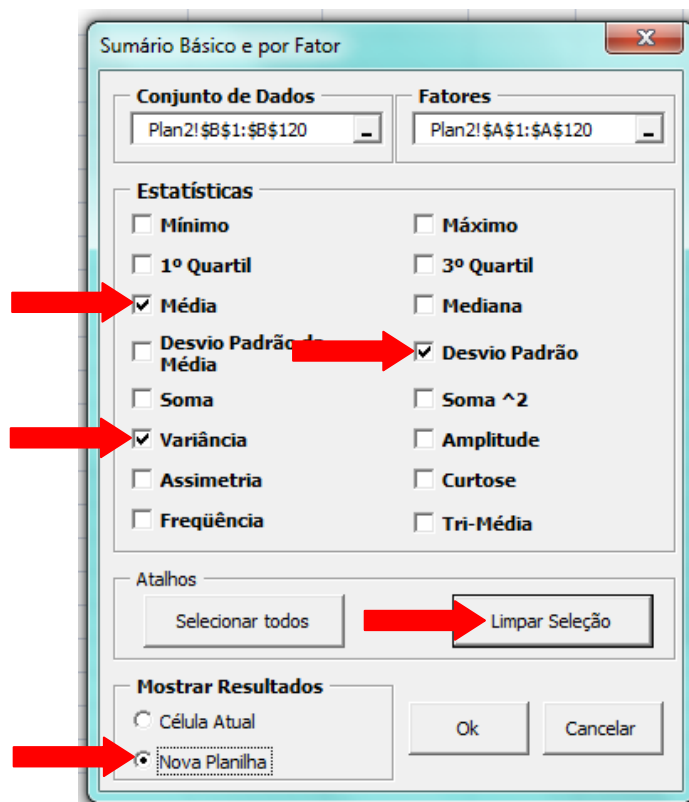
5. Clique no canto direito da caixa **Fatores**, conforme indicado na figura abaixo. Em seguida, selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A120.



6. Clique novamente no canto direito da caixa **Conjunto de Dados**, conforme indicado na figura abaixo.



7. Clique em **limpar seleção** e em seguida selecione as medidas estatísticas que deseja calcular, no nosso exemplo selecionaremos a **média**, a **variância** e o **desvio padrão**. Depois clique em **nova planilha** em seguida clique em **Ok**.



8. O *Action* gera uma tabela com os valores das medidas estatísticas para a variável Temperatura, fazendo a comparação entre os locais apresentados na série de dados, conforme figura abaixo.

	A	B	C	D
1	SUMÁRIO BÁSICO E POR FATOR			
2				
3	DADOS DO PROCESSO			
4				
5	Fator	Média	Desvio Padrão	Variância
6	Anuri	25,53	5,910952	34,939357
7	Bom_Lugar	29,938667	2,395227	5,737112
8	Cameta	30,536	2,752987	7,57894
9	Flor_ouro	28,422	1,433374	2,05456
10	Ig_Preto	29,378	3,42329	11,718917
11	Itapira	30,191333	2,576833	6,64007
12	Lago_Verde	28,717333	3,154895	9,953364
13	Santana	30,883333	3,668039	13,45451



Neste exemplo, os níveis são os locais (Anuri, Bom Lugar, Cameta, etc) de coleta da série de dados e o fator de comparação das medidas estatísticas são os lagos.

[Clique aqui](#) e assista ao vídeo tutorial sobre como calcular medidas estatísticas com fatores utilizando o *Action*.

Exercício Proposto

1. Imagine que um pesquisador queira saber qual a média de nitrogênio e fósforo da água de lagos de meandros abandonados da região do Rio Purus – Ac. Após a coleta, os dados foram organizados na tabela abaixo.

Lago	NT	PT	Lago	NT	PT	Lago	NT	PT
Itapira	1,28	0,381	Itapira	1,99	0,432	Itapira	2,832	0,23
Itapira	1,58	0,481	Itapira	1,99	0,432	Itapira	2,832	0,23
Itapira	1,58	0,668	Itapira	1,99	0,432	Itapira	2,832	0,23
Itapira	1,48	0,411	Itapira	1,99	0,432	Itapira	2,832	0,23
Itapira	2,58	0,281	Itapira	1,99	0,432	Itapira	2,832	0,23
Cameta	1,28	0,381	Cameta	1,401	0,205	Cameta	2,278	0,217
Cameta	1,58	0,481	Cameta	1,401	0,205	Cameta	2,278	0,217
Cameta	1,58	0,668	Cameta	1,401	0,205	Cameta	2,278	0,217
Cameta	1,48	0,411	Cameta	1,401	0,205	Cameta	2,278	0,217
Cameta	2,58	0,281	Cameta	1,401	0,205	Cameta	2,278	0,217
Santana	1,47	0,336	Santana	1,586	0,24	Santana	3,075	0,221
Santana	1,37	0,396	Santana	1,586	0,24	Santana	3,075	0,221
Santana	2,17	0,236	Santana	1,586	0,24	Santana	3,075	0,221
Santana	1,55	0,206	Santana	1,586	0,24	Santana	3,075	0,221
Santana	1,343	0,296	Santana	1,586	0,24	Santana	3,075	0,221
Bom_Lugar	1,47	0,336	Bom_Lugar	2,105	0,439	Bom_Lugar	2,896	0,16
Bom_Lugar	1,37	0,396	Bom_Lugar	2,105	0,439	Bom_Lugar	2,896	0,16
Bom_Lugar	2,17	0,236	Bom_Lugar	2,105	0,439	Bom_Lugar	2,896	0,16
Bom_Lugar	1,55	0,206	Bom_Lugar	2,105	0,439	Bom_Lugar	2,896	0,16
Bom_Lugar	1,343	0,296	Bom_Lugar	2,105	0,439	Bom_Lugar	2,896	0,16
Lago_Verde	1,482	0,211	Lago_Verde	1,257	0,269	Lago_Verde	1,257	0,269
Lago_Verde	1,482	0,211	Lago_Verde	1,257	0,269	Lago_Verde	2,382	0,196
Lago_Verde	1,482	0,211	Lago_Verde	1,257	0,269	Lago_Verde	2,382	0,196
Lago_Verde	1,482	0,211	Lago_Verde	1,257	0,269	Lago_Verde	2,382	0,196
Lago_Verde	1,482	0,211	Lago_Verde	1,257	0,269	Lago_Verde	2,382	0,196
Ig_Preto	1,63	0,254	Ig_Preto	1,655	0,201	Ig_Preto	3,09	0,671
Ig_Preto	1,13	0,274	Ig_Preto	1,655	0,201	Ig_Preto	2,89	1,871
Ig_Preto	1,73	0,234	Ig_Preto	1,655	0,201	Ig_Preto	2,99	3,251
Ig_Preto	1,83	0,214	Ig_Preto	1,655	0,201	Ig_Preto	3,31	1,251
Ig_Preto	1,93	0,294	Ig_Preto	1,655	0,201	Ig_Preto	2,34	0,251
Flor_ouro	1,482	0,211	Flor_ouro	1,551	0,197	Flor_ouro	3,202	0,249
Flor_ouro	1,482	0,211	Flor_ouro	1,551	0,197	Flor_ouro	3,202	0,249



Flor_ouro	1,482	0,211	Flor_ouro	1,551	0,197	Flor_ouro	3,202	0,249
Flor_ouro	1,482	0,211	Flor_ouro	1,551	0,197	Flor_ouro	3,202	0,249
Flor_ouro	1,482	0,211	Flor_ouro	1,551	0,197	Flor_ouro	3,202	0,249
Anuri	5,262	2,241	Anuri	1,984	0,347	Anuri	2,978	1,196
Anuri	5,862	1,641	Anuri	1,984	0,347	Anuri	2,278	1,106
Anuri	4,262	1,841	Anuri	1,984	0,347	Anuri	1,998	1,156
Anuri	4,992	1,241	Anuri	1,984	0,347	Anuri	2,178	1,116
Anuri	5,162	1,841	Anuri	1,984	0,347	Anuri	2,298	1,236

- a) Calcule a média a variância e o desvio padrão dos dados sobre nitrogênio e compare-os com os locais observados utilizando o *Action*.
- b) Calcule a média a variância e o desvio padrão dos dados sobre fósforo e compare-os com os locais observados utilizando o *Action*.
- c) A qual conclusão podemos chegar sobre os dados observados.

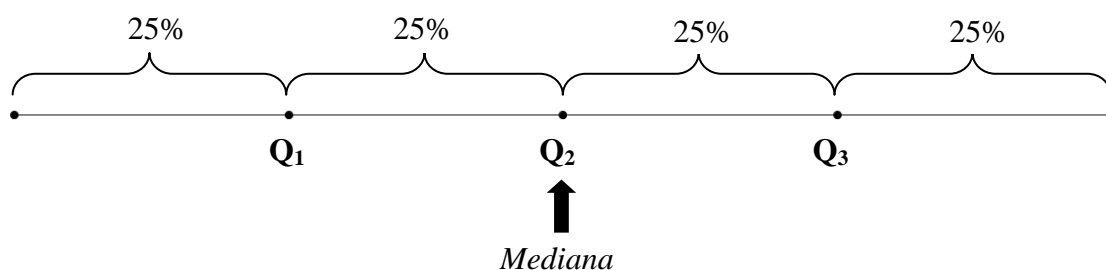
8.4 Quartil

Os quartis são medidas de dispersão estatística que dividem uma série de dados organizados em ordem crescente em quatro partes iguais. Dessa forma, existem três quartis: O primeiro quartil (Q_1), o segundo quartil (Q_2) que é a mediana, e o terceiro quartil (Q_3).

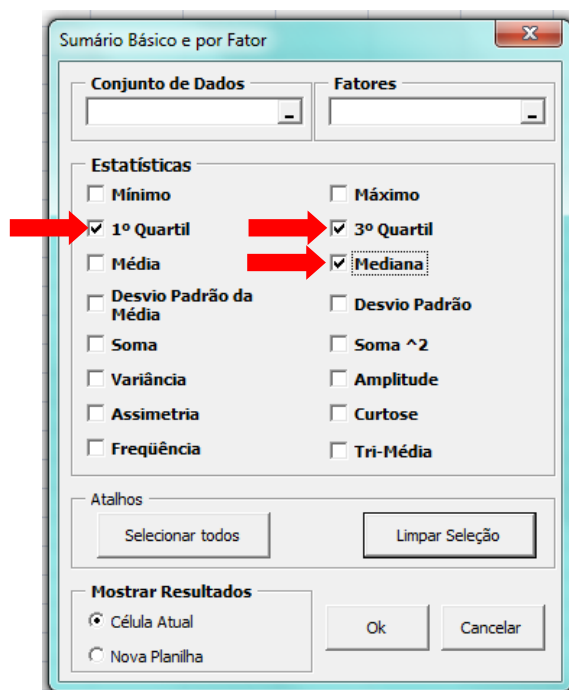
O primeiro quartil indica que 25% dos dados estão abaixo de seu valor e consequentemente 75% dos dados estão acima de seu valor.

O segundo quartil indica que 50% dos dados estão abaixo de seu valor e 50% estão acima.

O terceiro quartil indica que 75% dos dados estão abaixo de seu valor e 25% estão acima.



Para calcularmos os quartis utilizando o suplemento *Action* do *Excel* procedemos da mesma forma que calculamos as medidas estatísticas anteriores, incluindo, após a seleção dos dados, a seleção do 1º quartil, da mediana (2º quartil) e do 3º quartil, em *Estatísticas* na janela de *Sumário Básico e por Fator*, conforme figura abaixo.



Exercício Proposto

1. Para exercitar, calculemos os quartis dos dados sobre nitrogênio e fósforo da tabela do *Exercício Proposto* anterior.

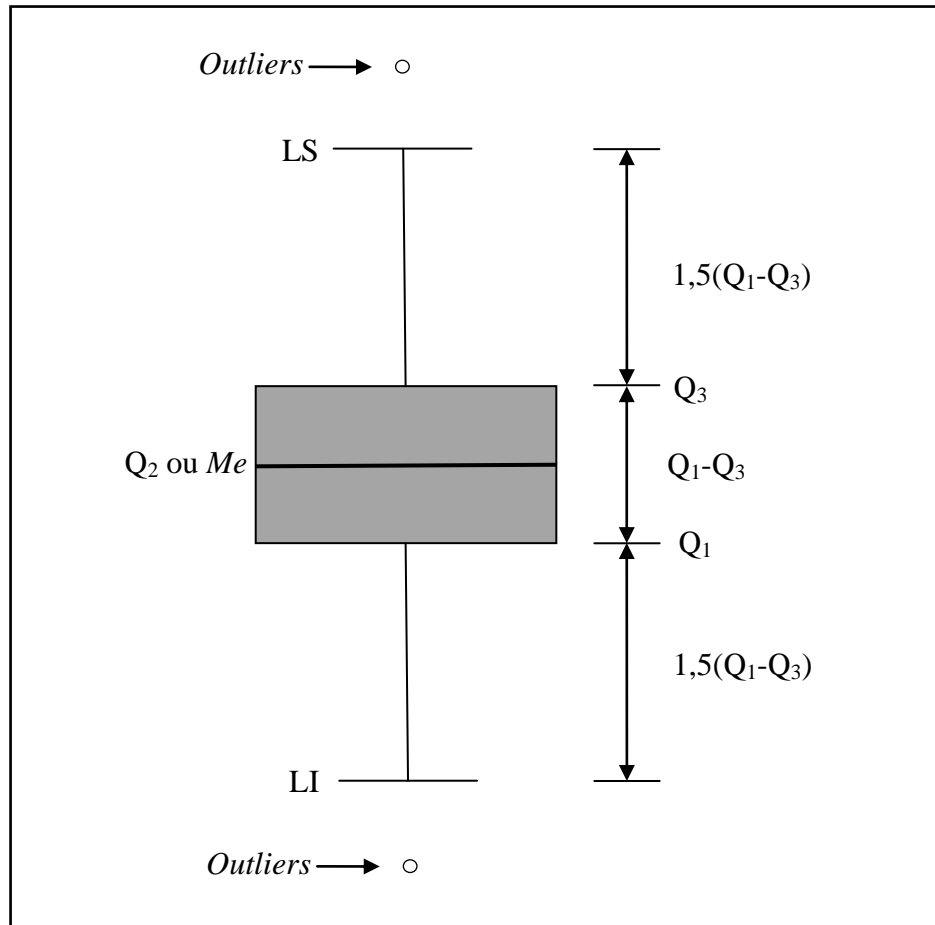
8.5 Box plot (Gráfico de Caixa)

O *Box plot* é um gráfico que resume as informações contidas em um conjunto de dados quantitativos em função de seus parâmetros. Através do *Box plot* podemos verificar:

- os valores máximos e mínimos;
- os valores dos quartis;
- se existe simetria ou assimetria entre os dados observados;
- a variância entre os dados ou entre grupos de dados;
- a presença de dados discrepante, também chamados de *outliers*².

Para compreender melhor o *box plot* apresentamos na figura abaixo um esquema sintetizado com o objetivo de instruí-lo na interpretação desse gráfico, que é uma ferramenta bastante útil para uma análise inicial de um conjunto de dados observados.

² Outliers são valores que destoam muito dos demais, neste material estaremos considerando que isso acontece quando um determinado valor do conjunto de dados estiver fora do intervalo $[Q_1 - 1,5(Q_3 - Q_1); Q_3 + 1,5(Q_3 - Q_1)]$, em que $(Q_3 - Q_1)$ é a distância interquartílica.



Onde:

Outliers: Dados discrepantes

LI: Limite inferior

LS: Limite superior

Q₁: 1º quartil

Q₂: 2º quartil

Q₃: 3º quartil

Me: Mediana

Q₁-Q₃: Distância interquartílica

1,5(Q₁-Q₃): Um e meio da distância interquartílica

O tamanho da caixa indica a variação dos dados, caixas maiores indicam que os dados estão mais dispersos, caixas menores indicam que os dados estão mais agrupados.

8.5.1 Gerando *Box plot* com o software *Action*

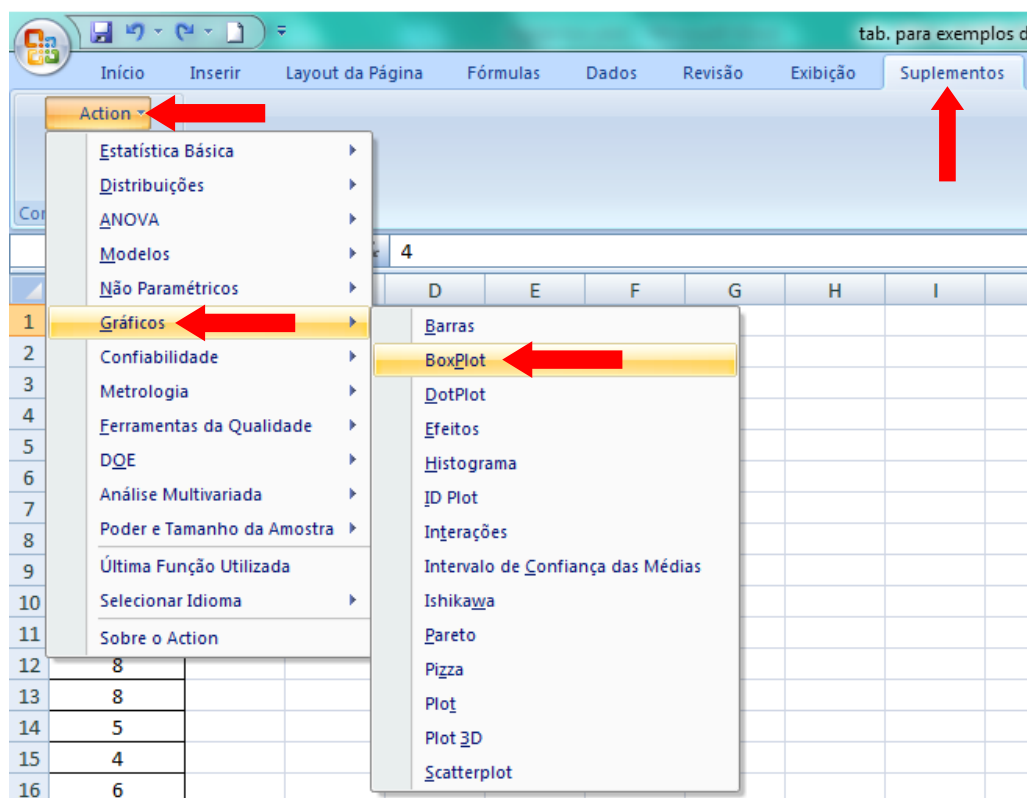
Para construirmos um *Box plot* utilizando o suplemento *Action* do *Excel*, procedemos da seguinte forma:

Tabela 13 – Pesquisa realizada com 50 alunos da UFAC, Campus Floresta, Cruzeiro do Sul – Acre, sobre a quantidade de irmãos.

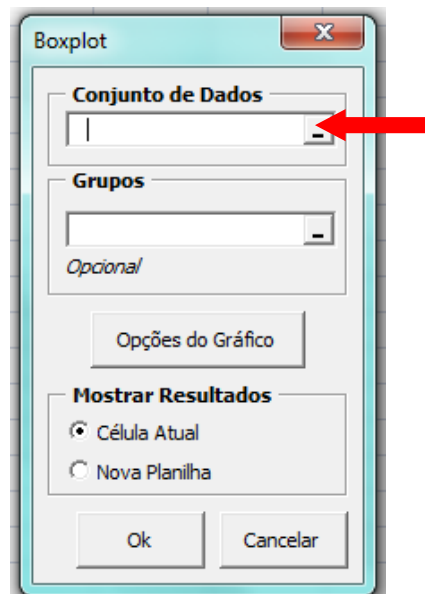
4	2	7	2	3
0	8	4	2	4
8	8	4	1	2
1	5	2	1	2
4	4	4	3	1
2	6	3	6	10
3	16	5	0	2
2	2	6	4	2
5	3	6	2	6
2	2	2	2	12

Fonte: Alunos do 3º período do Curso de Bacharelado em Ciências Biológicas-UFAC.

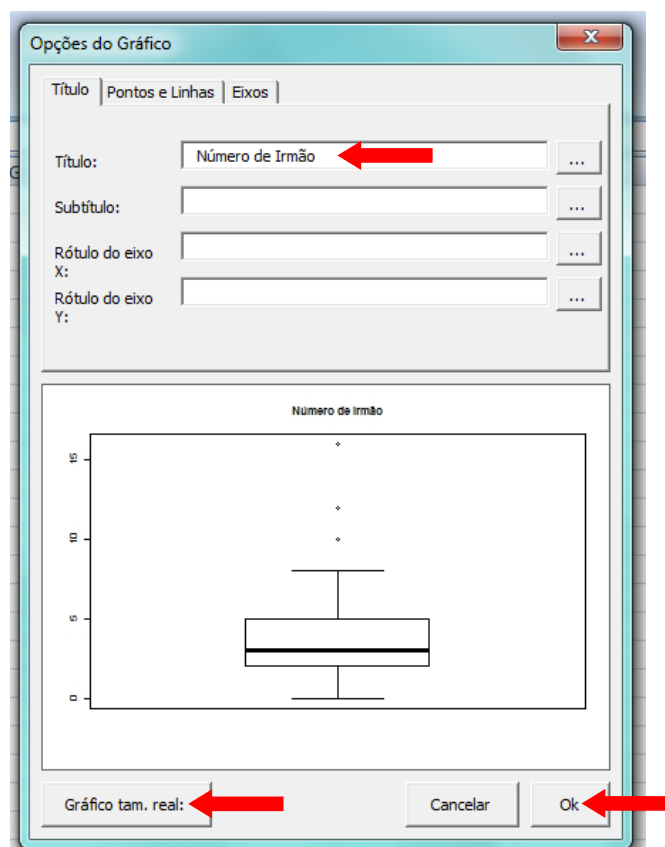
1. Digitamos os dados da Tabela 13 em uma coluna do Excel.
2. Clicamos em **Suplementos** na barra de ferramentas, em seguida, **Action**, **Gráficos** e **BoxPlot**



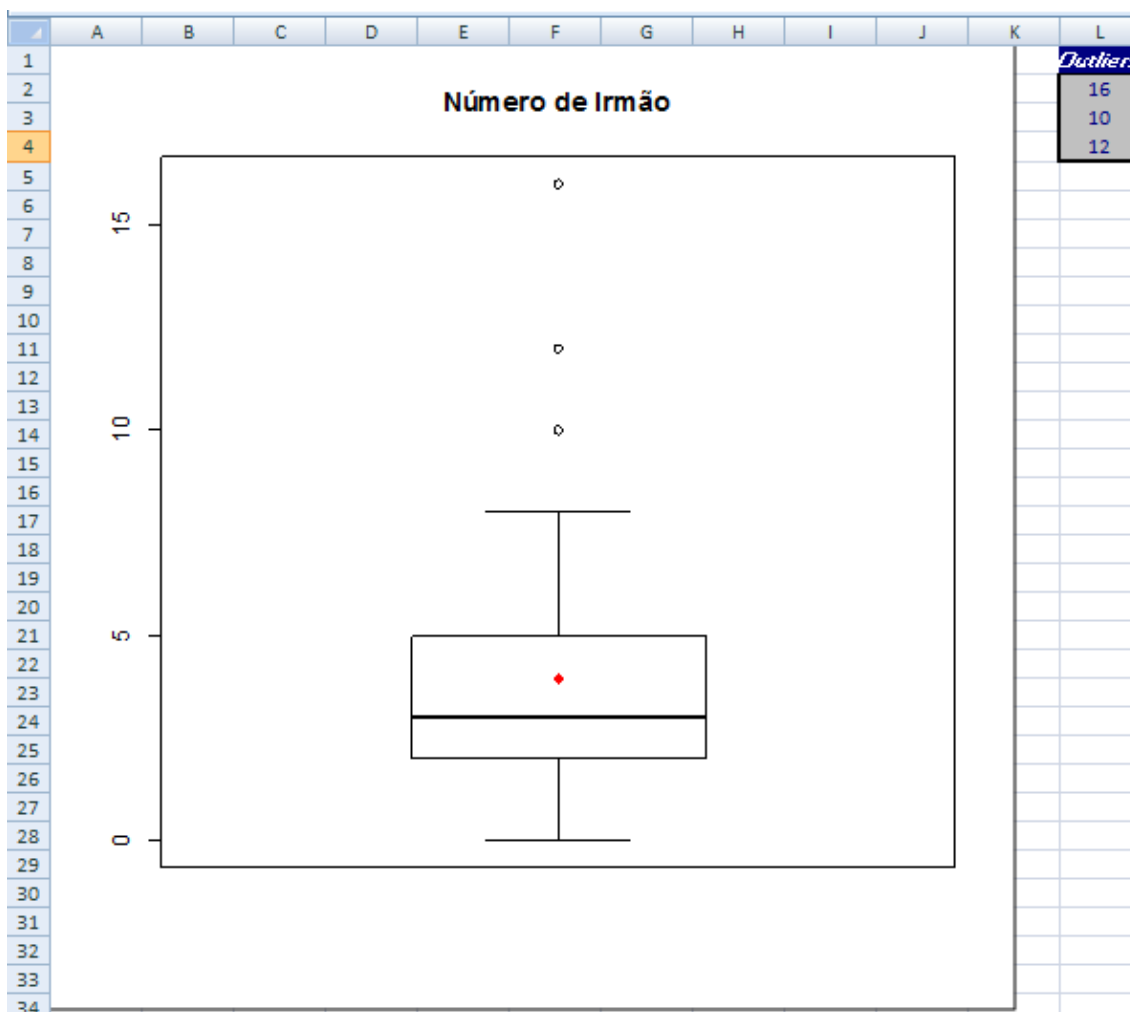
3. A janela *Boxplot* será aberta, clicamos na caixa *Conjunto de dados* e selecionamos os dados sobre o número de irmão.



4. Ao clicarmos em *Opções do Gráfico* podemos editar o *Box plot*, inserindo título, rótulo dos eixos, entre outras formatações. Ao clicarmos em *Gráfico tam. real:* podemos ter uma visualização previa de como ficara o *Box plot*. Após a edição clicamos em *Ok*.



5. O *Action* gera o *box plot* conforme a figura abaixo.



Além do *Box plot* o *Action* nos dá uma coluna contendo os valores discrepantes (*outliers*), se existirem.

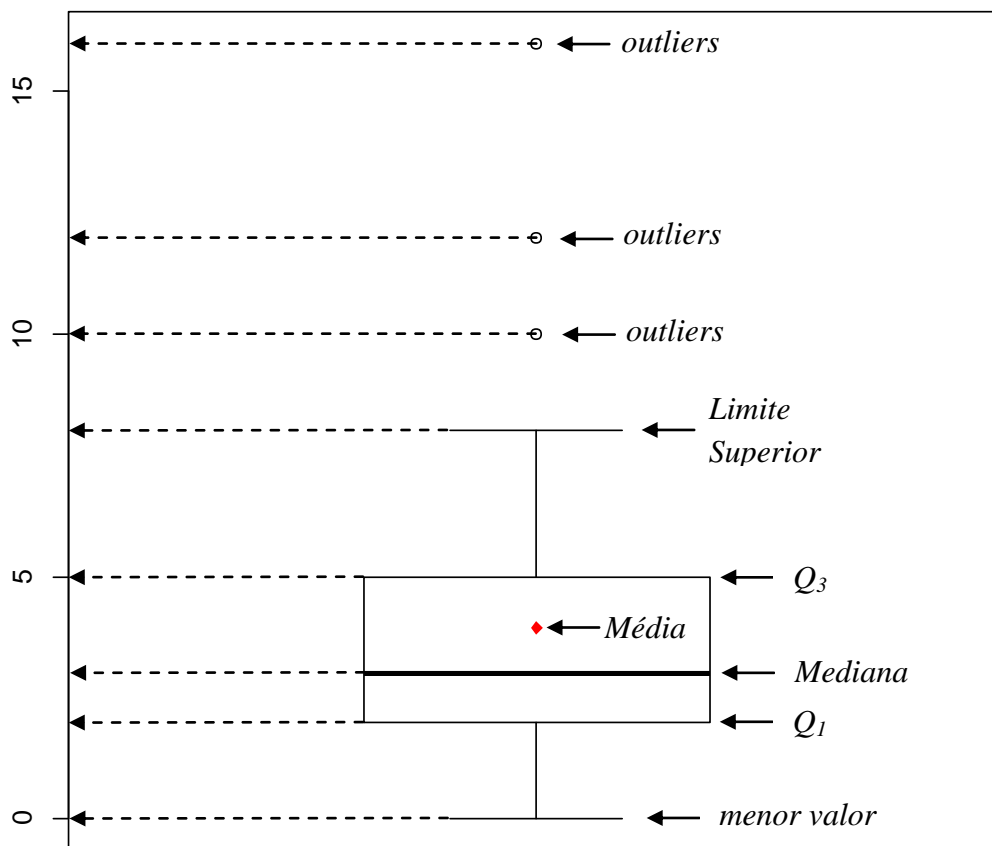
[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar o *Box plot* utilizando o suplemento *Action* do *Excel*.

Com a visualização do *Box plot* podemos chegar a algumas conclusões sobre os dados observados, como por exemplo: há uma leve assimetria entre os dados, pois a mediana (linha mais espessa) está um pouco abaixo do meio da caixa; há dados discrepantes, ou seja, que se diferem da maioria, representados por pontos fora da caixa; a média (ponto vermelho dentro da caixa) é diferente da mediana; 50% das pessoas entrevistadas tem entre 2 e 5 irmãos, representados pela altura da caixa.

O *Action* não representa a linha que indica os limites superiores e inferiores com um e meio do tamanho da caixa, pois seria desnecessário, uma vez que o menor valor observado é inferior a esse tamanho.

A seguir, apresentamos uma figura do *Box plot*, destacando as análises feitas anteriormente, através de seus pontos no gráfico.

Número de Irmão



Percebam que todas as análises são realizadas no eixo vertical, onde neste exemplo, está a quantidade de irmãos dos alunos entrevistados. Como o *Action* apresentou a quantidade de irmão em intervalos de 5 unidades, podemos chegar a valores aproximados dos quartis e da mediana, no entanto, se quisermos saber os valores reais destas estatísticas, precisamos apenas gerar o *sumário básico e por fator* (apresentado no item **8.4 Quartil**) selecionando as medidas estatísticas desejadas.

8.5.2 Gerando *Box plot* para a comparação entre dois ou mais grupos

Com o auxílio de *Action* também podemos gerar *Box plot* para comparar dois ou mais grupos de dados, isso é muito útil quando o pesquisador quer verificar algumas características que são comuns ou que divergem entre esses grupos.

Para exemplificar, imagine que um pesquisador queira verificar a variação dos dados sobre o peso entre homens e mulheres estudantes da UFAC em Cruzeiro do Sul – Acre, para isso ele fez uma pesquisa com 50 alunos. Os dados estão apresentados na tabela a seguir.

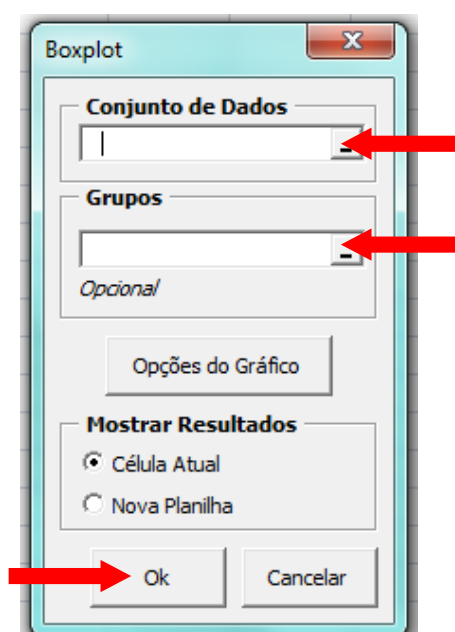
Tabela 14 – Dados sobre peso em kg de homens e mulheres estudantes da UFAC, em Cruzeiro do Sul – Acre

Sexo	Peso	Sexo	Peso	Sexo	Peso	Sexo	Peso
M	97	M	70	F	48	F	68
F	78	F	58	M	68	F	65
F	54	F	51	F	70	M	73
M	76	F	80	F	63	F	75
F	60	F	50	F	55	F	61
M	75	F	45	F	48	M	90
F	51	F	57	F	60	F	60
F	63	F	50	F	64	F	75
F	51	M	81	M	69	M	59
M	73	M	75	F	60	F	55
F	65	M	64	F	65	M	65
M	55	F	58	F	66		
F	65	M	74	F	79		

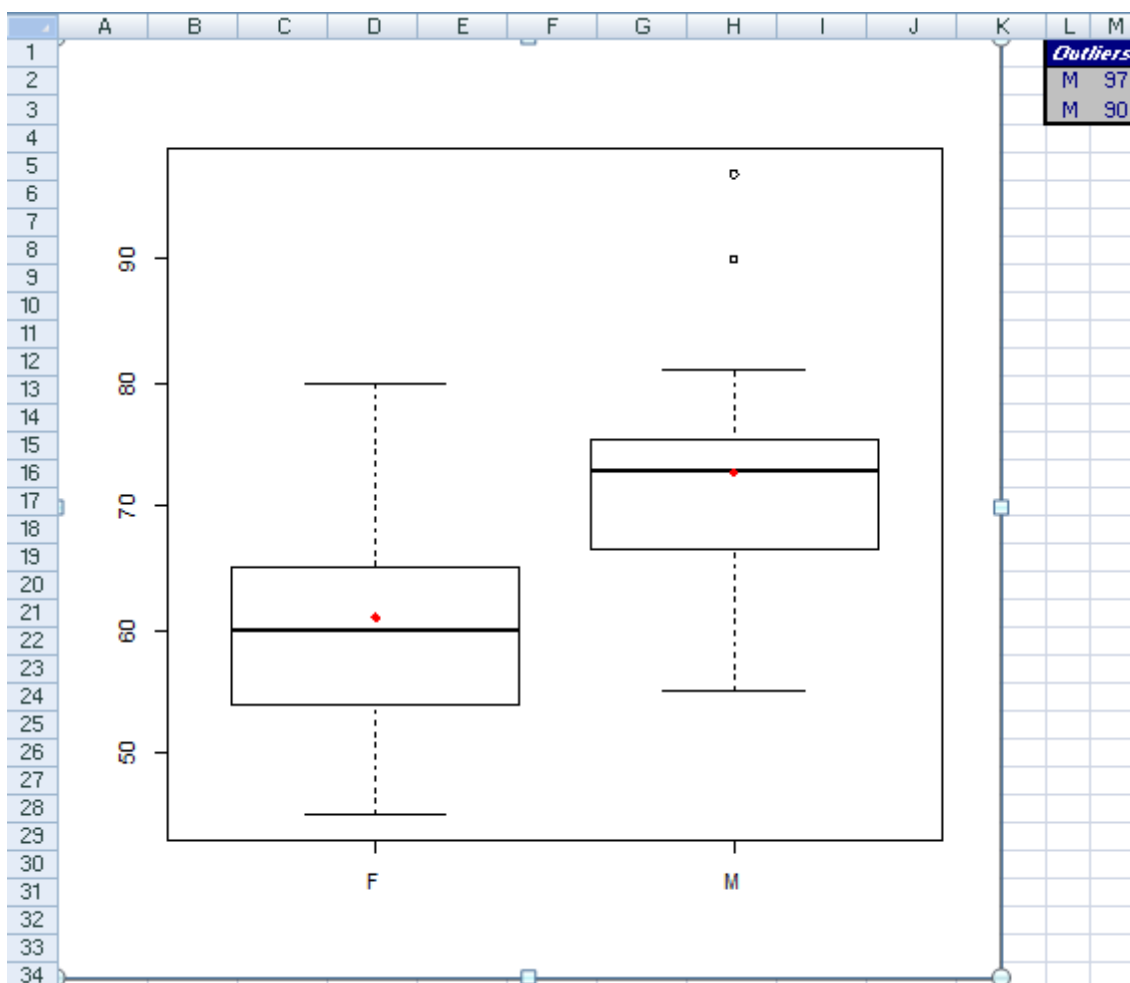
Fonte: Alunos do 3º período do Curso de Bacharelado em Ciências Biológicas-UFAC.

Para construirmos o *Box plot* utilizando o suplemento *Action* do *Excel*, procedemos da seguinte forma:

1. Digitamos os dados da Tabela 12 em uma coluna do *Excel*.
2. Clicamos em **Suplementos** na barra de ferramentas, em seguida, **Action**, **Gráficos** e **BoxPlot**.
3. A janela *Boxplot* será aberta, clicamos na caixa **Conjunto de dados** e selecionamos os dados numéricos sobre o peso dos alunos, em seguida, clicamos na caixa **Grupos** e selecionamos os dados sobre o sexo dos alunos entrevistados. Após a seleção clicamos em **Ok**.



5. O *Action* gera o *blox plot* conforme a figura abaixo.



[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar Box plot para comparação entre dois grupos, utilizando o suplemento *Action* do *Excel*.

A interpretação dos dados através do *Box plot* pode ser feita da seguinte forma:

- os homens são mais pesados que as mulheres, pois a caixa M (masculino) está acima da caixa F (feminino);
- os dados de peso entre as mulheres podem ser considerados simétricos, pois a linha mais espessa está quase no meio da caixa F, lembrando que uma leve assimetria pode ser considerada como simetria em consequência do tamanho da amostra;
- houve a presença de dados discrepante no grupo dos homens, representados pelos dois pontos fora da caixa M.
- o peso das mulheres variou mais que o peso dos homens, isso pode ser verificado devido a caixa F tem altura maior que a caixa M.

Estas são algumas das análises que podem ser feitas quando utilizamos o *Box plot* para comparar dois grupos. Também podemos comparar mais de dois grupos, no entanto,



quanto maior o número de grupos, mais demorada tende a ser a análise. Dessa forma, se o pesquisador precisar fazer a comparação entre a variabilidade de mais de dois grupos recomendamos a escolha de testes estatísticos mais eficientes, como a Análise de variância, entre outros, assunto que abordaremos em outra oportunidade.

Exercício proposto

1. A tabela a seguir contém os dados sobre a idade em anos e o sexo de 48 alunos obtidos em uma pesquisa realizada na UFAC, Campus Floresta, Cruzeiro do Sul – Acre.

Idade	Sexo	Idade	Sexo	Idade	Sexo
25	M	22	M	19	F
18	F	19	F	19	M
38	F	21	F	18	F
46	M	26	F	23	F
24	F	21	F	28	F
19	M	23	F	22	F
26	F	25	F	26	F
18	F	23	F	24	F
19	F	30	M	20	M
22	M	20	M	19	F
20	F	21	M	24	F
19	M	22	F	26	F
19	F	18	M	29	F
25	M	23	F	19	M
22	M	18	F	18	F
26	F	35	F	29	F

- Construa um *Box plot* no *Action* dos dados sobre a idade dos alunos.
- Faça uma análise sobre a variação dos dados e sobre a presença de dados discrepantes através do gráfico gerado no exercício anterior.
- Construa um *Box plot* no *Action* comparando o grupo dos homens e das mulheres.
- Faça uma análise do gráfico construído anteriormente e verifique quais as principais características entre os grupos.



9. NOÇÕES DE CORRELAÇÃO

Até agora vimos como organizar e resumir informações relacionadas a uma única variável. No entanto, na biologia, existem situações nas quais nos interessa estudar o comportamento conjunto de duas ou mais variáveis aleatórias. O objetivo deste estudo é verificar se existe relação entre duas variáveis quantitativas, para realizarmos esta verificação precisamos quantificá-la através do que chamamos de *Coefficiente de Correlação de Pearson*, isso facilitará a compreensão sobre o comportamento dos dados. O primeiro passo para a verificação de correlação entre duas variáveis é a construção de um gráfico que chamamos de *diagrama de dispersão*.

9.1 Diagrama de dispersão

O diagrama de dispersão é uma representação gráfica de dados de duas variáveis quantitativas, em que cada par de variáveis é representado por um ponto (x, y) num sistema de eixos coordenados.

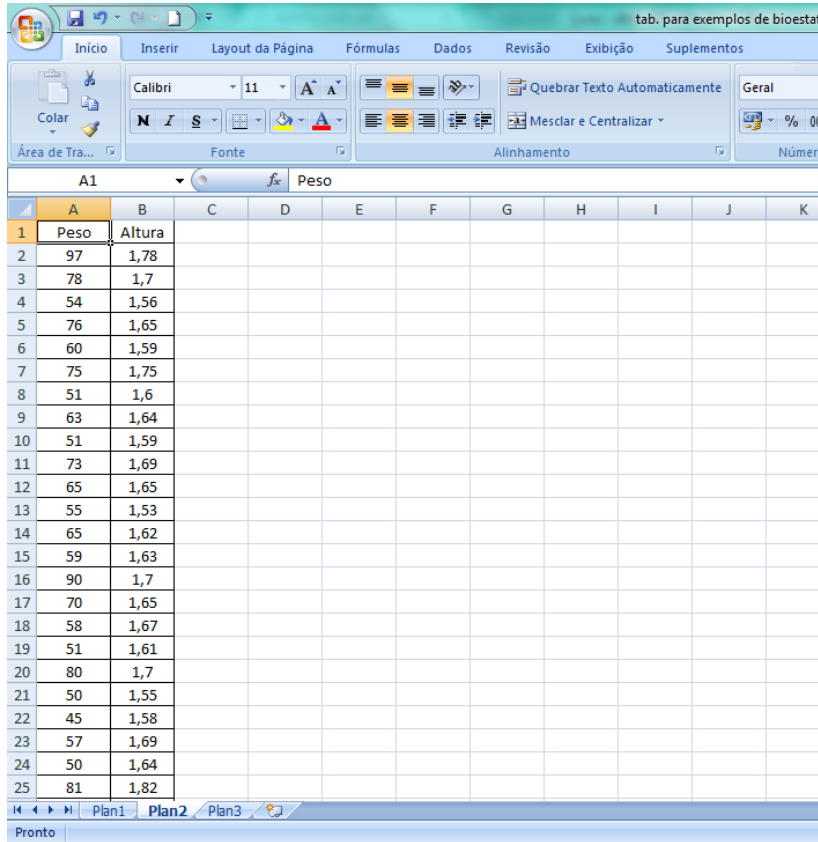
Para construirmos um diagrama de dispersão utilizando o *Microsoft Office Excel* realizamos os seguintes passos:

Tabela 15 – Dados sobre peso (Kg) e altura (m) de 50 alunos da Universidade Federal do Acre – UFAC, coletados na cidade de Cruzeiro do Sul – AC, em 2015.

Peso (kg)	Altura (m)	Peso (kg)	Altura (m)	Peso (kg)	Altura (m)
97	1,78	45	1,58	65	1,59
78	1,70	57	1,69	66	1,69
54	1,56	50	1,64	79	1,76
76	1,65	81	1,82	65	1,77
60	1,59	75	1,69	75	1,67
75	1,75	64	1,72	68	1,60
51	1,6	58	1,54	65	1,73
63	1,64	74	1,78	73	1,83
51	1,59	55	1,60	75	1,68
73	1,69	60	1,63	61	1,58
65	1,65	48	1,58	51	1,61
55	1,53	68	1,80	80	1,70
65	1,62	70	1,73	50	1,55
59	1,63	63	1,70	64	1,67
90	1,70	55	1,58	69	1,70
70	1,65	48	1,55	60	1,77
58	1,67	60	1,58		

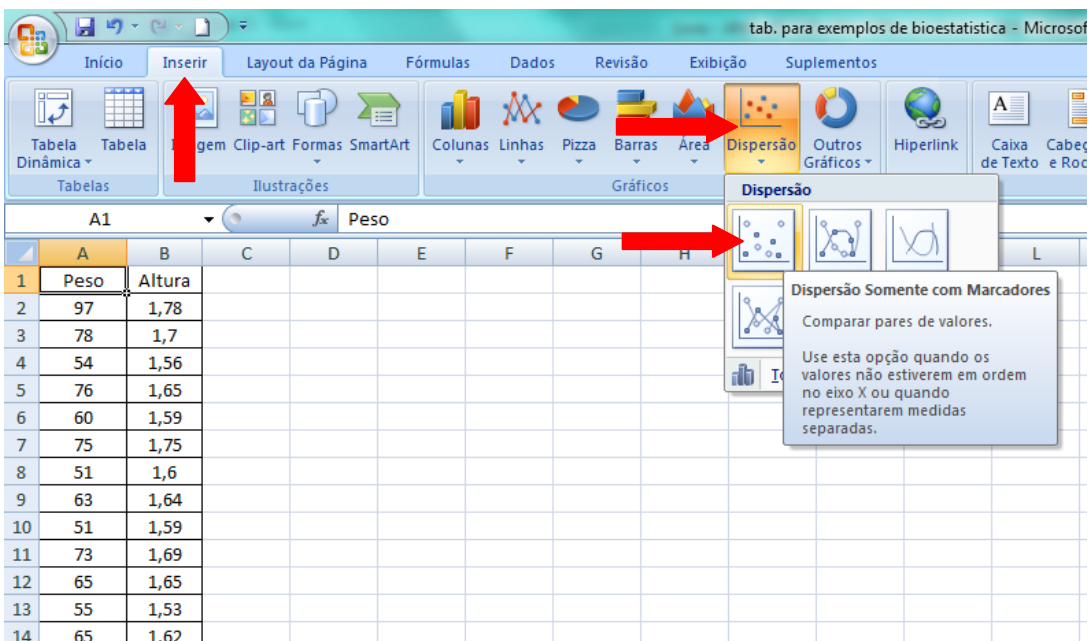
Fonte: Alunos do 3º período do curso de Bacharelado em Ciências Biológicas da UFAC – Campus Floresta.

1. Digitamos os dados sobre o peso dos alunos na coluna A e os dados sobre altura na coluna B do *Excel*.



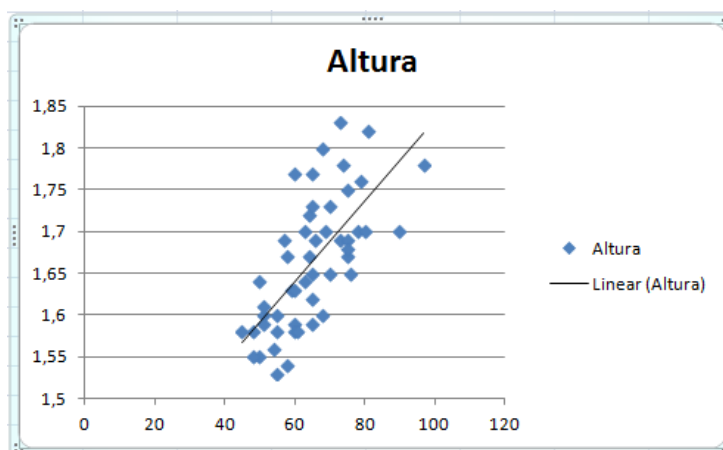
	A	B	C	D	E	F	G	H	I	J	K
1	Peso	Altura									
2	97	1,78									
3	78	1,7									
4	54	1,56									
5	76	1,65									
6	60	1,59									
7	75	1,75									
8	51	1,6									
9	63	1,64									
10	51	1,59									
11	73	1,69									
12	65	1,65									
13	55	1,53									
14	65	1,62									
15	59	1,63									
16	90	1,7									
17	70	1,65									
18	58	1,67									
19	51	1,61									
20	80	1,7									
21	50	1,55									
22	45	1,58									
23	57	1,69									
24	50	1,64									
25	81	1,82									

2. Clique em qualquer célula onde os dados estão inseridos, por exemplo, na célula A1, em seguida, clique no menu **Inserir** na barra de ferramentas, selecione o comando **Dispersão** e depois **Dispersão Somente com Marcadores**.



Dispersão Somente com Marcadores
Comparar pares de valores.
Use esta opção quando os valores não estiverem em ordem no eixo X ou quando representarem medidas separadas.

Com esses passos o *Excel* gera o diagrama de dispersão dos dados digitados na planilha.



Este gráfico nos auxilia a verificar se há uma relação linear entre os dados das variáveis em análise, ou seja, se o formato dos dados agrupados assemelha-se a uma reta.

Ao verificarmos a existência de uma relação linear entre os dados de duas variáveis, pode-se medir o grau de associação através do coeficiente de correlação.

[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar um diagrama de dispersão utilizando o *Excel*.

9.2 Coeficiente de Correlação

Existe uma medida para determinarmos o grau de correlação entre duas variáveis, essa medida é o *Coefficiente de Correlação de Pearson*, representado pela letra *r* e definido pela seguinte fórmula:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Onde:

$\sum xy$: somatório do produto entre *x* e *y*;

$\sum x$: somatório dos valores de *x*;

$\sum y$: somatório dos valores de *y*;

$\sum x^2$: somatório dos valores de *x* elevados ao quadrado;

$\sum y^2$: somatório dos valores de *y* elevados ao quadrado.

O coeficiente de correlação varia entre -1 e 1, ou seja, $-1 \leq r \leq 1$, se r for igual a um então existe uma correlação perfeita positiva entre as variáveis, e se r for igual a -1, então é uma correlação perfeita negativa. Não existe correlação se r for igual a zero.

Para entendermos como aplicar a fórmula da correlação, observe os exemplos de correlação positiva a seguir.

Imagine que um pesquisador queira verificar qual o grau de relação entre as variáveis da tabela a seguir:

x	y
1	1
2	2
3	4
4	5
5	8

Primeiro construímos a tabela de cálculos intermediários para a obtenção do coeficiente de correlação.

x	y	x^2	y^2	xy
1	1	1	1	1
2	2	4	4	4
3	4	9	16	12
4	5	16	25	20
5	8	25	64	40
15	20	55	110	77

Em seguida, substituímos os valores obtidos na tabela acima na fórmula da correlação, como podemos ver a seguir:

$$r = \frac{77 - \frac{15 \times 20}{5}}{\sqrt{\left(55 - \frac{15^2}{5}\right) \left(110 - \frac{20^2}{5}\right)}}$$

$$r = \frac{77 - 60}{\sqrt{(55 - 45)(110 - 80)}}$$

$$r = \frac{17}{\sqrt{300}}$$

$$r = 0,98$$

Podemos interpretar o resultado acima dizendo que as variáveis observadas apresentaram um percentual de correlação positiva de 98%. Isso significa que a medida que a variável x cresce, a variável y também cresce, e uma explica a outra em torno de 98%.

Agora observemos um exemplo de correlação negativa.

Imagine que um pesquisador queira verificar qual o grau de relação entre as variáveis da tabela a seguir:

x	y
1	7
2	4
3	4
4	3
5	1

Da mesma forma do exemplo anterior, primeiro construímos a tabela de cálculos intermediários para a obtenção do coeficiente de correlação.

x	y	x^2	y^2	xy
1	7	1	49	7
2	4	4	16	8
3	4	9	16	12
4	3	16	9	12
5	1	25	1	5
15	19	55	91	44

Em seguida, substituído os valores obtidos na tabela acima na fórmula da correlação.

$$r = \frac{44 - \frac{15 \times 19}{5}}{\sqrt{\left(55 - \frac{15^2}{5}\right)\left(91 - \frac{19^2}{5}\right)}}$$

$$r = \frac{44 - 57}{\sqrt{(55 - 45)(91 - 72,2)}}$$

$$r = \frac{13}{\sqrt{188}}$$

$$r = -0,95$$

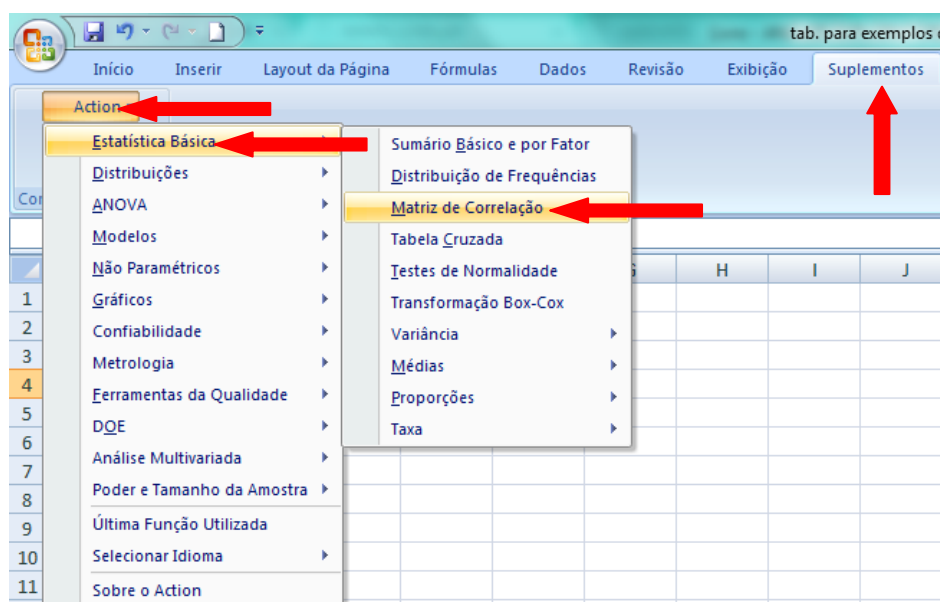
Podemos interpretar o resultado acima dizendo que as variáveis observadas apresentaram um percentual de correlação negativa de 95%. A correlação negativa indica comportamentos inversos com relação as variáveis, ou seja, a medida que uma cresce a outra tende a diminuir em termos de valores.

9.3 Cálculo do coeficiente de correlação com o *Software Action*.

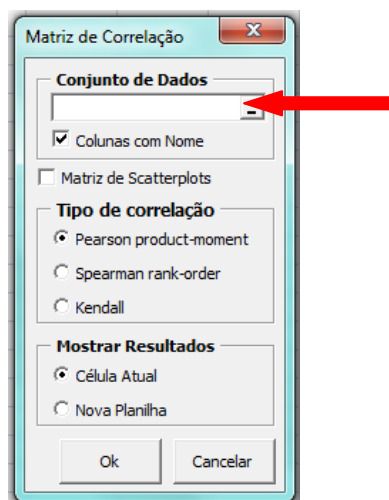
O *software Action* possibilita o cálculo do coeficiente de correlação com apenas alguns cliques, para exemplificar utilizaremos os dados da tabela 10, sobre o peso e a altura de 50 alunos da UFAC, em 2015.

1. Digitamos os dados sobre o peso dos alunos na coluna A e os dados sobre altura na coluna B do *Excel*, semelhante ao realizado na construção de diagramas de dispersão.

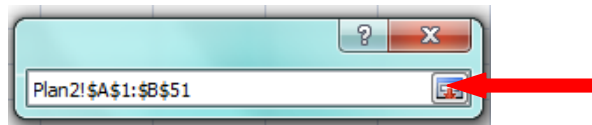
2. Clique no menu **Suplementos** na barra de ferramentas, selecione o comando, **Action**, **Estatística Básica** e em seguida **Matriz de Correlação**.



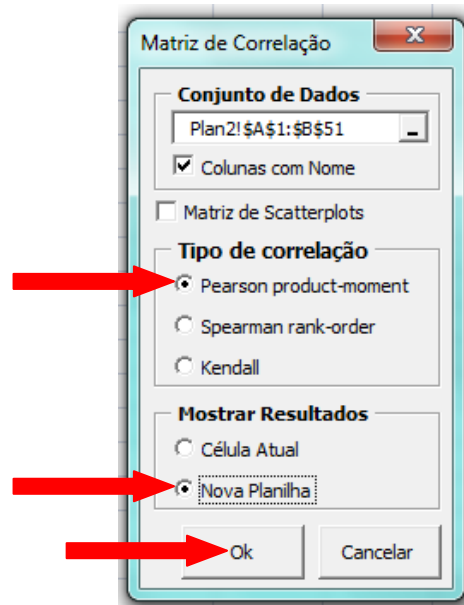
3. A janela de Matriz de Correlação será aberta, clique no canto direito da caixa **Conjunto de Dados** e selecione os dados das duas colunas, inclusive as células com os nomes das variáveis.



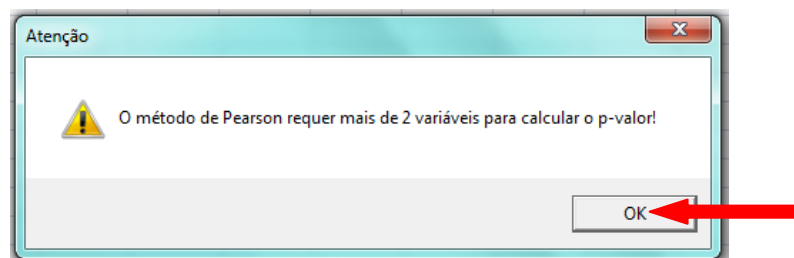
4. Clique novamente no canto direito da caixa **Conjunto de Dados**, conforme figura abaixo:



5. Selecione as opções **Pearson product-moment**, **Nova Planilha** e clique em **Ok**.



6. Aparecerá uma janela com um aviso sobre o método de Pearson requerer mais de 2 variáveis para calcular o p-valor, clique em Ok, pois não nos interessa neste instante o cálculo do p-valor.



O Action gera a Matriz de Correlação com o valor do coeficiente de correlação de Pearson, conforme figura abaixo:

	A	B	C
1	MATRIZ DE CORRELAÇÃO		
2			
3	DADOS DO PROCESSO		
4			
5	Matriz de Correlação: Pearson		
6		<i>Peso</i>	<i>Altura</i>
7	<i>Peso</i>	1	0,681749646
8	<i>Altura</i>	0,681749646	1



Em nosso exemplo, o coeficiente de correlação entre o peso e a altura dos alunos foi igual a 0,68. Isso significa que há uma relação entre as variáveis *peso* e *altura* de 68% nos dados observados.

[Clique aqui](#) e assista ao vídeo tutorial sobre como coeficiente de correlação utilizando o suplemente *Action* do *Excel*.

Exercício Proposto

Os dados da tabela 16 foram coletados com o objetivo de avaliar a relação entre riqueza e composição de espécies de peixes com as variáveis bióticas e abióticas dos lagos de meandro abandonado da região do rio Purus - Acre.

Utilizando o *Software Action* realize as seguintes análises.

1. Calcule o coeficiente de correlação e verifique qual o grau de relação existente entre as seguintes variáveis:

- a) Temperatura da água e profundidade;
- b) Profundidade e pH;
- c) Temperatura do ar e nitrogênio
- d) transparência da água e fósforo
- e) Riqueza de espécies de plantas aquáticas e pH;
- f) Riqueza de espécies de plantas aquáticas e temperatura da água;
- g) Riqueza de espécies de plantas aquáticas e profundidade;
- h) Riqueza de espécies de plantas aquáticas e transparência da água;
- i) Riqueza de espécies de plantas aquáticas e nitrogênio;
- j) Riqueza de espécies de plantas aquáticas e fósforo;
- k) Riqueza de espécies de peixes e pH;
- l) Riqueza de espécies de peixes e temperatura da água;
- m) Riqueza de espécies de peixes e profundidade;
- n) Riqueza de espécies de peixes e transparência da água;
- o) Riqueza de espécies de peixes e nitrogênio;
- p) Riqueza de espécies de peixes e fósforo;



2. Faça uma análise dos dados considerando o grau de relação entre as variáveis estudadas, em seguida descreva quais os principais fatores contribuem para uma maior riqueza de espécies de plantas aquáticas nos locais de coleta dos dados.
3. Descreva quais os principais fatores contribuem para uma maior riqueza de espécies de peixes nos locais de coleta dos dados.

Tabela 16 – Dados sobre a relação entre riqueza e composição de espécies de peixes com as variáveis bióticas e abióticas dos lagos de meandros abandonados da região do Rio Purus – Acre.

Lago	Coleta	Banco	pH	Cond	Temp.	Transp.	NT	PT	Temp. Ar	Conect.	Prof	Riq. Macro	Cober.	Área Macro	Riqueza de peixes
Itapira	1	1	6,21	84	26,12	70	1,28	0,381	26	274	1500	3	4,2	1425	14
Itapira	1	2	6,1	83	30,5	70	1,58	0,481	24,8	274	1000	3	3,1	1426	11
Itapira	1	3	6,8	78	29,2	70	1,58	0,668	24,8	274	500	3	4,2	1427	10
Itapira	1	4	6,9	90	28,9	70	1,48	0,411	30	274	1200	3	6	1428	9
Itapira	1	5	6,7	79	25,34	70	2,58	0,281	30	274	1000	3	8,6	1429	11
Cameta	1	1	5,9	78	28,6	65	1,28	0,381	24	1141	1500	6	3	2205	9
Cameta	1	2	6,06	78	29	65	1,58	0,481	24	1141	1000	7	3	2206	10
Cameta	1	3	5,8	65	28,35	65	1,58	0,668	25,33	1141	500	5	3,3	2207	10
Cameta	1	4	6,2	65	27	65	1,48	0,411	25,33	1141	1200	8	3	2208	5
Cameta	1	5	6,33	78	27,89	65	2,58	0,281	26	1141	1000	8	11,8	2209	9
Santana	1	1	6,13	87	31	38	1,47	0,336	25,6	587	1450	4	3,1	2719	4
Santana	1	2	5,42	91	39,8	38	1,37	0,396	25,6	587	1250	3	5,41	2720	9
Santana	1	3	5,86	92	28	38	2,17	0,236	29,3	587	850	4	4,4	2721	5
Santana	1	4	5,6	90	29,22	38	1,55	0,206	29,3	587	950	4	4,8	2722	7
Santana	1	5	6,9	91	30,6	38	1,343	0,296	38,5	587	880	1	2,9	2723	6
Bom_Lugar	1	1	6,11	57	28,7	50	1,47	0,336	26,6	2204	1450	11	14,8	2683	5
Bom_Lugar	1	2	6,35	58	27	50	1,37	0,396	26,6	2204	1250	6	3	2684	3
Bom_Lugar	1	3	6,32	58	26,66	50	2,17	0,236	24	2204	850	6	3	2685	2
Bom_Lugar	1	4	6,35	59	32	50	1,55	0,206	24	2204	950	7	3	2686	1
Bom_Lugar	1	5	6,01	58	33,02	50	1,343	0,296	25	2204	880	9	6	2687	7
Lago_Verde	1	1	6,48	61	25	50	1,482	0,211	25,5	917,631	1320	4	2,94	2091	2
Lago_Verde	1	2	6,48	61	25	10	1,482	0,211	25,5	917,631	1420	4	12,05	2092	5
Lago_Verde	1	3	6,68	62	25,9	10	1,482	0,211	24,5	917,631	1320	4	12	2093	5



Lago_Verde	1	4	6,7	63	25,8	10	1,482	0,211	24,5	917,631	920	3	2,45	2094	3
Lago_Verde	1	5	7,1	66	25,6	10	1,482	0,211	24,8	917,631	1020	3	5,45	2095	9
Ig_Preto	1	1	6,11	73	28,9	10	1,63	0,254	24,5	1007	345	2	4,5	154	5
Ig_Preto	1	2	5,99	65	24,4	10	1,13	0,274	21,9	1007	234	2	6,7	114	4
Ig_Preto	1	3	6,99	71	26,9	10	1,73	0,234	26,8	1007	213	3	2,3	89	4
Ig_Preto	1	4	6,89	74	26,9	10	1,83	0,214	26,8	1007	231	3	5,5	111	4
Ig_Preto	1	5	6,71	72	26,9	10	1,93	0,294	26,8	1007	124	1	2,1	83	3
Flor_ouro	1	1	6,1	76	26,41	30	1,482	0,211	26	411	1320	6	9,25	3902	6
Flor_ouro	1	2	6,17	85	29,72	30	1,482	0,211	26	411	1420	6	3,3	3903	6
Flor_ouro	1	3	6,8	87	30,1	30	1,482	0,211	27,8	411	1320	4	11,9	3904	2
Flor_ouro	1	4	6,9	84	28,8	30	1,482	0,211	27,8	411	920	5	6,6	3905	1
Flor_ouro	1	5	6,3	77	27,7	30	1,482	0,211	28,8	411	1020	4	2,84	3906	1
Anuri	1	1	6,48	61	26,8	20	5,262	2,241	31,8	2152	678	2	9	1831	3
Anuri	1	2	6,28	64	25,9	20	5,862	1,641	31,8	2152	468	2	6,3	3031	2
Anuri	1	3	6,43	76	25,11	20	4,262	1,841	31,8	2152	346	2	10,9	2831	3
Anuri	1	4	6,38	77	25,11	20	4,992	1,241	31,8	2152	401	2	6	1431	1
Anuri	1	5	6,8	78	25,11	20	5,162	1,841	31,8	2152	238	2	2,84	2631	3
Itapira	2	1	6,5	125,8	30	30	1,99	0,432	37	274	1420	2	3	475	2
Itapira	2	2	6,5	125,8	30	30	1,99	0,432	37	274	420	2	3	476	5
Itapira	2	3	6,5	125,8	30	30	1,99	0,432	39,8	274	1420	1	3	477	1
Itapira	2	4	6,5	125,8	30	30	1,99	0,432	39,8	274	920	1	3	478	3
Itapira	2	5	6,5	125,8	30	30	1,99	0,432	28,7	274	520	2	3	479	14
Cameta	2	1	6,58	98	28,9	65	1,401	0,205	30	1141	1420	8	2,4	2506	10
Cameta	2	2	6,628	99	34	65	1,401	0,205	30	1141	420	5	2,3	2507	12
Cameta	2	3	6,604	114	32	65	1,401	0,205	31	1141	1420	3	1,1	2508	4
Cameta	2	4	6,703	116	31	65	1,401	0,205	31	1141	920	3	3	2509	10
Cameta	2	5	6,932	119	30	65	1,401	0,205	30,2	1141	520	3	6	2510	12
Santana	2	1	6,1	103	31,9	70	1,586	0,24	30,2	587	1500	3	12	2788	14



Santana	2	2	7,37	109	31	70	1,586	0,24	32,3	587	1300	3	5	2789	12
Santana	2	3	6,75	112	31,5	70	1,586	0,24	32,3	587	1100	3	5,3	2790	9
Santana	2	4	6,2	109	32,1	70	1,586	0,24	36,1	587	1300	3	22,3	2791	11
Santana	2	5	6,7	109	29,33	70	1,586	0,24	34	587	1400	3	3	2792	7
Bom_Lugar	2	1	6,83	87,2	29,5	37	2,105	0,439	30	1018	1500	4	3	1217	4
Bom_Lugar	2	2	6,82	84,4	29,4	37	2,105	0,439	28	990	1300	5	3	1217	3
Bom_Lugar	2	3	6,82	86,3	28,5	37	2,105	0,439	29	975	1100	3	3	1217	7
Bom_Lugar	2	4	6,82	86,3	28,5	37	2,105	0,439	29	975	1300	2	3	1217	2
Bom_Lugar	2	5	6,82	86,3	28,5	37	2,105	0,439	29	975	1400	2	3	1217	3
Lago_Verde	2	1	6,77	141	31	65	1,257	0,269	25,1	917,631	1630	4	3	3807	7
Lago_Verde	2	2	6,66	147	29,66	65	1,257	0,269	25,1	917,631	1230	6	6,1	3808	4
Lago_Verde	2	3	6,66	147	30	65	1,257	0,269	34	917,631	1530	3	4	3809	1
Lago_Verde	2	4	6,7	140	28,5	65	1,257	0,269	34	917,631	1520	5	5,1	3810	2
Lago_Verde	2	5	6,5	141	28,8	65	1,257	0,269	33	917,631	1190	1	3	3811	10
Ig_Preto	2	1	6,87	17	26,91	55	1,655	0,201	38,8	1017	1630	3	2	114	9
Ig_Preto	2	2	7,25	19	27,74	55	1,655	0,201	45,5	1017	1230	3	3,3	228	12
Ig_Preto	2	3	6,6	22	26,82	55	1,655	0,201	48,2	1017	1530	3	3	68	7
Ig_Preto	2	4	6,83	15	27,7	55	1,655	0,201	48,2	1017	1520	3	6,6	69	12
Ig_Preto	2	5	6,5	19	29,9	55	1,655	0,201	48,2	1017	1190	4	3	70	11
Flor_ouro	2	1	6,23	124	28,5	70	1,551	0,197	29,6	411	800	4	3	5463	6
Flor_ouro	2	2	6,72	155	26,29	70	1,551	0,197	26	411	1200	4	3,4	5464	10
Flor_ouro	2	3	6,72	114	28,9	70	1,551	0,197	26	411	700	3	9,4	5465	3
Flor_ouro	2	4	6,84	129	26,36	70	1,551	0,197	27,8	411	870	4	12,4	5466	8
Flor_ouro	2	5	6,4	114	29,72	70	1,551	0,197	27,8	411	980	3	15,5	5467	8
Anuri	2	1	6,4	13	26,6	50	1,984	0,347	24,4	2146	1123	1	3	2831	5
Anuri	2	2	5,87	89	25,11	50	1,984	0,347	24,4	2146	145	2	3	2831	6
Anuri	2	3	5,98	13	27	50	1,984	0,347	24,1	2146	1226	3	6,04	2831	3
Anuri	2	4	5,84	14	24,9	50	1,984	0,347	24,1	2146	789	2	3	2831	6



Anuri	2	5	5,93	13	5,31	50	1,984	0,347	34,4	2146	680	4	3,58	2831	14
Itapira	3	1	7,34	137	30,01	20	2,832	0,23	28,9	274	280	3	3	950	7
Itapira	3	2	7,31	132	32	20	2,832	0,23	29	274	390	3	3,1	951	14
Itapira	3	3	7,24	139	30,8	20	2,832	0,23	31	274	450	3	3	952	9
Itapira	3	4	7,6	136	35	20	2,832	0,23	31	274	270	3	3	953	12
Itapira	3	5	7,6	136	35	20	2,832	0,23	31	274	256	3	2,6	954	12
Cameta	3	1	8,23	108	31,1	60	2,278	0,217	26,7	1141	120	4	3,5	2306	9
Cameta	3	2	8,12	109	29,1	60	2,278	0,217	26,7	1141	145	4	3,5	2307	5
Cameta	3	3	8,1	110	30,2	60	2,278	0,217	25,8	1141	178	4	2,9	2308	7
Cameta	3	4	8,2	109	37,4	60	2,278	0,217	25,8	1141	167	3	3	2309	2
Cameta	3	5	7,8	110	33,5	60	2,278	0,217	24,5	1141	167	4	3	2310	5
Santana	3	1	7,95	129	36,5	40	3,075	0,221	30	587	230	4	3	2370	9
Santana	3	2	7,88	129,3	31,5	40	3,075	0,221	30	587	245	4	3	2371	7
Santana	3	3	6,19	126	27,8	40	3,075	0,221	31,4	587	456	3	3	2372	7
Santana	3	4	7,82	129	24	40	3,075	0,221	31,4	587	346	4	3	2373	3
Santana	3	5	7,41	127	29	40	3,075	0,221	28,8	587	324	4	6	2374	2
Bom_Lugar	3	1	7,77	126	33,4	30	2,896	0,16	28,8	510	267	4	6,7	2202	7
Bom_Lugar	3	2	7,81	127	35,1	30	2,896	0,16	28,8	420	234	4	9,8	2202	6
Bom_Lugar	3	3	6,74	121	29,6	30	2,896	0,16	28,8	320	123	5	6,9	2204	6
Bom_Lugar	3	4	6,74	121	29,6	30	2,896	0,16	28,8	320	126	2	6,9	2204	1
Bom_Lugar	3	5	6,74	121	29,6	30	2,896	0,16	28,8	320	178	2	6,9	2204	0
Lago_Verde	3	1	6,4	145	32,2	65	1,257	0,269	26,6	917,631	189	4	3	3399	5
Lago_Verde	3	2	6,9	103	28,5	25	2,382	0,196	26,6	917,631	101	4	2,8	3400	4
Lago_Verde	3	3	7,64	105	34	25	2,382	0,196	26,6	917,631	256	3	3,05	3401	5
Lago_Verde	3	4	7,6	105	34,4	25	2,382	0,196	26,6	917,631	135	4	3	3402	3
Lago_Verde	3	5	6,81	104,2	26,4	25	2,382	0,196	26,6	917,631	156	3	3	3403	5
Ig_Preto	3	1	7,74	98,2	31,4	40	3,09	0,671	38,8	1017	22	3	4,5	98	3
Ig_Preto	3	2	7,92	69,9	36,9	40	2,89	1,871	40,5	1017	21	2	3,5	68	4



Ig_Preto	3	3	7,76	63,5	33,1	40	2,99	3,251	38,2	1017	11	2	2,2	47	3
Ig_Preto	3	4	7,76	63,5	33,1	40	3,31	1,251	40	1017	12	3	3,2	48	3
Ig_Preto	3	5	7,76	63,5	33,1	40	2,34	0,251	39	1017	21	4	1,2	38	4
Flor_ouro	3	1	6,23	84	26,41	60	3,202	0,249	30,7	411	123	5	3,3	4683	7
Flor_ouro	3	2	6,72	77	29,72	60	3,202	0,249	30,7	411	234	5	2,95	4684	2
Flor_ouro	3	3	6,72	76	30,1	60	3,202	0,249	31,4	411	356	5	2,4	4685	7
Flor_ouro	3	4	6,84	124	28,8	60	3,202	0,249	31,4	411	476	4	3,95	4686	2
Flor_ouro	3	5	6,4	155	28,8	60	3,202	0,249	27,8	411	357	4	3,8	4687	7
Anuri	3	1	6,25	26	29	50	2,978	1,196	24,4	2146	457	3	4,5	831	10
Anuri	3	2	6,45	28	27	50	2,278	1,106	24,1	2146	453	2	5,6	1831	11
Anuri	3	3	6,22	29	30	50	1,998	1,156	34,4	2146	234	2	6,6	2031	6
Anuri	3	4	6,22	29	30	50	2,178	1,116	34,4	2146	278	3	2,3	791	9
Anuri	3	5	6,22	29	30	50	2,298	1,236	34,4	2146	170	3	4,5	531	9

FONTE: Prof^o Lucena Rocha**Legendas**

Coluna	Significado	Unidade de medida
Lago	Local onde foi coletada a amostra.	-
Banco	Banco de macrofitas, vegetação aquática	unidade
pH	Potencial de hidrogênio	mol/L
Cond.	Condutividade elétrica da água	mS/cm
Temp.	Temperatura da água	°C
Transp.	Transparência da água	cm
NT	Nitrogênio	K
PT	Fósforo	
Temp. Ar	Temperatura do ar	°C
Conect.	Conexão hidrológica do rio com o lago	m
Prof.	Profundidade	m
Riq. Macro	Riqueza de espécies de plantas aquáticas nos bancos	unidade
Cobertura	Cobertura total de vegetação aquática	m
Área Macro	Área total de vegetação aquática	m ²
Riqueza	Riqueza de peixes	unidade



Referências

CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Artimed, Porto Alegre, 2003.

MEDRI, W. **Análise Exploratória de Dados**. Universidade Estadual de Londrina, Londrina, 2011.

ODUM, E. P. **Fundamentos de Ecologia**. 6ª ed. Fundação Calouste Gulbenkian, São Paulo, 2004.

VIEIRA, S. **Introdução a Bioestatística**. 4ª ed. Elsevier, Rio de Janeiro, 2011.

ZAROS, L. G. **Bioestatística**. 2ª ed. EDUFRN, 2011.