

TECNÓLOGO EM GESTÃO AMBIENTAL

ESTATÍSTICA

JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL

JOÃO MARIA FILGUEIRA

ESTATÍSTICA

**Natal
2012**



GESTÃO AMBIENTAL



GOVERNO DO BRASIL

Presidente da República
DILMA VANA ROUSSEFF

Ministro da Educação
JOSÉ HENRIQUE PAIM FERNANDES

Diretor de Educação a Distância da CAPES
JOÃO CARLOS TEATINI

Reitor do IFRN
BELCHIOR DE OLIVEIRA ROCHA

Pró-Reitor de Pesquisa e Inovação
JOSÉ YVAN PEREIRA LEITE

Coordenador da Editora do IFRN
PAULO PEREIRA DA SILVA

Diretor do *Campus* EaD/IFRN
ERIVALDO CABRAL

Diretora Acadêmica do *Campus* EaD/IFRN
ANA LÚCIA SARMENTO HENRIQUE

Coordenadora Geral da UAB /IFRN
ILANE FERREIRA CAVALCANTE

Coordenadora Adjunta da UAB/IFRN
MARLI TACCONI

Coordenadora do Curso
de Tecnologia em Gestão Ambiental
MARIA DO SOCORRO DIÓGENES PAIVA

ESTATÍSTICA
Material Didático

Professor Pesquisador/Conteudista
JOÃO MARIA FILGUEIRA

Diretora da Produção de Material Didático
ROSEMARY PESSOA BORGES

Coordenador da Produção de
Material Didático
LEONARDO DOS SANTOS FEITOZA

Revisão Linguística
HILANETE PORPINO DE PAIVA

Projeto Gráfico
BRENO XAVIER

Diagramação
GEORGIO NASCIMENTO
LUANNA CANUTO DA ROCHA
MARÍLIA DA COSTA PAIVA
VICTOR HUGO ROCHA
EMERSON LUA BEZERRA DOS SANTOS
ALEF SOUZA DA SILVA
JOACI NASCIMENTO DE PAULA

Ficha Catalográfica

F475e Filgueira, João Maria.
Estatística / João Maria Filgueira – Natal : IFRN Editora, 2012.
230f. : il. color.

ISBN 978-85-8333-075-2

1. Estatística – EaD. 2. Medidas e Softwares. 3. Frequência. 4. Apresentação Gráfica. 5. Padrão de Estimativa. I. Instituto Federal de Educação, Ciência e Tecnologia do RN. II. Título.

RN/IFRN/EaD

CDU 311

O Material Didático

Caro(a) Aluno(a):

Você está recebendo este material didático por meio do qual vai realizar a maior parte de seus estudos do curso de Tecnologia em Gestão Ambiental. Na Educação a Distância, o material didático é a mais importante ferramenta de estudo. Ele é o principal mediador entre você e os conhecimentos historicamente acumulados que foram escolhidos para compor cada aula que agora está em suas mãos.

O material didático na EaD é, ainda, substituto do professor no momento em que você o utiliza. Nesses textos, o professor se faz presente através da linguagem dialogada, das estratégias de mobilização dos conteúdos, das atividades, enfim, de tudo o que compõe esse material. É importante que você tenha clareza de que a sua aprendizagem depende, sobretudo, do seu empenho em estudá-lo, dedicando bastante atenção aos conteúdos de cada aula. Realizar cada uma das atividades, comunicar-se com seu tutor e/ou professor através das várias formas de interação e sanar as dúvidas que, por ventura, venham surgir durante o processo de utilização desse material, constituem-se elementos primordiais para o seu aprendizado.

Esse material foi concebido, escrito e finalizado com muita dedicação com um objetivo principal: a sua aprendizagem. Cada imagem, ícone ou atividade passou por um refinado processo de análise com o objetivo de que, no final de cada sessão de estudo, você tenha compreendido bem os conceitos, categorias ou postulados essenciais à sua formação como tecnólogo em Gestão Ambiental. Desejamos que o itinerário iniciado por você seja exitoso e que, ao final do curso, esse material tenha contribuído efetivamente para seu crescimento na condição de indivíduo, cidadão e profissional.



Bons estudos.

Diretoria de Produção de Material Didático

As seções

Com o objetivo de facilitar a sua aprendizagem, as aulas foram estruturadas didaticamente em seções que facilitam o seu itinerário de estudos. Essas seções cumprem, cada uma, um objetivo específico e estão articuladas entre si, de modo que, ao final de cada aula, você tenha compreendido o conteúdo e apreendido os conceitos principais. Vamos ver quais são essas seções e quais as suas funções nas aulas.



Apresentando a aula

Apresentação do conteúdo da aula e chamada para a importância dos temas que serão tratados nela.



Definindo objetivos

Apresentação dos objetivos de aprendizagem da aula.

Desenvolvendo o conteúdo

Desenvolve a temática da aula através da apresentação dos conteúdos propriamente ditos.



Atividade

Apresenta as atividades de fixação/percurso relativas aos conhecimentos estudados no bloco de conteúdo.



Lembre-se!

Apresenta informações complementares importantes para o aluno como biografemas ou mesmo Glossário.



Resumindo

Resumo da aula que você estudou.



Leituras complementares

Indicação de leituras complementares.



Avaliando seus conhecimentos

Espaço em que o professor sugere algumas maneiras de você se autoavaliar em relação ao seu aprendizado.

Conhecendo as referências

Apresenta as referências bibliográficas que foram utilizadas pelo professor para a elaboração da aula.

Índice

| | |
|---|---------|
| Significado da estatística: abordagem contextual e aplicações | Aula 01 |
| Estatística descritiva: séries estatísticas | Aula 02 |
| Distribuição de frequência | Aula 03 |
| Medidas de tendência central | Aula 04 |
| Medidas de variabilidade | Aula 05 |
| Medidas separatrizes | Aula 06 |
| Apresentação gráfica | Aula 07 |
| Assimetria e curtose | Aula 08 |
| Significado da análise de correlação linear | Aula 09 |
| Coefficiente de correlação linear | Aula 10 |
| Regressão linear | Aula 11 |
| Erro padrão da estimativa | Aula 12 |
| Distribuição de frequência com <i>software</i> estatístico | Aula 13 |
| Medidas estatísticas com <i>software</i> estatístico | Aula 14 |
| Correlação e regressão com <i>software</i> estatístico | Aula 15 |

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 01
SIGNIFICADO DA ESTATÍSTICA: ABORDAGEM
CONTEXTUAL E APLICAÇÕES

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Você já deve ter lido sobre a importância da educação para o desenvolvimento de um país, mas talvez você não tenha imaginado que vários estudos estatísticos foram realizados antes para comprovar que investimentos em educação reduzem problemas sociais, aumentam a riqueza de uma nação, e assim teríamos vários benefícios advindos da educação para mostrar aqui. É isso mesmo, é um trabalho estatístico ao longo do tempo, coletando dados e analisando se esses aspectos têm relação entre si.

Nesta aula, você estudará sobre a importância da Estatística para o nosso dia-a-dia, também poderá perceber que ela está presente em várias situações e pode contribuir bastante para solucionar vários problemas, inclusive questões ligadas à Gestão ambiental. Bons Estudos!

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- Entender o significado da Estatística a partir de alguns elementos próprios das fases de um trabalho estatístico.
- Compreender como a Estatística é importante para estudar fenômenos da vida real, em particular de gestão ambiental;

DESENVOLVENDO O CONTEÚDO

ASPECTOS BÁSICOS DA ESTATÍSTICA



Fig. 01 – O símbolo da Estatística

Fonte: [http://www.brasilecola.com/upload/e/simbolo_da_estatistica\[1\].jpg](http://www.brasilecola.com/upload/e/simbolo_da_estatistica[1].jpg)

envolvidas.

Imagine, por exemplo, que você queira analisar se o Índice de Desenvolvimento Humano – IDH - de sua cidade interfere na ocorrência de casos de dengue. A estatística pode lhe dar uma resposta bastante razoável. Basta que você tenha uma coleção desses pares de dados e utilize técnicas que vamos estudar ao longo de nosso curso.

Mas é importante que, antes, você consiga compreender alguns aspectos mais básicos da Estatística. Em particular, você precisa saber que essa ciência tem várias aplicações, desde há muito tempo. Precisa também saber conceitos universais que são necessários para delimitar bem a abrangência da Estatística. Alguns cientistas foram muito importantes para a evolução dessa ciência. Nomes como Gauss, Bernoulli e Tchebchev, tiveram uma contribuição decisiva na consolidação do pensamento estatístico.

Em relação à utilização da estatística, a partir do século XVI começaram a surgir as primeiras análises sistemáticas de registros diversos como os de nascimento, óbitos, riquezas, casamentos. Esses registros eram utilizados

Você já deve ter ouvido falar em pesquisas de opinião sobre eleições, em estimativas de crescimento econômico do Brasil, em estimativas de vendas para o comércio. E já deve ter utilizado as notas das disciplinas que cursou para calcular sua média escolar, para saber se seria aprovado. Em todas essas ocasiões, como em muitas outras que você conhece, e recorda, há técnicas estatísticas



Fig. 02 – Carl Friedrich Gauss - 1777-1855

Fonte: http://www.uibra.br/SITE_ULBRA/fo-to_s_noticia/a3af20915b565a6f.jpg

principalmente para cobrar impostos. Por isso você pode notar o quanto esses dados eram importantes para o planejamento das cidades e dos governos, de um modo geral. De posse dessas informações, seria possível planejar estimativas de receitas financeiras como coleta de impostos; ou seria possível ter um mínimo de estimativa sobre a população, a partir do histórico do número de casamentos e do número de nascimentos, por exemplo.

No século XVIII, esses estudos foram denominados como uma nova ciência com o nome de Estatística. Surgiram tabelas mais complexas, representações gráficas e cálculo de probabilidade. Com o advento dessas técnicas gráficas e de probabilidades, tornou-se possível identificar tendências no comportamento dos dados: já era possível saber se os nascimentos iriam crescer, ou estacionar, ou decrescer; e, com isso, poderiam ser propostas políticas ligadas ao matrimônio, no sentido de responder àquela situação detectada no comportamento gráfico. Ou seja, se havia decréscimo de casamentos, e era interessante para o governo que houvesse um número mais elevado, várias políticas governamentais poderiam ser definidas no sentido de reverter essa tendência. Como você pode perceber à luz desses dois momentos históricos da estatística, em sua essência, ela é a ciência que apresenta processos próprios para coletar, apresentar e interpretar adequadamente conjuntos de dados, sejam eles numéricos ou não. Pode-se dizer que seu objetivo é o de apresentar informações sobre dados em análise para que se tenha maior compreensão dos fatos que os mesmos representam.

A Estatística subdivide-se em três áreas: **descritiva**, **probabilística** e **inferencial**. A estatística descritiva, como o próprio nome já diz, preocupa-se em descrever os dados. A estatística inferencial, fundamentada na teoria das probabilidades, ocupa-se da análise desses dados e sua interpretação.

É importante também que você conheça alguns conceitos básicos referentes ao trabalho estatístico. São conceitos que precisam ser sempre bem aplicados quando você estiver planejando um estudo que tenha análise estatística. Vamos ver alguns desses conceitos?

- **População:** conjunto de todos os elementos relativos a um determinado fenômeno que possuem pelo menos uma característica em comum. Na sua cidade tem uma população, que pode ser expressa pelo número total de habitantes; também na sua cidade há pessoas que estudam, elas formam a população escolar da sua cidade.

- **Amostra:** é um subconjunto da população e deverá ser considerada finita. A amostra deve ser selecionada seguindo certas regras e deve ser representativa, de modo que ela represente todas as características da população como se fosse uma fotografia desta. Na população escolar, uma amostra poderia ser obtida apenas dos alunos que estudam o ensino fundamental: de cada escola, seriam importantes apenas aqueles alunos que estivessem no ensino fundamental.

- **Variável:** é aquilo que se deseja observar para se tirar algum tipo de conclusão, é uma característica que pode ser mensurada. Ainda considerando a situação da amostra de alunos do ensino fundamental de sua cidade, você poderia querer avaliar a idade desses alunos, como a idade é uma característica que possui valor, ela seria uma variável importante no seu estudo; assim também você poderia eleger como variável a renda familiar desses alunos.

As variáveis podem ser classificadas dos seguintes modos:

- **Qualitativas:** São características de uma população que não podem ser medidas. No exemplo da amostra de alunos do ensino fundamental, comentado antes, uma variável qualitativa poderia ser o turno que os alunos estudam, pois o turno é uma característica que não pode ser mensurada, é uma categoria tipo Matutino/Vespertino/Noturno.

- **Quantitativas:** são características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas. Na situação

que foi tratada anteriormente, a idade e a renda familiar seriam variáveis quantitativas, pois podem ser mensuradas. As variáveis quantitativas podem ter valores inteiros ou podem ter valores com partes fracionárias. A renda familiar fica mais bem representada com valores que podem conter partes fracionárias, tipo R\$ 750,35, ou R\$ 567,92; já o número de alunos em cada sala de aula do ensino fundamental não contém parte fracionária, tipo 40 alunos, 35 alunos. Então, as variáveis quantitativas que só possuem valor inteiro são denominadas variáveis discretas. Aquelas variáveis quantitativas que podem ter parte fracionária são denominadas variáveis contínuas.

- **Discretas:** são aquelas variáveis que pode assumir somente valores inteiros num conjunto de valores. São geradas pelo processo de contagem, como o número de veículos que passa em um posto de gasolina ou o número de estudantes numa sala de aula, por exemplo.
- **Contínuas:** são aquelas variáveis que podem assumir um valor dentro de um intervalo de valores. São geradas pelo processo de medição. Nesse caso serve como exemplo o volume de água em um reservatório ou o peso de um pacote de cereal.
- **Dados estatísticos:** são valores que as variáveis podem assumir. A variável renda familiar pode assumir o valor R\$ 750,35 ou algum outro. A variável número de alunos por sala de aula no ensino fundamental pode ter como valores 40 alunos ou algum outro valor discreto.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre Estatística até agora:

1. Identifique quais as variáveis que seriam necessárias para estudar a evasão escolar de sua cidade;
2. Explique a importância dessas variáveis para o estudo em questão.



FASES DO TRABALHO ESTATÍSTICO

Agora que você já deve estar bastante consciente da importância da estatística, e até já identifica, com clareza, variáveis importantes para analisar determinados fenômenos, inclusive fenômenos de interesse ambiental, é muito importante que você saiba as fases do trabalho estatístico. Com essas fases você certamente vai poder planejar adequadamente as análises estatísticas futuras para seus trabalhos, tanto do Curso de Gestão Ambiental, como outros trabalhos de cunho profissional e pessoal.

O trabalho estatístico, portanto, abrange as seguintes fases:

- a) Definição do Problema - consiste na formulação correta do problema, no exame de outros levantamentos realizados no mesmo campo, no que se pretende pesquisar, definindo o problema corretamente, por exemplo, identificando a problemática, a população, variáveis de interesse, fontes de dados.
- b) Planejamento – após a definição do problema, é preciso determinar o procedimento necessário para resolver o problema, por exemplo: levantar

informações; identificar tipos de levantamento - Censo (completo) ou Amostragem (parcial) - Cronograma, Custos.

c) Coleta ou levantamento dos dados – nessa fase procura-se obter os dados referentes ao trabalho que desejamos fazer. A coleta pode ser: direta - diretamente da fonte - ou indireta - feita através de outras fontes. Os dados podem ser obtidos pela própria pessoa (primários) ou se basear no registro de terceiros (secundários).

d) Apuração dos Dados ou sumarização – nesse momento procede-se a elaboração de um resumo dos dados através de contagem e agrupamento. É um trabalho de coordenação e de tabulação.

e) Apresentação dos dados - é a fase em que vamos mostrar os resultados obtidos na coleta e na organização. Essa apresentação pode ser: tabular ou gráfica.

f) Análise e interpretação dos dados – essa é a fase final, a mais importante e também a mais delicada. Obtém conclusões que auxiliam o pesquisador a resolver seu problema.

EXEMPLO APLICANDO FASES DO TRABALHO ESTATÍSTICO

Você deve estar se perguntando como seria possível realizar um estudo estatístico em sua cidade, um estudo que, por exemplo, pudesse avaliar a situação da evasão escolar em sua cidade. Esse parece ser um bom exemplo para você compreender a importância e o desenvolvimento dessas fases, e também para ter a certeza de que tendo conhecimento de técnicas estatísticas poderá responder a diversos questionamentos de forma adequada.

A problemática poderia ser avaliar a tendência de evasão entre alunos do nível médio escolar na sua cidade, no período de 2000 a 2007. A população seria composta de todos os alunos evadidos das escolas de nível médio de sua cidade. O número de alunos evadidos por escola, por ano, seria a variável principal de interesse.

Para responder a esta problemática, você deveria ter informações acadêmicas, provavelmente um Censo Escolar de todas as escolas que têm curso de nível médio, desde 2000 até 2007. Precisaria também ter o número de alunos evadidos de cada escola, em cada ano da série estudada. Você deveria também identificar quais técnicas estatísticas seriam necessárias, por exemplo, técnicas de tabulação? A média? O desvio-padrão? Apresentações gráficas? Poderia ser útil, ainda, definir um cronograma para as diversas atividades, como, por exemplo:

| Atividade | Agosto | Setembro | Outubro | Novembro |
|---------------------|--------|----------|---------|----------|
| Coleta | ----- | | | |
| Apuração | | ----- | ----- | |
| Apresentação | | | ----- | |
| Análise | | | | ----- |

A coleta de dados poderia ser eletrônica, pois provavelmente há censos educacionais em algum portal, como o IBGE, na Secretaria Estadual de Educação, ou no IDEMA. Caso contrário, você teria que utilizar os dados das secretarias das escolas de nível médio de sua cidade e registrar os dados um a um.

Para a apuração você, certamente, utilizaria técnicas de tabulação, como distribuição de frequências, medidas de tendência central, como a média, e medidas de variação, como o desvio-padrão.

Deveria, posteriormente, apresentar esses resultados sob a forma de tabelas, gráficos. Essa seria a apresentação dos resultados.

Finalmente, você deveria fazer um relatório analítico sobre a situação apresentada à luz dos resultados. Provavelmente você iria ter que afirmar que há uma tendência de crescimento/decrescimento da evasão, se ela é mais caracterizada no turno noturno, ou em alunos que trabalham. Essa seria a parte final do seu trabalho estatístico, seria a análise.

ATIVIDADE 02



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre as fases do trabalho estatístico até agora:

1. Identifique as fases de um trabalho estatístico para analisar o número de nascimentos em sua cidade. Quais seriam suas fontes para coleta de dados?

RESUMINDO



Nesta aula, você aprendeu um pouco sobre a história da estatística, seus conceitos básicos e suas possibilidades de aplicação em pesquisas de diversas naturezas. Aprendeu, ainda, como se dão as principais fases de elaboração de uma pesquisa estatística e como elas podem ser aplicadas à realidade de sua cidade e de seu curso.

LEITURAS COMPLEMENTARES



O texto recomendado permitirá que você compreenda como empresas, governos e sociedade podem se beneficiar do uso da Estatística na análise de seus dados, principalmente para utilizar melhor os recursos disponíveis.

COSTA, Eduardo et all. **A Estatística no Mundo Moderno**. Disponível em: < <http://www.administradores.com.br/producao-academica/a-estatistica-no-mundo-moderno/518/> >. Acessado em 01/12/2013.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já compreende bem a importância e as fases do trabalho Estatístico, descreva como uma análise do número mensal de casos de dengue poderia permitir uma melhoria na qualidade de vida dos moradores de sua cidade.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 02
ESTATÍSTICA DESCRITIVA: SÉRIES ESTATÍSTICAS

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Na maioria de suas leituras sobre estatística você deve ter visto bastante tabelas resumindo dados. É assim quando lemos sobre a intenção de votos de candidatos, sobre o número de alunos evadidos por turno, sobre casos de dengue por município.

Enfim, as análises estatísticas sempre se iniciam com um resumo de dados, para melhor agrupá-los e permitir assim uma melhor compreensão de aspectos que esses dados têm em comum e que isoladamente não seria possível identificar. É exatamente isto que você vai estudar nesta aula. Vamos lá?

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- Compreender o que é a estatística descritiva;
- Identificar séries estatísticas;
- Construir tabelas para representar séries estatísticas.

DESENVOLVENDO O CONTEÚDO

SÉRIE ESTATÍSTICA

Quando você inicia um trabalho com dados Estatísticos, muitas vezes o conjunto de valores originais mostra-se bastante desorganizado, seja porque são muito dados, porque os dados são muito diferentes entre si ou por outro motivo. Será necessário, então, utilizar alguns recursos que permitam resumir esses dados, agrupando-os de acordo com algum critério. Por exemplo, você já deve ter lido resultados de pesquisas de opinião sobre os principais problemas



Fig. 01 – Agrupando dados.

do Brasil, e certamente compreendeu que as pessoas consultadas mostraram que os principais problemas atualmente são a insegurança e a saúde. Esses dados, provavelmente, estavam resumidos em tabelas. Foram entrevistadas várias pessoas e uma equipe técnica utilizou alguns critérios para montar uma tabela bem simples de ser compreendida.

Os dados em geral fazem parte de uma série denominada série estatística. Uma série estatística, portanto, define-se como toda e qualquer coleção de dados estatísticos, ou seja, a série é uma seqüência de números que se refere a uma certa variável.

Assim, você pode ter uma série com os dados de evasão de sua cidade, por escola; pode ter uma série de casos de dengue notificados diariamente. Você já deve estar pensando que as notas dos alunos da disciplina de Estatística de uma turma de Gestão Ambiental formam também uma série; e de fato formam mesmo, muito bem!

Para obter representações bem explicativas, é muito importante resumir organizadamente os dados das séries, pois, como você já deve ter compreendido, isso facilita a análise inicial do leitor sobre aquele fenômeno em estudo.

Para um melhor entendimento de séries estatísticas, pode ser útil classificá-

las, cada uma classificação sendo caracterizada por um elemento agrupador. Você vai estudar os tipos que seguem:

Série Temporal

Uma série temporal, caracteriza-se por ter seus dados ao longo do tempo, formando uma série em que os valores são registrados por dia, por mês, por ano, ou por algum outro elemento temporal. Na sua cidade deve ter registros de casos de dengue por mês, e o conjunto desses dados formam uma série temporal da variável casos de dengue. Uma série temporal também pode ser denominada de série histórica.

Série Geográfica

Uma série geográfica, caracteriza-se por ter seus dados segundo a localidade de ocorrência, formando uma série onde os valores são registrados por local. Na sua cidade deve ter registros de alunos matriculados em escolas da área urbana e área rural, e o conjunto desses dados formam uma série geográfica da variável alunos matriculados. Uma série geográfica também pode ser denominada de série espacial.

Série Específica

Uma série específica, caracteriza-se por ter seus dados segundo um critério específico diferente do critério temporal e do geográfico. Na sua cidade deve ter dados de professores por titulação acadêmica, e o conjunto desses dados formam uma série específica da variável professores de sua cidade. Uma série específica também pode ser denominada de série categórica.

DISTRIBUIÇÃO DE Frequência

Na realidade, a distribuição de frequência é uma série estatística bastante peculiar pois nela os dados são agrupados em suas respectivas categorias numéricas. Você já deve ter compreendido que o número de alunos de sua cidade, que foi tratado anteriormente, registrado por área geográfica - que não é uma categoria numérica, pois não possui valor - formam uma séria geográfica. Porém pode ser necessário estudar o número de alunos de acordo com a renda familiar, que é um categoria numérica; neste caso, haverá para cada aluno um valor correspondente de renda familiar, neste caso para ser possível resumir conjuntamente alunos e rendas familiares será necessário construir uma distribuição de frequências. Essa distribuição conterà o número de alunos em cada faixa de renda estabelecida.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre séries estatísticas até agora:

1. Identifique quais os tipos das séries idade dos alunos da turma da disciplina de Estatística de Gestão Ambiental e número de habitantes da sua cidade na área urbana e rural.



APRESENTAÇÃO DE SÉRIES ESTATÍSTICAS

Como você já deve ter compreendido, as séries estatísticas são muito importantes para o início de um trabalho Estatístico, agora você vai aprender que a forma de apresentá-las também tem uma grande importância para garantir que sua leitura seja satisfatória, e assim permita que o fenômeno em estudo seja bem analisado.

Recomenda-se apresentar séries sob a forma de tabelas e gráficos. Você, a partir de agora, vai aprender as partes que compõem tabelas e gráficos; as partes a seguir apresentadas são comuns a ambos.

Há vários critérios para a elaboração de tabelas e gráficos; os critérios que serão aqui apresentados estruturam ambos em: cabeçalho, corpo e rodapé.

Nesta aula você vai compreender a representação de séries especificamente utilizando tabelas, em outras aulas você compreenderá sobre a representação gráfica.

No cabeçalho devem constar informações que permitam ao leitor identificar qual é a variável de interesse, bem como o período e o local aos quais ela se refere. Considere a tabela 1 a seguir contendo o número de matrículas no CEFET/RN em 2007.

Tabela 01 - Quadro Resumo das Matrículas do CEFET/RN, Em 2007

| Modalidade | Total |
|--|---------------|
| Pós-graduação | 286 |
| Superior | 1.751 |
| Técnicos Subsequentes | 2.567 |
| Técnicos Integrados | 2.573 |
| Ensino Médio | 49 |
| PROCEFET | 4.438 |
| Educação Inicial e Continuada de Trabalhadores | 4.307 |
| Total Geral | 15.971 |

Fonte: Relatório de Gestão 2007

Como você já deve ter identificado na tabela 1, o cabeçalho é composto pela descrição “Quadro Resumo de Matrículas do CEFET/RN, Em 2007”.

No corpo devem ser representados os dados, geralmente sob a forma de colunas: com as primeiras linhas sendo de títulos, e as demais abaixo com os dados.

Assim, ainda considerando a tabela 1, o corpo é formado pelas colunas modalidade e total, e as linhas são os valores Pós-Graduação com 286 matrículas, e assim sucessivamente até Educação Inicial e Continuada de Trabalhadores com 4.307 matrículas.

Já o rodapé deve conter informações sobre as fontes dos dados. Pode-se utilizar também o rodapé para apresentar legendas ou explicitar situações especiais, como por exemplo informar que determinados dados não puderam ser obtidos e o motivo.

Para a tabela 1, o rodapé é a fonte de onde foram obtidos os dados, neste caso, a fonte é o Relatório de Gestão de 2007.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre as fases do trabalho estatístico até agora construa uma tabela com número de habitantes de sua cidade de 2000 a 2010.



RESUMINDO



Nesta aula, você aprendeu sobre séries estatísticas, seus tipos e formas de representação tabular e gráfica. Aprendeu, ainda, que um trabalho Estatístico deve dedicar especial atenção a esses conteúdos, pois todas as técnicas serão aplicadas posterior sob esses representações, de modo que elas passam a ser a base para serem aplicadas as demais técnicas estatísticas.

LEITURAS COMPLEMENTARES



Neste texto você encontrará estatísticas sobre dengue, de janeiro a abril de 2008, e poderá compreender como as séries estatísticas nele representadas permitem ao leitor uma compreensão bastante satisfatória do fenômeno . Disponível em < <http://www.combateadengue.com.br/?p=271> >. Acessado em 01/12/2013.

AVALIANDO SEUS CONHECIMENTOS



Considerando-se que você já compreende bem a importância que as séries estatísticas representam para o trabalho Estatístico, e sabe como representá-las em tabelas, construa uma tabela com a população de sua cidade no último censo, organizada por faixas etárias. Você pode utilizar as faixas etárias que julgar mais conveniente.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 03
DISTRIBUIÇÃO DE FREQUÊNCIA

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Como já foi estudado na aula sobre Estatística Descritiva séries estatísticas, o resumo de dados é a base para a aplicação das demais técnicas estatísticas, você já sabe que as tabelas são a forma mais usual de resumir dados.

A partir de agora você deve estar se perguntando se é só organizar os dados em linhas de uma tabela e pronto, eles estarão bem representados e poderão ser compreendidos.

Quando você dispõe de poucos dados até que pode simplesmente criar suas próprias tabelas, sem critérios muito técnicos, mas a realidade é que a Estatística em quanto ciência está preparada e é recomendada para tratar grandes volumes de dados, e nestes casos é preciso utilizar-se de critérios técnicos para construir essas tabelas, que são denominadas especialmente de distribuição de frequências, e é o que você vai estudar nesta aula.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- Compreender as técnicas para resumo de conjuntos de dados;
- Construir distribuições de frequências;
- Interpretar as informações contidas em distribuições de frequências.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO DISTRIBUIÇÃO DE FREQUÊNCIAS

Na maioria das tomadas de decisão que envolvem dados numéricos, ou seja, variáveis quantitativas, a Estatística é utilizada, e a partir de suas técnicas de análise é que decide-se sobre o que fazer.

É assim quando a Estatística é utilizada para afirmar se uma turma de alunos de determinado Curso teve um rendimento esperado, a partir de uma análise das notas dos mesmos. É assim também quando uma empresa decide se instalar



Fig. 01 – Material de dados.

http://cdn2.freepik.com/fotos-gratis/forma-ve-torial-material-de-dados-estatisticos_15-8778.jpg

em sua cidade, ela busca dados representativos sobre consumo, mão de obra, condições de acesso e alguns outros dados que você já pode imaginar, e realiza análises estatísticas para avaliar se poderá produzir, vender, entregar de forma satisfatória.

Você já identificou claramente que a Estatística é uma ciência que utiliza-se de dados e tenta analisá-los de tal forma que seja possível identificar um comportamento, uma tendência, enfim, algum aspecto mais geral que possa explicar como todo o conjunto de dados pode ser compreendido. E para isso é preciso resumir os dados, o que geralmente ocorre com uma distribuição de frequências.

Você já deve estar querendo ver uma distribuição de frequências e tentar compreender ainda mais sua importância.

Então considere que você tem as notas de alunos de uma turma de Informática e precisa analisá-las utilizando uma distribuição de frequências. Utilizando-se de técnicas adequadas, pode-se construir uma distribuição de frequências como a apresentada na tabela 01.

Tabela 01 - Notas de uma turma de Informática.

| Notas | Alunos |
|-----------------|---------------|
| De 0,0 até 2,0 | 2 |
| De 2,1 até 4,0 | 6 |
| De 4,1 até 6,0 | 8 |
| De 6,1 até 8,0 | 10 |
| De 8,1 até 10,0 | 4 |
| Total | 30 |

Fonte: Secretaria Escolar, 2007

Você já deve ter compreendido com este simples exemplo que a turma tinha 30 alunos, que os dados foram obtidos junto à Secretaria Escolar e referem-se a 2007. Com certeza, já observou que apenas 2 de seus alunos obtiveram notas bastante baixas, de 0,0 até 2,0; e que há muitos alunos com notas de 6,1 a 8,0, são 10 alunos; que apenas 4 alunos obtiveram notas bastante altas de 8,1 a 10,0; e que 4 alunos obtiveram notas de 2,1 a 4,0. Já seria possível afirmar algo ou decidir sobre o rendimento dessa turma?

Considere agora que você está avaliando o nível econômico das famílias de determinada cidade e decidiu também iniciar o trabalho com a construção de uma distribuição de frequências. Utilizando-se de técnicas adequadas, construiu a distribuição de frequências apresentada na tabela 02.

Tabela 02 - Renda familiar em determinada cidade.

| Renda familiar (em salários mínimos) | Número de Famílias |
|---|---------------------------|
| De 0,0 até 3,0 | 10 |
| Acima de 3,0 até 6,0 | 20 |
| Acima de 6,0 até 9,0 | 45 |
| Acima de 9,0 até 12,0 | 15 |
| Acima de 12,0 | 10 |
| Total | 100 |

Fonte: IBGE, 2007

Pelo exemplo você analisou 100 famílias e suas respectivas rendas, em salários mínimos, e obteve os dados junto ao IBGE, e esses dados referem-se a 2007. Com certeza já observou que apenas 10 famílias têm renda acima de 12,0 salários mínimos, que 60 famílias têm renda acima de 3,0 até 12,0 salários mínimos, e que 30 famílias têm renda de 0,0 até 6 salários mínimos. Já seria possível afirmar algo ou decidir sobre o nível econômico dessas famílias?

Essas duas situações de notas e rendas familiares, na tabela 1 e na tabela 2, já lhe permite compreender como uma simples distribuição de frequências já exhibe com bastante clareza uma situação que cotidianamente você pode vivenciar e agora já sabe que pode dar uma possível solução para ela.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre distribuição de frequências até agora:

1. Utilize as idades apresentadas a seguir e construa uma distribuição de frequências para as mesmas. Considere que essas idades, em anos, são de turistas que visitaram, em 2012, uma reserva ambiental em sua cidade. Explique qual o critério utilizado para distribuir as idades, e porque o utilizou. Idades: 20, 28, 30, 45, 33, 29, 32, 25, 20, 35, 34, 22, 41, 38, 27.



CONSTRUINDO DISTRIBUIÇÃO DE FREQUÊNCIAS

Depois de ver exemplos e compreendê-los, você certamente já está querendo colocar em prática tudo o que já sabe. Então, a partir de agora, você vai aprender a construir distribuições de frequências.

Há várias metodologias para se construir uma distribuição de frequência, você irá aprender uma que é bastante utilizada, segue um roteiro relativamente simples e produz resultados bastante satisfatórios.

Para construir uma distribuição de frequências, será necessário aplicar as seguintes técnicas:

- Identificar o número de dados – **n**;
- Colocar o conjunto dos dados em ordem – esse conjunto ordenado recebe o nome de **ROL**;
- Calcular a amplitude total – **AT**

$$\mathbf{AT = MaiorValor - MenorValor;}$$

- Calcular o número de classes – **c**

$$c = \begin{cases} 5, & \text{se o número de dados } n \leq 25; \\ \sqrt{n}, & \text{se o número de dados } n > 25. \end{cases}$$

- Calcular a amplitude de classe – **a = AT/c**
- Construir cada classe, identificando seus **limites: Inferior e Superior**;
- Obter as **frequências** de cada classe, observando o ROL e os Limites de cada classe;
- Obter as **porcentagens** de cada classe, observando suas respectivas frequências.

Você já deve estar se perguntando: como colocar essas técnicas em prática. Você vai fazer isso a partir de agora com um exemplo que contém dados referentes à precipitação pluviométrica (em mm) durante vinte e dois

As precipitações pluviométricas (mm) diárias são: 18, 18, 17, 16, 19, 19, 17, 18, 20, 20, 22, 25, 23, 22, 21, 21, 19, 17, 18, 16, 15, 20.

As precipitações pluviométricas (mm) diárias são: 18, 18, 17, 16, 19, 19, 17, 18, 20, 20, 22, 25, 23, 22, 21, 21, 19, 17, 18, 16, 15, 20.

Como você pode perceber, no primeiro dia representado pelos nossos dados a precipitação foi de 18mm; no segundo dia foi de 18mm; e assim, sucessivamente, até que no penúltimo dia foi de 15mm e no último dia essa precipitação foi de 20mm.

Então, seguindo as técnicas, você deve primeiro identificar o número de dados: nesse caso, como temos 22 dias, $n = 22$. A próxima técnica é obter o **ROL**, ordenando o conjunto de dados: você vai começar pelo menor dado, e vai obter todos os demais até obter o maior deles. O menor é **15**, na ordem crescente o próximo é **16**, e assim sucessivamente, você construirá o seguinte **ROL**.

ROL: 15, 16, 16, 17, 17, 17, 18, 18, 18, 18, 19, 19, 19, 20, 20, 20, 21, 21, 22, 22, 23, 25.

A partir do **ROL** você vai calcular agora a Amplitude Total:

AT = MaiorValor – MenorValor, ou seja,

$$\mathbf{AT = 25 - 15 = 10}$$

Você precisa calcular o número de classes, para isso vai utilizar o número de dados $n=22$. A regra que foi definida anteriormente diz que se o número de dados for menor ou igual a 25, o número de classes será 5. Como nesse exemplo das precipitações pluviométricas, o número de dados é 22, e esse número é menor ou igual a 25, então:

$$\mathbf{c = 5.}$$

$$\mathbf{c = 5.}$$

Como você já sabe que sua distribuição de frequências terá 5 classes, precisa agora saber qual é a amplitude de cada classe. Para isso irá aplicar a técnica da amplitude de classe.

a = AT / c, ou seja,

a = 10 / 5, o que resulta em

$$\mathbf{a = 2.}$$

Você já tem os valores necessários para construir cada uma das 5 classes da distribuição de frequências. A primeira classe, como você já deve ter imaginado, terá como **LimiteInferior** o primeiro dado do **ROL**, e como **LimiteSuperior** esse valor somado com a amplitude de classe **a**. Ou seja,

$$\text{LimiteInferior} = 15 \text{ e}$$

$$\text{LimiteSuperior} = \text{LimiteInferior} + a, \text{ logo}$$

$$\text{LimiteSuperior} = 15 + 2 = 17$$

Portanto, a **primeira classe** conterá valores de 15 a 17 e será representada por: **15 |----- 17**.

A **segunda classe** começa exatamente onde a primeira classe termina, ou seja, o **LimiteInferior da segunda classe será igual ao LimiteSuperior da primeira classe**, portanto para a segunda classe

$$\text{LimiteInferior} = 17.$$

Como todas as classes têm a mesma amplitude de classe, o LimiteSuperior da segunda classe será obtido da seguinte forma:

$$\text{LimiteSuperior} = \text{LimiteInferior} + a, \text{ logo}$$

$$\text{LimiteSuperior} = 17 + 2 = 19.$$

Portanto, a **segunda classe** conterá valores de 17 a 19 e será representada por: **17 |----- 19**.

Seguindo esse mesmo raciocínio, você construirá as demais classes. A terceira classe conterá valores de 19 a 21, a quarta conterá valores de 21 a 23, e a quinta de 23 a 25. Desse modo, a distribuição de frequências terá as classes apresentadas na tabela 03.

Tabela 03 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | Número de Dias - Frequências |
|--------------------------------------|------------------------------|
| 15 ----- 17 | |
| 17 ----- 19 | |
| 19 ----- 21 | |
| 21 ----- 23 | |
| 23 ----- 25 | |

Fonte: Dados primários.

Você, ao observar a tabela 3, deve estar querendo obter as frequências referentes a cada uma das classes de precipitação pluviométrica. E essa é a próxima técnica que vamos utilizar. Para isso você deve recordar que precisará observar o ROL e os limites inferior e superior de cada classe, para contar quantos dados no ROL pertencem a cada classe.

Para a primeira classe, a frequência será o número de dados do ROL que são maiores ou iguais a 15 e menores que 17. Muito bem, você já entendeu que a representação **15 |----- 17** contém o 15, mas não contém o 17. Então, contando no ROL, você terá os seguintes dados para essa primeira classe: 15, 16, 16; **logo a frequência será igual a 3.**

Já a segunda classe, cuja representação é **17 |----- 19**, contém o 17, mas não contém o 19. Então, contando no ROL, você terá os seguintes dados para essa segunda classe: 17, 17, 17, 18, 18, 18, 18; **logo a frequência será igual a 7.**

Para a terceira classe, cuja representação é **19 |----- 21**, que contém o 19, mas não contém o 21. Então, contando no ROL, você terá os seguintes dados para essa terceira classe: 19, 19, 19, 20, 20, 20; **logo a frequência será igual a 6.**

Para a quarta classe, cuja representação é **21 |----- 22**, que contém o 21, mas não contém o 23. Então, contando no ROL, você terá os seguintes dados para essa quarta classe: 21, 21, 22, 22; **logo a frequência será igual a 4.**

Finalmente, para a quinta classe, cuja representação é **23 |-----| 25**, que contém o 23 e também contém o 25. É isso mesmo, na última classe, e apenas na última classe, o limite superior é incluído; não esqueça disso! Então, contando no ROL, você terá os seguintes dados para essa quinta classe: 23, 25 ; **logo a frequência será igual a 2.**

Assim, após a obtenção das frequências de cada classe, você terá sua distribuição de frequências como a apresentada na tabela 04.

Tabela 04 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | Frequência - f |
|--------------------------------------|----------------|
| 15 ----- 17 | 3 |
| 17 ----- 19 | 7 |
| 19 ----- 21 | 6 |
| 21 ----- 23 | 4 |
| 23 ----- 25 | 2 |
| Σ | 22 |

Fonte: Dados primários.

Para finalizar a construção da distribuição de frequências, é necessário calcular as porcentagens para cada classe. Para isso você deve utilizar a seguinte fórmula:

$$\% = \left(\frac{f}{\sum f} \right) * 100$$

onde f é a frequência de cada classe; e $\sum f$ é a soma das frequências.

Portanto, você precisa obter as porcentagens de cada classe. Para a primeira classe, você terá: $\% = (3/22)*100$, logo a primeira porcentagem será 13,63636, que arredondando com duas casas decimais resulta em 13,64%.

Desse modo, a distribuição de frequências completa está apresentada na tabela 5.

Tabela 05 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | Frequência - f | Porcentagens - % |
|--------------------------------------|----------------|------------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

Como você já verificou, a partir da tabela 5 pode-se concluir que há uma grande porcentagem (31,82%) de dias, neste município, que chove entre 17 mm e 19 mm. Como também há poucos dias (9,09%) com chuvas acima de 23 mm.

A partir de agora você poderá resumir dados para suas mais diversas aplicações.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

1. Identifique exemplos reais de utilização de distribuição de frequências e comente-as. Você deve dar especial atenção para as frequências e as porcentagens.



RESUMINDO



Nessa aula, você aprendeu a importância de resumir dados, e a identificar tendências, comportamentos e aspectos que possam melhor analisá-los. Aprendeu também que uma forma adequada de resumir dados, é construir distribuições de frequências, com classes, frequências e porcentagens.

LEITURAS COMPLEMENTARES



Neste texto você encontrará uma distribuição de frequências elaborada com as idades de internautas que acessaram o site. No texto, há uma breve análise sobre o perfil etário desses internautas.

Disponível em < <http://www.blogviche.com.br/2007/03/03/distribuicao-de-frequencias-pesquisa-sobre-faixa-etaria/>>. Acessado em 15/12/2013.

AVALIANDO SEUS CONHECIMENTOS



Considerando-se que você já sabe resumir dados, utilize os dados a seguir, referentes a notas de alunos de Estatística, para construir uma distribuição de frequências.

Notas: 8, 9, 5, 0, 10, 3, 2, 9, 7, 2, 1, 3, 6, 8, 9, 10, 4, 2, 6, 1, 10.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 04
MEDIDAS DE TENDÊNCIA CENTRAL

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL

APRESENTANDO A AULA



Como já foi estudado na aula sobre Distribuição de Frequências, é muito importante resumir os dados na forma de Frequências uma vez que os critérios utilizados não são subjetivos e sim fruto de técnicas independentes dos dados, o que garante o seu resumo mais efetivo. Você já sabe, por exemplo, que o número de classes não é algo pessoal e sim resultado de uma técnica que considera o número de dados: quanto maior o número de dados, maior o número de classes para representá-los.

Você deve estar se perguntando o que fazer para avaliar uma distribuição de Frequências, utiliza-se alguma medida para representar todos os dados resumidos? É um simples processo de observação das classes, das Frequências e das porcentagens?

De fato, a Estatística, como ciência, dispõe de várias técnicas para obter medidas e representar todos os dados resumidos. A idéia geral é que essas medidas teriam em torno de si a maioria dos dados resumidos, de modo que elas poderiam, por isso, representá-los de forma satisfatória, pois haveria mais dados em torno delas do que distantes delas. Essas medidas, que são denominadas especialmente de medidas de tendência central, e seu processo de obtenção é o que você vai estudar nesta aula.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender as técnicas para obter medidas de tendência central;
- calcular média, mediana e moda;
- interpretar as informações dos valores médio, central e mais frequente.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO MEDIDAS DE TENDÊNCIA CENTRAL

Em uma distribuição de frequências há informações importantes sobre o comportamento geral dos dados. Considere, por exemplo, a tabela 1, construída na aula 3 - Distribuição de Frequências, na qual se pode identificar que há pequena porcentagem de dias com altas precipitações pluviométricas e que, ao contrário, há grande porcentagem de dias com baixas precipitações pluviométricas.

Tabela 01 - Distribuição de Frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | Frequência - f | Porcentagens - % |
|--------------------------------------|----------------|------------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

De acordo com a tabela 1, são 45,46% (13,64+31,82) dos dias com chuva de 15mm a próximo de 19mm; enquanto que em apenas 27,27% (18,18+9,09) dos dias chove de 21mm a 25mm. Ou seja, chove pouco nesse município. Mas esse pouco pode ser representado por uma única medida? E essa medida seria representativa dos 22 dados de precipitação pluviométrica?

Você já deve estar querendo apresentar respostas para essas indagações. Muito bem, a seguir serão apresentadas três possíveis respostas para uma mesma situação: a seleção para o cargo de consultor ambiental de determinada empresa que realizou uma bateria de 5 avaliações com os candidatos. Suponha que o candidato vencedor obteve as seguintes notas 3, 4, 7, 8, 8.

Uma primeira solução certamente é a medida que você utiliza para representar seu rendimento escolar: por exemplo, quando você faz duas avaliações, na primeira obtém nota 9,0 e na segunda nota 10,0, o valor representativo dessas notas pode ser 9,5. No caso da seleção para o cargo de consultor ambiental, o valor médio do candidato vencedor poderia ser 6,0 $((3+4+7+8+8)/5)$. Essa medida 6,0, que não é igual à primeira nota, nem à segunda nota, nem à nenhuma outra, é uma medida de tendência central, denominada de média, pois ela traz informações centrais que são próximas de todas as notas.

Uma segunda possível resposta, seria identificar um valor que estivesse bem no centro dos dados. Nesse caso, a nota que está no centro das notas obtidas pelo vencedor da seleção é a nota 7. Por esta medida, denominada de mediana, o candidato teria um valor central igual a 7, pois as demais notas estão uma metade abaixo deste valor e a outra metade acima dele.

Finalmente, uma terceira resposta seria identificar qual foi a nota mais frequente do candidato. Nesse caso, a nota mais frequente dentre as notas obtidas pelo vencedor da seleção é a nota 8. Por esta medida, denominada de moda, o candidato teria um valor mais frequente igual a 7, pois esta nota foi a que se repetiu mais vezes; o candidato obteve duas notas 8.

Essas três medidas de tendência central, média, mediana e moda são bastante úteis para representar o comportamento geral dos dados. Ocorre, no entanto, que nem sempre o conjunto de dados é tão pequeno e simples como o que foi tratado nesse exemplo da seleção do consultor ambiental. Você já sabe que a Estatística foi aperfeiçoada ao longo do tempo para processar uma quantidade bastante grande de dados, de modo que para um conjunto com muitos dados, a identificação de média, mediana e moda não é uma simples observação como a que você já compreendeu no exemplo da seleção do consultor ambiental, para esses casos mais complexos há técnicas mais apuradas. É isso que você vai passar a estudar agora.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre medidas de tendência central até agora, utilizando os números de multas mensais aplicadas por um órgão ambiental em determinado município, apresentados a seguir, obtenha a média, mediana e moda.



| Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 9 | 9 | 6 | 2 | 3 | 4 | 9 | 2 | 4 | 6 | 6 | 4 |

CALCULANDO MÉDIA, MEDIANA E MODA

Agora que você já sabe o que são medidas de tendência central e sabe obtê-las para o pequeno conjunto de dados não agrupados, pois foi isto que você já compreendeu na seção anterior e já fez a atividade correspondente, você deve estar se perguntando como se deve calcular média, mediana e moda quando se tem muitos dados e eles estão agrupados em uma distribuição de frequências. Você vai começar a aprender esses cálculos a partir de agora.

Para calcular a **média para dados agrupados**, você utilizará a seguinte fórmula:

$$(4.1.) \quad \bar{x} = \frac{\sum (f * X)}{\sum f}$$

Onde, f representa a frequência de cada classes; e

X representa o ponto médio de cada classe, que é obtido da seguinte forma $(\text{LimiteInferior} + \text{LimiteSuperior})/2$.

Como você já entendeu, os pontos médios para cada classe da distribuição de frequências precisam ser calculados. Veja na tabela 2 a obtenção de cada ponto médio.

Tabela 02 - Distribuição de Frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % | X |
|--------------------------------------|-----------|------------|------------------|
| 15 ----- 17 | 3 | 13,64 | $(15+17)/2 = 16$ |
| 17 ----- 19 | 7 | 31,82 | $(17+19)/2 = 18$ |
| 19 ----- 21 | 6 | 27,27 | $(19+21)/2 = 20$ |
| 21 ----- 23 | 4 | 18,18 | $(21+23)/2 = 22$ |
| 23 ----- 25 | 2 | 9,09 | $(23+25)/2 = 24$ |
| Σ | 22 | 100 | |

Fonte: Dados primários.

Agora, de posse de frequências e pontos médios de cada classe, você poderá aplicar a fórmula (4.1.). Veja na tabela 3 a obtenção dos valores (f *X) que, depois de obtidos, precisam ser somados.

Tabela 03 - Distribuição de Frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % | X | (f *X) |
|--------------------------------------|-----------|------------|----|----------------|
| 15 ----- 17 | 3 | 13,64 | 16 | $(3*16) = 48$ |
| 17 ----- 19 | 7 | 31,82 | 18 | $(7*18) = 126$ |
| 19 ----- 21 | 6 | 27,27 | 20 | $(6*20) = 120$ |
| 21 ----- 23 | 4 | 18,18 | 22 | $(4*22) = 88$ |
| 23 ----- 25 | 2 | 9,09 | 24 | $(2*24) = 48$ |
| Σ | 22 | 100 | | |

Fonte: Dados primários.

Portanto, aplicando a fórmula (4.1.) você terá o valor da média a seguir:

$$(4.2.) \quad \bar{x} = \frac{\sum (f * X)}{\sum f} = \frac{430}{22} = 19,5454$$

Note que o valor da média 19,5454 pode ser arredondado para ficar com 2 casas decimais, desprezando a terceira casa decimal que é 5, e como você já sabe, quando o valor fracionário a ser desprezado é maior ou igual que 5 aumenta-se o valor da casa decimal que fica, portanto, a média para as precipitações pluviométricas é 19,55 mm.

Para calcular a moda para dados agrupados, você precisa primeiro identificar a classe que contém a maior frequência, que será denominada de classe modal, e depois utilizar a seguinte fórmula:

$$(4.3.) \quad \hat{X} = L + \left(\frac{Da}{Da + Dp} \right) * a$$

Onde:

L representa o limite inferior da classe modal;

Da = MaiorFrequência – FrequênciaAnterior;

Dp = MaiorFrequência – FrequênciaPosterior;

a representa a amplitude de classe.

Como você já entendeu, a Moda vai estar na classe que tiver a maior frequência. Veja na tabela 4 que a classe modal, ou seja, de maior frequência é a segunda classe **17** — **19**.

Tabela 04 - Distribuição de Frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % |
|--------------------------------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

É, pois, essa classe que você vai tomar como referência para aplicar a fórmula (4.3.). A seguir você terá o cálculo da Moda, primeiro você deve obter cada parcela individual que compõe a Moda.

L = 17 é o limite inferior da classe modal;

Da = MaiorFrequência – FrequênciaAnterior = 7 – 3 = 4;

Dp = MaiorFrequência – FrequênciaPosterior = 7 -6 = 1;

a = 2 é a amplitude de classe, obtida anteriormente.

Portanto, o valor final da Moda é:

$$\hat{X} = L + \left(\frac{Da}{Da + Dp} \right) * a = 17 + \left(\frac{4}{4 + 1} \right) * 2$$

$$\hat{X} = 17 + 0,8 * 2 = 17 + 1,6 = 18,6$$

Note que foi resolvido primeiro o cálculo da fração $\left(\frac{4}{4 + 1} \right) = 0,8$

Depois foi realizada a multiplicação $2 * 0,8 = 1,6$ e finalmente a soma:

$$17 + 1,6 = 18,6$$

O valor final da Moda, 18,6, representa que há uma frequência grande de dias com precipitação em torno de 18,6mm.

Para calcular a mediana para dados agrupados, você precisa primeiro identificar a primeira classe que a soma das porcentagens contém 50%, a qual será denominada de classe mediana, e depois utilizar a seguinte fórmula:

$$(4.4.) \quad \tilde{X} = L + \left(\frac{\left(\frac{\sum f}{2} \right) - Sant}{f} \right) * a$$

L representa o limite inferior da classe mediana;

$\sum f$ representa a soma das frequências;

Sant representa a soma das frequências anteriores;

f representa a frequência da classe mediana;

a representa a amplitude de classe.

Como você já entendeu, a Mediana vai estar na primeira classe que a soma das porcentagens contenha 50%. Veja na tabela 5 que até a primeira classe há 13,64%; até a segunda há 13,64+31,82=45,46%; até a terceira há 13,64+31,82+27,27=72,73% e, portanto, contém 50%, logo, é a terceira classe de **19** — **21** a classe mediana.

Tabela 05 - Distribuição de Frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % |
|--------------------------------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

É a terceira classe que você vai tomar como referência para aplicar a fórmula (4.4.). A seguir você terá o cálculo da Mediana, primeiro você deve obter cada parcela individual que compõe a Mediana.

L = 19 é o limite inferior da classe mediana;

Σf = 22 é a soma das Frequências;

Sant = 7 + 3 = 10 é a soma das Frequências anteriores;

f = 6 é Frequência da classe mediana;

a = 2 é a amplitude de classe, obtida anteriormente.

Portanto, o valor final da Mediana é:

$$\tilde{X} = L + \left(\frac{\left(\frac{\Sigma f}{2} \right) - Sant}{f} \right) * a = 19 + \left(\frac{\left(\frac{22}{2} \right) - 10}{6} \right) * 2$$

$$\tilde{X} = 19 + 0,17 * 2 = 19 + 0,34 = 19,34$$

Note que foi resolvido primeiro o cálculo da fração:

$$\left(\frac{\left(\frac{22}{2} \right) - 10}{6} \right) = 0,16666 \cong 0,17$$

Depois foi realizada a multiplicação **0,17 * 2 = 0,34** e, finalmente, a soma **19 + 0,34 = 19,34**. O valor final da Mediana 19,34 representa que em metade (50%) dos dias choveu abaixo de 19,34mm.

Como você pôde observar essas medidas, média, moda e mediana podem ser bastante úteis para avaliação global de um conjunto de dados. Nesse exemplo em particular, sobre as chuvas em determinado município, com média de 19,55 mm, moda de 18,6mm e mediana de 19,34mm, fica bastante claro que chove pouco nesse município, pois em metade dos dias choveu menos de 19,34mm e um grande número de dias choveu apenas 18,6mm, com uma média diária de 19,55mm.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre medidas de tendência central até agora, utilize a distribuição de Frequências a seguir, referente ao número diário de casos de dengue em determinado município, calcule média, moda e mediana.

Tabela 06 - Casos diários de dengue, em determinado município.

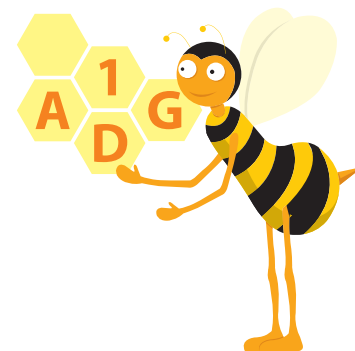
| Casos Dengue | f |
|--------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.



RESUMINDO

Nesta aula você aprendeu a importância que medidas de tendência central têm na compreensão global do fenômeno estudado. Aprendeu que a média, mediana e a moda são medidas de tendência central que permitem caracterizar o valor médio, o valor mais freqüente e o valor central, respectivamente, de um conjunto de dados; e também as aplicou em um exemplo.





LEITURAS COMPLEMENTARES

No artigo disponível no site < http://www.scielo.br/scielo.php?pid=S0034-737X2010000600004&script=sci_arttext > você encontrará análises quanto à qualidade de variedades de sementes de mamona, e observará como as medidas de tendência central média, moda e mediana foram bastante utilizadas. Você poderá comprovar, portanto, como essas medidas são bastante utilizadas para avaliação de variáveis. Quando entre elas não há uma variação, constata-se que o fenômeno está sob controle ou se diz que os dados são considerados normais.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe calcular medidas de tendência central. Considere a distribuição de Frequências a seguir, apresentada na tabela 7, referente renda mensal (em salários mínimos) de famílias de determinado município. Calcule média, moda e mediana para área Urbana ou Rural e afirme se os dados podem ser considerados Normais. Os dados são considerados normais quando os valores médio, mais freqüente e central são iguais.

Tabela 07 - Salários mínimos por área.

| RENDAS (Salários Mínimos) | ÁREA URBANA | ÁREA RURAL |
|---------------------------|-------------|------------|
| 1 ---- 8 | 3 | 2 |
| 8 ---- 15 | 7 | 9 |
| 15 ---- 22 | 2 | 5 |
| 22 ---- 29 | 4 | 2 |
| 29 ---- 36 | 1 | 1 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 05
MEDIDAS DE VARIABILIDADE

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Como já foi estudado na aula sobre Medidas de Tendência Central, há medidas em torno das quais se concentram uma grande parte dos dados, ou seja, essas medidas contêm bastante informação. Há, no entanto, uma pergunta a ser respondida: quanto de representatividade essas medidas têm? Para mensurar essa representatividade das medidas de tendência central, pode-se utilizar medidas de variabilidade que permitem identificar qual o percentual dos dados que está distante das medidas de tendência central. Por exemplo, se a média tiver uma variação de 10%, ela tem poucos dados longe dela, ou seja, tem uma representação bastante aceitável, ao contrário, por exemplo, se a média tiver uma variação de 90%, há muitos dados, a quase totalidade deles, distante da média.

Você deve estar se perguntando o que fazer para avaliar a representatividade de medidas de tendência central. Elas representam qual porcentagem dos dados? Representam 10% dos dados? Representam 90%? E o que pode se concluir diante de uma situação de alta ou baixa representatividade? Para responder a essas perguntas, você precisará aprender medidas de tendência de variabilidade. Isso é o que você vai estudar nesta aula.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender as técnicas para obter medidas de variabilidade;
- calcular variância, desvio padrão e coeficiente de variação;
- interpretar a representatividade do valor médio.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO MEDIDAS DE VARIABILIDADE

Quando são calculadas medidas de tendência central, como você aprendeu na Aula 4 – Medidas de Tendência Central, espera-se que elas possam conter muita informação, e que apenas utilizando-as seja possível representar todo o conjunto dos dados. Você recorda que a média é uma medida muito utilizada no dia-a-dia, e certamente você mesmo tem uma série de exemplos que podem evidenciar essa ampla utilização da média.

Considere o seguinte exemplo: três alunos foram avaliados com duas provas cada um, e obtiveram as notas apresentadas a seguir.

Tabela 01 - Notas dos alunos.

| ALUNO | NOTA 1 | NOTA 2 |
|------------|--------|--------|
| Aluno um | 0,0 | 10,0 |
| Aluno dois | 3,0 | 7,0 |
| Aluno três | 5,0 | 5,0 |

Fonte: Do autor.

A partir dessas notas, é preciso obter uma medida que expresse, de algo modo razoável, o rendimento acadêmico desses alunos; a média poderia ser uma medida sugerida. Então, veja a seguir os valores das médias de cada aluno.

Tabela 02 - Notas dos alunos com média.

| ALUNO | NOTA 1 | NOTA 2 | MÉDIA |
|------------|--------|--------|-------|
| Aluno um | 0,0 | 10,0 | 5,0 |
| Aluno dois | 3,0 | 7,0 | 5,0 |
| Aluno três | 5,0 | 5,0 | 5,0 |

Fonte: Do autor.

E agora? As médias são iguais, qual seria o aluno que teve melhor rendimento? São muitos os critérios, alguns subjetivos outros objetivos. Dentre os subjetivos, você pode afirmar que o ALUNO UM teve uma grande evolução. Dentre os objetivos, você pode utilizar medidas de variabilidade, por exemplo, pode-se calcular quanto as notas dos alunos estão distantes de suas médias: para o ALUNO 1, a variação poderia ser calculada como a soma de suas variações absolutas de cada nota em relação à média:

$$\text{VARIACÃO} = |\text{NOTA1-MÉDIA}| + |\text{NOTA2-MÉDIA}|$$

$$\text{VARIACÃO} = |0,0 - 5,0| + |10,0-5,0| = |-5| + |5|$$

$$\text{VARIACÃO} = 5 + 5 = 10.$$

Ou seja, o ALUNO 1 teve uma média acadêmica de 5,0 e uma variação de 10. O que significa que sua variação é muito maior, é o dobro da média. Você deve estar imaginando que essa média mostra pouca representatividade, pois está muito longe dos dados, sua variação é muito maior do que sua representatividade.

As demais médias e suas variações são apresentadas a seguir.

Tabela 03 - Notas dos alunos com variação.

| ALUNO | NOTA 1 | NOTA 2 | MÉDIA | VARIACÃO |
|------------|--------|--------|-------|----------|
| Aluno um | 0,0 | 10,0 | 5,0 | 10 |
| Aluno dois | 3,0 | 7,0 | 5,0 | 4 |
| Aluno três | 5,0 | 5,0 | 5,0 | 0 |

Fonte: Do autor.

Portanto, utilizando-se de critérios objetivos, pode-se concluir que o rendimento do AUNO TRÊS é o que pode ser melhor representado pela média. E por que? Porque para este aluno a média possui a menor variação, ela é ZERO, ou seja, a sua média está muito próxima dos dados, logo ela é uma excelente representante do rendimento do aluno.

Esse critério de representatividade da média é amplamente utilizado. As organizações o utilizam amplamente para avaliar seus processos, avaliar seus servidores, avaliar seus clientes.

Você já deve ter ouvido falar que existe uma medida chamada desvio padrão que é utilizada para avaliar o desempenho de candidatos em um vestibular; pois essa medida é justamente uma das que você vai estudar, ela avalia o quanto um candidato está próximo ou distante de seus concorrentes. Você já deve ter ouvido falar também que foram feitas pesquisas de opinião e que essas pesquisas tinham um erro de 5%; pois essa medida também é uma medida de variabilidade e expressa que a média obtida nessas pesquisas poderá variar no máximo 5%, para mais ou para menos. Com todas essas informações você já deve ter imaginado a importância das medidas de variabilidades e, por causa disso, já quer aprender como calculá-las. A partir de agora você vai aprender a calculá-las.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre medidas de variabilidade até agora, utilizando os números de multas mensais aplicadas por um órgão ambiental em determinado município, apresentados a seguir, utilize a média já obtida para esses dados na Aula 4 – Medidas de Tendência Central, e obtenha as variações absolutas de cada multa em relação à média. Você, com certeza, ao final, identificará quais as multas que estão mais próximas da média e quais estão mais distantes.

| Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 9 | 9 | 6 | 2 | 3 | 4 | 9 | 2 | 4 | 6 | 6 | 4 |



CALCULANDO VARIÂNCIA, DESVIO PADRÃO, COEFICIENTE DE VARIAÇÃO

Agora que você já sabe o que são medidas de variabilidade e sabe obtê-las para o pequeno conjunto de dados não agrupados, pois foi isto que você já compreendeu na seção anterior e já fez a atividade correspondente, você vai aprender a calcular medidas de variabilidade quando se tem muitos dados e eles estão agrupados em uma distribuição de frequências. Você vai aprender três medidas de variabilidade para dados agrupados, a variância, o desvio padrão e o coeficiente de variação; há outras medidas de variabilidade na literatura de Estatística, porém essas três são as mais amplamente utilizadas.

Todas essas três medidas avaliam a variação dos dados em relação à média. Ou seja, quanto menor for a variabilidade, mais os dados estão próximos da média e, portanto, ela poderá ser utilizada para representar todos os dados, pois eles estariam em torno dela. Caso a variabilidade seja grande, significa que os dados estão distantes da média e, portanto, ela não poderá ser utilizada para representar os dados, pois eles não estariam em torno dela.

Para calcular a **variância para dados agrupados**, você utilizará a seguinte fórmula:

$$(5.1.) \quad S^2 = \frac{\sum f * (X - \bar{X})^2}{(\sum f) - 1}$$

Onde:

f representa a frequência de cada classes;

X representa o ponto médio de cada classe;

\bar{X} representa a média.

Como você já entendeu, as frequências, os pontos médios e a média precisam estar calculados. E você já aprendeu a calculá-los.

Para calcular a variância, será utilizado o exemplo das precipitações pluviométricas, visto na Aula 4 – Medidas de Tendência Central. Neste exemplo, a média obtida para as precipitações pluviométricas foi de 19,55 mm, já os pontos médios estão na tabela 4, que contém também as frequências.

Tabela 04 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % | X |
|--------------------------------------|-----------|------------|----|
| 15 ----- 17 | 3 | 13,64 | 16 |
| 17 ----- 19 | 7 | 31,82 | 18 |
| 19 ----- 21 | 6 | 27,27 | 20 |
| 21 ----- 23 | 4 | 18,18 | 22 |
| 23 ----- 25 | 2 | 9,09 | 24 |
| Σ | 22 | 100 | |

Fonte: Dados primários.

Agora, de posse de frequências, pontos médios de cada classe e da média igual a 19,55 mm, você poderá aplicar a fórmula (5.1.). Veja na tabela 5 a obtenção dos valores $\sum f*(X - \bar{X})^2$ que, depois de obtidos, precisam ser somados.

Tabela 05 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % | X | $f*(X - \bar{X})^2$ |
|--------------------------------------|-----------|------------|----|--------------------------|
| 15 ----- 17 | 3 | 13,64 | 16 | $3*(16-19,55)^2 = 37,81$ |
| 17 ----- 19 | 7 | 31,82 | 18 | $7*(18-19,55)^2 = 16,82$ |
| 19 ----- 21 | 6 | 27,27 | 20 | $6*(20-19,55)^2 = 1,22$ |
| 21 ----- 23 | 4 | 18,18 | 22 | $4*(22-19,55)^2 = 24,01$ |
| 23 ----- 25 | 2 | 9,09 | 24 | $2*(24-19,55)^2 = 39,61$ |
| Σ | 22 | 100 | | 119,47 |

Fonte: Dados primários.

Para obter o primeiro valor, que é igual a 37,81, você com certeza fez os seguintes cálculos:

$(X - \bar{X}) = (16 - 19,55) = -3,55$, depois você elevou este valor ao quadrado, que resultou em:

$(X - \bar{X})^2 = (-3,55)^2 = 12,6025$, o qual finalmente será multiplicado pela frequência, resultando em:

$f * (X - \bar{X})^2 = 3 * 12,6025 = 37,8075$, que foi arredondado com duas casas decimais para 37,81.

Seguindo esse mesmo raciocínio você obteve todos os demais valores, cuja soma foi igual a 119,47. Portanto, aplicando a fórmula (5.1.) você terá o valor da variância a seguir:

$$(5.2.) \quad S^2 = \frac{\sum f * (X - \bar{X})^2}{(\sum f) - 1} = \frac{119,47}{22 - 1} = 5,689 \cong 5,69$$

Note que o valor da variância 5,689 pode ser arredondado para ficar com 2 casas decimais, desprezando a terceira casa decimal que é 9, e como você já sabe, quando o valor fracionário a ser desprezado é maior ou igual que 5 aumenta-se o valor da casa decimal que fica, portanto, a variância para as precipitações pluviométricas é de 5,69 mm². É isso mesmo, a unidade de medida da variância é ao quadrado, pois os valores obtidos são ao quadrado.

Para calcular o **desvio padrão para dados agrupados**, você utilizará a fórmula (5.3.), apresentada a seguir. Como você já observou, o desvio padrão é obtido a partir da variância.

$$(5.3.) \quad S = \sqrt{S^2}$$

Onde:

S^2 representa a variância.

Para calcular o desvio padrão, será utilizado o exemplo das precipitações pluviométricas, visto na Aula 4 – Medidas de Tendência Central. Você acabou de calcular a variância para esse exemplo, e o valor obtido em (5.2.) foi $S^2 = 5,69$, logo o valor do desvio padrão será:

$$(5.4.) \quad S = \sqrt{S^2} = \sqrt{5,69} = 2,385372088 \cong 2,39$$

Como você já entendeu, o valor do desvio padrão em (5.4.) foi arredondado e ficou igual a 2,39mm. É isso mesmo, o desvio padrão tem a mesma unidade de medida dos dados e tem uma grande importância na análise da representatividade da média, ele representa o erro da média: ou seja, a Média de 19,55mm está sujeita a um erro de 2,39mm, para mais ou para menos. Você inclusive pode até afirmar que a média poderá variar no intervalo de:

$$[\bar{X} - S; \bar{X} + S] = [19,55 - 2,39 ; 19,55 + 2,39] = [17,16; 21,94]$$

Este intervalo expressa uma informação muito concisa sobre a representatividade da média. Você já compreendeu que ela está em torno dos valores de 17,16mm até 21,94mm; os demais valores podem ser considerados distantes dela.

Para calcular o coeficiente de variação para dados agrupados, você utilizará a fórmula (5.5.), apresentada a seguir. Como você já observou, o coeficiente de variação é obtido a partir do desvio padrão e da média.

$$(5.5.) \quad Cv = \left(\frac{S}{\bar{X}} \right) * 100$$

Onde:

S representa o desvio padrão;

\bar{X} representa a média.

Para calcular o coeficiente de variação, será utilizado o exemplo das precipitações pluviométricas visto na Aula 4 – Medidas de Tendência Central. Você acabou de calcular o desvio padrão para esse exemplo e o valor obtido em (5.4.) foi $S = 2,39$, e já havia obtido o valor da média $\bar{X} = 19,55$, logo, o valor do coeficiente de variação será

$$(5.6.) \quad C_v = \left(\frac{S}{\bar{X}} \right) * 100 = \left(\frac{2,39}{19,55} \right) * 100 = 12,225 \cong 12,23\%$$

Como você já entendeu, o valor do coeficiente de variação é expresso em percentual e, nesse caso, em (5.6.) foi arredondado e ficou igual a 12,23%. É isso mesmo, esse coeficiente de variação expressa que 12,23% dos dados variam em relação à média, ou seja, podem ser considerados distantes dela; ao passo que 87,77% (100-12,23) dos dados podem ser considerados próximos da média.

Você inclusive pode até afirmar que a média poderá representar 87,77% das precipitações estudadas. E esse valor permite que você conclua, com bastante segurança, que essa média é uma boa representação de todas as precipitações pluviométricas.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre medidas de variabilidade até agora, utilize a distribuição de frequências das tabelas a seguir, referente ao número diário de casos de dengue em determinado município, utilize a média já obtida para esses dados na Aula 4 – Medidas de Tendência Central, e calcule variância, desvio padrão e coeficiente de variação.



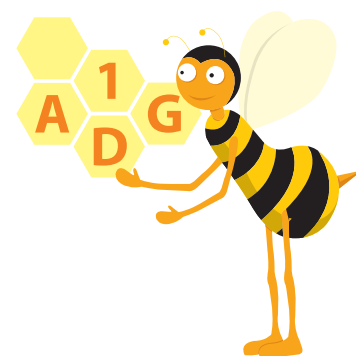
Tabela 06 - Casos diários de dengue em determinado município.

| Casos Dengue | f |
|--------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.

RESUMINDO

Nesta aula você aprendeu a importância que medidas de variabilidade têm na avaliação da representatividade da média, e aplicou-as em um exemplo. Aprendeu também que quando a variabilidade é pequena, significa que ela está próxima dos dados e, portanto, é uma boa representação para os mesmos.



LEITURAS COMPLEMENTARES

No artigo disponível no site http://www.abep.nepo.unicamp.br/site_eventos_alap/PDF/ALAP2004_388.PDF você encontrará análises quanto a alterações sócio-ambientais ocorridas na comunidade pesqueira Canto do Mangue, no município de Canguaretama-RN, e observará como medidas de variabilidade foram utilizadas para realizar análise quantitativa da situação. Você poderá comprovar, também, como essas medidas são bastante utilizadas para avaliação de variáveis, em particular no caso da avaliação do comportamento da produção de caranguejo.





AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe calcular medidas de variabilidade, observe a distribuição de frequências a seguir, apresentada na tabela 4, referente à renda mensal (em salários mínimos) de famílias de determinado município. Utilizando a média, já obtida na Aula 4 – Medidas de Tendência Central, calcule a variância, o desvio padrão e o coeficiente de variação para área Urbana ou Rural e afirme se os dados podem ser representados pela média.

Tabela 04 - Salários mínimos por área.

| RENDAS (Salários Mínimos) | ÁREA URBANA | ÁREA RURAL |
|---------------------------|-------------|------------|
| 1 ---- 8 | 3 | 2 |
| 8 ---- 15 | 7 | 9 |
| 15 ---- 22 | 2 | 5 |
| 22 ---- 29 | 4 | 2 |
| 29 ---- 36 | 1 | 1 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 06
MEDIDAS SEPARATRIZES

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Em muitas situações da vida cotidiana, é necessário organizar um grupo de dados para permitir que dentro desses grupos haja certa similaridade entre eles. Por exemplo, você já deve ter participado ou lido resultados de pesquisas de avaliação, nas quais havia conceitos tipo ÓTIMO, BOM, REGULAR, RUIM, PÉSSIMO; esses conceitos são muito importantes, porém pode ser necessário ter valores que indiquem quando a variável é ÓTIMA, será que ela é ÓTIMA quando seu valor estiver entre 9,0 e 10,0? Ela seria BOA quando seu valor variasse entre 7,0 e 9,0? As separatrizes são medidas que permitem organizar os dados dessa forma, utilizando-se de valores numéricos. São elas que você vai começar a estudar a partir de agora.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender as técnicas para obter medidas separatrizes;
- calcular quartis, decis, percentis;
- organizar os dados em grupos similares.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO MEDIDAS SEPARATRIZES

Você já deve estar querendo compreender melhor qual a utilidade de medidas separatrizes. Considere que você tem notas de alunos da disciplina de Estatística e está querendo organizar essas notas em quatro grupos, conforme o Quadro 1, a seguir.

Quadro 1 - Grupos de Separatrizes.

| 1º Grupo- 25% | 2º Grupo- 50% | 3º Grupo- 75% | 4º Grupo- 100% |
|--------------------|--|--|--------------------|
| 25% menores Notas. | 25% Notas entre as menores notas e a nota central. | 25% Notas entre a nota central e as maiores notas. | 25% maiores Notas. |

Fonte: Do autor.

Conforme você pode verificar no Quadro 1, o 1º grupo contém as 25% menores notas; o 2º grupo contém as 25% notas que estão entre as menores e a nota central, representada por 50%, que podem ser consideradas notas regulares; o 3º grupo contém as 25% notas que estão entre a nota central e as melhores notas; já o 4º contém as 25% maiores notas.

Deste modo, você poderia organizar melhor as notas, utilizando-se de critérios mais objetivos, pois o elemento que está permitindo a organização não é uma categoria, é um valor. Esse critério das separatrizes é bastante útil quando se tem conjuntos de valores numéricos e tem-se a necessidade de agrupá-los.

Mas você deve estar se perguntando: e os grupos só podem ser quatro? Na realidade, você pode criar seus próprios grupos, e você terá oportunidade de aprender como fazer essa definição mais adiante ainda nesta aula.

Você deve ter notado que no exemplo do Quadro 1, há quatro grupos com 25% de dados em cada grupo, de modo que eles conjuntamente totalizam 100%. E é justamente esse o critério para compor grupos: é preciso que eles juntos totalizem 100%. Portanto, você poderá formar dez grupos de 10%; cinco grupos de 20%; não poderá formar grupos com 15%, pois eles juntos não totalizam 100%.

Esta condição de que os grupos juntos totalizem 100% é muito importante e você deve tê-la sempre em mente. Aliás, compreendendo essa condição, você poderá criar suas próprias Separatrizes de acordo com as peculiaridades de cada situação.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Separatrizes até agora, utilizando os números de multas mensais aplicadas por um órgão ambiental em determinado município, apresentados a seguir, utilize algum critério pessoal e organize os dados em cinco grupos: o 1º Grupo conterà as menores multas, o 5º Grupo conterà as maiores multas; e entre eles deve haver os demais grupos, o 2º, o 3º e o 4º. Você, com certeza, ao final, identificará que para um número maior de dados, a dificuldade de identificar as Separatrizes é muito grande.



| Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 9 | 9 | 6 | 2 | 3 | 4 | 9 | 2 | 4 | 6 | 6 | 4 |

CALCULANDO QUARTIS, DECIS, PERCENTIS

Agora que você já compreende o que são separatrizes, e já percebe que para um conjunto muito grande de dados você precisará de uma técnica mais objetiva para calculá-las, deve estar querendo utilizar essas técnicas quando os dados estiverem agrupados em uma distribuição de frequências. Como você já sabe, os Quartis são separatrizes que organizam os dados em quatro grupos, conforme Quadro 2.

Quadro 2 - Grupos dos Quartis Q_1 , Q_2 , Q_3 .

| Q_1 | Q_2 | Q_3 | |
|--------------------------|--|--|---------------------------|
| 1º Grupo- 25% | 2º Grupo- 50% | 3º Grupo- 75% | 4º Grupo- 100% |
| 25% menores Notas. | 25% Notas entre as menores notas e a nota central. | 25% Notas entre a nota central e as maiores notas. | 25% maiores Notas. |

Fonte: Do autor.

Conforme o Quadro 2, o Quartil Q_1 é o valor que abaixo dele há os 25% menores valores; o Q_2 é o valor abaixo do qual há 50% dos valores; e o Q_3 é o valor que abaixo dele há os 75% maiores valores.

Para calcular **Quartis para dados agrupados**, você utilizará a seguinte fórmula:

$$(6.1.) \quad Q_i = L + \left(\frac{i * \left(\frac{\sum f}{4} \right) - Sant}{f} \right) * a$$

Onde:

L representa o limite inferior da classe do Quartil;

i representa o índice do Quartil – 1, 2 ou 3;

$\sum f$ representa a soma das frequências;

Sant representa a soma das frequências anteriores;

f representa a frequência da classe do Quartil;

a representa a amplitude de classe.

Para cada Quartil, você precisa primeiro identificar qual a primeira classe que contém a soma das porcentagens correspondente ao Quartil. Ou seja, para obter o Q_1 você precisa primeiro identificar qual a primeira classe cuja soma das porcentagens contém 25%; para o Q_2 , a primeira classe que a soma das porcentagens contém 50%; e para o Q_3 , a primeira classe que a soma das porcentagens contém 75%. A partir de agora, você vai calcular Q_1 , Q_2 e Q_3 para o exemplo das precipitações pluviométricas, cuja distribuição de frequências está apresentada na Tabela 1.

O Quartil Q_1 , como você já entendeu, vai estar na primeira classe cuja soma das porcentagens contenha 25%. Veja na Tabela 1 que até a primeira classe há 13,64%; até a segunda há $13,64+31,82=45,46\%$ e, portanto, contém 25%, logo, é a segunda classe de **17** — **19** a classe do Q_1 .

Tabela 1 - Distribuição de frequências de precipitações pluviométricas (em mm)

| Precipitações | f | % |
|----------------------------|-----------|--------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

É a segunda classe que você vai tomar como referência para aplicar a fórmula (6.1.). A seguir você terá o cálculo do Q_1 , primeiro você deve obter cada parcela individual que compõe o Q_1 .

L = 17 é o limite inferior da classe do Q_1 ;

i = 1 é o índice do Q_1 ;

$\Sigma f = 22$ é a soma das frequências;

Sant = 3 é a soma das frequências anteriores;

f = 7 é frequência da classe do Q_1 ;

a = 2 é a amplitude de classe, obtida anteriormente.

Portanto, o valor final do Q_1 é:

$$Q_1 = L + \left(\frac{i * \left(\frac{\sum f}{4} \right) - Sant}{f} \right) * a = 17 + \left(\frac{1 * \left(\frac{22}{4} \right) - 3}{7} \right) * 2$$

$$Q_1 = 17 + 0,36 * 2 = 17 + 0,72 = 17,72$$

Note que foi resolvido primeiro o cálculo da fração:

$$\left(\frac{1 * \left(\frac{22}{4} \right) - 3}{7} \right) = 0,3571 \cong 0,36$$

Depois foi realizada a multiplicação $0,36 * 2 = 0,72$ e, finalmente, a soma $17 + 0,72 = 17,72$. O valor final do Q_1 é 17,72 mm e representa que há 25% de dias com chuvas abaixo de 17,72 mm.

O Quartil Q_2 , como você já entendeu, vai estar na primeira classe cuja soma das porcentagens contenha 50%. Veja na Tabela 2 que até a primeira classe há 13,64%; até a segunda há $13,64+31,82=45,46\%$; até a terceira há $13,64+31,82+27,27=72,73\%$ e, portanto, contém 50%, logo, é a terceira classe de **19** — **21** a classe do Q_2 .

Tabela 2 - Distribuição de frequências de precipitações pluviométricas (em mm)

| Precipitações | f | % |
|----------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

É a terceira classe que você vai tomar como referência para aplicar a fórmula (6.1.). A seguir você terá o cálculo do Q_2 , primeiro você deve obter cada parcela individual que compõe o Q_2 .

$L = 19$ é o limite inferior da classe do Q_2 ;

$i = 2$ é o índice do Q_2 ;

$\Sigma f = 22$ é a soma das frequências;

$Sant = 3+7 = 10$ é a soma das frequências anteriores;

$f = 6$ é frequência da classe do Q_2 ;

$a = 2$ é a amplitude de classe, obtida anteriormente.

Portanto, o valor final do Q_2 é:

$$Q_2 = L + \left(\frac{i * \left(\frac{\Sigma f}{4} \right) - Sant}{f} \right) * a = 19 + \left(\frac{2 * \left(\frac{22}{4} \right) - 10}{6} \right) * 2$$

$$Q_2 = 19 + 0,17 * 2 = 19 + 0,34 = 19,34$$

Tabela 3 - Distribuição de frequências de precipitações pluviométricas (em mm)

| Precipitações | f | % |
|----------------------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

O valor final do Q_2 é 19,34 mm e representa que há 50% dos dias com chuvas abaixo de 19,34 mm.

O Quartil Q_3 , como você já entendeu, vai estar na primeira classe cuja soma das porcentagens contenha 75%. Veja na Tabela 3 que até a primeira classe há 13,64%; até a segunda há 13,64+31,82=45,46%; até a terceira há 13,64+31,82+27,27=72,73%; até a quarta classe há 13,64+31,82+27,27+18,19=90,91% e, portanto, contém 75%, logo, é a quarta classe de **21 |----- 23** a classe do Q_3 .

$$Q_3 = L + \left(\frac{i * \left(\frac{\Sigma f}{4} \right) - Sant}{f} \right) * a = 21 + \left(\frac{3 * \left(\frac{22}{4} \right) - 16}{4} \right) * 2$$

$$Q_3 = 21 + 0,125 * 2 = 21 + 0,25 = 21,25$$

O valor final do Q_3 é 21,25 mm e representa que há 75% de dias com chuvas abaixo de 21,25 mm.

Finalmente, no Quadro 3, você já tem os Quartis Q_1 , Q_2 e Q_3 , que organizam as precipitações pluviométricas em quatro grupos.

Quadro 3 - Quartis Q_1 , Q_2 , Q_3 para as precipitações pluviométricas.

| | 17,72 | 19,32 | 21,25 | |
|--|---|---|---|---|
| | Chuvas Muito Baixas | Chuvas Baixas | Chuvas Altas | Chuvas Muito Altas |
| | 1º Grupo - 25% dos valores estão abaixo de 17,72mm. | 2º Grupo - 50% dos valores estão abaixo de 19,34mm. | 3º Grupo - 75% dos valores estão abaixo de 21,25mm. | 4º Grupo - 25% dos valores estão abaixo de 21,25mm. |

Finalmente, a partir do Quadro 3, você tem quatro grupos formados com critérios objetivos utilizando-se os valores das separatrizes: o primeiro pode ser denominado de Chuvas Muito Baixas, o segundo de Chuvas Baixas, o terceiro de Chuvas Altas e o último de Chuvas Muito Altas.

Em relação aos Decis, como você já sabe, eles são separatrizes que organizam os dados em dez grupos de 10% em 10%, conforme Quadro 4.

Quadro 4 - Grupos dos Decis D_1 , D_2 , D_3 , D_4 , D_5 , D_6 , D_7 , D_8 , D_9 .

| D_1 | D_2 | D_3 | D_4 | D_5 | D_6 | D_7 | D_8 | D_9 | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |

Fonte: Do autor.

Conforme o Quadro 4 o Decil:

- D_1 é o valor abaixo do qual há 10% dos valores;
- D_2 é o valor abaixo do qual há 20% dos valores;
- D_3 é o valor abaixo do qual há 30% dos valores;
- D_4 é o valor abaixo do qual há 40% dos valores;
- D_5 é o valor abaixo do qual há 50% dos valores;
- D_6 é o valor abaixo do qual há 60% dos valores;
- D_7 é o valor abaixo do qual há 70% dos valores;
- D_8 é o valor abaixo do qual há 80% dos valores;
- D_9 é o valor abaixo do qual há 90% dos valores;

Para calcular Decis para dados agrupados, você utilizará a seguinte fórmula:

$$(6.2.) \quad D_i = L + \left(\frac{i * \left(\frac{\sum f}{10} \right) - Sant}{f} \right) * a$$

L representa o limite inferior da classe do Decil;

i representa o índice do Decil – 1,2,4,5,6,7,8, ou 9;

$\sum f$ representa a soma das frequências;

Sant representa a soma das frequências anteriores;

f representa a frequência da classe do Decil;

a representa a amplitude de classe.

Para cada Decil, você precisa primeiro identificar qual a primeira classe que contém a soma das porcentagens correspondente ao Decil.

Em relação aos **Percentis**, como você já sabe, eles são separatrizes que organizam os dados em cem grupos, de 1% em 1%.

Para calcular **Percentis para dados agrupados**, você utilizará a seguinte fórmula:

$$(6.3.) \quad P_i = L + \left(\frac{i * \left(\frac{\sum f}{100} \right) - Sant}{f} \right) * a$$

L representa o limite inferior da classe do Percentil;

i representa o índice do Percentil – 1,2,397,98, ou 99;

$\sum f$ representa a soma das frequências;

Sant representa a soma das frequências anteriores;

f representa a frequência da classe do Percentil;

a representa a amplitude de classe.

Para cada Percentil, você precisa primeiro identificar qual a primeira classe que contém a soma das porcentagens correspondente ao Percentil. Por exemplo, para obter o Percentil P_{38} , você terá que primeiro identificar qual a primeira classe que contém a soma das porcentagens correspondente 38%. E assim sucessivamente para obter o Percentil desejado.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Separatrizes até agora, utilize a distribuição de frequências a seguir, referente ao número diário de casos de dengue em determinado município, organize a distribuição em quatro grupos calculando os Quartis Q_1 , Q_2 e Q_3 .

Tabela 03 - Casos diários de dengue, em determinado município.

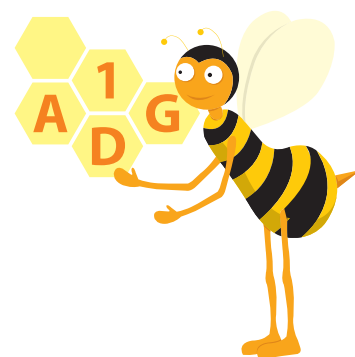
| Casos Dengue | f |
|---------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.



RESUMINDO

Nesta aula você aprendeu a importância que medidas separatrizes têm para definir os grupos de dados que organizem a distribuição. Aprendeu que podem ser definidos grupos de 25%, de 10% e de 1%; além disso você aprendeu que pode criar seus próprios grupos, desde que eles conjuntamente somem 100%.





LEITURAS COMPLEMENTARES

No artigo disponível no site <http://www.dieese.org.br/dieese/esp/cju/anote29.pdf> você encontrará análises quanto às diferenças existentes na distribuição de renda e na situação das famílias brasileiras e ratificam, a partir da utilização das separatrizes Q_1 , Q_2 e Q_3 , a conhecida situação de disparidade existente no Brasil. Você poderá chegar a várias conclusões, por exemplo, concluirá utilizando o $Q_1=84,71$ de Recife, que 25% da população tem renda familiar abaixo de R\$ 84,71; enquanto para São Paulo, que teve $Q_1=179,29$, que 25% da população tem renda familiar abaixo de R\$ 179,29.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe calcular Separatrizes, observe a distribuição de frequências a seguir, apresentada na tabela 4, referente à renda mensal (em salários mínimos) de famílias de determinado município. Calcule os Percentis P_{25} , P_{50} e P_{75} para área Urbana ou Rural e afirme o que eles representam.

Tabela 04 - Salários mínimos por área.

| RENDAS (Salários Mínimos) | ÁREA URBANA | ÁREA RURAL |
|---------------------------|-------------|------------|
| 1 ---- 8 | 3 | 2 |
| 8 ---- 15 | 7 | 9 |
| 15 ---- 22 | 2 | 5 |
| 22 ---- 29 | 4 | 2 |
| 29 ---- 36 | 1 | 1 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 07
APRESENTAÇÃO GRÁFICA

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Você já estudou bastante sobre a Estatística e suas técnicas para descrever fenômenos, desde a construção da distribuição de frequências, passando pelas medidas de tendência central, pelas medidas de variabilidade até as medidas separatrizes.

Foi um caminho com muitos cálculos e com os correspondentes comentários sobre o significado dos resultados obtidos. Mas você precisa também aprender a representar resultados de forma visualmente apresentável, de modo que seja possível compreender aspectos que nem sempre são traduzidos pelos cálculos.

Por exemplo, muitas vezes há uma tendência nos dados, e os cálculos pontuais, como média, moda e mediana, não são capazes de identificar; podem ocorrer também variações bastante sutis nos dados, porém importantes na interpretação global do fenômeno, e medidas de variabilidade como variância, desvio-padrão e coeficiente de variação não são capazes de detectar. Nesses casos, você precisará do auxílio de gráficos para complementar sua análise estatística. Nesta aula você construirá gráficos e perceberá como eles são úteis para fundamentar melhor suas análises, em complemento aos diversos cálculos de medidas estatísticas.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender a importância das apresentações gráficas;
- identificar e construir diversos gráficos;
- construir histograma e polígono de frequência, associados a uma distribuição de frequência.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO APRESENTAÇÃO GRÁFICA

Você já aprendeu a apresentar dados através de tabelas, as quais têm um nível de detalhamento de informações bastante razoável. No entanto, nas mais diversas áreas de conhecimento, é necessário expandir as informações apresentadas de modo que seja possível ter uma compreensão geral do fenômeno estudado, do seu comportamento, de sua tendência, dentre outros aspectos.

Também é muito importante ter em mente que a apresentação gráfica é um produto visual, que deve ressaltar clareza, organização entre outros aspectos que trazem atrativos visuais capazes de garantir uma leitura adequada.

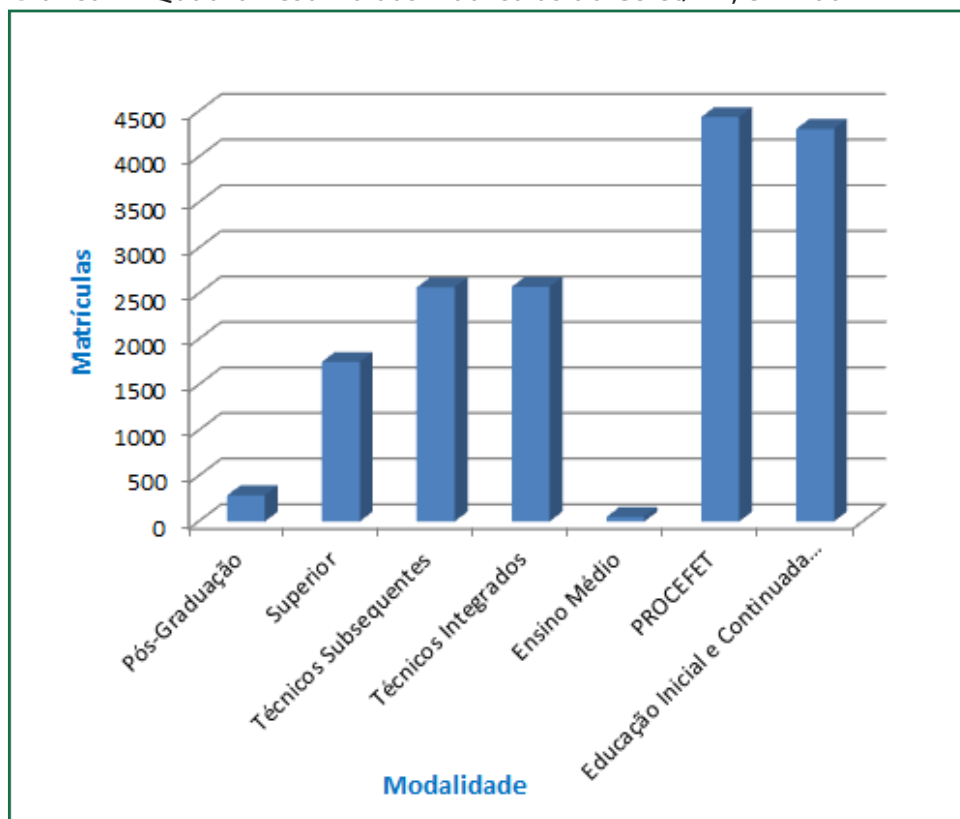
Você deve estar pensando, então, que a escolha do tipo do gráfico também é importante para garantir sua compreensão. É isso mesmo, há gráficos mais adequados para um conjunto pequeno de dados, outros mais adequados para um grupo maior de dados; há gráficos mais adequados para uma única variável, outros para mais de uma variável. E tudo isso, de fato, influi no resultado final e na compreensão do gráfico.

Há também um aspecto menos visível do gráfico, mas que é muito importante, que é a sua escala. Ela é o valor entre os números que são marcados no gráfico, ou seja, é o espaçamento entre os valores, por exemplo você pode criar escala com espaçamento de 5 em 5, ou de 10 em 10, ou com outro valor de sua escolha; o critério é que a escala sempre possui o mesmo valor, ou sempre 5, ou sempre 10, ou outro valor. A escala exerce um papel fundamental no grau de clareza dos dados: uma escala pequena vai aproximar os dados, uma escala longa vai distanciar os dados, e você deve defini-la pensando nisso. A escala horizontal é da esquerda para a direita de quem lê; a escala vertical é de cima para baixo.

Na representação de um gráfico, você pode considerar os mesmos aspectos que você já aprendeu para tabelas, quando estudou a **AULA 2 - Estatística Descritiva: séries estatísticas**. Os critérios para a elaboração de tabelas, que foram vistos anteriormente, são: cabeçalho, corpo e rodapé.

No cabeçalho devem constar informações que permitam ao leitor identificar qual é a variável de interesse, bem como o período e o local a que ela se refere. Considere a gráfico 1 a seguir, contendo o número de matrículas no CEFET/RN em 2007.

Gráfico 1 - Quadro Resumo das Matrículas do Cefet/RN, em 2007



Fonte: Relatório de Gestão 2007

Como você já deve ter identificado no gráfico 1, o cabeçalho é composto pela descrição "QUADRO RESUMO DE MATRÍCULAS DO CEFET/RN, EM 2007".

No corpo deve ser representado o gráfico propriamente, neste caso com dois eixos: o das modalidades e o das matrículas. Note que no eixo das modalidades não há escala porque a variável é qualitativa, pois seus valores são atributos e não números.

Já no eixo das matrículas, a escala utilizada foi 500, ou seja, o espaçamento entre os valores foi de 500 em 500.

Já o rodapé deve conter informações sobre as fontes dos dados. Pode-se utilizar também o rodapé para apresentar legendas ou explicitar situações especiais, como, por exemplo, informar que determinados dados não puderam ser obtidos e o motivo.

Para o gráfico 1, o rodapé é a fonte de onde foram obtidos os dados, neste caso a fonte é o Relatório de Gestão de 2007.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Apresentação Gráfica até agora, identifique gráficos com informações ligadas à Gestão Ambiental e conclua se eles são compreensíveis ou não; você certamente vai perceber a importância da escala para essa compreensão.



CONSTRUINDO GRÁFICOS

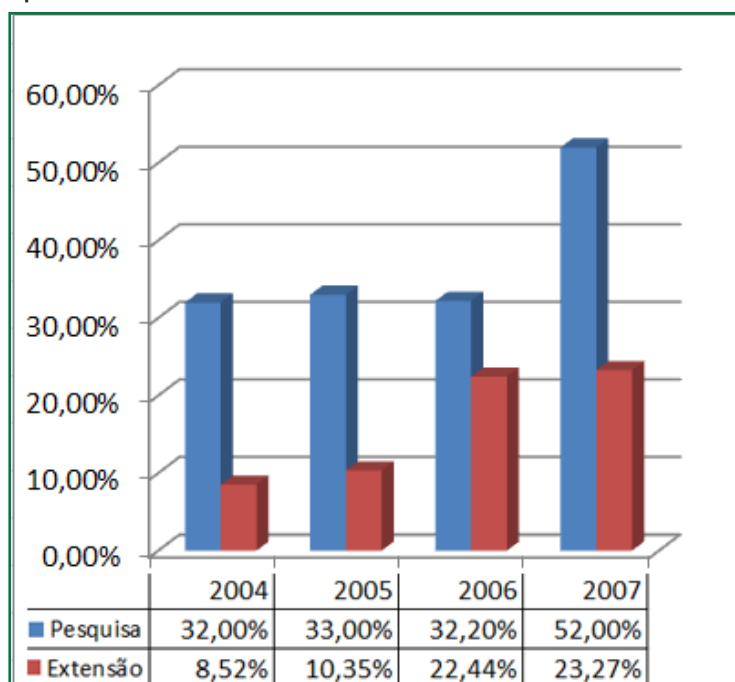
Há uma grande variedade de gráficos, aqui serão apresentados aqueles que mais frequentemente você precisará utilizar, são eles: gráfico de colunas, gráfico de barras, gráfico de setores, gráfico de linha, histograma e polígono de frequências. Com o conhecimento desses gráficos, você poderá aprender sobre outros sem dificuldade.

Gráfico de colunas

É um gráfico composto por dois eixos, um com variáveis qualitativas e outro com a frequência/porcentagem de ocorrência da variável qualitativa; interligando esses eixos há retângulos verticais, que são as colunas. O gráfico 1, que você já observou anteriormente, é um gráfico de colunas, onde a variável qualitativa é modalidade de ensino, a frequência neste caso é o número de matrículas para cada modalidade e as colunas interligam os eixos.

Em um gráfico de colunas é possível ter mais de uma variável qualitativa. Veja o gráfico 2, nele você pode observar que são apresentadas duas variáveis: o grau de participação docente em atividades de pesquisa e o grau de participação docente em atividades de extensão, ambas no CEFET/RN.

Gráfico 2 – Grau de envolvimento de Docentes com Pesquisa e Extensão



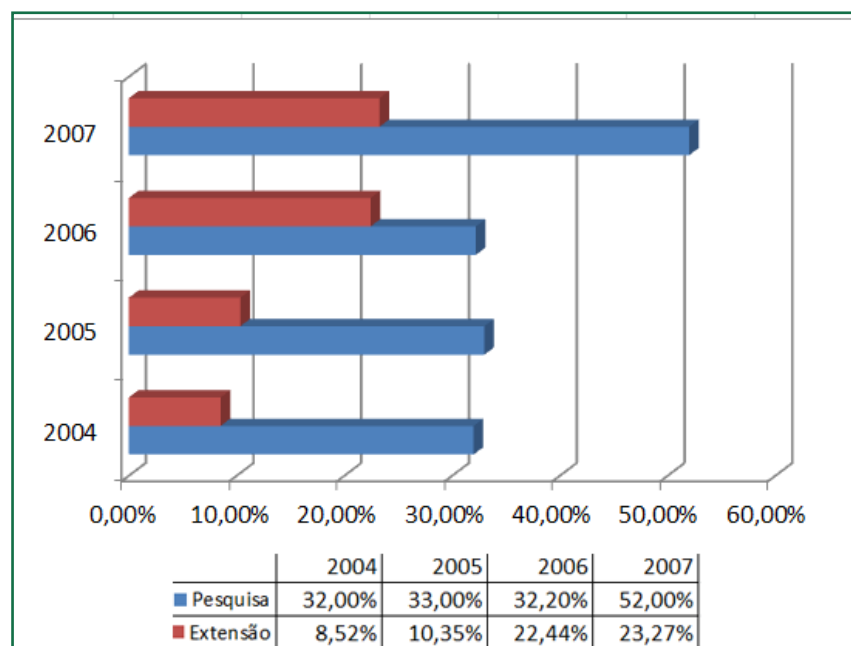
Ainda em relação ao gráfico 2, você pode observar que as informações foram obtidas junto ao Relatório de Gestão de 2007 do CEFET/RN. Você também pode verificar que a escala do eixo de porcentagem é de 10% em 10%; e no rodapé observa-se uma legenda para Pesquisa outra para Extensão e o conjunto desses números, por ano.

Gráfico de barras

É um gráfico composto por dois eixos, um com variáveis qualitativas e outro com a frequência/porcentagem de ocorrência da variável qualitativa; interligando esses eixos há retângulos horizontais, que são as barras. Ou seja, esse gráfico difere do de colunas apenas pelo sentido da barra que é horizontal e da coluna que é vertical.

Veja o gráfico 3, nele você pode observar que são apresentadas as mesmas variáveis, o grau de participação docente em atividades de pesquisa e o grau de participação docente em atividades de extensão, ambas no CEFET/RN.

Gráfico 3 – Grau de envolvimento de Docentes com Pesquisa e Extensão



Fonte: Relatório de Gestão 2007

Ainda em relação ao gráfico 3, você pode observar que as informações foram obtidas junto ao Relatório de Gestão de 2007 do CEFET/RN. Você também pode verificar que a escala do eixo de porcentagem é de 10% em 10%; e no rodapé observa-se uma legenda para Pesquisa outra para Extensão e o conjunto desses números, por ano.

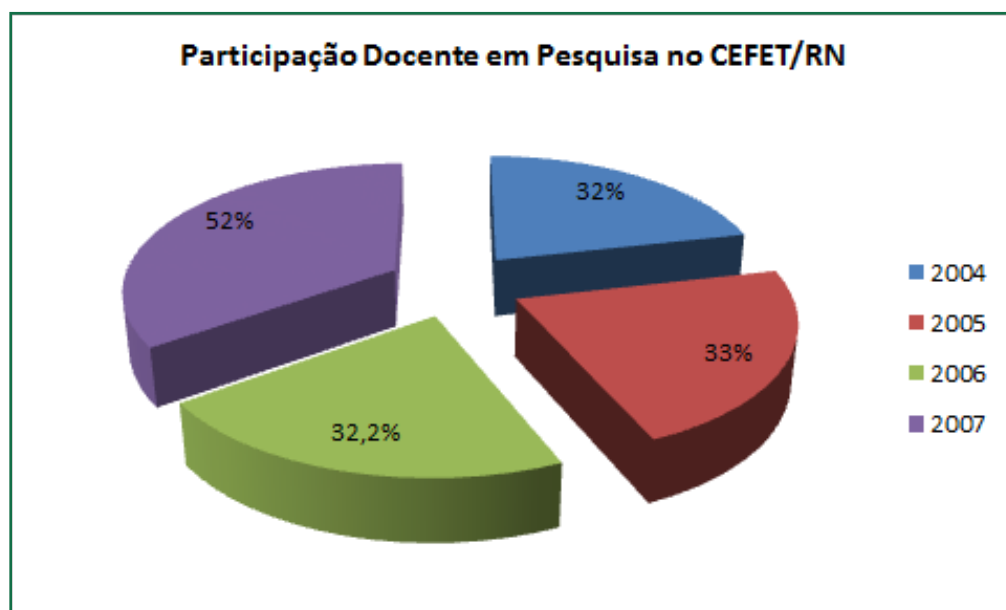
Gráfico de setores

O gráfico de setores é utilizado para comparar partes que compõem um todo. Geralmente é utilizado para representar uma variável, explorando o quanto cada uma de suas partes contribuem para o total.

Do ponto de vista físico, o gráfico de setores é um gráfico circular com 360°, e cada uma de suas partes será uma proporção correspondente, que somadas totalizaram 360°.

Por exemplo, no caso do gráfico 2, onde temos duas variáveis, uma de Participação Docente em Pesquisa e outra de Participação em Extensão no CEFET/RN, para representá-las em gráfico de setor seria melhor utilizar dois gráficos de setor um para cada participação docente.

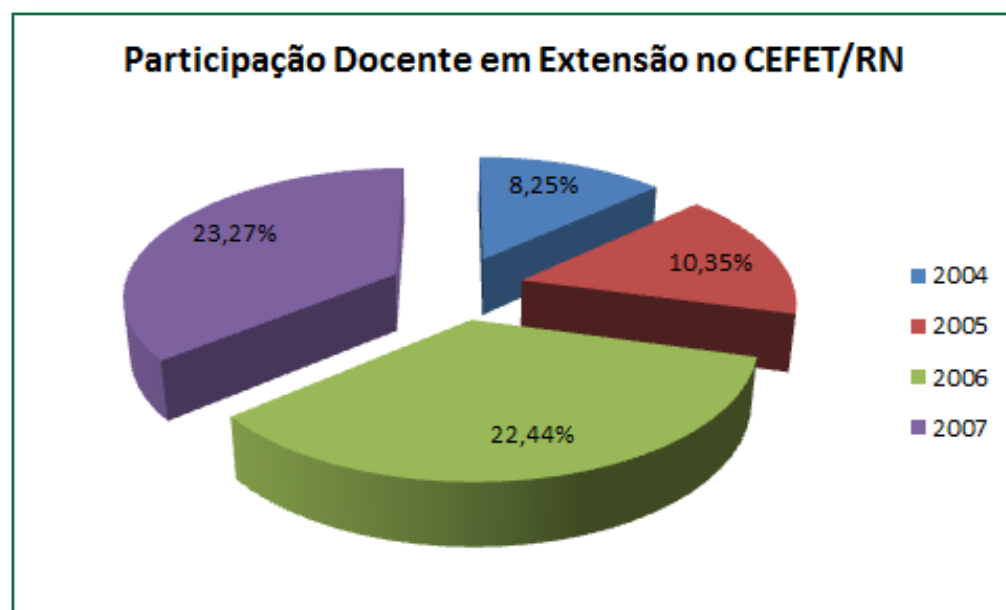
Gráfico 4 – Grau de envolvimento de Docentes com Pesquisa



Fonte: Relatório de Gestão 2007

Veja que no gráfico 4 há informações que permitem compreender que em 2007 houve o maior valor de participação docente em pesquisa para o período 2004 a 2007, que foi de 52%.

Gráfico 5 – Grau de envolvimento de Docentes com Extensão



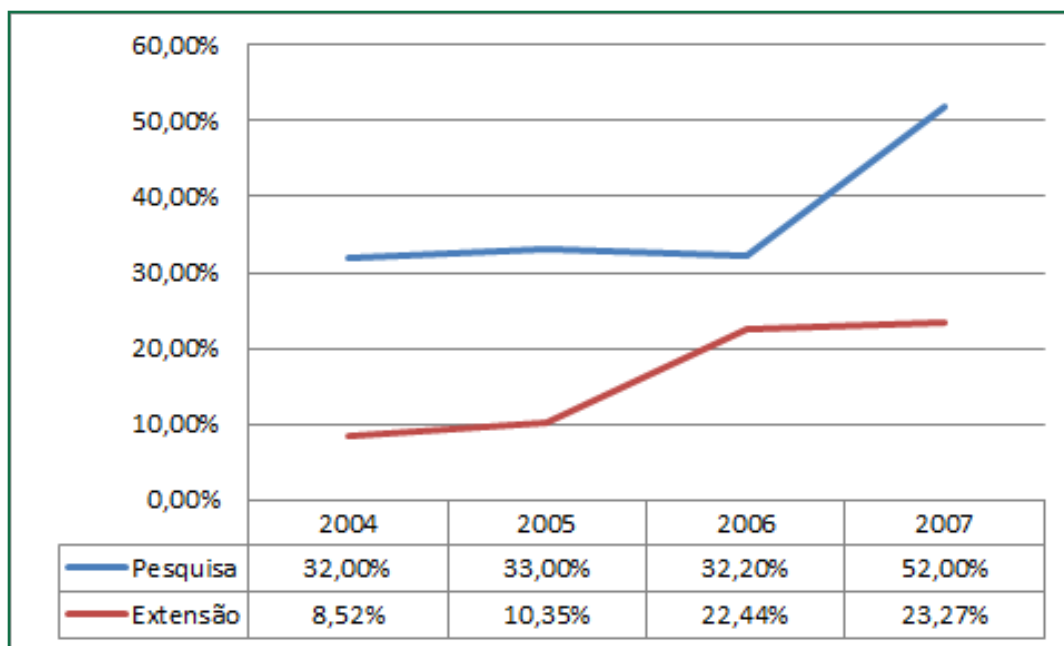
Fonte: Relatório de Gestão 2007

Veja que no gráfico 5 há informações que permitem compreender que em 2006 e 2007 houve grande participação de docentes em extensão, em relação ao período 2004 a 2007: em 2006 a participação foi de 22,44% e em 2007 foi de 23,27%.

Gráfico de linha

É um gráfico composto por dois eixos, eixo dos X e eixo dos Y, que são interligados por linhas que ligam os pontos de cada par (X,Y). Em um gráfico de linhas é possível representar mais de uma variável. Veja o gráfico 6, nele você pode observar que são apresentadas duas variáveis: o grau de participação Docente em atividades de Pesquisa e o grau de participação Docente em atividades de Extensão, ambas no CEFET/RN.

Gráfico 6 – Grau de envolvimento de Docentes com Pesquisa e Extensão.



Fonte: Relatório de Gestão 2007

Ainda em relação ao gráfico 6, você pode observar que as informações foram obtidas junto ao Relatório de Gestão de 2007 do CEFET/RN. Você também pode verificar que a escala do eixo de porcentagem é de 10% em 10%; e no rodapé observa-se uma legenda para Pesquisa outra para Extensão e o conjunto desses números, por ano.

A literatura recomenda que o gráfico de linha seja utilizado mais predominantemente em séries temporais ou históricas, aquelas que possuam muitos valores, pois só com um grande conjunto de dados sua representatividade, visual inclusive, é melhor.

Histograma

Já o Histograma, que é um gráfico de coluna, tem sua maior aplicação para representar distribuições de frequências. Nos seus dois eixos, marcam-se as classes no eixo X e as porcentagens no eixo Y, interligando-as com colunas.

Considere a distribuição de frequências das precipitações pluviométricas, já obtida anteriormente, na tabela 1.

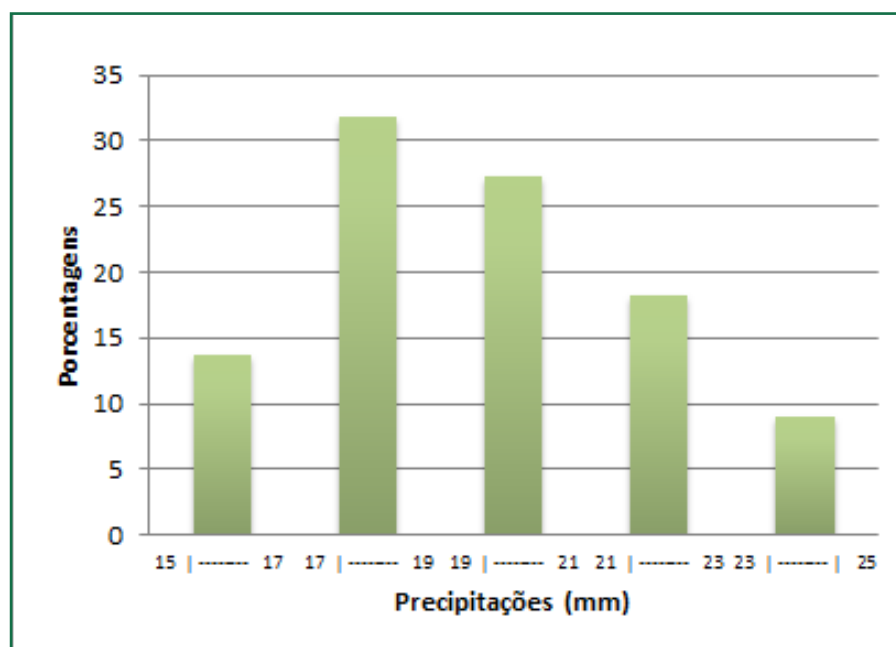
Tabela 1 - Distribuição de frequências de precipitações pluviométricas (em mm)

| Precipitações | f | % |
|----------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

Para você construir um histograma para a tabela 1, terá que marcar nos dois eixos escalas a seu critério. Por exemplo, para o eixo X, onde vão ficar as classes com as precipitações, pode-se usar a escala já da distribuição que é de 2 em 2, pois de 15 para 17 a escala é 2, de 17 para 19 é 2, e assim para todas as classes; já para o eixo Y, onde vão ficar as porcentagens, pode-se usar a escala de 5% em 5%, começando em 5%, passando para 10%, para 15%, para 20%, para 25%, para 30% e para 25%. Veja o gráfico 7.

Gráfico 7 – Distribuição de precipitações pluviométricas (em mm)



Fonte: Dados primários.

Veja que a escala do eixo X é 2, que é diferente da escala 5% do eixo Y. As escalas precisam ser iguais dentro do eixo. Lembra-lhe também que qualquer outra escala seria válida, desde que fosse sempre a mesma dentro do eixo.

Polígono de frequências

O Polígono de frequência, que é um gráfico de linha, tem sua maior aplicação para representar distribuições de frequências. Nos seus dois eixos, marcam-se pontos médios no eixo X e as porcentagens no eixo Y, interligando-as com linhas.

Considere a distribuição de frequências das precipitações pluviométricas, já obtida anteriormente, na tabela 2.

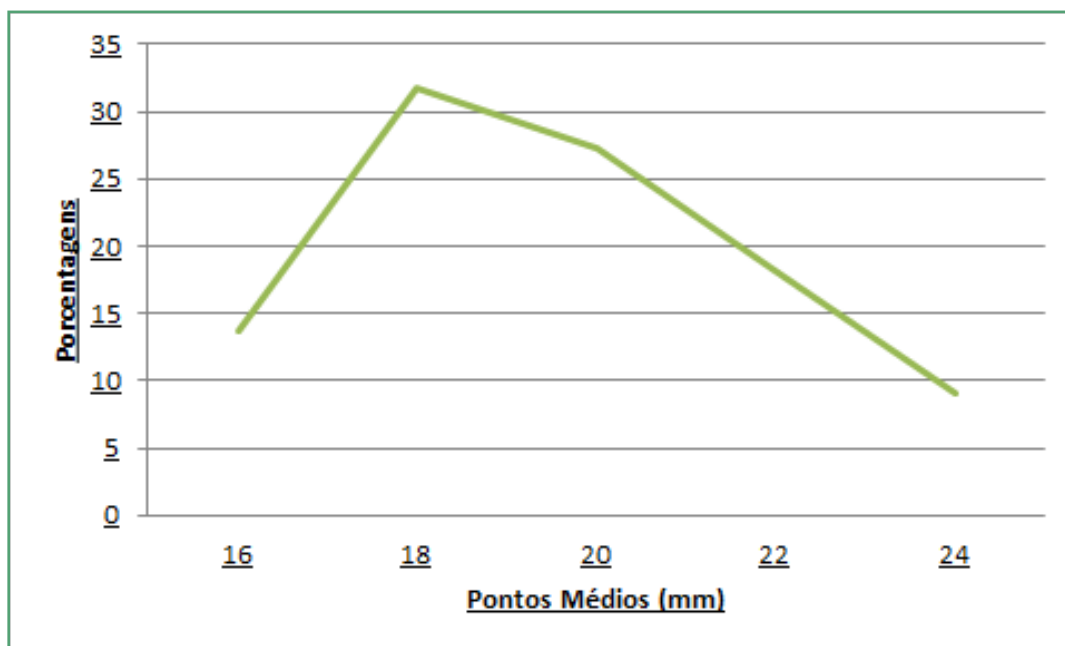
Tabela 02 - Distribuição de frequências de precipitações pluviométricas (em mm).

| Precipitações pluviométricas (em mm) | f | % | X |
|--------------------------------------|-----------|------------|----|
| 15 ----- 17 | 3 | 13,64 | 16 |
| 17 ----- 19 | 7 | 31,82 | 18 |
| 19 ----- 21 | 6 | 27,27 | 20 |
| 21 ----- 23 | 4 | 18,18 | 22 |
| 23 ----- 25 | 2 | 9,09 | 24 |
| Σ | 22 | 100 | |

Fonte: Dados primários.

Para você construir um polígono de frequência para ela, terá que marcar nos dois eixos escalas a seu critério. Por exemplo, para o eixo X, onde vão ficar os pontos médios, pode-se usar a escala já da distribuição que é de 2 em 2, pois de 16 para 18 a escala é 2, de 18 para 20 é 2, e assim para todos os pontos médios; já para o eixo Y, onde vão ficar as porcentagens, pode-se usar a escala de 5% em 5%, começando em 5%, passando para 10%, para 15%, para 20%, para 25%, para 30% e para 25%. Veja o gráfico 8.

Gráfico 8 – Distribuição de precipitações pluviométricas (em mm)



Fonte: Dados primários.

Veja que a escala do eixo X é 2, que é diferente da escala 5% do eixo Y. As escalas precisam ser iguais dentro do eixo. Lembro-lhe também que qualquer outra escala seria válida, desde que fosse sempre a mesma dentro do eixo.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Apresentação Gráfica até agora, utilize a distribuição de frequências a seguir, referente ao número diário de casos de dengue em determinado município, e construa um Histograma e um Polígono de Frequências.



Tabela 03 - Casos diários de dengue, em determinado município.

| Casos Dengue | f |
|--------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.



RESUMINDO

Nessa aula você aprendeu a importância que as apresentações gráficas têm para a compreensão de tendência/comportamento dos dados, bem como para melhor apresentá-los sob o aspecto visual. Aprendeu também a construir vários gráficos, em especial o Histograma e o Polígono de Frequência, que são utilizados junto com distribuições de frequências.



LEITURAS COMPLEMENTARES

No artigo disponível no site <http://www.unifor.br/notitia/file/1053.PDF> você encontrará análises sobre a influência das condições ambientais e organizacionais de trabalho na qualidade de vida dos trabalhadores e sua relação com a qualidade dos serviços oferecidos em uma indústria no estado do Ceará. Para consubstanciar a análise foram utilizados vários histogramas como um meio de melhorar a visualização dos resultados e proporcionar uma análise mais expressiva. Em particular, você perceberá que os trabalhadores avaliaram muito ruim a qualidade do ambiente térmico, que teve um índice 0,25; e avaliaram como ruim a qualidade do ambiente sonoro, que teve um índice de 0,50; os índices variariam de 0 a 1.

AVALIANDO SEUS CONHECIMENTOS

Considerando que você já sabe construir gráficos, observe a distribuição de frequências a seguir, apresentada na tabela 4, referente à renda mensal (em salários mínimos) de famílias de determinado município. Construa o Histograma e o Polígono de frequências para a zona urbana e para a zona rural; comente os mesmos.



Tabela 04 - Salários mínimos por área.

| RENDAS (Salários Mínimos) | ÁREA URBANA | ÁREA RURAL |
|---------------------------|-------------|------------|
| 1 ---- 8 | 3 | 2 |
| 8 ---- 15 | 7 | 9 |
| 15 ---- 22 | 2 | 5 |
| 22 ---- 29 | 4 | 2 |
| 29 ---- 36 | 1 | 1 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 08
ASSIMETRIA E CURTOSE

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

A análise gráfica de uma distribuição, como você estudou na Aula 7 – Apresentação Gráfica, permite identificar tendências/comportamentos nos dados. Há, no entanto, um comportamento desejável do ponto de vista essencialmente estatístico, que é o comportamento denominado de normal; para identificar esse comportamento, você pode utilizar-se da Assimetria e da Curtose, que são medidas complementares utilizadas para identificar se uma distribuição de frequências pode ser considerada normal. Se uma distribuição for considerada normal, várias outras técnicas para realizar inferências sobre os dados podem ser utilizadas. Portanto, nesta aula, você vai estudar técnicas que permitirão identificar aspectos comportamentais de uma distribuição de frequências, que darão suporte para a utilização de outras técnicas mais refinadas da estatística, as quais só poderão ser utilizadas se os dados forem considerados normais.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

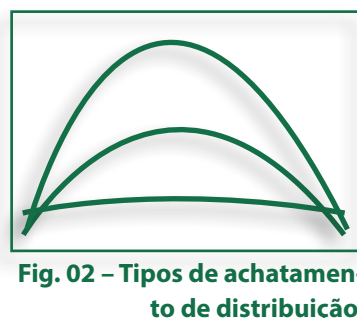
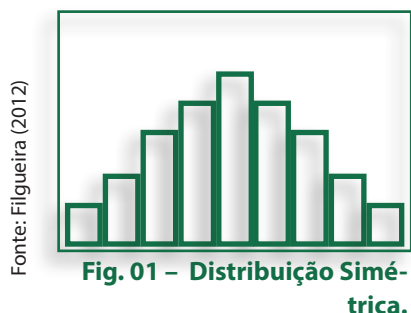
- compreender a importância comportamental da Assimetria e da Curtose;
- calcular Assimetria e Curtose;
- identificar a normalidade ou não de uma distribuição de frequências .

DESENVOLVENDO O CONTEÚDO

ENTENDENDO ASSIMETRIA E CURTOSE

Uma distribuição de frequências contém muitas informações, dela podem ser obtidas média, moda, mediana, variância, desvio padrão, coeficiente de variação, separatrizes. Até o comportamento/tendência da distribuição pode ser obtido através de uma apresentação gráfica. Mas há outras técnicas que só poderão ser utilizadas se os dados forem considerados normais.

Esse critério equivale ao fato da distribuição ter um comportamento simétrico, conforme você pode observar na figura 1, de modo que a maior parte dos dados se concentre em torno do centro dos dados, uma pequena quantidade esteja bem abaixo desse centro e outra pequena quantidade bem acima dele.



A situação desejável é aquela em a distribuição é simétrica e tem grau de achatamento uniforme, nem muito mais estreito que a distribuição normal, nem mais largo do que ela. Você vai passar agora a aplicar esses conceitos de forma concreta, utilizando suas equações.

ATIVIDADE 01



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre Assimetria e Curtose até agora:

1. Utilize o gráfico histograma ou polígono de frequência, construído na Aula 7 – Apresentação Gráfica para o exemplo dos casos de dengue, e afirme se a distribuição pode ser considerada normal.

CALCULANDO ASSIMETRIA E CURTOSE

Agora que você já deve estar bastante consciente da importância da Assimetria e da Curtose, você vai aprender como calcular essas medidas e interpretá-las adequadamente.

A **Assimetria** tem por finalidade quantificar o deslocamento/afastamento da distribuição em relação às suas medidas centrais. Você vai utilizar uma equação de Assimetria, a (8.1). Há outras equações, mas todas têm uma representatividade similar.

$$(8.1) \text{ Ass} = \frac{(\bar{X} - \hat{X})}{S}$$

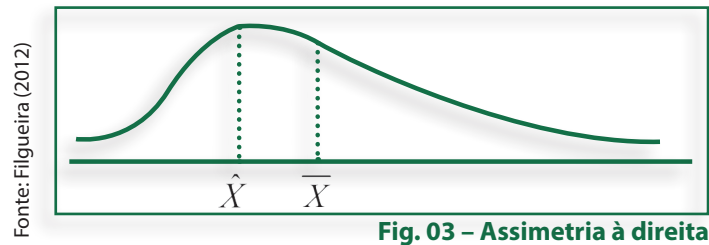
Onde:

\bar{X} representa a média;

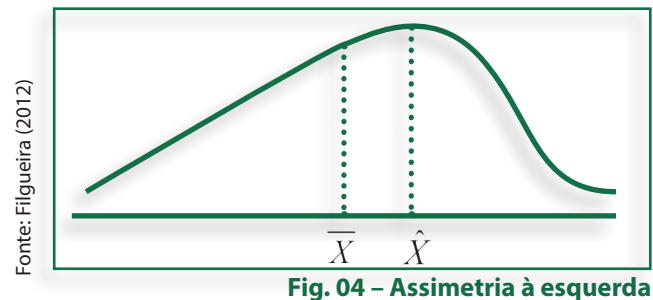
\hat{X} representa a moda;

S representa o desvio padrão.

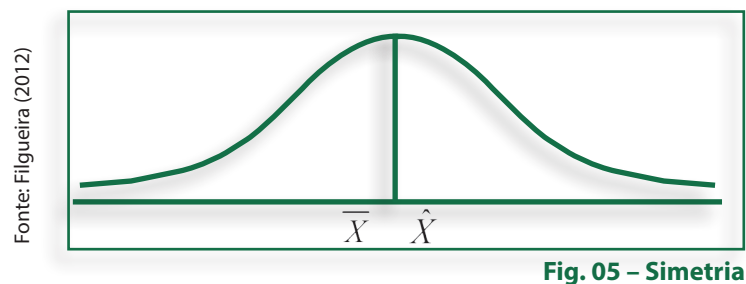
A literatura apresenta as seguintes situações inerentes à Assimetria. A distribuição pode ser Assimétrica à direita, conforme você pode observar na figura 3, nesta situação você perceberá que a Moda é menor que a Média, o que desloca a distribuição para o início. Neste caso, o valor da Assimetria é positivo.



Uma outra situação pode ser vista na figura 4, onde a distribuição é Assimétrica à esquerda. Nesta situação, você perceberá que a Moda é maior que a Média, o que desloca a distribuição para o final. Neste caso, o valor da Assimetria é negativo.



E, finalmente, a situação desejável é a da figura 5, onde a distribuição é Simétrica em torno da Média, nesta situação, a Moda e a Média são iguais, o que gera uma grande concentração em torno da Média. Neste caso, o valor da Assimetria é igual a zero.



Para calcular a Assimetria, será utilizado o exemplo das precipitações pluviométricas, visto na Aula 4 – Medidas de Tendência Central. Para este exemplo, na Aula 4 – Medidas de Tendência Central, você já calculou a Média e a Moda; na Aula 5 – Medidas de Variabilidade, já calculou o Desvio Padrão. Os valores são:

Média igual a 19,55 mm;

Moda igual a 18,6mm;

Desvio Padrão igual a 2,39mm.

Desse modo, aplicando a fórmula (8.1), obtem-se:

$$(8.2) \quad Ass = \frac{(\bar{X} - \hat{X})}{S} = \frac{(19,55 - 18,6)}{2,39} = \frac{0,95}{2,39} = 0,397 \cong 0,40$$

Portanto, o valor final da Assimetria é positivo e vale 0,40, o que significa que a distribuição é levemente Assimétrica à direita. Inclusive, essa situação já ficou evidenciada quando, na Aula 7 – Apresentação Gráfica, você construiu o Polígono de frequências e o Histograma, o qual está apresentado novamente aqui na figura 6.

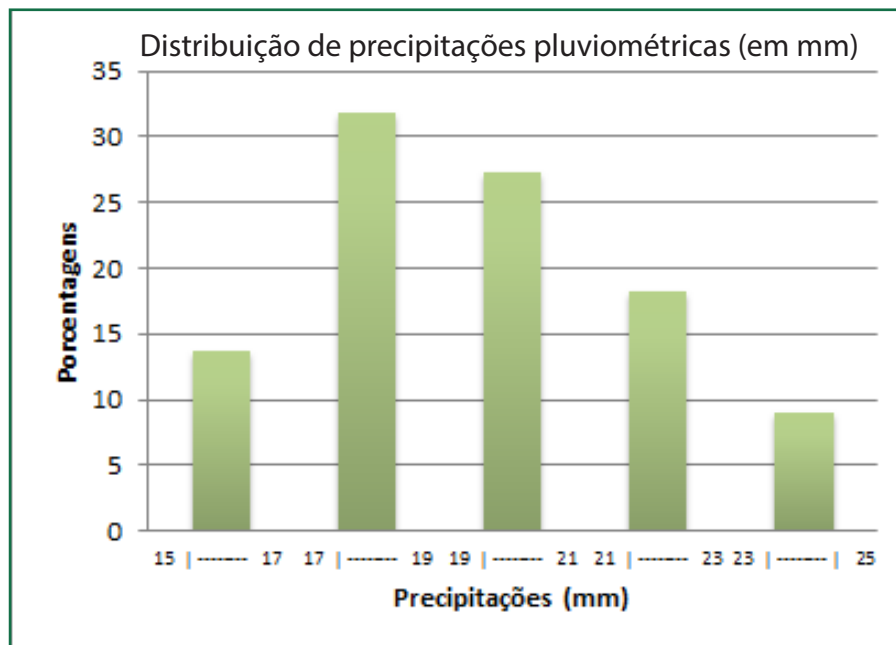


Fig. 06 – Histograma

Como você já observou na Figura 6, a segunda classe de 17 |----- 19, que tem a maior frequência e, portanto, a Moda, deslocou um pouco a distribuição para a direita. No entanto, como os valores da Média e da Moda são bastante próximos, o valor final da Assimetria ficou muito próximo de zero, ou seja, a Distribuição é levemente Assimétrica à direita.

A Curtose tem por finalidade quantificar o achatamento/variação da distribuição em relação à distribuição simétrica. Você vai utilizar a seguinte equação de Curtose:

$$(8.3) \quad K = \frac{(Q_3 - Q_1)}{2 * (P_{90} - P_{10})}$$

Onde:

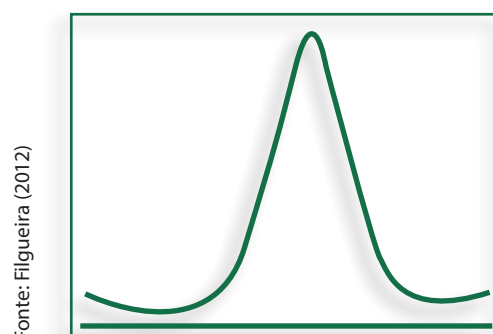
Q_3 representa o terceiro Quartil;

Q_1 representa o primeiro Quartil;

P_{90} representa o 90º Percentil;

P_{10} representa o 10º Percentil.

A literatura apresenta várias situações inerentes à Curtose. Em uma dessas situações, a distribuição pode ser Leptocúrtica, conforme você pode observar na figura 7. Neste caso, o valor da Curtose é menor que 0,263.



Fonte: Filgueira (2012)

Fig. 07 - Distribuição Leptocúrtica

Uma outra situação pode ser vista na figura 8, onde a distribuição é Platicúrtica. Neste caso, o valor da Curtose é maior que 0,263.

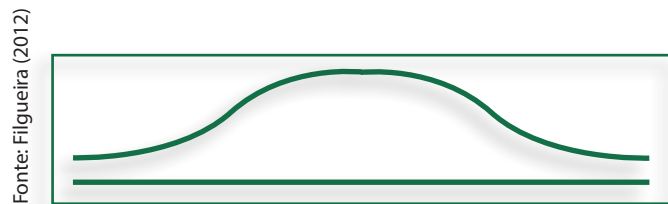


Fig. 08 - Distribuição Platicúrtica

E, finalmente, a situação desejável é a da figura 9, onde a distribuição é Mesocúrtica. Neste caso, o valor da Curtose é igual a 0,263.

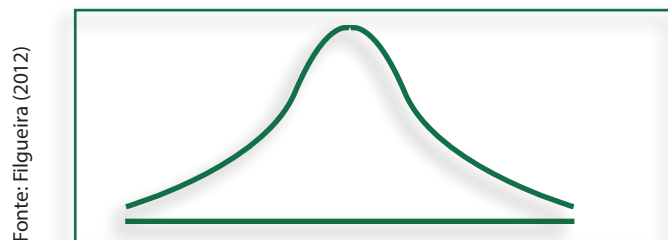


Fig. 09 - Distribuição Mesocúrtica

Para calcular a Curtose, será utilizado o exemplo das precipitações pluviométricas, visto na Aula 6 – Medidas Separatrizes. Para este exemplo, na Aula 6 – Medidas Separatrizes, você já calculou o Q_1 e o Q_3 , cujos valores foram Q_1 igual a 17,72 mm e Q_3 igual a 21,25 mm. Faltam apenas calcular P_{90} e P_{10} , que serão calculados a seguir, utilizando a tabela 1.

Tabela 1 - Distribuição de frequências de precipitações pluviométricas (em mm)

| Precipitações | f | % |
|----------------------------|-----------|------------|
| 15 ----- 17 | 3 | 13,64 |
| 17 ----- 19 | 7 | 31,82 |
| 19 ----- 21 | 6 | 27,27 |
| 21 ----- 23 | 4 | 18,18 |
| 23 ----- 25 | 2 | 9,09 |
| Σ | 22 | 100 |

Fonte: Dados primários.

Para calcular o P_{90} , inicialmente você precisa identificar a primeira classe que a soma das porcentagens contenha 90%, que é a classe **21 |----- 23**, que contém 90,91% (Vide Tabela 1). Você irá aplicar a fórmula (8.4), tomando como referência essa classe.

$$(8.4) \quad P_{90} = L + \left(\frac{90 * \left(\frac{\sum f}{100} \right) - Sant}{f} \right) * a$$

Onde:

L = 21 representa o limite inferior da classe do P_{90} ;

i = 90 representa o índice do P_{90} ;

$\sum f = 22$ representa a soma das frequências;

Sant = 6+7+3 = 16 representa a soma das frequências anteriores;

f = 4 representa a frequência da classe do P_{90} ;

a = 2 representa a amplitude de classe.

Portanto, o valor do P_{90} será

$$P_{90} = L + \left(\frac{90 * \left(\frac{\sum f}{100} \right) - Sant}{f} \right) * a$$

$$P_{90} = 21 + \left(\frac{90 * \left(\frac{22}{100} \right) - 16}{4} \right) * 2 = 21 + 1,9 = 22,9$$

Logo, há 90% dos dias com chuvas abaixo de 22,9 mm. Para o P_{10} , inicialmente você precisa identificar a primeira classe que a soma das porcentagens contenha 10%, que é a classe **15 |----- 17**, que contém 13,64% (Vide Tabela 1).

Você irá aplicar a fórmula (8.5), tomando como referência essa classe.

$$(8.5) \quad P_{10} = L + \left(\frac{10 * \left(\frac{\sum f}{100} \right) - Sant}{f} \right) * a$$

Onde:

L = 15 representa o limite inferior da classe do P_{10} ;

i = 10 representa o índice do P_{10} ;

$\sum f = 22$ representa a soma das frequências;

Sant = 0 representa a soma das frequências anteriores;

f = 3 representa a frequência da classe do P_{10} ;

a = 2 representa a amplitude de classe.

Portanto, o valor do P_{10} será

$$P_{10} = L + \left(\frac{10 * \left(\frac{\sum f}{100} \right) - Sant}{f} \right) * a$$

$$P_{10} = 15 + \left(\frac{10 * \left(\frac{22}{100} \right) - 0}{3} \right) * 2 = 15 + 1,466 \cong 16,47$$

Logo, há 10% dos dias com chuvas abaixo de 16,47 mm.

Portanto, o valor da Curtose será

$$K = \frac{(Q_3 - Q_1)}{2 * (P_{90} - P_{10})} = \frac{(21,25 - 17,72)}{2 * (22,9 - 16,47)} = \frac{3,53}{12,86} = 0,274 \cong 0,27$$

Portanto, o valor final da Curtose vale 0,27, que é maior que 0,263, o que significa que a distribuição é levemente Platicúrtica. Inclusive essa situação já ficou evidenciada quando, na Aula 7 – Apresentação Gráfica, você construiu o Polígono de frequências e o Histograma, o qual está apresentado novamente aqui na figura 10.

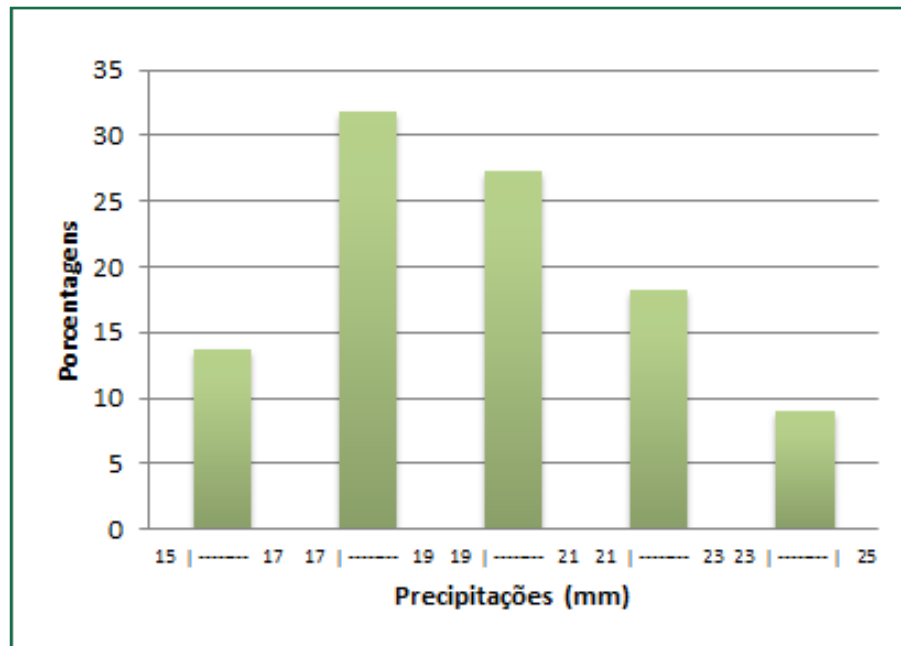


Fig. 10 – Histograma

Ou seja, considerando-se o valor da Assimetria, o valor da Curtose e o Histograma, pode-se afirmar que a distribuição das precipitações pluviométricas pode ser considerada normal.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Assimetria e Curtose até agora, utilize a distribuição de frequências a seguir, referente ao número diário de casos de dengue em determinado município, calcule Assimetria.



Tabela 02 - Casos diários de dengue, em determinado município.

| Casos Dengue | f |
|--------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.



RESUMINDO

Nesta aula você aprendeu a importância que a Assimetria e a Curtose têm na identificação do comportamento da distribuição. Aprendeu também que essas medidas complementam uma análise gráfica, e todas elas conjuntamente permitem avaliar quando uma distribuição pode ser considerada normal.



LEITURAS COMPLEMENTARES

O artigo apresentado no site <http://ppe.ipea.gov.br/index.php/ppe/article/viewFile/24/> demonstra que o crescimento da produtividade setorial da indústria brasileira deve-se não somente ao aumento da produtividade das firmas, mas também à sua dinâmica intra-setorial. Na análise, são utilizados os coeficientes de Assimetria e de Curtose, além de várias apresentações gráficas. O artigo apresenta muitos gráficos que demonstram, de um modo geral, comportamento simétrico, possuindo, portanto, distribuição normal, o que é imprescindível para aplicação das demais técnicas utilizadas.

AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe calcular Assimetria e Curtose, considere a distribuição de frequências a seguir, apresentada na tabela 3, para a qual você já calculou a Assimetria na ATIVIDADE 2, calcule a Curtose e afirme se a distribuição pode ser considerada normal.



Tabela 03 - Casos diários de dengue, em determinado município.

| Casos Dengue | f |
|--------------------|---|
| 0,0 ----- 3,4 | 5 |
| 3,4 ----- 6,8 | 3 |
| 6,8 ----- 10,2 | 4 |
| 10,2 ----- 13,6 | 2 |
| 13,6 ----- 17,0 | 6 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 09
SIGNIFICADO DA ANÁLISE DE CORRELAÇÃO LINEAR

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL

APRESENTANDO A AULA



Você já estudou bastante sobre a Estatística Descritiva e suas técnicas para descrever e apresentar fenômenos. Em todas as situações estudadas você utilizou apenas uma variável para análise: foi assim nas precipitações pluviométricas, nos casos de dengue, nas rendas mensais. Mas você já está imaginando que nem sempre apenas uma variável será suficiente para analisar adequadamente um fenômeno, você com certeza precisará estudar técnicas que permitam analisar mais de uma variável. Para isso, há várias técnicas, dentre elas uma de bastante utilização e que produz bons resultados é a Análise de Correlação Linear, que você vai estudar nesta aula. Com ela, por exemplo, você poderá descobrir se a chuva é responsável pela ocorrência de dengue, se o número de alunos matriculados é responsável pela redução da violência, se o desmatamento é responsável pela redução na produção de leite, se o investimento em saúde de fato é responsável pela redução de doenças; enfim, com a Análise de Correlação Linear, utilizando-se de duas ou mais variáveis, você poderá responder a diversas questões acadêmicas, técnicas, sociais, políticas, só para citar algumas aplicações. Por questões didático-pedagógicas, você vai estudar técnicas aplicadas para duas variáveis; mas a expansão para um número maior de variáveis é bastante simples, se você compreender bem a aplicação dessas técnicas para duas variáveis.

DEFININDO OBJETIVOS

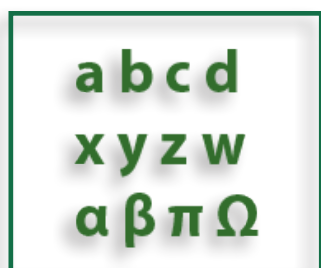
Ao final desta aula, você deverá ser capaz de:

- compreender a importância da Análise de Correlação Linear;
- identificar variáveis independente e dependente;
- construir diagrama de dispersão.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO ANÁLISE DE CORRELAÇÃO LINEAR

Você já deve ter ouvido falar que os alunos que mais faltam às aulas têm notas mais baixas, ou seja, intervindo nas faltas seria possível obter melhores notas. Já ouviu também que quando há mais chuva é provável que as produções agrícolas tendam a ser maiores, ou seja, com um aumento apenas da água irrigada haveria um aumento nas produções agrícolas. Há várias outras situações do dia-a-dia, nas quais uma determinada variável interfere na ocorrência de outra variável, de tal forma que, para que se possa controlar a ação conjunta de ambas, basta intervir em apenas uma delas.



Fonte: PAIVA (2013)

Fig. 01 - Representação de variáveis

Você deve estar imaginando como seria possível estudar variáveis dessa forma e controlar adequadamente o comportamento global do fenômeno. Com a Análise de Correlação Linear é perfeitamente possível identificar o comportamento das variáveis e decidir se elas se comportam de tal modo que uma interfira diretamente ou indiretamente na ocorrência da outra.

No exemplo citado anteriormente, entre notas e faltas, é provável que as faltas interfiram inversamente nas notas, pois espera-se que quanto maior for o número de faltas menor será o valor da nota. Nesse caso haveria uma correlação inversa entre faltas e notas. No outro exemplo citado anteriormente, entre irrigação e produção agrícola, é provável que a quantidade de água irrigada interfira diretamente na produção agrícola, pois espera-se que quanto maior for a quantidade irrigada maior será a produção agrícola. Nesse caso haveria uma correlação direta entre irrigação e produção agrícola.

Como você já percebeu, em cada um dos exemplos citados há duas variáveis: uma pode ser chamada de **independente e ser denotada por X**, e a outra pode ser chamada de **dependente e ser denotada por Y**. É muito importante você definir corretamente as variáveis X e Y, pois elas serão utilizadas mais adiante nessa aula para a construção de um gráfico.

No caso do exemplo das faltas e das notas, você com certeza já identificou que as notas dependem das faltas, logo a variável **dependente é Y-Notas** e a variável **independente é X-Faltas**. No outro, das irrigações e produções agrícolas, a variável **dependente é Y-Produções** e a variável **independente é X-Irrigações**.

Como vão haver somente duas variáveis em nosso estudo, identificar a que é Y-dependente e a que é X-independente é uma tarefa muito simples.

ATIVIDADE 01



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre Análise de Correlação Linear até agora:

1. Identifique um fenômeno ligado à Gestão Ambiental;
2. defina um par de variáveis que possam representá-lo;
3. afirme qual é a variável dependente e qual é a variável Independente.

Você certamente vai perceber a simplicidade dessa definição, embora elas sejam grande importância para o estudo do fenômeno.

CONSTRUINDO DIAGRAMA DE DISPERSÃO

Após a definição das variáveis X e Y, é preciso avaliar se os pares (X,Y) dessas variáveis se comportam de forma linear. Aqui cabe uma observação, os valores de X e Y precisam referir-se ao mesmo período de tempo, ou ao mesmo local de coleta, ou a alguma outra condição que seja comum às duas variáveis, só desse modo poderão ser utilizadas na Análise de Correlação.

Por exemplo, na tabela 1 constam as notas Estatística de alunos de uma turma de Gestão Ambiental, no primeiro e no segundo período.

Tabela 1 – Notas do 1º e 2º período de Estatística

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8,0 |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Para este exemplo, seria possível estudar as variáveis 1ª Nota e 2ª Nota, pois os alunos que têm 1ª Nota também têm 2ª Nota, ou seja, está caracterizado o emparelhamento entre 1ª Nota e 2ª Nota.

Considere um outro exemplo, representado na tabela 2, que contém Notas da 1ª Unidade de alunos de Estatística e de Matemática, todos de Gestão Ambiental.

Tabela 2 – Notas da 1ª Unidade de Estatística e de Matemática

| Aluno | Estatística | Matemática |
|-------|-------------|------------|
| 1 | 4,5 | ---- |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Legenda: ---- aluno não cursa a disciplina.

Para este exemplo, não seria possível estudar as variáveis Estatística e Matemática, pois nem todos os alunos que cursam Estatística cursam também Matemática, ou seja, não está caracterizado o emparelhamento entre as variáveis Estatística e Matemática.

Muito bem, agora que você já sabe em que condições as variáveis X e Y podem ser utilizáveis para um estudo de Correlação Linear, você precisa saber

identificar quando elas têm comportamento linear.

Para essa identificação o meio mais adequado é o meio gráfico, utilizando-se um gráfico de pontos (X,Y) denominado Diagrama de Dispersão. Antes, porém, é bom você lembrar-se que a variável X-Independente será representada no eixo horizontal do gráfico, e a variável Y-Dependente será representada no eixo vertical do gráfico.

Em relação ao Diagrama de Dispersão, ele pode apresentar vários comportamentos, que serão discutidos a partir de agora.

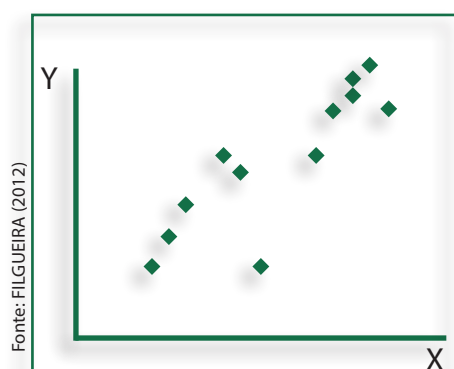


Fig. 02 – Correlação Linear Positiva

Considere, por exemplo, a figura 2. Nela é possível compreender que o gráfico é de pontos, tem os eixos X e Y, e apresenta uma correlação bem positiva, pois à medida que os valores de X crescem os de Y também crescem.

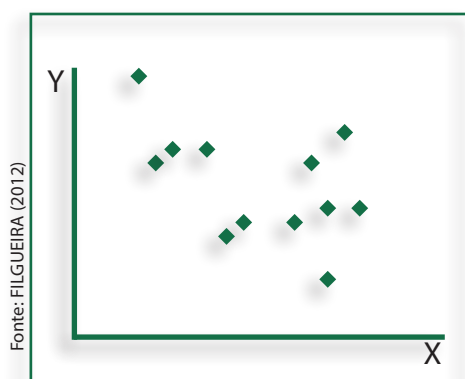


Fig. 03 – Correlação Linear Negativa

Já a figura 3 apresenta uma correlação bem negativa, pois à medida que os valores de X crescem os de Y decrescem.

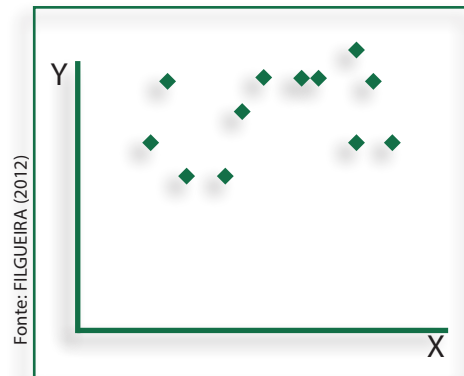


Fig. 04 – Ausência de Correlação Linear

A figura 4, por sua vez, apresenta uma situação em que não é possível associar o comportamento da variável Y ao comportamento da variável X, pois **não está claro se quando a variável X cresce a variável Y cresce ou decresce**. Nesse caso, diz-se que não há correlação linear entre X e Y.

Agora você passará a construir diagramas de dispersão, e para isso você deve recordar que, na AULA 7 - Apresentação Gráfica, foi discutido que um aspecto muito importante na construção de um gráfico é a **escala**. Assim, você precisa estar atento na definição da escala do eixo X e na do eixo Y, considerando que **o critério é que a escala sempre possua o mesmo valor, ou sempre 5, ou sempre 10, ou outro valor**.

Considerando o exemplo das notas de estatística contidas na tabela 1, apresentada anteriormente, você vai construir um diagrama de dispersão e analisar se ele apresenta uma provável correlação. Inicialmente, é necessário definir a variável dependente e a variável independente; depois é preciso definir a escala de cada eixo; depois deve-se marcar o diagrama de dispersão; e, finalmente, avaliar se o comportamento do diagrama exprime ou não correlação linear entre as variáveis.

Em relação às variáveis, é de se esperar que a 2ª nota dependa de certo modo da 1ª nota, assim a variável dependente será Y-2ª Nota e a variável independente será X-1ª Nota.

Em relação à escala, pode ser utilizada uma **escala de 0,5 para o eixo X**, e uma **escala de 1,0 para o eixo Y**.

Com essas informações, é possível construir o Diagrama de Dispersão apresentado na figura 5.

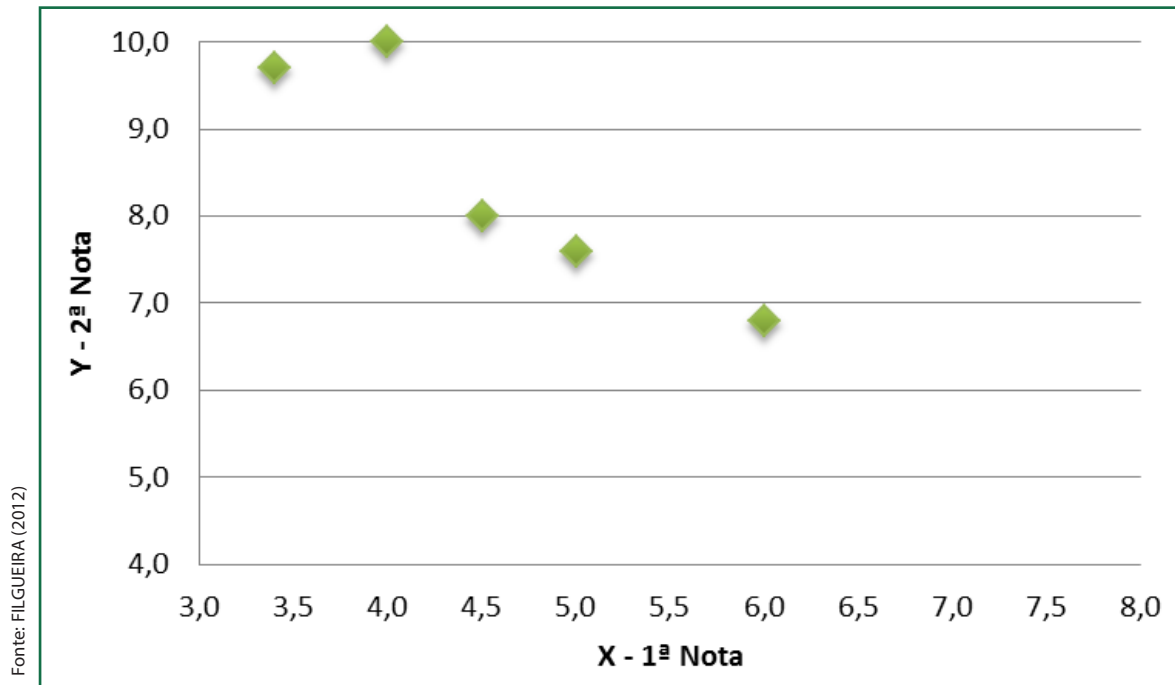


Fig. 05 – Diagrama de Dispersão entre 1ª Nota e 2ª Nota

É possível concluir, à luz do gráfico da figura 5, que há uma correlação negativa entre X e Y, pois à medida que os dados de X crescem os de Y decrescem. Portanto, seria razoável concluir que correlação entre a 1ª Nota e a 2ª Nota é negativa, quanto maior for a primeira nota, menor tenderá a ser a segunda nota.

ATIVIDADE 02



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

Com base no que estudamos sobre Análise de Correlação até agora, utilize os dados a seguir para construir um Diagrama de Dispersão e comentar o comportamento dos dados.

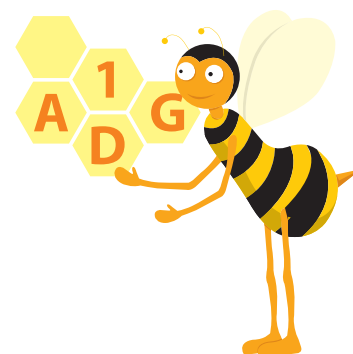
Tabela 3: Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos.

| MÊS | FATURAMENTO (R\$) | PESSOAS ATENDIDAS |
|----------|-------------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

RESUMINDO

Nessa aula você aprendeu a importância que a Análise de Correlação Linear tem para a compreensão de tendência/comportamento de pares (X,Y) de dados, bem como para identificar a relação de dependência entre as variáveis X e Y . Aprendeu também a construir o Diagrama de Dispersão, gráfico utilizado amplamente para estudos de Correlação Linear.



LEITURAS COMPLEMENTARES

No artigo disponível no site http://www.economia-aplicada.ufv.br/revista/pdf/2007/01_Artigo_%206.pdf você encontrará várias aplicações do Diagrama de Dispersão no estudo entre o volume de açúcar exportado pelo porto de Santos e o valor do frete negociado para a distribuição de fertilizante ao cliente final, no Estado de São Paulo. Como resultados, pôde-se perceber que em 2004 há uma pequena correlação entre o volume de açúcar exportado pelo porto de Santos e o valor do frete. Em 2005 já há uma definição mais expressiva dessa correlação.



AVALIANDO SEUS CONHECIMENTOS



Considerando-se que você já sabe a importância da Análise de Correlação Linear e sabe também construir gráficos de Dispersão, considere a tabela 4, contendo os valores de testes realizados junto a funcionários de uma indústria antes e depois de um treinamento de requalificação. Construa o Diagrama de Dispersão e comente sobre a Correlação entre as variáveis. Os valores dos testes são de zero a dez.

Tabela 4: Testes antes e após treinamento de requalificação.

| Funcionário | Teste antes | Teste após |
|-------------|-------------|------------|
| 1 | 8,4 | 7,1 |
| 2 | 4,6 | 1,9 |
| 3 | 2,5 | 9,7 |
| 4 | 9,8 | 7,4 |
| 5 | 1,5 | 7,1 |
| 6 | 7,4 | 5,7 |
| 7 | 2,0 | 5,9 |
| 8 | 7,0 | 6,0 |
| 9 | 0,1 | 9,7 |
| 10 | 3,9 | 7,0 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 10
COEFICIENTE DE CORRELAÇÃO LINEAR

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL

APRESENTANDO A AULA



Você estudou na aula passada como construir um diagrama de dispersão e perceber várias situações em que a **variável independente X** pode ter uma relação com a **variável dependente Y**. Ou seja, você aprendeu a identificar graficamente possíveis correlações entre X e Y e isso é muito importante, mas é, de certo modo, subjetivo. Você certamente deseja aprender técnicas mais objetivas que complementem a análise gráfica, mensurando concretamente se a **variável X** relaciona-se com a **variável Y**, e até mesmo deseja saber quanto vale essa correlação, se é baixa, se é razoável, se é alta. Por exemplo: será que é alta a correlação entre chuva e casos de dengue? Se for, é preciso intensificar as medidas preventivas sempre que haja informação de que haverá muita chuva. Será que é alta a correlação entre o número de alunos matriculados e a redução da violência? Se for, é preciso intensificar as medidas que aumentem o número de alunos matriculados para garantir a redução nos casos de violência. Será que é alta a correlação entre o desmatamento e a redução na produção de leite? Se for, é preciso intensificar as medidas que reduzam o desmatamento para garantir que continue havendo produção de leite. Enfim, com o **Coefficiente de Correlação Linear** você poderá mensurar o grau de relacionamento entre variáveis em diversas situações acadêmicas, técnicas, sociais, políticas, só para citar algumas aplicações.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender a importância do Coeficiente de Correlação Linear;
- calcular o Coeficiente de Correlação Linear;
- calcular o Coeficiente de explicação.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO O COEFICIENTE DE CORRELAÇÃO LINEAR

Na Aula 9 - Significado da Análise de Correlação Linear - você aprendeu que, a partir de um Diagrama de Dispersão, é possível identificar se as variáveis X-Independente e Y-Dependente são correlacionadas. Embora essa identificação seja bastante importante, ela é subjetiva, e pode ser muito bem complementada pela utilização do Coeficiente de Correlação Linear, que mensura a proximidade dos pares (X,Y) e verifica se eles têm um comportamento aceitavelmente linear, como você vai aprender a partir de agora.

Muitas vezes, ao identificar um comportamento aceitavelmente linear em um Diagrama de Dispersão, você precisa de um valor que mensure adequadamente o quanto é linear esse comportamento: será que as variáveis X-independente e Y-dependente são pouco correlacionadas, são muito correlacionadas, enfim, de quanto vale essa correlação? O Coeficiente de Correlação Linear mensura com bastante qualidade essa correlação.



Fonte: FILGUEIRA (2012)

Fig. 01 - Diagrama de Dispersão com baixa correlação, $r = 0,4$

Por exemplo, considere a figura 1, nela o diagrama de dispersão está apresentando uma razoável correlação entre X e Y, porém quando foi calculado o Coeficiente de Correlação, que você vai aprender mais adiante, ainda nesta aula, o valor obtido demonstra uma fraca correlação.



Fonte: FILGUEIRA (2012)

Fig. 02 - Diagrama de Dispersão com alta correlação, $r = 0,7$

Na figura 2, o diagrama de dispersão está apresentando uma razoável correlação entre X e Y, e quando foi calculado o Coeficiente de Correlação, o valor obtido demonstrou uma correlação alta.

Na figura 3, o diagrama de dispersão está apresentando uma grande correlação entre X e Y, e quando foi calculado o Coeficiente de Correlação, o valor obtido demonstrou uma correlação muito alta.

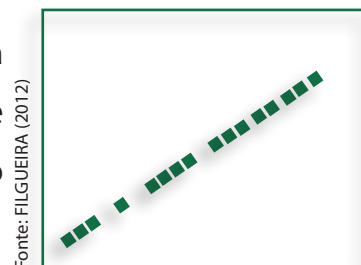


Fig. 03 – Diagrama de Dispersão com correlação muito alta, $r = 1,0$

Na figura 4, o diagrama de dispersão está apresentando uma razoável correlação inversa entre X e Y, e quando foi calculado o Coeficiente de Correlação, o valor obtido demonstrou uma correlação baixa.



Fig. 04 – Diagrama de Dispersão com alta correlação, $r = -0,3$

Portanto, conforme foi observado em todas as figuras, há um indicador de Correlação, que é o r expressando o quanto as variáveis X e Y estão correlacionadas. Na figura 1, por exemplo, há pequena correlação, $r=0,4$; na figura 2, já há uma correlação melhor, $r=0,7$; na figura 3, há uma correlação perfeita $r=1,0$; na figura 4, há uma correlação inversa bem baixa $r=-0,3$.

Como você já sabe, quanto maior o Coeficiente de Correlação Linear, maior será a influência da variável X na variável Y. Portanto, você precisa aprender a calcular esse Coeficiente de Correlação Linear.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Coeficiente de Correlação Linear até agora, utilize os dados a seguir e o Diagrama de Dispersão já construído na Aula 9 - Significado da Análise de Correlação Linear - e comente sobre o provável valor do Coeficiente de Correlação Linear.



Tabela 1: Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

CALCULANDO O COEFICIENTE DE CORRELAÇÃO LINEAR E O COEFICIENTE DE EXPLICAÇÃO

Para calcular o Coeficiente de Correlação Linear, você vai utilizar a seguinte fórmula.

$$(10.1) \quad r = \frac{n \sum (X * Y) - (\sum X) * (\sum Y)}{\sqrt{(n * (\sum X^2) - (\sum X)^2)} * \sqrt{(n * (\sum Y^2) - (\sum Y)^2)}}$$

Onde:

X é a variável independente;

Y é a variável dependente;

N é o número de pares de dados.

É importante lembrar que os dados precisam estar emparelhados, para poder ser utilizado o Coeficiente de Correlação Linear.

Outro aspecto importante a ser lembrado, é que o Coeficiente de Correlação Linear é composto de três elementos: no numerador da fórmula (10.1) encontra-se a Covariância, que é a variação que é comum a X e a Y; no denominador

encontram-se, dentro de raízes, o desvio-padrão de X e o desvio-padrão de Y. Os desvios-padrões de X e de Y jamais serão negativos, muito dificilmente iguais a zero; a covariância pode ser negativa, nula ou positiva.

Segundo a literatura, os valores possíveis de r variam de -1 a $+1$, e têm a seguinte representatividade.

Tabela 2: representatividade do Coeficiente de Correlação Linear

| Valor de r | Representatividade |
|----------------|------------------------|
| De 0,00 a 0,19 | Correlação muito baixa |
| De 0,20 a 0,39 | Correlação baixa |
| De 0,40 a 0,69 | Correlação moderada |
| De 0,70 a 0,89 | Correlação forte |
| De 0,90 a 1,00 | Correlação muito forte |

Fonte: Dados primários.

A tabela 2 tem a mesma interpretação quando o valor de r é negativo. Ou seja, valores de $-0,19$ a $0,00$ representam correlação muito baixa; valores de $-0,90$ a $-1,00$ representam correlação muito forte.

Você já deve estar querendo calcular o coeficiente de correlação linear. Considere, por exemplo, a tabela 3, que contém notas de Estatística de alunos de uma turma de Gestão Ambiental no primeiro e no segundo períodos.

Tabela 3 – Notas do 1º e 2º período de Estatística

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8,0 |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Você já havia construído o Diagrama de Dispersão para esses dados na Aula 9 - Significado da Análise de Correlação Linear. Ele está apresentado na figura 5.

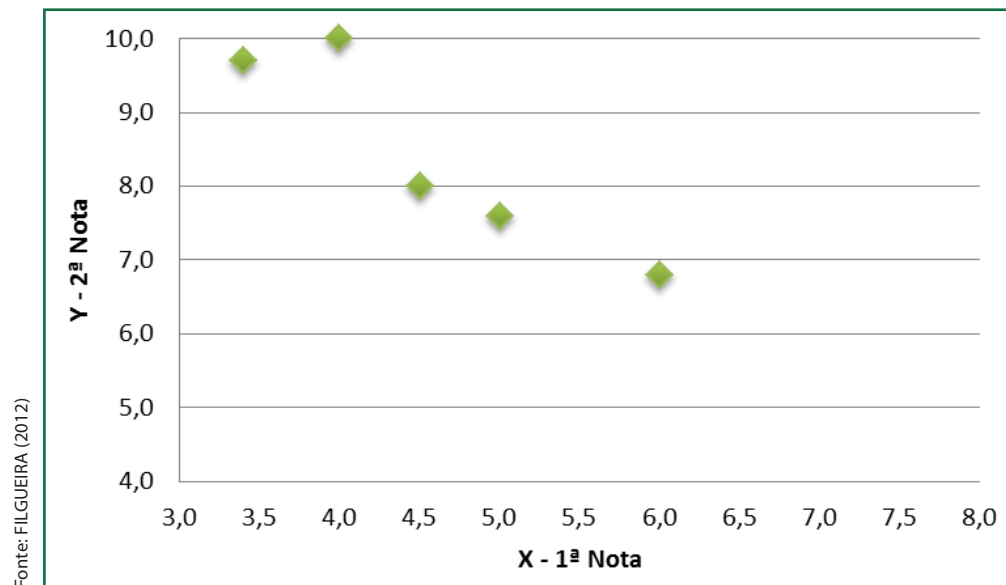


Fig. 05 – Diagrama de Dispersão entre 1ª Nota e 2ª Nota

Pela análise puramente gráfica, percebe-se uma correlação linear inversa. Mas você vai calcular e mensurar o valor real dessa Correlação.

A tabela 4 contém os dados que você precisa para calcular o Coeficiente de Correlação Linear r .

Tabela 4 – Dados para calcular Coeficiente de Correlação Linear

| Aluno | X-1ª Nota | X-2ª Nota | X*Y | X ² | Y ² |
|-------|-----------|-----------|--------|----------------|----------------|
| 1 | 4,50 | 8,00 | 36,00 | 20,25 | 64,00 |
| 2 | 5,00 | 7,60 | 38,00 | 25,00 | 57,76 |
| 3 | 3,40 | 9,70 | 32,00 | 11,56 | 94,09 |
| 4 | 6,00 | 6,80 | 40,80 | 36,00 | 46,24 |
| 5 | 4,00 | 10,00 | 40,00 | 16,00 | 100,00 |
| SOMA | 22,90 | 42,10 | 187,78 | 108,81 | 362,09 |

Fonte: Secretaria Escolar

Apenas uma leitura rápida da tabela 4, para obter o valor $X*Y$ para o Aluno 1, multiplicou-se $4,50*8,00 = 36,00$. Assim também para obter o valor X^2 para o Aluno 3, elevou-se ao quadrado $3,40^2 = 11,56$. Finalmente, para obter o valor Y^2 para o Aluno 5, elevou-se ao quadrado $10,00^2 = 100,00$. Com esse mesmo procedimento você obterá todos os demais valores $X*Y$, X^2 e Y^2 .

De posse das parcelas individuais da fórmula (10.1), apresentadas a seguir, é possível calcular o valor final de r .

Covariância de X e Y:

$$\begin{aligned} n \sum (X * Y) - (\sum X) * (\sum Y) \\ = 5 * 187,78 - (22,90 * 42,10) \\ = 938,9 - 964,09 = -25,19 \end{aligned}$$

Desvio-padrão de X:

$$\begin{aligned} \sqrt{(n * (\sum X^2) - (\sum X)^2)} &= \sqrt{(5 * 108,81) - (22,90)^2} \\ \sqrt{544,05 - 524,41} &= \sqrt{19,64} = 4,43 \end{aligned}$$

Desvio-padrão de Y:

$$\begin{aligned} \sqrt{(n * (\sum Y^2) - (\sum Y)^2)} &= \sqrt{(5 * 362,09) - (42,10)^2} \\ \sqrt{1810,45 - 1772,41} &= \sqrt{38,04} = 6,17 \end{aligned}$$

Portanto, o valor final de r é:

$$(10.2) \quad r = \frac{-25,19}{(4,43 * 6,17)} = \frac{-25,19}{27,33} = -0,9216$$

Portanto, pelo resultado (10.2) pode-se afirmar que há uma correlação inversa muito forte entre a 1ª Nota e a 2ª Nota. Ou seja, quanto maior for a 1ª Nota, menor tende a ser a 2ª Nota.

Agora, de posse do Coeficiente de Correlação Linear $r = -0,9216$ é possível calcular o Coeficiente de Explicação, o qual expressará o quanto da variável Y pode ser explicado pela variável X.

O Coeficiente de Explicação é calculado pela seguinte equação.

$$(10.3) \quad E = r^2 \cdot 100, \text{ onde}$$

R é o Coeficiente de Correlação Linear;

Portanto, para nosso exemplo das Notas de Estatística, tem-se o Coeficiente de Explicação igual a

$$(10.4) \quad E = r^2 \cdot 100 = (-0,9216)^2 \cdot 100 = 84,93\%$$

Finalmente, pode-se afirmar que 84,93% da variação da 2ª Nota são explicados pela 1ª Nota; apenas 15,07% deve-se a outros fatores.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Coeficiente de Correlação Linear até agora, utilize os dados a seguir e o Diagrama de Dispersão, já construído Aula 9 - Significado da Análise de Correlação Linear - e calcule o Coeficiente de Correlação Linear e o Coeficiente de Explicação.



Tabela 5: Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos.

| MÊS | FATURAMENTO (R\$) | PESSOAS ATENDIDAS |
|----------|-------------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

RESUMINDO



Nessa aula você aprendeu a importância que o Coeficiente de Correlação Linear e o Coeficiente de Explicação têm para analisar a influência que determinadas variáveis independentes X têm sobre variáveis dependentes Y . Aprendeu também a calcular o Coeficiente de Correlação Linear e o Coeficiente de Explicação.

LEITURAS COMPLEMENTARES



No artigo disponível no site http://www.economia-aplicada.ufv.br/revista/pdf/2007/01_Artigo_%206.pdf você encontrará várias aplicações do Coeficiente de Correlação Linear no estudo entre o volume de açúcar exportado pelo porto de Santos e o valor do frete negociado para a distribuição de fertilizante ao cliente final, no Estado de São Paulo. Como resultados, pôde-se perceber que em 2006 há uma pequena correlação, com um valor de $r=-0,31$, o que demonstra que há uma pequena relação inversa entre Mercado para Exportação de Açúcar e Frete de Fertilizantes, ou seja, quando os valores de Açúcar crescem os do Frete tendem a decrescer.

AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe a importância do Coeficiente de Correlação Linear e do Coeficiente de Explicação, considere a tabela 6, contendo os valores de testes realizados junto a funcionários de uma indústria antes e depois de um treinamento de requalificação. Utilize os dados dessa tabela 6 e do Diagrama de Dispersão, já construído na Aula 9 - Significado da Análise de Correlação Linear - e calcule o Coeficiente de Correlação Linear e o Coeficiente de Explicação. Os valores dos testes são de zero a dez.



Tabela 6: Testes antes e após treinamento de requalificação.

| Funcionário | Teste antes | Teste após |
|-------------|-------------|------------|
| 1 | 8,4 | 7,1 |
| 2 | 4,6 | 1,9 |
| 3 | 2,5 | 9,7 |
| 4 | 9,8 | 7,4 |
| 5 | 1,5 | 7,1 |
| 6 | 7,4 | 5,7 |
| 7 | 2,0 | 5,9 |
| 8 | 7,0 | 6,0 |
| 9 | 0,1 | 9,7 |
| 10 | 3,9 | 7,0 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 11
REGRESSÃO LINEAR

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL

APRESENTANDO A AULA



Você estudou na aula 10 - Coeficiente de Correlação Linear - como mensurar o grau de Correlação entre X e Y, calculou também o Coeficiente de Explicação e aprendeu que eles permitem identificar se uma variável X pode explicar a variação de uma variável Y, de tal modo que quanto maior for o Coeficiente de Explicação, maior será a importância da variável X na previsão da variável Y. Nesta aula, você vai estudar modelos que permitirão construir uma equação que descreva estatisticamente como Y e X se relacionam. Assim, será possível prever a variável Y utilizando-se valores da variável X; por exemplo, se X representar o investimento em saúde de seu município e Y representar o número de casos de dengue, você pode estudar X e Y, de modo que seja possível prever quantos casos de dengue ocorreriam quando fosse investido determinado valor em saúde. Haveria uma série de exemplos que poderiam ser apresentados aqui, mas o fato é que após calcular o **Coeficiente de Correlação** e o **Coeficiente de Explicação**, sendo eles altos, é possível utilizar modelos estatísticos que expressem a relação entre Y e X e assim poder **prever valores de Y a partir de valores de X**.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender a importância da Regressão Linear;
- calcular os Coeficientes da Regressão Linear;
- calcular previsões para valores de Y a partir de valores de X.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO A REGRESSÃO LINEAR

Na aula 10 - Coeficiente de Correlação Linear - você aprendeu a mensurar o grau de correlação e o de explicação entre duas variáveis, X e Y, você já sabe que quanto maior for o grau de explicação entre X e Y maior será a importância que a variável X tem na explicação da variação da variável Y.

Ademais, você aprendeu que se a explicação for muito alta, perto de 100%, apenas a variável X é suficiente para explicar a variação de Y. Isto é um aspecto muito importante, e você não deve esquecer.

Então, sempre que o grau de explicação for alto, a Estatística garante que há um modelo que expressa a relação entre X e Y. Esse modelo é uma função que pode ser obtida com a utilização de determinadas técnicas que você vai aprender nesta aula.

A técnica de obter essa função chama-se Regressão, pois ela permite **obter todos os valores da variável Y, a partir da variável X**, sempre que o grau de explicação for alto. Na obtenção da função é importante também observar o comportamento do diagrama de dispersão, pois ele é quem vai permitir identificar qual seria a função mais apropriada para explicar a Correlação entre X e Y.

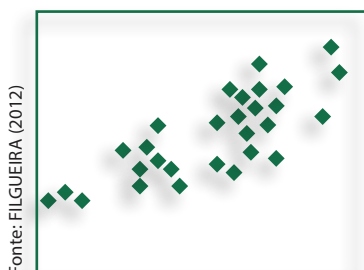


Fig. 01 – Diagrama de Dispersão Correlação $r = 0,7$ e Explicação $E = 49\%$

Por exemplo, na figura 1, o diagrama de dispersão está apresentando uma razoável correlação entre X e Y, e o Coeficiente de Correlação obtido foi $r = 0,7$, o que implica um Coeficiente de Explicação $E = 100*(0,7)^2 = 49\%$.

Analisando a figura 1, é possível identificar que há um comportamento linear entre X e Y, pois o gráfico permite concluir que uma reta passa entre os pontos. Deste modo, além de considerar a Correlação e a Explicação, o Diagrama

de Dispersão permite concluir que entre os pontos passa uma reta tipo $Y = a + b * X$, conforme se observa na Figura 2. E essa é uma função que você certamente já conhece de outras aplicações matemáticas.

A obtenção dos parâmetros **a** e **b** dessa função é o objetivo da Regressão. Nesse caso específico, quando a função a ser obtida é uma reta, diz-se que a **Regressão é Linear**.

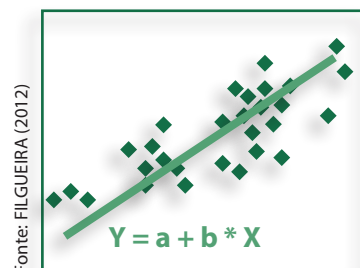


Fig. 02 – Regressão Linear entre X e Y

ATIVIDADE 01



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Regressão Linear até agora, utilize os dados a seguir e o Diagrama de Dispersão já construído na aula 9 - Significado da Análise de Correlação Linear - e comente se há uma reta $Y = a + b * X$, passando pelos pontos (X,Y) .

Tabela 1: Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

CALCULANDO OS COEFICIENTES DA REGRESSÃO LINEAR

A partir de agora você vai utilizar a reta de Regressão $Y = a + b * X$, para representar o comportamento das variáveis X e Y, quando o Coeficiente de Explicação for alto. Portanto, você vai aprender a calcular os parâmetros **a** e **b** dessa reta.

Para calcular o valor do **parâmetro b**, que é denominado coeficiente angular da reta, você vai utilizar a seguinte fórmula.

$$(11.1) \quad b = \frac{n \sum (X * Y) - (\sum X) * (\sum Y)}{n * (\sum X^2) - (\sum X)^2}$$

Onde:

X é a variável independente;

Y é a variável dependente;

n é o número de pares de dados.

Já para calcular o valor do parâmetro **a**, que é denominado coeficiente linear da reta, você vai utilizar a seguinte fórmula.

$$(11.2) \quad a = \bar{Y} - b * \bar{X}$$

Onde:

$$\bar{Y} = \frac{\sum y}{n} \quad \text{é a média de Y;}$$

$$\bar{X} = \frac{\sum x}{n} \quad \text{é a média de X;}$$

n é o número de pares de dados;

b é o coeficiente angular, obtido pela fórmula (11.1);

Você já deve estar querendo calcular os parâmetros da reta de regressão. Considere, por exemplo, a tabela 2, que contém notas de Estatística de alunos de uma turma de Gestão Ambiental, no primeiro e no segundo períodos.

Tabela 2 – Notas do 1º e 2º período de Estatística

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8,0 |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Na aula 10 - Coeficiente de Correlação Linear -, você deve recordar, foram calculados o Coeficiente de Correlação Linear e o Coeficiente de Explicação, além de ser mostrado o comportamento das variáveis X e Y em um Diagrama de Dispersão. Os valores foram: o Coeficiente de Correlação Linear $r = -0,9216$; o Coeficiente de Explicação $E = 84,93\%$.

Portanto, há uma alta explicação entre a 1ª e a 2ª nota de Estatística. O que permitirá construir um modelo de regressão para essas variáveis.

Para isso, você precisará utilizar a tabela 3, que já foi construída na aula 10 - Coeficiente de Correlação Linear.

Tabela 3 – Dados para calcular Coeficiente de Correlação Linear

| Aluno | X-1ª Nota | X-2ª Nota | X*Y | X ² | Y ² |
|-------|-----------|-----------|--------|----------------|----------------|
| 1 | 4,50 | 8,00 | 36,00 | 20,25 | 64,00 |
| 2 | 5,00 | 7,60 | 38,00 | 25,00 | 57,76 |
| 3 | 3,40 | 9,70 | 32,00 | 11,56 | 94,09 |
| 4 | 6,00 | 6,80 | 40,80 | 36,00 | 46,24 |
| 5 | 4,00 | 10,00 | 40,00 | 16,00 | 100,00 |
| SOMA | 22,90 | 42,10 | 187,78 | 108,81 | 362,09 |

Fonte: Secretaria Escolar

Inicialmente você deve calcular b. De posse das parcelas individuais da fórmula (11.1), apresentadas a seguir, é possível calcular o valor final de **b**.

$$n \sum (X * Y) - (\sum X) * (\sum Y) = 5 * 187,78 - (22,90 * 42,10)$$

$$938,9 - 964,09 = -25,19$$

$$\left(n * (\sum X^2) - (\sum X)^2 \right) = (5 * 108,81) - (22,90)^2$$

$$544,05 - 524,41 = 19,64$$

Portanto, o valor final de **b** é:

$$(11.3) \quad b = \frac{-25,19}{19,64} = -1,282587 \cong -1,28$$

Agora você deve calcular a. De posse das parcelas individuais da fórmula (11.2), apresentadas a seguir, é possível calcular o valor final de **a**.

$$\bar{Y} = \frac{\sum y}{n} = \frac{42,10}{5} = 8,42$$

$$\bar{X} = \frac{\sum x}{n} = \frac{22,90}{5} = 4,58$$

(11.4)

$$a = \bar{Y} - b * \bar{X} = 8,42 - (-1,28) * 4,58 = 8,42 + 5,87 = 14,29$$

Finalmente, a reta de regressão que expressa a relação entre X e Y é:

$$\bar{Y} = 14,29 - 1,28 X.$$

A partir de agora, é possível prever valores para Y-2ª Nota, com base em X-1ª Nota; por exemplo, caso um aluno tenha obtido, na 1ª Unidade, nota 8,7, a

partir da reta:

$$\bar{Y} = 14,29 - 1,28 X.$$

Pode-se prever que ele terá na 2ª Unidade uma nota:

$$\bar{Y} = 14,29 - (1,28 * 8,7) = 14,29 - 11,14 = 3,15.$$

E assim, poderiam ser previstos vários valores para a variável Y, a partir da variável X.

ATIVIDADE 02



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Coeficientes de Regressão Linear até agora, utilize os dados a seguir, o Coeficiente de Correlação Linear e o Coeficiente de Explicação, já calculados na aula 10 - Coeficiente de Correlação Linear -, para obter a reta de regressão $Y = a + b * X$.

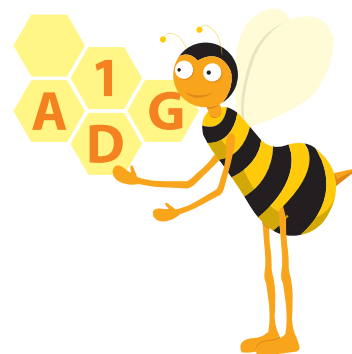
Tabela 4: Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos.

| MÊS | FATURAMENTO (R\$) | PESSOAS ATENDIDAS |
|----------|-------------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

RESUMINDO

Nesta aula você aprendeu a importância que a Regressão Linear e seus parâmetros a e b têm para prever valores da variável Y a partir de valores da variável X . Aprendeu também a calcular esses parâmetros, e utilizou-os para calcular previsões de valores para Y .



LEITURAS COMPLEMENTARES

No artigo disponível no site http://www.economia-aplicada.ufv.br/revista/pdf/2007/01_Artigo_%206.pdf você encontrará, na seção Ajuste Linear, o modelo de regressão linear entre X -Quantidade exportada de açúcar e Y -Valor do frete rodoviário. Esse modelo permite prever quanto seria o Y -Valor do frete, a partir do X -Quantidade exportada.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe a importância do Modelo de Regressão Linear, considere a tabela 5, contendo os valores de testes realizados junto a funcionários de uma indústria antes e depois de um treinamento de requalificação.

Utilize os dados dessa tabela 5, Coeficiente de Correlação Linear e o Coeficiente de Explicação, obtidos na aula 10 - Coeficiente de Correlação Linear-, para obter a reta de regressão $Y = a + b * X$. Os valores dos testes são de zero a dez.



Tabela 5: Testes antes e após treinamento de requalificação.

| Funcionário | Teste antes | Teste após |
|-------------|-------------|------------|
| 1 | 8,4 | 7,1 |
| 2 | 4,6 | 1,9 |
| 3 | 2,5 | 9,7 |
| 4 | 9,8 | 7,4 |
| 5 | 1,5 | 7,1 |
| 6 | 7,4 | 5,7 |
| 7 | 2,0 | 5,9 |
| 8 | 7,0 | 6,0 |
| 9 | 0,1 | 9,7 |
| 10 | 3,9 | 7,0 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 12
ERRO PADRÃO DA ESTIMATIVA

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Você estudou na aula 11 - Regressão Linear - como construir um modelo de Regressão Linear para expressar a relação entre as variáveis X e Y , $Y = a + b * X$. Você aprendeu também a utilizar esse modelo para calcular previsões para a variável Y a partir da variável X , ou seja, você aprendeu a obter estimativas para Y . Mas essas estimativas estão sujeitas a um erro, pois os valores obtidos pelo modelo $Y = a + b * X$, podem ser um pouco maiores ou menores que os valores originais de Y , o que exige que toda estimativa seja calculada e também o erro ao qual ela está submetida. Assim, nesta aula você vai estudar como calcular esse erro, que será chamado a partir de agora de Erro Padrão da Estimativa. Com a estimativa e seu erro correspondente, haverá mais segurança quando da utilização de estimativas para fins práticos. Por exemplo, se você estivesse estudando as variáveis X - o investimento em saúde de seu município e Y - o número de casos de dengue, você poderia estimar quantos casos de dengue ocorreriam quando fosse investido determinado valor em saúde, e ainda afirmaria qual o erro dessa estimativa, o que proporcionaria um resultado bastante seguro quanto à veracidade da estimativa.

DEFININDO OBJETIVOS

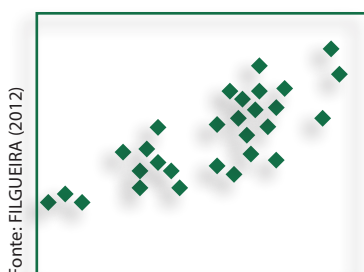
Ao final desta aula, você deverá ser capaz de:

- compreender a importância do Erro Padrão da Estimativa;
- calcular o Erro Padrão da Estimativa;
- construir intervalos para as estimativas, corrigindo-as com o Erro Padrão da Estimativa.

DESENVOLVENDO O CONTEÚDO

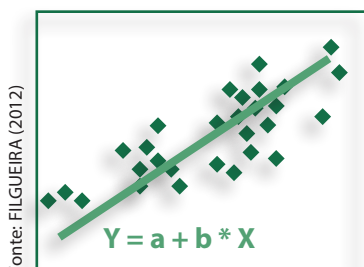
ENTENDENDO O ERRO PADRÃO DA ESTIMATIVA

Na aula 11 - Regressão Linear - você aprendeu a construir um modelo de Regressão Linear para expressar a relação entre as variáveis X e Y, $Y = a + b * X$, e observou que esse modelo expressa, na forma de uma reta, a relação entre as variáveis X e Y. Assim, sob esse modelo, é possível obter estimativas de Y a partir de X.



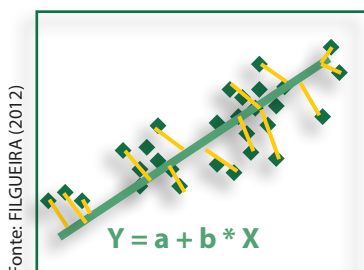
Fonte: FILGUEIRA (2012)

Fig. 01 - Diagrama de Dispersão entre X e Y.



Fonte: FILGUEIRA (2012)

Fig. 02 - Regressão Linear entre X e Y



Fonte: FILGUEIRA (2012)

Fig. 03 - Erros entre valores originais e os obtidos da regressão $Y = a + b * X$

Considere, por exemplo, a figura 1 representando um diagrama de dispersão. Ele está apresentando uma razoável correlação entre X e Y, o que permitiria afirmar que passa uma reta $Y = a + b * X$ por entre os pontos; sendo essa reta um Modelo de Regressão Linear.

Ou seja, o Diagrama de Dispersão da Figura 1 permite concluir que entre os pontos passa uma reta tipo $Y = a + b * X$, conforme se observa na Figura 2.

Conforme é possível observar na Figura 3, a reta de regressão $Y = a + b * X$ está passando entre os valores originais, e portanto há **erros** entre os valores originais e os valores da reta de regressão. E são justamente esses erros que precisam ser mensurados, pois eles definirão o Erro Padrão da Estimativa, conforme você vai calcular mais adiante nesta aula.

ATIVIDADE 01



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Erro Padrão da Estimativa até agora, utilize os dados a seguir e o Diagrama de Dispersão já construído na aula 9 - Significado da Análise de Correlação Linear - e trace uma reta entre os pontos, identificando os erros entre os valores originais e os valores da reta $Y = a + b * X$.

Tabela 1- Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

CALCULANDO O ERRO PADRÃO DA ESTIMATIVA

Agora você vai aprender a calcular o Erro Padrão da Estimativa, a partir a reta de Regressão $Y = a + b * X$ e dos valores originais da variável Y. A rigor, o que você vai calcular é o erro entre cada par (Y original, Y previsto).

Você vai utilizar a seguinte fórmula.

$$(12.1) \quad Se = \sqrt{\frac{\sum (Y - Y_p)^2}{n - 2}}$$

Y representa os valores originais da variável Y;

Y_p representa os valores previstos obtidos da reta $Y = a + b * X$;

n é o número de pares de dados.

Você já deve estar querendo calcular o Erro Padrão Da Estimativa. Considere, por exemplo, a tabela 2 que contém notas de Estatística dos alunos de uma turma de Gestão Ambiental no primeiro e no segundo períodos.

Tabela 2 – Notas do 1º e 2º período de Estatística

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8,0 |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Na aula 11 - Regressão Linear -, você deve recordar, foram calculados, para esse exemplo, os parâmetros **a** e **b** da reta de regressão $Y = a + b * X$, e a reta obtida foi $Y = 14,29 - 1,28 * X$.

Considerando que a fórmula (12.1) calcula a diferença entre cada par de Y e Y_p, é necessário calcular cada um dos Y_p. E esses valores serão calculados agora, na Tabela 3.

Tabela 3 – Cálculo de cada Y_p, utilizando a reta $Y = 14,29 - 1,28 * X$

| Aluno | 1ª Nota | 2ª Nota | $Y_p = 14,29 - 1,28 * X$ |
|-------|---------|---------|--|
| 1 | 4,5 | 8,0 | $Y_p = 14,29 - 1,28 * 4,5 = 14,29 - 5,76 = 8,53$ |
| 2 | 5,0 | 7,6 | 7,89 |
| 3 | 3,4 | 9,7 | 9,94 |
| 4 | 6,0 | 6,8 | 6,61 |
| 5 | 4,0 | 10,0 | 9,17 |
| SOMA | 22,9 | 42,1 | 42,14 |

Fonte: Secretaria Escolar

Como você pode observar na Tabela 3, para obter o primeiro Y_p , foi utilizado o primeiro valor de X , 4,5 na reta $Y = 14,29 - 1,28 * 4,5$, obtendo-se $Y_p = 8,53$. Repetindo o mesmo método para os demais valores, completou-se a Tabela 3.

De posse dos valores dos Y previstos, os Y_p , na Tabela 3 pode-se calcular os erros entre cada Y e cada Y_p , ou seja, pode-se calcular $(Y - Y_p)^2$. Esses valores estão na tabela 4.

Tabela 4 – Cálculo de cada erro $(Y - Y_p)^2$

| Aluno | Y | Y _p | (Y - Y _p) ² |
|-------|------|----------------|------------------------------------|
| 1 | 8,0 | 8,53 | 0,2809 |
| 2 | 7,6 | 7,89 | 0,0841 |
| 3 | 9,7 | 9,94 | 0,0576 |
| 4 | 6,8 | 6,61 | 0,0361 |
| 5 | 10,0 | 9,17 | 0,6889 |
| SOMA | 42,1 | 42,14 | 1,1476 |

Fonte: Secretaria Escolar

Portanto, o valor final de **Se** é:

(12.2)

$$Se = \sqrt{\frac{\sum (Y - Y_p)^2}{n - 2}} = \sqrt{\frac{1,1476}{5 - 2}} = \sqrt{0,382533} = 0,618 \cong 0,62$$

Assim, cada previsão da variável Y estará sujeita a um Erro Padrão em torno de **0,62**.

Por exemplo, caso um aluno tenha obtido, na 1ª Unidade, nota 8,7, a partir da reta $Y = 14,29 - 1,28 * X$, pode-se prever que ele terá, na 2ª Unidade, uma nota $Y = 14,29 - 1,28 * 8,7 = 14,29 - 11,14 = 3,15$. Essa previsão está sujeita a um erro de **0,62**, de modo que com esses valores pode-se construir um intervalo para essa previsão, corrigindo a nota prevista 3,15 com o erro 0,62, para menos e para mais.

Conforme você verá a seguir, a nota prevista na 2ª unidade para esse aluno poderia variar de 2,53 a 3,77.

$$[3,15 - 0,62; 3,15 + 0,62] = [2,53 ; 3,77]$$

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Erro Padrão da Estimativa até agora, utilize os dados a seguir e a Reta de Regressão, já calculada na aula 11 - Regressão Linear - para obter o Erro Padrão da Estimativa.

Tabela 5 - Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos.

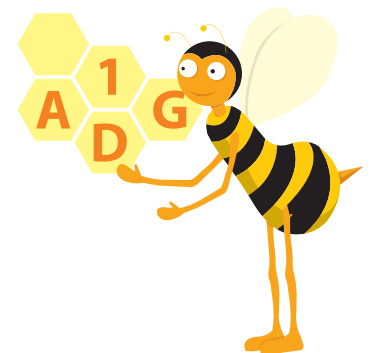
| MÊS | FATURAMENTO (R\$) | PESSOAS ATENDIDAS |
|----------|-------------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.



RESUMINDO

Nesta aula você aprendeu a importância do Erro Padrão da Estimativa para corrigir os valores previstos para menos e para mais. Aprendeu também a calcular o Erro Padrão da Estimativa e a construir um intervalo para as previsões de valores para Y.





LEITURAS COMPLEMENTARES

O artigo disponível em http://media.wix.com/ugd/5ce39e_ea3429bca2691b0bb1e0c5752f70b7fe.pdf estuda modelos de regressão que expressem o gasto energético da caminhada. Você observará na Tabela 5 vários modelos de regressão, e concluirá que os modelos que apresentam maior Coeficiente de Explicação (R^2) são os de menor Erro Padrão da Estimativa (EPE), ou seja, do ponto de vista Estatístico são os melhores modelos.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe a importância do Erro Padrão da Estimativa, considere a tabela 6, contendo os valores de testes realizados junto a funcionários de uma indústria antes e depois de um treinamento de requalificação. Utilize os dados dessa tabela 6 e a Reta de Regressão, já calculada na aula 11 - Regressão Linear - para obter o Erro Padrão da Estimativa. Os valores dos testes são de zero a dez.

Tabela 6 - Testes antes e após treinamento de requalificação.

| Funcionário | Teste antes | Teste após |
|-------------|-------------|------------|
| 1 | 8,4 | 7,1 |
| 2 | 4,6 | 1,9 |
| 3 | 2,5 | 9,7 |
| 4 | 9,8 | 7,4 |
| 5 | 1,5 | 7,1 |
| 6 | 7,4 | 5,7 |
| 7 | 2,0 | 5,9 |
| 8 | 7,0 | 6,0 |
| 9 | 0,1 | 9,7 |
| 10 | 3,9 | 7,0 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

DISCIPLINA
ESTATÍSTICA

AULA 13
DISTRIBUIÇÃO DE FREQUÊNCIA COM
SOFTWARE ESTATÍSTICO

AUTOR
JOÃO MARIA FILGUEIRA



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Durante as aulas anteriores, você estudou várias técnicas Estatísticas e aplicou-as em exemplos. A partir de agora, você terá a oportunidade de expandir esse conhecimento utilizando-se de software de apoio Estatístico. Em particular, nesta aula, você utilizará o Software Excel para construir Distribuições de Frequências, seguindo todos os passos já estudados na aula 3 - Distribuição de Frequências. É importante ressaltar que você deve ter domínio do conteúdo da aula 3 para poder ter êxito na aplicação das técnicas daquela aula no Excel. Ademais, você também poderá utilizar o arquivo produzido no Excel para outras situações, em particular para aquelas onde o conjunto de dados é bastante grande.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de:

- compreender a importância do Excel para aplicações Estatísticas;
- construir distribuição de Frequência utilizando o Excel;
- interpretar os resultados obtidos com o Excel.

DESENVOLVENDO O CONTEÚDO

ENTENDENDO A IMPORTÂNCIA DO SOFTWARE EXCEL

Dentre os vários *Softwares* de apoio Estatístico, o mais facilmente disponível e simples de ser utilizado é o Excel, que como você já sabe vem com o Sistema Operacional Windows. Essa ressalva não invalida a utilização de outros *Softwares* de apoio Estatístico, é simplesmente uma constatação da disponibilidade fácil do Excel.

Agora, é preciso compreender alguns aspectos básicos que serão explorados mais tarde nesta aula. Um deles refere-se à entrada de dados neste *software*: para entrar com textos não há restrição, basta digitá-lo na célula e pronto; para entrar com números também não há restrições, basta digitá-los na célula e pronto; já para o caso de ser necessário realizar cálculos, a célula que conterà os cálculos terá que ser iniciada com o símbolo = (igual), seguido da fórmula desejada.

Considere o exemplo apresentado na Tabela 1, sobre notas de alunos, digitada no Excel. Nessa tabela, há várias colunas e várias linhas: as colunas são A, B, C, D, e E, com os respectivos títulos Alunos, Nota 1, Nota 2, Soma e Média; as linhas são 1, 2, 3, 4, 5 e 6, com os respectivos títulos Alunos, Um, Dois, Três, Quatro e Fonte.

Tabela 1 - Exemplo de Notas de alunos

| | A | B | C | D | E |
|---|-------------------------|---------------|---------------|-------------|--------------|
| 1 | Alunos | Nota 1 | Nota 2 | Soma | Média |
| 2 | Um | 70 | 60 | = B2+C2 | =D2/2 |
| 3 | Dois | 50 | 45 | = B3+C3 | =D3/2 |
| 4 | Três | 60 | 84 | = B4+C4 | =D4/2 |
| 5 | Quatro | 70 | 95 | = B5+C5 | =D5/2 |
| 6 | Fonte: Dados primários. | | | | |

Observa-se na Tabela 1 que o aluno Um teve a 1ª Nota igual a 70 e a 2ª Nota igual a 60; a Soma está na coluna da Soma, com a fórmula **=B2+C2**, e Média está

na coluna da Média, com a fórmula **=D2/2**. Como você já entendeu, a Soma **=B2+C2** resultará em **=70+60=130**; a Média **=D2/2** resultará em **=130/2=65**.

Ou seja, as fórmulas obtiveram os valores que estavam nas células. Por exemplo, na fórmula **=B2+C2** foi obtido o valor de B2, que é 70 e o valor de C2, que é 60; e devido ao símbolo =, o Excel entende que esta célula é uma fórmula, e ela valerá, portanto, **=70+60=130**.

No caso da Média, foi utilizada a fórmula **=D2/2**, como na célula D2 está o valor 130, a fórmula resultará em **=130/2=65**.

Com este mesmo procedimento, você certamente obterá a Tabela 2, contendo os valores de todas as Somas e de todas as Médias.

Tabela 2 – Todas as Somas e todas as Médias

| | A | B | C | D | E |
|---|-------------------------|---------------|---------------|-------------|--------------|
| 1 | Alunos | Nota 1 | Nota 2 | Soma | Média |
| 2 | Um | 70 | 60 | 130 | 65 |
| 3 | Dois | 50 | 45 | 95 | 47,5 |
| 4 | Três | 60 | 84 | 144 | 72 |
| 5 | Quatro | 70 | 95 | 165 | 82,5 |
| 6 | Fonte: Dados primários. | | | | |

Portanto, para realizar cálculos no Excel você precisará identificar as células, referenciando-as por coluna e linha; e utilizar fórmulas iniciando-as nas células com o símbolo = (igual), seguindo-se da fórmula propriamente dita.

Em relação à Apresentação Gráfica, o Excel tem uma funcionalidade muito grande e simplicidade de uso. Há várias formas de construir gráficos com o Excel, vamos apresentar sucintamente uma forma de construí-los.

Considere a Tabela 3, com as notas e médias de alunos. Vamos apresentar em um gráfico as notas de cada aluno.

Inicialmente você deve **selecionar as colunas Alunos, Nota 1 e Nota 2**, conforme a **seleção** apresentada na Tabela 3.

Tabela 3 – Notas e Médias de Alunos

| | A | B | C | D | E |
|---|-------------------------|---------------|---------------|-------------|--------------|
| 1 | Alunos | Nota 1 | Nota 2 | Soma | Média |
| 2 | Um | 70 | 60 | 130 | 65 |
| 3 | Dois | 50 | 45 | 95 | 47,5 |
| 4 | Três | 60 | 84 | 144 | 72 |
| 5 | Quatro | 70 | 95 | 165 | 82,5 |
| 6 | Fonte: Dados primários. | | | | |

Após selecionar, conforme a Tabela 3, você vai clicar na Guia Inserir, e clicar na opção Gráfico Linhas. Conforme a Figura 1 - Selecionando Gráfico Linhas.

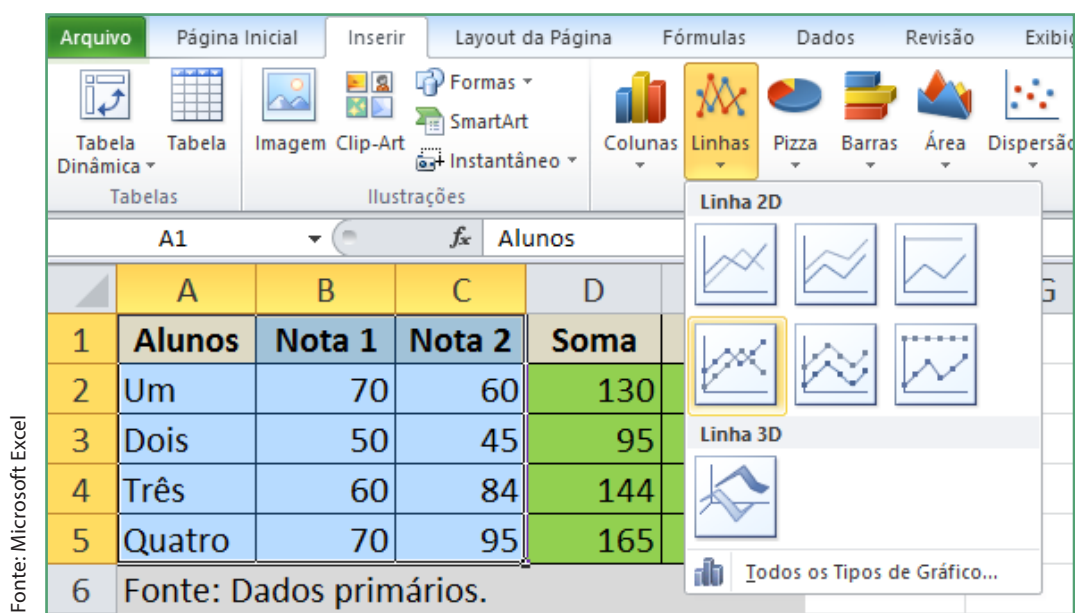


Fig. 01 – Selecionando Gráfico Linhas.

Após selecionar entre as opções do Gráfico de Linhas, surgirá o gráfico já construído. Conforme a Figura 2 - Gráfico com as notas dos Alunos. Nele você perceberá claramente o comportamento das notas dos alunos com a primeira e a segunda nota.

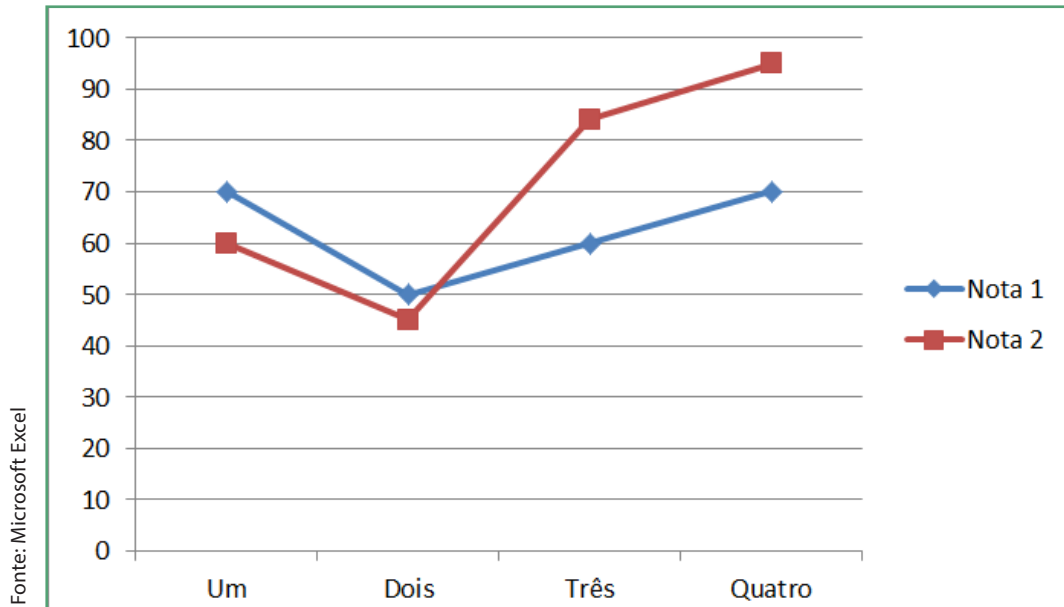


Fig. 02 - Gráfico com as notas dos Alunos.

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Distribuição de Frequência com Software Estatístico até agora, utilize os dados da Tabela 4, a seguir, e calcule no Excel a média do Faturamento e das Pessoas Atendidas.

Tabela 1- Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.



CONSTRUINDO DISTRIBUIÇÃO DE FREQUÊNCIA COM EXCEL

Conforme você estudou na aula 3 - Distribuição de Frequências -, para construir uma distribuição de frequências, será necessário aplicar as seguintes técnicas:

- Identificar o número de dados – **n**;
- Colocar o conjunto dos dados em ordem – esse conjunto ordenado recebe o nome de **ROL**;

- Calcular a amplitude total – **AT**

$$\mathbf{AT = MaiorValor - MenorValor;}$$

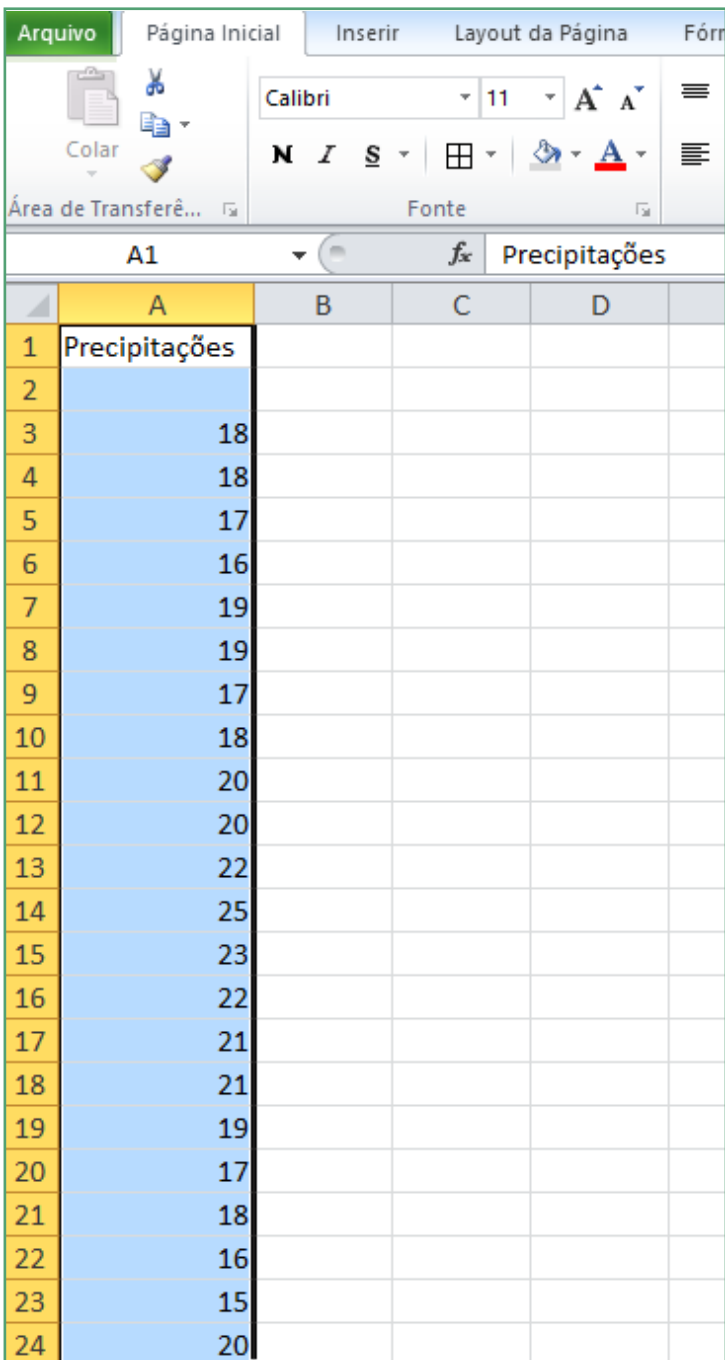
- Calcular o número de classes – **c**

$$c = \begin{cases} 5, & \text{se o número de dados } n \leq 25; \\ \sqrt{n}, & \text{se o número de dados } n > 25. \end{cases}$$

- Calcular a amplitude de classe – **a = AT/c**;
- Construir cada classe, identificando seus **Limites Inferior e Superior**;
- Obter as **Frequências** de cada classe, observando o ROL e os Limites de cada classe;
- Obter as **Porcentagens** de cada classe, observando suas respectivas frequências.

Então, considerando o exemplo das Precipitações Pluviométricas, cujos dados estão apresentados a seguir, você precisará aplicar as técnicas já citadas anteriormente. As precipitações pluviométricas (mm) diárias são: 18, 18, 17, 16, 19, 19, 17, 18, 20, 20, 22, 25, 23, 22, 21, 21, 19, 17, 18, 16, 15, 20.

Inicialmente você precisa digitar esses dados no Excel. Utilize a coluna A linha 1 para digitá-las (Veja a Figura 3).



Fonte: Microsoft Excel

| | A | B | C | D |
|----|---------------|---|---|---|
| 1 | Precipitações | | | |
| 2 | | | | |
| 3 | 18 | | | |
| 4 | 18 | | | |
| 5 | 17 | | | |
| 6 | 16 | | | |
| 7 | 19 | | | |
| 8 | 19 | | | |
| 9 | 17 | | | |
| 10 | 18 | | | |
| 11 | 20 | | | |
| 12 | 20 | | | |
| 13 | 22 | | | |
| 14 | 25 | | | |
| 15 | 23 | | | |
| 16 | 22 | | | |
| 17 | 21 | | | |
| 18 | 21 | | | |
| 19 | 19 | | | |
| 20 | 17 | | | |
| 21 | 18 | | | |
| 22 | 16 | | | |
| 23 | 15 | | | |
| 24 | 20 | | | |

Fig. 03 - Precipitações no Excel.

De posse desses dados, você precisará obter o Rol. Para isto, veja na Figura 4, na coluna Rol, todos os dados que estão copiados e selecionados e, com o mouse, foi selecionado a Guia Dados com a opção Classificar.

Fonte: Microsoft Excel

The screenshot shows the Microsoft Excel interface. The 'Dados' ribbon is active, and the 'Classificar' button is highlighted. The spreadsheet contains the following data:

| | A | B | C | D | E | F |
|----|---------------|-----|---|---|---|---|
| 1 | Precipitações | Rol | | | | |
| 2 | | | | | | |
| 3 | 18 | 18 | | | | |
| 4 | 18 | 18 | | | | |
| 5 | 17 | 17 | | | | |
| 6 | 16 | 16 | | | | |
| 7 | 19 | 19 | | | | |
| 8 | 19 | 19 | | | | |
| 9 | 17 | 17 | | | | |
| 10 | 18 | 18 | | | | |
| 11 | 20 | 20 | | | | |
| 12 | 20 | 20 | | | | |
| 13 | 22 | 22 | | | | |
| 14 | 25 | 25 | | | | |
| 15 | 23 | 23 | | | | |
| 16 | 22 | 22 | | | | |
| 17 | 21 | 21 | | | | |
| 18 | 21 | 21 | | | | |
| 19 | 19 | 19 | | | | |
| 20 | 17 | 17 | | | | |
| 21 | 18 | 18 | | | | |
| 22 | 16 | 16 | | | | |
| 23 | 15 | 15 | | | | |
| 24 | 20 | 20 | | | | |

Fig. 04 - Menu Dados, opção Classificar.

Agora aparecerá o aviso da Figura 5, você escolherá **Continuar com a seleção atual** e clicará em Classificar.

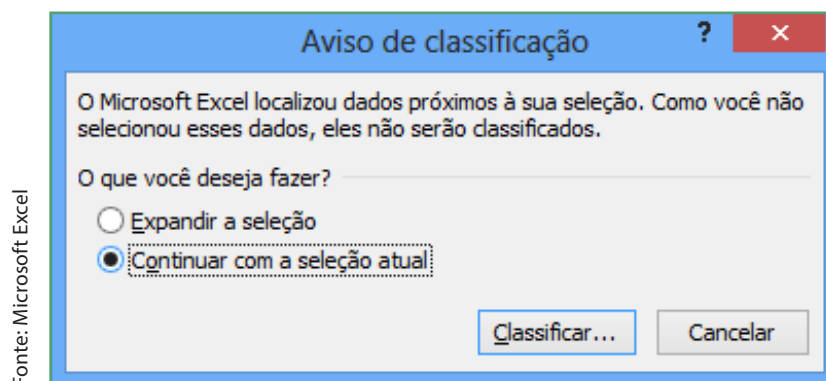


Fig. 05 – Aviso de classificação.

Surgirá uma nova tela, conforme a Figura 6. Nela você irá Classificar por **Colunas B**, e escolherá a opção Crescente, e clicará em OK.

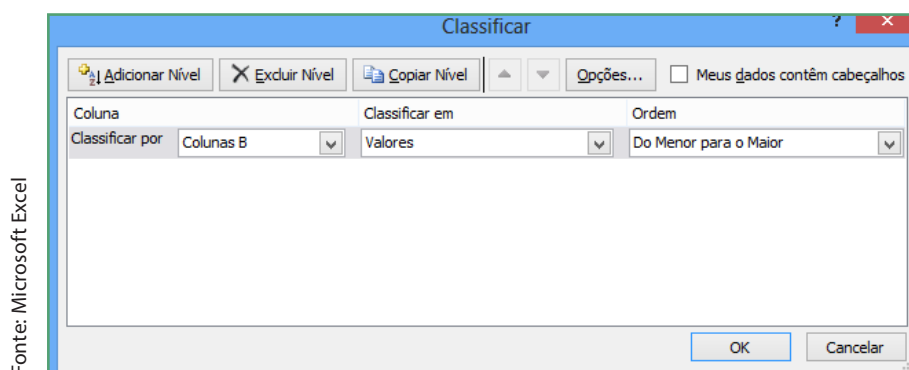


Fig. 06 – Classificar.

A partir de agora, na Coluna B, você terá o Rol finalizado, conforme você pode ver na Figura 7.

Fonte: Microsoft Excel

| | A | B | C | D |
|----|---------------|-----|---|---|
| 1 | Precipitações | Rol | | |
| 2 | | | | |
| 3 | 18 | 15 | | |
| 4 | 18 | 16 | | |
| 5 | 17 | 16 | | |
| 6 | 16 | 17 | | |
| 7 | 19 | 17 | | |
| 8 | 19 | 17 | | |
| 9 | 17 | 18 | | |
| 10 | 18 | 18 | | |
| 11 | 20 | 18 | | |
| 12 | 20 | 18 | | |
| 13 | 22 | 19 | | |
| 14 | 25 | 19 | | |
| 15 | 23 | 19 | | |
| 16 | 22 | 20 | | |
| 17 | 21 | 20 | | |
| 18 | 21 | 20 | | |
| 19 | 19 | 21 | | |
| 20 | 17 | 21 | | |
| 21 | 18 | 22 | | |
| 22 | 16 | 22 | | |
| 23 | 15 | 23 | | |
| 24 | 20 | 25 | | |

Fig. 07 - Rol finalizado.

Após obter o Rol, você vai calcular AT, c, a. No excel, conforme a Figura 8, estão calculados:

- na célula D1=B24-B3, é a fórmula de AT;
- na célula D2=5, é o valor de c;
- na célula D3=D1/D2, é a fórmula de a.

Fonte: Microsoft Excel

| | A | B | C | D |
|----|---------------|-----|----|---------|
| 1 | Precipitações | Rol | AT | =B24-B3 |
| 2 | | | C | 5 |
| 3 | 18 | 15 | a | =D1/D2 |
| 4 | 18 | 16 | | |
| 5 | 17 | 16 | | |
| 6 | 16 | 17 | | |
| 7 | 19 | 17 | | |
| 8 | 19 | 17 | | |
| 9 | 17 | 18 | | |
| 10 | 18 | 18 | | |
| 11 | 20 | 18 | | |
| 12 | 20 | 18 | | |
| 13 | 22 | 19 | | |
| 14 | 25 | 19 | | |
| 15 | 23 | 19 | | |
| 16 | 22 | 20 | | |
| 17 | 21 | 20 | | |
| 18 | 21 | 20 | | |
| 19 | 19 | 21 | | |
| 20 | 17 | 21 | | |
| 21 | 18 | 22 | | |
| 22 | 16 | 22 | | |
| 23 | 15 | 23 | | |
| 24 | 20 | 25 | | |

Fig. 08 - Formulas de AT, c, a.

Naturalmente, os valores finalizados serão:

$$AT = 10; c = 5; a = 2.$$

A partir de agora, você vai construir a Distribuição de Frequências. Na Figura 9, você verá que o primeiro Limite Inferior é o primeiro valor do Rol, que está na célula B3; já o primeiro Limite Superior é o Primeiro Limite Inferior + a, ou seja, será C7+D3. E assim sucessivamente, conforme mostra a Figura 10.

Fonte: Microsoft Excel

| | A | B | C | D | E |
|----|---------------|-----|-----------------|---------|-----------------|
| 1 | Precipitações | Rol | AT | =B24-B3 | |
| 2 | | | C | 5 | |
| 3 | 18 | 15 | a | =D1/D2 | |
| 4 | 18 | 16 | | | |
| 5 | 17 | 16 | | | |
| 6 | 16 | 17 | Limite Inferior | | Limite Superior |
| 7 | 19 | 17 | =B3 | ----- | =C7+D3 |
| 8 | 19 | 17 | =B7 | ----- | =C8+D3 |
| 9 | 17 | 18 | =B13 | ----- | =C9+D3 |
| 10 | 18 | 18 | =B19 | ----- | =C10+D3 |
| 11 | 20 | 18 | =B21 | ----- | =C11+D3 |
| 12 | 20 | 18 | | Soma | |
| 13 | 22 | 19 | | | |
| 14 | 25 | 19 | | | |
| 15 | 23 | 19 | | | |
| 16 | 22 | 20 | | | |
| 17 | 21 | 20 | | | |
| 18 | 21 | 20 | | | |
| 19 | 19 | 21 | | | |
| 20 | 17 | 21 | | | |
| 21 | 18 | 22 | | | |
| 22 | 16 | 22 | | | |
| 23 | 15 | 23 | | | |
| 24 | 20 | 25 | | | |

Fig. 09 - Fórmulas para montar a Distribuição de Frequências.

A distribuição finalizada possuirá os seguintes valores, conforme a Figura 10.

Fonte: Microsoft Excel

| | A | B | C | D | E | F |
|----|---------------|-----|-----------------|-------|-----------------|---|
| 1 | Precipitações | Rol | AT | 10 | | |
| 2 | | | C | 5 | | |
| 3 | 18 | 15 | a | 2 | | |
| 4 | 18 | 16 | | | | |
| 5 | 17 | 16 | | | | |
| 6 | 16 | 17 | Limite Inferior | | Limite Superior | |
| 7 | 19 | 17 | 15 | ----- | 17 | |
| 8 | 19 | 17 | 17 | ----- | 19 | |
| 9 | 17 | 18 | 19 | ----- | 21 | |
| 10 | 18 | 18 | 21 | ----- | 23 | |
| 11 | 20 | 18 | 22 | ----- | 24 | |
| 12 | 20 | 18 | Soma | | | |
| 13 | 22 | 19 | | | | |
| 14 | 25 | 19 | | | | |
| 15 | 23 | 19 | | | | |
| 16 | 22 | 20 | | | | |
| 17 | 21 | 20 | | | | |
| 18 | 21 | 20 | | | | |
| 19 | 19 | 21 | | | | |
| 20 | 17 | 21 | | | | |
| 21 | 18 | 22 | | | | |
| 22 | 16 | 22 | | | | |
| 23 | 15 | 23 | | | | |
| 24 | 20 | 25 | | | | |

Fig. 10 - Distribuição finalizada.

Finalmente, você precisará contar no ROL quantos valores estão em cada classe. Veja, na Figura 11, que os valores do ROL estão marcados de acordo com as frequências das classes.

Fonte: Microsoft Excel

| | A | B | C | D | E | F |
|----|---------------|-----|-----------------|-------|-----------------|----|
| 1 | Precipitações | Rol | AT | 10 | | |
| 2 | | | C | 5 | | |
| 3 | 18 | 15 | a | 2 | | |
| 4 | 18 | 16 | | | | |
| 5 | 17 | 16 | | | | |
| 6 | 16 | 17 | Limite Inferior | | Limite Superior | f |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 |
| 12 | 20 | 18 | Soma | | | 22 |
| 13 | 22 | 19 | | | | |
| 14 | 25 | 19 | | | | |
| 15 | 23 | 19 | | | | |
| 16 | 22 | 20 | | | | |
| 17 | 21 | 20 | | | | |
| 18 | 21 | 20 | | | | |
| 19 | 19 | 21 | | | | |
| 20 | 17 | 21 | | | | |
| 21 | 18 | 22 | | | | |
| 22 | 16 | 22 | | | | |
| 23 | 15 | 23 | | | | |
| 24 | 20 | 25 | | | | |

Fig. 11 - Obtenção das frequências.

Assim, você perceberá que a frequência da primeira classe é 3 e foi obtida contando os valores do ROL que são maiores ou iguais a 15 e menores que 17. Os valores foram **15,16,16** perfazendo uma frequência 3. E assim sucessivamente.

Para complementar a distribuição de frequências, você precisa calcular as porcentagens. Veja na Figura 13, as fórmulas para esse cálculo. Lembre-se que cada porcentagem é obtida pela fórmula:

$$\% = \left(\frac{f}{\sum f} \right) * 100$$

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|-----------------|---------------|-----------------|---------------|---------------|
| 1 | Precipitações | Rol | AT | =B24-B3 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | =D1/D2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % |
| 7 | 19 | 17 | =B3 | ----- | =C7+D3 | 3 | =F7/F12*100 |
| 8 | 19 | 17 | =B7 | ----- | =C8+D3 | 7 | =F8/F12*100 |
| 9 | 17 | 18 | =B13 | ----- | =C9+D3 | 6 | =F9/F12*100 |
| 10 | 18 | 18 | =B19 | ----- | =C10+D3 | 4 | =F10/F12*100 |
| 11 | 20 | 18 | =B21 | ----- | =C11+D3 | 2 | =F11/F12*100 |
| 12 | 20 | 18 | | Soma | | =SOMA(F7:F11) | =SOMA(G7:G11) |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | | | |
| 17 | 21 | 20 | | | | | |
| 18 | 21 | 20 | | | | | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | | | |
| 21 | 18 | 22 | | | | | |
| 22 | 16 | 22 | | | | | |
| 23 | 15 | 23 | | | | | |
| 24 | 20 | 25 | | | | | |

Fig. 12 - Fórmulas para calcular as porcentagens.

Na Figura 13, finalmente, você terá toda a distribuição de frequências, com classes, frequências e porcentagens.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G | H | I |
|----|---------------|-----|-----------------|---------------|-----------------|----|--------|---|---|
| 1 | Precipitações | Rol | AT | 10 | | | | | |
| 2 | | | C | 5 | | | | | |
| 3 | 18 | 15 | a | 2 | | | | | |
| 4 | 18 | 16 | | | | | | | |
| 5 | 17 | 16 | | | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % | | |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 | | |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 | | |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 | | |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 | | |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 | | |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 | | |
| 13 | 22 | 19 | | | | | | | |
| 14 | 25 | 19 | | | | | | | |
| 15 | 23 | 19 | | | | | | | |
| 16 | 22 | 20 | | | | | | | |
| 17 | 21 | 20 | | | | | | | |
| 18 | 21 | 20 | | | | | | | |
| 19 | 19 | 21 | | | | | | | |
| 20 | 17 | 21 | | | | | | | |
| 21 | 18 | 22 | | | | | | | |
| 22 | 16 | 22 | | | | | | | |
| 23 | 15 | 23 | | | | | | | |
| 24 | 20 | 25 | | | | | | | |

Fig. 13 - Distribuição de Frequências.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Identifique exemplos reais de utilização do Excel para realizar cálculos Estatísticos e comente-os. Você perceberá que ele produz resultados rápidos e permite até formatação desses dados, melhorando também a apresentação dos cálculos Estatísticos.





RESUMINDO

Nesta aula você aprendeu a importância de utilizar um *software* de apoio Estatístico, neste caso você utilizou o Excel. Você observou como inserir texto, fórmulas e construir gráficos no Excel. Aprendeu também a praticar nesse *software* o conteúdo de Distribuição de Frequências visto na Aula 3.

LEITURAS COMPLEMENTARES



O artigo <http://ctec.unicruz.edu.br/revista/artigos/31.pdf> apresenta um estudo com o objetivo de analisar a relação entre o processo saúde/doença com os determinantes histórico-sociais da população de um bairro da cidade de Cruz Alta-RS. A análise dos dados foi realizada com o *Software* de apoio Estatístico Excel. Há vários gráficos e tabelas caracterizando os principais resultados e, para cada um deles, há um breve comentário.

AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe utilizar o Excel para construir Distribuição de Frequências, utilize os dados a seguir, referentes a notas de alunos de Estatística para construir uma Distribuição de Frequências no Excel.

Notas: 8, 9, 5, 0, 10, 3, 2, 9, 7, 2, 1, 3, 6, 8, 9, 10, 4, 2, 6, 1, 10.



CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

NEUFELD, John L. **Estatística aplicada à administração usando Excel**. Tradução José Luiz Celeste; revisão técnica Cyro C. Patarra. São Paulo: Prentice Hall, 2003

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

**DISCIPLINA
ESTATÍSTICA**

**AULA 14
MEDIDAS ESTATÍSTICAS COM SOFTWARE ESTATÍSTICO**

**AUTOR
JOÃO MARIA FILGUEIRA**



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Na Aula 13 – Distribuição de Frequências com *Software* Estatístico - você utilizou o *Software* Excel para construir Distribuições de Frequências. Nesta aula você vai continuar utilizando esse *software*, agora para calcular Medidas de Tendência Central, Medidas de Variabilidade. As Medidas de Tendência Central que você utilizará nesta aula são a Média, a Mediana e a Moda. As Medidas de Variabilidade que você utilizará são a Variância, o Desvio Padrão e o Coeficiente de Variação.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de utilizar o Excel para:

- calcular Média, Mediana, Moda;
- calcular Variância, Desvio Padrão, Coeficiente de Variação.

DESENVOLVENDO O CONTEÚDO

CALCULANDO MÉDIA, MEDIANA, MODA NO EXCEL

Para o cálculo da Média, Mediana e Moda, o Excel possui funções específicas. Há também outras funções Estatísticas que o Excel provê, e você poderá aprender a utilizá-las sem muita dificuldade.

Essas funções estão disponíveis na Guia **Fórmulas**, clicando na opção **Inserir Função**, conforme apresentado na Figura 1.

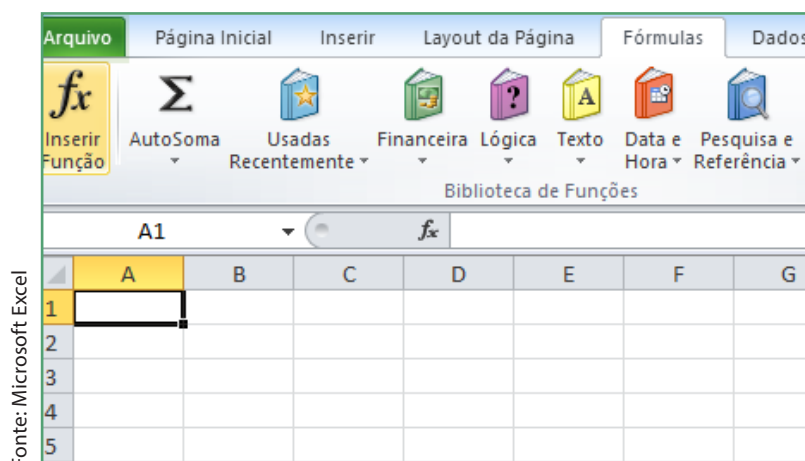


Fig. 01 – Inserir Função.

Após clicar em função, será apresentada uma janela, Inserir Função, nela você vai escolher a categoria Estatística, conforme a Figura 2.

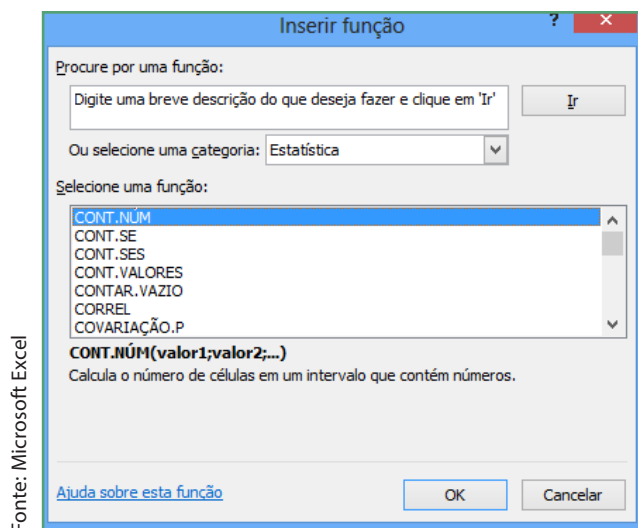


Fig. 02 – Janela Inserir Função.

Você vai utilizar bastante essa janela, pois nela estão as funções dessa aula e as demais que foram citadas anteriormente, as quais você poderá aprender sem dificuldades. Lembre-se sempre dos procedimentos para chegar até essa janela, você vai utilizá-la bastante.

Agora que você já sabe obter as funções Estatísticas no Excel, resta apenas saber os nomes das funções que você vai calcular para utilizá-las adequadamente.

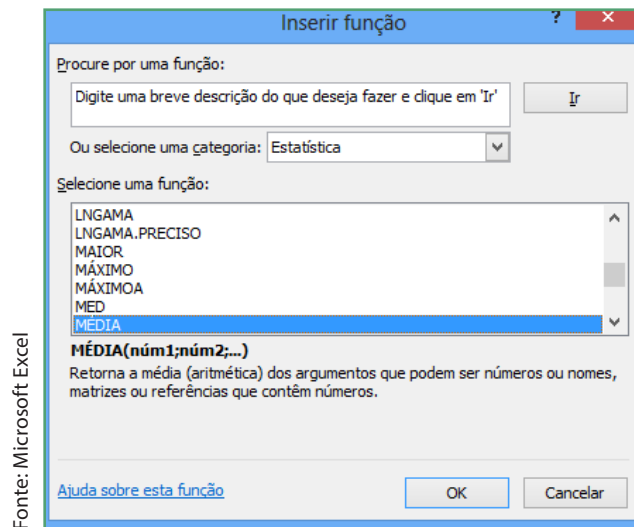
Para calcular a Média, inicialmente, você deve digitar na célula **E16** a palavra **Média**, depois deve ir para a célula **F16**, pois nesta célula será calculada a Média. Veja como ficará sua planilha na Figura 3.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|-----------------|---------------|-----------------|----|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | | |
| 17 | 21 | 20 | | | | | |
| 18 | 21 | 20 | | | | | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | | | |
| 21 | 18 | 22 | | | | | |
| 22 | 16 | 22 | | | | | |
| 23 | 15 | 23 | | | | | |
| 24 | 20 | 25 | | | | | |

Fig. 03- Células E16 e F16.

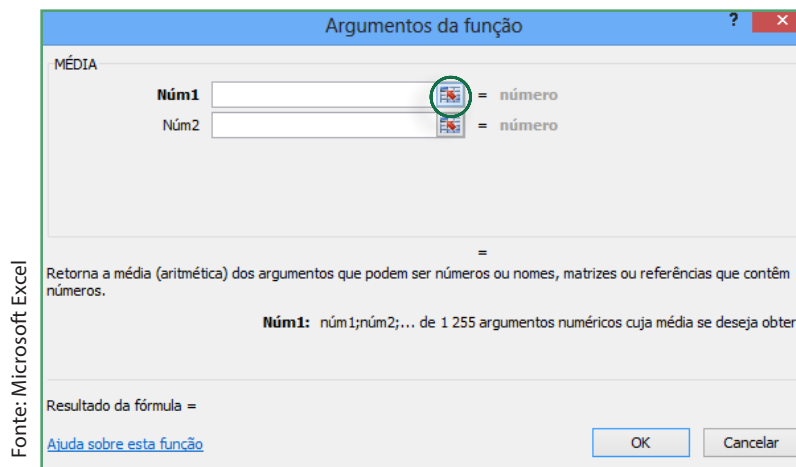
Após definir que a Média será calculada na célula F16, você escolherá a função Média na janela de Inserir Funções, categoria Estatísticas, conforme está apresentado na Figura 4. Você deve, após selecionar a função Média, clicar em OK.



Fonte: Microsoft Excel

Fig. 04 – Função Média.

Após clicar em OK, será apresentada uma janela para você selecionar os dados para os quais você deseja calcular a Média. Essa janela está apresentada na Figura 5.



Fonte: Microsoft Excel

Fig. 05 – Argumentos da Média.

Quando você clicar no Botão indicado, você vai voltar para a planilha e deve selecionar todos os dados do Rol. Conforme a Figura 6, após selecionar os dados do Rol, você deve clicar no Botão indicado para voltar à janela da função Média.

Fonte: Microsoft Excel

| Limite Inferior | Precipitações | Limite Superior | f | % |
|-----------------|---------------|-----------------|----|--------|
| 15 | 3 | 17 | 3 | 13,64 |
| 16 | 7 | 19 | 7 | 31,82 |
| 16 | 6 | 21 | 6 | 27,27 |
| 17 | 4 | 23 | 4 | 18,18 |
| 18 | 2 | 24 | 2 | 9,09 |
| Soma | | | 22 | 100,00 |

The dialog box 'Argumentos da função' shows the range B3:B24 selected. The spreadsheet cell A16 contains the formula =MÉDIA(B3:B24).

Fig. 06 - Selecionar os dados do Rol.

Para concluir você deve clicar em OK, para que o valor da Média seja inserido na Planilha. Na Figura 7 você já percebe o valor da média, basta clicar OK.

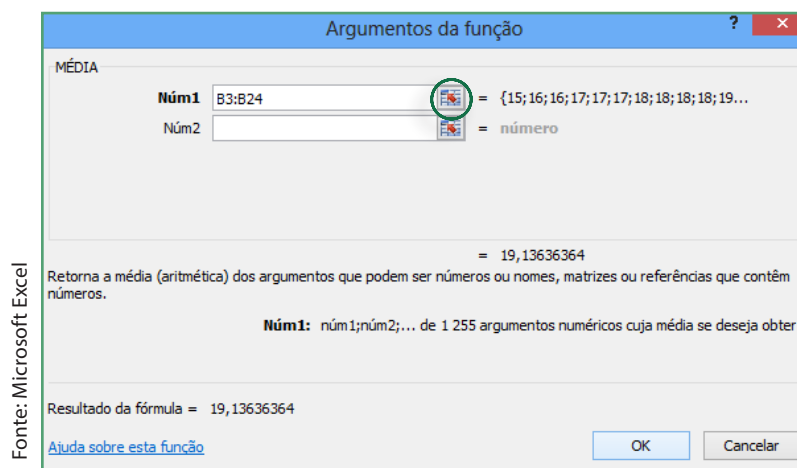


Fig. 07 - Valor da Média.

Finalmente, na célula **F16**, você terá esse valor da Média, conforme está apresentado na Figura 8.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|-----------------|---------------|-----------------|-------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | | | |
| 18 | 21 | 20 | | | | | |

Fig. 08 - Valor final da Média.

Para calcular a Mediana, inicialmente você deve digitar na célula **E17** a palavra **Mediana**, depois deve ir para a célula **F17**, pois nesta célula será calculada a Mediana. Veja como ficará sua planilha na Figura 9.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|-----------------|---------------|-----------------|-------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | | |
| 18 | 21 | 20 | | | | | |

Fig. 09 - Células E17 e F17.

Após definir que a Mediana será calculada na célula **F17**, você escolherá a função Med na janela de funções Estatísticas, conforme está apresentado na Figura 10. Você deve, após selecionar a função Med, clicar em OK.

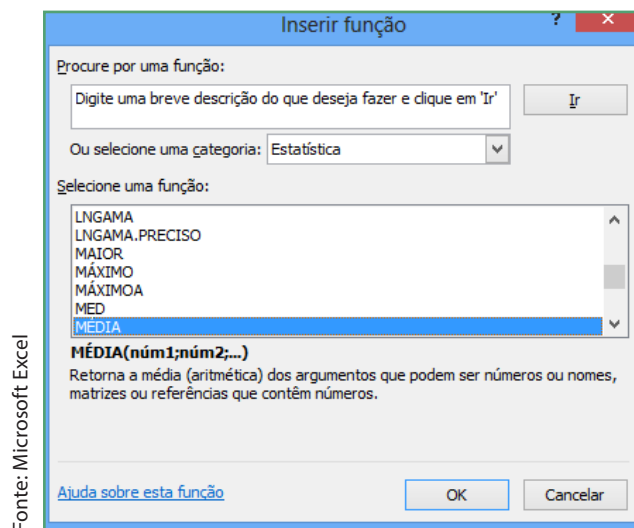


Fig. 10 – Mediana.

Seguindo os mesmos procedimentos que você já utilizou para obter a Média, você agora precisará selecionar todos os dados do Rol e o valor final da Mediana 19 na célula F17, conforme a figura 11.

Fonte: Microsoft Excel

| F17 | | fx | | =MED(B3:B24) | | | |
|-----|---------------|-----|-----------------|---------------|-----------------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | Limite Inferior | Precipitações | Limite Superior | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | | | |

Fig. 11 - Valor Final da Mediana.

Para calcular a Moda, inicialmente, você deve digitar na célula **E18** a palavra **Moda**, depois deve ir para a célula **F18**, pois nesta célula será calculada a Moda. Veja como ficará sua planilha na Figura 12.

Fonte: Microsoft Excel

| F18 | | fx | | | | | |
|-----|---------------|-----|------|---------------|---------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | | |

Fig. 12 - Células E18 e F18.

Após definir que a Moda será calculada na célula **F18**, você escolherá a função Modo na janela de funções Estatísticas, conforme está apresentado na Figura 13. Você deve, após selecionar a função Modo, clicar em OK.

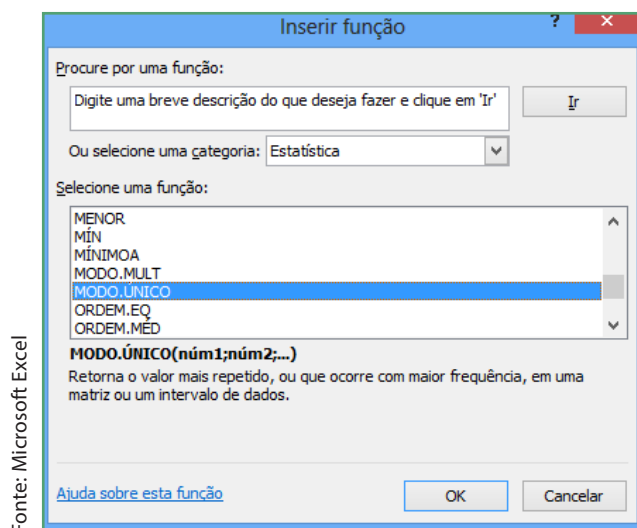


Fig. 13 – Moda.

Seguindo os mesmos procedimentos que você já utilizou para obter a Média e a Mediana, você agora precisará selecionar todos os dados do Rol e o valor final da Moda 18, na célula F18, conforme a figura 14.

Fonte: Microsoft Excel

| F18 | | fx | | =MODO.ÚNICO(B3:B24) | | | |
|-----|---------------|-----|----|---------------------|---------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |

Fig. 14 - Valor Final da Moda.

Portanto, você já calculou a Média, a Mediana e a Moda. Como você pôde perceber, para calcular a Média a função utilizada foi =MÉDIA(B3:B24); para calcular a Mediana a função utilizada foi =MED(B3:B24); finalmente para calcular a Moda a função utilizada foi =MODO(B3:B24), ou =MODO.ÚNICO(B3:B24). Esta é outra forma para calcular os valores: você digitar o nome da função e, entre parêntesis, a primeira e a última células que contêm os valores, as células devem ser separadas por: (dois pontos).

ATIVIDADE 01

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre o cálculo de Medidas Estatísticas com *Software* Estatístico até agora, utilize os dados a seguir, referentes a notas de alunos de Estatística, para calcular Média, Mediana e Moda no Excel.

Notas: 8, 9, 5, 0, 10, 3, 2, 9, 7, 2, 1, 3, 6, 8, 9, 10, 4, 2, 6, 1, 10.



CALCULANDO VARIÂNCIA, DESVIO PADRÃO, COEFICIENTE DE VARIAÇÃO NO EXCEL

Para calcular a Variância, inicialmente, você deve digitar na célula **E20** a palavra **Variância**, depois deve ir para a célula **F20**, pois nesta célula será calculada a Variância. Veja como ficará sua planilha na Figura 15.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|------|---------------|-----------|-------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | | |
| 21 | 18 | 22 | | | | | |

Fig. 15 - Células E20 e F20.

Após definir que a Variância será calculada na célula **F20**, agora você escolherá a função Var na janela de funções Estatísticas, conforme está apresentado na Figura 16. Você deve, após selecionar a função Var, clicar em OK.

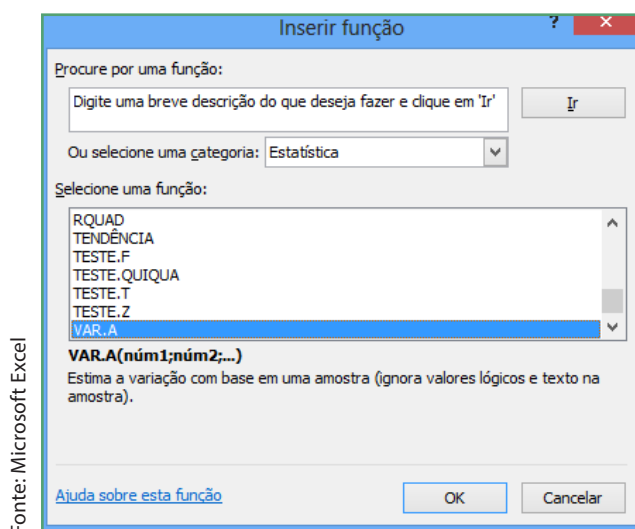


Fig. 16 - Variância.

Seguindo os mesmos procedimentos que você já utilizou para obter a Média, a Mediana e a Moda, você agora precisará selecionar todos os dados do Rol e o valor final da Variância 6,22, na célula F20, conforme a figura 17.

Fonte: Microsoft Excel

| F20 | | fx | | =VAR.A(B3:B24) | | | |
|-----|---------------|-----|----|----------------|-----------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | | | |

Fig. 17 - Valor Final da Variância.

Para calcular o Desvio Padrão, inicialmente, você deve digitar na célula **E21** as palavras **Desvio Padrão**, depois deve ir para a célula **F21**, pois nesta célula será calculada a Variância. Veja como ficará sua planilha na Figura 18.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|----|---------------|---------------|-------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | Desvio Padrão | | |
| 22 | 16 | 22 | | | | | |

Fig. 18 - Células E21 e F21.

Após definir que o Desvio padrão será calculado na célula **F21**, você escolherá a função Desvpad na janela de funções Estatísticas, conforme está apresentado na Figura 19. Você deve, após selecionar a função Desvpad, clicar em OK.

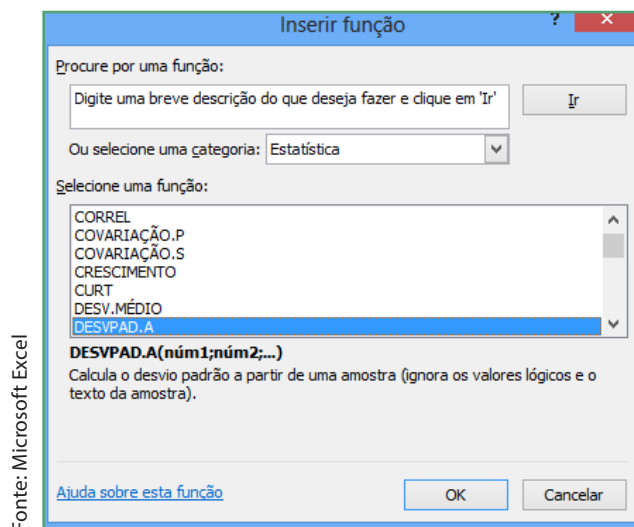


Fig. 19 – Desvio Padrão.

Seguindo os mesmos procedimentos que você já utilizou para obter a Variância, você agora precisará selecionar todos os dados do Rol e o valor final do Desvio Padrão 2,49, na célula F21, conforme a figura 20.

Fonte: Microsoft Excel

| F21 | | fx | | =DESVPAD.A(B3:B24) | | | |
|-----|---------------|-----|------|--------------------|---------------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | Soma | | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | Desvio Padrão | 2,493715044 | |

Fig. 20 - Valor Final do Desvio Padrão.

Para calcular o Coeficiente de Variação, inicialmente, você deve digitar na célula **E22** a palavra **CoefVariação**, depois deve ir para a célula **F22**, pois nesta célula será calculada a Variância. Veja como ficará sua planilha na Figura 21.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|----|-------|---------------|-------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | | Precipitações | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | Desvio Padrão | 2,493715044 | |
| 22 | 16 | 22 | | | CoefVariação | | |

Fig. 21 - Células E22 e F22.

Após definir que o Coeficiente de Variação será calculado na célula **F22**, agora você digitará a fórmula **= F21/F16*100**, conforme está apresentado na Figura 22.

Fonte: Microsoft Excel

| | A | B | C | D | E | F | G |
|----|---------------|-----|----|-------|---------------|--------------|--------|
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | | Precipitações | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | Desvio Padrão | 2,493715044 | |
| 22 | 16 | 22 | | | CoefVariação | =F21/F16*100 | |

Fig. 22 - Coeficiente de Variação.

O valor final do Coeficiente de Variação 13,03 será apresentado na célula F22, conforme a figura 23.

Fonte: Microsoft Excel

| F22 | | fx = =F21/F16*100 | | | | | |
|-----|---------------|-------------------|----|---------------|---------------|-------------|--------|
| | A | B | C | D | E | F | G |
| 1 | Precipitações | Rol | AT | 10 | | | |
| 2 | | | C | 5 | | | |
| 3 | 18 | 15 | a | 2 | | | |
| 4 | 18 | 16 | | | | | |
| 5 | 17 | 16 | | | | | |
| 6 | 16 | 17 | | Precipitações | | f | % |
| 7 | 19 | 17 | 15 | ----- | 17 | 3 | 13,64 |
| 8 | 19 | 17 | 17 | ----- | 19 | 7 | 31,82 |
| 9 | 17 | 18 | 19 | ----- | 21 | 6 | 27,27 |
| 10 | 18 | 18 | 21 | ----- | 23 | 4 | 18,18 |
| 11 | 20 | 18 | 22 | ----- | 24 | 2 | 9,09 |
| 12 | 20 | 18 | | Soma | | 22 | 100,00 |
| 13 | 22 | 19 | | | | | |
| 14 | 25 | 19 | | | | | |
| 15 | 23 | 19 | | | | | |
| 16 | 22 | 20 | | | Média | 19,13636364 | |
| 17 | 21 | 20 | | | Mediana | 19 | |
| 18 | 21 | 20 | | | Moda | 18 | |
| 19 | 19 | 21 | | | | | |
| 20 | 17 | 21 | | | Variância | 6,218614719 | |
| 21 | 18 | 22 | | | Desvio Padrão | 2,493715044 | |
| 22 | 16 | 22 | | | CoefVariação | 13,03129001 | |

Fig. 23 - Valor Final do Coeficiente de Variação.

ATIVIDADE 02



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre o cálculo de Medidas Estatísticas com *Software* Estatístico até agora, utilize os dados a seguir, referentes a notas de alunos de Estatística, para calcular Variância, Desvio Padrão e Coeficiente de Variação no Excel.

Notas: 8, 9, 5, 0, 10, 3, 2, 9, 7, 2, 1, 3, 6, 8, 9, 10, 4, 2, 6, 1, 10.



RESUMINDO

Nesta aula você aprofundou os conhecimentos sobre o uso de funções Estatísticas com o Excel. Você calculou Média, Mediana e Moda; e calculou também Variância, Desvio Padrão e Coeficiente de Variação.

LEITURAS COMPLEMENTARES

O artigo http://www.cpafr.embrapa.br/embrapa/attachments/114_3bp_012003_solos_roraima_valdinar.pdf apresenta um estudo Estatístico da distribuição dos valores de pH, Ca²⁺, Mg²⁺, K⁺, Al³⁺, P, Soma de bases, textura e carbono orgânico de 87 perfis de solos levantados no Estado de Roraima, identificando as principais limitações desses solos para o uso agrícola. Os dados foram analisados utilizando técnicas de estatística básica com o apoio do Software Excel. Foram calculadas as médias, amplitudes e desvio padrão. Os principais resultados demonstram uma grande predominância de solos de baixa fertilidade natural.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe utilizar o Excel para calcular Medidas Estatísticas, obtenha dados reais e utilize o Excel para calcular Média, Mediana, Moda, Variância, Desvio Padrão e Coeficiente de Variação e comente-os. Como sugestão você pode utilizar dados ligados à Gestão Ambiental.



CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

NEUFELD, John L. **Estatística aplicada à administração usando Excel**. Tradução José Luiz Celeste; revisão técnica Cyro C. Patarra. São Paulo: Prentice Hall, 2003

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3 a ed. São Paulo: Makron Books, 1993 – Coleção Schaum.

TECNÓLOGO EM GESTÃO AMBIENTAL

**DISCIPLINA
ESTATÍSTICA**

**AULA 15
CORRELAÇÃO E REGRESSÃO COM
SOFTWARE ESTATÍSTICO**

**AUTOR
JOÃO MARIA FILGUEIRA**



GESTÃO AMBIENTAL



APRESENTANDO A AULA

Na aula 14 - Medidas Estatísticas com *Software* Estatístico - você utilizou o *Software* Excel para calcular Medidas Estatísticas. Nesta aula você vai continuar utilizando esse *software*, agora para calcular Coeficientes de Correlação Linear e Coeficientes de Regressão Linear, além do Erro Padrão da Estimativa. Você vai aprender, também, a construir o Diagrama de Dispersão utilizando o Excel.

DEFININDO OBJETIVOS

Ao final desta aula, você deverá ser capaz de utilizar o Excel para:

- construir Diagrama de Dispersão;
- calcular Coeficiente de Correlação Linear;
- calcular Coeficiente de Explicação;
- calcular Coeficientes de Regressão Linear;
- calcular Erro Padrão da Estimativa.

DESENVOLVENDO O CONTEÚDO

CALCULANDO COEFICIENTES DE CORRELAÇÃO LINEAR E CONSTRUINDO DIAGRAMA DE DISPERSÃO NO EXCEL

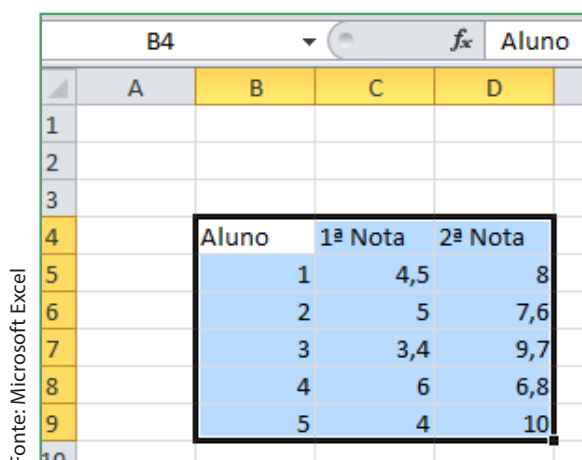
Você vai utilizar, nesta aula, o exemplo das notas de Estatística dos alunos de uma turma de Gestão Ambiental no primeiro e no segundo períodos, cujos dados estão apresentados a seguir na tabela 1.

Tabela 1 – Notas do 1º e 2º período de Estatística

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8,0 |
| 2 | 5,0 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6,0 | 6,8 |
| 5 | 4,0 | 10,0 |

Fonte: Secretaria Escolar

Inicialmente você precisa digitar esses dados no Excel. Você deve utilizar as células de B4 a D9 para digitá-las (Veja a Figura 1).



Fonte: Microsoft Excel

| | A | B | C | D |
|----|---|-------|---------|---------|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | Aluno | 1ª Nota | 2ª Nota |
| 5 | | 1 | 4,5 | 8 |
| 6 | | 2 | 5 | 7,6 |
| 7 | | 3 | 3,4 | 9,7 |
| 8 | | 4 | 6 | 6,8 |
| 9 | | 5 | 4 | 10 |
| 10 | | | | |

Fig. 01 – Notas dos alunos.

Agora você vai preparar a planilha para receber os valores calculados pelas funções do Excel. Veja como ficará sua planilha na Figura 2. Na célula **B11** você deve digitar Correlação, na **B12** Explicação, e assim você deve deixar sua planilha como a apresentada na Figura 2.

Fonte: Microsoft Excel

| | | G16 | | | | | | | |
|----|--|-----------------------|---------|---------|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H |
| 1 | | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | | | | | |
| 4 | | Aluno | 1ª Nota | 2ª Nota | | | | | |
| 5 | | 1 | 4,5 | 8 | | | | | |
| 6 | | 2 | 5 | 7,6 | | | | | |
| 7 | | 3 | 3,4 | 9,7 | | | | | |
| 8 | | 4 | 6 | 6,8 | | | | | |
| 9 | | 5 | 4 | 10 | | | | | |
| 10 | | | | | | | | | |
| 11 | | Correlação | | | | | | | |
| 12 | | Explicação | | | | | | | |
| 13 | | | | | | | | | |
| 14 | | Y=a+bX | Y = | | + | | X | | |
| 15 | | | | | | | | | |
| 16 | | ErroPadrão | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | Diagrama de dispersão | | | | | | | |
| 19 | | | | | | | | | |

Fig. 02 – Preparação da planilha.

Seguindo o que você aprendeu na aula 10 - Coeficiente de Correlação Linear -, você deve primeiro obter o Diagrama de Dispersão, depois o Coeficiente de Correlação Linear e o Coeficiente de Explicação.

Para construir o Diagrama de Dispersão, você deve selecionar as Notas, ou seja, você deve selecionar as células de C5:D9, conforme você pode ver na Figura 3.

Fonte: Microsoft Excel

| | A | B | C | D |
|----|---|-------|---------|---------|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | Aluno | 1ª Nota | 2ª Nota |
| 5 | | 1 | 4,5 | 8 |
| 6 | | 2 | 5 | 7,6 |
| 7 | | 3 | 3,4 | 9,7 |
| 8 | | 4 | 6 | 6,8 |
| 9 | | 5 | 4 | 10 |
| 10 | | | | |

Fig. 03 – Notas selecionadas.

Você precisa agora inserir um gráfico e, para isso, você deve clicar na Guia Inserir, e clicar na opção Gráfico Dispersão. Conforme a Figura 4 - Selecionando Gráfico Dispersão.

Fonte: Microsoft Excel

The screenshot shows the Microsoft Excel interface with the 'Inserir' (Insert) ribbon selected. The 'Gráficos' (Charts) group is active, and the 'Dispersão' (Scatter) icon is highlighted. A dropdown menu for 'Dispersão' is open, showing several chart options. The background shows the same data table as in Figure 3, with the '1ª Nota' and '2ª Nota' columns selected.

Fig. 04 – Inserir diagrama de dispersão.

Agora, o Diagrama de Dispersão está concluído e você pode arrastá-lo para onde quiser. Deixe-o conforme a Figura 5. Lembre-se que você poderá formatá-lo como quiser, mas faça-o posteriormente.

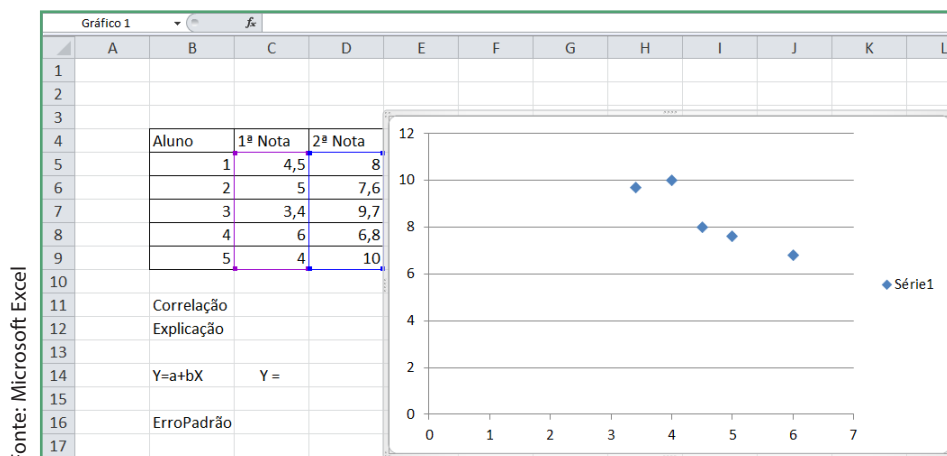


Fig. 05 – Diagrama de dispersão.

Uma breve análise do Diagrama de Dispersão na Figura 5 permite concluir que há uma razoável relação linear entre a 1ª Nota e a 2ª Nota. Você vai mensurá-la a partir de agora, calculando o Coeficiente de Correlação Linear e o Coeficiente de Explicação. Você deve ir para a célula C11, conforme a Figura 6.

Fonte: Microsoft Excel

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8 |
| 2 | 5 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6 | 6,8 |
| 5 | 4 | 10 |

Correlação

Fig. 06 – Célula C11.

Para o cálculo do Coeficiente de Correlação Linear, você vai utilizar as funções que estão disponíveis na Guia Fórmulas, clicando na opção Inserir Função, escolhendo a categoria Estatística, escolhendo a função CORREL e clicando OK, conforme a Figura 7. Você deve lembrar-se bem desse procedimento, pois já o utilizou bastante nas aula 13 - Distribuição de Frequência com *Software* Estatístico e aula 14 - Medidas Estatísticas com *Software* Estatístico.

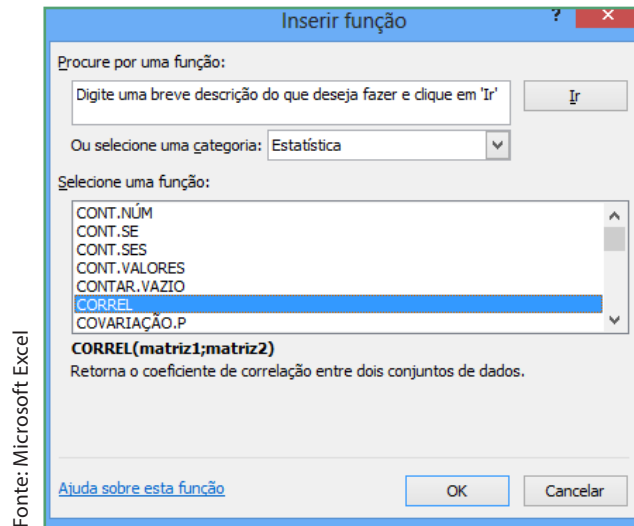


Fig. 07 – Inserir função.

Surgirá então a janela da Figura 08, para você selecionar os valores da Variável X (na Matriz1) e da Variável Y (na Matriz2).

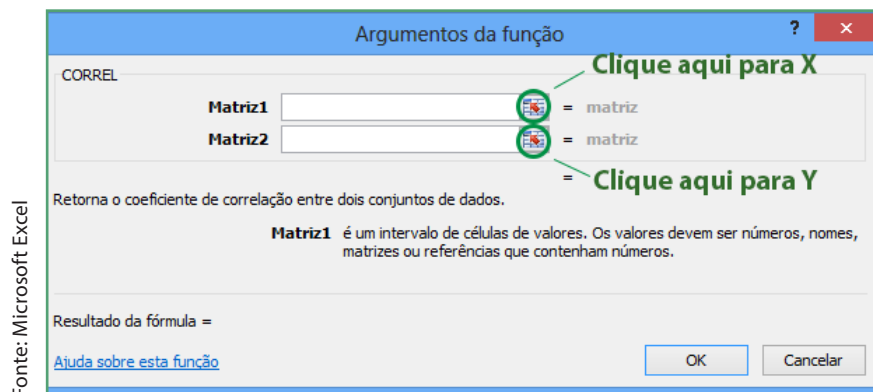


Fig. 08 – Matriz1 e Matriz2.

Seguindo a orientação da Figura 08, você clicará no local indicado para selecionar os valores da Variável X.

Agora você já está na planilha e vai selecionar os valores de X, conforme a Figura 09. Após selecionar os valores de X, clique conforme indicado na Figura 09 para voltar à janela Matriz1 e Matriz2.

Fonte: Microsoft Excel

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8 |
| 2 | 5 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6 | 6,8 |
| 5 | 4 | 10 |

Correlação (C5:C9)

Fig. 09 – Valores da 1ª Nota.

Seguindo a orientação da Figura 08, você clicará no local indicado para selecionar os valores da Variável Y. Agora você já está na planilha e vai selecionar os valores de Y, conforme a Figura 10. Após selecionar os valores de Y, clique conforme indicado na Figura 10 para voltar à janela Matriz1 e Matriz2.

Fonte: Microsoft Excel

| Aluno | 1ª Nota | 2ª Nota |
|-------|---------|---------|
| 1 | 4,5 | 8 |
| 2 | 5 | 7,6 |
| 3 | 3,4 | 9,7 |
| 4 | 6 | 6,8 |
| 5 | 4 | 10 |

Correlação ;D5:D9)

Fig. 10 – Valores da 2ª Nota.

Agora, você terá a janela da Figura 11, na qual você simplesmente clicará em OK, pois a Matriz1, a Matriz2 já estão com os valores respectivos de X e de Y; e o resultado da fórmula já está calculado.

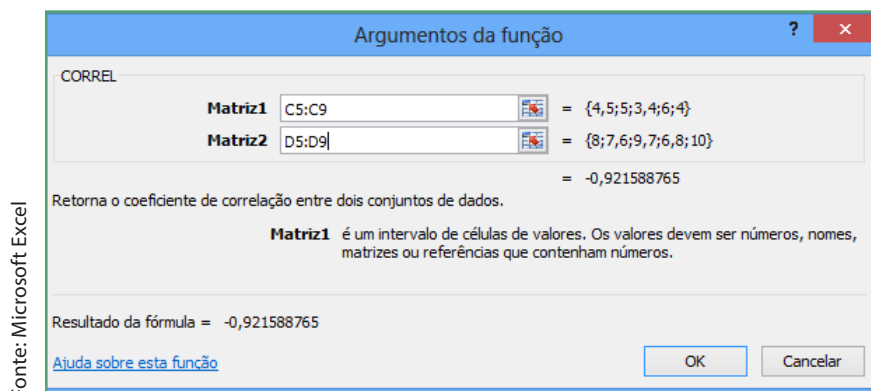


Fig. 11 – Valor do Coeficiente de Correlação Linear.

Para o cálculo do Coeficiente de Explicação, você deve ir para a célula C12, lá você vai utilizar as funções que estão disponíveis na Guia Fórmula, clicando na opção Inserir Função, escolhendo a categoria Estatística, escolhendo a função RQUAD e clicando OK, conforme a Figura 12.

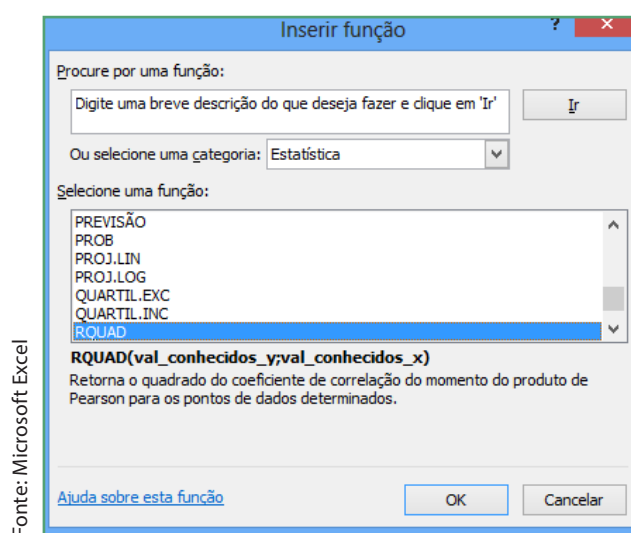


Fig. 12 – Inserir função.

Seguindo os procedimentos já conhecidos para selecionar os valores de Y e de X, você obterá uma janela como a da Figura 13, na qual você simplesmente clicará em OK, pois os valores de Y e de X já estão disponíveis; e o resultado da fórmula já está calculado.

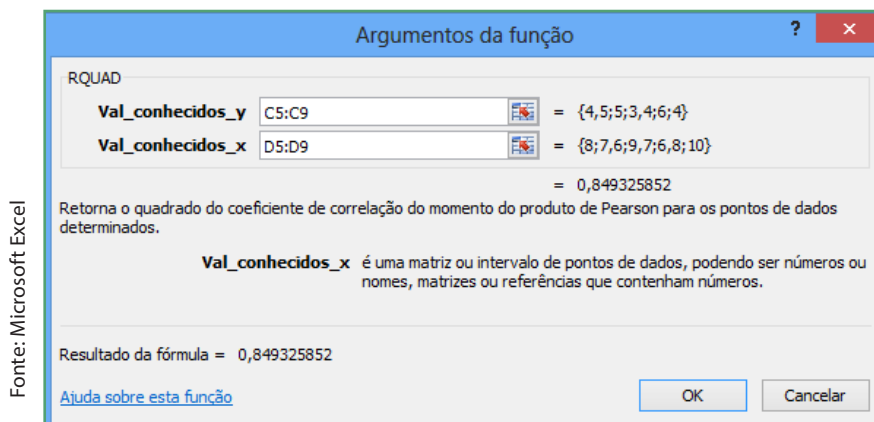


Fig. 13 – Valor do Coeficiente de Explicação.

Portanto, com este resultado, pode-se afirmar que a Variável 1ª Nota explica 84,93% da variação da 2ª Nota.

ATIVIDADE 01



A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la, retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Diagrama de Dispersão, Coeficiente de Correlação Linear e Coeficiente de Explicação com *Software* Estatístico até agora, utilize os dados a seguir, para construir um Diagrama de Dispersão e calcular o Coeficiente de Correlação Linear e o Coeficiente de Explicação utilizando o Excel.

Tabela 2 - Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

CALCULANDO COEFICIENTES DE REGRESSÃO LINEAR E ERRO PADRÃO DA ESTIMATIVA NO EXCEL

Para o cálculo dos Coeficientes de Regressão Linear e do Erro Padrão da Estimativa, você vai utilizar os dados da Tabela 1 e a planilha já preparada na Figura 2.

Para o cálculo do parâmetro **a** da reta de regressão $Y = a + b * X$, você deve ir para a célula **D14**, e lá utilizar as funções que estão disponíveis na Guia Fórmula, clicando na opção Inserir Função, escolhendo a categoria Estatística, escolhendo a função INTERCEPÇÃO e clicando OK, conforme a Figura 14.

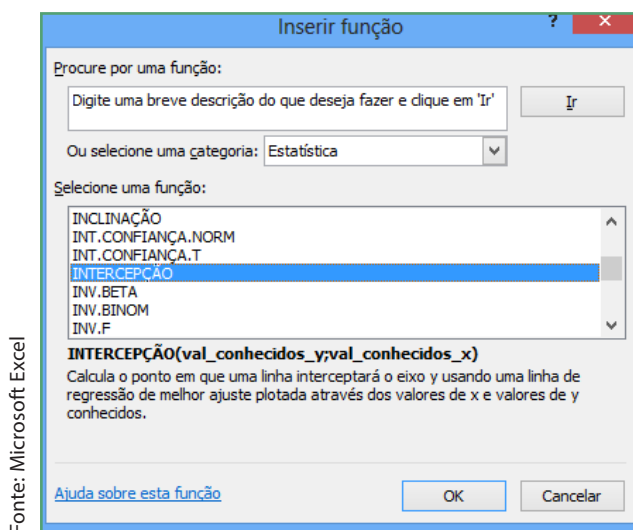


Fig. 14 – Inserir função.

Seguindo os procedimentos já conhecidos para selecionar os valores de Y e de X, você obterá uma janela como a da Figura 15, na qual você simplesmente clicará em OK, pois os valores de Y e de X já estão disponíveis e o resultado da fórmula já está calculado.

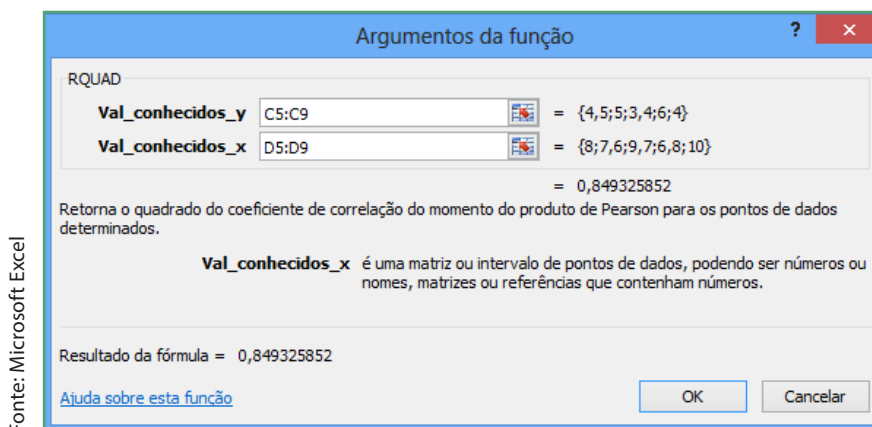


Fig. 15 – Parâmetro a da reta de regressão.

Para o cálculo do parâmetro **b** da reta de regressão $Y = a + b * X$, você deve ir para a célula **F14**, e lá utilizar as funções que estão disponíveis na Guia Fórmula, clicando na opção Inserir Função, escolhendo a categoria Estatística, escolhendo a função INCLINAÇÃO e clicando OK, conforme a Figura 16.

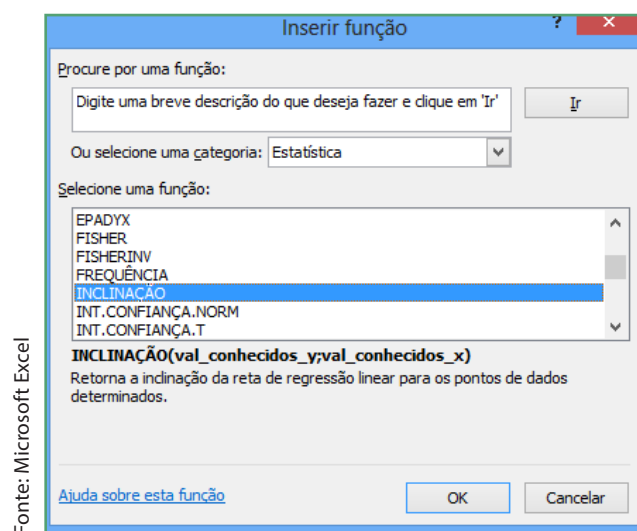


Fig. 16 – Inserir função.

Seguindo os procedimentos já conhecidos para selecionar os valores de Y e de X, você obterá uma janela como a da Figura 17, na qual você simplesmente clicará em OK, pois os valores de Y e de X já estão disponíveis e o resultado da fórmula já está calculado.

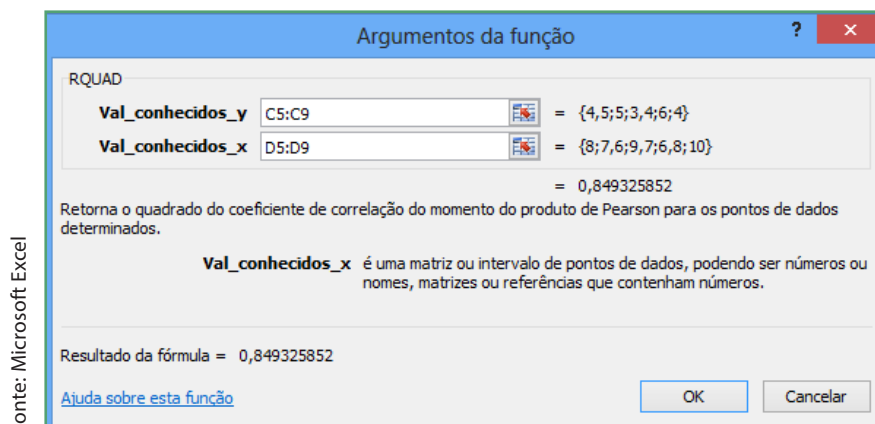


Fig. 17 – Parâmetro b da reta de regressão.

Portanto, você já tem a reta de regressão $Y = 14,29 - 1,28 * X$. E pode realizar qualquer previsão de 2ª Nota a partir da 1ª Nota, conforme aprendeu na aula 11 - Regressão Linear.

Para o cálculo do Erro Padrão da Estimativa, você deve ir para a célula **C16** e lá utilizar as funções que estão disponíveis na Guia Fórmula, clicando na opção Inserir Função, escolhendo a categoria Estatística, escolhendo a função EPADYX e clicando OK, conforme a Figura 18.

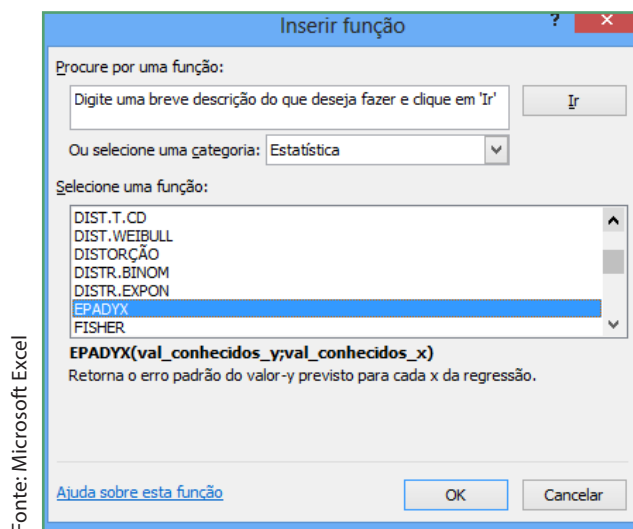


Fig. 18 – Inserir função.

Seguindo os procedimentos já conhecidos para selecionar os valores de Y e de X, você obterá uma janela como a da Figura 19, na qual você simplesmente clicará em OK, pois os valores de Y e de X já estão disponíveis e o resultado da fórmula já está calculado.

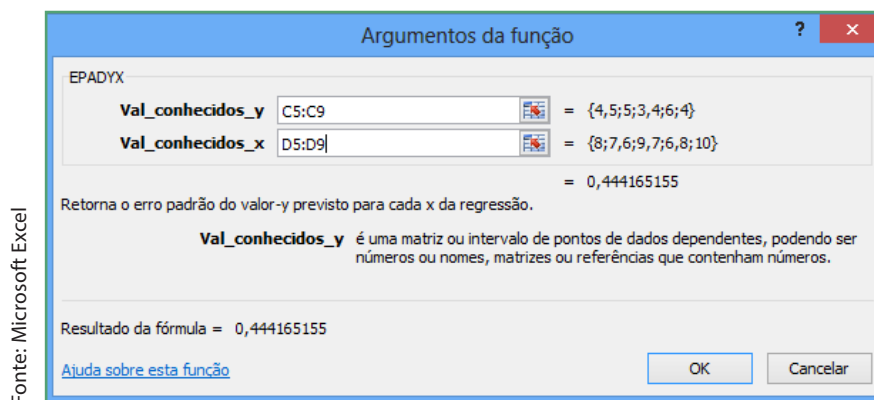


Fig. 19 – Erro Padrão da Estimativa.

Portanto, para cada previsão que você realizar para a 2ª Nota a partir da 1ª Nota, conforme aprendeu na aula 12 - Erro Padrão da Estimativa - você poderá cometer um erro em torno de 0,62; que é um erro muito pequeno considerando-se notas de alunos que variam de ZERO a DEZ.

ATIVIDADE 02

A atividade a seguir deve ser feita antes de prosseguir nos estudos da aula. Caso você não consiga respondê-la retome a leitura do conteúdo a que ela se refere.

1. Com base no que estudamos sobre Coeficientes de Regressão Linear e Erro Padrão da Estimativa com Software Estatístico até agora, utilize os dados a seguir, para construir calcular os parâmetros **a** e **b** da reta de regressão ($Y = a + b * X$) e o Erro Padrão da Estimativa, utilizando o Excel.

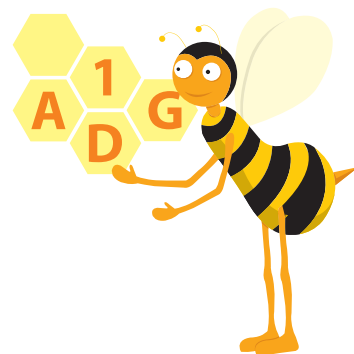
Tabela 2 - Pessoas atendidas e Faturamento mensal, de determinada loja de eletrodomésticos

| Mês | Faturamento R\$ | Pessoas atendidas |
|-----|-----------------|-------------------|
| 1 | 2001 | 804 |
| 2 | 2048 | 829 |
| 3 | 1998 | 797 |
| 4 | 2030 | 815 |
| 5 | 1992 | 805 |
| 6 | 2013 | 811 |

Fonte: Dados primários.

RESUMINDO

Nesta aula você aprofundou os conhecimentos sobre o uso de funções Estatísticas com o Excel. Você, utilizando Excel, calculou Coeficiente de Correlação Linear, Coeficiente de Explicação e construiu Diagrama de Dispersão; calculou também Coeficientes de Regressão Linear e o Erro Padrão da Estimativa.



LEITURAS COMPLEMENTARES

No artigo disponível em: <http://engema.up.edu.br/arquivos/engema/pdf/PAP0388.pdf> foram analisados os balanços sociais e as demonstrações de resultados de dezoito empresas, com o objetivo de avaliar se há correlação entre o gasto com as práticas sócio-ambientais e o desempenho das empresas. Após a análise, obteve-se um Coeficiente de Explicação $E=R^2=0,9921$, evidenciando que os Investimentos Sociais guardam forte sintonia com as Vendas e Lucros, explicando 99,21% destes últimos.



AVALIANDO SEUS CONHECIMENTOS

Considerando-se que você já sabe utilizar o Excel para calcular Parâmetros de Correlação e de Regressão, utilize a tabela 4, contendo os valores de testes realizados junto a funcionários de uma indústria antes e depois de um treinamento de requalificação, e calcule com o Excel o Coeficiente de Explicação e o Erro Padrão da Estimativa; construa também um Diagrama de Dispersão. Os valores dos testes são de zero a dez.



Tabela 4 - Testes antes e após treinamento de requalificação.

| Funcionário | Teste antes | Teste após |
|-------------|-------------|------------|
| 1 | 8,4 | 7,1 |
| 2 | 4,6 | 1,9 |
| 3 | 2,5 | 9,7 |
| 4 | 9,8 | 7,4 |
| 5 | 1,5 | 7,1 |
| 6 | 7,4 | 5,7 |
| 7 | 2,0 | 5,9 |
| 8 | 7,0 | 6,0 |
| 9 | 0,1 | 9,7 |
| 10 | 3,9 | 7,0 |

Fonte: Dados primários.

CONHECENDO AS REFERÊNCIAS

CONCEIÇÃO, Gleice Margarete de Souza et all. **Noções Básicas de Estatística**. Curso de Capacitação em Epidemiologia Básica e Análise da Situação de Saúde Ministério da Saúde Secretaria de Vigilância em Saúde.

FONSECA, Jairo Simon. **Curso de Estatística**. São Paulo: Atlas, 1990.

LARSON, Ron. **Estatística Aplicada**. Tradução e revisão técnica Cyro de Carvalho Patarra. São Paulo: Prentice Hall, 2004.

NEUFELD, John L. **Estatística aplicada à administração usando Excel**. Tradução José Luiz Celeste; revisão técnica Cyro C. Patarra. São Paulo: Prentice Hall, 2003

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Consentino, 3ª ed. São Paulo: Makron Books, 1993 – Coleção Schaum.