

**UNIVERSIDADE FEDERAL DO ACRE
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA NATUREZA
MESTRADO PROFISSIONAL EM ENSINO DE CIÊNCIAS E MATEMÁTICA**

ROGÉRIO LOPES CRAVEIRO

PLATAFORMA ON-LINE PARA O ENSINO DE BIOESTATÍSTICA: Ambiente virtual de aprendizagem e *software* estatístico como ferramentas de apoio ao ensino.

**Rio Branco – Ac
2016**

ROGÉRIO LOPES CRAVEIRO

PLATAFORMA ON-LINE PARA O ENSINO DE BIOESTATÍSTICA: Ambiente virtual de aprendizagem e *software* estatístico como ferramentas de apoio ao ensino.

Dissertação apresentada ao Mestrado Profissional em Ensino de Ciências e Matemática da Universidade Federal do Acre, como parte dos requisitos para a obtenção do título de Mestre em Ensino de Ciências e Matemática.

Orientador: Profº Dr. Edcarlos Miranda de Souza
Co-orientador: Profº Dr. Itamar Miranda da Silva

Rio Branco – Ac
2016

ROGÉRIO LOPES CRAVEIRO

PLATAFORMA ON-LINE PARA O ENSINO DE BIOESTATÍSTICA: Ambiente virtual de aprendizagem e *software* estatístico como ferramentas de apoio ao ensino.

Dissertação apresentada ao Curso de Mestrado Profissional em Ensino de Ciências e Matemática da Universidade Federal do Acre, como parte dos requisitos para a obtenção do título de Mestre em Ensino de Ciências e Matemática.

Aprovado em: ____/____/____

Banca Examinadora

Prof^o Dr. Edcarlos Miranda de Souza
Universidade Federal do Acre
Orientador

Prof. Dr. José Ronaldo Melo
Universidade Federal do Acre
Membro Interno

Prof. Dr. Sérgio Brazil Júnior
Universidade Federal do Acre
Membro Externo

**Rio Branco – Ac
2016**

DEDICATÓRIA

Dedico esse trabalho à minha família, professores e amigos pela compreensão e incentivo que tornaram possíveis a sua realização.

AGRADECIMENTOS

Ao Divino Mestre pelo dom da vida.

Aos meus Pais pelos valores morais e espirituais que sempre me ensinaram.

A minha esposa Roquilene que sempre me incentivou a não desistir e não mediu esforços para realizar meus sonhos, sendo uma companheira e conselheira nos momentos que mais precisei.

Aos meus filhos Samiris, Elis e Roger pelo carinho e apoio em minha caminhada para a conclusão dessa etapa de minha vida, vocês são o melhor presente que Deus me deu.

Ao meu irmão Jesus Nazareno e sua esposa Kátya Fernandes pela acolhida em sua residência, fazendo com que me sentisse em minha própria casa.

Ao meu irmão José Orlean que é um grande amigo.

Ao meu irmão Paulo Onofre pelo carinho que sempre teve por mim em todos os momentos de minha vida.

Ao amigo Jânio Melo, pelo incentivo em trilhar os caminhos da matemática.

Ao amigo Genivaldo a quem tenho uma imensa gratidão pelo auxílio em tudo que precisei.

Aos professores e amigos da primeira turma do MPECIM, pela troca de experiências.

Aos amigos Orleilson, Vilma, Paulo José, Valquírio e Ivaneide, pela amizade que firmamos.

Aos amigos Gastão, Jonas, Ramon e José Alves pela compreensão e apoio.

Ao meu co-orientador Itamar Miranda, pelas dicas sempre pertinentes para a melhoria do trabalho.

Ao meu orientador e amigo Edcarlos Miranda de Souza pelo compromisso que teve com a qualidade do nosso produto, incentivando-me a superar limites e buscar o melhor para a realização deste trabalho.

RESUMO

Este estudo tem como objetivo apresentar uma proposta do uso do *software* estatístico *Action* em ambientes virtuais de aprendizagem como a Plataforma *Moodle* no processo de ensino e aprendizagem sobre a exploração de dados biológicos. O princípio norteador do presente trabalho é a oportunidade de refletir, analisar e lidar com o ensino de Estatística em ambiente mediados por recursos tecnológicos. Buscamos verificar de que maneira o uso de *softwares* em ambientes virtuais de aprendizagem podem contribuir para o processo de aquisição do conhecimento. Levantamos um referencial teórico focando em especial pesquisadores que abordam mais especificamente a inserção de tecnologias na educação, e em alguns autores que abordam as teorias da aprendizagem. A pesquisa foi desenvolvida tendo como proposta de público alvo os alunos da Universidade Federal do Acre – UFAC, Campus Floresta, na cidade de Cruzeiro do Sul – Acre, em uma disciplina de Bioestatística. A pesquisa resultou em dois produtos educacionais: Produto I – Tutorial de Utilização da Plataforma *Moodle* para o Curso de Bioestatística I e Produto II - Livro Análise Exploratória de Dados Biológicos, com recursos de vídeos tutoriais sobre a utilização do *software Action 2.5*. A pesquisa serviu principalmente para modificar nosso entendimento sobre o processo de aquisição do conhecimento, sobre as ideias que estão sendo apresentadas nos estudos desta área e sobre como podemos modificar nossa postura como professores de uma geração caracterizada pela manipulação tecnológica.

Palavras-chave: *moodle*, *action*, bioestatística.

ABSTRACT

This study aims to present a proposal for the use of statistical software *Action* in virtual learning environments such as *Moodle* Platform in the process of teaching and learning on the exploration of biological data. The guiding principle of this work is the opportunity to reflect, analyze and deal with the Statistics teaching in environment mediated by technological resources. We have seek to verify how the use of *software* in virtual learning environments can contribute to the process of knowledge acquisition. We researched a theoretical framework focusing in particular researchers who address more specifically the inclusion of technology in education, and in some authors who discuss learning theories. The research was developed focusing the proposal public target the students of the Universidade Federal do Acre - UFAC, Campus Floresta, in Cruzeiro do Sul city – Acre, in one of Biostatistics. The Research has resulted in two educational products: Product I - Tutorial of using of the *Moodle* Platform for the *Biostatistics* Course *I* and Product II - Exploratory Analysis Book of Biological Data with resources of tutorial videos on using the *Action 2.5 software*. The research served primarily to modify our understanding of the acquiring knowledge process and, about the ideas that are being presented in the studies in this area and how we can change our attitude as teachers of a generation characterized by technological manipulation.

Keywords: moodle, action, biostatistics.

LISTA DE ILUSTRAÇÕES

Figura 3-1 Propt de comando do programa R.....	30
Figura 4-1 - Pagina inicial da plataforma <i>Moodle</i> UFAC.....	33
Figura 4-2 - Página de cadastro de novo usuário.....	33
Figura 4-3 - Conclusão do cadastro de usuário.....	34
Figura 4-4 - Identificando usuário na plataforma.....	34
Figura 4-5 - Buscando cursos na plataforma.....	35
Figura 4-6 - Acessando a disciplina de Bioestatística I/CZS.....	35
Figura 4-7 - Página do curso de Bioestatística I.....	36
Figura 4-8 - Recursos iniciais do curso.....	36
Figura 4-9 - Forum de apresentação.....	37
Figura 4-10 - Página para acesso ao Chat.....	37
Figura 4-11 - Chat de interação da turma.....	38
Figura 4-12 - Pré-teste de avaliação do curso.....	38
Figura 4-13 - Diário de Bordo.....	39
Figura 4-14 - Glossário de Biologia e Estatística.....	39
Figura 4-15 - Inserindo termos no glossário.....	40
Figura 4-16 - Wiki de Bioestatística.....	40
Figura 4-17 - Fórum de Dúvidas.....	41
Figura 4-18 - Inserindo perguntas no forum de dúvidas.....	42
Figura 4-19 - Página do <i>Youtube</i> com vídeo tutorial.....	43
Figura 4-20 - Atividade 1.....	44
Figura 4-21 - Atividade para envio de arquivos.....	45
Figura 4-22 - Envio de arquivos.....	45
Figura 4-23 - Selecionando o arquivo para envio.....	46
Figura 4-24 - Link para participantes.....	47
Figura 4-25 - Links de navegação.....	48
Figura 4-26 - Notas do usuário.....	49
Figura 4-27 - Pesquisar nos fóruns.....	49
Figura 4-28 – Mensagens.....	50
Figura-5-1 - Vasos sanguíneos do corpo humano.....	57
Figura-5-2 - Microscopia eletrônica mostrando as hemácias (em vermelho) e um glóbulo branco (em branco).....	58

Figura-5-3 - Peixes, algas, plantas e seres microscópios do mar.....	58
Figura 5-4 - Diagrama da Estatística.....	59
Figura -5-5 Tubo de ensaio com amostra sanguínea.....	59
Figura -5-6 Biólogo coletando espécies de cobras para análise.....	59
Figura-5-7 Coleta de amostra de nadadeira de peixes	60
Figura 5-8 - Diagrama das variáveis estatísticas.....	61

SUMÁRIO

1. INTRODUÇÃO	14
1.1 A Relação com a Pesquisa.....	15
1.2 Fundamentação Teórica	16
1.2.1 A Contribuição das Teorias da Aprendizagem para a Construção de Ambientes Informatizados	16
1.2.2 O Uso das Tecnologias de Informação e Comunicação no Ensino de Matemática ...	17
1.2.3 O Ambiente Virtual de Aprendizagem Plataforma <i>Moodle</i>	19
1.3 Revisão Bibliográfica.....	22
1.4 Justificativa da pesquisa	24
1.5 Questão de investigação	25
1.6 Hipótese da Pesquisa	25
2. OBJETIVOS	26
2.1 Objetivo Geral	26
2.2 Objetivos Específicos.....	26
3. METODOLOGIA.....	27
3.1 Sujeitos da Pesquisa	27
3.2 Questionário Sobre o Uso das Tecnologias	27
3.3 Implementação da disciplina no <i>Moodle</i>	27
3.4 Introdução à Análise Exploratória dos Dados.....	27
3.5 O <i>Software</i> Estatístico <i>Action</i>.....	29

4. PRODUTO 1 - TUTORIAL DE UTILIZAÇÃO DA PLATAFORMA MOODLE PARA O CURSO DE BIOESTATÍSTICA I	32
4.1 Introdução	32
4.2 Realizando o cadastro na plataforma Moodle Ufac.....	32
4.3 Acessando a disciplina Bioestatística I/CZS.....	34
4.4 Explorando os recursos introdutórios da disciplina.....	35
4.4.1 Fórum de Notícias	36
4.4.2 Fórum de Apresentação	37
4.4.3 Chat Interação da Turma – Aberto 24h.....	37
4.4.4 Pré-teste de Avaliação do Curso	38
4.4.5 Plano de Curso de Bioestatística I:.....	38
4.4.6 Diário de bordo:.....	39
4.4.7 Glossário de Biologia e Estatística:.....	39
4.4.8 Wiki:.....	40
4.5 Tópicos de Conteúdos.....	41
4.5.1 Fórum de Dúvidas	41
4.5.2 Livro – Análise Exploratória de Dados Biológicos.....	42
4.5.3 Tabelas do Excel para download	43
4.5.4 Atividades.....	44
4.5.5 Vídeos tutoriais na plataforma	46
4.5.6 Arquivo de instalação do Action 2.5	46
4.6 Outros recursos da plataforma Moodle UFAC	47
4.6.1 Participantes	47
4.6.2 Navegação	47
4.6.3 Administração do Curso - Notas	48
4.6.4 Pesquisar nos Fóruns	49
4.6.5 Últimas Notícias	49
4.6.6 Mensagens	50
4.6.7 Próximos Eventos.....	50
4.6.8 Atividades Recentes	50

4.7	Considerações Finais	50
5.	PRODUTO II – LIVRO ANÁLISE EXPLORATÓRIA DE DADOS	
	BIOLÓGICOS	51
5.1	INTRODUÇÃO	54
5.2	VISÃO GERAL DA ANÁLISE ESTATÍSTICA	56
5.2.1	Estatística Descritiva	56
5.2.2	Estatísticas Inferenciais	56
5.3	VOCABULÁRIO BÁSICO ESTATÍSTICO	57
5.3.1	População	57
5.3.2	Amostra	58
5.3.3	Variável	60
5.3.4	Tipos de Variáveis.....	60
5.3.5	Exercício Proposto	61
5.4	PLANEJAMENTO DA COLETA DE DADOS	62
5.4.1	Técnicas de Amostragem	62
5.4.2	Obtenção de Amostras.....	62
5.4.3	Amostra Aleatória ou Probabilística	63
5.4.4	Amostra Aleatória Simples	63
5.4.5	Amostra Aleatória Estratificada	65
5.5	ORGANIZAÇÃO DOS DADOS	67
5.5.1	Dados Brutos	67
5.5.2	Rol	67
5.5.3	Tabelas de Distribuição de Frequência	69
5.5.4	Interpretação da Tabela de Distribuição de Frequência Gerada pelo <i>Action</i>	74
5.5.5	Tabelas de Distribuição de Frequência com Intervalos de Classe	77
5.5.6	Interpretação da tabela de distribuição de frequência com intervalos de classe gerada pelo <i>Action</i>	80
5.6	GRÁFICOS	83
5.6.1	Gráficos para variáveis qualitativas	83

5.6.2	Gráficos para Variáveis Quantitativas.....	87
5.6.3	Histograma	91
5.7	MEDIDAS ESTATÍSTICAS	95
5.7.1	Medidas de Tendência Central	95
5.7.2	Medidas de Dispersão	98
5.7.3	Cálculo de medidas estatísticas utilizando o <i>software Action</i>	101
5.7.4	Quartil.....	110
5.7.5	<i>Box plot</i> (Gráfico de Caixa).....	111
5.8	NOÇÕES DE CORRELAÇÃO	120
5.8.1	Diagrama de dispersão	120
5.8.2	Coefficiente de Correlação	123
5.9	Referências	134
6.	CONSIDERAÇÕES FINAIS.....	135
7.	REFERÊNCIAS	137

1. INTRODUÇÃO

Atualmente, tem-se pensado bastante em meios de tornar o conhecimento estatístico mais significativo para o aluno, de modo a colaborar com a sua formação em termos de saberes, habilidades e competências, e deste modo, contribuir para o desenvolvimento integral do aluno. Neste viés, é importante que o professor que ensina estatística reflita sobre o seu exercício profissional, as suas concepções, crenças e representações sobre o ensino e aprendizagem. O desafio de ensinar estatística nos dias atuais, agregando seus conhecimentos às tecnologias de informação e comunicação (TIC's), que estão cada vez mais integradas ao cotidiano dos alunos por meio dos *smartphones*, *tablets*, *notebooks* e outras ferramentas é uma questão que aflige muitos professores desta área.

Meio as pesquisas que têm como foco o ensino e a aprendizagem, parece que as tecnologias podem contribuir para amenizar as dificuldades de interação dos alunos nas aulas de Estatística, tornando o trabalho do professor supostamente mais dinâmico. Nessa perspectiva, os suportes tecnológicos como os computadores, os softwares e a Internet serão apresentados neste trabalho como ferramentas de apoio ao ensino presencial, aliando-se na tentativa de superar alguns problemas presentes nas práticas de ensino.

Existem vários esforços para o uso eficiente das TIC's na educação, como por exemplo, o Programa Nacional de Tecnologias Educacionais (Proinfo), que é um programa educacional com o objetivo de promover o uso pedagógico da informática na rede pública de educação básica. Valente (1997) defende que o uso inteligente do computador na educação é justamente aquele que tenta provocar mudanças na abordagem pedagógica vigente ao invés de colaborar com o professor para tornar mais eficiente o processo de transmissão de conhecimento.

Nos dias atuais existem sistemas para apoiar cursos utilizando a internet, são os sistemas gerenciadores de cursos, que na literatura podem receber diferentes denominações e que por simplicidade, chamaremos de Ambientes Virtuais de Aprendizagem (AVA). Esses ambientes são utilizados na Educação à Distância (EaD) que é principalmente caracterizada pela separação física entre professor e aluno por meio de comunicações em tempo real ou por meio de mensagens que podem ser acessadas posteriormente. A internet pode permitir a interatividade entre os participantes, por isso, os ambientes virtuais de aprendizagem se destacam entre as demais mídias utilizadas na EaD. Supõe-se que o ensino a distância via internet possa possibilitar que o ensino seja personalizado, com a construção de cursos com diferentes níveis, e a navegação pelo aluno é realizada segundo seu ritmo de aprendizagem pessoal e cabível aos horários mais compatíveis. Aliando-se a isso, existe a flexibilidade dos

materiais didáticos produzidos no formato digital, que permite uma constante atualização de dados e informações.

Este trabalho visa analisar o uso de técnicas de Introdução à Análise Exploratória de Dados organizadas em um livro digital com exemplos reais de pesquisas já realizadas por pesquisadores da Universidade Federal do Acre, apresentados em um ambiente virtual de aprendizagem, com o auxílio de um software estatístico, como ferramentas de apoio ao ensino, especialmente de Bioestatística para alunos da Universidade Federal do Acre – UFAC, Campus Floresta.

1.1 A Relação com a Pesquisa

As concepções que temos hoje sobre ensino e aprendizagem são influenciadas pelas experiências que tivemos em nossa vida como aluno e professor. Essas experiências moldam nossa personalidade e traduzem nossos interesses e perspectivas. Para que possamos entender o que influenciou meu interesse pela pesquisa passo a relatar um pouco de minha história.

Como aluno do ensino fundamental, sempre vi a matemática como uma disciplina muito difícil, suas “mágicas” e definições sempre eram expostas de uma maneira que a meu ver, passavam a ideia de que a matemática não era acessível a todas as pessoas. Assim, esta forma de ver a matemática, contribuía para aumentar ainda mais minhas dificuldades em compreender o que era ensinado.

No ensino médio as dificuldades continuaram, pois muitos conteúdos que não compreendi anteriormente eram requisitos para a compreensão dos novos conteúdos. Terminei o ensino médio em 1995, mas só tive a oportunidade de prestar vestibular em minha cidade em 2000, no Programa Especial de Formação de Professores para a Educação Básica, oferecido pela Universidade Federal do Acre.

Quando tive a oportunidade de cursar uma faculdade de licenciatura em matemática e paralelamente ministrar aulas de matemática para alunos do ensino médio, procurei fazer diferente daquilo que tinha vivido. Daí então, trabalhei de forma a tentar despertar nos alunos o gosto pela matemática, utilizando recursos tecnológicos quando possíveis, materiais concretos, experimentos e outros recursos que poderiam melhorar o desempenho dos alunos na disciplina.

Em 2008 iniciei o curso de Licenciatura em Música pela Universidade de Brasília – UnB, onde tive o primeiro contato com o Ambiente Virtual de Aprendizagem Plataforma *Moodle*, o curso era semipresencial, tínhamos aulas com o tutor no começo e fim das

disciplinas, os demais momentos eram todos na plataforma, com isso pude perceber as deficiências e principalmente as potencialidades do ensino a distância.

Durante os dez anos que ministrei aulas em escolas públicas, participei de alguns cursos de formação continuada de professores promovidos pela Secretaria Estadual de Educação – SEE, neles os formadores traziam soluções para os problemas enfrentados em sala de aula de forma pronta, muitas vezes sem agregar o conhecimento prático dos professores ao conhecimento acadêmico apresentado, promovendo com isso um distanciamento entre a teoria e a prática. Paralelo a isso, se constatava ainda, muitos alunos com dificuldades na aprendizagem. As coordenações pedagógicas apresentavam algumas alternativas como aulas de reforço em horários diferentes, recuperações paralelas de conteúdos, aprendizagem colaborativa com a organização de grupos de estudos, entre outras atividades de apoio, na busca de resolver tais problemas. No entanto, a maioria dos alunos continuava com baixo rendimento na disciplina. Essas inquietações com a prática profissional é que me impulsionaram na pesquisa em educação.

1.2 Fundamentação Teórica

1.2.1 A Contribuição das Teorias da Aprendizagem para a Construção de Ambientes Informatizados

Refletimos neste trabalho sobre como as teorias da aprendizagem podem contribuir significativamente no processo de construção de propostas de ensino utilizando as TIC's, buscando emergir a relação tão desejada com a prática docente. Essas teorias são construções humanas que buscam interpretar, sistematizar e explicar como o indivíduo aprende.

A pouca prática articulada com a insuficiente fundamentação teórica sobre a construção de Ambientes Informatizados e/ou as concepções dos professores que a utilizam, podem fazer com que as ferramentas computacionais sirvam apenas como meros transmissores de informação, limitando os alunos frente às possíveis potencialidades dessas tecnologias.

Verifica-se que a maioria dos cursos de licenciatura em matemática e estatística no Brasil contemplam disciplinas que abordam as teorias da aprendizagem, com o objetivo de proporcionar aos futuros professores, conhecimentos sobre essas teorias, possibilitando aos mesmos uma visão crítica sobre as concepções de pessoas, sociedades, conhecimento, educação, escola e relação professor-aluno. Com tudo, há um indicativo de que alguns professores não valorizam essas teorias, um dos motivos desta desvalorização pode estar na própria concepção de ensino destes professores, que não conseguem perceber a conexão entre estas teorias e a realidade da sala de aula.

Fiorentini (1995) defende que o professor que concebe a Matemática como uma ciência exata, logicamente organizada e a-histórica ou pronta e acabada, certamente terá uma prática pedagógica diferente daquele que a concebe como uma ciência viva, dinâmica e historicamente sendo construída pelos homens, atendendo a determinados interesses e necessidades sociais.

Da mesma forma, o professor que acredita que o aluno aprende Matemática através de memorização de fatos, regras ou princípios transmitidos pelo professor ou pela repetição exaustiva de exercícios, também terá uma prática diferenciada daquele que entende que o aluno aprende construindo os conceitos a partir de ações reflexivas sobre materiais e atividades, ou a partir de situações-problema e problematizações do saber matemático (FIORENTINI 1995).

Esses distanciamentos entre as teorias da aprendizagem e as práticas de ensino de Estatística nos fazem refletir sobre como essas teorias podem contribuir na construção de propostas pedagógicas que valorizem a aprendizagem significativa, tecendo saberes entre a utilização de TIC's e o ensino de Estatística.

1.2.2 O Uso das Tecnologias de Informação e Comunicação no Ensino de Matemática

Um dos objetos de investigação em Educação Matemática na atualidade é o estudo das relações que envolvem o uso das tecnologias de informação e comunicação (TIC) no ensino e aprendizagem de Matemática. Sequências didáticas que utilizam recursos computacionais são cada vez mais comuns em trabalhos de pesquisa nesta área.

Segundo Borba e Penteadó (2003, p. 8), “a informática se constituiu em uma das principais tendências da Educação Matemática”. Neste mesmo sentido Miranda e Laudares (2007, p. 73) afirmam que:

A Educação Matemática não se realiza sem a Educação Tecnológica. A escola é uma instituição social. Daí “sofre” intervenção do contexto no qual se insere e, se a sociedade moderna está definida e estruturada pela tecnologia, todas as suas instituições passam a ser delineadas com parâmetros tecnológicos.

Seguindo por esta trilha, Borba (2012) argumenta que, embora existam pesquisas sobre o uso de softwares em Educação Matemática há mais de 30 anos, este conjunto de pesquisas não necessariamente resultou na incorporação da tecnologia computacional nas salas de aula de Matemática.

Há uma tendência que aponta que a utilização de recursos tecnológicos como ferramentas de apoio ao ensino de matemática pode contribuir no desenvolvimento de

competências e habilidades necessárias a aprendizagem, como no uso de *softwares* de Geometria Dinâmica (LOUREIRO, 2001; ALVES; SOARES, 2003), de aspectos lúdicos no ensino de conteúdos matemáticos com a utilização de jogos eletrônicos (ROSA, 2004), do uso de plataformas para comunicação em cursos *on-line* a distância, com ênfase nas discussões matemáticas sobre geometria ou em discussões sobre as tendências em Educação Matemática (ZULATTO, 2006; GRACIAS, 2003), dentre outros. Com isso, deve-se pensar que a informática pode abrir novas possibilidades para a Educação, mas não basta só ensinar Matemática e disponibilizar computadores na sala de aula, mas sim compor um movimento de repensar a Educação (DAWBOR, 1996).

Neste sentido, Os Parâmetros Curriculares Nacionais do Ensino Médio (2000, p. 5) citam que:

A denominada “revolução informática” promove mudanças radicais na área do conhecimento, que passa a ocupar um lugar central nos processos de desenvolvimento, em geral. É possível afirmar que, nas próximas décadas, a educação vá se transformar mais rapidamente do que em muitas outras, em função de uma nova compreensão teórica sobre o papel da escola, estimulada pela incorporação das novas tecnologias.

No entanto, alguns cuidados devem ser tomado com relação a inserção do computador nas aulas de matemática. Para Silva (2001, p. 32),

“a introdução desses equipamentos sem uma prévia reflexão sobre o porquê, como e para que utiliza-los provoca o abandono do computador, pois o seu uso depende do contexto no qual se opera, da capacidade criativa do professor, do software disponível e, sobretudo, dos objetivos que se desejam alcançar.”

Segundo Valente (1998):

A verdadeira função do aparato educacional não deve ser a de ensinar, mas sim a de criar condições de aprendizagem. Isto significa que o professor deve deixar de ser o repassador do conhecimento e passar a ser o criador de ambientes de aprendizagem e o facilitador do processo de desenvolvimento intelectual do aluno.

É de fundamental importância que o professor conheça e avalie cuidadosamente a ferramenta tecnológica que pretende utilizar, sobre isso, Freire *et al.* (1998) nos diz que:

É natural que o professor, no início do seu trabalho na área de informática na educação, eleja algumas ferramentas computacionais, para ele mais adequadas e/ou mais amigáveis. Para alguns professores, o aprendizado de algumas ferramentas pode parecer custoso, a princípio. Aos poucos começa a compreender o funcionamento da mesma e atinge um momento de estabilidade que lhe permite usa-la com crescente desembaraço. Esse momento é absolutamente necessário, pois não se pode aprender tudo de uma vez e usar tudo com os

alunos sem uma avaliação cuidadosa. O aprimoramento atingido pelo professor em relação a sua prática pedagógica permite que ele avalie a pertinência do uso de outros recursos tecnológicos que colaborem para a consolidação do seu trabalho com os alunos.

Neste trabalho refletimos sobre as possibilidades de inserir as tecnologias de informação e comunicação na Educação Estatística de forma intencional, possibilitando que professores e alunos explorem suas potencialidades.

1.2.3 O Ambiente Virtual de Aprendizagem Plataforma Moodle

As teorias da aprendizagem deram o suporte pedagógico para muitas ferramentas tecnológicas utilizadas atualmente na educação. Um exemplo disso são os ambientes virtuais de aprendizagem, também conhecidos como AVA's. Nestes ambientes, a interação entre aluno e professor é bastante valorizada como também a interação entre aluno e aluno. Ferramentas como fóruns, *chat* e *wikis* contribuem significativamente para uma interação dinâmica do conhecimento. Isso nos lembra das ideias de Vygotsky sobre a interação social, que considera esta interação fundamental para o desenvolvimento cognitivo e linguístico de qualquer indivíduo (MOREIRA, 2009, p. 20).

O Ambiente Virtual de Aprendizagem Plataforma Moodle – objeto de nossa pesquisa – é um *software* com uma metodologia pedagógica baseada na teoria construtivista social.

Sobre isto, Rosa (2013) nos diz que:

Esse sistema foi desenvolvido, em 1999, pelo educador e cientista computacional Martin Dougiamas, na *Curtin University of Technology*, em *Perth*, na Austrália, como resultado de sua tese de doutorado intitulada *The Use of Open Source Software to Support a Social Constructionist Epistemology of Teaching and Learning within Internet-based Communities of Reflective Inquiry*, que examinou a utilização de softwares livres para apoiar a epistemologia social construtivista para o ensino e a aprendizagem de comunidades on-line, por meio do desenvolvimento de atividades reflexivas. Assim, Dougiamas (1999) criou a plataforma Moodle com o intuito de fomentar um ambiente de colaboração no qual os usuários podem intercambiar saberes ao experimentar e criar novas interfaces para serem utilizadas em comunidades abertas.

O termo construtivismo está relacionado à teoria do desenvolvimento cognitivo elaborada por Piaget, o qual tem como princípio básico que o conhecimento se constrói a partir das ações do sujeito.

Os trabalhos publicados por Piaget e Vygotsky dentre outros, são exemplos de paradigma integracionista, dentro da filosofia construtivista, pois partem do pressuposto que o indivíduo em interação com o meio externo deve ser construtor do seu próprio conhecimento por meio de experiências autênticas. Piaget enfatiza em seus trabalhos os fatores internos,

individuais e genéticos, na gênese da lógica dos conceitos e na sua explicação cognitiva. Enquanto, Vygotsky focaliza os fatores externos, sociais e adquiridos enfatizando o meio social como formador das funções psicológicas (RESENDE 2009).

Nessa visão educacional, há a necessidade da realização de atividades significativas, autênticas e contextualizadas, com base também na abordagem sócio-histórica de Vygotsky (1984), que afirma que os indivíduos são seres sociais inseridos em um ambiente definido historicamente e culturalmente.

Outro elemento citado por Vygotsky que pode nos auxiliar na construção de nossa proposta didática para o ensino e aprendizagem de Estatística é a Zona de Desenvolvimento Proximal (ZDP). A zona de desenvolvimento proximal é definida por Vygotsky como a distância entre o nível de desenvolvimento cognitivo real do indivíduo, tal como medido por sua capacidade de resolver problemas independentemente, e o seu nível de desenvolvimento potencial, tal como medido através da solução de problemas sob orientação de alguém (um adulto, no caso de uma criança) ou em colaboração com companheiros mais capazes (Vygotsky, 1984, p. 97). A zona de desenvolvimento proximal define as funções que ainda não amadureceram, mas que estão no processo de maturação. É uma medida do potencial de aprendizagem; representa a região na qual o desenvolvimento cognitivo ocorre; é dinâmica e está constantemente mudando (MOREIRA, 2009, p. 21).

Quando percebemos o que o nosso aluno já sabe e o que ele pode aprender com a ajuda de um parceiro mais experiente, visualizamos com mais nitidez quais competências e habilidade podem ser potencializadas, quais atividades podem ser sugeridas, quais conteúdos podem ser aprendidos com mais facilidade. A resolução de problemas é um dos mecanismos que podemos utilizar para medir o potencial de aprendizagem do desenvolvimento cognitivo.

Segundo Vygotsky (1984), o desenvolvimento cognitivo dos alunos acontece por meio da "internalização das atividades socialmente enraizadas e historicamente construídas" (p. 64) pelos indivíduos, em um processo colaborativo de ensino e aprendizagem. Rosa (2003) nos diz que a teoria do construtivismo afirma que os alunos constroem ativamente o conhecimento ao mesmo tempo em que interagem com o ambiente.

A interação social que provoca a aprendizagem deve ocorrer dentro da zona de desenvolvimento proximal, mas, ao mesmo tempo, tem um papel importante na determinação dos limites dessa zona. O limite inferior é, por definição, fixado pelo nível real de desenvolvimento do aprendiz. O superior é determinado por processos instrucionais que podem ocorrer no brincar, no ensino formal ou informal, ou no trabalho. Independentemente do contexto, o importante é a interação social (Driscoll, 1995, p. 233).

Para Vygotsky, o único bom ensino é aquele que está à frente do desenvolvimento cognitivo e o lidera. Analogamente, a única boa aprendizagem é aquela que está avançada em relação ao desenvolvimento. A aprendizagem orientada para níveis de desenvolvimento já alcançados não é efetiva do ponto de vista do desenvolvimento cognitivo do aprendiz (MOREIRA, 2009, p. 22).

Para Dougiamas (1999), o construtivismo social destaca o desenvolvimento individual do significado por meio da comunicação, da construção ativa do conhecimento e do compartilhamento de produtos sociais nas interações com outros indivíduos em ambientes virtuais de aprendizagem. De acordo com Vygotsky (1984), é pela linguagem conversacional utilizada no contexto social que o conhecimento emerge no ambiente de aprendizagem, para que as ideias sejam negociadas e possam adquirir significado.

Diante desse cenário, Valente, Moreira e Dias (2009) concluem:

(...) todas as plataformas de apoio ao e-learning se afirmam promotoras de modelos construtivistas de aprendizagem e favoráveis à criação de comunidades de prática (p. 39).

Neste aspecto, esses autores admitem que:

Os ambientes de aprendizagem considerados construtivistas preconizam que o aluno participe ativamente na resolução de problemas, que utilize o pensamento crítico sobre as atividades de aprendizagem que mais significam para si e que construa o seu próprio conhecimento, cabendo ao professor o papel de (...) orientador, facilitador, conselheiro, tutor e aprendiz (VALENTE, MOREIRA e DIAS, 2009, p. 41).

Não pretendemos encerrar aqui a discussão sobre a inserção da TIC na educação. Precisamos analisar cuidadosamente essa nova compreensão teórica sobre o papel da escola, estimulada pela incorporação das novas tecnologias. Propomos-nos a refletir sobre o modo como ensinamos estatística em ambientes mediados pelos recursos tecnológicos, buscando conhecer e avaliar cuidadosamente estas ferramentas, verificando se a TIC pode realmente abrir novas possibilidades para a Educação e promover mudanças na área do conhecimento.

Embasamos nossa proposta de construção de uma plataforma *on-line* para o ensino de Bioestatística no construtivismo social, por acreditarmos que os alunos constroem ativamente o conhecimento ao mesmo tempo em que interagem com o ambiente, por meio da comunicação, da construção ativa do conhecimento e do compartilhamento de produtos sociais nas interações com outros indivíduos.

Nessa perspectiva é que passaremos a relatar a metodologia da pesquisa, onde buscamos construir atividades significativas e contextualizadas sobre Introdução a Análise Exploratória

de Dados Biológicos no Ambiente Virtual de Aprendizagem Plataforma *Moodle*, com o auxílio do *software* estatístico *Action*.

1.3 Revisão Bibliográfica

Apresentamos a partir de agora um panorama, referente à utilização da plataforma *Moodle* e o ensino de Bioestatística para alunos de graduação. Na seleção dos trabalhos para comporem nossa pesquisa, agrupamos os que o resumo mencionava as palavras: *Moodle*, Ambientes Virtuais de Aprendizagem, Bioestatística e Educação. Foram selecionadas três dissertações (duas de mestrado profissional e uma de mestrado acadêmico). Neste conjunto, destacamos em cada uma, a partir da leitura dos resumos e do texto completo: os objetivos; tipo de metodologia e os principais resultados.

Daroit (2009) aborda o uso de aplicativos computacionais como ferramenta auxiliar no processo de ensino e aprendizagem através da utilização de materiais didáticos próprios elaborados a partir de alguns pressupostos da Teoria da Aprendizagem Significativa de David P. Ausubel (1980 – 2003). O trabalho realizado buscou investigar de que forma esta estratégia didática oportuniza o desenvolvimento de habilidades e competências dos estudantes e contribui na construção de significados e formação de conceitos.

A pesquisa teve como objetivo geral explorar o uso de ferramentas computacionais através de materiais didáticos próprios, buscando verificar se estes possuíam potencial de serem significativos e de que modo contribuiriam para o desenvolvimento de habilidades e competências dos estudantes. Neste sentido, a pesquisadora utilizou a metodologia de definir dois aspectos considerados como indicativos da ocorrência de uma aprendizagem significativa: o aspecto motivacional, relacionado à predisposição do estudante para a aprendizagem e o aspecto cognitivo, relacionado à aprendizagem do conteúdo.

Considerou-se positivo o resultado da utilização de aplicativos computacionais para o desenvolvimento de competências e habilidades, em vista da serie de alternativas estruturais apresentadas pelo software. Entretanto, constatou-se que o domínio da linguagem, uma competência que reflete diretamente na interpretação dos dados e na análise dos resultados obtidos, não pode ser desenvolvida diretamente através da utilização de aplicativos computacionais.

O produto apresentado na dissertação foi uma proposta de ensino que utiliza o software Bioestat 5.0 e o Ambiente Virtual de Aprendizagem (AVA) TelEduc.

Apesar de ter citado o ambiente virtual de aprendizagem TelEduc no desenvolvimento da proposta de ensino, o mesmo só foi recomendado para a utilização dos recursos de correio eletrônico e acesso aos materiais didáticos.

Esta pesquisa contribuiu para dar evidências de que a utilização de softwares estatísticos pode desenvolver nos alunos aspectos motivacionais e cognitivos, se utilizados em propostas de ensino que priorizem a aprendizagem significativa. Verificou-se ainda, a possibilidade de uma contribuição para a ampliação desse campo de pesquisa, no sentido de utilizar essa ferramenta computacional agregada a plataforma *Moodle*, explorando não apenas seus recursos de correio eletrônico e acesso a materiais didáticos, mas também suas ferramentas de interação (fóruns, *chats* e *wikis*) e de suporte à avaliação *on-line*.

Lucas (2009) objetivou a construção de um Ambiente Virtual de Aprendizagem para o ensino significativo de Geometria Analítica, com a utilização de visualizadores 3D idealizados no GeoGebra, permitindo ao aluno visualizar e manipular o espaço tridimensional e seus objetos sob vários “pontos de vista”.

A metodologia aplicada na pesquisa teve como princípio o desenvolvimento de um Ambiente Virtual de Aprendizagem no *Moodle*. O ambiente foi idealizado no sentido de que as atividades fizessem com que os alunos pensassem e refletissem sobre o conhecimento adquirido na visualização geométrica conectada à álgebra.

Segundo o autor, houve uma melhora na participação em sala e uma maior dedicação aos estudos que ocasionou numa melhora de conceitos (notas) e de motivação.

O produto da dissertação foi a construção do Ambiente Virtual de Aprendizagem no *Moodle* sobre conceitos fundamentais de Geometria Analítica.

As contribuições desta pesquisa deram suporte para a utilização de ambientes virtuais de aprendizagem em disciplinas ministradas presencialmente, evidenciando as vantagens da plataforma *Moodle*, no sentido de dar ao aluno a oportunidade de imprimir seu próprio ritmo de estudo, contribuindo para a aquisição de saberes.

Verificou-se que a plataforma *Moodle* pode ser agregada a softwares educacionais, potencializando seu uso e ampliando sua dinâmica de interação entre aluno-professor e entre aluno-aluno.

A pesquisa de **Delgado (2009)** analisa uma experiência desenvolvida na Escola Politécnica da Universidade Federal do Rio de Janeiro, envolvendo o uso da Ferramenta *Moodle* no apoio ao ensino presencial da disciplina Princípios da Ciência dos Materiais do sexto período da graduação do curso de Engenharia.

O problema abordado se baseia em pesquisas recentes que mostram que a utilização de ambientes virtuais de aprendizagem requer uma mudança de comportamento não só dos alunos, mas principalmente dos professores, que serão condutores do processo de mudança.

O objetivo central foi construir uma reflexão em torno das práticas pedagógicas na utilização de um ambiente virtual de aprendizagem, o *Moodle*, como apoio ao ensino presencial de uma disciplina de graduação.

Na apresentação dos dados da pesquisa foi feita uma descrição de como o professor construiu o Ambiente Virtual de Aprendizagem no *Moodle*, quais os recursos utilizados, suas funcionalidades e como foram explorados durante a disciplina.

Os dados apresentados nos levam a considerar que a interação no AVA ficou restrita a relação professor-aluno. A interação aluno-aluno foi praticamente nula. A análise também aponta como possíveis causas, a ausência de atividades em grupo, a limitação do editor de texto, que não possui fórmulas matemáticas ou ainda a quantidade de tarefas, que levava os alunos a focar nos prazos dos exercícios deixando em segundo plano a interação. Nos desafios, a interação também se restringiu a relação professor-aluno, não foi observada na plataforma a interação entre os alunos a cerca do conteúdo. Esse fato pode ter ocorrido por uma limitação da plataforma, do desenho ou mesmo de tempo de dedicação dos alunos, portanto merece aprofundamentos.

Os trabalhos analisados nos fizeram refletir sobre as potencialidades e limitações da utilização das tecnologias de informação e comunicação no ensino, dando-nos uma perspectiva do que vem sendo pesquisado, principalmente sobre ambientes virtuais de aprendizagem e aplicativos computacionais, mostrando-nos algumas alternativas estruturais apresentadas por alguns *softwares*. No entanto, precisamos pensar e refletir um pouco mais sobre o conhecimento adquirido até o momento, de que forma se dará a dinâmica de interação nesses ambientes, a construção de significados e a aquisição de saberes para a formação de conceitos em matemática. Para isso, buscamos no capítulo seguinte um embasamento teórico em pesquisadores que abordam sobre esses temas para fortalecermos nossa proposta de ensino.

1.4 Justificativa da pesquisa

Alguns artigos já foram publicados sobre a utilização da plataforma *moodle* no ensino de matemática, no entanto, não conseguimos informações específicas sobre os resultados reais dessa aplicação no desenvolvimento de competências e habilidades dos conteúdos estatísticos com a utilização de softwares.

Dessa forma, este estudo tem como princípio norteador a oportunidade de refletir, analisar e lidar com o ensino de Estatística, fornecendo dados que justifiquem a utilização da plataforma *Moodle* e do *software* estatístico *Action* na melhoria da qualidade do ensino, possibilitando aos alunos, a utilização de um espaço de aprendizagem dinâmico, que se baseia no construtivismo social para atender suas necessidades.

No entanto, temos a clareza de que para os objetivos serem de fato alcançados é preciso criar mecanismos e fazer com que esse trabalho seja dialógico, cooperativo e dinâmico, visto que o papel do professor é oportunizar aos alunos procedimentos metodológicos que auxiliem no desenvolvimento do raciocínio e na discussão da resolução de situações-problema através de exercícios, que façam com que os alunos estejam capacitados para agir de modo transformador sobre si mesmo e sobre a realidade que os cerca.

1.5 Questão de investigação

Como organizar um curso presencial de Bioestatística em ambientes virtuais de aprendizagem como a plataforma *Moodle* para alunos da graduação, utilizando materiais didáticos com recursos digitais e com a utilização do *software* estatístico *Action*?

1.6 Hipótese da Pesquisa

As tecnologias de informação e comunicação podem contribuir como ferramentas para amenizar algumas dificuldades dos alunos nas aulas de Estatística, tornando o trabalho do professor supostamente mais dinâmico.

2. OBJETIVOS

2.1 Objetivo Geral

Propor o uso do *software* estatístico *Action* em ambientes virtuais de aprendizagem como a plataforma *Moodle* no processo de ensino e aprendizagem sobre a exploração de dados e informações em uma disciplina de bioestatística.

2.2 Objetivos Específicos

- Elaborar um tutorial para uso da plataforma *Moodle*, viabilizando a interação dos alunos com o ambiente virtual;
- Produzir material didático escrito em formato de livro digital sobre os principais conceitos envolvendo Introdução à Análise Exploratória de Dados, com exemplos práticos voltados a área da biologia que mereçam um tratamento estatístico;
- Produzir vídeos tutoriais sobre a utilização das ferramentas do *software* estatístico *Action* nas análises de dados biológicos.

3. METODOLOGIA

A proposta de uso do software estatístico *Action* no ambiente virtual de aprendizagem plataforma *Moodle* foi desenvolvida seguindo algumas etapas definidas, organizadas com o objetivo de minimizar os problemas advindos de imprevistos que poderiam surgir no decorrer do trabalho.

3.1 Sujeitos da Pesquisa

Os produtos deste trabalho foram elaborados com o propósito de serem aplicados na disciplina de Bioestatística I, para turmas dos cursos de Licenciatura e Bacharelado em Ciências Biológicas da Universidade Federal do Acre – UFAC – Campus Floresta, na cidade de Cruzeiro do Sul – Acre.

3.2 Questionário Sobre o Uso das Tecnologias

Inicialmente, será aplicado um questionário para os alunos matriculados na disciplina, sobre o uso de computadores e internet, para verificar quais serão as potencialidades e dificuldades na inserção da Plataforma *Moodle* e do *software Action* como ferramentas auxiliares nas aulas de Bioestatística. Nas análises serão verificados se os alunos tem familiaridade com computadores e acesso a internet em casa ou na universidade. Isso contribuirá para que as propostas possam ser utilizadas sem nenhuma restrição para os alunos.

3.3 Organização da disciplina no Moodle

A Universidade Federal do Acre (UFAC) gerencia um provedor do *Moodle*, o *Moodle UFAC*. O primeiro passo para a organização da disciplina, consiste em buscar a autorização necessária junto ao órgão responsável para a utilização da plataforma, que no caso da UFAC fica a cargo do Núcleo de Interiorização e Educação a Distância - NIEAD. Logo após a autorização ser concedida, foi disponibilizada uma página na plataforma onde o professor pode estruturar o curso da forma que achar convenientes.

A estrutura da plataforma foi pensada de forma a criar um ambiente dinâmico com interação contínua entre os participantes.

3.4 Introdução à Análise Exploratória dos Dados

A estatística se desenvolveu na história da humanidade através da resolução de problemas de ordem prática. Por isso, a experimentação concreta é essencial para que o aluno desenvolva competências e habilidade para seus trabalhos futuros em diferentes áreas de sua

atuação profissional e social. Dessa forma, a resolução de problemas pode ser um recurso eficiente para o ensino de estatística.

O Produto II – Análise Exploratória de Dados Biológicos, baseia-se na filosofia da Análise Exploratória de Dados, que foi introduzida inicialmente por John Wilder Tukey em 1977, no seu livro: *Exploratory Data Analysis*.

Também chamada de Estatística Descritiva, a Análise Exploratória de Dados constitui o que a maioria das pessoas entende por Estatística, e sem perceber usam no seu dia a dia. Consiste em coletar, resumir e organizar os dados em tabelas, gráficos ou medidas numéricas, e a partir daí procurar alguma regularidade ou padrão nestas observações. Com essa interpretação inicial pode-se identificar se os dados seguem algum modelo conhecido, que permita a chegada de uma conclusão sobre os fenômenos estudados.

Para Batanero, Estepa e Godino (1991), o trabalho com a filosofia da Análise Exploratória de Dados, precisa:

- ✓ possibilitar a geração de situações de aprendizagem contextualizadas em temas que sejam de interesse para o aluno;
- ✓ lançar mão de um forte apoio as representações gráficas (que facilitam a percepção da variabilidade no conjunto de dados observados);
- ✓ empregar, preferencialmente, as estatísticas de ordem (que aportam maior facilidade na atribuição de significado pelo aluno);
- ✓ utilizar diferentes escalas de categorização das variáveis para o estudo dos dados observados;
- ✓ não utilizar (por não ser necessário) ferramentas matemáticas complexas.

Para a introdução dos conteúdos específicos, Gal (2002) aponta cinco elementos que são assumidos como conhecimentos estatísticos de base:

- ✓ conhecimento da razão da necessidade dos dados e como estes podem ser produzidos;
- ✓ familiaridade com termos e ideias básicas da estatística descritiva;
- ✓ familiaridade com termos e ideias básicas da representação gráfica e tabular;
- ✓ compreensão de noções básicas de probabilidade;
- ✓ conhecimento de métodos de obtenção de inferências ou conclusões estatísticas.

Com base nestas ideias, introduzimos a filosofia de Análise Exploratória de Dados, abordando os conteúdos descritos no Plano de Curso de Bioestatística I, apresentados no Anexo II.

Os conteúdos serão apresentados de forma que os alunos percebam a problematização envolvida na coleta de informações, que eles mesmos deverão realizar, dando significado aos resultados obtidos com os cálculos estatísticos, objetivando a análise dos mesmos. Com isso, os alunos poderão entender corretamente o significado dos valores encontrados.

3.5 O *Software Estatístico Action*

As tecnologias de informação e comunicação foram fundamentais para os avanços nos estudos estatísticos, o uso de computadores possibilitou análises que até então seriam praticamente impossíveis sem eles. Atualmente algumas universidades do Brasil – como a Universidade Federal de Lavras (UFLA) – já utilizam softwares como o SISVAR¹ para o ensino de estatística nos cursos de graduação e pós-graduação. Essas ferramentas tecnológicas mudam de certa forma o significado da estatística por que introduzem novas representações e modificam a forma com que a utilizamos. Tais mudanças implicam na necessidade de uma revisão das atividades de ensino. No entanto, ainda existe uma forte tendência para que os professores valorizem sobretudo o cálculo ao invés da análise dos dados e sua interpretação. Para Batanero (2002), se os professores querem acompanhar, de algum modo, a evolução da educação estatística tem de aceitar a rápida mudança tecnológica que permitirá libertá-los dos cálculos enfadonhos, rotineiros e descontextualizados.

Existem diversos pacotes estatísticos, que são programas especificamente desenvolvidos para efetuar análises estatísticas. Neste trabalho, utilizaremos o *Action*. O motivo pelo qual escolhemos o *Action* como ferramenta de apoio nas análises estatísticas foi pela sua possibilidade de utilização integrada ao *Microsoft Office Excel* que é um *software* bastante conhecido entre os alunos, muitos já o utilizam, mesmo sem explorarem seus recursos avançados de cálculos.

O *Action*² é um *software* de estatística desenvolvido sob plataforma R. A linguagem R é um ambiente para computação estatística e gráfica, R é uma série integrada de instalações de *softwares* para manipulação de dados, cálculo e exibição gráfica.

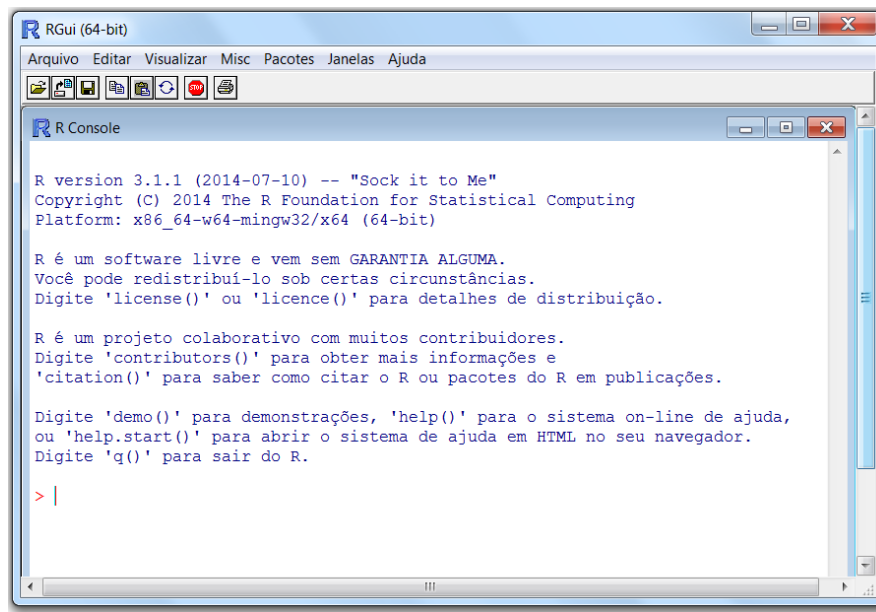
Segundo seu sitio oficial³ o R (Figura 4-17) começou a ser desenvolvido por Robert Gentleman e Ross Ihaka do Departamento de Estatística da Universidade de Auckland em Nova Zelândia, mais conhecidos por “R & R”, apelido do qual se originou o nome R do

¹ <http://www.dex.ufla.br/~danielff/software.htm>

² <http://www.portalaction.com.br/>

³ <http://www.r-project.org>

programa. O objetivo inicial de “R & R”, em 1991, era produzir um *software* para as suas aulas de laboratório baseado na já revolucionária linguagem S, utilizada pelo *software* comercial S-Plus criado por Jonh M. Chambers no *Bell Laboratories* (antiga AT & T, agora *Lucent Technologies*) que atualmente vem contribuindo para o aperfeiçoamento e ampliação das análises estatísticas do R.



```

RGui (64-bit)
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda

R Console

R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> |

```

Figura 3-1 Propt de comando do programa R

Desde então o R vem ganhando cada vez mais adeptos em todo o mundo, em parte devido ao fato de ser totalmente gratuito e também por ser um programa que exige do usuário o conhecimento das análises que está fazendo, diminuindo assim as chances de uma interpretação errada dos resultados.

No entanto, o R apresenta algumas desvantagens com relação a sua utilização por leigos em programação, pois seus usuários precisam conhecer seus comandos para sua perfeita utilização, os comandos são realizados por meio de operações e de funções que precisam ser bem definidas para que o programa execute a ação corretamente.

O *Action* foi criado por uma consultoria estatística e em qualidade, situada em São Carlos - SP, a *Estatcamp*⁴, que o desenvolveu sobre plataforma R, mas com uma *interface* de simples acesso que qualquer pessoa com o mínimo de conhecimento em informática pode opera-lo. Por isso, o *software Action* parece ser mais adequado para os alunos que queiram apenas realizar análises estatísticas para sua atuação profissional, pois com apenas alguns

⁴ <http://www.estatcamp.com.br/empresa/>

cliques os mesmos podem realizar operações complexas de cálculos estatísticos e construção de gráficos que precisariam de várias linhas de comandos se fossem realizados no R.

Segundo Maia (2011), independente do programa, a utilização de *softwares* estatísticos pode ser resumida em três etapas: entradas de dados; execução da análise estatística; interpretação dos resultados.

[...] A Entrada de Dados deve assumir certas convenções, que podem variar de um programa para outro, e tem como efeito a criação de uma matriz, em que, normalmente, cada linha corresponde a uma unidade experimental (elemento da população ou amostra), e cada coluna a uma variável.

A Execução da Análise Estatística pressupõe o conhecimento de como o programa que está sendo utilizado trabalha as informações: as funcionalidades disponíveis, a linguagem de programação, as convenções, entre outros.

Após a execução da análise estatística, chega-se à etapa de Interpretação de Resultados. O detalhe importante é: o computador não interpreta os resultados.[...] (MAIA, 2011, pg. 3 à 5)

Neste trabalho, o *software* estatístico *Action* servirá como ferramenta de otimização do tempo do aluno, dando condições para que ele possa refletir sobre os resultados obtidos, sem precisar utilizar o tempo escolar em cálculos desnecessários.

As aulas de introdução ao *software* serão realizadas no laboratório de informática da Universidade Federal do Acre – UFAC, onde os alunos poderão realizar os mesmos cálculos feitos manualmente com o auxílio do computador. Com isso, espera-se mostrar para os alunos as vantagens da utilização das ferramentas computacionais nos cálculos estatísticos.

4. PRODUTO 1 - TUTORIAL DE UTILIZAÇÃO DA PLATAFORMA MOODLE PARA O CURSO DE BIOESTATÍSTICA I

4.1 Introdução

A plataforma *Moodle* é um ambiente virtual de aprendizagem que possibilita a alunos e professores uma interação virtual para a troca de saberes. M.O.O.D.L.E. é o acrônimo de Modular Object-Oriented Dynamic Learning Environment, que significa *Ambiente Modular de Aprendizagem Dinâmica Orientada a Objetos*. É um *software* livre com código aberto, ou seja, qualquer pessoa com algum conhecimento de programação pode alterar sua estrutura, dessa forma, o *Moodle* permite o acréscimo de novas funcionalidades.

Atualmente o *Moodle* tem uma das maiores bases de usuários do mundo. Muitas universidades estruturam seus cursos exclusivamente pelo *Moodle*, pois ele permite a criação de cursos *on-line*, ou suporte *on-line* de cursos presenciais.

Este tutorial tem o objetivo de instruir os alunos a navegar pelos recursos da plataforma *Moodle* na disciplina de Bioestatística I.

4.2 Realizando o cadastro na plataforma Moodle UFAC

Para começar a acessar a plataforma *Moodle* UFAC deve-se realizar o cadastro de usuário, para isso, digite na barra de endereço do seu navegador de internet o link: uab.ufac.br/moodle.

Será aberta a página de acesso ao *Moodle* UFAC conforme figura 4-1, onde o aluno deverá realizar seu cadastro na plataforma. O objetivo do cadastro é restringir o acesso aos materiais apenas para os alunos participantes, pois como serão realizadas atividades avaliativas o professor deve identificar os alunos matriculados em sua disciplina.

Na página do *moodle* UFAC deve-se clicar em **Criar uma conta** no canto esquerdo da página, conforme aponta a seta em vermelho. Em seguida, abrirá uma nova página com os campos obrigatórios para serem preenchidos com os dados dos participantes da disciplina a que se deseja cursar, conforme figura 4-2.

Após o preenchimento dos campos obrigatórios clique em **Cadastras este novo usuário** conforme figura 4.3.

Depois da realização do cadastro, o **Administrador Moodle** enviará um *e-mail* automático para o endereço cadastrado, o novo usuário deverá acessar sua conta de e-mail e

abrir a mensagem com o nome **Moodle UFAC: confirmação de cadastramento de novo usuário**.

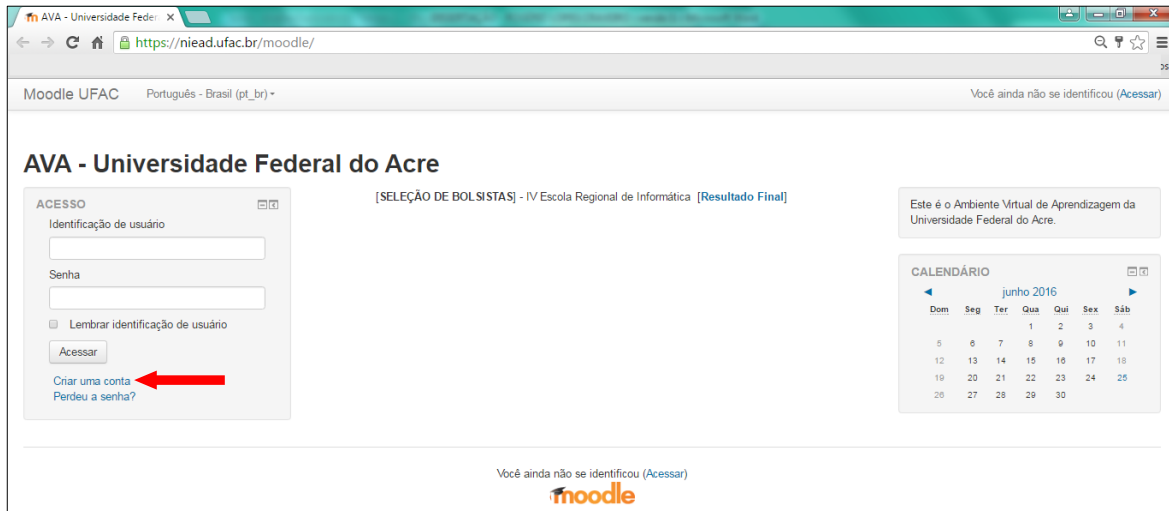


Figura 4-1 - Página inicial da plataforma Moodle Ufac

Figura 4-2 - Página de cadastro de novo usuário

Para confirmar o cadastro deve-se clicar no link disponível, isso o direcionará para a plataforma *Moodle UFAC*, onde já será possível visualizar no canto superior direito da página o nome do usuário cadastrado, isso indica que o cadastro foi realizado com sucesso.

Após o cadastro confirmado, o professor deve inscrever o aluno na disciplina para que o mesmo tenha acesso aos materiais disponíveis no curso.

Figura 4-3 - Conclusão do cadastro de usuário

4.3 Acessando a disciplina Bioestatística I/CZS

Para acessar a disciplina de Bioestatística I/CZS na plataforma *Moodle* UFAC, após o professor responsável já ter confirmado a inscrição do usuário na mesma, execute os seguintes procedimentos:

1. Acesse a plataforma *Moodle* UFAC através do link: uab.ufac.br/moodle/, digite sua identificação de usuário e senha nos campos específicos, após clique em **Acessar**, conforme figura 4-4;

Figura 4-4 - Identificando usuário na plataforma

2. Clique na caixa Buscar cursos e digite: **Bioestatística I/CZS**, depois clique em **Vai**, conforme figura 4-5.

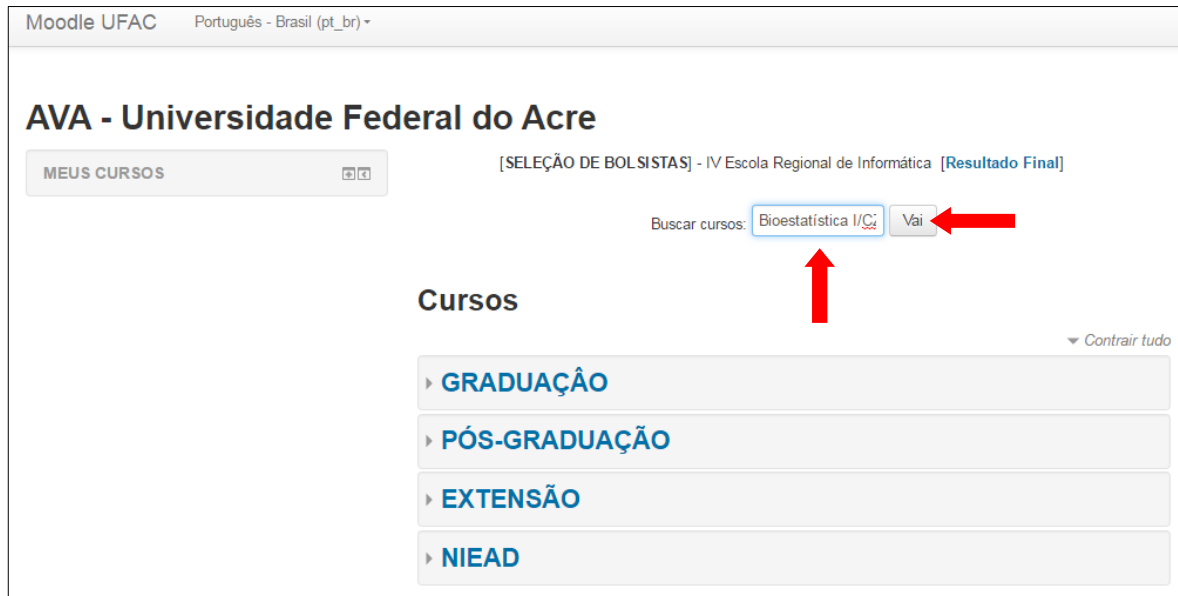


Figura 4-5 - Buscando cursos na plataforma

3. No resultado da pesquisa aparecerá o link para acessar a disciplina, clique em **Bioestatística I/CZS** para visualizar os recursos disponíveis, conforme figura 4-6.



Figura 4-6 - Acessando a disciplina de Bioestatística I/CZS

4. Ao clicar no link Bioestatística I/CZS a página do curso será aberta.

4.4 Explorando os recursos introdutórios da disciplina

Ao clicar na figura do sapo (figura 4-7), o usuário é direcionado para uma página com jogos voltados para a área biológica do site www.sobiologia.com.br. Este recurso opcional serve como entretenimento entre as atividades da disciplina.

Moodle UFAC Rogério Lopes Craveiro

Bioestatística I/CZS

Página inicial ▸ GRADUAÇÃO ▸ CMULTI ▸ CMULT735 Ativar edição

PARTICIPANTES

Participantes

NAVEGAÇÃO

Página inicial

- Painel
- Páginas do site
- ▾ Curso atual
 - CMULT735
 - Participantes
 - Bem-vindo a Bioestatística
 - INTRODUÇÃO A BIOESTATÍSTICA
 - ORGANIZAÇÃO DE DADOS EM TABELAS E GRÁFICOS
 - MEDIDAS ESTATÍSTICAS
 - Medidas de Dispersão para uma Amostra
 - NOÇÕES DE CORRELAÇÃO
 - Noções sobre Probabilidade
 - Distribuição Binomial

Bem-vindo a Bioestatística

PESQUISAR NOS FÓRUNS

Vai

[Pesquisa Avançada](#)

ÚLTIMAS NOTÍCIAS

[Acrescentar um novo tópico...](#)

Tarefa para a aula do dia 16 de junho de 2016

16 Jun, 08:40 Rogério Lopes Craveiro

[Horário de início e termino das aulas](#)

11 Mai, 17:52 Rogério Lopes Craveiro

[Bem-vindos a Bioestatística](#)

3 Mar, 16:11 Rogério Lopes Craveiro

[Início da Disciplina de Bioestatística I na Plataforma Moodle](#)

4 Nov, 10:59 Rogério Lopes Craveiro

[Chat sábado dia 07/06/14 15h00min](#)

3 Jun, 08:31 Rogério Lopes Craveiro

Figura 4-7 - Página do curso de Bioestatística I

Logo abaixo da figura do sapo no início da página de apresentação (figura 4-8), são disponibilizados alguns recursos que ao serem clicados direcionam o usuário para outras áreas da plataforma, estas áreas serão descritas a seguir:

4.4.1 Fórum de Notícias

Este item é utilizado para o professor dar informes relevantes sobre o andamento da disciplina. Ao clicar neste recurso o usuário será direcionado para uma nova página onde poderá visualizar as informações em forma de tópicos.

Curso atual

- ▾ CMULT735
 - Participantes
 - Bem-vindo a Bioestatística
 - INTRODUÇÃO A BIOESTATÍSTICA
 - ORGANIZAÇÃO DE DADOS EM TABELAS E GRÁFICOS
 - MEDIDAS ESTATÍSTICAS
 - Medidas de Dispersão para uma Amostra
 - NOÇÕES DE CORRELAÇÃO
 - Noções sobre Probabilidade
 - Distribuição Binomial
 - Distribuição Normal
 - Teste Qui-quadrado
 - Teste t
 - Análise de Variância (ANAVA)
 - Links úteis
- Meus cursos
- Cursos

- Fórum de Notícias
- Fórum de Apresentação
- [Chat] Interação da Turma - aberto 24h
- Pré-teste de avaliação do curso
- Plano de Curso de Bioestatística I
- Diário de Bordo
- Glossário de Biologia e Estatística
- Wiki de Bioestatística

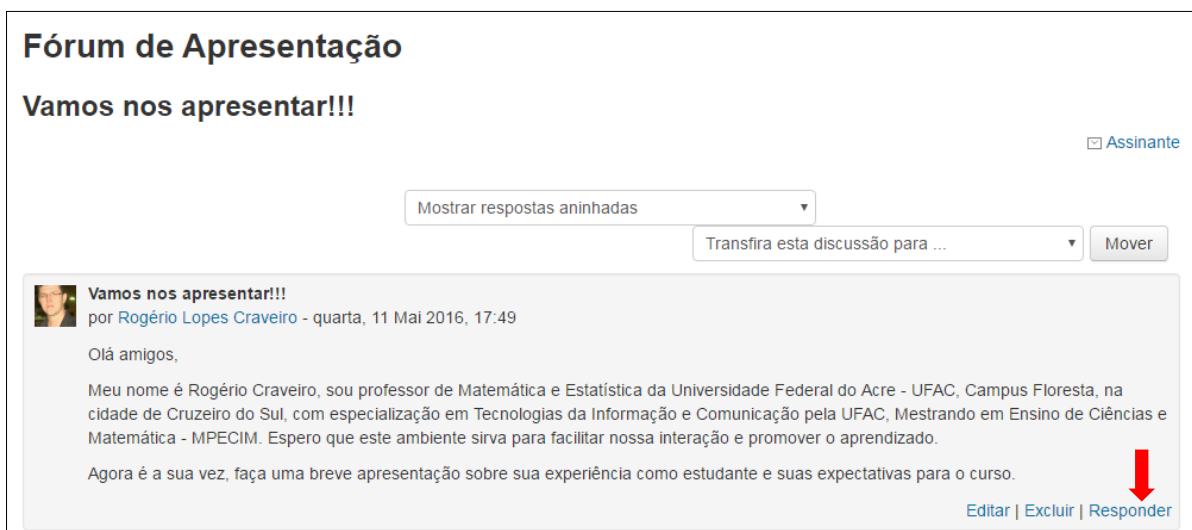
ADMINISTRAÇÃO

- ▾ Administração do curso
 - Ativar edição

Figura 4-8 - Recursos iniciais do curso

4.4.2 Fórum de Apresentação

Neste item é proposta ao usuário a realização de um breve relato sobre sua vida pessoal e suas expectativas sobre a disciplina. Para inserir um comentário o usuário deve clicar em responder na mensagem do professor, conforme figura 4-9.



Fórum de Apresentação

Vamos nos apresentar!!!

Assinante

Mostrar respostas aninhadas

Transfira esta discussão para ... Mover

Vamos nos apresentar!!!
por Rogério Lopes Craveiro - quarta, 11 Mai 2016, 17:49

Olá amigos,

Meu nome é Rogério Craveiro, sou professor de Matemática e Estatística da Universidade Federal do Acre - UFAC, Campus Floresta, na cidade de Cruzeiro do Sul, com especialização em Tecnologias da Informação e Comunicação pela UFAC, Mestrando em Ensino de Ciências e Matemática - MPECIM. Espero que este ambiente sirva para facilitar nossa interação e promover o aprendizado.

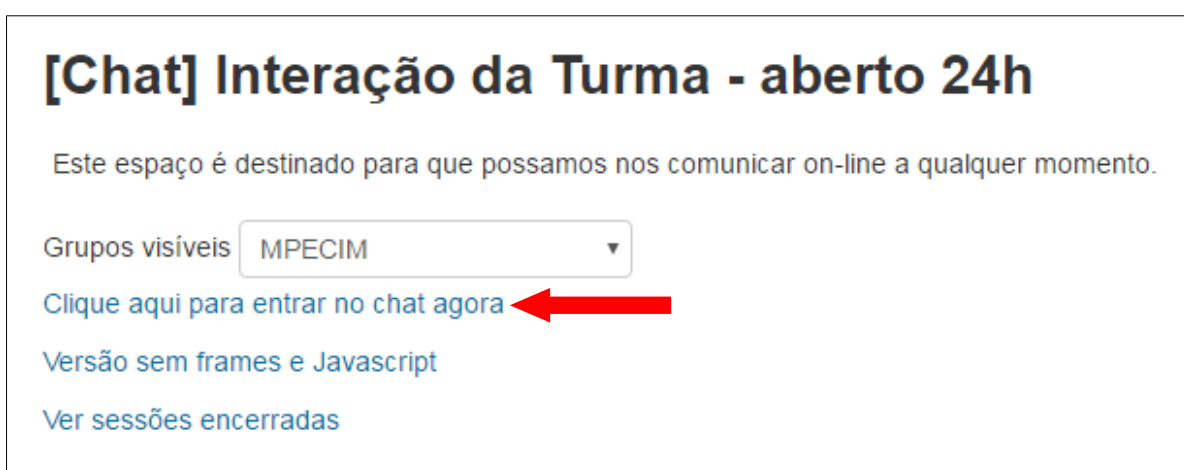
Agora é a sua vez, faça uma breve apresentação sobre sua experiência como estudante e suas expectativas para o curso.

[Editar](#) | [Excluir](#) | [Responder](#)

Figura 4-9 - Forum de apresentação

4.4.3 Chat Interação da Turma – Aberto 24h

Este item permite que os alunos possam ter um espaço onde tirar dúvidas em tempo real com o professor sobre tópicos abordados em aulas anteriores. Para iniciar o chat o usuário deve clicar em **Clique aqui para entrar no chat agora**.



[Chat] Interação da Turma - aberto 24h

Este espaço é destinado para que possamos nos comunicar on-line a qualquer momento.

Grupos visíveis

[Clique aqui para entrar no chat agora](#)

[Versão sem frames e Javascript](#)

[Ver sessões encerradas](#)

Figura 4-10 - Página para acesso ao Chat

Ao clicar no link indicado na figura 4-10, será aberta uma janela para o usuário iniciar sua participação no chat.

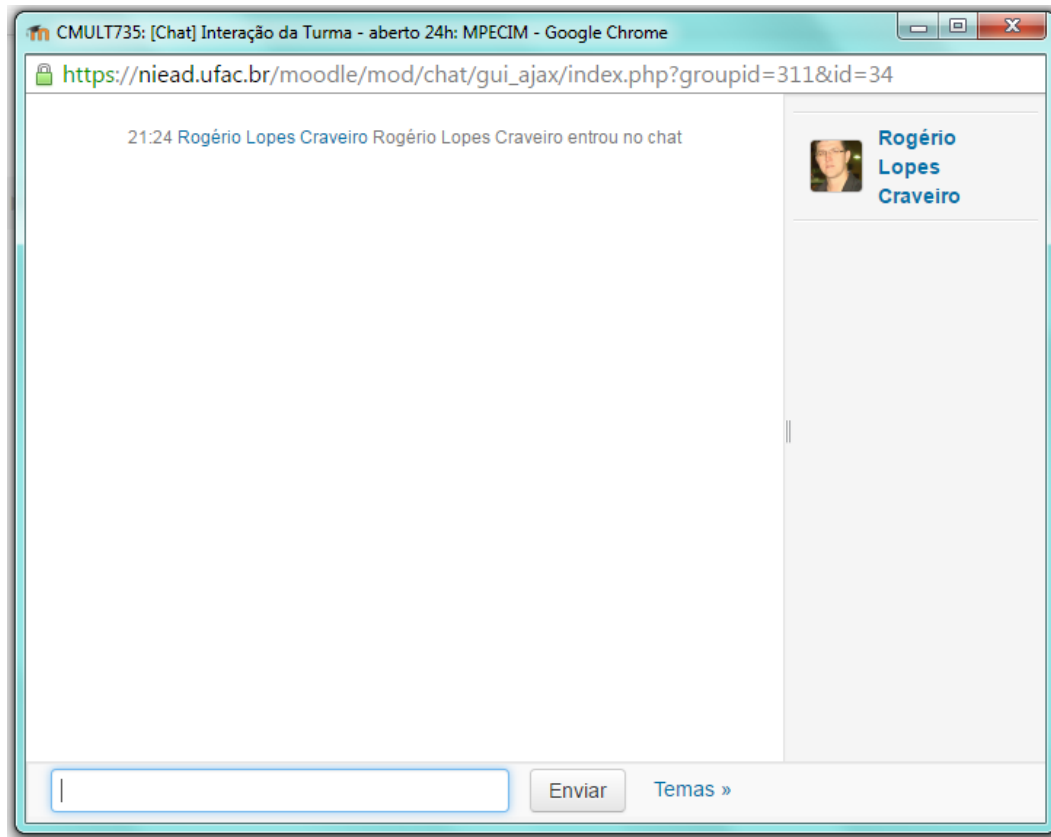


Figura 4-11 - Chat de interação da turma

4.4.4 Pré-teste de Avaliação do Curso

Este link direciona o aluno a um questionário sobre a relevância da disciplina para a formação em ciências biológicas, a proposta é que os usuários respondam as questões baseados em suas expectativas quanto ao seu aprendizado no curso. As questões deste pré-teste de avaliação do curso estão disponíveis no anexo I.

Pré-teste de avaliação do curso [Ver 76 respostas](#)

COMO RESPONDER AO QUESTIONÁRIO O objetivo deste questionário é fazer uma pesquisa sobre as suas expectativas em relação ao seu processo de aprendizagem neste curso. Cada frase abaixo é uma declaração que descreve situações que podem caracterizar o seu processo de aprendizagem neste curso. Reflita sobre cada declaração e, depois, selecione a opção que descreve melhor a frequência em que você gostaria que as situações descritas ocorressem neste curso (frequência desejada). Não há respostas certas ou erradas; nós estamos interessados apenas na sua opinião. Todas as respostas dadas são confidenciais e não tem nenhum impacto sobre a avaliação do participante. Agradecemos a sua colaboração.

Todas as questões são obrigatórias e devem ser respondidas

Figura 4-12 - Pré-teste de avaliação do curso

4.4.5 Plano de Curso de Bioestatística I:

Ao clicar neste item o usuário pode baixar o plano de curso da disciplina. O plano de curso está disponível no Anexo II deste material.

4.4.6 Diário de bordo:

Este item servirá para que o usuário registre sua evolução na disciplina, servindo como um diário, onde o mesmo pode descrever seus problemas e sua evolução no aprendizado de Bioestatística. Apenas o usuário e o professor podem ter acesso a este texto. Para inserir um texto no Diário clique em **Editar envio**.

Diário de Bordo

Caro(a) aluno(a),

Este espaço foi criado para que você reflita sobre sua participação na disciplina, comece respondendo como a plataforma moodle está contribuindo para seu aprendizado em Bioestatística, quais as vantagens da utilização do software Action em estatísticas, suas dificuldades no decorrer do curso, seus avanços e suas dicas para a melhoria da disciplina.

Status de envio

Status de envio	Enviado para avaliação
Status da avaliação	Não há notas
Última modificação	quarta, 11 Mai 2016, 17:56
Texto online	

Comentários ao envio ▶ Comentários (0)

←

Figura 4-13 - Diário de Bordo

4.4.7 Glossário de Biologia e Estatística:

Este item serve como um dicionário de termos biológicos e estatísticos que surgirão nas atividades propostas posteriormente. O usuário pode inserir o significado de termos biológicos e estatísticos que surgirão no decorrer da disciplina, principalmente contidos nos exemplos e exercícios propostos. Para inserir o significado de um termo o usuário deve clicar em **Inserir novo item**.

Glossário de Biologia e Estatística

Caro(a) aluno(a),

Este é o espaço para que possamos construir juntos um glossário de biologia e estatística, com o significado de termos que aparecem em nossos textos, exemplos e exercícios na disciplina de bioestatística. Não deixem de informar a fonte das informações disponibilizadas.

Buscar em todo o texto

←

Figura 4-14 - Glossário de Biologia e Estatística

Ao clicar em inserir novo item o usuário será direcionado para a página seguinte que permitira que o conceito e a definição do termo biológico ou estatístico sejam inseridos. Após a inserção clique em **Salvar mudanças**.

Glossário de Biologia e Estatística

Caro(a) aluno(a),

Este é o espaço para que possamos construir juntos um glossário de biologia e estatística, com o significado de termos que aparecem em nossos textos, exemplos e exercícios na disciplina de bioestatística. Não deixem de informar a fonte das informações disponibilizadas.

[▶ Expandir tudo](#)

▼ Geral

Conceito*

Definição*

Parágrafo

B *I* ✂ 📄 📁 ☰ ☷ 🔑 🔗 🔗 🖼 📺 📄 <>

Caminho: p

Outras palavras que serão linkadas ao mesmo item

Figura 4-15 - Inserindo termos no glossário

4.4.8 Wiki:

Este item propõe a construção de um texto produzido em grupo com um tema predefinido pelo professor. Para criar um parágrafo deste texto clique em **Criar página**.

Wiki de Bioestatística

O wiki é uma ferramenta de produção de textos colaborativos. O objetivo é que cada participante do grupo escreva sobre o tema "A importância da estatística para o aprendizado de biologia". Assim, ao final da atividade teremos construído um texto em conjunto. Bom trabalho!

▼ Nova página

Novo título da página*

Formato ?

* Formato HTML

* Formato Creole

* Formato NWiki

Grupo MPECIM

Figura 4-16 - Wiki de Bioestatística

4.5 Tópicos de Conteúdos

Os conteúdos estão estruturados na plataforma de acordo com o plano de curso da disciplina. A seguir passamos a relatar as funcionalidades de cada uma das ferramentas disponíveis nestes tópicos.

4.5.1 Fórum de Dúvidas

No início de cada tópico de conteúdo, após a figura ilustrativa, disponibilizamos a ferramenta Fórum de Dúvidas.



Esta ferramenta permite que os usuários possam postar suas dúvidas sobre os assuntos estudados em sala de aula ou sobre a leitura do material disponível na plataforma. As postagens podem ser visualizadas pelo professor e pelos colegas, dessa forma, todos podem responder as dúvidas coletivamente, proporcionando uma interação entre professor-aluno e entre aluno-aluno.

Ao clicar em Fórum de Dúvidas o usuário será direcionado para a página ilustrada na figura 4-17, onde deverá clicar em **A acrescentar um novo tópico de discussão** para inserir uma nova pergunta.

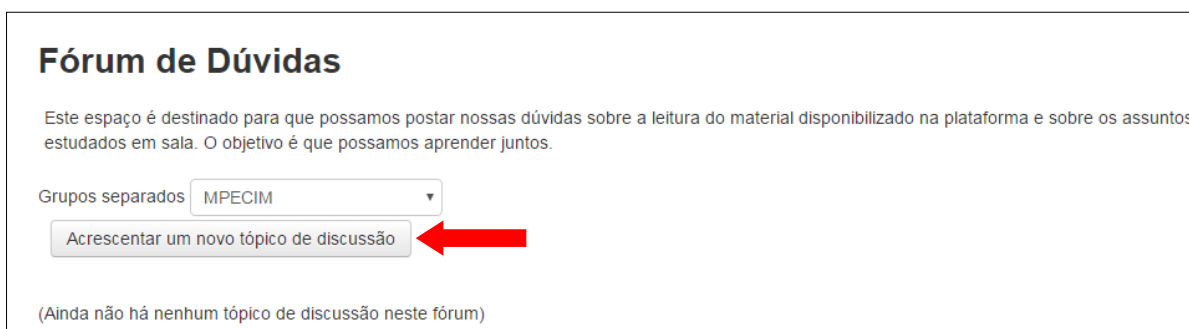


Figura 4-17 - Fórum de Dúvidas

Em seguida o usuário deverá digitar apenas o título da pergunta na caixa **Assunto**. O conteúdo da pergunta deve ser digitado na caixa **Mensagem**, conforme ilustra a figura 4-18. Para postar a pergunta clique em **Enviar mensagem ao fórum**.

Fórum de Dúvidas

Este espaço é destinado para que possamos postar nossas dúvidas sobre a leitura do material disponibilizado na plataforma e sobre os assuntos estudados em sala. O objetivo é que possamos aprender juntos.

▼ Novo tópico de discussão

Assunto* Como acrescentar perguntas no fórum de dúvidas? ←

Mensagem*

Parágrafo ▾ B I ✂ 📄 📎 ☰ ☷ 🔗 🔗 🔗 🖼️ 🎥 📄 ⏪ ⏩

Olá colegas,
 Como devo fazer para acrescentar perguntas no Fórum de dúvidas? ←
 Abraços.
 Caminho: p

Assinatura de discussão ⓘ

Anexo ⓘ Tamanho máximo para novos arquivos: 500Kb, máximo de anexos: 9

📎 Arquivos

↓
 Você pode arrastar e soltar arquivos aqui para adicioná-los.

Enviar email em seguida

Enviar uma cópia a todos os grupos

ⓘ

Grupo MPECIM ▾

Enviar mensagem ao fórum Cancelar ←

Figura 4-18 - Inserindo perguntas no forum de dúvidas

4.5.2 Livro – Análise Exploratória de Dados Biológicos

Em cada tópico de conteúdo é disponibilizado o livro *Análise Exploratória de Dados Biológicos*, que foi dividido em cinco tópicos na seguinte ordem: Introdução a Análise Exploratória de Dados Biológicos, Organização de Dados em Tabelas, Gráficos, Medidas Estatísticas e Noções de Correlação. Ao clicar em um desses títulos, será aberta uma página contendo o arquivo do livro em formato PDF⁵.

 [Introdução a Análise Exploratória de Dados Biológicos](#)

⁵ A sigla inglesa **PDF** significa *Portable Document Format* (Formato Portátil de Documento), um formato de arquivo criado pela empresa *Adobe Systems* para que qualquer documento seja visualizado, independente de qual tenha sido o programa que o originou.

Recomendamos ao usuário clicar com o botão direito do *mouse* sobre o arquivo do livro e selecionar a opção **Abrir link em uma nova guia**, dessa forma, o navegador de internet abrirá o arquivo de texto em uma nova guia e a página da plataforma permanecerá visível, permitindo a navegação entre o texto e a plataforma simultaneamente.

O livro contém *links* após a apresentação de alguns exemplos que direcionam o usuário a vídeos tutoriais no *youtube*⁶ sobre a utilização de ferramentas computacionais para cálculos estatísticos.

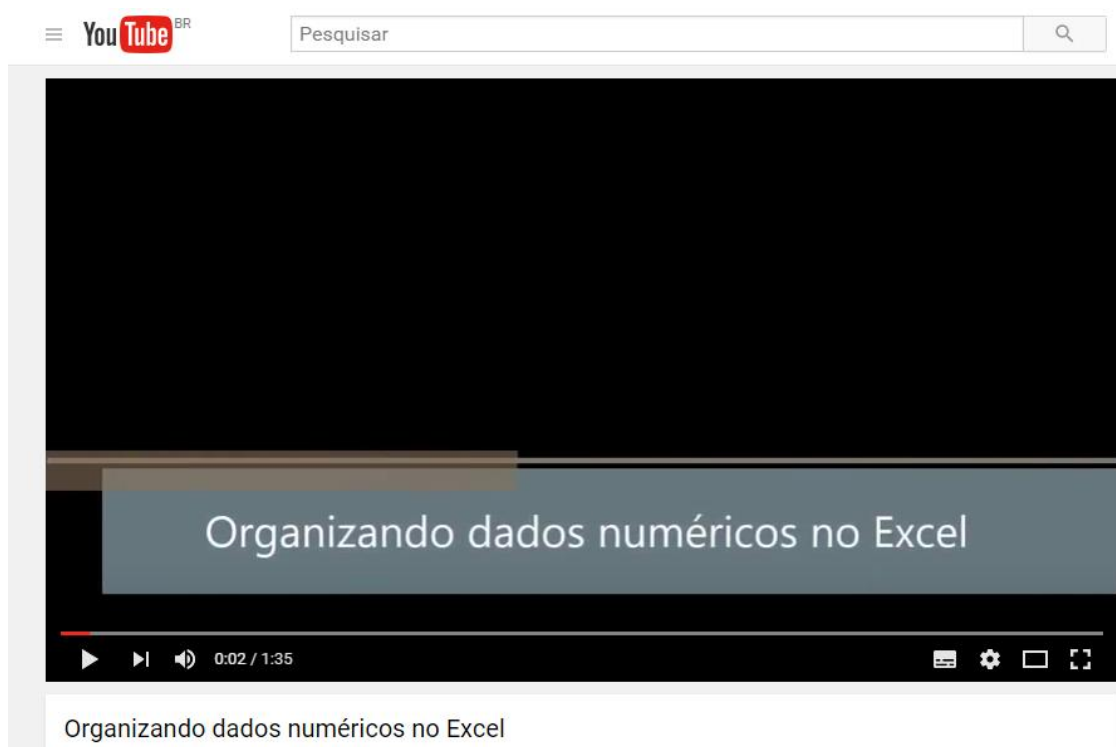


Figura 4-19 - Página do *Youtube* com vídeo tutorial

4.5.3 Tabelas do *Excel* para download

Alguns exemplos do livro *Análise Exploratória de Dados Biológicos* apresentam tabelas com dados de pesquisas realizadas por professores e alunos da UFAC, estes dados são disponibilizados para download na plataforma em arquivos do *Microsoft Office Excel* com o objetivo de aperfeiçoar o tempo de aula, pois os dados não precisarão ser digitados pelos alunos, priorizando a manipulação das ferramentas computacionais e a análise dos resultados.

 [Tabela para exemplos de amostras](#)

⁶ <https://www.youtube.com/>

Ao clicar em um desses *links* será aberta uma janela para a escolha do local onde o usuário deseja salvar o arquivo, escolha o local e clique em salvar. Para abri-lo, selecione a pasta onde salvou e de um clique duplo sobre o mesmo.

4.5.4 Atividades

Ao concluirmos alguns tópicos de conteúdo apresentamos atividades que deverão ser realizadas na plataforma *Moodle* UFAC, em alguns casos estas atividades serão avaliadas automaticamente após o envio e a nota do usuário já poderá ser visualizada. É o caso da “Atividade 1” sobre variáveis estatísticas.

[Atividade 1] - Variáveis estatísticas

Ao clicar no *link* desta atividade, o usuário será direcionado para a página ilustrada na figura 4-20 onde poderá iniciar a resolução da tarefa clicando em **Tentar responder o questionário agora**.



Figura 4-20 - Atividade 1

Na página seguinte o usuário poderá responder a um questionário de múltipla escolha sobre as variáveis estatísticas. Esta primeira atividade em especial permite a realização de várias tentativas até conseguir a resposta certa. Para finalizar a tarefa clique em **Enviar tudo e terminar**.

As demais atividades poderão ser realizadas *off-line* (sem precisar estar conectado a internet) com o uso de aplicativos computacionais e enviadas pela plataforma para a correção do professor.

[Atividade 2] Tabelas de frequência

Para enviar as atividades clique no link disponível, o usuário será direcionado para a página ilustrada na figura 4-21, clique em **Editar envio**.

[Atividade 2] Tabelas de frequência

Caro aluno,

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*.

Salve o arquivo em formato do Excel e envie com o seguinte nome:

Atividade2_nomesobrenome

Bons estudos!!!

Status de envio

Status de envio	Enviado para avaliação
Status da avaliação	Não há notas
Última modificação	quinta, 12 Mai 2016, 14:40
Comentários ao envio	▶ Comentários (0)



[Editar envio](#) 


Figura 4-21 - Atividade para envio de arquivos

Na janela seguinte clique em **Adicionar**, conforme indicado com a seta vermelha na figura 4-22. O usuário também pode arrastar o arquivo para a área indicada com a seta azul.

Envio de arquivos Tamanho máximo para novos arquivos: 10Mb, máximo de anexos: 1

▶ Arquivos


Você pode arrastar e soltar arquivos aqui para adicioná-los.

[Salvar mudanças](#) [Cancelar](#)

Figura 4-22 - Envio de arquivos

Ao clicar em **adicionar**, será aberta uma janela conforme ilustrada na figura 4-22, clique em **Escolher arquivo** e selecione o local onde está o arquivo que deseja enviar para avaliação. Após digite o nome do arquivo em **Salvar como** e o nome do usuário em **Autor**. Em seguida, clique em **Enviar este arquivo**, conforme figura 4-23.

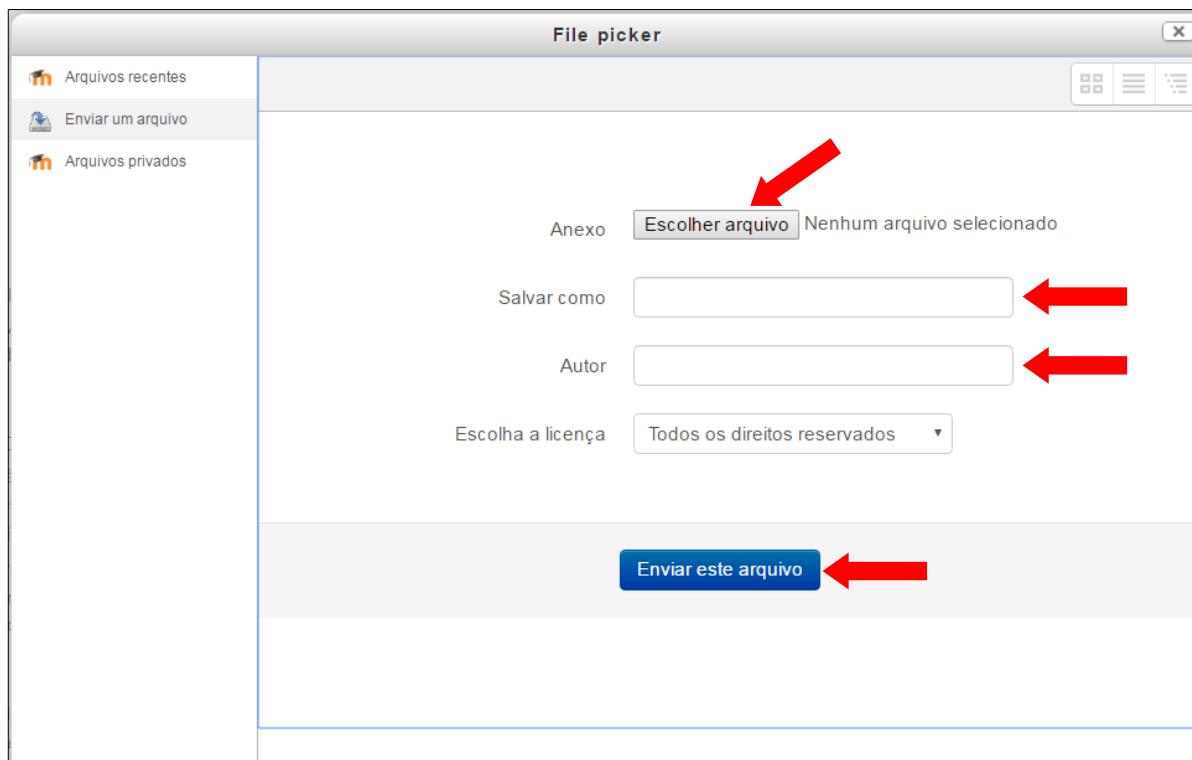



Figura 4-23 - Selecionando o arquivo para envio

A confirmação de conclusão da atividade será enviada para o *e-mail* cadastrado na plataforma. Após a correção da tarefa pelo professor o usuário receberá uma nova mensagem contendo sua nota e comentários.

4.5.5 Vídeos tutoriais na plataforma

No tópico *Organização de Dados em Tabelas* está disponível um link para acesso a vídeos tutoriais sobre como baixar, instalar e configurar o *software Action 2.5* em seu computador. As configurações mínimas necessárias para a utilização do aplicativo são as seguintes: Windows 7/8/10, Microsoft Office Excel 2007/2010/2013.

 [Vídeo - Como baixar, instalar e configurar o Action](#)

Para acessar os vídeos tutoriais clique no link disponível, escolha o vídeo e assista-o direto da plataforma *moodle UFAC*.

4.5.6 Arquivo de instalação do *Action 2.5*

No tópico *Organização de Dados em Tabelas* também é disponibilizado ao usuário o arquivo de instalação do *software* estatístico *Action 2.5*, para baixa-lo basta clicar sobre o *link* e escolher o local onde deseja salvá-lo.

 [Arquivo de Instalação do Action 2,5](#)

Os procedimentos para a instalação e configuração do *software Action 2.5* estão dispostos em vídeos tutoriais disponíveis na plataforma.

4.6 Outros recursos da plataforma Moodle UFAC

A plataforma *Moodle UFAC* disponibiliza alguns recursos complementares para auxiliar seus usuários em suas navegações pelos cursos disponíveis, estes recursos serão apresentados a seguir:

4.6.1 Participantes

No canto esquerdo da página do curso podemos encontrar o *link Participantes*, que ao ser clicado, direciona o usuário a uma página onde se pode visualizar o perfil de todos os usuários matriculados na disciplina, com endereço de *e-mail* e outras informações.



Figura 4-24 - Link para participantes

4.6.2 Navegação

Abaixo do *link* de **Participantes** podemos encontrar o *link Navegação*, onde estão disponíveis todos os tópicos da disciplina. Ao clicar nos *links* o usuário pode alternar entre os tópicos, facilitando o acesso às ferramentas da disciplina, conforme figura 4-25.

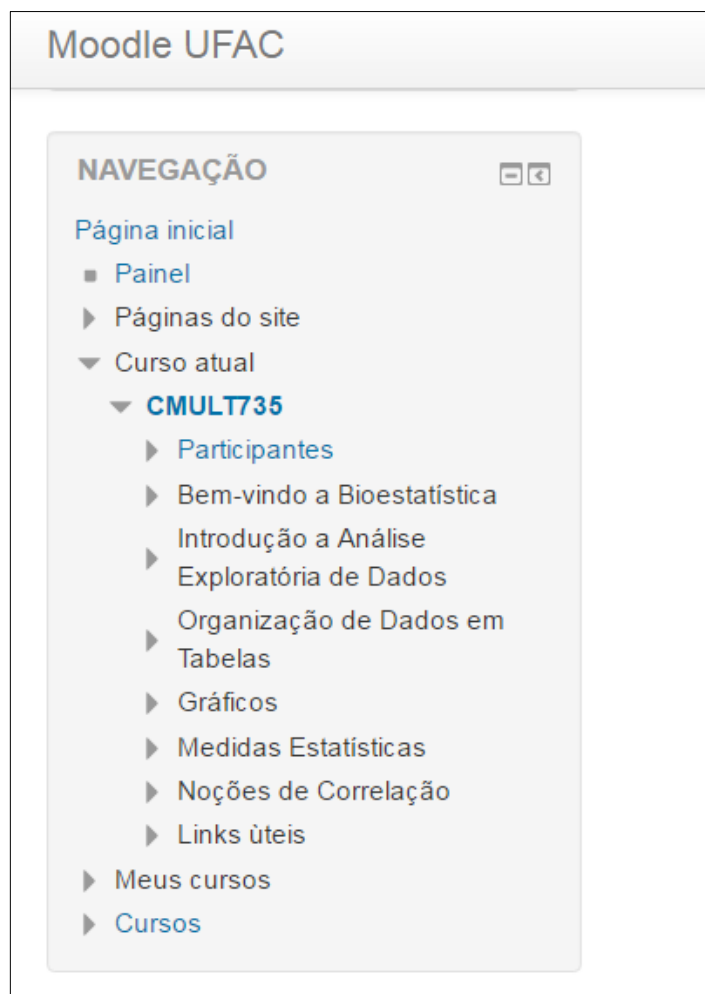


Figura 4-25 - Links de navegação

4.6.3 Administração do Curso - Notas

Abaixo do *link* **Navegação** encontramos o *link* **Notas**, que ao ser clicado direciona o usuário para a página ilustrada na figura 4-26, onde o aluno pede visualizar seu conceito nas atividades realizadas na disciplina de Bioestatística I, com o peso calculado, a nota, o intervalo, a porcentagem, a avaliação e a contribuição para o total do curso. Ao final podemos encontrar ainda a média ponderada das notas.

Item de nota	Peso calculado	Nota	Intervalo	Porcentagem	Avaliação	Contribuição para o total do curso
Bioestatística I/CZS						
[Atividade 2] Tabelas de frequência	-	-	0-100	-	-	-
[Atividade 5] Histograma	-	-	0-100	-	-	-
[Atividade 6] Medidas estatísticas	-	-	0-100	-	-	-
Diário de Bordo	-	-	0-100	-	-	-
[Atividade 9] - Correlação	-	-	0-100	-	-	-
[Atividade 1] - Variáveis estatísticas	-	-	0-10	-	-	-
[Atividade 3] - Gráfico em Setores	-	-	0-100	-	-	-
[Atividade 4] - Gráfico em Barras	-	-	0-100	-	-	-
[Atividade 7] - Quartis	-	-	0-100	-	-	-
[Atividade 8] - Box Plot	-	-	0-100	-	-	-
Total do curso Média ponderada das notas.	-	-	0-100	-	-	-

Figura 4-26 - Notas do usuário

4.6.4 Pesquisar nos Fóruns

Do lado direito da página da disciplina de Bioestatística I, o usuário poderá pesquisar os assuntos que estão sendo discutidos nos fóruns, esta ferramenta auxilia o aluno a verificar se sua dúvida já foi levantada por outro usuário ou se sua pergunta já foi respondida pelo professor ou pelos colegas.

Para pesquisar clique na caixa **Pesquisar nos fóruns**, digite o assunto desejado e clique em **Vai**.



Figura 4-27 - Pesquisar nos fóruns

4.6.5 Últimas Notícias

Abaixo do *link* **Pesquisar nos Fóruns** está a ferramenta de **Últimas Notícias** que informa o usuário sobre os últimos eventos disponibilizados na disciplina, como inserção de tópicos no Fórum de notícias ou no Fóruns de dúvida, atividades recentes, etc.

4.6.6 Mensagens

Em seguida está o tópico de mensagens, com esta ferramenta o usuário pode visualizar se há mensagens recentes em sua caixa postal e quem a enviou. O aluno também pode enviar mensagens para o professor ou para os colegas de curso clicando no link disponível.



Figura 4-28 – Mensagens

4.6.7 Próximos Eventos

Em próximos eventos o usuário pode visualizar se há ou não atividades futuras a serem realizadas na disciplina, buscando organizar-se para o cumprimento dos prazos estabelecidos pelo professor.

4.6.8 Atividades Recentes

Esta ferramenta possibilita ao usuário monitorar seu acesso à disciplina, gerando um relatório completo de atividades recentes, contendo a data e a hora que foram realizadas. A plataforma *Moodle* UFAC grava todos os acessos dos usuários e o tempo em que eles ficaram conectados, isso possibilita ao professor monitorar o tempo que cada aluno disponibiliza para a realização das atividades de sua disciplina.

4.7 Considerações Finais

Este tutorial é apenas uma introdução a utilização da plataforma *Moodle* UFAC na disciplina de Bioestatística I, vários outros recursos que são disponibilizados na plataforma não foram citados aqui, no entanto, acreditamos que com estas informações os usuários poderão acessar os recursos principais da disciplina e realizar as atividades propostas.

As dúvidas que surgirem no decorrer do acesso podemos ser apresentadas nos *Fóruns de Dúvidas* ou sanados nas aulas presenciais com o professor da disciplina.

5. PRODUTO II – LIVRO ANÁLISE EXPLORATÓRIA DE DADOS BIOLÓGICOS

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)

5.1 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.

5.2 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.

5.2.1 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.

5.2.2 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, como é o caso em estudos de casos, ou pode ser um método que antecede uma boa análise inferencial.

5.3 VOCABULÁRIO BÁSICO ESTATÍSTICO

5.3.1 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.

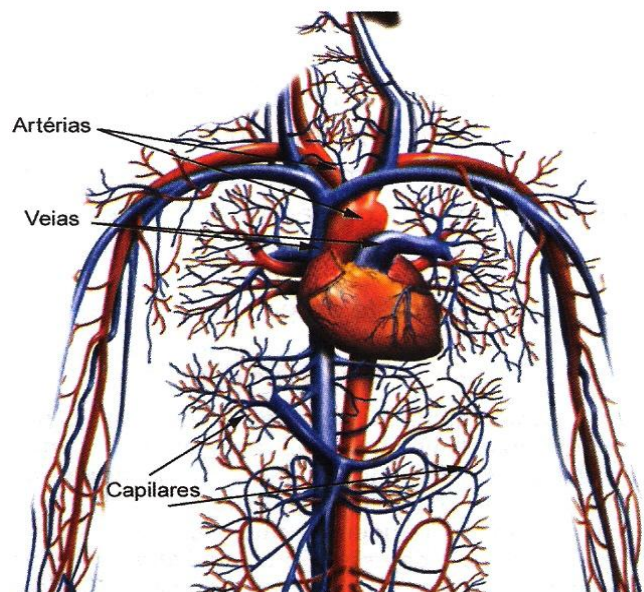


Figura-5-1 - Vasos sanguíneos do corpo humano.

Forte: www.colegiocardeal.com.br

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

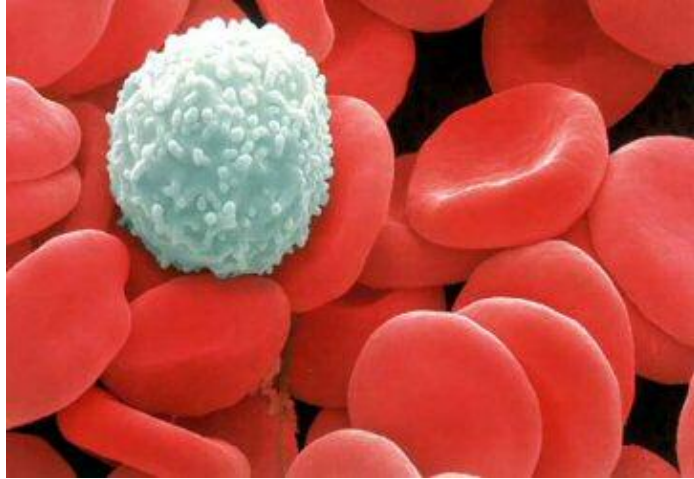


Figura-5-2 - Microscopia eletrônica mostrando as hemácias (em vermelho) e um glóbulo branco (em branco).

Fonte: www.sobiologia.com.br

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



Figura-5-3 - Peixes, algas, plantas e seres microscópicos do mar.

Fonte: www.sobiologia.com.br

5.3.2 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.

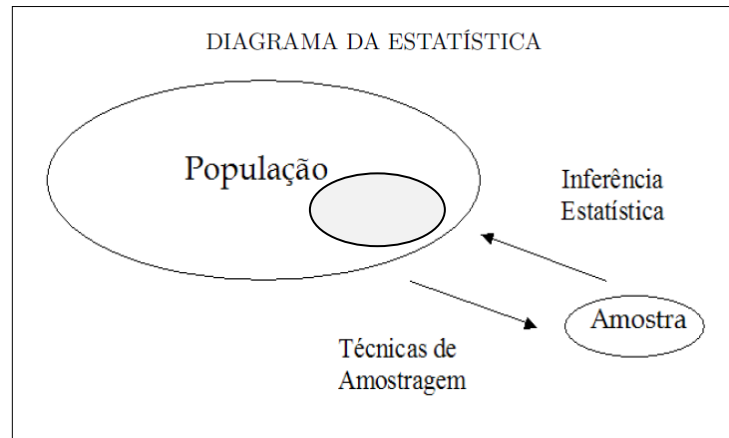


Figura 5-4 - Diagrama da Estatística

Exemplos de amostras em biologia

Amostra de sangue para exames laboratoriais.

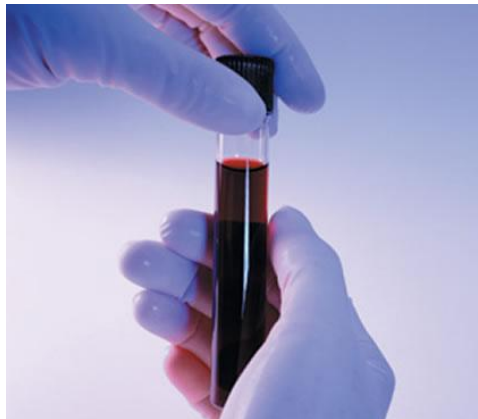


Figura -5-5 Tubo de ensaio com amostra sanguínea

Fonte: <http://www.mundodastribos.com>

Amostra de cobras



Figura -5-6 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-5-7 Coleta de amostra de nadadeira de peixes

Fonte: www.embrapa.br

5.3.3 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.

5.3.4 Tipos de Variáveis

5.3.4.1 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.

5.3.4.2 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.

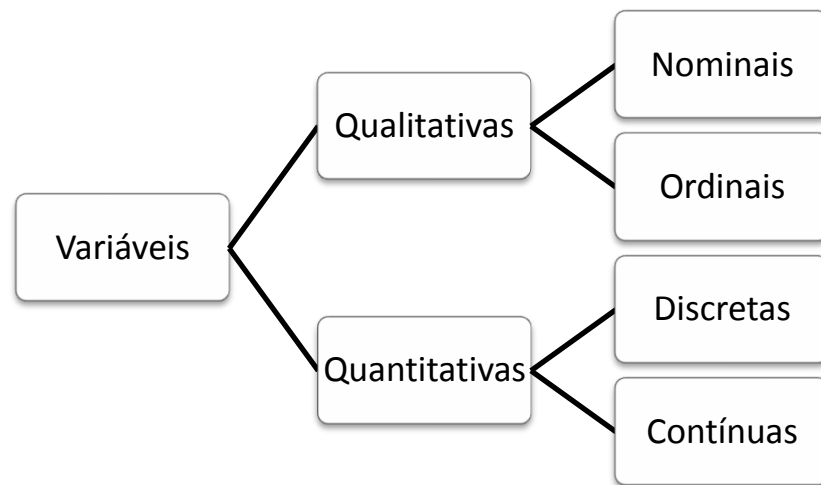


Figura 5-8 - Diagrama das variáveis estatísticas

5.3.5 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.

5.4 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.

5.4.1 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.

5.4.2 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 utilizadas 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.

5.4.3 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 a amostra.

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

5.4.4 Amostra Aleatória Simples

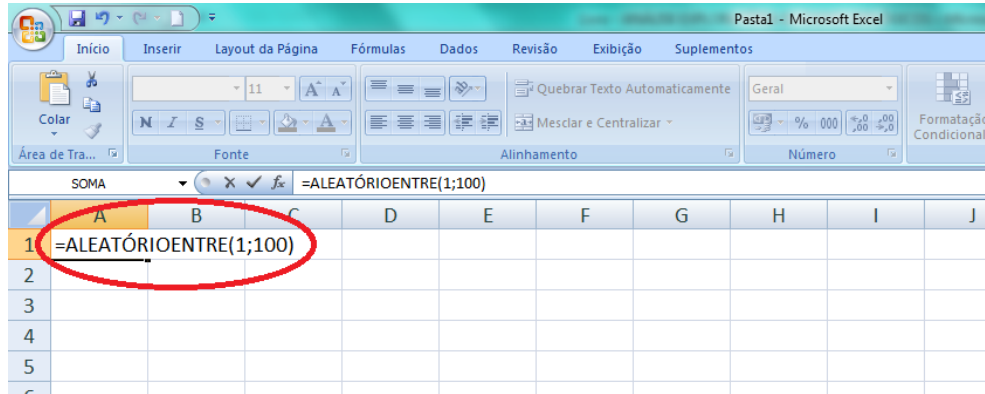
Quando temos uma população homogênea para a variável em estudo, utilizamos a *amostra aleatória simples* que é obtida através de sorteio das unidades da população. Neste cenário, cada ponto da população tem igual probabilidade de ser sorteado. Existem alguns procedimentos que podem ser utilizados para 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 simples⁷ 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.

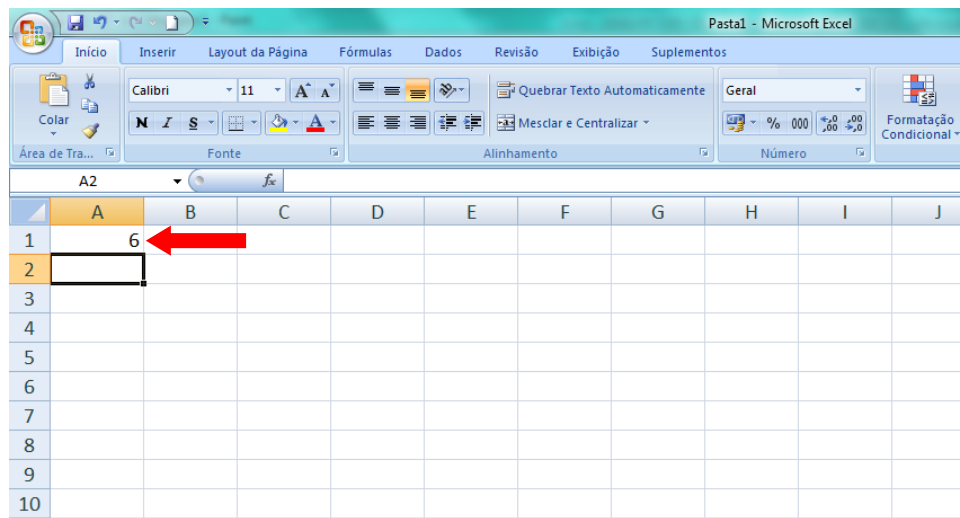
Para gerarmos os números aleatórios no *Excel* procedemos da seguinte forma:

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

⁷ 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.

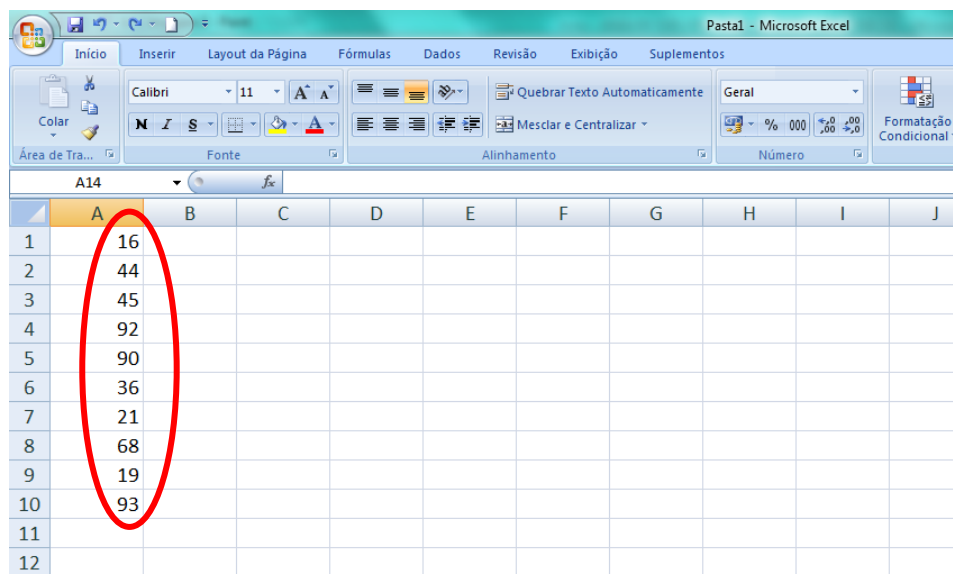


2. Tecele **ENTER**.



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.

5.4.5 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 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.

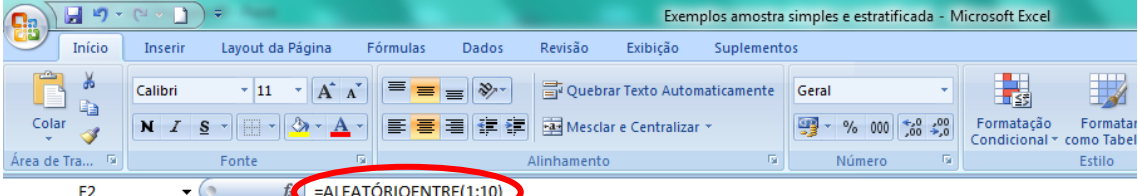
Para realizar o sorteio primeiramente dividi-se os indivíduos em dois grupos ou estratos, no nosso caso, homens e mulheres, em seguida, sorteia-se metade dos homens e metade das mulheres.

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.

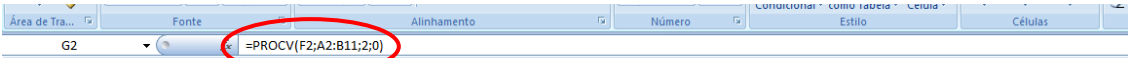
	A	B	C	D	E
1	Nº	HOMENS	Nº	MULHERES	
2	1	Alberto Costa	11	Adriana Santos	
3	2	Carlos Faria	12	Alexandra Freitas	
4	3	João Almeida	13	Andreia Cabral	
5	4	Júlio de Jesus	14	Carla Cardoso	
6	5	Luís Conde Santos	15	Carmen Gomes	
7	6	Manuel Beldade	16	Conceição Mesquita	
8	7	Miguel Coutinho	17	Cristina Rebelo	
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.

5.5 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.

5.5.1 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.

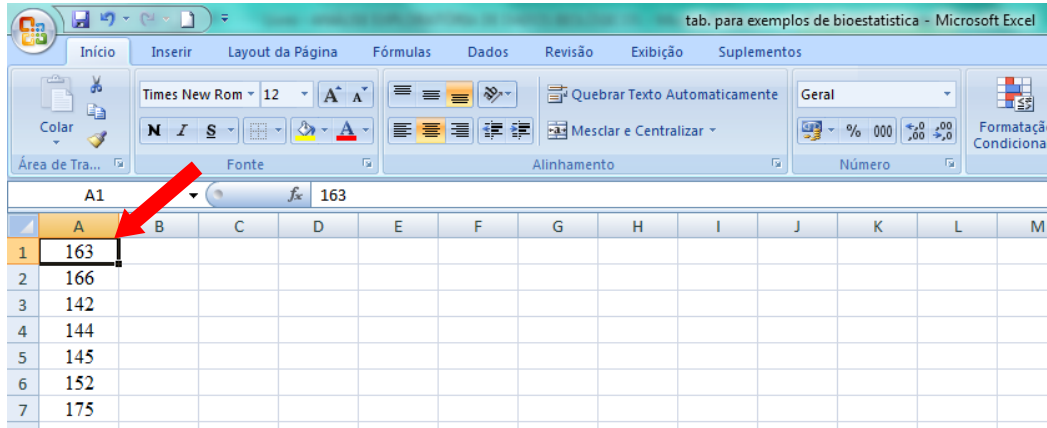
5.5.2 Rol

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.

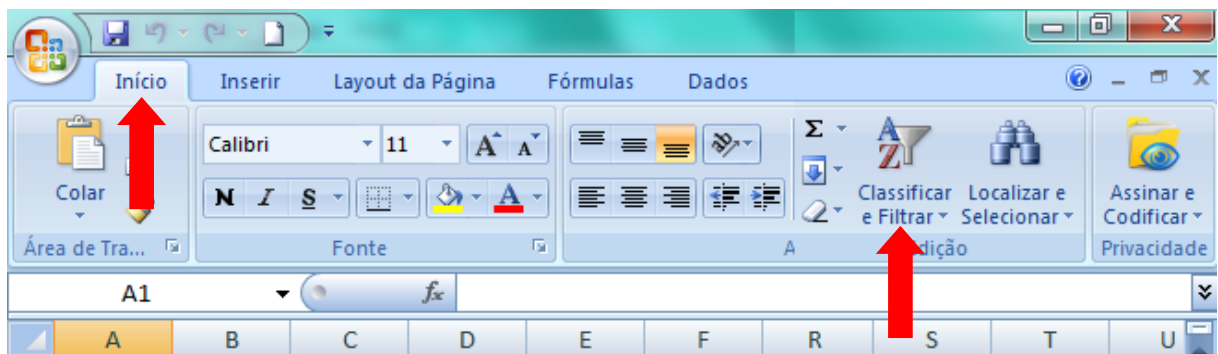
5.5.2.1 Organizando Dados Numéricos no Excel

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

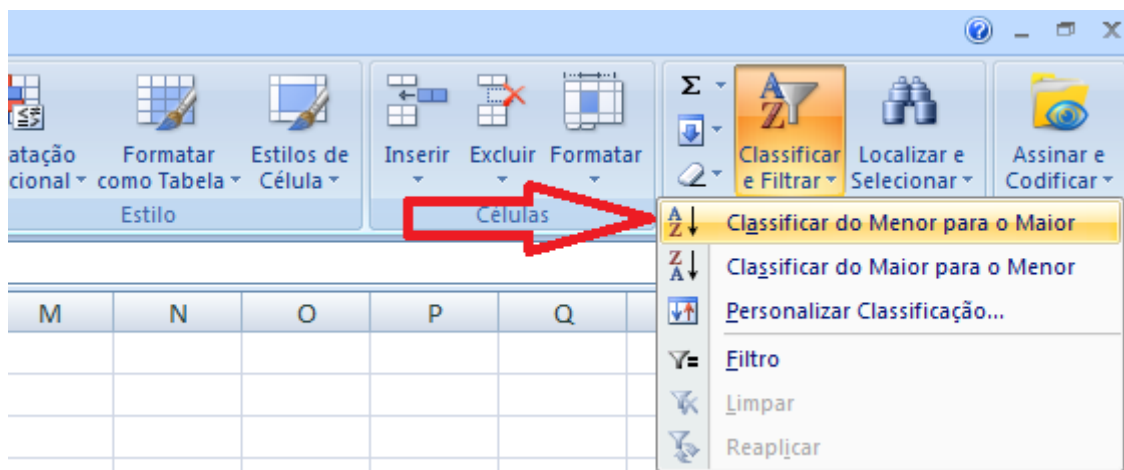
1. Digite os dados da Tabela 1 na coluna A, como mostra a figura abaixo.



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.

5.5.3 Tabelas de Distribuição de Frequência

5.5.3.1 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.

5.5.3.2 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

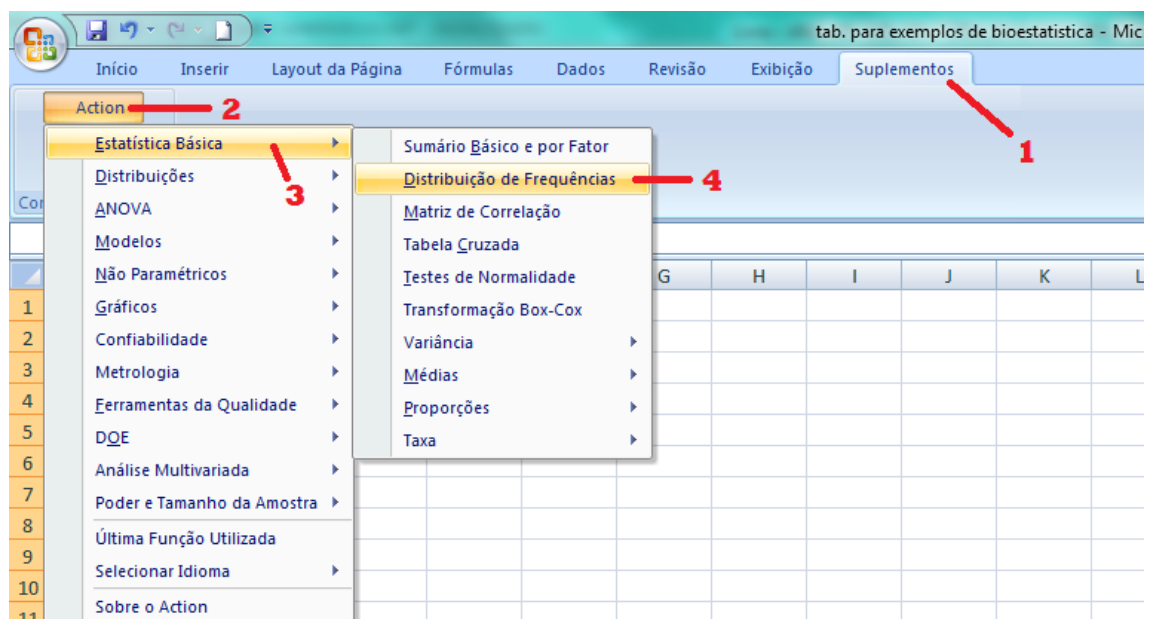
Fonte: Alunos do 3º período do curso de Bacharelado em Ciências Biológicas da UFAC – Campus Floresta

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										

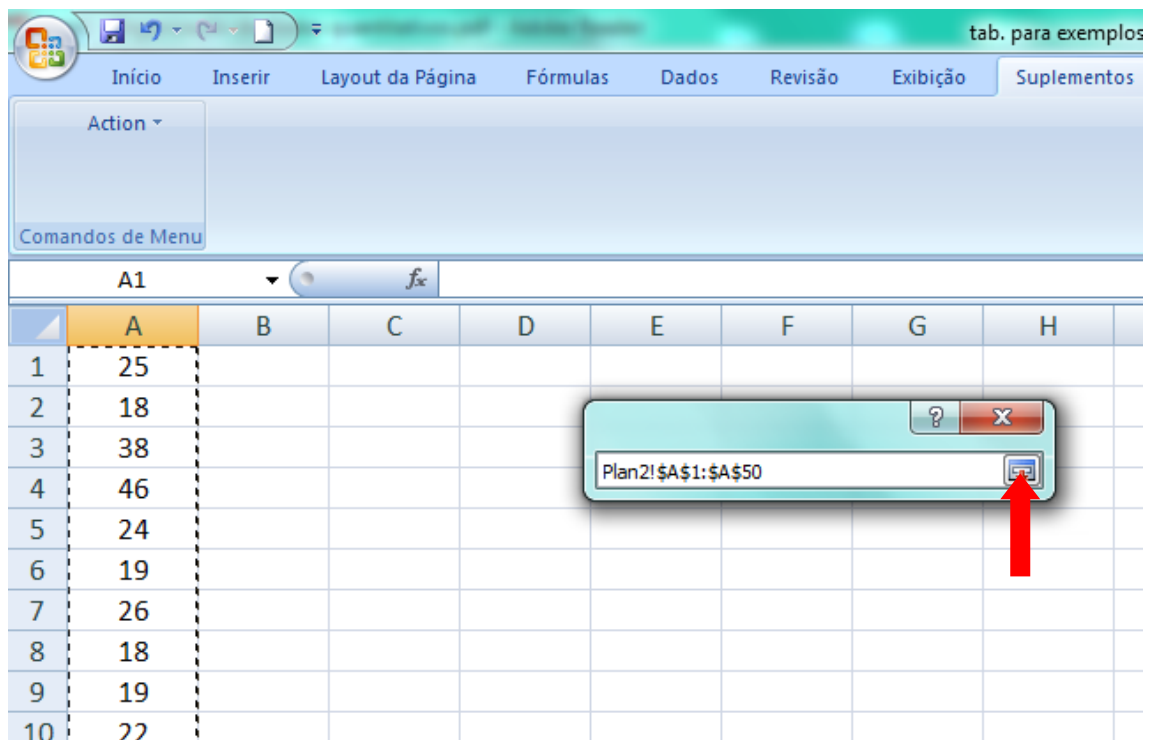
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.



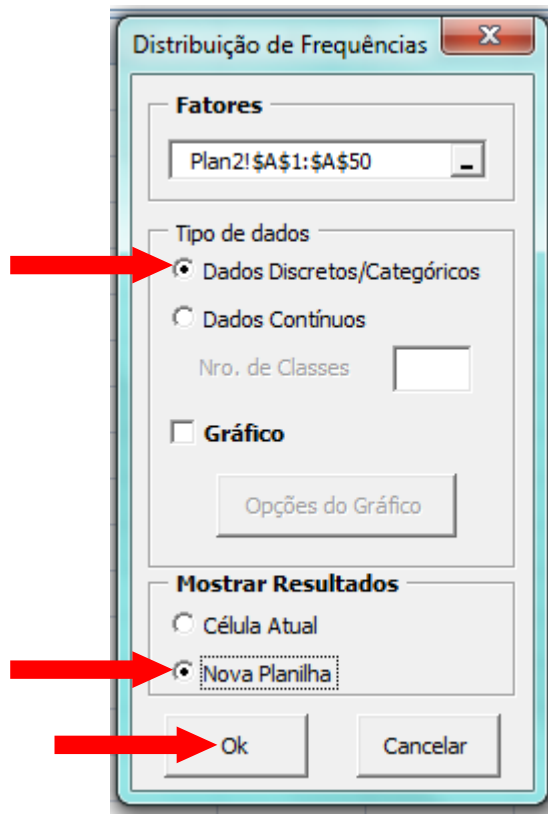
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique no canto direito da caixa de **Fatores** conforme indicado na figura abaixo, em seguida selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A50.



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



5. Clique em Dados **Discretos/Catagóricos**, 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.

	A	B	C	D	E	F	G
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/Categóricos**, 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.

	A	B	C	D	E
1	DISTRIBUIÇÃO DE FREQUÊNCIAS				
2					
3	DADOS DO PROCESSO				
4					
5	Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
6	Eufriesea	1	0,033333333	3,333333333	3,333333333
7	Euglossa	22	0,733333333	73,333333333	76,66666667
8	Eulaema	7	0,233333333	23,333333333	100
9	Total	30			
10					

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

5.5.3.3 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*.

5.5.4 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*.

5.5.4.1 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.

5.5.4.2 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

5.5.4.3 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.

5.5.4.4 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.

5.5.4.5 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

5.5.5 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 estudada 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.

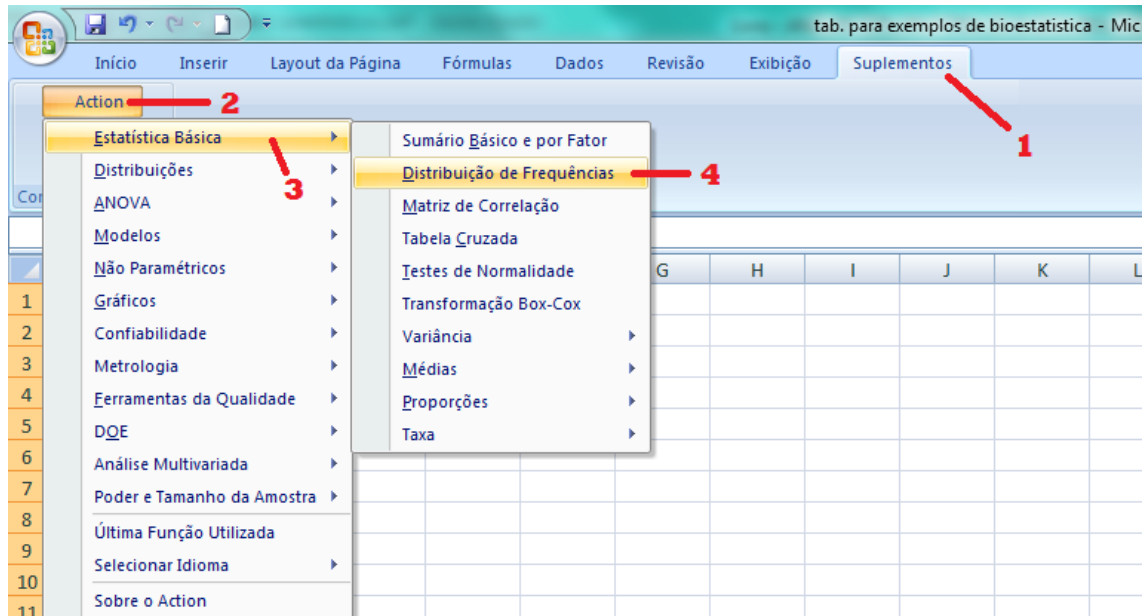
1,78	1,65	1,58	1,58	1,59
1,7	1,53	1,69	1,8	1,69
1,56	1,62	1,64	1,73	1,76
1,65	1,63	1,82	1,7	1,77
1,59	1,7	1,69	1,58	1,67
1,75	1,65	1,72	1,55	1,6
1,6	1,67	1,54	1,58	1,73
1,64	1,61	1,78	1,67	1,83
1,59	1,7	1,6	1,7	1,68
1,69	1,55	1,63	1,77	1,58

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

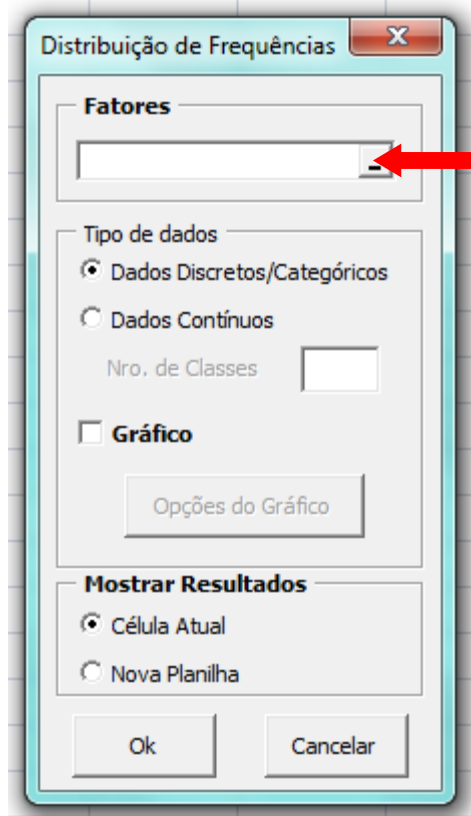
Para construirmos a tabela de distribuição de frequência com intervalos de classe dos dados da Tabela 5, utilizando o *Action* siga os seguintes passos:

1. Digite os dados da Tabela 5 em uma coluna do *Excel*;

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



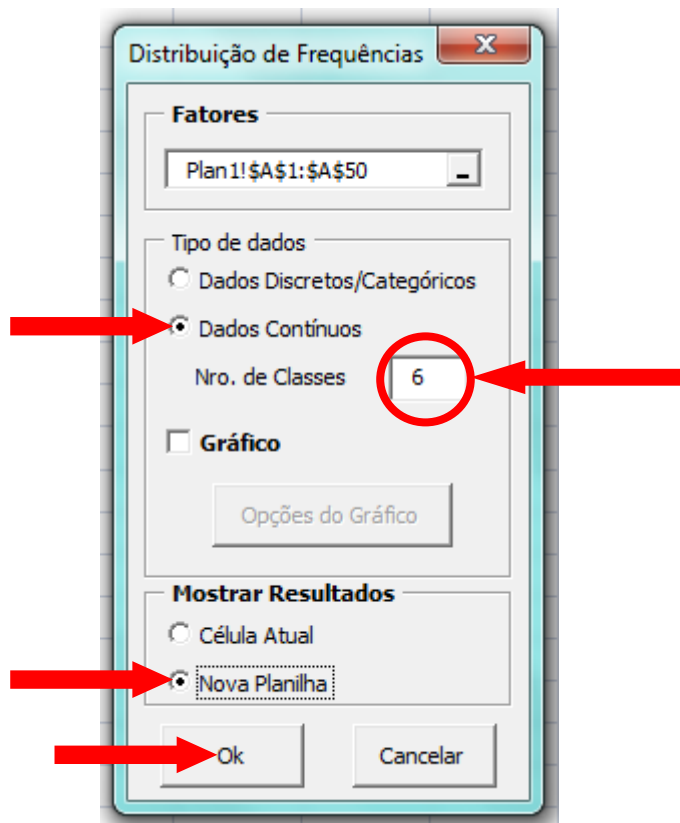
3. A janela de Distribuição de Frequência do Action será aberta. Clique no canto direito da caixa de **Fatores** conforme indicado na figura a seguir, após selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A50.



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*.

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

5.5.6.1 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:

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

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)

5.5.6.2 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.

5.5.6.3 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.

5.5.6.4 Ponto Médio

O ponto médio é a média entre o limite inferior 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

5.5.6.5 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.

5.6 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.

5.6.1 Gráficos para variáveis qualitativas

5.6.1.1 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:

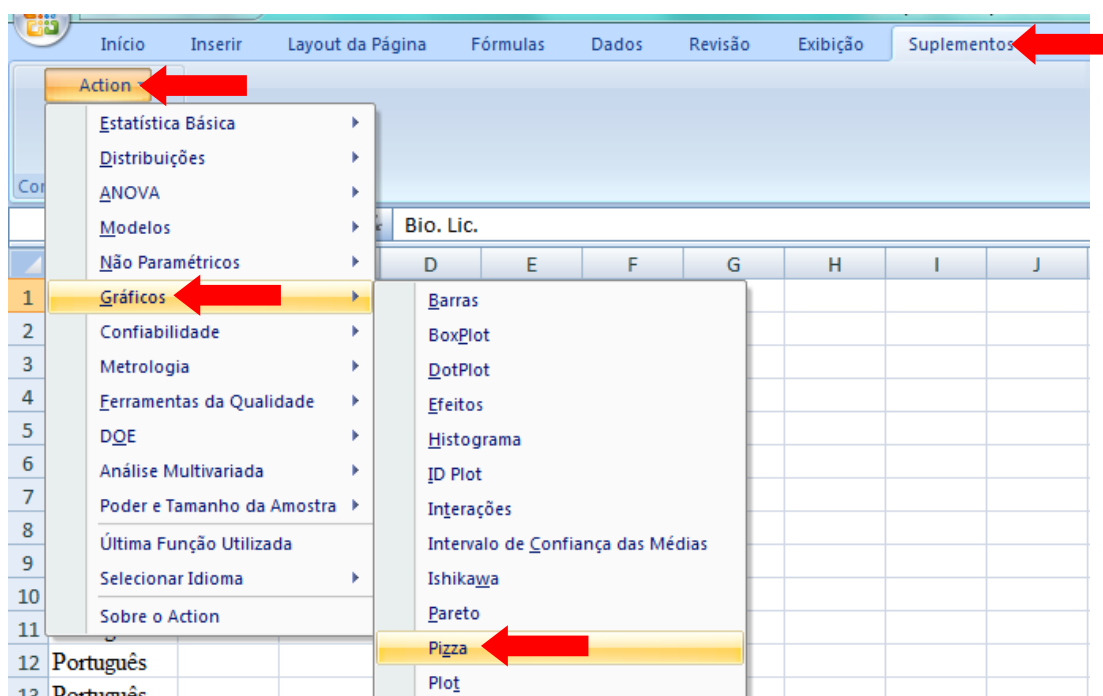
Tabela 7 – Pesquisa realizada com 50 alunos da Universidade Federal do Acre – UFAC – Campus Floresta, na cidade de Cruzeiro do Sul – Acre.

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;



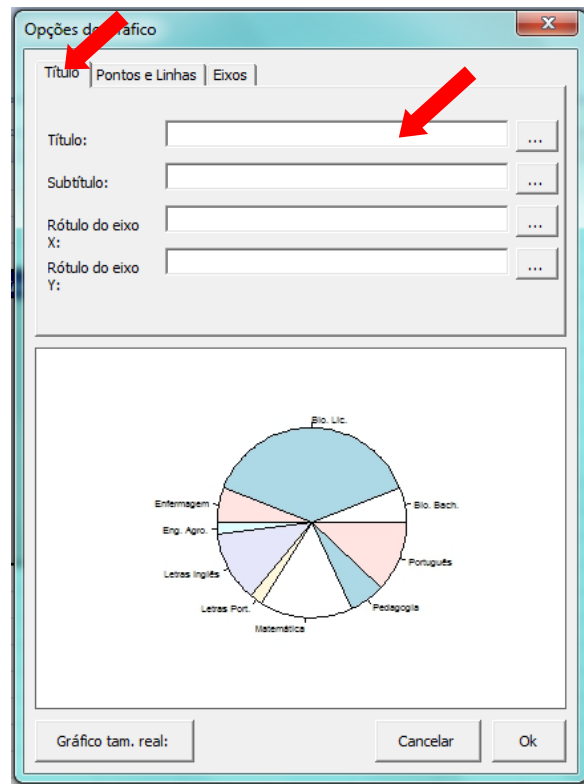
4. A janela **Gráfico de Pizza** será aberta. Clicamos na caixa **Conjunto de Dados** e selecionamos os dados da coluna *frequências* sem o Total;

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			

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

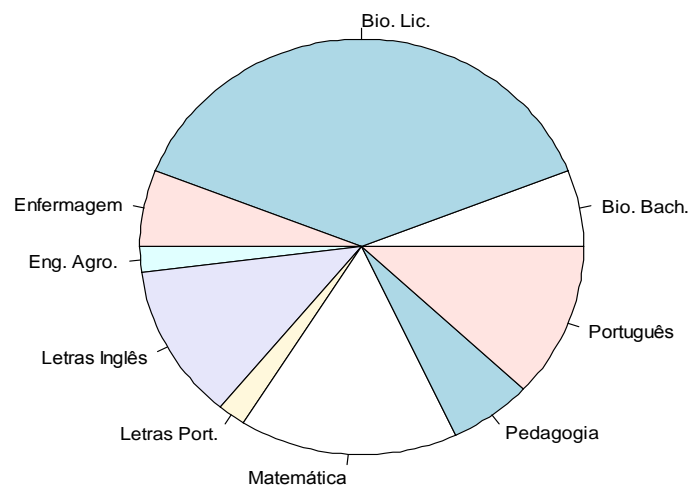
Fatores	Frequências	Freq. Rel.	Freq. Perc.	Freq. Acum.
Bio. Bach.	3	0,06	6	6
Bio. Lic.	19	0,38	38	44
Enfermagem	3	0,06	6	50
Eng. Agro.	1	0,02	2	52
Letras Inglês	6	0,12	12	64
Letras Port.	1	0,02	2	66
Matemática	8	0,16	16	82
Pedagogia	3	0,06	6	88
Português	6	0,12	12	100
Total	50			

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*.

5.6.1.2 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.

5.6.2 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.

5.6.2.1 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.

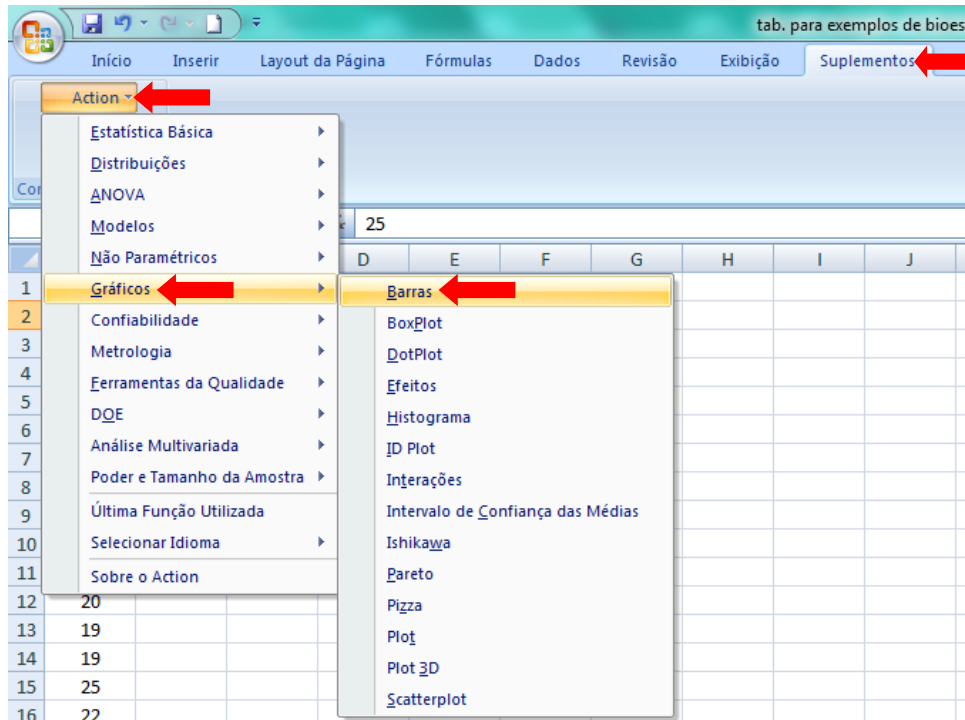
Para construirmos um gráfico em Barras para a variável “Idade” da tabela 8 abaixo, utilizando o *Action*, procedemos da seguinte forma:

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

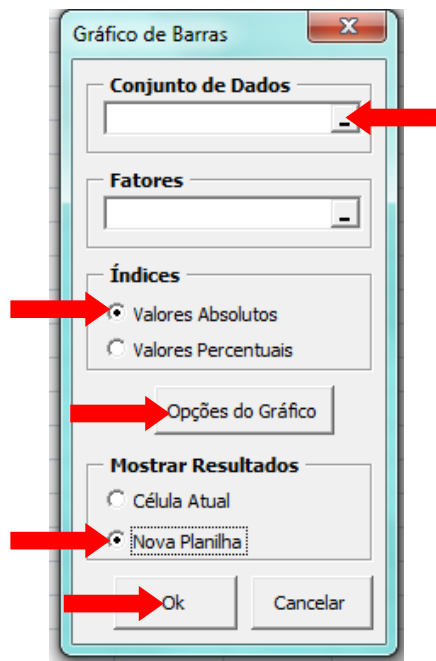
Idade (anos)	Nº de Filhos	Nº de Irmãos	Idade (anos)	Nº de Filhos	Nº de Irmãos
25	0	4	21	1	3
18	0	0	22	2	5
38	2	8	18	0	6
46	2	1	23	0	6
24	0	4	18	0	2
19	1	2	19	0	2
26	1	3	19	0	2
18	0	2	18	0	1
19	0	5	23	1	1
22	1	2	28	1	3
20	0	2	22	2	6
19	0	8	26	2	0
19	0	8	24	2	4
25	1	5	20	0	2
22	0	4	19	0	2
22	0	6	24	2	3
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.

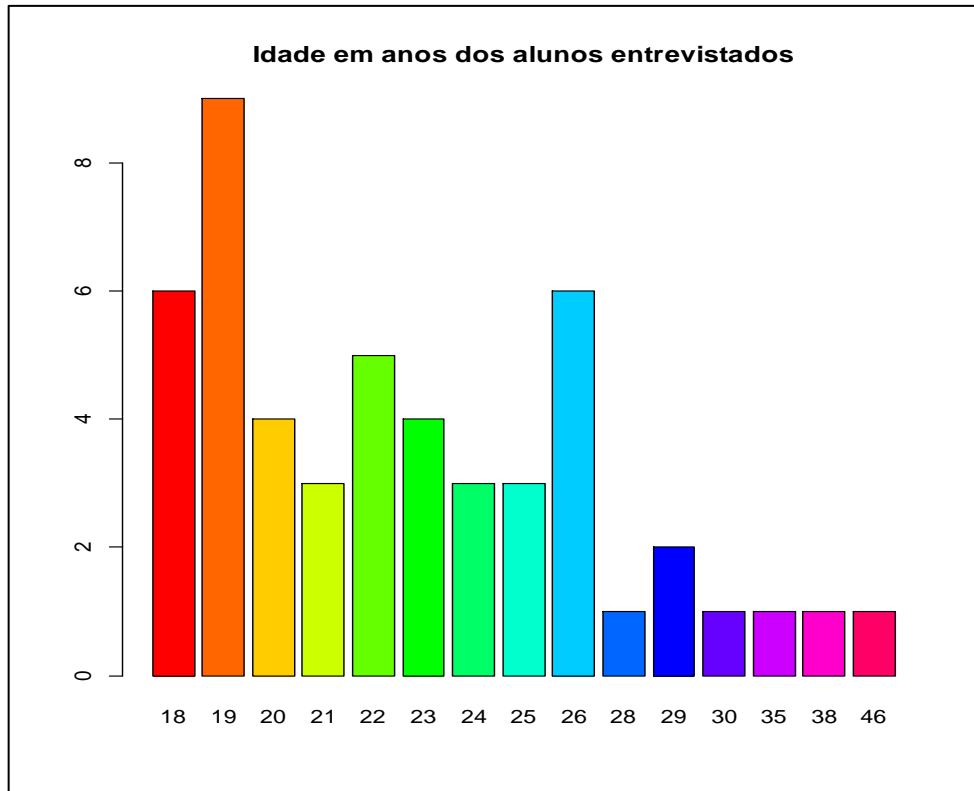
1. Digitamos os dados em uma coluna do Excel;
2. Clicamos em **Suplementos, Action, Gráficos e Barras**, conforme figura abaixo;



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.

5.6.2.2 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.

5.6.3 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).

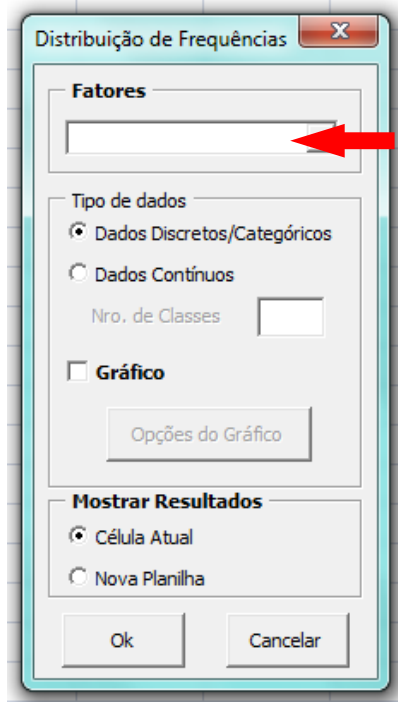
Para exemplificar, construiremos um histograma no Excel, utilizando o software Action dos dados coletados sobre o pH de lagos de meandro abandonado da região do Rio Purus – Acre.

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

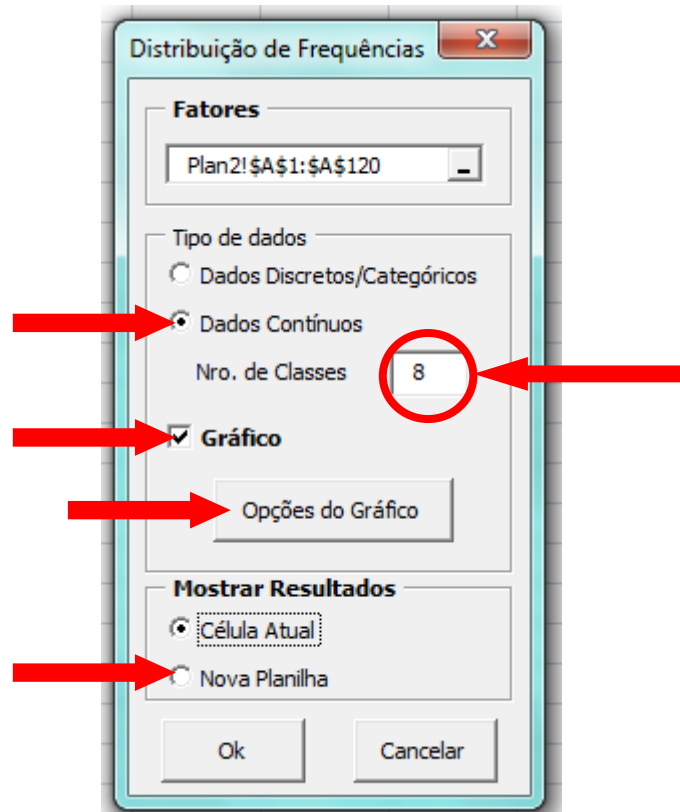
6,21	6,48	6,50	6,77	7,34	6,40
6,10	6,48	6,50	6,66	7,31	6,90
6,80	6,68	6,50	6,66	7,24	7,64
6,90	6,70	6,50	6,70	7,60	7,60
6,70	7,10	6,50	6,50	7,60	6,81
5,90	6,11	6,58	6,87	8,23	7,74
6,06	5,99	6,63	7,25	8,12	7,92
5,80	6,99	6,60	6,60	8,10	7,76
6,20	6,89	6,70	6,83	8,20	7,76
6,33	6,71	6,93	6,50	7,80	7,76
6,13	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

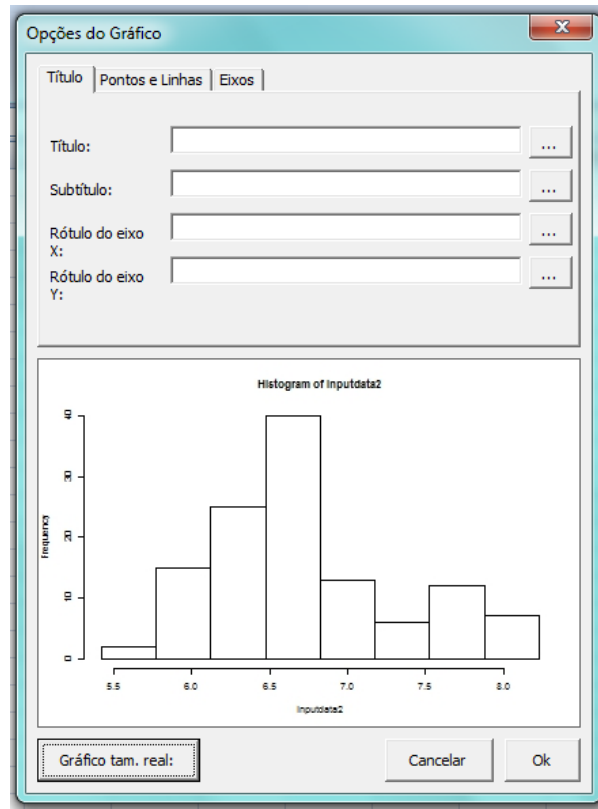
1. Digitamos os dados na coluna A do Excel.
2. Clique em (1) **Suplementos** na barra de ferramentas, em seguida em (2) **Action**, (3) **Estatística Básica** e (4) **Distribuição de Frequência**.
3. A janela de Distribuição de Frequência do *Action* será aberta. Clique na caixa fatores e selecione os dados da coluna A, clicando na célula A1 e arrastando o ponteiro do mouse até a célula A120.



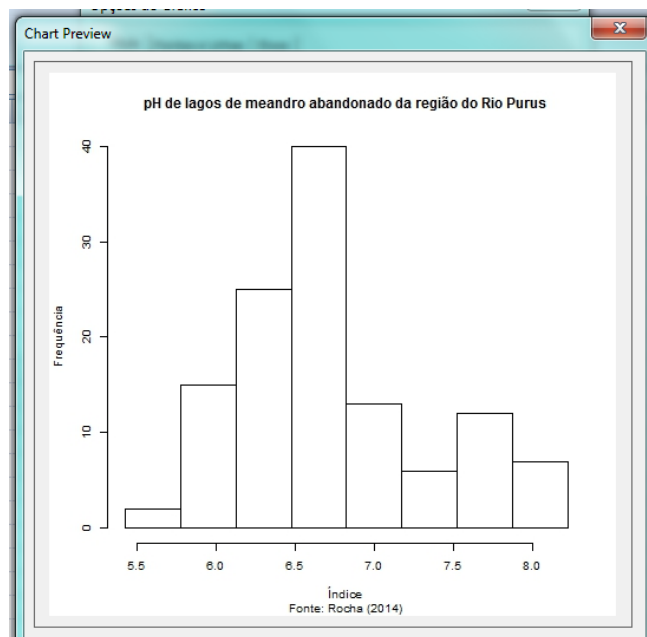
4. Clique em **Dados Contínuos** e escolha quantas classes terá a tabela. Neste exemplo, decidimos que nossa tabela terá oito linhas com intervalos de classe, então digitamos o número 8 na caixa **Nro. De Classe**, depois clicamos em **Gráfico** e em seguida em **Opções do Gráfico**.



5. 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*.

5.6.3.1 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.

5.7 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*).

5.7.1 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.

5.7.1.1 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.

$$m\acute{e}dia = \frac{Soma\ de\ todos\ os\ dados}{n\acute{u}mero\ de\ observa\c{c}o\es}$$

Símbolos matemáticos

Para representarmos esta fórmula usamos os seguintes símbolos matemáticos:

Média da população: μ (letra grega *miú* 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:

$$4,2 \quad 3,1 \quad 4,2 \quad 6 \quad 8,6$$

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 influencia 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 influencia dos valores extremos é a mediana.

5.7.1.2 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.

5.7.1.3 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

$$Mo = 5$$

5.7.2 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*.

5.7.2.1 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:

Local A: $a = 45 - 45 \quad a = 0$

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

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.

5.7.2.2 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
<hr/>	
$\sum x = 18$	$\sum x^2 = 88$

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.

5.7.2.3 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:

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

$$s = 2,4$$

5.7.2.4 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 10, procedemos da seguinte forma:

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

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

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

5.7.3 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.

5.7.3.1 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.

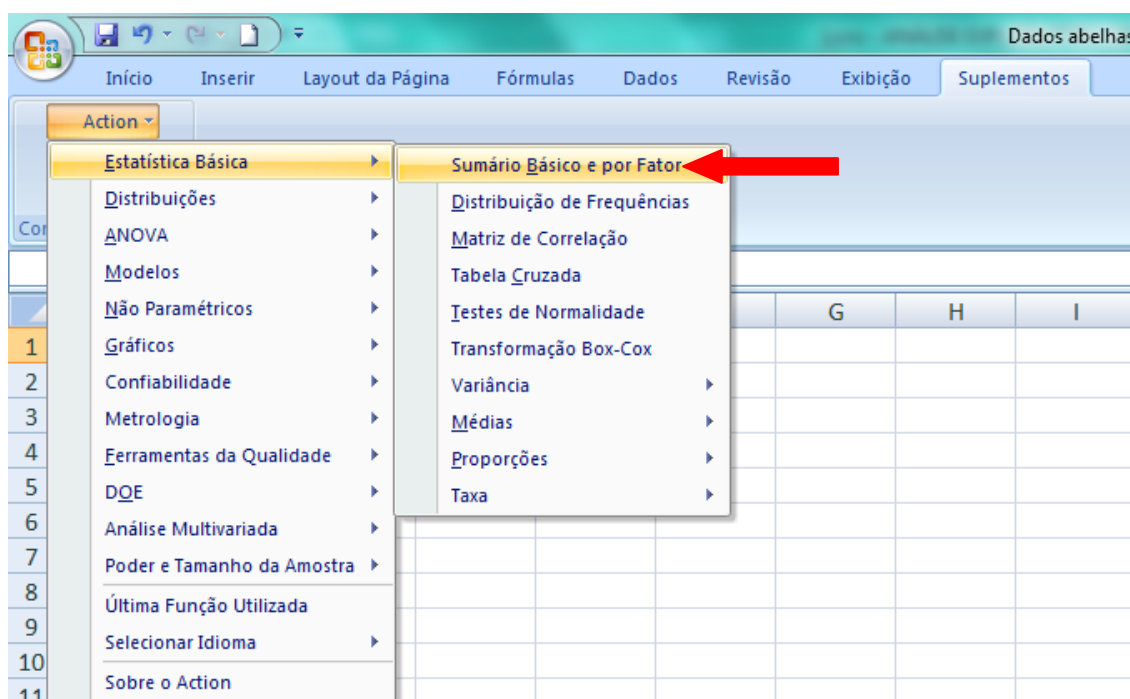
Tabela 11 – Dados de precipitação em *mm*, coletados entre os meses de agosto a dezembro de 2010, na região de Cruzeiro do Sul – AC.

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

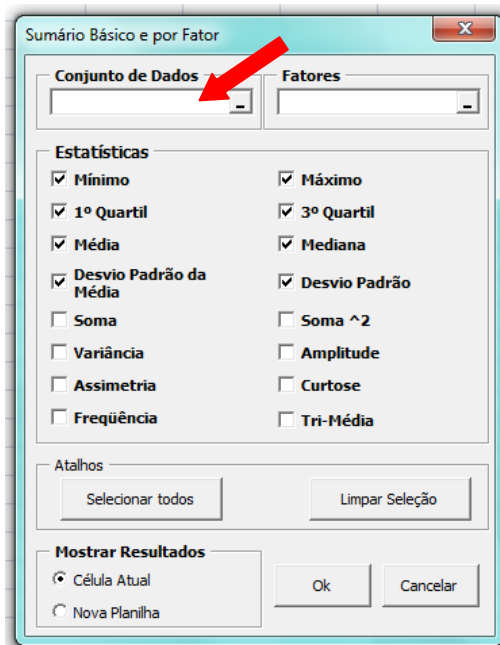
Fonte: Prof^o Rogério Oliveira Souza

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

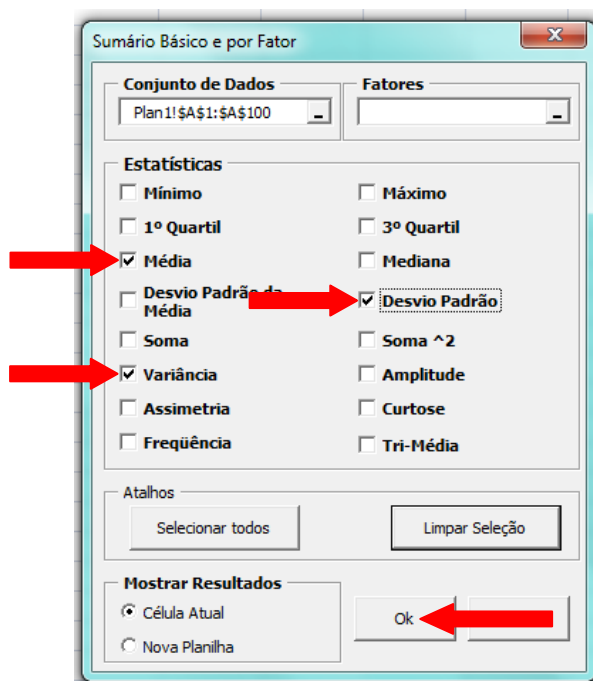
1. Digitamos os dados na coluna A do Excel.
2. Clique em Suplementos na barra de ferramentas, em seguida em *Action*, Estatística Básica e Sumário básico e por fator.



3. A janela de Sumário Básico e por Fator do *Action* será aberta. Clique na 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 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**.



5. 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*.

5.7.3.2 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*.

Veamos 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.

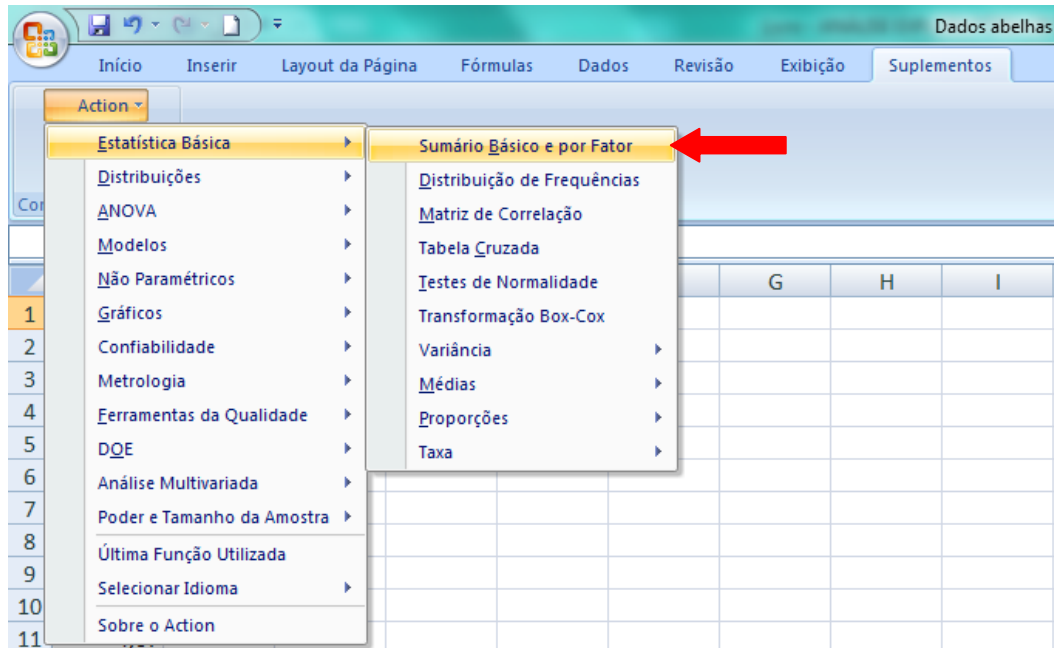
Lago	Temp.	Lago	Temp.	Lago	Temp.	Lago	Temp.
Itapira	26,12	Flor_ouro	26,41	Lago_Verde	31	Santana	36,5
Itapira	30,5	Flor_ouro	29,72	Lago_Verde	29,66	Santana	31,5
Itapira	29,2	Flor_ouro	30,1	Lago_Verde	30	Santana	27,8
Itapira	28,9	Flor_ouro	28,8	Lago_Verde	28,5	Santana	24
Itapira	25,34	Flor_ouro	27,7	Lago_Verde	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

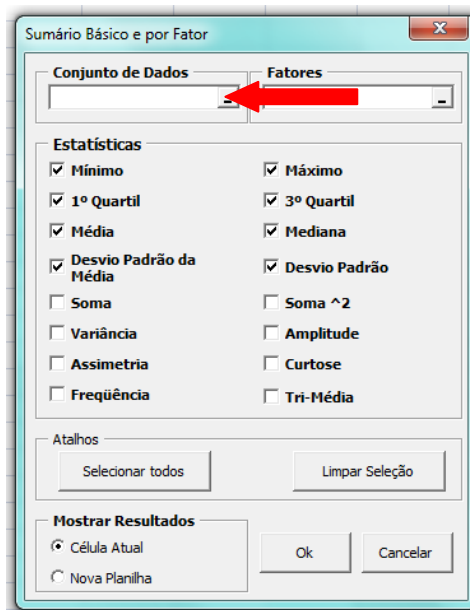
Fonte: Prof^ª. Lucena Rocha

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

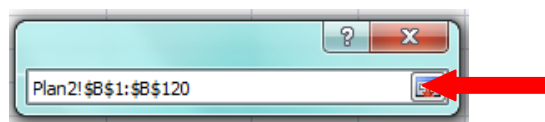
1. Digitamos os dados da variável lago na coluna A e os dados da variável temperatura na coluna B do Excel.
2. Clique em Suplementos na barra de ferramentas, em seguida em *Action*, Estatística Básica e Sumário básico e por fator.



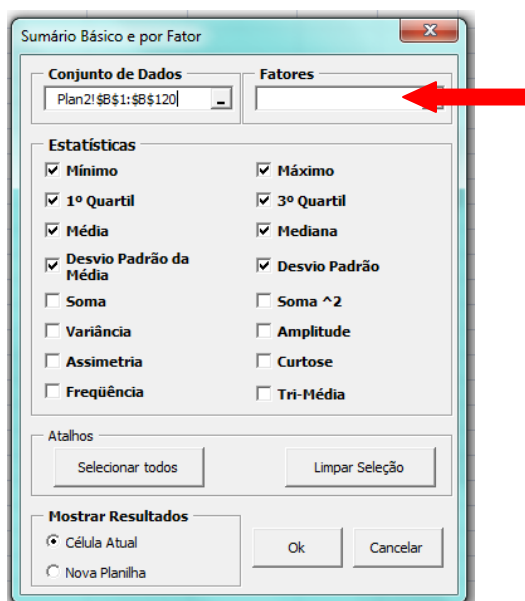
3. A janela de Sumário Básico e por Fator do *Action* será aberta. Clique no canto direito da caixa **Conjunto de Dados**, conforme indicado na figura abaixo. Em seguida, selecione os dados da coluna B, clicando na célula B1 e arrastando o ponteiro do mouse até a célula B120.



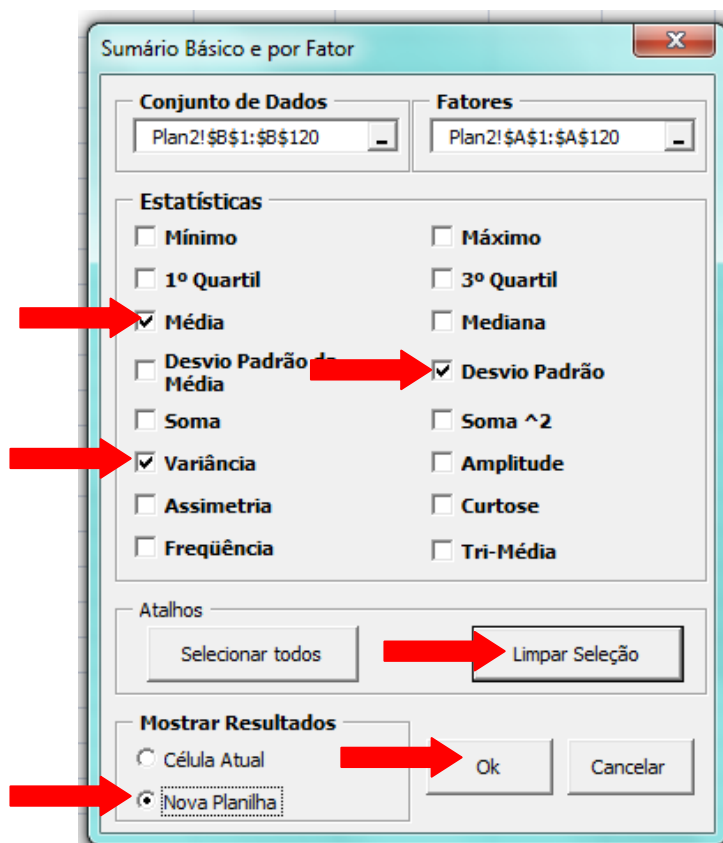
4. Clique novamente no canto direito da caixa **Conjunto de Dados**, conforme indicado na figura abaixo.



5. Clique na 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.



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*.

5.7.3.3 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.

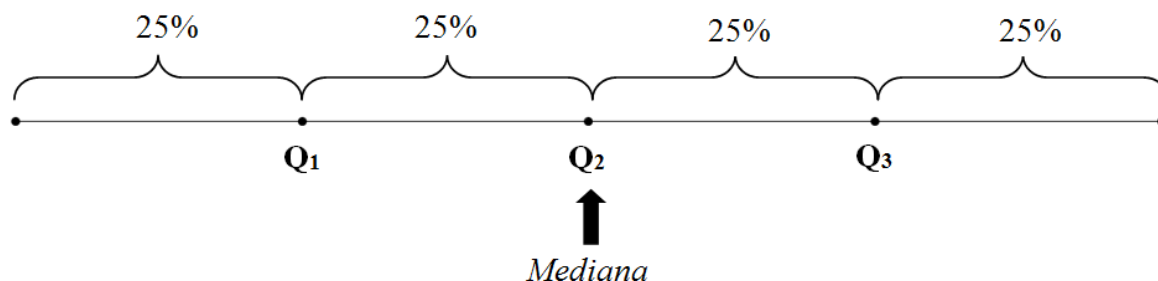
5.7.4 Quartil

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).

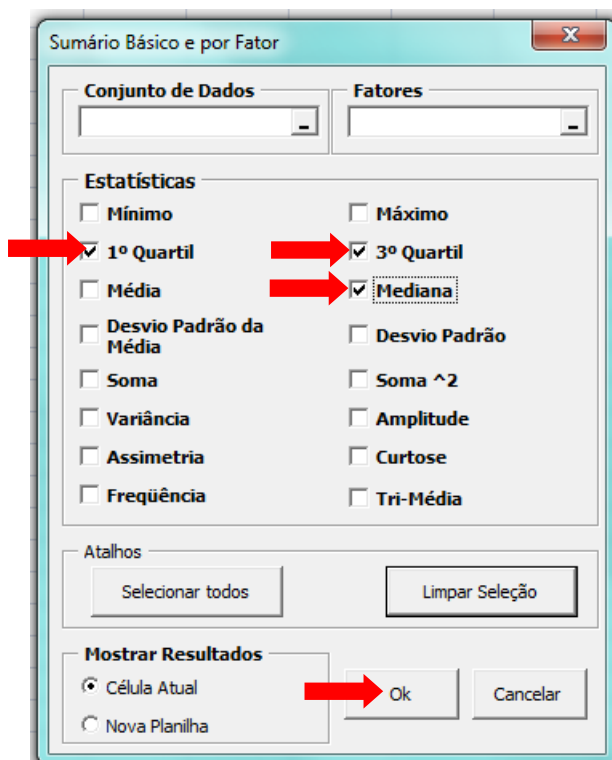
O primeiro quartil indica que 25% dos dados estão abaixo de seu valor e consequentemente 75% dos dados estão acima de seu valor.

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.



5.7.4.1 Exercício Proposto

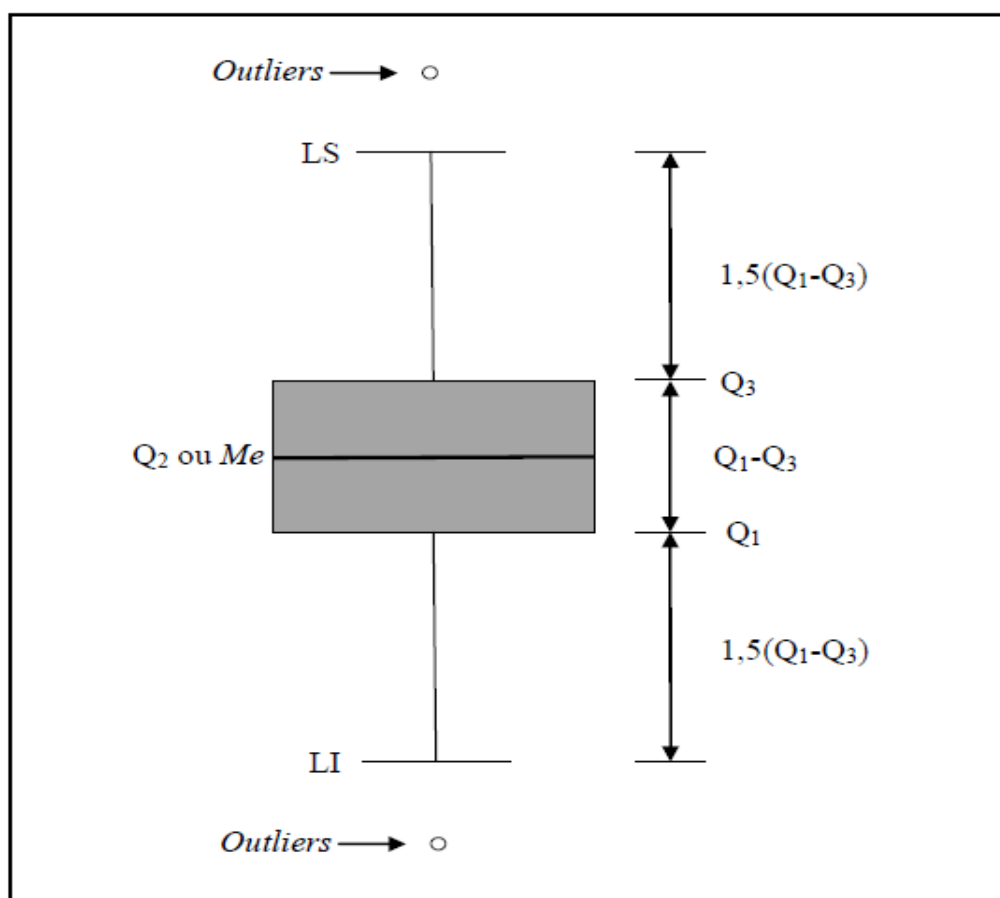
1. Para exercitar, calculemos os quartis dos dados sobre nitrogênio e fósforo da tabela do *Exercício Proposto* anterior.

5.7.5 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*⁸.

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.



⁸ 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 interquartilica.

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.

5.7.5.1 Gerando Box plot com o software Action

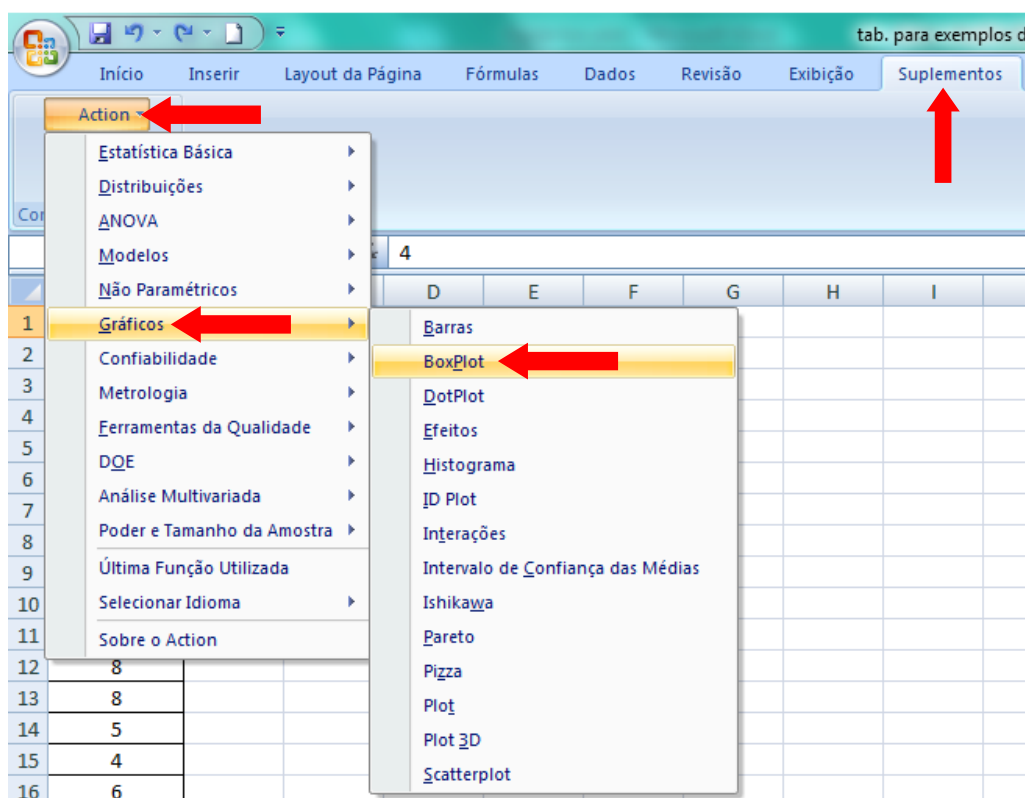
Para construirmos um Box plot utilizando o suplemento *Action* do *Excel*, procedemos da seguinte forma:

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

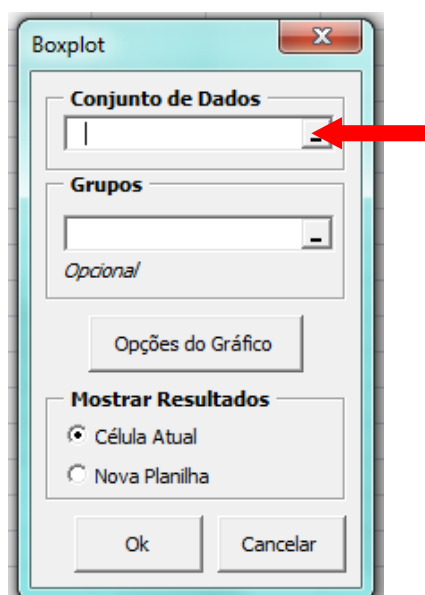
4	2	7	2	3
0	8	4	2	4
8	8	4	1	2
1	5	2	1	2
4	4	4	3	1
2	6	3	6	10
3	16	5	0	2
2	2	6	4	2
5	3	6	2	6
2	2	2	2	12

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

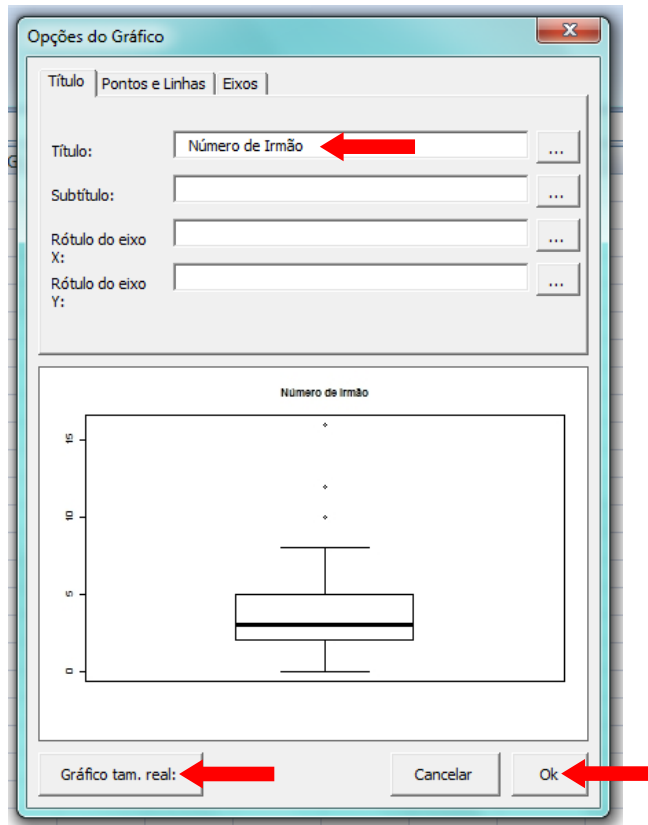
1. Digitamos os dados da Tabela 13 em uma coluna do Excel.
2. Clicamos em **Suplementos** na barra de ferramentas, em seguida, **Action**, **Gráficos** e **BoxPlot**



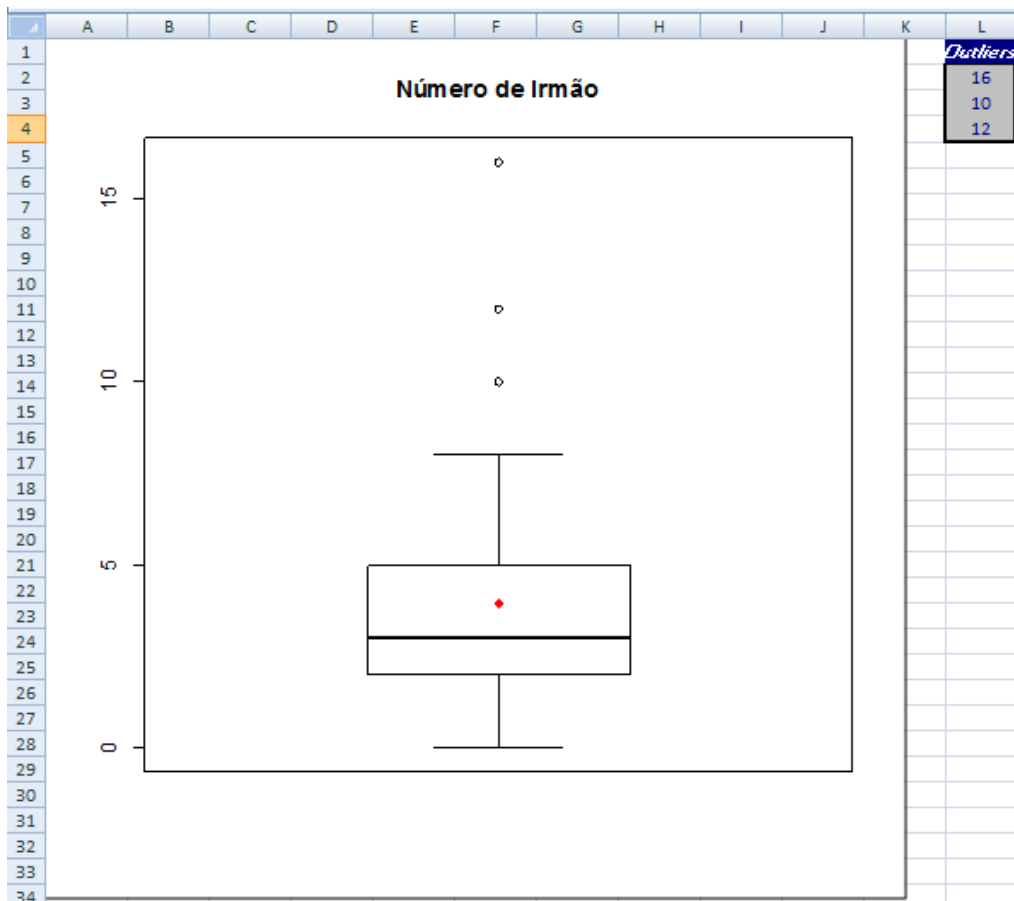
3. A janela *Boxplot* será aberta, clicamos na caixa *Conjunto de dados* e selecionamos os dados sobre o número de irmão.



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.

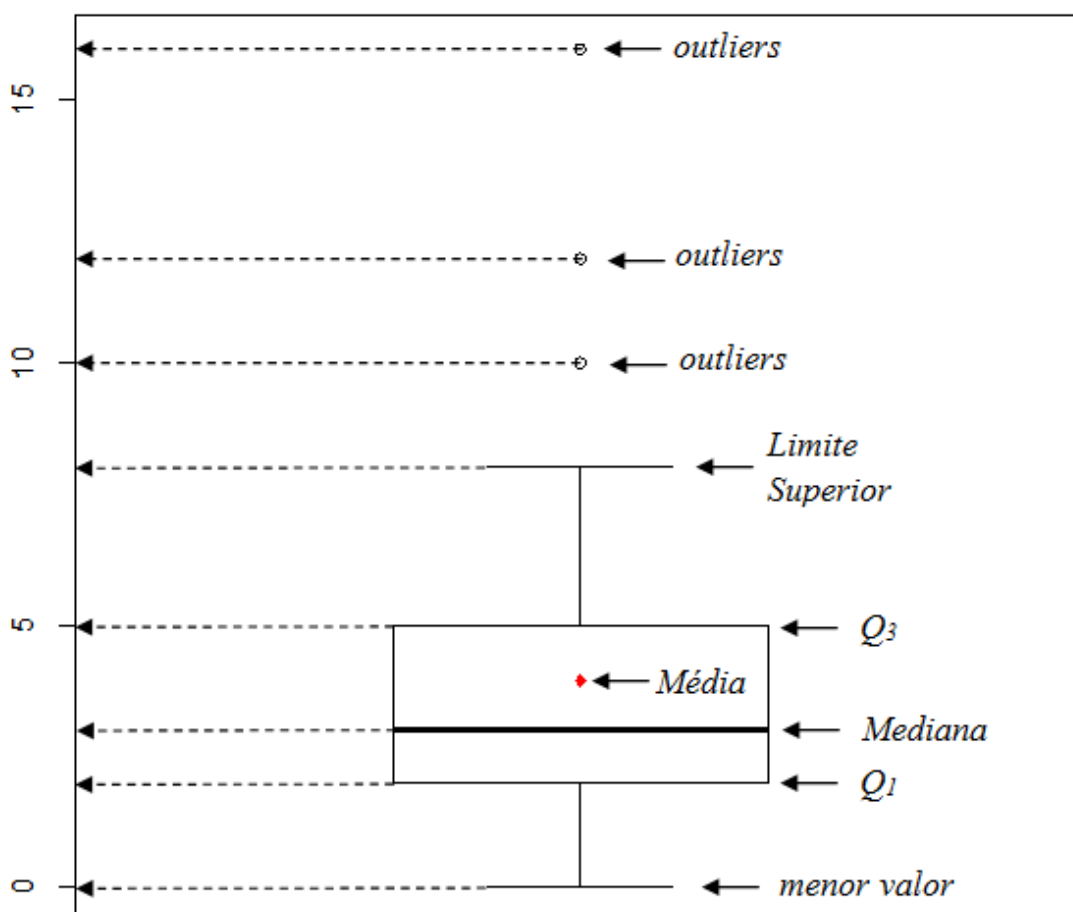
[Clique aqui](#) e assista ao vídeo tutorial sobre como gerar o *Box plot* utilizando o suplemento *Action* do *Excel*.

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.

5.7.5.2 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.

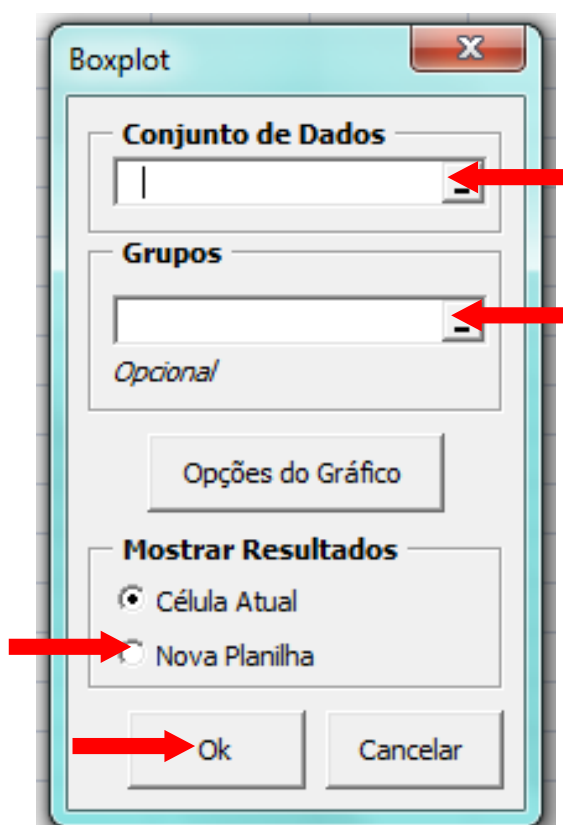
Tabela 14 – Dados sobre peso em kg de homens e mulheres estudantes da UFAC, em Cruzeiro do Sul – Acre

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		

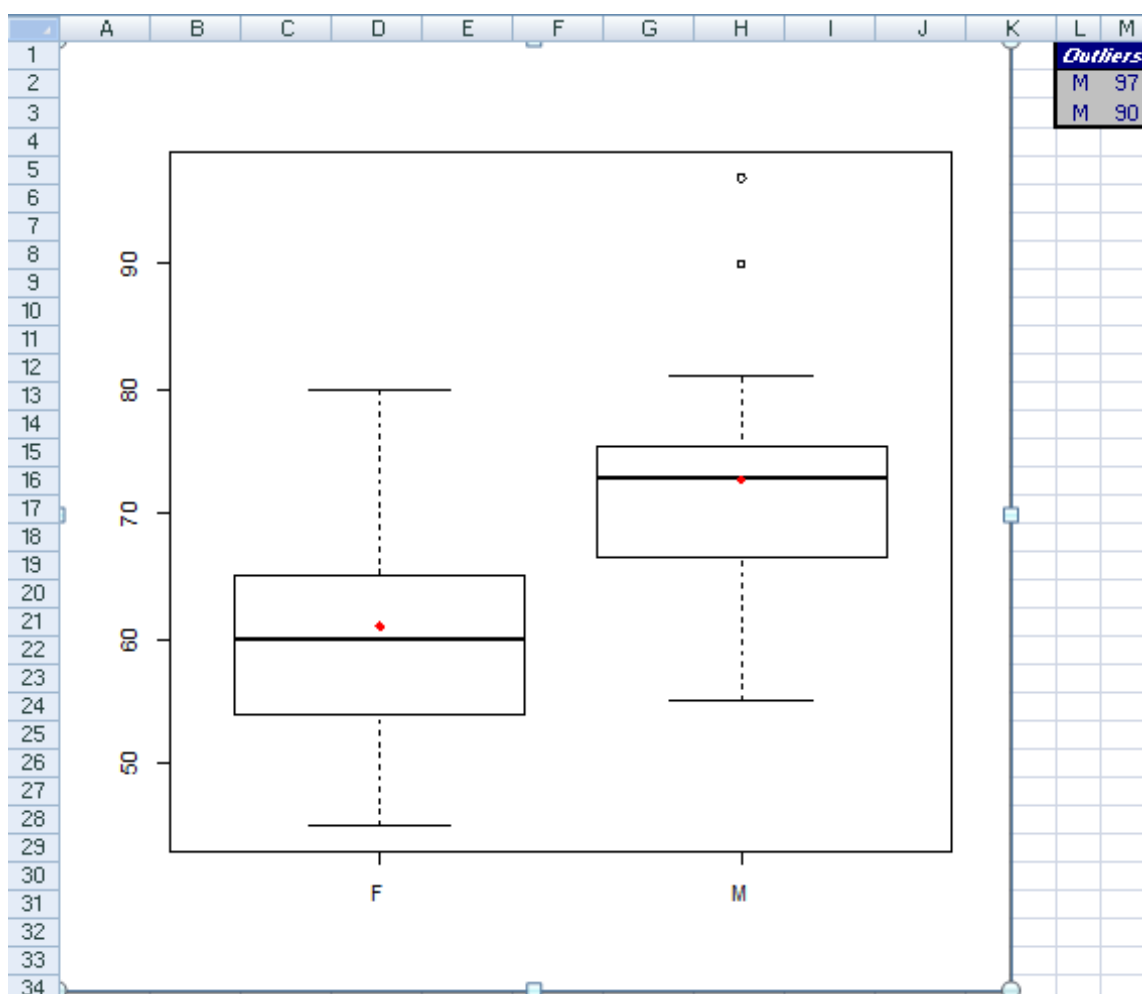
Fonte: Alunos do 3º período do Curso de Bacharelado em Ciências Biológicas-UFAC.

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 *box plot* conforme a figura a seguir, comparando os dois grupos através do gráfico de caixas, no canto superior esquerda da planilha do *Excel*, contam os dados discrepantes para os dois grupos.



[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.

5.7.5.3 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.

5.8 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*.

5.8.1 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:

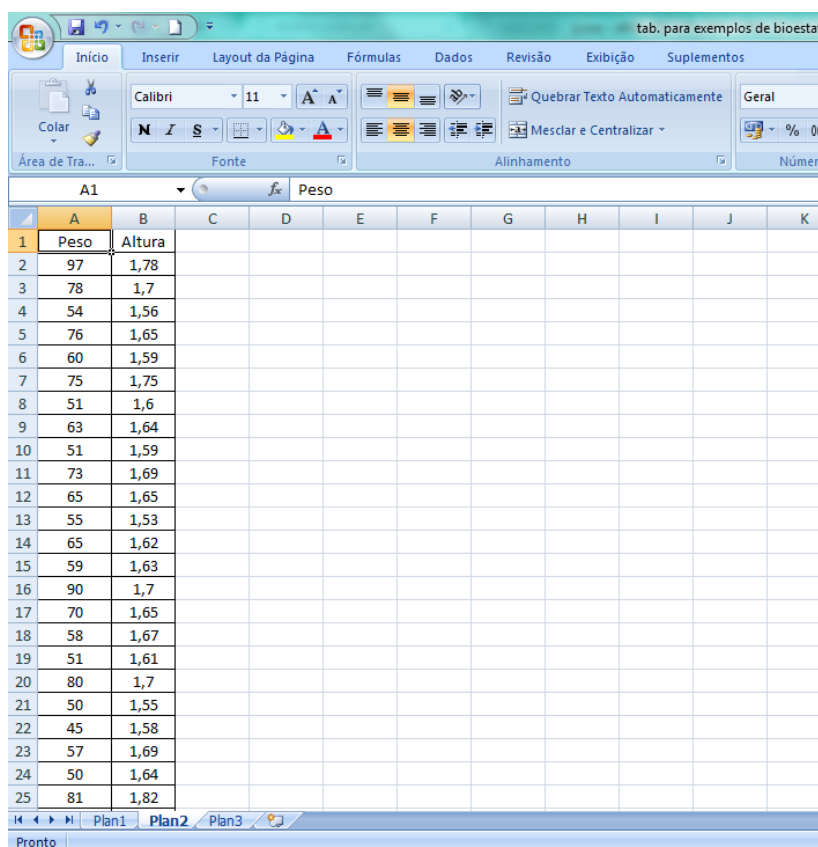
Tabela 15 – Dados sobre peso (Kg) e altura (m) de 50 alunos da Universidade Federal do Acre – UFAC, coletados na cidade de Cruzeiro do Sul – AC, em 2015.

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.

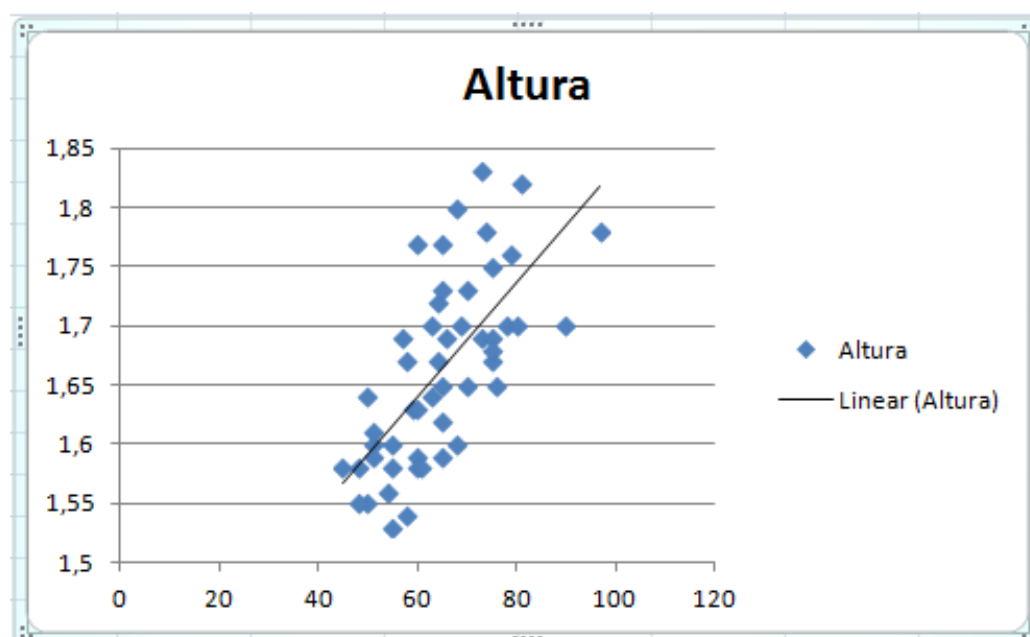
1. Digitamos os dados sobre o peso dos alunos na coluna A e os dados sobre altura na coluna B do *Excel*.



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**.

	A	B	C	D	E	F	G	H	L
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							

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*.

5.8.2 Coeficiente de Correlação

Existe uma medida para determinarmos o grau de correlação entre duas variáveis, essa medida é o *Coeficiente 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:

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

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

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

$$\mathbf{r = 0,98}$$

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.

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

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

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

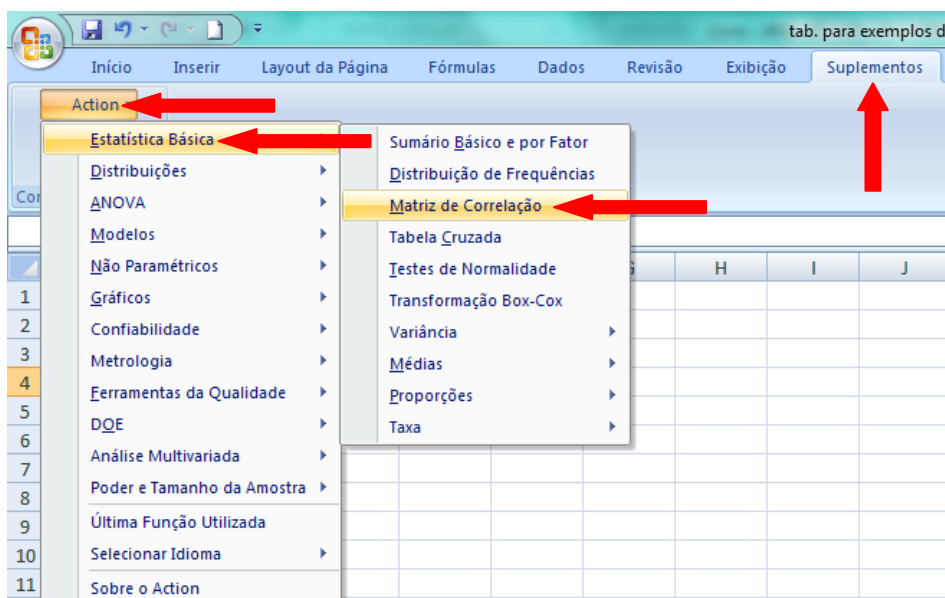
$$r = -0,95$$

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.

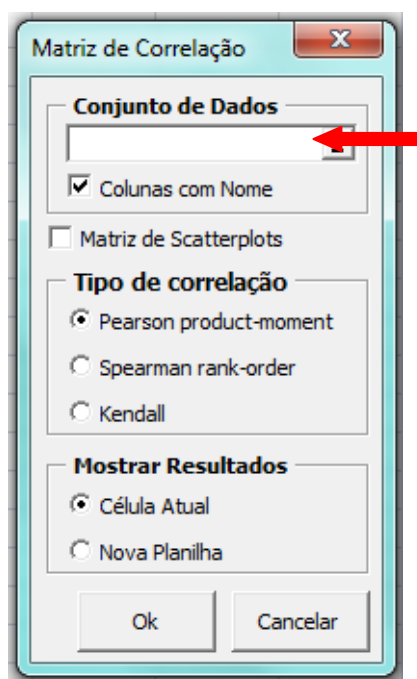
5.8.2.1 Cálculo do coeficiente de correlação com o *Software Action*.

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.

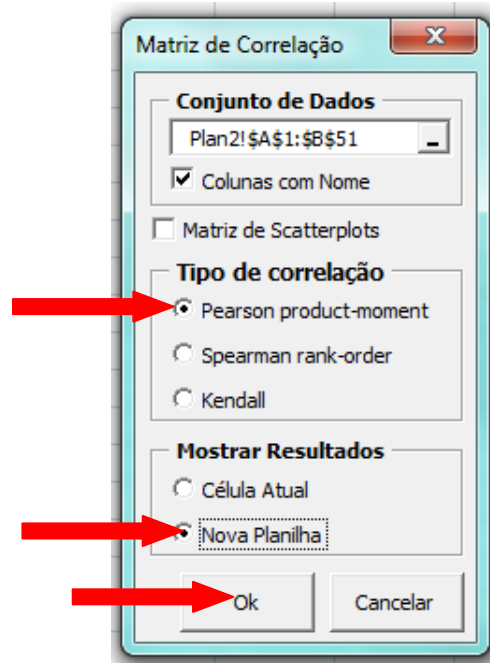
1. Digitamos os dados sobre o peso dos alunos na coluna A e os dados sobre altura na coluna B do *Excel*, semelhante ao realizado na construção de diagramas de dispersão.
2. Clique no menu **Suplementos** na barra de ferramentas, selecione o comando, **Action, Estatística Básica** e em seguida **Matriz de Correlação**.



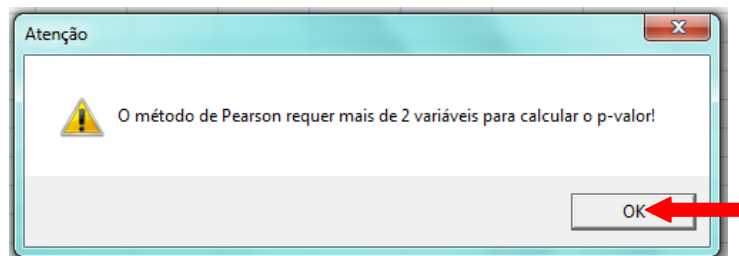
3. A janela de Matriz de Correlação será aberta, clique na caixa **Conjunto de Dados** e selecione os dados das duas colunas, inclusive as células com os nomes das variáveis.



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		Peso	Altura
7	Peso	1	0,681749646
8	Altura	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*.

5.8.2.2 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

5.9 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.

6. CONSIDERAÇÕES FINAIS

Nossa pesquisa foi realizada buscando como referenciais teóricos uma metodologia pedagógica baseada na teoria construtivista social, para a construção de uma proposta de ensino de Bioestatística em ambientes virtuais, utilizando *softwares* como ferramentas de apoio ao ensino. Para isso, usamos das ideias de autores que defendem a utilização de ambientes informatizados como forma de alteração na formação do indivíduo.

A hipótese de nosso estudo se propôs a investigar se as tecnologias de informação e comunicação podem contribuir como ferramentas para amenizar algumas dificuldades dos alunos nas aulas de Estatística, tornando o trabalho do professor supostamente mais dinâmico.

A redução do tempo na realização dos cálculos estatísticos é uma vantagem que pode ser evidenciada com a utilização de *softwares* no ensino de Estatística, amenizando o problema da priorização do cálculo sobre a interpretação e análise dos resultados.

O ambiente virtual de aprendizagem plataforma *Moodle* se propõe a servir como uma extensão da sala de aula, um lugar onde os alunos poderão tirar suas dúvidas com o professor e com os próprios colegas através de ferramentas como o fórum e o *chat*, possibilitando diminuir problemas de interação entre alunos e professores.

Nossa questão de investigação era como organizar um curso presencial de Bioestatística em ambientes virtuais de aprendizagem como a plataforma *Moodle* para alunos da graduação, utilizando materiais didáticos com recursos digitais e com a utilização do software estatístico *Action*.

Os produtos apresentados neste trabalho são uma forma de contribuir com o ensino e aprendizagem em Estatística. A pesquisa serviu principalmente para modificar nosso entendimento sobre o processo de aquisição do conhecimento, sobre as ideias que estão sendo apresentadas nos estudos desta área e sobre como podemos modificar nossa postura como professores de uma geração caracterizada pela manipulação tecnológica.

No decorrer da pesquisa tivemos algumas dificuldades que influenciaram no resultado final dos produtos apresentados, como por exemplo, a utilização de dados reais de pesquisas em áreas biológicas realizadas por pesquisadores da UFAC – Campus Floresta. Apesar de termos entrado em contato com todos os docentes do curso de Biologia, apenas dois professores se dispuseram a disponibilizar dados brutos coletados em pesquisas. Isso prejudicou a riqueza de exemplos voltados para a área biológica no livro *Análise Exploratória de Dados Biológicos*.

Não foi possível aplicar a versão final dos produtos educacionais propostos nesta pesquisa nas turmas de Bacharelado e Licenciatura em Ciências Biológicas da UFAC – Campus Floresta, por motivos temporais. No Entanto, nos propomos a realizar estudos sobre a impressão dos discentes com a utilização destas ferramentas de apoio ao ensino, que serão objetos de artigos científicos posteriormente.

Esperamos, assim, ter contribuído com nossa pesquisa para incentivar a utilização das tecnologias da informação e comunicação no ensino de estatística aplicado as áreas biológicas, proporcionando novas possibilidades de investigação nesta área do conhecimento.

7. REFERÊNCIAS

ALVES, G. S.; SOARES, A. B. **Geometria Dinâmica: um estudo de seus recursos, potencialidades e limitações através do software *Tabulae***. XXIII Congresso da Sociedade Brasileira de Computação. Campinas, 2003.

BATANERO C.; ESTEPA A.; GODINO J. D. *Analisis exploratorio de datos: sus posibilidades en la enseñanza secundaria*. Suma nº 9, 1991. Disponível em: <http://www.researchgate.net/profile/Carmen_Batanero/publication/255765681_ANALISIS_E_XPLORATORIO_DE_DATOS_SUS_POSIBILIDADES_EN_LA_ENSEANZA_SECUNDA_RIA/links/02e7e520a590f13ba1000000.pdf> Acesso em: 16 abr. 2015, 22:30:30.

BORBA, M. C.; PENTEADO, M. G. **Informática e educação matemática**. Coleção: Tendências em Educação Matemática. 3 ed. Belo Horizonte: Autêntica Editora, 2003.

BORBA, M. C. *Humans-With-Media and continuing education for mathematics teachers in on-line environments*. ZDM Mathematics Education, 2012.

DAROIT, L. **Utilização de Ferramentas Computacionais Visando a uma Aprendizagem Significativa em Bioestatística**. Dissertação (Mestrado em Ensino de Ciências Exatas) – Centro Universitário Univates, Lajeado, 2009.

DAWBOR, L. **Educação, tecnologia e cidadania**. Revista do Cogei, [s.i.], n. 8, p. 9-22, jun. 1996.

DELGADO, L. M. M. **Uso da Plataforma Moodle como Apoio ao Ensino Presencial: Um Estudo de Caso**. Dissertação (Programa de Interdisciplinar Pós-graduação em Linguística Aplicada) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

DOUGIAMAS, M. *An exploration of the use of an open source software called Moodle to support a social constructionist epistemology of teaching and learning within Internetbased communities of reflective inquiry*. Tese de Doutorado. Science and Mathematics Education Centre Curtin University of Technology Perth, Western Australia, 1999.

DRISCOLL, M. P. *Psychology of learning and instruction*. Boston, MA: Allyn and Bacon. p. 409, 1995.

FIORENTINI, D. **Alguns Modos de Ver e Conceber o Ensino da Matemática no Brasil**. Revista Zetetiké, ano 3, nº 4, p. 4-5. 1995.

FREIRE, F. M. P. et al. **A Implantação da Informática no Espaço Escolar: questões emergentes ao longo do processo**. Revista Brasileira de Informática na Educação, Santa Catarina, n. 3, p 45-62, set. 1998.

GAL, I. *Adults' statistical literacy: meanings, components, responsibilities*. International Statistical Review, Netherlands: ISI, v. 70, n.1, p. 1-50, abr. 2002.

GRACIAS, T. A. S. **A reorganização do pensamento em um curso a distância sobre Tendências em Educação Matemática**. 2003. Tese (Doutorado em Educação Matemática) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2003.

GUZZO, R. S. L. **Psicologia em instituições escolares e educativas: apontamentos para um debate**. In Ano da psicologia na educação. Textos geradores (pp. 53-61). Brasília, DF: Conselho Federal de Psicologia. 2008.

LOUREIRO, C. **Computadores no ensino de Geometria**. In: Ensino de Geometria no virar do milênio. p. 43-50, 2001.

LUCAS, R. D. **GeoGebra e Moodle no Ensino de Geometria Analítica**. Dissertação (Mestrado Profissional em Ensino de Ciências Exatas) – Universidade Federal de São Carlos, São Carlos, 2009.

MAIA, R. D. **Introdução à Análise Exploratória de Dados**. Notas de Aula. Universidade Federal do Minas Gerais, Belo Horizonte, 2011.

MIRANDA, D. F. e LAUDARES, J. B. Informatização no Ensino de Matemática: Investindo no Ambiente de Aprendizagem. São Paulo. Zetetiké, vol. 15. nº 27. jul. 2007.

MOREIRA, M. A. **Subsídios Teóricos para o Professor Pesquisador em Ensino de Ciências: Comportamentalismo, Construtivismo e Humanismo**. Porto Alegre, Ed do Autor, 2009.

RESENDE, A. C. A. **Subjetividade: a contribuição da psicanálise ao debate**. In: Miranda, M. G; Resende, A. C. A. (orgs.). **Escritos de Psicologia, Educação e Cultura**. Goiânia, Editora da Universidade Católica, 2009.

ROSA, M. **Role Playing Game Eletrônico: uma tecnologia lúdica para aprender e ensinar matemática**. 2004. Dissertação (Mestrado em Educação Matemática) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2004.

ROSA, M. e OREY, D. C. **O Construtivismo Como um Embasamento Teórico-Filosófico para o Ambiente Virtual Moodle de Aprendizagem**. X Congresso Brasileiro de Ensino Superior a Distância. Belém. jun. 2013.

SILVA, N. O. **Educação Matemática: epistemologia, princípios e evolução**. In texto elaborado para o Curso de Especialização em Educação Matemática do NPADC/UFPA, 2001.

VALENTE, J. A. **O uso inteligente do computador na educação**. Pátio Revista Pedagógica. Editora: Artes Médicas Sul, ano 1, no 1, págs. 19-21. 1997.

VALENTE, J. A. **Diferentes Usos do Computador na Educação**. In: Valente (org). **Computadores e Conhecimento: Repensando a Educação**. Campinas, 1998.

VALENTE, J. A.; MOREIRA, P.; DIAS, P. **Moodle: moda, mania ou inovação na formação?** In Alvez, L., Barros, D., Okada, A. (Orgs.). **Moodle: estratégias pedagógicas e estudo de caso**. Salvador, BA: EDUNEB, 2009. p 35-54.

VYGOTSKY, L. S. *Thought and language*. Cambridge, MA: MIT Press, 1984.

VYGOTSKY, L. S. **A Formação Social da Mente: o desenvolvimento dos processos psicológicos superiores**. ed. 7. Martins Fontes, São Paulo. 1984.

ZULATTO, R. B. A. **Concepções norteadoras do *design* de um curso a distância em Matemática**. In: ENCONTRO BRASILEIRO DE ESTUDANTES DE PÓSGRADUAÇÃO EM EDUCAÇÃO MATEMÁTICA, 10, 2006, Belo Horizonte. *Anais...* Belo Horizonte: Universidade Federal de Minas Gerais. 2006.

Apoio dos Colegas

Respostas	Ainda não respondeu	Quase nunca	Raramente	Algumas vezes	Freqüentemente	Quase sempre
Neste curso...						
17 Os outros participantes me encorajam a participar.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18 Os outros participantes elogiam as minhas contribuições.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19 Os outros participantes estimam as minhas contribuições.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20 Os outros participantes demonstram empatia quando me esforço para aprender.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Compreensão

Respostas	Ainda não respondeu	Quase nunca	Raramente	Algumas vezes	Freqüentemente	Quase sempre
Neste curso...						
21 Eu compreendo bem as mensagens dos outros participantes.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22 Os outros participantes compreendem bem as minhas mensagens.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23 Eu compreendo bem as mensagens do tutor.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24 O tutor compreende bem as minhas mensagens.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

25 Em quanto tempo você concluiu este questionário?

Escolher... ▼

26 Você quer fazer outros comentários?

Continuar

ANEXO II

Plano de Curso de Bioestatística I

1. Ementa

Dados e variáveis. Apresentação tabular e gráfica de dados. Medidas descritivas. Probabilidades. Modelos e distribuições de probabilidades. Estimação de parâmetros. Teste de hipóteses sobre parâmetros. Análise de contingência. Normalidade e homocedasticidade. Transformações de dados. O teste t. Delineamentos. Análise de variância. Teste de Tukey. Análise de correlação. Regressão linear simples. Uso de programas estatísticos.

2. Objetivo Geral

Despertar no aluno, o interesse pela estatística, principalmente no que diz respeito ao uso dos conceitos da disciplina para as aplicações dentro das ciências biológicas e também em outros ramos científicos.

3. Objetivos Específicos

Compreender os conceitos básicos e os objetivos da bioestatística, de modo a auxiliá-lo no planejamento de pesquisas e exercício de suas habilidades profissionais;

Compreender os conceitos de Estatística Descritiva, abordando os conceitos que proporcionem informações pertinentes e interessantes ao aluno;

Conduzir o acadêmico para a utilização da teoria das probabilidades, levando a levantar hipóteses e fazer previsões sobre objetos específicos de pesquisa;

Conduzir o acadêmico a construir uma visão crítica dos dados pesquisados, buscando conclusões futuras através dos testes estatísticos e teoria da estimação;

Analisar dados presentes e produzir estimativas futuras através de regressão e correlação, obtendo modelos estatísticos eficientes.

4. Conteúdo Programático

4.1 Unidade Temática I – A Natureza da Estatística

4.1.1 – Panorama histórico;

4.1.2 – Método estatístico;

4.1.3 – Fases do método estatístico.

4.2 Unidade Temática II – População e Amostra

4.2.1 – Variáveis;

4.2.2 – População e amostra;

4.2.3 – Técnicas de amostragem.

4.3 Unidade Temática III – Apresentação de Dados em Tabelas

4.3.1 – Componentes das tabelas;

4.3.2 – Tabelas de contingência;

4.3.3 – Tabelas de distribuição de frequência.

4.4 Unidade Temática IV – Apresentação de Dados em Gráficos

4.4.1 – Gráfico de Barras;

4.4.2 – Gráfico de setores;

4.4.3 – Histograma;

4.4.4 – Polígono de frequência.

4.5 Unidade Temática V – Medidas de Tendência Central

4.5.1 – Média aritmética;

4.5.2 – Média de dados em tabelas de distribuição de frequência;

4.5.3 – Mediana;

4.5.4 – Moda.

4.6 Unidade Temática VI – Medidas de Dispersão para uma Amostra

4.6.1 – Amplitude;

4.6.2 – Variância;

4.6.3 – Desvio padrão;

4.6.4 – Coeficiente de variação.

4.7 Unidade Temática VII – Noções sobre Correlação

4.7.1 – Diagrama de dispersão;

4.7.2 – Correlação positiva e correlação negativa;

4.7.3 – Coeficiente de correlação.

5. Procedimentos Metodológicos

Aulas expositivas dialogadas, leituras analíticas, avaliações individuais e em equipe, vídeos e pesquisas.

6. Recursos Didáticos

Material de uso comum, textos e livros, datashow, notebook, softwares estatísticos dentre outros.

7. Bibliografia Básica

BEIGUELMAN, B. Curso Prático de Bioestatística. Ribeirão Preto: Fundação de Pesquisas Científicas de Ribeirão Preto, 2002.

CALLEGARI-JACQUES, S. M. Bioestatística: Princípios e Aplicações. Porto Alegre: Artmed, 2003. CRESPO, A. A. Estatística Fácil. São Paulo: Saraiva, 1991.

NAGHETTINI, M. C.; PINTO, E. J. A. Hidrologia Estatística. Belo Horizonte: CPRM, 2007.

Bibliografia Complementas

ANDRADE, D. F.; OGLIARI, P. F. Estatística Aplicada às Ciências Agrárias e Biológicas. Florianópolis: Editora da UFSC, 2001.

LEVINE, D. M. et al. Estatística: Teoria e aplicações. Rio de Janeiro: LTC, 2011.

OLIVEIRA, Francisco Estevam M. Estatística e Probabilidade. 2.ed. São Paulo: Atlas, 1999.

RIBEIRO JUNIOR, J. I. Análises estatísticas no Excel: guia prático. Viçosa: UFV, 2004.

SPIEGEL, M. R. Estatística. São Paulo: Makron Books, 1993.

VIEIRA, S. Introdução à Bioestatística. 3.ed. Rio de Janeiro: Campus, 1980.

VIEIRA, S. Análise de variância (Anova). São Paulo: Atlas, 2006