UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

INSTITUTO DE MATEMÁTICA

PROGRAMA DE PÓS GRADUAÇÃO

MATEMÁTICA EM REDE NACIONAL

MESTRADO PROFISSIONAL

ANDRÉ LUIZ SÁ FIRMINO

ESTATÍSTICA COM EXCEL E APLICAÇÕES

CAMPO GRANDE-MS

FEVEREIRO DE 2015

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

INSTITUTO DE MATEMÁTICA

PROGRAMA DE PÓS GRADUAÇÃO

MATEMÁTICA EM REDE NACIONAL

MESTRADO PROFISSIONAL

ANDRÉ LUIZ SÁ FIRMINO

ANDRÉ LUIZ SÁ FIRMINO

ESTATÍSTICA COM EXCEL E APLICAÇÕES

Orientadora: Prof.ª Dr.ª Rúbia Mara de Oliveira Santos

Dissertação apresentada ao Programa de Pós-Graduação em Matemática em Rede Nacional do Instituto de Matemática, da Universidade Federal do Mato Grosso do Sul, como parte dos requisitos para obtenção do Título de Mestre.

CAMPO GRANDE-MS

FEVEREIRO DE 2015

ESTATÍSTICA COM EXCEL E APLICAÇÕES

ANDRÉ LUIZ SÁ FIRMINO

Dissertação submetida ao Programa de Mestrado em Matemática em Rede Nacional do Instituto de Matemática, da Universidade Federal de Mato Grosso do Sul, como parte dos requisitos para obtenção do título de Mestre.

Aprovado pela Banca Examinadora:

Prof.ª Dr.ª Rúbia Mara de Oliveira Santos - UFMS
Prof.ª Dr.ª Selma Helena Marchiori Hashimoto - UFGD
Prof.ª Dr.ª Lilian Milena Ramos Carvalho - UFMS
Prof.ª Dr.ª Fabiana Villa Alves - EMBRAPA GADO DE CORTE

CAMPO GRANDE-MS

FEVEREIRO DE 2015

Ao meu pai José Edésio Firmino (in memorian) e
a minha mãe Maurisbela de Sá Firmino, dedico.

Epígrafe

Sabemos que todas as coisas cooperam para o bem daqueles que amam a Deus, daqueles que são chamados segundo o seu propósito.

Romanos 8:28.

AGRADECIMENTOS

> "Porque, assim como descem a chuva e a neve dos céus e para lá não tornam, sem que primeiro reguem a terra, e a fecundem, e a façam brotar, para dar semente ao semeador e pão ao que come, assim será a palavra que sai da minha boca: não voltará vazia, mas fará o que me apraz e prosperará no que a designei." (Is 55:10-11)

Meu primeiro agradecimento ao meu amado Jesus que me mostra o caminho da cruz, que vem antes das bênçãos para quem não quiser pegar atalhos, ao Espírito Santo que me capacitou em tudo que fiz até aqui, e ao meu Deus Misericordioso.

Aos meus colegas dessa turma maravilhosa meu agradecimento por momentos de muita força e ajuda mútua, sem vocês eu não conseguiria chegar até aqui. Em especial pelo meu amigo Edgar José dos Santos Arinos sempre dando mais que uma mão em meu auxílio; e ao meu irmão em Cristo Sílvio Rogério Alves Esquinca, não só por sempre se prontificar e dedicar do seu tempo para me ajudar, mas também pelas conversas que sempre me edificaram e me exortaram nessa caminhada. Agradeço meus professores e professoras que me acrescentaram e muito em minha profissão. E sou grato a Dr.ª Fabiana Villa Alves pela confiança em me incluir em projetos da EMBRAPA para analisar seus dados. E como agradecimento especial à minha orientadora, Prof.ª Dra Rúbia Mara de Oliveira Santos, que com sua paciência e dedicação infinita me ajudou a trilhar esse caminho com esmero, a todo momento me incentivando e auxiliando muito mais do que eu merecia, pois sei que mais que suas palavras, e sim suas orações, me ajudaram a concluir essa etapa em minha vida. Ao meu amigo mais chegado que um irmão Sérgio Martins de Lima meu apreço infinito pelas palavras de encorajamento e suporte nos nossos tererés de fim de tarde. Em memória agradeço ao meu pai que me ensinou o verdadeiro valor da busca pelo conhecimento, e a minha mãe, dona Maurisbela, maior exemplo de perseverança que eu poderia ter.

Resumo

Esse trabalho mostra um problema conhecido pelos professores do ensino básico que é a lacuna deixada pelos livros didáticos quanto ao ensino da estatística com o uso de uma planilha eletrônica. Aborda o referencial teórico e exemplos de situações didáticas, inclusive usando a interdisciplinaridade, no qual é sugerido um esquema no qual a estatística poderia ser melhor aproveitada no processo de ensino aprendizagem, tendo o uso da planilha eletrônica MICROSOFT® EXCEL como sua ferramenta de auxílio para uso do professor, na sala de tecnologia da escola. Por fim há uma contribuição da estatística em um experimento da Embrapa, no qual é analisado o comportamento ingestivo de bovinos de corte, o microclima em sistemas de integração lavoura-pecuária-floresta (ILPF) e a sombra de espécies arbóreas, sendo que os testes estatísticos nos mostram aonde estão as diferenças significativas entre as variáveis e quais decisões podem ser tomadas frente a hipóteses levantadas.

Palavras-chave: Estatística, MICROSOFT® EXCEL, Embrapa.

Abstract

This work shows a problem known for primary school teachers is that the gap left by textbooks on the teaching of statistics using a spreadsheet. Approaches the theoretical referential and examples of teaching situations, including using interdisciplinarity, which suggested a scheme in which the statistic could be better utilized in the process of teaching and learning, and the use of spreadsheet MICROSOFT® EXCEL as a auxiliary tool for use of teacher, at school technology room. Finally there is a contribution of statistics in an experiment of Embrapa, in which we analyze the feeding behavior of beef cattle, the microclimate of tillage-livestock-forest integration systems (ITLF) and the shadow of tree species, and the tests where are shown in the statistical significant differences between the variables and decisions which can be taken against the raised hypotheses.

Keywords: Statistics, MICROSOFT® EXCEL, Embrapa.

Lista de Tabelas

Lista de Figuras

Sumário

Capítulo 1

Introdução

A estatística é o ramo da matemática que analisa um agrupamento metódico de dados para melhor entende-los e representa-los. Surgiu da necessidade de conhecimento e manipulação de dados coletados, e de como extrair informações de interesse coletivo, pois ela descreve uma ou mais populações, sendo usada na atividade científica, no gerenciamento de sistemas, e em levantamentos de dados de uma forma geral. Assim sendo, a estatística trata da coleta de dados informativos e da interpretação destes dados, facilitando se chegar a conclusões confiáveis sobre algum fenômeno que esteja sendo alvo de estudos.

O ensino da estatística é visto pelo aluno desde as séries iniciais do ensino básico, promovendo ao aluno conhecimentos que o habilita a reconhecer informações de gráficos e tabelas, os quais tem contato durante todos os anos da educação básica em outros conteúdos ministrados. Esse conhecimento também é dado ao aluno para que ele desenvolva senso crítico quanto as pesquisas que tem contato, o que ocorrerá além da sala de aula, como pesquisas eleitorais, características de uma população, análise de cenários econômicos e escolhas sensatas para adquirir um produto ou serviço. O tratamento das informações vem sendo presente cada vez mais frequente nas avaliações externas, principalmente no ENEM e nos vestibulares de todo o país, sendo a universidade pública ou particular.

No ensino superior, além das áreas exatas, a estatística também é utilizada em humanas e da saúde.

Nos Parâmetros Curriculares Nacionais (PCN) há os temas transversais envolvendo a matemática, com temas de referência social, que devem ser aplicados a grupos específicos, uma vez que ela participa da formação básica do cidadão.

Assim sendo, este trabalho apresenta o estudo da estatística com a utilização de uma planilha eletrônica (EXCEL) na sala de tecnologia e suas aplicações para a interpretação dos dados nela apresentados. Pois, o Microsoft EXCEL, além de ser uma planilha eletrônica que possui significativos recursos de uma planilha, bem como de programação, e de alta compatibilidade com outros aplicativos, tem sido utilizado como ferramenta de análise de dados.

O presente estudo tem por objetivo apreciar a importância do conteúdo de estatística na vida cotidiana do aluno dentro e fora da escola, sendo apresentada a legislação em vigor para conhecimento de como a estatística é apresentada ao aluno no decorrer do ensino básico, e situando a estatística em um contexto histórico. Como auxílio no desenvolvimento desse conteúdo será apresentada a ferramenta Microsoft EXCEL como planilha eletrônica que servirá de apoio no processo de ensino aprendizagem, através de situações didáticas já utilizadas por mim em sala de aula, que despertem o interesse do aluno vendo a estatística de uma forma atrativa.

Também há o objetivo de dar suporte técnico ao professor que trabalha a estatística básica, a partir de seus princípios. Sabendo que a estatística é amplamente utilizada em áreas além da matemática, não serão definidos princípios ou métodos avançados, nesse momento, por se tratarem de temas muito extensos que fogem ao objetivo desse trabalho.

Por último objetiva-se exemplificar a aplicação da estatística em um experimento real. Dois problemas serão analisados a partir da coleta de dados que ocorreu na Embrapa Gado de Corte em Campo Grande–MS, com o gado Nelore criados em sistemas de integração lavoura-pecuária-floresta (ILPF). A primeira análise será do comportamento do animal, procurando-se levantar informações sobre como os animais gastam seu tempo em pastagens diferentes, no sol ou na sombra, em estações climáticas diferentes, durante a manhã e à tarde, e em três sistemas de integração, tendo as conclusões informações cruzadas desses itens. A segunda análise será em torno de índices de temperatura, umidade e carga térmica ao qual o gado é submetido no clima de Campo Grande(MS), sendo essa análise importante para concluir se o gado vive em um ambiente de estresse causado pelo calor, temperatura e carga térmica recebidas. Nesse contexto três espécies arbóreas (Cambará, Cumbaru e Eucalipto) terão esses mesmos índices sendo levantados em suas sombras, para se determinar se há alguma delas que proporciona melhor conforto térmico para o gado.

1.1 Organização do Texto

As propostas dessa dissertação se apresentam em duas partes. A primeira finalidade nesse trabalho, no capítulo 2, é sugerir aos professores de matemática como usar uma planilha eletrônica, o EXCEL, para tornar a estatística atrativa e real aos alunos. Exemplos comprovados em sala de aula por mim, e quando trabalhei como coordenador de matemática. Os aspectos relacionados a legislação tem enfoque na estatística como tratamento da informação por ser, assim apresentada no PCN de matemática do ensino básico. A estatística foi abordada historicamente para que o professor tenha familiaridade com situações e nomes que influenciaram no desenvolvimento do estudo da probabilidade e, por seguinte, no da estatística.

No capítulo 3, como parte fundamental para contextualizar a estatística nos problemas que se seguirão, sendo uma ferramenta útil para o professor do ensino básico que queira se aprofundar no estudo da estatística, sendo alguns conceitos mais avançados, os quais trabalho há 15 anos fazendo análises estatísticas em trabalhos e pesquisas. Alguns conceitos não são aprofundados, pois essa não é uma finalidade dessa dissertação.

No capítulo 4, é apresentado o problema do desconforto térmico que o gado sofre podendo levá-lo ao estresse com a troca do calor que é feita com o meio exterior. A finalidade desse problema é mostrar a aplicação da estatística em um problema real, pois inúmeras são suas aplicações nas áreas de exatas, humanas e de saúde, o professor pode ilustrar ao aluno que haverá grande chance dele ter contato com a estatística no ensino superior, de acordo com a carreira profissional que escolher. Os métodos e cálculos usados têm aqui uma profundidade maior em suas apresentações, mostrando a estatística como ferramenta fundamental na resolução de um problema real.

Os dados analisados foram colhidos por uma equipe da Embrapa – Gado de Corte de Campo Grande-MS, sob responsabilidade da Dra Fabiana Villa Alves. No projeto uma análise faz referência aos tipos de sombra de algumas espécies arbóreas, pois os índices térmicos e até a carga térmica de radiação podem variar de espécie para espécie, influenciando a produtividade dos animais. A outra análise faz referência ao comportamento ingestivo de bovinos da raça nelore, pois com isso pode-se inferir sobre o grau de desconforto térmico de cada sistema de produção avaliado.

Capítulo 2

O Ensino da Estatística via Excel

Neste capítulo será apresentado como decorre o ensino da estatística no ensino básico, conforme a legislação em vigor. A estatística será mostrada em um contexto histórico. Propor-se-á técnicas de ensino da estatística na sala de aula utilizando a planilha eletrônica Microsoft Excel. Apresentar-se-á algumas situações didáticas, que podem enriquecer o trabalho do professor em sala de aula, ajudando-o a ministrar a estatística de uma forma didática ao aluno.

2.1 Conteúdos e Legislação

Atualmente o ensino da matemática tem sido visto de formas diferentes pelo professor e pelo aluno: o professor tem a consciência da importância da matemática no cotidiano do aluno; já o aluno está insatisfeito com esse ensino, pois não consegue ver essa importância e relacioná-la a outros conteúdos vistos em sala de aula.

Os objetivos, as metodologias e os recursos utilizados pelo professor podem ser repensados e reformulados para que a distância entre o que o professor ensina e o que o aluno aprende diminuía de forma gradativa.

A matemática serve como base para que o aluno saiba manipular números e fórmulas para a construção de conhecimentos necessários em outras áreas de ensino. Ela permite que problemas cotidianos sejam resolvidos com exatidão e precisão, sendo que as aplicações no mundo atual são inumeráveis.

A estatística passa a ser fundamental na interação da matemática com outras disciplinas, o que tem se tornado comum depois da LDB 9.394/96 e os Parâmetros Curriculares impulsionarem a interdisciplinaridade dos conteúdos.

De acordo com os PCN'S (Parâmetros Curriculares Nacionais) a matemática participa da formação básica do cidadão brasileiro, sendo a estatística de suma importância.

> A compreensão e a tomada de decisões diante de questões políticas e sociais também dependem da leitura e interpretação de informações complexas, muitas vezes contraditórias, que incluem dados estatísticos e índices divulgados pelos meios de comunicação. Ou seja, para exercer a cidadania, é necessário saber calcular, medir, raciocinar, argumentar, tratar informações estatisticamente, etc. (Redação dada pela Lei nº 12.796, de 2013)

A estatística se faz necessária na Saúde. Como citado em Matemática e os Temas Transversais. As informações sobre saúde, muitas vezes apresentadas em dados estatísticos, permitem o estabelecimento de comparações e provisões, que contribuem para o autoconhecimento, possibilitam o autocuidado e ajudam a compreender aspectos sociais relacionados a problemas de saúde.

Nos blocos de conteúdos: noções de estatística, probabilidade e análise combinatória, compõe o Tratamento da Informação.

A estatística ajuda o aluno a construir procedimentos para coletar, organizar, comunicar e interpretar dados utilizando tabelas, diagramas e gráficos, com medidas de tendência central e variabilidade representando situações do seu dia a dia.

De acordo com o PCNEM (livro dos Parâmetros Curriculares Nacionais para o Ensino Médio) o ensino da estatística contempla três grandes competências:

- representação e comunicação, que envolvem a leitura, a interpretação e a produção de textos nas diversas linguagens e formas textuais características dessa área do conhecimento;
- investigação e compreensão, competência marcada pela capacidade de enfrentamento e resolução de situações-problema, utilização dos conceitos e procedimentos peculiares do fazer e pensar das ciências;
- contextualização das ciências no âmbito sociocultural, na forma de análise crítica das ideias e dos recursos da área e das questões do mundo que podem ser respondidas ou transformadas por meio do pensar e do conhecimento científico.

As competências desenvolvidas pelos alunos são relevantes para proporcionar ao aluno instrumentos à vida, que serão utilizados na era da tecnologia, informação e globalização.

A cada conteúdo a ser abordado, é necessário lembrar que estas competências serão o objetivo final, ou seja, o aluno deverá tê-las como habilidade, e executá-las em sua plenitude.

As unidades temáticas do tratamento da informação (eixo que é composto pela estatística) do Referencial Curricular da Rede Estadual de Ensino de Mato Grosso do Sul, com as habilidades e competências, descritas em seguida, que devem ser buscadas nos alunos do ensino fundamental do 6º ao 9º ano e do 3º ano do ensino médio, são elas:

6º ano - Dados, tabelas, gráficos de colunas, barras, setores e linhas, média aritmética e produção de textos.

- Ler e analisar informações de dados apresentadas em diferentes tipos de tabelas.

- Comparar e interpretar informações de dados apresentadas em diferentes tipos de tabelas.
- Ler e interpretar informações de dados em gráficos de barras e colunas simples e duplas.
- Coletar dados (pesquisa).
- Organizar as informações de dados de pesquisa em tabela.
- Construir gráfico de barras e colunas com as informações de dados das tabelas.
- Produzir textos a partir da leitura de tabelas diversas.
- Produzir textos a partir de leitura de gráficos.
- Calcular a média aritmética.
- Resolver problemas envolvendo a média aritmética.
- Analisar e debater criticamente as informações apresentadas em dados, tabelas e gráficos.

7º ano - Dados, tabelas, gráficos de colunas, barras, setores e linhas, média aritmética e ponderada, produção de textos e probabilidade.

Além das competências e habilNa obtenção do conhecimento científico há experimentos em áreas como: biologia, filosofia, sociologia, geografia, história, química e física que, em suas observações de fatos e no estudo de teorias relacionadas a eles, tem hipóteses levantadas para sua veracidade. Para que as conclusões sejam válidas as experiências seguem condições previamente estabelecidas e são conduzidas por meio de determinados critérios. No experimento acontece a coleta de dados que ao serem analisados podem ou não confirmar a hipótese levantada. Nesse processo de análise, exploração, manipulação e descrição de dados a estatística toma seu papel fundamental de servir como base para a conclusão do experimento científico. idades citadas anteriormente, são acrescentadas:

- Resolver problemas envolvendo a média aritmética e/ou ponderada como um indicador de tendência central de uma pesquisa.
- Elaborar gráficos de setores.
- Construir o espaço amostral, utilizando o princípio multiplicativo e a indicação da probabilidade de um evento por meio de uma razão.
- Construir diagramas e árvores de possibilidades, usando material concreto (moeda e dado).
- Verificar probabilidades previstas através da elaboração de experimentos e simulações.

8º ano - Gráficos de linhas e histograma, elementos da estatística, média, moda e mediana, mapa amostral, medidas de estatística, produção de textos e probabilidade.

- Construir diagramas com amostragens a partir de informações que estuda as "chances" de um determinado resultado acontecer.
- Identificar e construir tabelas, gráfico de linhas e histograma.
- Construir gráficos e tabelas com dados obtidos a partir de pesquisa ou através de leitura de texto informativo.
- Determinar taxas e índices em tabelas e/ou gráficos (médias e porcentagens).
- Produzir textos escritos a partir de dados estatísticos.
- Elaborar conclusões a partir de tabelas e gráficos.
- Compreender o conceito de amostragem por meio da razão.
- Identificar os elementos da medida da estatística.
- Compreender os dados de um gráfico (médias e porcentagens).
- Reconhecer a média como uma medida estatística de tendência central.
- Determinar a média aritmética, média ponderada, moda e mediana.
- Resolver problemas envolvendo a probabilidade.

9º ano - Gráficos de colunas, barras, linhas e setores, coleta de dados, dados estatísticos (tabelas e gráficos), média, moda e mediana, probabilidade, frequência e frequência relativa.

- Identificar a importância da estatística na atividade humana.
- Analisar informações apresentadas em tabelas contextualizadas.
- Pesquisar, coletar e organizar dados estatísticos em tabelas e gráficos.
- Ler e interpretar informações apresentadas em diferentes tipos de gráficos: coluna, barra, linhas e setores.
- Analisar informações apresentadas em diferentes tipos de gráficos: coluna, barra, linhas e setores.
- Determinar a média aritmética, ponderada, mediana e moda.
- Identificar medidas estatísticas de tendência central em tabelas e gráficos.
- Elaborar textos a partir de dados estatísticos.
- Elaborar gráficos a partir de uma pesquisa estatística simples.
- Construir o gráfico de setores a partir dos dados de uma tabela.

- Analisar gráficos de setores.
- Analisar dados estatísticos representados em tabelas.
- Transcrever gráficos de linhas, barras e colunas para gráfico de setores.
- Utilizar o conhecimento de regras e com senso comum debater criticamente as informações de tabelas e gráficos.
- Resolver problemas envolvendo a probabilidade.
- Observar e relacionar as possibilidades de ocorrer um evento.
- Calcular a probabilidade de ocorrências de alguns eventos, por meio da razão.

3º ano (ensino médio) – Probabilidade, espaço amostral, espaço amostral, frequências, representações gráficas, média, moda e mediana.

- Resolver situação problema que envolva conhecimentos de estatística e probabilidade.
- Identificar diferentes formas de quantificar dados numéricos para decidir se a resolução de um problema requer cálculo exato, aproximado, probabilístico ou análise de médias.

Como auxílio do professor tem se a disposição o material didático, pelo qual o processo de aprendizagem torna-se mais eficaz, pois ele possui recursos que auxiliam o aluno, como: explicações teóricas, exercícios resolvidos e propostos, dicas para aprendizagem, ilustrações, tabelas e gráficos; os quais pelo tempo disponível em sala de aula o professor não poderia expor apenas usando o quadro, sendo uma extensão do que ele aprende em sala de aula. Definir a importância do livro didático para o processo de ensino aprendizagem na matemática depende de vários fatores como: a metodologia empregada, os recursos disponíveis, a continuidade de conteúdos de forma lógica e gradual, o uso de figuras com aplicação do conteúdo no cotidiano do aluno, a disponibilidade dele para o aluno, etc.

Segundo a pedagoga Bromberg: "O material didático colabora para a transformação social na medida em que favorece a elaboração constante do conhecimento como resultado de experiências interativas, propiciando o crescimento de um cidadão criativo, crítico e produtivo, pronto a enfrentar a vida com mais segurança" [25].

Dada à importância do material didático a Comissão de Educação, Cultura e Esporte aprovou por unanimidade o projeto de lei do Senado (PLS 415/11) para regulamentar a escolha, compra e distribuição de material didático-escolar às escolas de educação básica do País. O objetivo do projeto foi o de preencher o vazio que existe na área educacional pela falta de uma lei específica sobre os programas de material didático.

Outro importante auxílio para o professor no processo de ensino aprendizagem é a tecnologia. De acordo com a LDB 9.394/96, no desenvolvimento do aluno no ensino básico deve-se incluir o uso da tecnologia, pois o aluno levará esse conhecimento para sua vida e o aplicará como no trabalho.

Os livros didáticos do ensino básico incluem como uso de tecnologia para o aprendizado do aluno apenas a calculadora. No ensino da estatística não há a inclusão de uso de uma planilha eletrônica, como o Microsoft Excel e o OpenOffice.org Calc.

O Excel como planilha eletrônica oferece ao professor recursos na aprendizagem do ensino da estatística, seja na facilidade, rapidez e precisão dos cálculos ou na apresentação gráfica. Não convém substituir os cálculos que os alunos têm de aprender, e sim apresentá-los como complemento para o estudo (como será apresentado a frente). É necessário que o aluno aprenda a calcular, por exemplo, o desvio padrão de uma amostra e não apenas o comando da função no Excel que retorna esse resultado, porque, se assim fosse, o aluno provavelmente não aprenderia a interpretar o que o resultado desse cálculo significa, e o que poderá se concluir com esse valor. Com relação aos gráficos e tabelas a análise a ser feita é análoga aos cálculos estatísticos, associada com a utilização de recursos gráficos pelo aluno que ganhará tempo para escolher a forma de apresentar os dados.

Com todos os recursos e ferramentas de que o professor dispõe no Excel, ele deve não apenas expô-los ao aluno, mas deve fazer isso de maneira dinâmica, em problemas que façam o aluno associar seu cotidiano com o conteúdo aprendido, de forma desafiadora e capaz de despertar o interesse do aluno levando-o a usar o Excel fora da sala de aula.

As escolas públicas e privadas do ensino básico que são munidas de salas de tecnologias poderiam utilizar como metodologia do professor, na aprendizagem de estatística, a combinação de aula expositiva no quadro com a aplicação no computador dos conteúdos de forma prática.

Com isso o aluno terá um aprendizado que vai além da matemática do ensino básico, pois o conhecimento do uso da planilha eletrônica, em especial o EXCEL, é um dos requisitos exigidos para os jovens que querem adentrar no campo de trabalho. Ou mesmo se prepararem para os concursos que, quase a totalidade, cobram como conteúdo básico noções de informática.

Em um ano como esse (2014), que tivemos a copa do mundo de futebol no Brasil e as eleições, inclusive para o principal cargo do país (a Presidência da República), pode-se facilmente atrair a atenção do aluno para o ensino da estatística. Por mais que ele tenha dificuldade ou até mesmo não sinta prazer em estudar matemática, com certeza ele acompanhou as estatísticas dos jogos da copa do Mundo e das pesquisas eleitorais de intenção de votos.

Pode utilizar da interdisciplinaridade para tornar mais interessante o estudo da estatística utilizando a tecnologia. Um exemplo (que foi realizado na escola em que trabalho nesse ano) será o cálculo do IMC dos alunos (se houver balança na escola). O educador físico poderá auxiliar na coleta de dados da altura e massa dos alunos. A partir daí o professor de matemática poderá pedir aos alunos calculem cada um seu IMC, apresentando a fórmula do cálculo do IMC em kg/m2, que é dada por: $\text{IMC} = \dfrac{m}{h^2}$, sendo m a massa (em quilogramas) e h é a altura (em metros). Com o IMC calculado utiliza-se a tabela 2.1.1 para classificação nutricional do aluno.

Tabela 2.1.1: Classificação do IMC

IMC	CLASSIFICAÇÃO	RISCO DE DOENÇA
Menos de 18,5	Magreza	Elevado
Entre 18,5 e 24,9	Normal	
Entre 25 e 29,9	Sobrepeso	Elevado
Entre 30 e 39,9	Obesidade	Muito elevado
Pelo menos 40	Obesidade Grave	Muitíssimo elevado

Classificado o IMC de cada aluno deve-se reunir todos os dados para se formar uma tabela calculando o percentual da sala de aula em cada nível.

O resultado desse levantamento será de grande importância para que a escola tome decisões quanto à necessidade de acompanhamento da alimentação dos alunos, sendo necessária até uma intervenção por parte de um órgão de saúde, como agentes de saúde ou estagiários de universidades na área de nutrição, para que os alunos passem por uma reeducação alimentar.

Note que será mais interessante para o aprendizado do aluno se o professor utilizar a planilha eletrônica para que ele distribua os dados, colocando-os em linhas e colunas, aplicando a fórmula de forma direta, classificando percentualmente o IMC dos alunos, podendo representar os resultados de uma forma correta e elegante como em um gráfico de setores. A culminância desse estudo pode-se dar com a impressão dos gráficos em um mural da escola, para que todos os alunos tenham conhecimento da importância de uma vida equilibrada para que não corram risco de saúde.

2.2 Aspectos Históricos da Estatística

A definição de estatística no dicionário Michaelis é dada por: "Ciência que tem por objetivo a coleção, análise e interpretação de dados numéricos a respeito de fenômenos coletivos ou de massa, bem como a indução das leis a que tais fenômenos cabalmente obedecem e, ainda, a representação numérica e comparativa, em tabelas ou gráficos, dos resultados da análise desses fenômenos" [23].

O vocábulo estatística vem da palavra status (estado, em latim). Na Roma antiga, tais levantamentos buscavam o registro de todos os indivíduos de alguma camada social na sociedade, bem como o inventário de seus bens e propriedades, com a finalidade de se determinar como, quem e quanto deveria ser taxado e/ou convocado para o serviço militar. Estes levantamentos extensivos eram chamados censos, sendo promovidos por um magistrado chamado censor, cargo esse criado em 443 a.C. Posteriormente, o cargo passou a compreender outras funções. Há indícios de até 3000 anos antes de Cristo esses censos serem feitos na Babilônia, Egito, China e até mesmo há registro no livro Números do Velho Testamento onde Deus ordena a Moisés fazer o censo do povo hebreu para saber quantos homens que tinham mais de 20 anos sairiam à guerra contra os povos inimigos.

A estatística começou a ser vista por causa de jogos de azar. Girolano Cardano ganhava dinheiro com o que mais gostava: partidas de xadrez e jogos de azar. Apesar do xadrez depender da habilidade do jogador em relembrar jogadas e posições das peças no decorrer da partida isso não acontecia nos jogos de azar. Na ânsia de ganhar mais ou ganhar sempre Cardano utilizou seus talentos matemáticos em prol de melhorar seu desempenho nos jogos. Se ele pudesse compreender melhor a relação das chances que tinha, a probabilidade de ganhar ou perder, teria informações privilegiadas que seus oponentes não tinham, ganhando uma vantagem sobre eles [3].

Cardano via nas chances proporcionais as respostas uniformes dos resultados possíveis no jogo como fundamento principal nesse estudo. Achava que o meio mais seguro de trapacear - ganhar mediante um subterfúgio, seria honesto se você conhecesse as probabilidades melhores que seu adversário. Existem exemplos em que as chances reais de um jogo de azar são muito diferentes do que a maior parte das pessoas imaginam.

A teoria das probabilidades avançou quando obteve a atenção de Blaise Pascal, que se correspondia com Fermat acerca de um problema matemático relacionado com jogos de azar: como as apostas numa série de jogos de azar deveriam ser divididas se a competição tivesse de ser abandonada antes do término determinado. A questão existia desde a Idade Média, e dessa vez ela teve solução. Pascal e Fermat durante esse processo criaram um novo ramo da matemática: a teoria da probabilidade. Surgiu um conceito novo que hoje é chamado de expectativa.

A teoria das probabilidades foi reconhecida na área da matemática quando em 1713, Jacob Bernoulli [3] que começou com a definição prática usual de probabilidade de um evento: a proporção de vezes em que ele ocorre, a longo prazo, quase o tempo todo. Considerando como exemplo o lançamento seguido de uma moeda honesta onde só há dois resultados possíveis, cara ou coroa, na maior parte do tempo há uma sequência de resultados aparentemente aleatória de caras e coroas. Se a moeda for lançada por tempo suficientemente grande haverá cara em cerca da metade dos lançamentos. No entanto, é raro obter cara exatamente metade das vezes; aliás seria impossível em um número ímpar de lançamentos. Tentar modificar a definição se inspirando no cálculo de modo que a probabilidade de obter cara seja o limite da proporção de caras quando o número de lançamentos tende ao infinito, é preciso provar que esse limite existe. E pode ser que ele não exista.

Surgindo, assim, o pensamento de uma sequência de resultados ser improvável, mas para definir improvável tem que definir probabilidade, que é o que se espera que o limite possibilitasse. Então mesmo que o limite exista, poderia não ser o valor "correto" de 0,5 para a face superior da moeda mostrar cara. Poderia acontecer um caso extremo da moeda ter na face superior sempre cara, sendo agora o limite igual a 1. E poderia acontecer outro caso extremo de não sair na face superior uma cara sequer nos lançamentos da moeda, sendo o limite zero. Isso é altamente improvável, mas possível.

Bernoulli resolveu abortar toda a questão da direção oposta. Começou definindo a palavra probabilidade de cara ou coroa como sendo um número p entre 0 e 1, e provou o teorema básico da lei dos grandes números, o qual afirma que a longo prazo, com exceção de uma fração de tentativas que se torna arbitrariamente pequena, a proporção de obter caras nas faces superiores tem sim, um limite, e o limite é p. Logo, assume o ponto de vista de que os números atribuídos como probabilidades fornecem um modelo matemático consistente do lançamento de uma moeda repetidas vezes.

Esse modelo matemático que tem um padrão numérico utiliza os números representados no que se conhece como triângulo de Pascal, onde os números são citados em linhas, seguindo algumas regras. A primeira linha seria a linha 0, a segunda seria a linha 1, e assim por diante. O triângulo de Pascal pode ser escrito, em suas primeiras linhas, como:

$$
\begin{array}{lllll}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1
\end{array}
$$

no qual todas as linhas começam e terminam com 1, e cada número é a soma do número

imediatamente acima dele e seu antecessor. Esses números são chamados de coeficientes binomiais, porque surgem na álgebra da expressão binomial $(p+q)^n$, ou seja,
$(p+q)^0 = 1$
$(p+q)^1 = 1p + 1q$
$(p+q)^2 = 1p^2 + 2pq + 1q^2$
$(p+q)^3 = 1p^3 + 3p^2q + 3pq^2 + 1q^3$
$(p+q)^4 = 1p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + 1q^4$

sendo o triângulo de Pascal os coeficientes dos termos separados.

Bernoulli observou que, ao lançar uma moeda n vezes, se a probabilidade de sair cara for igual a p, então a probabilidade de um número específico de dar cara é o termo correspondente $(p+q)^n$, em que $q = 1 - p$.

O problema é que ao aumentar o número n de jogadas torna-se muito difícil e trabalhoso o cálculo da probabilidade, o que só pode ser feito com exatidão a partir de 1960, e mais popularmente a partir do final dos anos de 1980, com o uso de computadores.

Para encontrar boas aproximações, por volta de 1730, Abraham De Moivre [3] deduziu uma fórmula aproximada para as probabilidades envolvidas em lançamentos repetitivos de uma moeda viciada. Isso levou à função erro, também chamada de função erro de Gauss, que foi desenvolvida para calcular a integral da distribuição normal, muitas vezes chamada de "curva de sino" por causa de seu formato. Moivre provou e definiu a distribuição normal $\Phi(x)$ com média μ e variância σ^2 pela fórmula

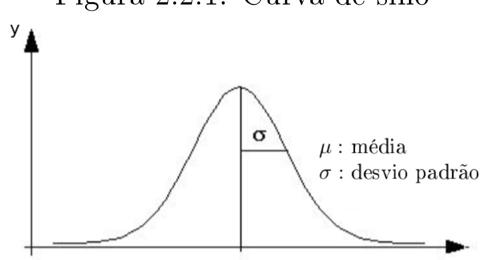
$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma}} \frac{e^{-(x-\mu)^2}}{2\sigma^2}$$

Assim, para um n grande a probabilidade de obter m caras em n lançamentos de uma moeda viciada é muito próxima de $\Phi(x)$, quando

$$x = \frac{m}{n} - p \qquad \mu = np \qquad \sigma = npq$$

A figura 2.2.1 mostra como o valor de $\Phi(x)$ varia com x. A curva se parece com um sino, daí seu nome informal. A curva do sino é um exemplo de distribuição de probabilidade; isto significa que a probabilidade de obter dados entre dois valores dados é igual a área abaixo da curva e entre as linhas verticais correspondentes a esses valores. A área total sob a curva é 1.

Figura 2.2.1: Curva de sino



Em 1835 Adolphe Quetelet [14] utilizou métodos quantitativos em sociologia, tornando esses métodos mais conhecidos e utilizados por outros. Em cada aplicação a curva de sino mostrava o cenário mais possível, tendo bom êxito nos resultados vendo a ser utilizada na teoria das probabilidades e na estatística.

Para resolver o problema de dados imprecisos devido a erros possivelmente na coleta, em 1805 o matemático francês Adrian-Marie Legendre [24] criou uma fórmula para cálculo do erro na predição de resultados, que recebeu o nome de Método dos Mínimos Quadrados.

Essas objeções fariam sentido se alguém procurasse predisser quando e quem iria se matar. Mas quando Quetelet concentrou-se em questões estatísticas, tais como a proporção de suicídios em vários grupos de pessoas, vários locais e diferentes anos, começou a estudar padrões. Pois qualquer que fosse o motivo, pessoas em massa se comportam mais previsivelmente do que indivíduos. É claro que as pessoas exercem seu livre-arbítrio, mas as influências para o suicídio têm variáveis comuns a eles: problemas financeiros, problemas de relacionamento, estado mental e passado religioso. Em todo caso, a curva de sino não faz predições exatas; ela simplesmente afirma qual cenário é mais provável.

Rapidamente a curva de sino se tornou um ícone na teoria das probabilidades e em especial em sua aplicação na estatística. Os astrônomos a utilizaram para seus cálculos imensos onde havia grande parte de erros nos dados levantados, seja por falha humana ou movimento de correntes de ar na atmosfera. Para solução desse problema prático em 1805, o matemático francês Adrian-Marie Legendre descobriu uma fórmula simples para cálculo do erro na predição de resultados, chamando-o de Método dos Mínimos Quadrados. Essa descoberta foi tão importante que Gauss fez desse método a base de seu trabalho em mecânica celeste.

A inferência estatística teve boa base com os estudos de Thomas Bayes [3] com o conceito de probabilidade inversa em situações que se caminha do particular para o geral.

Em 1810, Laplace utilizou a transformada de Fourier para provar que a média de muitas observações astronômicas é descrita por uma curva de sino, ainda que as observações individuais não sejam. Seu resultado, o teorema do limite central, foi importante para o estudo da probabilidade e estatística, pois forneceu uma justificativa teórica para o uso da distribuição favorita dos matemáticos, a curva de sino, na análise de erros observacionais.

Com as ferramentas estatísticas a disposição de quem queria procurar padrões a partir de dados, muitas foram as aplicações, principalmente nas ciências sociais.

Em 1865, Francis Galton [16] estudou como a altura de uma criança se associa a altura de seus pais, para entender a hereditariedade. Para determinar a eficácia de drogas e procedimentos médicos e outras aplicações foi introduzido o Teste de Hipóteses por Ronald Aylmer Fisher e Karl Pearson [16].

Outra importante contribuição foi desenvolvida por William Gosset que trabalhava na Cervejaria Guinness e precisava de uma distribuição que pusesse ser analisada em pequenas amostras. Como a empresa que ele trabalhava não permitia a publicação de resultados de pesquisas, Gosset publicou com o pseudônimo de Student.

2.3 Sugestões Didáticas para o Ensino da Estatística

O livro didático é um instrumento a favor do processo de ensino aprendizagem. As escolas públicas podem adquirir o material didático através do Programa Nacional do Livro Didático (PNLD), que tem como principal objetivo subsidiar o trabalho pedagógico dos professores por meio da distribuição de coleções de livros didáticos aos alunos da educação básica.

São muitos os livros didáticos oferecidos pelas editoras com apresentação e didática diferentes dos conteúdos da matemática. Em relação ao ensino de estatística no ensino básico tem-se um gama de apresentações de como estudar dados coletados em experiências e pesquisas, representando-os através de medidas de tendência central e variabilidade, e apresentando-os através de tabelas e gráficos, que servirão para previsões e tomada de decisões na análise desses dados.

Há uma lacuna na utilização de uma ferramenta de produtividade como uma planilha eletrônica nos livros didáticos, já que as escolas possuem sala de tecnologia com acesso aos computadores a disposição dos professores e alunos.

Tendo em vista a omissão por parte dos livros didáticos, nessa seção serão apresentadas propostas didáticas para serem utilizadas em sala de aula, munindo o professor do Ensino Médio em um caminho onde ele poderá trabalhar com estatística na sala de aula e na sala de tecnologia. Será utilizado o Microsoft Excel como uma planilha eletrônica que possui recursos eletrônicos específicos para ser mais um instrumento de auxílio ao professor no processo de ensino aprendizagem. O professor poderá utilizá-lo para planejamento de suas aulas tendo um fator motivador com o aluno, cujo interesse na tecnologia é cada dia maior e se faz presente em seu dia a dia.

Os computadores que são usados pelas escolas em sua maioria vêm com o sistema operacional Linux, pois ele não necessita de licença paga para seu uso. Apesar disso são poucos os profissionais de educação que fazem uso dele, ou por não ter formação básica para usar seus softwares, ou por não achar que ele seria a melhor escolha para uso educacional. É sabido pelos profissionais da educação que o sistema operacional Windows está instalado em praticamente todas as escolas, apesar de seu uso necessitar de licença paga. Sendo assim, os professores preferem trabalhar com os softwares do Windows, mesmo porque o utilizam em seus computadores pessoais, tendo, portanto, conhecimento básico no momento de usá-lo didaticamente.

Apesar do Excel possuir inúmeros recursos para cálculos matemáticos, a utilização dele para o ensino da estatística no ensino básico, será apenas dos recursos disponíveis para a análise de dados aplicados na estatística descritiva, onde os dados coletados serão analisados e representados.

A estatística é um conjunto de técnicas e métodos de pesquisa que utiliza a indução lógica, isto é, a generalização de características observadas em uma parte da coletividade para se conhecer o todo, podendo assim, representa-lo, analisa-lo e até fazer previsões e tomada de decisões sobre o todo analisado.

A apresentação da estatística pode ser iniciada com situações do cotidiano do aluno, situações estas que tenham necessidade de serem analisadas para melhor serem conhecidas. Essas situações podem ser encontradas em toda a parte, como por exemplo, nos jornais, revistas, livros. Para iniciarmos a discussão do estudo de estatística neste trabalho, será utilizada uma situação problema comum do dia a dia da escola. O exemplo 1 foi adaptado das salas de aula que sou professor e o exemplo 2 foi adaptado de um exemplo citado por Triola (2005).

Exemplo 1. Com o intuito de saber se há diferenças relevantes no aprendizado das salas de primeiro ano do Ensino Médio, o professor André pegou as notas do 1º bimestre das cinco salas de aula que ministra Matemática (de A até E), cada uma com 20 alunos, e formou uma tabela com elas em ordem crescente (Tabela 2.3.1).

Tabela 2.3.1: Notas bimestrais de cada turma

1º A	1º B	1º C	1º D	1º E
4,0	4,5	2,0	2,0	1,5
4,5	5,0	4,0	3,5	2,0
5,0	5,0	4,0	4,0	2,0
5,0	5,5	4,0	4,0	4,0
5,0	5,5	4,5	4,0	4,5
5,5	7,0	4,5	5,5	4,5
5,5	7,0	5,0	5,5	4,5
6,0	7,0	5,0	5,5	5,0
6,0	7,0	5,0	5,0	5,0
6,0	7,5	5,5	5,5	5,5
6,0	8,0	6,0	6,0	5,5
6,0	8,5	6,5	6,0	6,0
6,0	8,5	7,0	7,0	7,0
6,0	8,5	7,0	7,0	7,0
6,0	8,5	7,5	7,0	7,0
6,0	9,0	7,5	7,5	8,0
6,5	9,0	7,5	7,5	8,0
7,0	9,5	8,5	9,0	9,5
7,0	9,5	9,5	9,0	9,5
7,0	10,0	9,5	9,5	10,0

Com a tabela criada o professor André pode obter diversas informações importantes, tais como:

- Quais as maiores e menores notas;
- Qual a amplitude das notas;
- Qual a nota que representam cada sala de aula;
- Se há variações significativas nas notas em cada sala de aula.

O professor pode comentar que com a coleta desses dados a estatística já está em andamento, observando que os dados foram citados de uma forma organizada em sequência formando um rol.

Para complementar a situação pode-se ressaltar que não seria fácil estudar todos os dados ao mesmo tempo, comparando-os um a um, o que leva a entender que será necessário representar todas as notas de cada turma através de uma nota apenas, utilizando as medidas de tendência central, tais como: média, mediana e moda. Deve-se explicar que é importante verificar se na sala de aula há diferença significativas entre as notas tiradas pelos alunos, utilizando as medidas de variabilidade ou dispersão, tais como: desvio absoluto médio, desvio padrão e variância. E, para finalizar, dizer que uma maneira de representar essas medidas seria utilizando os gráficos. E os melhores gráficos para expor esses resultados seriam o gráfico de colunas e o Box-Plot.

Pode-se comentar que a média é, em geral, a mais importante de todas as medidas usadas para descrever dados, e o desvio padrão é o mais importante depois da média. Para deixar claro pode pedir ao aluno citar tudo que ele já ouviu de informações com o nome média, e nas pesquisas, sobre a margem de erro (que é calculada usando principalmente a média e o desvio padrão).

Para sistematizar a representação da média amostral é feita pela letra x com uma barra em cima dela ($\overline{x}$), e o desvio padrão amostral pela letra s.

Com isso o professor dominará as ferramentas básicas para mensurar e descrever as características do conjunto de dados coletados, isso implica que poderá assegurar se as salas têm notas iguais ou não, ou seja, estatisticamente se as amostras vêm de uma mesma população.

Com os dados e a tabela prontos, poderia surgir o problema de calcular a média representativa das notas da sala, sendo necessário relacionar os dados com soma e subtração, e a partir dessa relação começar a exposição do algoritmo do cálculo da média de um conjunto de dados.

Portanto, na resolução do exercício será necessário voltar a tabela 2 e responder de que maneira pode-se determinar a média de cada sala de aula?

Para o cálculo da média devemos somar todos os valores da amostra e dividir pelo número total de dados da amostra. Logo, para a fórmula da média amostral, tem-se:

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_{n-1} + x_n}{n}$$

sendo $x_1,\ x_2,\ x_3, ...,\ x_{n-1}$ e x_n são os valores coletados na amostra e n é o total de dados coletados na amostra.

Como o aluno está familiarizado com o cálculo da média das suas notas desde os primeiros anos do Ensino Fundamental do Ensino Básico, não será difícil para ele conseguir fazer esse cálculo já que é algo intuitivo, onde ele somará todas as notas de sua sala e dividirá pelo número total de alunos. Como resultado obtém-se:

1) Cálculo da média do 1º A:

$$\overline{x} = \frac{4 + 4,5 + ... + 7 + 7 + 7}{20} = \frac{116}{20} = 5,8$$

2) Cálculo da média do 1º B:

$$\overline{x} = \frac{4,5 + 5 + 5 + ... + 9,5 + 9,5 + 10}{20} = \frac{150}{20} = 7,5$$

3) Cálculo da média do 1º C:

$$\overline{x} = \frac{2+4+4+...+8,5+9,5+9,5}{20} = \frac{120}{20} = 6,0$$

4) Cálculo da média do 1º D:

$$\overline{x} = \frac{2+3,5+4+...+9+9+9,5}{20} = \frac{120}{20} = 6,0$$

5) Cálculo da média do 1º E:

$$\overline{x} = \frac{1,5+2+2+...+9,5+9,5+10}{20} = 5,8$$

Após a resolução do problema, cabe ao professor expor a definição de média de um conjunto de dados. Com a sequência compreendido pelo aluno, é importante fazer com que o aluno compreenda a importância de analisar o valor calculado comparando-o com os valores das médias das outras salas.

O aluno deve concluir que a média das notas do 1º B é maior que a média do 1º C e 1º D, que por sua vez é maior que a média do 1º A e do 1º E. Então não há diferenças entre as notas que as turmas do 1º C e o 1º D tiraram, e do 1º A e do 1º E, conforme a nota de avaliação bimestral que essas turmas tiveram?

Quando o professor for utilizar esses dados para auxiliá-lo na análise do aprendizado de seus alunos, se ele chegar à conclusão que não há diferenças significativas com relação, por exemplo, as turmas do 1º A e do 1º E, e deverá tomar medidas parecidas em cada sala, para recuperar os alunos que estão com notas menores nas avaliações, e que não estão conseguindo acompanhar o aprendizado tirando notas, pelo menos, acima da média.

Essa conclusão embasada apenas da média pode nos ocultar algumas informações importantes sobre os valores coletados. Pois a média pode ser drasticamente afetada por valores extremos.

Entendendo isso, o professor pode seguir com a análise dos dados pedindo aos alunos que verifiquem se há grandes variações entre as notas das turmas que têm mesma média. Com isto, provavelmente conseguirá responder ao questionamento anterior. Nesse momento, pode-se pedir ao aluno observar a variação entre as notas das turmas calculando a amplitude total das notas das turmas que têm mesma média. Após observação, o aluno verá que não há diferença entre a amplitude das notas do 1º C e do 1º D, mas que há entre as notas do 1º A e o 1º E. Precisamos de uma medida de dispersão ou variabilidade que nos dê uma boa interpretação da variação das notas em torno da média. É de extrema importância deixar o aluno concluir, por si só, que haverá maior variabilidade no conjunto de dados que têm notas mais longe das médias. Se isso não ocorrer, ele pode utilizar um exemplo simples para ajudar o aluno.

Exemplo 2. Tempos atrás, muitos bancos exigiam que clientes esperassem em filas separadas para cada caixa, mas muitos mudaram agora para o sistema de fila única. Porque fizeram essa mudança? O tempo médio de espera não mudou, porque a configuração da fila não afeta a eficiência dos atendentes. Eles mudaram para a fila única porque os clientes preferem tempos de espera que sejam consistentes com menos variação. Milhares de bancos fizeram uma mudança que resultou em menor variação mesmo que a média não tenha sido afetada. Consideremos tempos de espera (em minutos) específicos de 3 clientes em 2 bancos.

1º - Banco do Paralelepípedo – 4, 7 e 7.
2º - Banco HMST – 1, 3 e 14.

Verifica-se que a média de espera nos 2 bancos é igual a 6 minutos. E verifica-se visualmente, que os tempos de espera de 4, 7 e 7 variam bem menos que os tempos de espera de 1, 3 e 14.

Observada a necessidade, por parte dos alunos, de uma medida de variabilidade dos dados o professor tem a possibilidade de expor como deve ser calculado o desvio padrão de cada amostra. O aluno depois de compreender o exemplo das variações dos tempos médios tenderá a calcular a distância de cada valor a média. E faz sentido começar a calcular as quantidades individuais pelas quais cada valor se afasta da média. Para um valor particular x, a quantidade de desvio (variação) é $x - \overline{x}$, que é a diferença entre o valor individual de x e a média. No exemplo 2, os desvios dos tempos do Banco do Paralelepípedo a partir da média são -2, 1 e 1; e no Banco HMST são -5, -3 e 8. Se conseguíssemos combinar esses valores em um único coletivo seria muito fácil ter um valor para compararmos. A simples soma dos desvios não serve, porque essa soma é sempre zero. Para obtermos uma estatística que meça a variação (em vez de sempre zero), precisamos evitar o cancelamento dos valores positivos e negativos.

Depois de verificar que o aluno compreendeu e se certificou que esses desvios somados sempre zeram, podemos lembrar o aluno do valor absoluto de um número. Importante que o professor nesse momento apresente a estatística como um conteúdo da matemática que necessita da aplicação do valor absoluto, já que um dos maiores questionamentos dos nossos alunos é em relação a aplicação dos conteúdos aprendidos em matemática.

Um método para chegarmos a nossa medida de desvio é somar os valores absolutos, como em $\sum_{i=1}^{n}(x_i - \overline{x})$. Se acharmos a média dessa soma, obteremos o Desvio Médio Absoluto (DMA), que é a distância média dos dados até a média. Logo,

$$\text{DMA} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})}{n}$$

Cabe ao aluno calcular o desvio médio absoluto do exemplo 2, obtendo como DMA do Banco Paralelepípedo $\dfrac{2+1+1}{3} = 1,3$, e do Banco HMST $\dfrac{5+3+8}{3} = 16,3$.

Tendo o aluno uma compreensão da medida de variabilidade, é provável que o professor seja questionado em relação ao porquê de não se usar o desvio médio absoluto como medida de variabilidade dos dados e querer continuar com outros cálculos para se chegar ao desvio padrão. O uso de valores absolutos cria dificuldades algébricas nos métodos de inferência estatística, por exemplo, o desvio médio absoluto não possui a propriedade aditiva, sendo um valor viesado. Isto significa que ao se encontrarem valores absolutos médios de amostras, não se tende a atingir o valor médio da população.

O cálculo do desvio padrão baseia-se em uma raiz quadrada de uma soma de quadrados, assemelhando-se a fórmulas de distâncias encontradas na álgebra, tendo o aluno familiaridade com elas, como exemplo, a fórmula da distância entre dois pontos no plano cartesiano. Elevamos ao quadrado os desvios absolutos médios porque as amostras mais distantes da média aritmética irão adquirir maior peso do que aquelas que estão mais perto.

Assim, esses valores serão não negativos e ao serem somados não resultarão em zero como na soma dos desvios absolutos médios.

Depois de encontrar todos os valores individuais de $(x-\overline{x})^2$, iremos combiná-los através de sua soma e a seguir obtemos uma média dividindo por $n-1$. O resultado da raiz quadrada desse valor nos dá o desvio padrão da amostra. Mas, antes de prosseguir, o professor deve explicar ao aluno porque é escolhido $n-1$ e não n que é o número de dados coletados na amostra. Dividimos por n–1 porque há apenas n–1 valores independentes. Isto é, dada uma média, apenas n – 1 valores podem ser associados a qualquer número antes que o último valor seja determinado. Só utilizamos a divisão por n quando tivermos toda a população sendo analisada, ou seja, no cálculo do desvio padrão populacional.

Logo, a fórmula para o cálculo do desvio padrão amostral será:

$$s=\sqrt{\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+...+(x_{n-1}-\overline{x})^2+(x_n-\overline{x})^2}{n-1}}$$

na qual $x_1,\ x_2,\ x_3,...,\ x_{n-1}$ e x_n são os valores coletados na amostra, $\overline{x}$ é a média amostral e n é o total de dados coletados na amostra.

Continuando com nosso exemplo bancário o aluno deverá aplicar a fórmula para calcular o desvio padrão, e como resultado deve obter:

1) Cálculo do desvio padrão do tempo de espera na fila pelos clientes do Banco do Paralelepípedo:

$$s=\sqrt{\frac{(4-6)^2+(7-6)^2+(7-6)^2}{3-1}}=\sqrt{3}=1,7\text{ minuto.}$$

2) Cálculo do desvio padrão do tempo de espera na fila pelos clientes do Banco HMST:

$$s=\sqrt{\frac{(1-6)^2+(3-6)^2+(14-6)^2}{3-1}}=\sqrt{49}=7,0\text{ minutos}$$

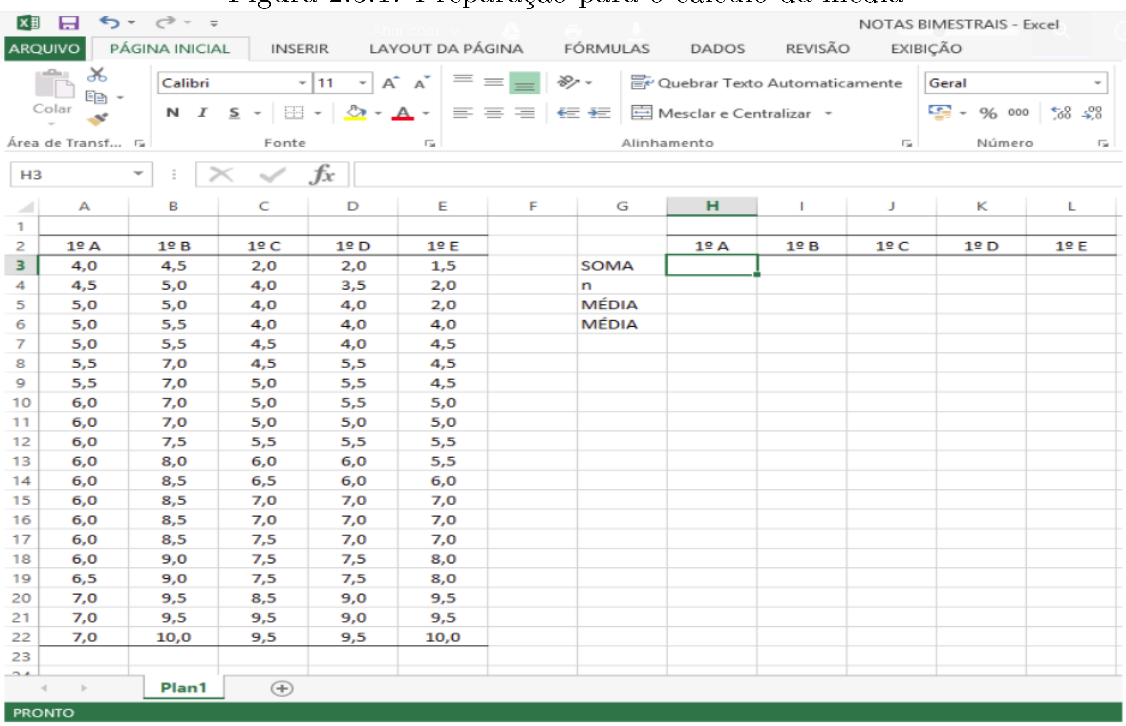
Agora o professor pode pedir aos alunos que tirem suas conclusões a partir do resultado calculado. E a conclusão que devemos chegar é de que o tempo de espera na fila para ser atendido no Banco do Paralelepípedo é bem mais consistente (ou tem menor variação) que no Banco HMST.

A partir da conclusão sobre a medida de variabilidade relacionada ao cálculo feito do desvio padrão, o aluno estará apto a calcular o desvio padrão das notas da tabela 2. O aluno irá desanimar perto dessa tarefa imensa, pois são 30 valores em cada amostra! É nesse momento que o professor sugere a utilização de uma ferramenta extremamente importante que irá auxiliar o aluno, não só nesse cálculo, mas em outros cálculos estatísticos: o Excel.

O Excel permite cálculo das medidas de tendência central: média, mediana e moda; de variabilidade: amplitude, desvio médio absoluto, desvio padrão e a variância; e a realização de gráficos para representar as amostras. O Excel possui mais funções associadas a essas medidas, como por exemplo, cálculo de média geométrica, média harmônica e coeficiente de variação. Mas são funções que fogem ao referencial teórico do Ensino Básico.

Utilizando os dados com as notas bimestrais da tabela 2, o professor poderá abrir um arquivo e disponibilizá-lo em rede para que todos os computadores estejam com os dados tabulados exatamente como a tabela 2. Ela pode estar exposta nas células no intervalo A2:E22, tendo a linha 2 as respectivas turmas, representadas nas colunas de A até E, e as notas citadas da linha 3 até a linha 22. Iniciando a atividade será proposto aos alunos que copiem o intervalo A3:E3 no espaço na mesma linha 3 nas colunas de H até L, citando SOMA na célula G3, n na célula G4 e média na célula G5 e G6. (Figura 2.3.1).

Figura 2.3.1: Preparação para o cálculo da média



	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	1º A	1º B	1º C	1º D	1º E			1º A	1º B	1º C	1º D	1º E
3	4,0	4,5	2,0	2,0	1,5		SOMA					
4	4,5	5,0	4,0	3,5	2,0		n					
5	5,0	5,0	4,0	4,0	2,0		MÉDIA					
6	5,0	5,5	4,0	4,0	4,0		MÉDIA					
7	5,0	5,5	4,5	4,0	4,5							
8	5,5	7,0	4,5	5,5	4,5							
9	5,5	7,0	5,0	5,5	4,5							
10	6,0	7,0	5,0	5,5	5,0							
11	6,0	7,0	5,0	5,0	5,0							
12	6,0	7,5	5,5	5,5	5,5							
13	6,0	8,0	6,0	6,0	5,5							
14	6,0	8,5	6,5	6,0	6,0							
15	6,0	8,5	7,0	7,0	7,0							
16	6,0	8,5	7,0	7,0	7,0							
17	6,0	8,5	7,5	7,0	7,0							
18	6,0	9,0	7,5	7,5	8,0							
19	6,5	9,0	7,5	7,5	8,0							
20	7,0	9,5	8,5	9,0	9,5							
21	7,0	9,5	9,5	9,0	9,5							
22	7,0	10,0	9,5	9,5	10,0							
23												

Com o cursor na célula H3 (Tabela 2.3.2) o aluno deverá escrever =SOMA(A3:A22) e clicar ENTER, com isso obterá a soma total das notas do 1º A. Na célula H4 (onde está o cursor) deverá escrever =CONT.NÚM(A3:A22) e clicar ENTER, com isso obterá o número de células no intervalo citado que contem números (o aluno pode digitar 20, mas essa é uma fórmula do EXCEL que ele provavelmente irá utilizar em outros exercícios). Agora, ele digitará em H5 (onde parou o cursor) =H3/H4 e ENTER, que retornará o valor da média da amostra que contém as notas da turma do 1º A. Para finalizar essa parte na célula H6 (onde parou o cursor), o professor deve pedir ao aluno para digitar =MÉDIA(A3:A22), que deverá conter o mesmo valor que a média calculada em H5. Obviamente o aluno não ficará feliz de ter feito os cálculos das células H3, H4 e H5 para chegar na média, já que poderia utilizar direto a fórmula em H6. Cabe ao professor explicar a importância dele compreender todo o processo para se chegar ao valor da média fixando, assim, o conteúdo com maior precisão e podendo explicar como o algoritmo fez o cálculo da média amostra. Esse processo de ensino-aprendizagem facilita muito, pois é um meio alternativo onde o aluno sai do ambiente de sala de aula com quadro branco e pincel, tornando-se mais receptivo e aberto para se esforçar a alcançar um novo conhecimento.

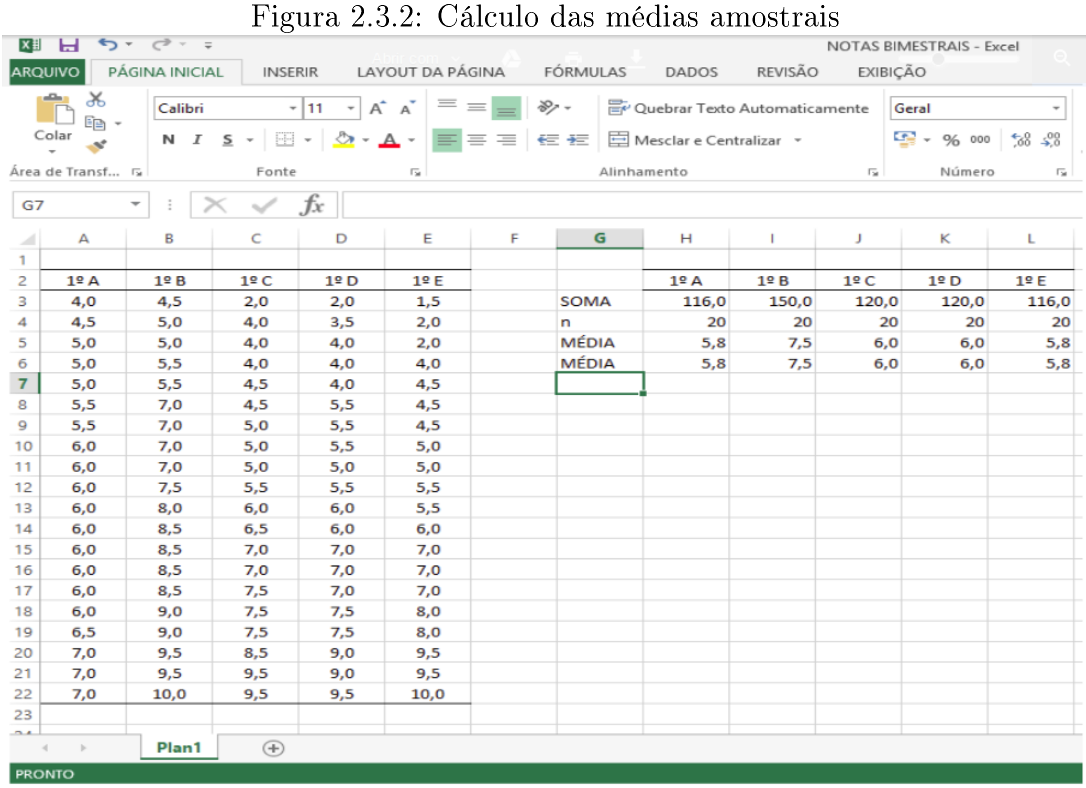
O intervalo H2:L6 que conterá as fórmulas deverá ser conforme a tabela 2.3.2, com o acréscimo do símbolo de igualdade na frente de cada função digitada, contendo todos os comandos dados.

Tabela 2.3.2: Modelo para cálculo da média das notas das turmas

1° A	1° B	1° C	1° D	1° E
SOMA(A3:A22)	SOMA(B3:B22)	SOMA(C3:C22)	SOMA(D3:D22)	SOMA(E3:E22)
CONT.NÚM(A3:A22)	CONT.NÚM(B3:B22)	CONT.NÚM(C3:C22)	CONT.NÚM(D3:D22)	CONT.NÚM(E3:E22)
H3/H4	I3/I4	J3/J4	K3/K4	L3/L4
MÉDIA(A3:A22)	MÉDIA(B3:B22)	MÉDIA(C3:C22)	MÉDIA(D3:D22)	MÉDIA(E3:E22)

Para fazer esses cálculos nas outras turmas o aluno poderá, de forma análoga, seguir os mesmos passos. Poderá utilizar uma ferramenta bastante útil que servirá para cálculos futuros: deverá selecionar o intervalo H3:H6 (que possui os resultados calculados) e posicionar o cursor em cima de um quadrado que aparecerá no canto inferior direito do intervalo selecionado, o cursor ficará como uma cruz escura; após isso, só precisa arrastar o intervalo ainda selecionado até a coluna L (ainda na mesma linha 6) e soltá-lo. Automaticamente os mesmos comandos dados no intervalo H3:H6 aparecerão no intervalo I3:L6, tendo a coluna I os cálculos da turma do 1º B, o intervalo J3:J6 do 1º C, o intervalo K3:K6 do 1º D e o intervalo L3:L6 do 1º E (Figura 2.3.2). Essa primeira etapa do exercício terá o cálculo das médias das turmas que deverá ser confirmado pelos alunos com os cálculos feitos anteriormente no caderno.

Figura 2.3.2: Cálculo das médias amostrais



	A	B	C	D	E
2	1º A	1º B	1º C	1º D	1º E
3	4,0	4,5	2,0	2,0	1,5
4	4,5	5,0	4,0	3,5	2,0
5	5,0	5,0	4,0	4,0	2,0
6	5,0	5,5	4,0	4,0	4,0
7	5,0	5,5	4,5	4,0	4,5
8	5,5	7,0	4,5	5,5	4,5
9	5,5	7,0	5,0	5,5	4,5
10	6,0	7,0	5,0	5,5	5,0
11	6,0	7,0	5,0	5,0	5,0
12	6,0	7,5	5,5	5,5	5,5
13	6,0	8,0	6,0	6,0	5,5
14	6,0	8,5	6,5	6,0	6,0
15	6,0	8,5	7,0	7,0	7,0
16	6,0	8,5	7,0	7,0	7,0
17	6,0	8,5	7,5	7,0	7,0
18	6,0	9,0	7,5	7,5	8,0
19	6,5	9,0	7,5	7,5	8,0
20	7,0	9,5	8,5	9,0	9,5
21	7,0	9,5	9,5	9,0	9,5
22	7,0	10,0	9,5	9,5	10,0

G	H	I	J	K	L
	1º A	1º B	1º C	1º D	1º E
SOMA	116,0	150,0	120,0	120,0	116,0
n	20	20	20	20	20
MÉDIA	5,8	7,5	6,0	6,0	5,8
MÉDIA	5,8	7,5	6,0	6,0	5,8

Continuando com a medida de tendência central o aluno calculará a mediana e a moda. A mediana pode ser usada quando temos valores muito distantes dos outros valores, causando uma variabilidade muito grande, esses valores são chamados de *outliers* (ou valores discrepantes). Para o cálculo da mediana o professor pode explicar ao aluno que essa medida é literalmente a medida de centro dos dados, pois é a medida que se encontra no meio dos dados quando colocados em ordem crescente (ou decrescente) de magnitude. E com isso explicar que a mediana não fica afetada por dados discrepantes, como acontece com a média aritmética. Quando o número de dados da amostra é ímpar fica mais fácil saber a mediana, pois será exatamente o termo do meio; mas quando o número de dados é ímpar, deve-se fazer a média entre os dois termos do meio.

Já a moda é bastante útil quando os dados não são numéricos. Para o cálculo da moda deve-se pegar o dado que tiver maior frequência, podendo a amostra ser: modal, bimodal (dois dados ocorrem com a mesma frequência), multimodal (mais de dois dados ocorrem com a mesma frequência) ou amodal (quando nenhum dado se repete).

Como em cada turma da nossa atividade são 30 alunos, já tendo os dados em ordem crescente, o aluno deverá calcular a média entre o 15º e o 16º dado. As medianas calculadas devem ser: 6,0 – 1º A, 7,8 – 1º B, 5,8 – 1º C, 5,8 – 1º D e 5,5 – 1º E. E, a moda de cada turma deve ser: 6,0 – 1º A, 7,0 ou 8,5 (bimodal) – 1º B, 4,0 ou 5,0 ou 7,5 (multimodal) – 1º C, 5,5 – 1º D e 4,5 ou 7,0 (bimodal) – 1º E.

No EXCEL a função que retorna o valor da mediana é MED e a moda é MODO. Segue como atividade o aluno utilizar o EXCEL para confirmar seus cálculos, é importante o professor deve esclarecer que quando existe mais de um valor se repetindo com mesma frequência (bimodal ou multimodal) o EXCEL retornará o menor valor como moda.

Analisando as medidas centrais calculadas podemos compará-las graficamente. A figura 2.3.3 é um exemplo da comparação das médias das notas bimestrais das turmas dos 1º anos. Na figura 2.3.4 além das médias também citamos as medianas.

Figura 2.3.3: Notas médias bimestrais das turmas dos 1º anos

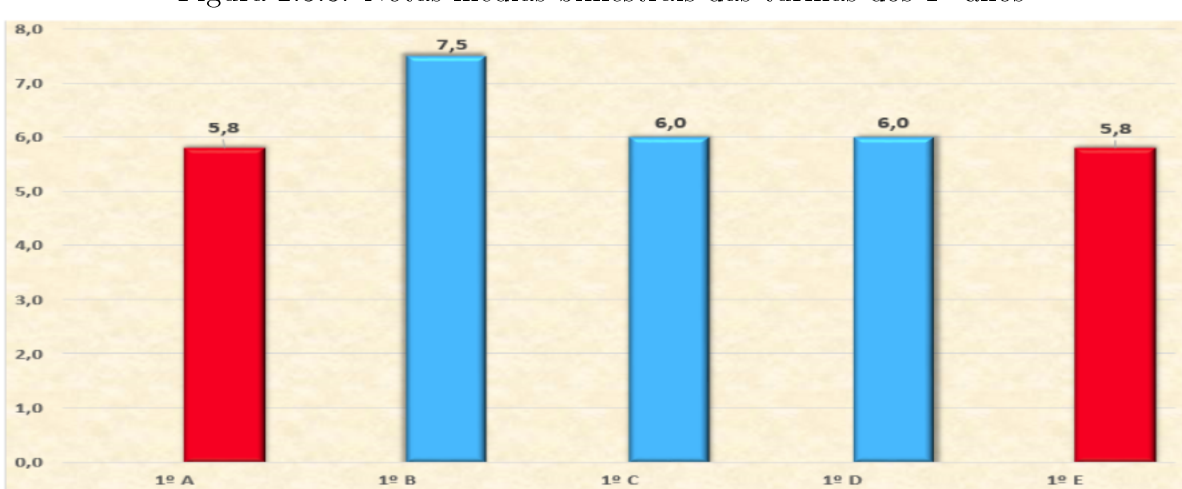
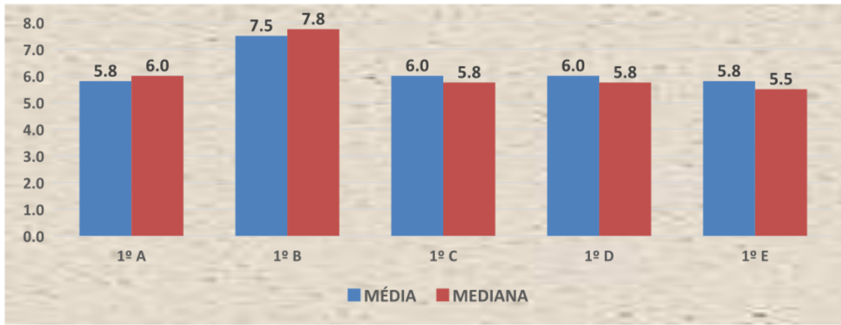


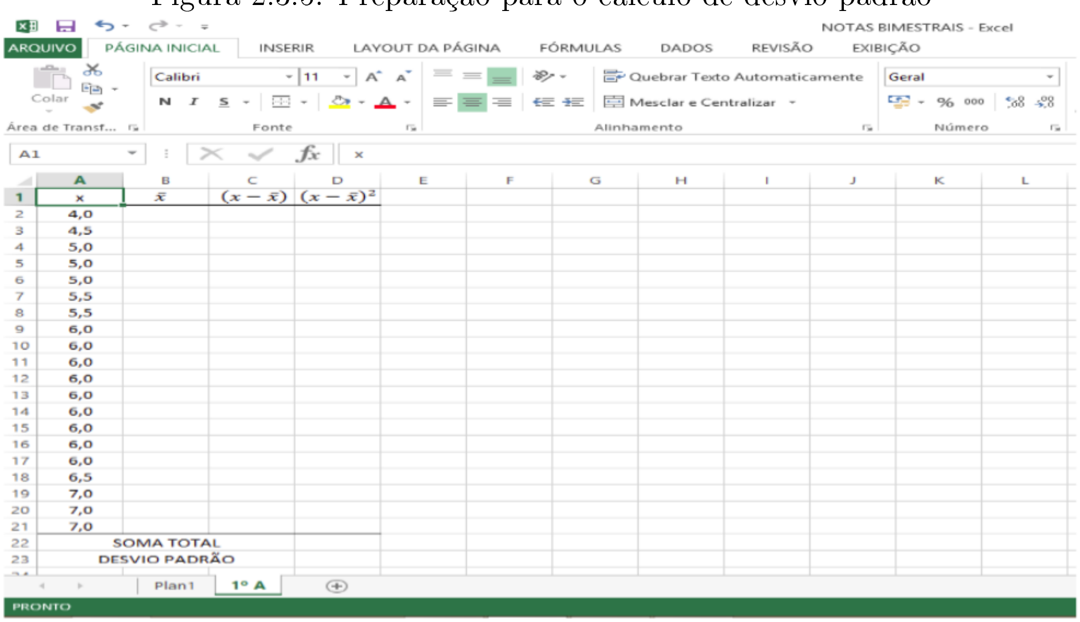
Figura 2.3.4: Notas médias e medianas bimestrais das turma dos 1º anos



Na segunda parte da atividade o aluno calculará o desvio padrão. O professor poderá pedir ao aluno que abra uma nova aba com as notas de cada turma separadas por quatro colunas, ou utilizar uma aba para cada turma (faremos dessa maneira sem prejuízo ao processo de ensino-aprendizagem).

Pode-se começar com as notas da turma 1º A (nomeando-a na aba) colando as notas no intervalo A2:A21, e citando na linha 1 o que representa cada coluna: x, $\overline{x}$, $x - \overline{x}$ e $(x - \overline{x})^2$, nas colunas A, B, C e D, respectivamente, e na linha 22 mesclando as colunas A, B e C escrever SOMA TOTAL, cujo resultado estará na célula D22, e na linha 23 mesclando DESVIO PADRÃO, com resultado na célula D23 (Figura 2.3.5).

Figura 2.3.5: Preparação para o cálculo de desvio padrão



Na coluna B da linha 2 até a linha 21 deve-se citar o valor da média já calculada; na coluna C, iniciando na linha 2, deve-se citar a fórmula =A3-B3, e na coluna D, iniciando na linha 2, deve-se citar a fórmula =C3ˆ2. Com isso, temos o desvio médio calculado na coluna C e o seu quadrado na coluna D. De forma análoga ele repetirá as fórmulas nas próximas linhas até a linha 21. Espera-se que o aluno não queira repetir esse processo até a última linha (linha 21), mas que utilize o conhecimento adquirido anteriormente e selecionando o intervalo C2:D2 ele arraste o cursor até a linha 21, fazendo com que o EXCEL repita esses cálculos com os outros valores. A célula D22 conterá a soma dos quadrados dos desvios médios com a fórmula =SOMA(D2:D21). Lembrando a fórmula do desvio padrão deve-se citar na célula D23 a fórmula =RAIZ((D22)/(CONT.NÚM(A2:A21)-1))).

O intervalo A1:D23 que conterá as fórmulas deverá ser conforme a tabela 2.3.3, com o acréscimo do símbolo de igualdade na frente de cada função digitada, contendo todos os comandos dados.

Tabela 2.3.3: Modelo para o cálculo do desvio padrão das notas do 1º A

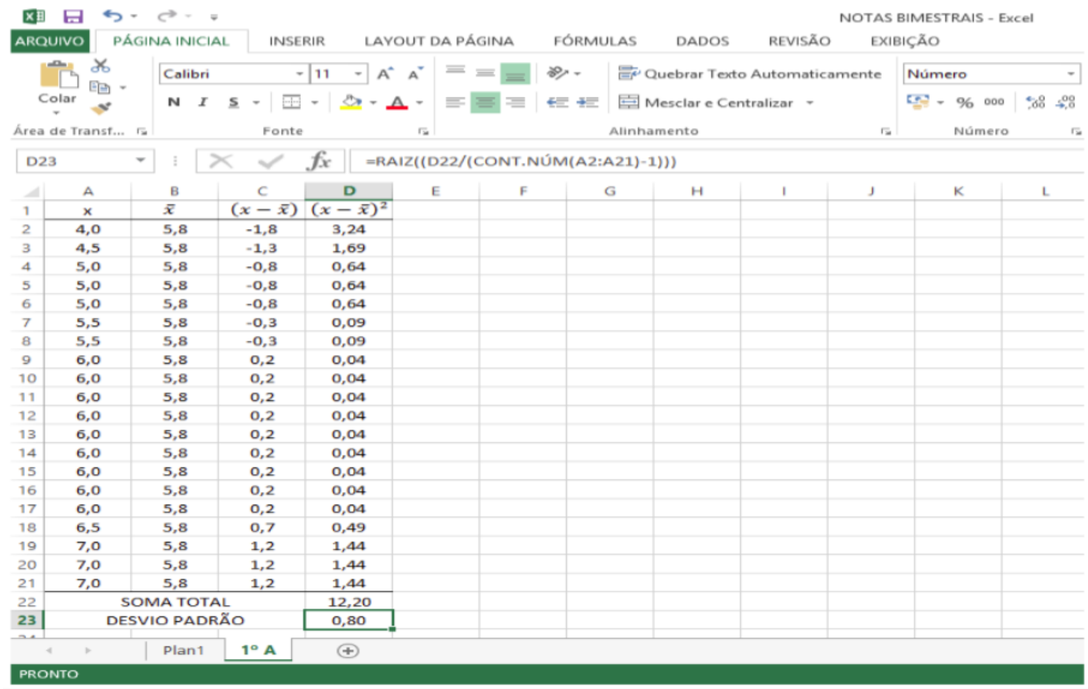
	A	B	C	D
1	x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
2			A2-B2	C2^2
3			A3-B3	C3^2
4			A4-B4	C4^2
5			A5-B5	C5^2
6			A6-B6	C6^2
7			A7-B7	C7^2
8			A8-B8	C8^2
9			A9-B9	C9^2
10			A10-B10	C10^2
11			A11-B11	C11^2
12			A12-B12	C12^2
13			A13-B13	C13^2
14			A14-B14	C14^2
15			A15-B15	C15^2
26			A16-B16	C16^2
17			A17-B17	C17^2
18			A18-B18	C18^2
19			A19-B19	C19^2
20			A20-B20	C20^2
21			A21-B21	C21^2
22	SOMA TOTAL			SOMA(D2:D21)
23	DESVIO PADRÃO			RAIZ((D22/(CONT.NÚM(A2:A21)-1)))

Depois de todos os comandos dados, o resultado do desvio padrão da turma do 1º A será igual a 0,8. (Figura 2.3.6).

Com o valor do desvio padrão calculado o aluno pode digitar em qualquer célula não utilizada na planilha a fórmula para o cálculo do desvio padrão, sendo que se o EXCEL instalado no computador estiver na versão 2007 ou anterior, então deve digitar =DESVPAD(A2:A21); caso contrário =DESVPAD.A(A2:A21).

Agora devem ser comparados os resultados que têm de ser iguais.

Figura 2.3.6: Cálculo do desvio padrão das notas da turma do 1º A

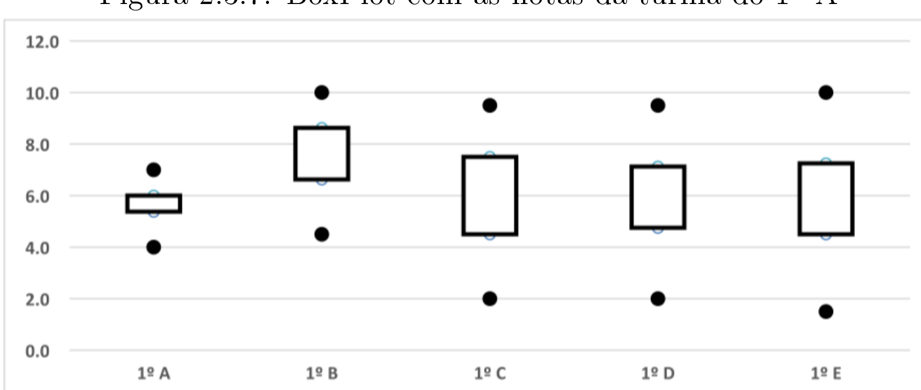


	A	B	C	D
1	x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
2	4,0	5,8	-1,8	3,24
3	4,5	5,8	-1,3	1,69
4	5,0	5,8	-0,8	0,64
5	5,0	5,8	-0,8	0,64
6	5,0	5,8	-0,8	0,64
7	5,5	5,8	-0,3	0,09
8	5,5	5,8	-0,3	0,09
9	6,0	5,8	0,2	0,04
10	6,0	5,8	0,2	0,04
11	6,0	5,8	0,2	0,04
12	6,0	5,8	0,2	0,04
13	6,0	5,8	0,2	0,04
14	6,0	5,8	0,2	0,04
15	6,0	5,8	0,2	0,04
16	6,0	5,8	0,2	0,04
17	6,0	5,8	0,2	0,04
18	6,5	5,8	0,7	0,49
19	7,0	5,8	1,2	1,44
20	7,0	5,8	1,2	1,44
21	7,0	5,8	1,2	1,44
22	SOMA TOTAL			12,20
23	DESVIO PADRÃO			0,80

Seguindo o mesmo processo o desvio padrão das turmas do 1º B, 1º C, 1º D e 1º E, são, respectivamente: 1,7; 2,0; 2,0; e 2,5.

Como não há um gráfico pronto para a média e o desvio padrão no EXCEL podemos utilizar o gráfico chamado BoxPlot, que utiliza a mediana, os valores máximos e mínimos, o primeiro e o terceiro quartis dos dados, e que pode ser construído em uma planilha com esses 5 valores. (Figura 2.3.7).

Figura 2.3.7: BoxPlot com as notas da turma do 1º A



Fazendo uma observação visual pode-se ver facilmente que o valor central do 1º B é maior e o restante é quase o mesmo. Já comparando o 1º A tem menor variabilidade que o 1º E, como verificamos em nossos cálculos.

É comum nos livros didáticos a situação problema acabar por aqui com o cálculo das medidas de tendência central e de variabilidade de um conjunto de dados. O desvio padrão não é de fácil compreensão para os alunos por se tratar de um algoritmo extenso, desde modo o professor tem que ocupar um tempo após o cálculo para investigar se os alunos compreenderam e auxiliá-los para suas conclusões sejam corretas. Sobre a variabilidade dos dados pode-se concluir que:

1. Mesmo tendo maior média entre as turmas o 1º B tem um desvio padrão de 1,7, observando-se que é alto, pois há alunos que estão bem distante da média da turma;

2. Apesar de terem notas diferentes a turma do 1º C e do 1º D têm mesma média e desvio padrão. Fato curioso que deve ser interpretado como algo que dificilmente acontece, mas que como aqui, pode acontecer;

3. As turmas do 1º A e do 1º E têm médias iguais, mas a variação das notas é significativamente diferente. O 1º A foi a turma com menor variação enquanto o 1º E teve a maior, mostrando que é a turma mais heterogênea. Com essa informação o professor poderia, por exemplo, utilizar da monitoria nessa turma aproveitando os alunos com notas maiores, que se destacaram, fazendo trabalhos em grupo na sala de aula, deixando cada um aluno desses alunos em grupos com os alunos que estão com menores notas.

2.4 Considerações Finais

Ao que foi posto neste capítulo, a aplicação da planilha eletrônica (EXCEL), no estudo da estatística, pode facilitar a compreensão dos alunos na leitura, análise e interpretação dos dados, tendo o professor sugestões para iniciar essa aplicação.

Capítulo 3

Fundamentação Teórica de Estatística

Neste capítulo são definidas variáveis, tratamentos, tipos de delineamentos, tipos de amostragens, tipos de dados, alguns elementos da estatística descritiva e alguns testes estatísticos.

3.1 Definição da Estatística

Estatística é uma coleção de métodos para o planejamento de experimentos, obtenção de dados e, consequente organização, resumo, apresentação, análise, interpretação e elaboração de conclusões baseadas nos dados. Desse processo todo faz parte a estatística descritiva e a inferência estatística [18].

Estatística Descritiva pode ser definida como os métodos que envolvem a coleta, a apresentação e a caracterização de um conjunto de dados de modo a descrever apropriadamente as várias características deste conjunto [11].

Inferência Estatística pode ser definida como os métodos que tornam possível a estimativa de uma característica de uma população ou a tomada de uma decisão referente à população com base somente em resultado de amostras [11].

3.2 Definição dos Modelos

Atividade Científica é o esforço para a coleta de dados na natureza e a construção de modelos que expliquem fenômenos naturais ou permitam a sua previsão.

Método Científico é o processo de abordagem científica com postulação de um problema ou fenômeno a ser investigado, seguido da formulação de um modelo e observação ou coleta de dados, sendo concluído com a verificação desse modelo.

Modelo Determinístico é o modelo que determina uma relação matemática precisa entre as variáveis envolvidas, devido ao grau de conhecimento relativamente alto do sistema sob estudo.

Modelo Não-Determinístico é o modelo que procura prever um resultado, em geral usando probabilidades, sem se preocupar com a especificação de agentes causais ou determinantes desse resultado.

Experimento é o processo que permite a coleta das observações sob determinadas condições.

Dados Experimentais são as observações obtidas dos experimentos.

3.3 Definição das Variáveis e Tratamento

Variável é a característica pela qual deseja-se que a população seja descrita, ou pela qual decisões serão tomadas.

Variável Fator é a variável cujo efeito se deseja conhecer e avaliar no experimento.

Variável Resposta é a variável a ser mensurada ou avaliada no experimento, gerando os dados experimentais.

Variável Condição Experimental que representa cada uma das condições em que será realizado o experimento.

Tratamento é utilizado para caracterizar os valores (tipos ou níveis) que um fator pode assumir em um determinado experimento.

Parcela Experimental é uma mesma quantidade de material experimental aplicado em cada tratamento.

Erro Experimental é a variabilidade entre parcelas que receberam o mesmo tratamento.

3.4 Tipos de Delineamento

Delineamento Inteiramente Casualizado (DIC) é utilizado quando a variabilidade entre as parcelas experimentais for pequena, isto é, praticamente inexistente. Por isso é utilizado onde as condições experimentais podem ser bem controladas. O modelo estatístico para o DIC é dado por:

$$y_{ij} = \mu + t_i + e_{ij}$$

sendo:
y_{ij} é a observação feita na parcela para o tratamento i na repetição j;
μ representa uma constante inerente a toda parcela;
t_i representa o efeito no tratamento j;
e_{ij} é o erro experimental na parcela i, j.

Delineamento em Blocos Casualizado (DBC) tem seu material experimental dividido em grupos homogêneos, cada grupo constituindo uma repetição. Cada repetição ou bloco deve conter uma vez cada tratamento, no caso de blocos completos. O objetivo em todas as etapas do experimento, é manter o erro, dentro de cada bloco, tão pequeno quanto seja possível na prática. O modelo estatístico para o DBC é dado por:

$$y_{ij} = \mu + t_i + b_i + e_{ij}$$

sendo:
y_{ij} é a observação feita na parcela para o tratamento i na repetição j;
μ representa uma constante inerente a toda parcela;
t_i representa o efeito no tratamento j;
b_i é o efeito do bloco j;
e_{ij} é o erro experimental na parcela i, j.

Delineamento Quadrado Latino (DQL) tem os tratamentos agrupados nas repetições de duas maneiras distintas. Essa sistematização dos blocos em duas direções, designadas por linhas e colunas, permite eliminar os efeitos de duas fontes de variação do erro experimental. O modelo estatístico para o DQL é dado por:

$$y_{i\ jk} = \mu + t_i + c_j + l_k + e_{i\ jk}$$

sendo:
$y_{i\ jk}$ é a observação da parcela na coluna j e na linha k, recebendo o tratamento i;
μ representa uma constante inerente a todas as parcelas;
t_i é o efeito no tratamento j;
c_i é o efeito na coluna j;
l_k é o efeito na linha k;
$e_{i\,jk}$ é o erro aleatório nas parcelas i, j e k, sendo i, j, k = 1, 2, ..., l.

3.5 Definição dos Elementos da Amostragem

População corresponde ao sistema ou ao todo que se quer descrever. É sempre um conjunto de elementos com característica em comum. Geralmente representada por N.

Amostra é um subconjunto ou parte da população, cujos elementos são avaliados utilizando uma ou mais variáveis descritoras. Geralmente representada por n.

Amostragem é o estudo das relações existentes entre a amostra e a população de onde esta foi extraída. Ela é usualmente realizada com o objetivo de estimar parâmetros da população.

Parâmetro é um valor numérico que descreve alguma característica de uma população.

Estimador é uma função ou processo numérico que permite a geração de estimativas.

Estimativa é a aproximação numérica particular para um parâmetro associado a função gerada pelo estimador.

3.6 Tipos de Amostragem

Amostragem Aleatória Simples (AAS) é obtida quando todos os elementos da população têm a mesma probabilidade, diferente de zero, de pertencer à amostra. Esse tipo de amostragem é recomendado quando as populações são homogêneas. Para a obtenção de uma AAS de tamanho n, procede-se ao sorteio com ou sem reposição, dentre os N elementos de uma população, até que se complete a amostra.

Amostra Aleatória Estratificada (AAE) subdivide-se a população em subpopulações de forma que dentro das subpopulações haja homogeneidade. Este processo chama-se estratificação da população, sendo cada subpopulação um estrado.

Amostra Aleatória Sistemática (AAS) a população deve estar ordenada de tal forma que cada um de seus elementos possa ser unicamente identificado pela sua localização ou por outro critério qualquer. Para efetuar a seleção da amostra, procede-se o sorteio de um ponto de partida entre 1 e o fator de expansão k, definido pela razão entre o número de elementos da população e o número de elementos da amostra, isto é, $k = N/n$.

Amostra Aleatória por Conglomerado (AAC) é utilizada quando a população é muito dispersa, inviabilizando a organização de um rol com todos elementos. Na AAC a população é dividida em subpopulações ou conglomerados, sendo alguns destes sorteados para constituir a amostra.

3.7 Tipos de Dados

Dados Qualitativos relatam uma qualidade que é uma característica não-numérica do conjunto de dados.

Dados Qualitativos Ordenativos ou Ranks são dados não-numéricos que podem ser classificados de acordo com uma ordem crescente ou decrescente de valores dos dados.

Dados Qualitativos Categóricos são dados não-numéricos que não podem ser classificados em alguma ordem, podem ser dicotômicos (duas categorias mutuamente exclusivas) ou multinomiais (mais de duas categorias).

Dados Quantitativos ou Numéricos consistem em números que representam contagem ou medidas do conjunto de dados.

Dados Numéricos Discretos surgem do processo de contagem, isto é, só podem assumir valores naturais.

Dados Numéricos Contínuos resultam de infinitos valores possíveis reais em escala que não apresenta intervalo vazio de números.

3.8 Estatística Descritiva

A estatística descritiva analisa e interpreta o conjunto de dados estudado, caracterizando-o por medidas de tendência central, de variabilidade e representando-o por um gráfico ou tabela. Se essas medidas descritivas forem calculadas através de uma amostra de dados, elas serão chamadas de estatísticas; caso sejam calculadas através de uma população de dados, elas serão chamadas de parâmetros. Nas **medidas de tendência central** pode-se incluir a média, a mediana, a moda.

Média aritmética é calculada somando-se todas as observações de um conjunto de dados e dividindo-se pelo número total de dados envolvidos. Ela é representada por $\overline{x}$, e sua fórmula é dada por:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n},$$

sendo $\sum$ a adição dos dados e n o total de dados se os dados forem amostrais; se os dados forem de uma população a média será representada por μ e o total de dados por N.

Mediana é o valor do meio em uma sequência de dados arranjada em ordem crescente (ou decrescente), e é representada por $\widetilde{x}$. Seu cálculo dependo do número de dados, se for ímpar a fórmula do ponto de posicionamento será dada por:

$$\frac{n+1}{2}$$

e, se for par, será calculada pela média aritmética entre os valores dos dados que estão no meio da sequência numérica, e sua fórmula será a média dos valores que estiverem nas posições:

$$\frac{n}{2} \text{ e } \frac{n+1}{2}$$

Moda é o valor que ocorre com maior frequência, sendo representado por $\boldsymbol{M}$. Pode ocorrer que não exista um valor que mais se repita nos dados analisados, passando o conjunto de dados a ser chamado de **amodal**; se ocorrer dois valores que mais repetem entre os dados o conjunto de dados será chamado **bimodal**; e, se mais de dois valores se repetirem com a mesma frequência o conjunto de dados será chamado **multimodal**.

Nas **medidas de variabilidade** pode-se incluir a amplitude, o desvio padrão e a variância.

Amplitude é a diferença entre o maior e o menor valor de um conjunto de dados. Para seu cálculo pode-se usar a fórmula:

$$A = x_{maior} - x_{menor}$$

sendo x_{maior} o maior valor e x_{menor} o menor valor do conjunto de dados.

Desvio padrão é a medida de variação dos dados amostrais em torno da média. É representado por s; se os dados forem de uma população será representado por σ. Seu cálculo é igual a raiz quadrada da soma das diferenças ao quadrado em torno da média aritmética dividida pelo tamanho da amostra (**n**) menos 1, se for populacional será dividido pelo tamanho da população ($\boldsymbol{N}$). Sua fórmula é dada por:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} \text{ e } \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}}$$

Variância é a medida de variação igual ao quadrado do desvio padrão. É representado por (amostral) e (populacional). Para calcular a variância deve-se utilizar a fórmula do desvio padrão elevado ao quadrado.

3.9 Testes de Hipóteses

Teste de Kolmogorov-Smirnov – Destinado a medir o grau de concordância entre a distribuição acumulada de valores observados da amostra com os dados teoricamente esperados. É, portanto, um teste de normalidade. É calculado por:

$$D = MÁXIMO\,|F_0(x) - S_n(x)|,$$

sendo:

$F_0(x)$ a distribuição teórica acumulada e $S_n(x)$ a distribuição amostral acumulada.

Teste de Shapiro-Wilk – Utilizado para testar a normalidade dos dados, comparando a distribuição dos valores dos dados com os valores aproximadamente normais. É calculado por:

$$W = \frac{b^2}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2},$$

em que x_i são os valores de amostras ordenados (x_1 é o menor). A constante b é determinada da seguinte forma

$$b = \begin{cases} \sum\limits_{i=1}^{\frac{n}{2}} \quad a_{n-i+1} * (x_{(n-1)} - x_i), \ \ se\ n\ é\ par \\ \\ \sum\limits_{i=1}^{\frac{(n+1)}{2}} \quad a_{n-i+1} * (x_{(n-1)} - x_i), \ \ se\ n\ é\ ímpar, \end{cases}$$

em que a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal.

Teste Binomial para duas proporções – Mede a diferença entre as proporções de duas amostras independentes. É calculado por:

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} \text{ e } \hat{q} = 1 - \hat{p}, \text{ substituindo } \hat{p} \text{ e } \hat{q} \text{ em } z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

sendo:

$\hat{p}$ a proporção amostral;
$\hat{p_1}$ a proporção da primeira amostra e $\hat{p}_2$ da segunda;
n_1 o tamanho da primeira amostra e n_2 da segunda.

Teste t de Student – Analisa a diferença entre duas amostras, dependentes ou independentes, de mesmo tamanho ou desiguais. O tamanho das amostras deve ser no máximo iguais a 30. É calculado por:

$$s_{dif} = \sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \text{ e logo após, } t = \frac{\overline{X}_1 - \overline{X}_2}{s_{dif}},$$

sendo:

n_1 o tamanho da primeira amostra e n_2 da segunda;

s_1^2 a variância amostral da primeira amostra e s_2^2 da segunda amostra;

$\overline{X}_1$ a média amostral da primeira amostra e $\overline{X}_2$ a média amostral da segunda amostra, com graus de liberdade igual a $n_1 + n_2 - 2$.

Teste de Mann-Whitney – Testa a hipótese de que duas populações têm a mesma distribuição. É um teste Não-Paramétrico alternativo ao teste t de Student para amostras independentes. É calculado por:

se $n_1 = n_2$, então calcula-se apenas $U = \sum R_1 - \left[\dfrac{n_1(n_1+1)}{2}\right]$,

mas se $n_1 \neq n_2$, então também é necessário calcular $U = \sum R_2 - \left[\dfrac{n_2(n_2+1)}{2}\right]$,

sendo n_1 o número de dados da primeira amostra e n_2 da segunda amostra.

Escolhe-se o menor valor de e o substitui no cálculo de z dado por

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}.$$

Teste de Wilcoxon - Testa a hipótese de que duas populações têm a mesma distribuição, desde que estejam pareadas. É um teste Não-Paramétrico alternativo ao teste t de Student para amostras dependentes. É calculado por:

$$z = \frac{\sum R}{\sqrt{\sum R^2}}.$$

sendo R os opostos assinalados.

Teste z – Destinado a analisar a diferença estatística entre duas amostras independentes, do mesmo tamanho ou desiguais. O tamanho das amostras deve ser no mínimo iguais a 30. É calculado como no teste t sem os graus de liberdade.

Teste x^2 (Qui-Quadrado) – É um teste Não-Paramétrico destinado a comprovar se dois grupos diferentes diferem em relação a determinada categoria, ou seja, suas frequências observadas e esperadas são diferentes. É calculado por:

$$x^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

sendo:

r o número de linhas;

k o número de colunas e

o grau de liberdade dado por $g_l = (r-1)(k-1)$.

Teste ANOVA - Destinado a analisar a diferença estatística entre três ou mais amostras independentes, do mesmo tamanho ou desiguais. É calculado por:

com $QM_{TRATAMENTOS} = \dfrac{SQ_{TRATAMENTOS}}{g_{l_{TRATAMENTOS}}}$ e $QM_{ERRO} = \dfrac{SQ_{ERRO}}{g_{l_{ERRO}}}$, calcula-se $F = \dfrac{QM_{TRATAMENTOS}}{QM_{ERRO}}$.

Teste de Kruskal-Wallis – Serve para testar se três ou mais amostras têm a mesma distribuição. É um teste Não-Paramétrico alternativo ao teste ANOVA. Tendo k grupos observados e $n_1,\ n_2, ...,\ n_k$ casos, respectivamente, é calculado por

$$H = \frac{12}{n(n+1)}\left[\frac{\left(\sum R_1\right)^2}{n_1} + \frac{\left(\sum R_2\right)^2}{n_2} + ... + \frac{\left(\sum R_k\right)^2}{n_k}\right] - 3(n+1),$$

em que $\sum R_1,\ \sum R_2, ...,\ \sum R_k$ são as somas dos postos dos grupos 1, 2, ..., k, respectivamente; $n_1,\ n_2, ...,\ n_k$ representam o número de casos nos respectivos grupos; e n é o número total de casos, isto é, $n_1 + n_2 + ... + n_k$.

Teste de Tukey – Utilizado quando o teste ANOVA mostra que há diferenças significativas entres as amostras, sendo um teste post-hoc. Ele compara de duas em duas as amostras para mostrar onde estão as diferenças significativas. É calculado por:

$$DMS = q_{(l,\ v)}\sqrt{\frac{QM_{ERRO}}{J}},$$

sendo:

$q_{(l,\ v)}$ a amplitude total estudentizada com l tratamentos, v graus de liberdade e J número de repetições.

Teste de Student-Newman-Keuls - Utilizado quando o teste de Kruskal-Wallis mostra que há diferenças significativas entres as amostras, sendo um teste post-hoc. Ele compara de duas em duas as amostras para mostrar onde estão as diferenças significativas. É calculado por:

$$DMS = q_i\sqrt{\frac{QM_{ERRO}}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)},$$

sendo:

$i + p + 2$, com p médias existentes entre as duas médias comparadas.

Teste de Friedman – Serve para testar se vários grupos relacionados têm, todos, a mesma distribuição. É um teste Não-Paramétrico alternativo ao teste ANOVA com dupla entrada. É calculado por:

$$X_r^2 = \frac{12}{Nk(k+1)}\sum R^2 - 3N(k+1),$$

sendo:

N o número de linhas; k o número de grupos; $\sum R = \left(\sum R_1\right)^2 + \left(\sum R_2\right)^2 + ... + \left(\sum R_k\right)^2$ a soma dos quadrados dos postos dos grupos 1, 2,..., k, respectivamente.

Teste de Correlação de Pearson – Mostra o grau de associação (ou dependência) entre duas variáveis. É calculado por:

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} * \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}} \text{ e } t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}},$$

sendo:

r o coeficiente linear de Pearson;

n o número de pares dos dados presentes;
x e y valores das variáveis.

Teste de Correlação de Spearman – É um teste Não-Paramétrico que mostra o grau de associação (ou dependência) entre duas variáveis, ou seja, é uma alternativa para o teste de correlação de Pearson quando a distribuição dos dados não for numérica ou não se apresentar de forma normal. É calculado por:

$$r_s = 1 - \left[\frac{6\sum D^2}{n\left(n^2-1\right))}\right] \text{ e } t = \frac{r_s}{\sqrt{\dfrac{1-r_s^2}{n-2}}},$$

sendo:

r_s o coeficiente de Spearman;
D a diferença entre os pares de postos;
n o número de pares dos dados presentes.

3.10 Considerações Finais

O referido estudo desse capítulo fornece ao professor elementos que permitem a compreensão de alguns aspectos da estatística essenciais para o desenvolvimento da análise de dados.

Capítulo 4

Aplicações da Estatística Não-Paramétrica em um Experimento da Área de Ciências Agrárias

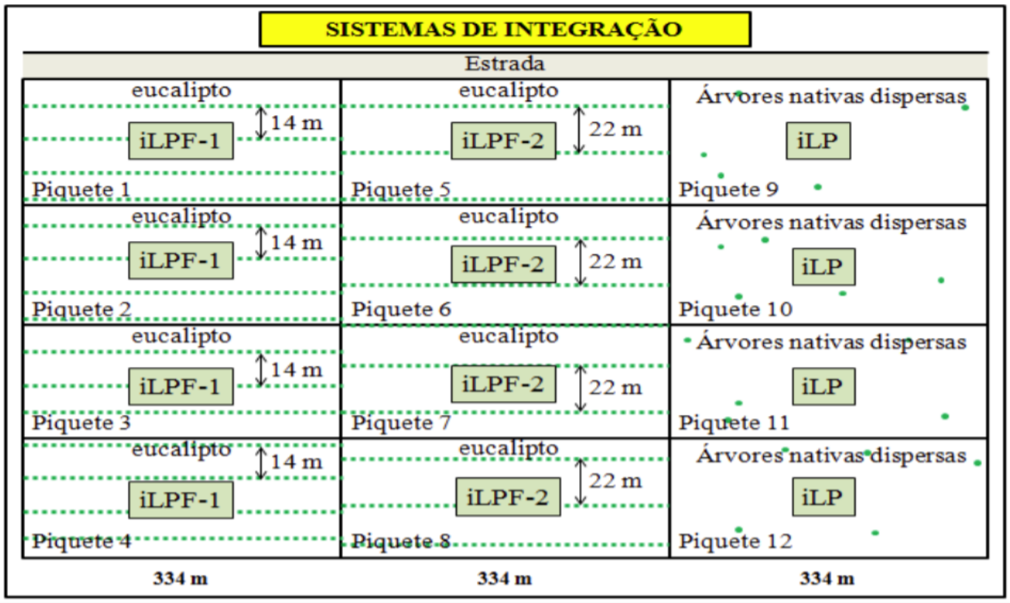
Este capítulo destina-se a aplicação do cálculo estatístico em um projeto da Embrapa. A primeira análise foi a investigação do microclima em sistemas de integração e características quanti-qualitativas de espécies arbóreas nativas e cultivada no cerrado, e a segunda, o comportamento ingestivo diurno de bezerras Nelore. O experimento ocorreu na Embrapa Gado de Corte em Campo Grande-MS, de julho de 2013 à a abril de 2014.

A Embrapa Gado de Corte em Campo Grande – MS, tem como missão 'viabilizar soluções tecnológicas sustentáveis para a cadeia produtiva da pecuária de corte em benefício da sociedade brasileira'. Para que isso se concretize seus pesquisadores e equipes conduzem pesquisas visando conhecer e aperfeiçoar a tecnologia existente na área de bovinos de corte.

O conforto térmico dos animais, em regime de pastagem, é condição essencial para garantir a boa produtividade dos rebanhos, sendo que, animais expostos a forte radiação solar e, consequentemente, a altas temperaturas, entram em estado de estresse térmico, alterando sua frequência respiratória e cardíaca e, principalmente, sua taxa metabólica, ou seja, sua capacidade de transformar em carne, leite ou energia o que ingerem.

O experimento foi realizado em área experimental com três sistemas de integração (iLPF-1, com 357 árvores/ha, iLPF-2, com 227 árvores/ha e iLP, com 5 árvores nativas remanescentes/ha), em pastagens de *Brachiaria brizantha* cv. BRS Piatã, associada ao *Eucalyptus grandis* x *urophilla* (híbrido H13) nos sistemas de iLPF e espécies nativas no iLP (Figura 4.0.1). Os sistemas de integração, que em sua composição apresentam o componente arbóreo para fornecimento de sombra, têm demonstrado eficiência em trocar calor com o ambiente, sendo que as características da copa influenciam na quantidade de sombra projetada.

Figura 4.0.1: Representação esquemática da área experimental



4.1 Comportamento ingestivo de bovino de corte

Essa pesquisa foi realizada pela Me. Caroline Carvalho de Oliveira (Universidade Federal dos Vales do Jequitinhonha e Mucuri) orientada pela pesquisadora Dr.ª Fabiana Villa Alves da Embrapa, que atua na área de Sistemas de Produção Sustentáveis (Manejo Animal).

No estudo do comportamento ingestivo de bovino de corte foram utilizadas 80 novilhas nelores, com peso médio inicial de 157 kg, observadas em dois períodos, manhã (06h00 às 12h00) e tarde (12h01min às 17h00) (Figura 4.1.1). Em cada piquete foram mantidos dois animais-teste e inserido animais reguladores para ajuste das alturas de pastejo. As avaliações de comportamento ingestivo diurno foram realizadas por meio de observação instantânea dos animais-teste (método "scan-sampling"; SETZ, 1991), com uso de binóculos, a intervalos de 10 minutos, anotando-se as atividades em planilhas específicas. As variáveis comportamentais estudadas foram: pastejo, ruminação, ócio, ingestão de água, ingestão de sal mineral, roçando árvores e outras atividades, em duas diferentes posições corporais (em pé e deitado) e duas posições espaciais (ao sol e a sombra). Foram considerados expostos ao sol, quando 50% ou mais de seu corpo encontravam-se sob o sol, e, à sombra, quando 51% ou mais de seu corpo estivesse exposto à sombra, durante quatro dias consecutivos, em cada estação do ano (seca e águas). As frequências das variáveis comportamentais foram transformadas em minutos (min), em cada período do dia, totalizando 300 min. por período.

Figura 4.1.1: Sistemas de integração (ILPF)

Na comparação das variáveis foi utilizado um método para se testar a igualdade de três ou mais médias populacionais através da análise das variâncias amostrais, pois são nove as variáveis comparadas: pastejo no sol (PASTSOL), pastejo na sombra (PASTSOM), ruminação no sol (RUMISOL), ruminação na sombra (RUMISOM), ócio no sol (OCISOL), ócio na sombra (OCISOM), bebendo água (BEB), mineralizando (MIN), e roçando árvores (ROCA).

Quando os dados são comparados com a aplicação de testes estatísticos segue-se fundamentos conceituais da metodologia do teste de hipóteses [18]. Primeiro escolhe-se as hipóteses: nula e alternativa; a nula especifica a afirmação que contém uma igualdade e a alternativa, a hipótese que deve ser verdadeira caso a nula seja falsa. Nos problemas levantados nesse projeto a hipótese nula sempre foi a de igualdade entre as variáveis comparadas. Segue-se a escolha de um nível de significância que é a probabilidade de ocorrer a rejeição da hipótese nula quando ela é verdadeira, isso pode ocorrer porque utilizamos uma estimativa como parâmetro da população utilizando dados amostrais. O nível de significância selecionado para as comparações foi estabelecido em 0,05.

O próximo passo é a escolha do melhor teste para aplicar nas comparações. Existem testes para comparar uma proporção, duas proporções, a média com desvio padrão conhecido e desconhecido, duas médias dependentes ou independentes, análise de variâncias, correlações, entre outros. Antes de escolher o teste deve-se verificar se os testes serão paramétricos ou não, se houver suposições sobre a natureza da população, como se a distribuição dos dados é normal ou os dados são numéricos, os testes devem ser paramétricos, caso contrário utiliza-se os testes não paramétricos. Apesar da lógica e execução dos testes não paramétricos ser mais fácil do que os paramétricos, eles têm desvantagens: são menos poderosos, ou seja, têm menor probabilidade, do que os paramétricos, de rejeitar a hipótese de nulidade quando esta é falsa [18].

Após aplicação do teste com os dados amostrais utiliza-se uma tabela com a estatística de teste para comparar o valor crítico, que é qualquer valor que separa a região crítica da tabela onde rejeitamos a hipótese nula. Outro método para comparação do resultado é o valor p que é a probabilidade de se obter um valor da estatística de teste no mínimo tão extremo quanto o que representa os dados amostrais, supondo que a hipótese nula seja verdadeira [18]. Esse valor é amplamente usado porque é de fácil interpretação do resultado para a conclusão de rejeitar a hipótese nula. Nos testes aplicados a seguir quando p for extremamente pequeno, tendo valor no máximo igual a 0,05 rejeitamos a hipótese nula.

Como muitos testes estatísticos são extremamente extensos e cansativos em seus cálculos, utiliza-se softwares estatísticos que auxiliam o usuário a chegar a uma conclusão sobre as comparações feitas. Nos testes aplicados com os dados da Embrapa utilizou-se o programa estatístico BioEstat versão 5.3 [26].

O teste de Shapiro-Wilk é um dos Testes usados para testar a Normalidade das amostras e que pode mostrar se há possibilidade de aplicação de um teste paramétrico. Ele é baseado na estatística W, e é dado por

$$W = \frac{b^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

em que x_i são os valores de amostras ordenados (x_1 é o menor). A constante b é determinada da seguinte forma

$$b = \begin{cases} \sum_{i=1}^{\frac{n}{2}} a_{n-i+1} * (x_{n-1} - x_i), & \text{se } n \text{ é par} \\ \\ \sum_{i=1}^{\frac{(n+1)}{2}} a_{n-i+1} * (x_{n-1} - x_i), & \text{se } n \text{ é ímpar} \end{cases}$$

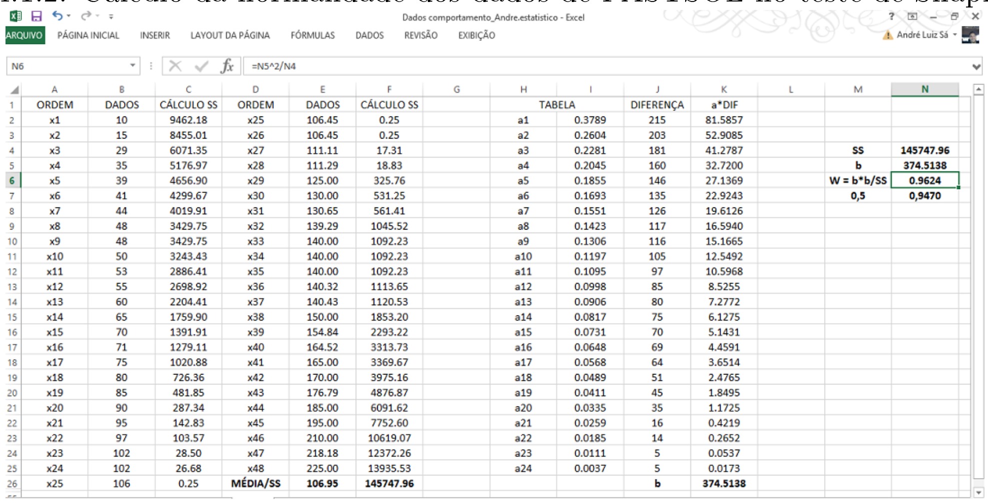
em que a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal.

No início do teste deve-se colocar em ordem crescente os elementos de cada variável, sendo que cada uma tem n = 48 amostras. Com os dados organizados deve-se calcular SS como se segue:

$$\mathrm{SS} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Na variável PASTSOL o valor encontrado para SS foi de 145.747,96 (Figura 4.1.2). Como n é par o cálculo de b terá até n/2 = 24 parcelas. Os valores de a_1 até a_{24} são encontrados na tabela de Shapiro-Wilk. Será calculada a diferença entre x_{48} e $x_1, ..., x_{25}$ e x_{24} é o i-ésimo elemento de dados na ordem de classificação. Após multiplicar a diferença pelo valor correspondente de a_n, somam-se esses valores obtendo o valor de b = 374,51. Dividindo b^2 por SS obtém-se o valor da estatística W = 0,9623. Como o valor crítico, na tabela de Shapiro-Wilk para W para $\alpha = 0,05$, é de 0,9470, temos que a estatística W calculada é maior, não rejeitamos a hipótese nula de que os dados são normalmente distribuídos.

Figura 4.1.2: Cálculo da normalidade dos dados de PASTSOL no teste de Shapiro-Wilk



ORDEM	DADOS	CÁLCULO SS	ORDEM	DADOS	CÁLCULO SS	TABELA		DIFERENÇA	a*DIF		
x1	10	9462.18	x25	106.45	0.25	a1	0.3789	215	81.5857		
x2	15	8455.01	x26	106.45	0.25	a2	0.2604	203	52.9085		
x3	29	6071.35	x27	111.11	17.31	a3	0.2281	181	41.2787	**SS**	**145747.96**
x4	35	5176.97	x28	111.29	18.83	a4	0.2045	160	32.7200	**b**	**374.5138**
x5	39	4656.90	x29	125.00	325.76	a5	0.1855	146	27.1369	**W = b*b/SS**	**0.9624**
x6	41	4299.67	x30	130.00	531.25	a6	0.1693	135	22.9243	**0,5**	**0,9470**
x7	44	4019.91	x31	130.65	561.41	a7	0.1551	126	19.6126		
x8	48	3429.75	x32	139.29	1045.52	a8	0.1423	117	16.5940		
x9	48	3429.75	x33	140.00	1092.23	a9	0.1306	116	15.1665		
x10	50	3243.43	x34	140.00	1092.23	a10	0.1197	105	12.5492		
x11	53	2886.41	x35	140.00	1092.23	a11	0.1095	97	10.5968		
x12	55	2698.92	x36	140.32	1113.65	a12	0.0998	85	8.5255		
x13	60	2204.41	x37	140.43	1120.53	a13	0.0906	80	7.2772		
x14	65	1759.90	x38	150.00	1853.20	a14	0.0817	75	6.1275		
x15	70	1391.91	x39	154.84	2293.22	a15	0.0731	70	5.1431		
x16	71	1279.11	x40	164.52	3313.73	a16	0.0648	69	4.4591		
x17	75	1020.88	x41	165.00	3369.67	a17	0.0568	64	3.6514		
x18	80	726.36	x42	170.00	3975.16	a18	0.0489	51	2.4765		
x19	85	481.85	x43	176.79	4876.87	a19	0.0411	45	1.8495		
x20	90	287.34	x44	185.00	6091.62	a20	0.0335	35	1.1725		
x21	95	142.83	x45	195.00	7752.60	a21	0.0259	16	0.4219		
x22	97	103.57	x46	210.00	10619.07	a22	0.0185	14	0.2652		
x23	102	28.50	x47	218.18	12372.26	a23	0.0111	5	0.0537		
x24	102	26.68	x48	225.00	13935.53	a24	0.0037	5	0.0173		
x25	106	0.25	**MÉDIA/SS**	**106.95**	**145747.96**			**b**	**374.5138**		

De forma análoga as outras variáveis tiveram valor da estatística W igual a: 0,9395 (PASTSOM), 0,9084 (RUMISOM), 0,8639 (RUMISOL), 0,8681 (OCISOL), 0,8201 (OCISOM), 0,7817 (BEB), 0,3564 (MIN) e 0,3767 (ROCA). Como todos os valores são menores que o valor crítico, rejeitamos a hipótese nula de que essas distribuições são normalmente distribuídas.

Utilizando o programa estatístico obtém-se esses resultados e o valor p, que confirma apenas a distribuição dos dados de PASTOL como normalmente distribuída, sendo os valores p iguais a: 0,2592 (PASTSOM), 0,0274 (RUMISOM), 0,0098 (RUMISOL), 0,0093 (OCISOL), 0,0093 (OCISOM), 0,0088 (BEB), 0,0038 (MIN) e 0,0041 (ROCA).

Logo, aplica-se o teste de Kruskal-Wallis que é um teste Não-Paramétrico, ou seja, uma alternativa para a análise de variância com uma classificação. Esse teste também é chamado de ANOVA by ranks test, pois faz as comparações com postos ou ranks (como em todos testes Não-Paramétricos), substituindo os valores numéricos das amostras em ordem e em sequência pelos postos. Se houver diferenças significativas entre as variáveis deve-se aplicar o teste post hoc SNK (Student-Newman-Keuls), que irá comparar de duas em duas as variáveis para mostrar quais variáveis têm diferenças.

O teste de Kruskal-Wallis com k grupos observados e $n_1,\ n_2, ..., n_k$ casos, respectivamente, é dado por

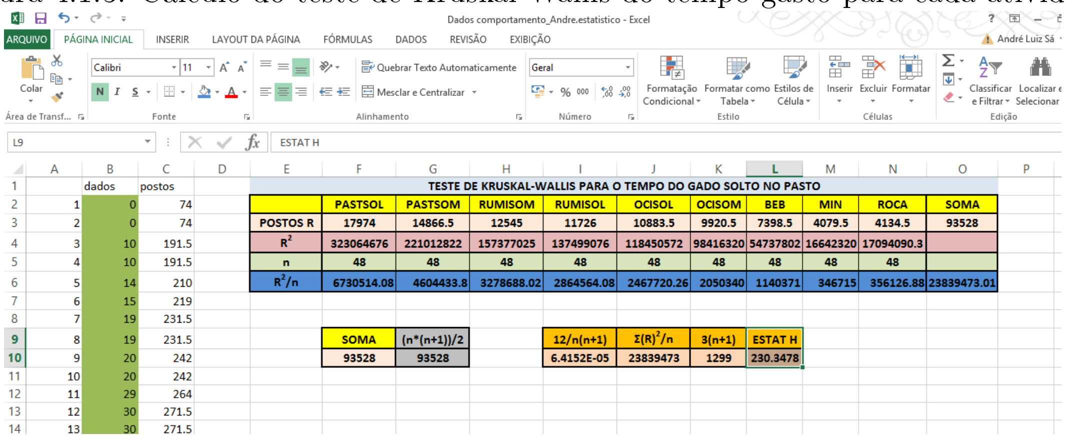
$$H = \frac{12}{n(n+1)}\left[\frac{(\sum R_1)^2}{n_1} + \frac{(\sum R_2)^2}{n_2} + ... + \frac{(\sum R_k)^2}{n_k}\right] - 3(n+1)$$

em que $\sum R_1$, $\sum R_2$,..., $\sum R_k$ são as somas dos postos dos grupos 1, 2,..., k, respectivamente; n_1, n_2,..., n_k representam o número de casos nos respectivos grupos; e n é o número total de casos, isto é, $n_1 + n_2 + ... + n_k$.

Na aplicação do teste primeiro junta-se os 48*9 $=$ 432 dados em um só conjunto para atribuição dos postos de acordo com a ordem crescente dos dados (Figura 4.1.3). Deve-se anotar o somatório dos postos de cada grupo (F3:N3) e calcular o quadrado de cada um

(F4:N4), lembrando que cada grupo contém n = 48 amostras (F5:N5). Antes de prosseguir é importante conferir se os cálculos estão corretos, sabendo que a soma dos postos (03 e F10) sempre será igual a metade do produto do número de dados pelo seu sucessor imediato (G10). De acordo com a fórmula citada acima, o valor da estatística H é dado pelo produto representado em I10 por J10 subtraído por K10, conforme o resultado em L10.

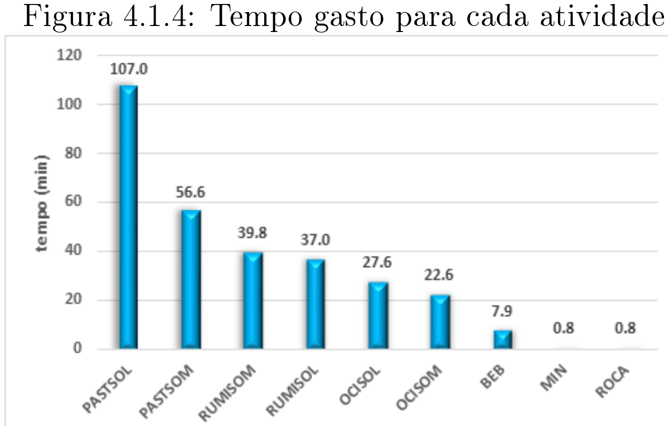
Figura 4.1.3: Cálculo do teste de Kruskal-Wallis do tempo gasto para cada atividade



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		dados	postos		TESTE DE KRUSKAL-WALLIS PARA O TEMPO DO GADO SOLTO NO PASTO										
2	1	0	74			PASTSOL	PASTSOM	RUMISOM	RUMISOL	OCISOL	OCISOM	BEB	MIN	ROCA	SOMA
3	2	0	74		POSTOS R	17974	14866.5	12545	11726	10883.5	9920.5	7398.5	4079.5	4134.5	93528
4	3	10	191.5		R^2	323064676	221012822	157377025	137499076	118450572	98416320	54737802	16642320	17094090.3	
5	4	10	191.5		n	48	48	48	48	48	48	48	48	48	
6	5	14	210		R^2/n	6730514.08	4604433.8	3278688.02	2864564.08	2467720.26	2050340	1140371	346715	356126.88	23839473.01
7	6	15	219												
8	7	19	231.5												
9	8	19	231.5			SOMA	(n*(n+1))/2		12/n(n+1)	$\Sigma(R)^2/n$	3(n+1)	ESTAT H			
10	9	20	242			93528	93528		6.4152E-05	23839473	1299	230.3478			
11	10	20	242												
12	11	29	264												
13	12	30	271.5												
14	13	30	271.5												

O valor crítico na tabela de Kruskal-Wallis para 8 graus de liberdade e para $\alpha = 0,05$ é de 15,507. Utilizando o programa estatístico obtém-se como valor da estatística H o valor 230,5217. Logo, pode-se afirmar que há diferença significativa no tempo gasto pelo gado durante o dia pelo gado solto no pasto (p<0,0001). Apesar de poder-se ter essa mesma conclusão olhando a figura 4.1.4, é necessário a confirmação do teste estatístico. Como há diferença significativa entre as variáveis, aplica-se o teste SNK que compara de duas em duas as variáveis para mostrar quais têm diferenças significativas entre si (Tabela 4.1.1).

Figura 4.1.4: Tempo gasto para cada atividade



Com o teste de SNK pode-se concluir que o tempo gasto pelo gado pastando no sol (35,7% do tempo coletado) é maior que nas outras circunstâncias (Tabela 4.1.1).

Tabela 4.1.1: Comparação entre as variáveis pelo tempo de acordo com o comportamento ingestivo de bovino de corte

Variável	Médias (em minutos)	Comparações
PASTSOL	107,0	a
PASTSOM	56,6	b
RUMISOL	39,8	bc
RUMISOM	37,0	cd
OCISOL	27,6	d
OCISOM	22,6	d
BEB	7,9	e
MIN	0,8	f
ROCA	0,8	f

Na comparação do turno de coleta dos dados (manhã e tarde), da estação (seca e nas águas) e na altura do pastejo (15 cm e 30 cm), tem-se a comparação das variáveis de duas em duas. Portanto, o teste para essas comparações é o teste de Mann-Whitney (que é a alternativa de teste independente Não-Paramétrico para o teste t de Student).

No cálculo do teste de Mann-Whitney é necessário primeiro calcular U, da seguinte forma

$$se\ n_1 = n_2,\ então\ calcula-se\ apenas\ U = \sum R_1 - \left[\frac{n_1(n_1+1)}{2}\right]$$

$$mas\ se\ n_1 \neq n_2,\ então\ também\ é\ necessário\ calcular\ U = \sum R_2 - \left[\frac{n_2(n_2+1)}{2}\right]$$

sendo n_1 o número de dados da primeira amostra e n_2 o da segunda amostra.

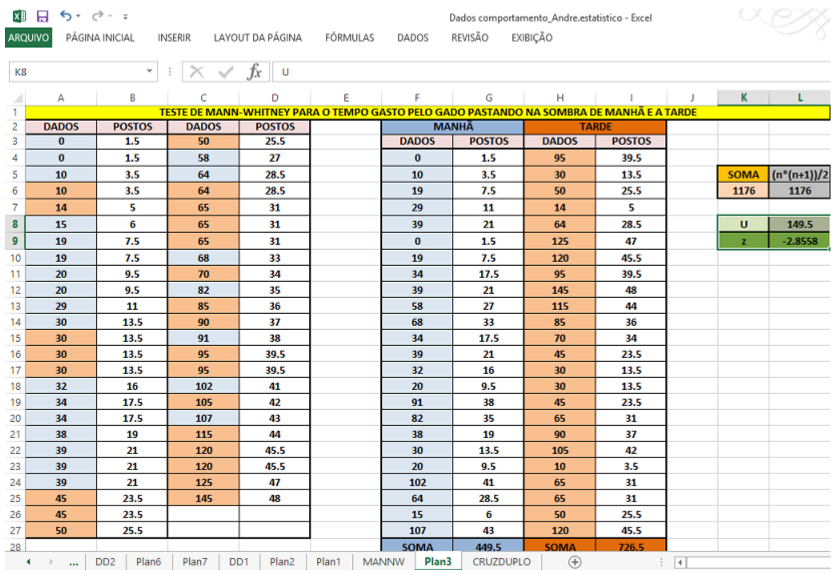
Escolhe-se o menor valor de U e o substitui no cálculo de z dado por

$$z = \frac{U - \dfrac{n_1 n_2}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

que tem, aproximadamente, distribuição normal padronizada.

Para o cálculo da estatística U durante o tempo que o gado pastou na sombra (PASTSOM), ordena-se os dados dos dois grupos em ordem crescente para cálculo dos postos (Figura 4.1.5). Após isso soma-se os postos de cada grupo e usa-se para cálculo de U o valor menor encontrado na soma dos postos (G28), e utilizando a fórmula para cálculo de U acima, tem-se que U = 149,5 (L8). Antes de prosseguir é importante conferir se os cálculos estão corretos, sabendo que a soma dos postos (K6) sempre será igual a metade do produto do número de dados pelo seu sucessor imediato (L6). Substituindo o valor encontrado de U na fórmula para cálculo de z, citada acima, com $n_1 = n_2 = 24$, obtém-se z = -2,8558 (L9). O valor crítico na tabela z para $\alpha = 0,05$, é igual a 1,96. Como z calculado excede o valor crítico, há diferença significante entre manhã e tarde, sendo que o gado pasta mais a tarde quando está na sombra. Utilizando o programa estatístico encontra-se o mesmo valor e p =0,0043.

Figura 4.1.5: Cálculo do teste de Mann-Whitney no período do dia que o gado pasta na sombra



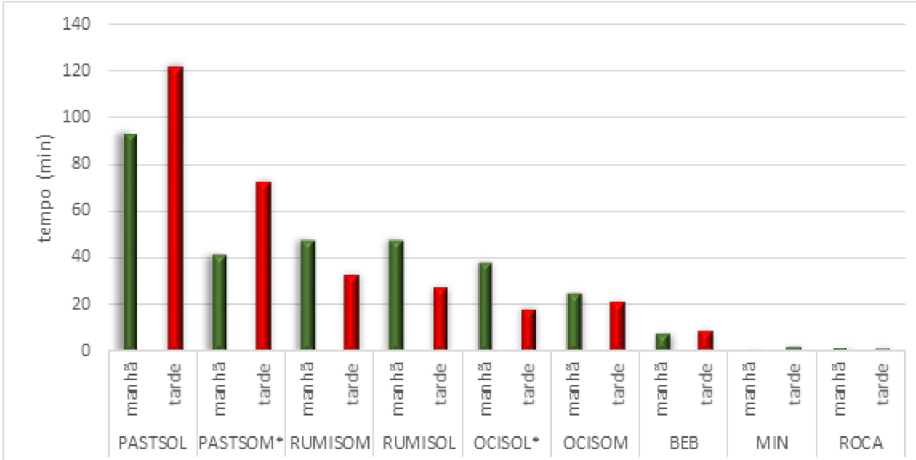
	A	B	C	D	E	F	G	H	I	J	K	L
1	TESTE DE MANN-WHITNEY PARA O TEMPO GASTO PELO GADO PASTANDO NA SOMBRA DE MANHÃ E A TARDE											
2	DADOS	POSTOS	DADOS	POSTOS		MANHÃ		TARDE				
3	0	1.5	50	25.5		DADOS	POSTOS	DADOS	POSTOS			
4	0	1.5	58	27		0	1.5	95	39.5			
5	10	3.5	64	28.5		10	3.5	30	13.5		SOMA	(n*(n+1))/2
6	10	3.5	64	28.5		19	7.5	50	25.5		1176	1176
7	14	5	65	31		29	11	14	5			
8	15	6	65	31		39	21	64	28.5		U	149.5
9	19	7.5	65	31		0	1.5	125	47		z	-2.8558
10	19	7.5	68	33		19	7.5	120	45.5			
11	20	9.5	70	34		34	17.5	95	39.5			
12	20	9.5	82	35		39	21	145	48			
13	29	11	85	36		58	27	115	44			
14	30	13.5	90	37		68	33	85	36			
15	30	13.5	91	38		34	17.5	70	34			
16	30	13.5	95	39.5		39	21	45	23.5			
17	30	13.5	95	39.5		32	16	30	13.5			
18	32	16	102	41		20	9.5	30	13.5			
19	34	17.5	105	42		91	38	45	23.5			
20	34	17.5	107	43		82	35	65	31			
21	38	19	115	44		38	19	90	37			
22	39	21	120	45.5		30	13.5	105	42			
23	39	21	120	45.5		20	9.5	10	3.5			
24	39	21	125	47		102	41	65	31			
25	45	23.5	145	48		64	28.5	65	31			
26	45	23.5				15	6	50	25.5			
27	50	25.5				107	43	120	45.5			
28						SOMA	449.5	SOMA	726.5			

Comparando o tempo que o gado passa ocioso no sol, o tempo que ele gasta de manhã é maior do que a tarde (p=0,0126). Apesar do bovino passar mais tempo pastando no sol a tarde que de manhã essa diferença não foi significativa (p=0,1078), o que também aconteceu com as outras variáveis (Figura 4.1.6 e Tabela 4.1.2).

Tabela 4.1.2: Comparação do comportamento ingestivo de bovino de corte em relação ao tempo gasto de manhã e à tarde. (*p<0,05 – diferenças significativas)

COMPORTAMENTO	PERÍODO	TEMPO (MIN)	VALOR p
PASTSOL	MANHÃ	92.6	0,1078
	TARDE	121.3	
PASTSOM*	MANHÃ	41.2	0,0043*
	TARDE	72.0	
RUMISOM	MANHÃ	47.5	0,1609
	TARDE	32.1	
RUMISOL	MANHÃ	47.3	0,0762
	TARDE	26.8	
OCISOL*	MANHÃ	37.8	0,0126*
	TARDE	17.3	
OCISOM	MANHÃ	24.6	0,4455
	TARDE	20.5	
BEB	MANHÃ	7.6	0,8447
	TARDE	8.1	
MIN	MANHÃ	0.2	0,4394
	TARDE	1.5	
ROCA	MANHÃ	1.2	0,6062
	TARDE	0.4	

Figura 4.1.6: Comparação do comportamento ingestivo de bovino de corte em relação ao tempo gasto de manhã e à tarde. (*p<0,05 – diferenças significativas)

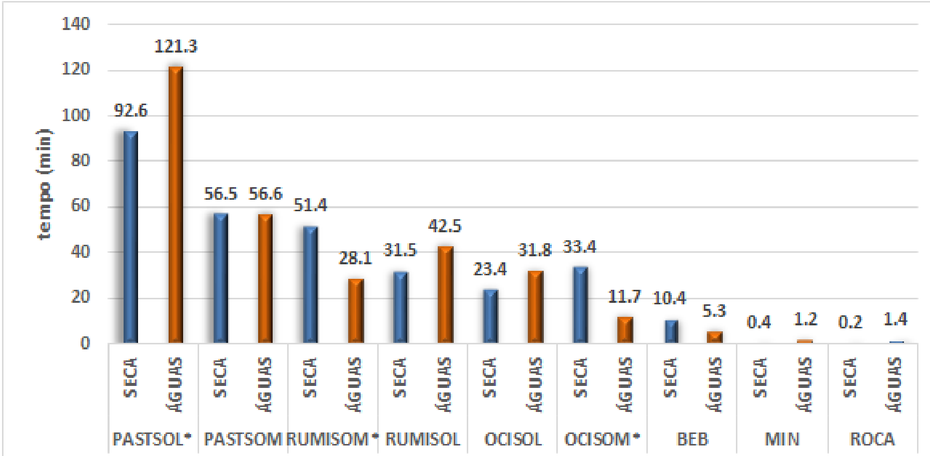


Com relação às estações o gado passa mais tempo pastando no sol nas águas (p=0,0328), ruminando na sombra na seca (p=0,0103) e ocioso na sombra na seca (p=0,0122). As outras variáveis não mostraram diferenças significativas entre as estações (Figura 4.1.7 e Tabela 4.1.3).

Tabela 4.1.3: Comparação do comportamento ingestivo de bovino de corte em relação à estação climática. (*p<0,05 – diferenças significativas)

COMPORTAMENTO	PERÍODO	TEMPO (MIN)	VALOR p
PASTSOL	MANHÃ	92.6	0,1078
	TARDE	121.3	
PASTSOM*	MANHÃ	41.2	0,0043*
	TARDE	72.0	
RUMISOM	MANHÃ	47.5	0,1609
	TARDE	32.1	
RUMISOL	MANHÃ	47.3	0,0762
	TARDE	26.8	
OCISOL*	MANHÃ	37.8	0,0126*
	TARDE	17.3	
OCISOM	MANHÃ	24.6	0,4455
	TARDE	20.5	
BEB	MANHÃ	7.6	0,8447
	TARDE	8.1	
MIN	MANHÃ	0.2	0,4394
	TARDE	1.5	
ROCA	MANHÃ	1.2	0,6062
	TARDE	0.4	

Figura 4.1.7: Comparação do comportamento ingestivo de bovino de corte em relação a estação. (*p<0,05 – diferenças significativas)



A comparação da altura do pastejo não mostrou diferença significativa no comportamento ingestivo de bovino de corte (p>0,05). (Tabela 4.1.4).

Tabela 4.1.4: Comparação do comportamento ingestivo de bovino de corte com a altura do pastejo

COMPORTAMENTO	ESTAÇÃO	TEMPO (MIN)	VALOR P
PASTSOL	15 cm	110,5	0,7105
	30 cm	103,4	
PASTSOM	15 cm	59,8	0,4642
	30 cm	53,3	
RUMISOM	15 cm	36,1	0,5227
	30 cm	43,4	
RUMISOL	15 cm	33,0	0,9753
	30 cm	41,1	
OCISOL	15 cm	28,3	0,9343
	30 cm	26,9	
OCISOM	15 cm	21,9	0,9015
	30 cm	23,2	
BEB	15 cm	9,1	0,5777
	30 cm	6,6	
MIN	15 cm	0,2	0,4394
	30 cm	1,5	
ROCA	15 cm	1,0	0,9753
	30 cm	0,6	

Há diferenças significativas no tempo gasto pelo gado pastando ao sol (p<0,0001), sendo que no sistema iLP o gado passa mais tempo pastando que no sistema iLPF14 e que no iLPF22 (p<0,05). Apesar do tempo pastando no sol ser maior no iLPF22 que no iLPF14, essa diferença não é significativa (p=0,1555). (Figura 4.1.8 e Tabela 4.1.5)

Os três sistemas de integração foram comparados pelos testes de Kruskal-Wallis e o teste *post hoc* SNK. Há diferenças significativas no tempo gasto pelo gado pastando na sombra (p=0,0077), sendo que no sistema iLPF22 o gado passa mais tempo pastando que no sistema iLP (p<0,05). Apesar do tempo pastando na sombra ser maior no iLPF22 que no iLPF14, essa diferença não é significativa (p>0,05), o que também aconteceu com iLPF14 e iLP (p>0,05). (Figura 4.1.8 e Tabela 4.1.5). Há diferenças significativas no tempo gasto pelo gado ocioso na sombra (p=0,0163), sendo que no sistema iLP o gado passa menos tempo que no sistema iLPF14 e iLPF22 (p<0,05). Não há diferença do tempo ocioso na sombra entre os sistemas iLPF14 e iLPF22 (p>0,05). As outras variáveis não mostraram diferenças significativas na comparação pelo sistema de integração (p>0,05). (Figura 4.1.8 e Tabela 4.1.5)

Figura 4.1.8: Comportamento ingestivo de bovino de corte em relação ao sistema de integração. (*p<0,05 – diferenças significativas)

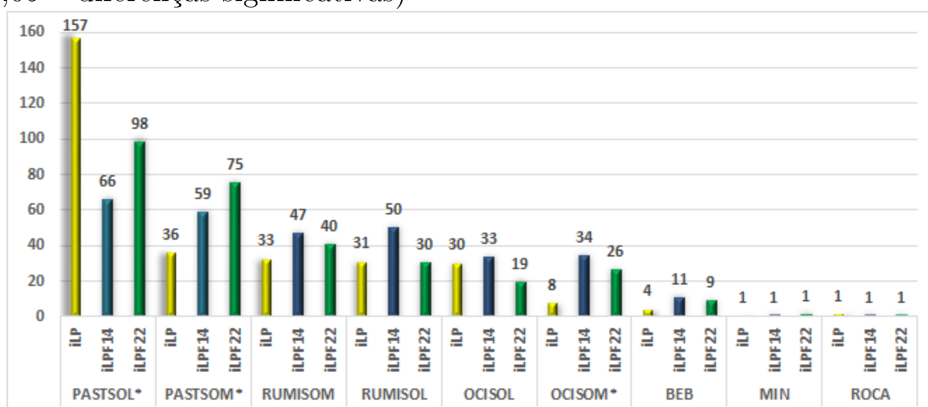


Tabela 4.1.5: Comparação do comportamento ingestivo de bovino de corte com os sistemas de integração. (*p<0,05 – diferenças significativas)

COMPORTAMENTO	ESTAÇÃO	TEMPO (MIN)	VALOR p
PASTSOL*	iLP	156.7	<0,0001*
	iLPF14	65.6	
	iLPF22	98.5	
PASTSOM*	iLP	36.2	0,0077*
	iLPF14	58.5	
	iLPF22	75.0	
RUMISOM	iLP	32.6	0,4661
	iLPF14	46.5	
	iLPF22	40.2	
RUMISOL	iLP	31.1	0,5357
	iLPF14	49.8	
	iLPF22	30.3	
OCISOL	iLP	30.0	0,4794
	iLPF14	33.3	
	iLPF22	19.5	
OCISOM*	iLP	7.5	0,0163*
	iLPF14	34.2	
	iLPF22	26.0	
BEB	iLP	4.1	0,1269
	iLPF14	10.9	
	iLPF22	8.6	
MIN	iLP	0.6	0,9457
	iLPF14	0.6	
	iLPF22	1.3	
ROCA	iLP	1.2	0,9990
	iLPF14	0.6	
	iLPF22	0.6	

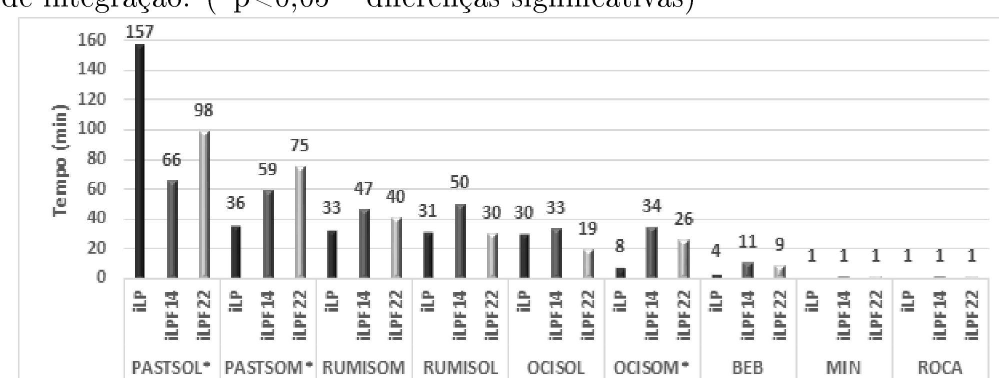
Como nos sistemas de integração e nas estações existiram diferenças significativas, utilizou-se o teste de Mann-Whitney para comparação do tempo gasto pelo gado em seu comportamento ingestivo diurno.

Pode-se afirmar que há diferenças significativas ($p<0{,}05$) nas variáveis: PASTSOL - durante o pastejo no sol o tempo gasto pastando no sol na estação das águas é maior do que na seca nos sistemas iLPF14 e iLPF22; RUMISOM – o tempo ruminando na sombra durante a seca é maior que na estação das águas no sistema iLP; OCIOSOL – o tempo ocioso do bovino no sol é maior na estação das águas que na seca no sistema iLP; OCIOSOM – o tempo ocioso do bovino na sombra é maior na seca do que na estação das águas nos sitemas iLPF14 e iLPF22; e BEB – o tempo bebendo água é maior na seca do que nas águas no sistema iLPF22. (Figura 4.1.9 e Tabela 4.1.6).

Tabela 4.1.6: Comparação do comportamento ingestivo de bovino de corte nas estações climáticas de acordo com os sistemas de integração. (*p<0,05 – diferenças significativas)

	Estação	iLP	iLPF14	iLPF22
PASTSOL *	SECA	155,3	46,6	76,1
	ÁGUAS	158,2	84,7	120,9
	valor p	0,7527	0,0274*	0,0357*
PASTSOM	SECA	30,8	62.0	76,7
	ÁGUAS	41,5	55.0	73,4
	valor p	0,1415	0,7527	0,8336
RUMISOM	SECA	53,2	58,6	42,5
	ÁGUAS	12,0	34,5	38,0
	valor p	0,0209*	0,1722	0,6365
RUMISOL	SECA	32,0	31,0	31,6
	ÁGUAS	30,1	68,5	28,9
	valor p	0,9581	0,1278	0,8336
OCISOL*	SECA	18,9	34,8	16,5
	ÁGUAS	41,1	31,7	22,5
	valor p	0,0274*	0,8336	0,3446
OCISOM*	SECA	4,3	53,5	42,6
	ÁGUAS	10,7	14,9	9,4
	valor p	0,2480	0,0074*	0,0157*
BEB*	SECA	5,5	13,5	12,3
	ÁGUAS	2,6	8,3	5,0
	valor p	0,2271	0,8748	0,0357*

Figura 4.1.9: Comportamento ingestivo de bovino de corte nas estações de acordo com o sistema de integração. (*p<0,05 – diferenças significativas)



4.2 Levantamento dos Índices de Temperatura e Umidade e Comparação das Espécies Arbóreas

Essa pesquisa foi realizada pelo Me. Nivaldo Karvatte Junior (Universidade Estadual do Oeste do Paraná) orientado pela pesquisadora da Embrapa Drª Fabiana Villa Alves.

Foram coletados dados de temperatura do ar (Ta), temperatura de bulbo úmido (Tbu), temperatura de globo negro (Tgn), umidade relativa do ar (UR) e velocidade do vento (Vv), ao sol e a sombra, das 08h00 às 17h00, com os tipos de árvores escolhidos. Posteriormente, foram calculados os índices de conforto térmico, índice de temperatura e umidade (ITU), índice de temperatura de globo e umidade (ITGU) e a carga térmica de radiação (CTR) para se avaliar qual espécie arbórea tem o melhor conforto térmico, entre o Cambará, o Cumbaru e o Eucalipto (Figura 4.2.1).

Figura 4.2.1: Cambará, cumbaru e eucalipto, respectivamente

Verificando-se pelo teste de Shapiro-Wilk que os dados não apresentam uma distribuição normal. Logo, deve-se usar um teste Não-Paramétrico. Para comparar as variáveis: Vv, Ta, Tbu, Tgn e UR, utilizou-se o teste de Friedman que serve para testar a hipótese de que vários grupos relacionados têm, todos, a mesma distribuição. Esses grupos têm os valores pareados que são os mesmos horários de coleta em cada um. E para a apresentação gráfica foi utilizado o Box-Plot.

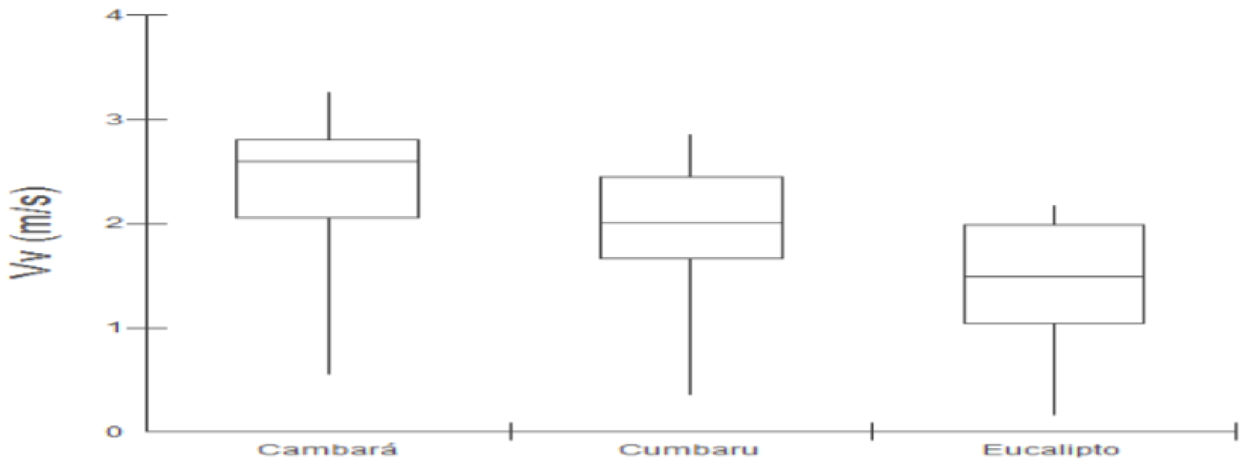
O teste de Friedman é calculado por

$$X_r^2 = \frac{12}{Nk(k+1)} \sum R^2 - 3N(k+1)$$

sendo N é o número de linhas; k o número de grupos; $\sum R^2 = \left(\sum R_1\right)^2 + \left(\sum R_2\right)^2 + ... + \left(\sum R_k\right)^2$ a soma dos quadrados dos postos dos grupos 1, 2, ..., k, respectivamente.

Como resultados das comparações a velocidade do vento foi maior no cambará (2,3 m/s) do que no eucalipto (1,4 m/s), não havendo diferenças para o cumbaru (1,9 m/s) com p=0,0007. (Figura 4.2.2).

Figura 4.2.2: Velocidade do vento nas espécies arbóreas

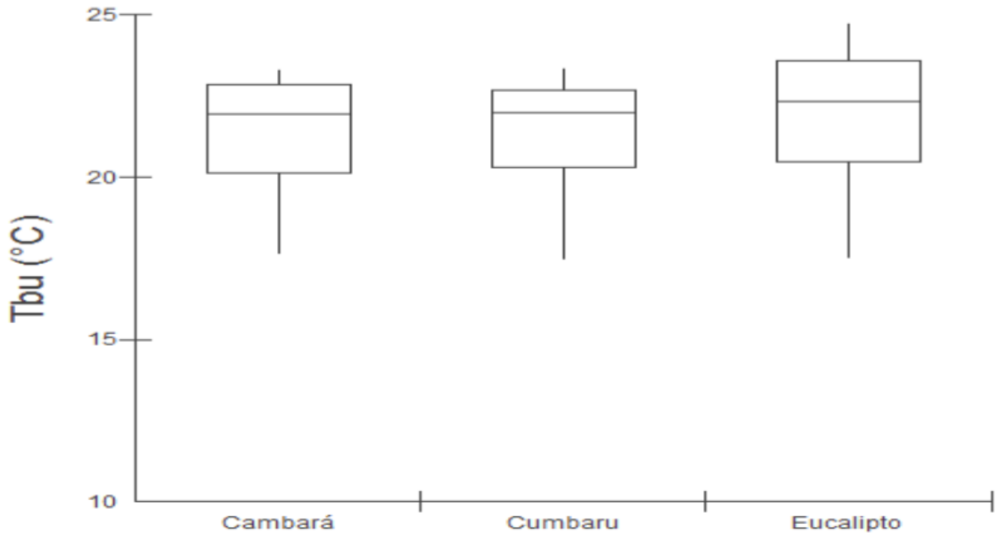


Apesar da temperatura ambiente ter sido maior no eucalipto (30,5°C), essa diferença não foi significativa para o cambará (30,0 °C) e para o cumbaru (29,8° C), com p=0,0821. (Figura 4.2.3).

Figura 4.2.3: Temperatura ambiente das espécies arbóreas

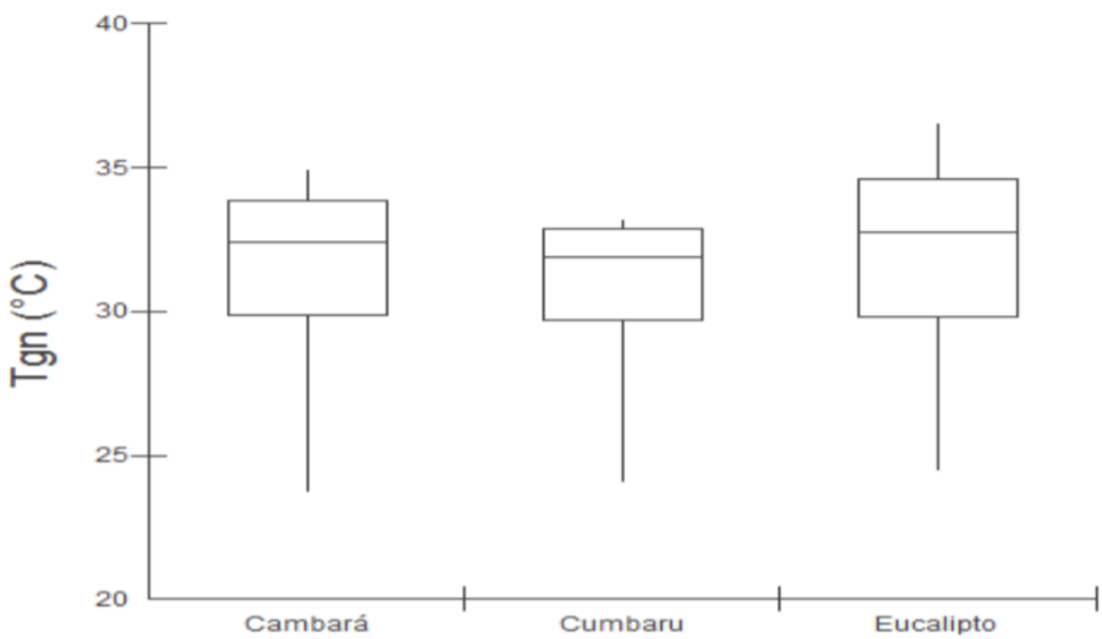
Na temperatura do bulbo úmido pode-se afirmar que no eucalipto a Tbu (22,0° C) é maior que no cambará (21,4° C) e que no cumbaru (21,4° C), com p=0,0074. (Figura 4.2.4).

Figura 4.2.4: Temperatura do bulbo úmido nas espécies arbóreas



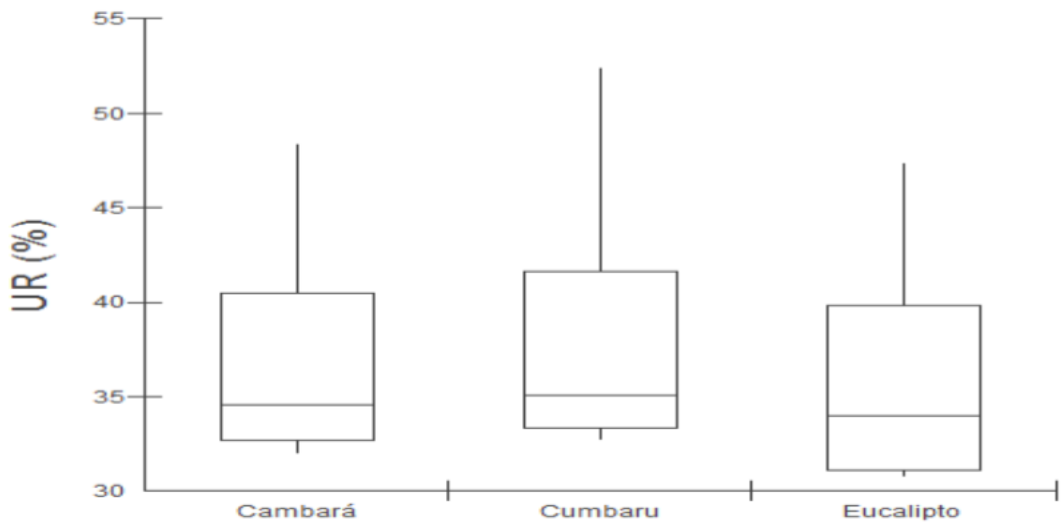
Apesar da temperatura do globo negro ser maior no eucalipto (24,4° C) do que no cumbaru (24,1° C) e do que no cambará (23,7° C), essas diferenças não são significativas com p=0,0821. (Figura 4.2.5).

Figura 4.2.5: Temperatura do globo negro nas espécies arbóreas



A umidade relativa foi maior no cumbaru (38,0%) do que no cambará (36,8%) e do que no eucalipto (36,0%), com p=0,0001. (Figura 4.2.6).

Figura 4.2.6: Umidade relativa do ar nas espécies arbóreas



Os tipos de espécies arbóreas para comparação do ITU, ITGU e CTR continuam sendo o cambará, o cumbaru e o eucalipto, mas para essas comparações tem-se o eucalipto com espaçamento de 14 metros (eucalipto i) e o de 22 metros (eucalipto ii). A classificação do índice de temperatura e umidade segundo Hahn e Mader (1997), é dada pelos valores de ITU $\leq$ 70 que são indicadores de um ambiente não estressante; entre 71 e 78 são críticos; de 79 a 83 a situação é de perigo; e acima de 83, situação de emergência. A classificação do índice de temperatura do globo e umidade segundo National Weather Service (1976), são delimitados por valores até 74, situação de conforto; entre 75 e 78, situação de alerta; 79 a 84, perigo; e acima deste, situação crítica.

O ITU é obtido pela seguinte equação

$$\text{ITU} \ = t_{bs} + 0,36 t_{po} + 41,5$$

onde t_{bs} é a temperatura do bulbo seco (°C) e t_{bo} é a temperatura do ponto de orvalho (°C).

O ITGU é obtido pela seguinte equação

$$\text{ITGU} \ = t_{gn} + 0,36 t_{po} + 41,5$$

sendo t_{gn} a temperatura do globo negro (°C) e t_{po} a temperatura do ponto de orvalho (°C).

A CTR é calculada pela equação de Stefan-Boltzmann, dada por

$$CTR = \sigma (TRM)^4$$

onde a CTR é a carga térmica de radiação, em $W.m^{-2}$, σ é a constante de Stefan-Boltzmann é igual a $5,67 \times 10^{-8}$ $W.m^{-2}.K^{-4}$ e TRM é a Temperatura Radiante Média (K). A TRM pode ser obtida pela equação

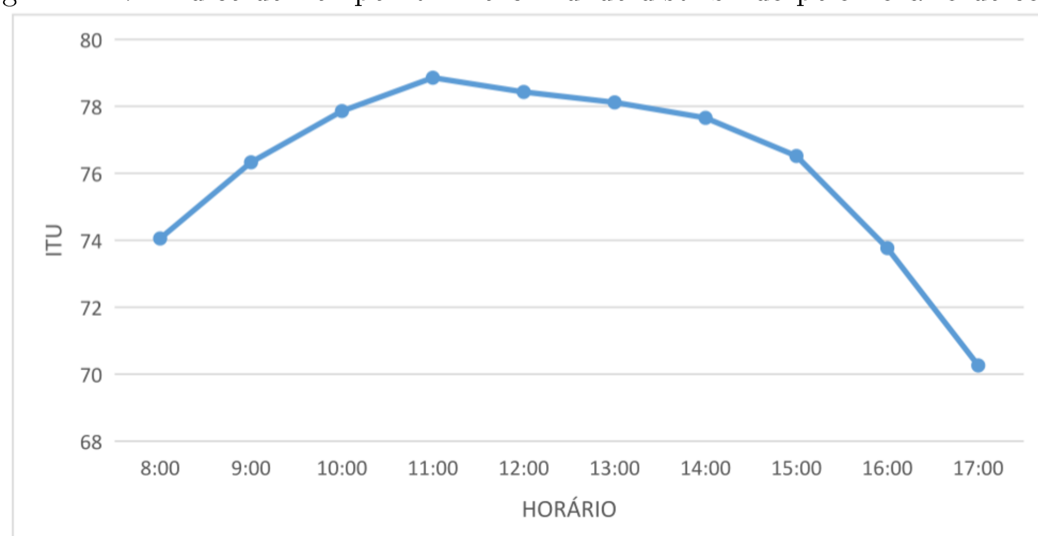
$$TRM = 100\sqrt{2,51 * \sqrt{V_v} * (t_{gn} - t_{bs}) + \left(\frac{t_{gn}}{100}\right)^4}$$

sendo V_v a velocidade do vento $(m.s^{-1})$, t_{bs} a temperatura do bulbo seco (k) e t_{gn} a temperatura do globo negro (K).

Após o teste de Shapiro-Wilk pode-se afirmar que a distribuição dos dados do ITU, do ITGU e da CTR é normal. Portanto, será utilizado o teste ANOVA, e se necessário o teste *post hoc* de Tukey.

Observou-se diferença significativa (p<0,0001) no índice de temperatura e umidade entre os horários observados, com menor índice as 17:00 horas (64,9) e maior as 11:00 horas (84,6), sendo a maior média as 11:00 horas (78,9) e a menor as 17:00 horas (70,3). (Figura 4.2.7).

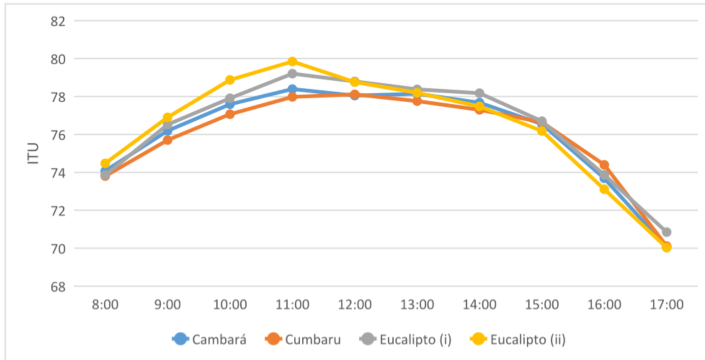
Figura 4.2.7: Índice de Temperatura e Umidade distribuído pelo horário de coleta



Considerando-se os limites do ITU as 8:00 horas os animais já se encontram em uma faixa-crítica, enquanto que, no horário mais quente (11:00 horas) os animais se encontram em situação de perigo, voltando a faixa-crítica das 12:00 as 16:00 horas. Eles ficaram em um ambiente não estressante (zona de conforto térmico) apenas as 17:00 horas.

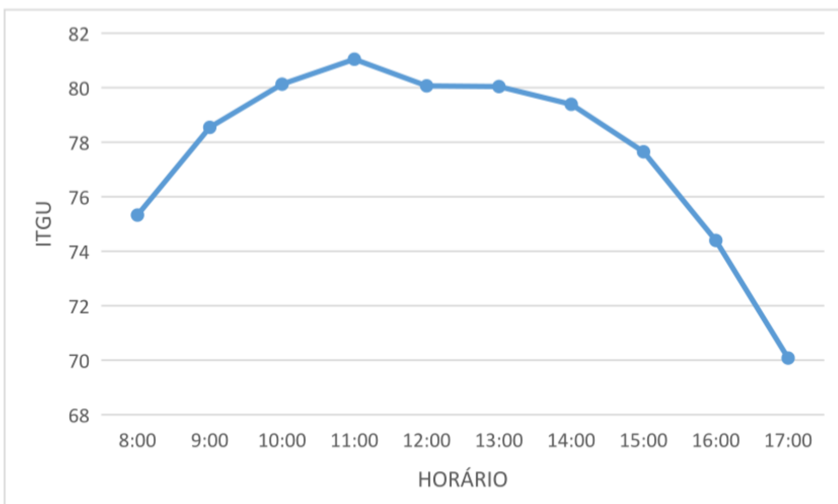
As espécies arbóreas apresentam altos valores médios de ITU durante o período estudado, tendo 76,4 para o eucalipto i e ii, 76,0 cambará e 75,9 cumbaru, mas não há diferenças significativas entre si (p=0,7739). (Figura 4.2.8).

Figura 4.2.8: Índice de Temperatura e Umidade nas espécies arbóreas



Observou-se diferença significativa (p<0,0001) no índice de temperatura de globo e umidade entre os horários observados, com menor índice as 17:00 horas (64,9) e maior as 10:00 horas (90,9), sendo a maior média as 11:00 horas (81) e a menor as 17:00 horas (70,1). (Figura 4.2.9).

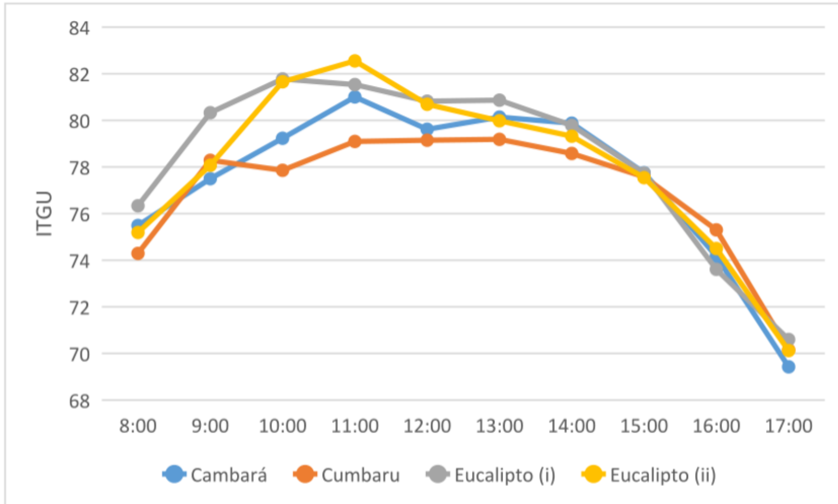
Figura 4.2.9: Índice de Temperatura do Globo e Umidade distribuído pelo horário de coleta



Considerando-se os limites do ITGU as 8:00 horas os animais já se encontram em situação de alerta, enquanto que no horário das 9:00 horas as 14:00 horas a situação é de perigo, voltando a situação de alerta das 15:00 as 16:00 horas. Eles estavam em uma situação de conforto térmico apenas as 17:00 horas.

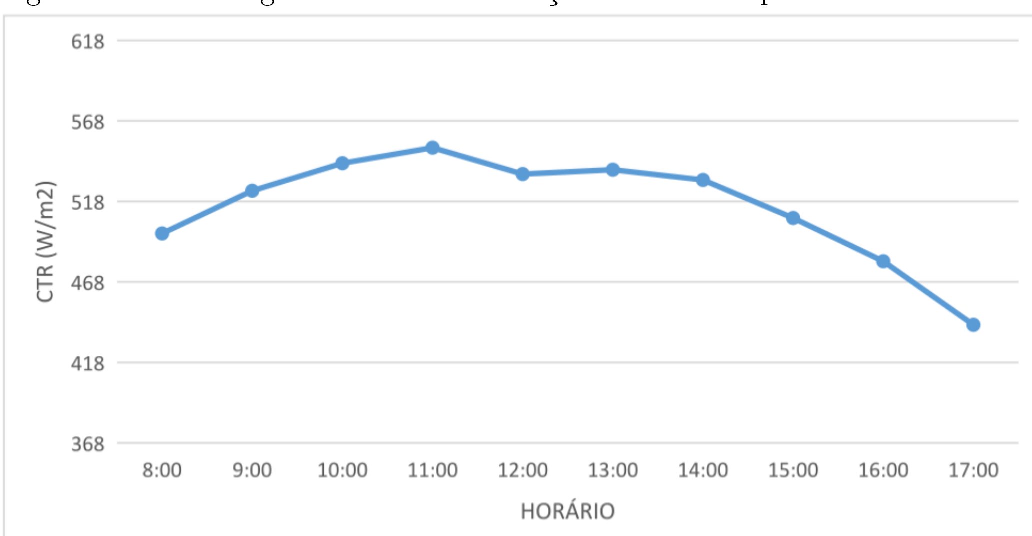
As espécies arbóreas apresentam altos valores de ITGU durante o período estudado, tendo 78,3 para o eucalipto i, 77,9 no eucalipto ii, 77,4 no cambará e 76,9 cumbaru, mas não há diferenças significativas entre si (p=0,3072). (Figura 4.2.10).

Figura 4.2.10: Índice de Temperatura do Globo e Umidade nas espécies arbóreas



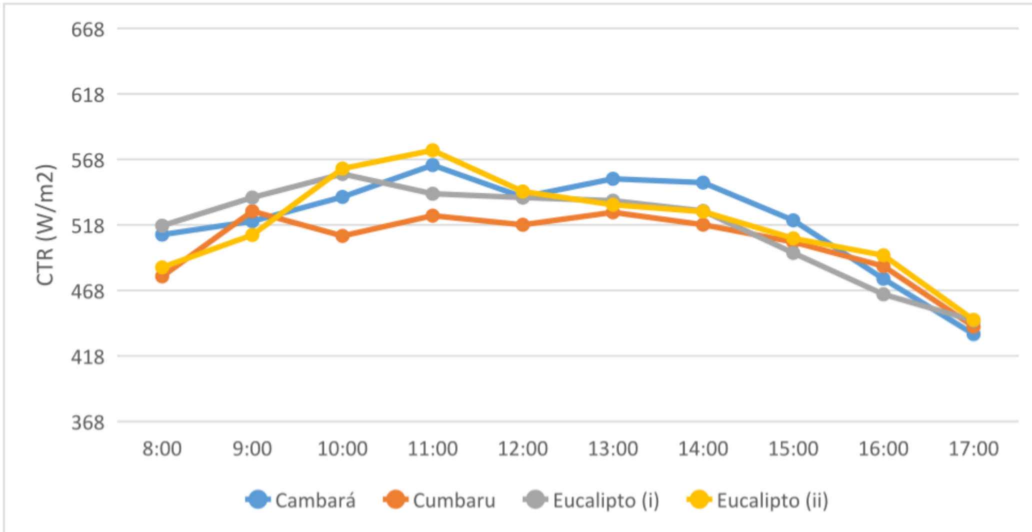
Quanto a carga térmica de radiação (CTR) o menor valor encontrado foi as 17:00 horas (404,3 W/m2) e o maior as 11:00 horas (720,6 W/m2). Houve diferença significativa (p<0,0001) nos horários estudados, sendo a menor média de todas as 17:00 horas (441,4 W/m2), seguida das 16:00 horas (480,8 W/m2) e das 8:00 horas (498,1 W/m2) que não diferem entre si, mas são menores que os outros valores estudados. (Figura 4.2.11).

Figura 4.2.11: Carga Térmica de Radiação distribuído pelo horário de coleta



As espécies apresentam valores de CTR sem diferenças significativas (p=0,2555) durante o período estudado, tendo 521,0 W/m2 para o cambará, 518,5 W/m2 no eucalipto ii, 516,6 W/m2 no eucalipto i e 503,8 W/m2 no cumbaru. (Figura 4.2.12).

Figura 4.2.12: Carga Térmica de Radiação nas espécies arbóreas



4.3 Considerações Finais

No sistema de integração iLP onde as árvores estão mais dispersas o gado sofre mais com o calor pastando no sol o mesmo tempo na estação da seca como nas águas. O gado passa mais tempo ruminando na sombra no tempo da seca, e na estação das águas ele fica ocioso muito tempo no sol, praticamente não havendo diferenças nos sistemas de integração iLPF14 e iLPF22.

As temperaturas do microclima estão acima do ideal para a produtividade e sobrevivência do bovino, a umidade relativa está abaixo da faixa ideal e a velocidade do vento não demonstra preocupação.

Apesar das espécies arbóreas Cambará, Cumbaru e Eucalipto apresentarem elevados índices de ITU, ITGU e CTR, ainda assim proporcionam condições mais favoráveis de microclima e conforto térmico, viabilizando as trocas de calor do bovino com o ambiente.

A exemplificação do uso da estatística para interpretar os dados das pesquisas apresentadas ilustra a viabilidade da aplicação da planilha eletrônica (EXCEL) como ferramenta útil e necessária.

Capítulo 5

Conclusão Final

Todos os dias nos meios de comunicação nos deparamos com informações que são dadas em cima de pesquisas e levantamento de dados sobre diversos setores da sociedade. São informações que influenciam o comportamento do ser humano na sociedade e a tomada de decisão frente a situações importantes. Essas informações influenciam os setores de educação, saúde, econômico, administrativo, financeiro, tributário e tantos outros, tamanha a influência das análises estatísticas de dados levantados. Tendo, pois, uma importância a ser considerada na vida do aluno dentro e fora da escola, fica evidente a necessidade de se procurar novas ideias para o ensino de estatística em sala de aula, procurando tornar o processo de ensino aprendizagem mais atrativo ao aluno.

Por isso, tendo visto a falta de um material de apoio ao professor com o uso da planilha eletrônica EXCEL, utilizei de experiências bem sucedidas em minhas aulas e análises de notas de outros conteúdos na escola, para propor ao professor sugestões para o ensino da estatística na sala de tecnologia usando a planilha eletrônica. É claro que são alguns exemplos, mas que tem por intuito instigar o professor, que é o principal agente executor nessa busca, a procurar novas situações didáticas que o auxiliem no desenvolvimento dos conteúdos de estatística em sala de aula. Inclusive utilizando da interdisciplinaridade, onde o aluno começa a ver aplicações dos conteúdos de matemática em outros conteúdos escolares.

Para munir o professor quanto aos conteúdos básicos da estatística se fez necessário a apresentação deles, mas com a certeza de que ele deve estar buscando se aperfeiçoar obtendo mais e mais conhecimento dos conteúdos que ministra em sala de aula. Logo, esse trabalho assessora o professor na exposição dos conteúdos estatísticos, mas o não leva a dominar toda estatística descritiva e a inferência.

A aplicação de conhecimentos mais avançados de estatística foi feita de acordo com meu conhecimento técnico oriundo de vários anos fazendo análises estatísticas de monografias até teses de doutorado. E demonstrei a importância desse conhecimento aplicando-o em um experimento da Embrapa. Com os dados já coletados tive a oportunidade de ser incluído em dois projetos onde a estatística se fez extremamente necessária para se concluir algo em relação ao que se estava estudando. Com a aplicação de testes estatísticos consegui classificar os dados quanto a sua normalidade ou não, optando pelos testes não-paramétricos de acordo com o resultado calculado. E a partir disso, cada comparação foi feita por um estudo minucioso das hipóteses levantadas, provando ou não se havia diferenças significativas entre as variáveis, podendo escolher entre a rejeição ou não da hipótese nula em cada caso. Foi utilizado um programa estatístico o BioEstat versão 5.3. Apesar disso, para enriquecimento desse trabalho fiz os cálculos dos testes utilizando o MICROSOFT® EXCEL e demonstrando passo a passo como deve-se aplicar cada um dos testes utilizados, não em todas as comparações, mas citando um exemplo de cada teste. Utilizando um nível de significância de 0,05 verifiquei, principalmente, que nas comparações quanto ao comportamento ingestivo de bovino de corte no sistema de integração iLP o gado sofre mais com o calor pastando no sol, o que faz com que o gado sofra mais com a temperatura expondo-o ao estresse que o leva a produzir menos leite e carne. O que leva à conclusão que o sistema de integração iLP é o menos indicado em nossa região para criação de gado de corte. Também pude afirmar que as temperaturas em nossa região estão acima do ideal, já que os índices de temperatura, umidade e carga térmica de radiação mostram valores fora do ideal, fazendo com que o gado diminua a produtividade. Em uma última análise verifiquei que a poucas diferenças de temperatura, umidade e carga térmica de radiação entre as espécies arbóreas Cambará, Cumbaru e o Eucalipto, que possuem valores altos para esses índices, mas que apesar desses índices as espécies arbóreas se fazem necessárias, porque proporcionam condições mais favoráveis de conforto térmico, ajudando o gado na troca de temperatura com o ambiente em que vive.

Crédito das imagens

http://www.nelorems.org/upload/noticia_imagem/1323369219.jpg
http://redeglobo.globo.com/sp/eptv/terra-da-gente/platb/flora/cambaragochnatia-polymor pha/
http://edpquatro.blogspot.com.br/2011-09-01-archive.html
www.canaldoprodutor.com.br/comunicacao/noticias/cresce-demanda-por-capacitacao-sobre-cultivo-de-eucaliptos-em-ms

Referências Bibliográficas

[1] BANZATTO, D.A. e KRONKA, S.N. Experimentação Agrícola. Jaboticabal, FUNEP, 1992.

[2] BARROS NETO, J.C. e Outros. Planejamento e Otimização de Experimentos. Campinas, UNICAMP, 1995.

[3] BOYER, Carl B. História da Matemática.São Paulo: Editora Edgard Blucher Ltda, 1996.

[4] CAMPOS, H. DE. Amostragem I. Notas de Aulas, ESALQ/USP: Piracicaba, SP, 1980.

[5] CAMPOS, H. DE. Amostragem II. Notas de Aulas, ESALQ/USP: Piracicaba, SP, 1980.

[6] COCHRAN, W. G. Técnicas de Amostragem. Editora Fundo de Cultura, 1.ed., 1965.

[7] FELLER,W. Introdução à Teoria das Probabilidades e Suas Aplicações. parte 1. espaços amostrais discretos. São Paulo,Edgard Blücher,1976.236p.

[8] GAMERMAN, D. & LOPES, H. F. (2006) Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference, 2nd edn. Chapman & Hall.

[9] JAMES, B.R. Probabilidade: um curso em nível intermediário. Rio de Janeiro, IMPA, 1981, 304p.

[10] LAPPONI, J.C. Estatística Usando o EXCEL. Versões 4 e 5. São Paulo, Lapponi Trein. e Ed. Ltda, 1995.

[11] LEVINE, DAVID M. Estatística: Teoria e Aplicações usando Microsoft EXCEL em português. LTC, 1.ed, 1998.

[12] MEYER, P.L. Probabilidade: aplicações à estatística. Rio de Janeiro, LTC, 2.ed. 1983, 426p.

[13] MURTEIRA, B.J.F. Análise Exploratória de Dados. Estatística Descritiva. Mc Graw-Hill, Lisboa,1993.

[14] QUÉTELET, ADOLPHE. Sur l'homme et le developpement de ses facultés, essai d'une physique sociale.[S.l.: s.n.].

[15] SANTOS, J.P.O.; MELO, M.P. e MURARI, I.T.C. Introdução à Análise Combinatória. Campinas, Editora da UNICAMP, 1995. 265p.

[16] STEWART IAN. 17 Equações que Mudaram o Mundo. Rio de Janeiro: Zahar, 2013.

[17] TONHASCA Jr., A. The three “capital sins” of statistics used in biology. Ciência e Cultura, v.46, n.6, p.417-422, 1991.

[18] TRIOLA, MARIO F. Introdução à Estatística. LTC, 9.ed.2005.

[19] TUKEY, J.W. Exploratory Data Analysis. Addison-Wesley, Reading,1970.

[20] VIEIRA, S. Bioestatística: Tópicos Avançados. Ed. Elsevier. 3.ed. 2011.

[21] VIEIRA, S. Introdução a Bioestatística. Ed. Elsevier. 1.ed. 2004.

[22] WEIBULL, W. A Statistical Function of the Applicability. Journal of Applied Mechanics, p.293-297, December, 1951.

[23] http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues&palavra=estat%EDstica, acessado em 15/03/2014.

[24] http://pt.wikipedia.org/wiki/Adrien-Marie_Legendre, acessado em 20/08/2014.

[25] http://www.hiperatividade.com.br/article.php?sid=90, acessado em 25/11/2014.

[26] http://www.mamiraua.org.br/pt-br/downloads/programas/bioestat-versao-53/, acessado em 13/12/2014.