



REVISÃO DE ESTATÍSTICA BÁSICA

PROF. IVAN BEZERRA ALLAMAN

CRONOGRAMA

1. Síntese de dados qualitativos e quantitativos
2. Medidas de tendência central e dispersão
3. Probabilidade
4. Distribuição normal e normal padrão
5. Distribuição t de student

SÍNTESE DE DADOS QUALITATIVOS E QUANTITATIVOS

SÍNTESE DE DADOS QUALITATIVOS

Lembre-se que dados qualitativos se refere a variáveis cujo a resposta é um atributo, cujo a escala pode ser nominal ou ordinal

- A exploração dos dados deve começar com a redução dos mesmos por meio de tabelas de distribuição de frequências (tdf).
- Conforme estudado em estatística básica, a tdf possui basicamente 6 colunas a saber: classe ou intervalo de classes, frequência absoluta (F.a), frequência relativa (F.r), frequência percentual (F.rp (%)), frequência acumulada (F.ac) e frequência acumulada percentual (F.ac (%)).

Classe	F.a	F.r	F.rp (%)	F.ac	F.ac (%)
A					
B					
.					
.					
.					

SÍNTESE DE DADOS QUANTITATIVOS

Não esqueçam de que quando a variável é discreta e tem poucos elementos distintos, a tabulação pode ser feita da mesma forma que vimos para dados qualitativos

- No caso de variáveis quantitativas, o **intervalo de classes** pode ser elaborado por meio do critério de Sturges, que é o mais utilizado. Existem outros métodos que não abordaremos por se tratar de uma revisão.

1. Ordene os dados (crescente)
2. Calcule o limite inferior da primeira classe (LI)

$$LI = X_1 - \frac{|X_1|}{100}$$

3. Calcule o limite superior da última classe (LS)

$$LS = X_n + \frac{|X_n|}{100}$$

4. Calcule o número de classes (no caso será pelo Sturges)

$$k = 1 + \ln_2 n$$

5. Calcule o comprimento de classe (h)

$$h = \frac{LS - LI}{k}$$

MEDIDAS DE TENDÊNCIA CENTRAL E DISPERSÃO

MÉDIA ARITMÉTICA SIMPLES

Não se esqueçam do rigor da notação estatística, ou seja, \bar{x} para o estimador e μ para o parâmetro

- É a soma de todas as observações dividido pela quantidade de dados

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

MÉDIA HARMÔNICA

É uma medida utilizado com menos frequência. No melhoramento genético é aplicado em algum cálculo para componentes da variância.

- É calculado como:

$$\overline{x_h} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

MÉDIA GEOMÉTRICA

É apropriada para calcular médias de razões, taxas de variações, taxas de crescimento em dados financeiros, índices econômicos e taxas de crescimento de microorganismos.

- O cálculo é efetuado da seguinte forma:

$$\bar{x}_g = \sqrt[n]{(x_1) * (x_2) * \dots * (x_n)}$$

VARIÂNCIA

É a mais utilizada dentre as medidas de dispersão

- A média por si só não é capaz de descrever as características de um conjunto de dados embora, erroneamente, as pessoas tenham tal costume.
- Se faz necessário uma medida que informe o quanto os dados estão dispersos ao redor de sua tendência (média no caso).
- O cálculo da variância amostral é dado da seguinte forma:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

DESVIO PADRÃO

- Surgiu com o intuito de termos uma interpretabilidade factível acerca da variabilidade dos dados, uma vez que, a unidade de medida resultante da variância é elevada ao quadrado.
- Podemos calcular da seguinte forma:

$$S = \sqrt{S^2}$$

PROBABILIDADE

NÃO PODEMOS ESQUECER JAMAIS

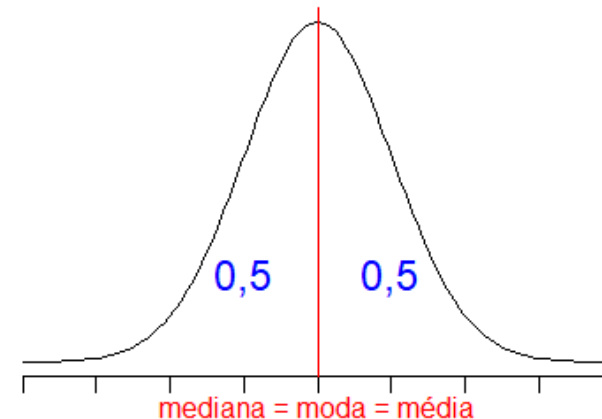
- A probabilidade de um evento aleatório qualquer é um número entre 0 e 1.
- Portanto, não existe probabilidade negativa e muito menos maior que 1.
- Na Estatística Aplicada iremos utilizar frequentemente o termo **significância** ou **confiança** cujo o alicerce remete ao entendimento do que é probabilidade.

DISTRIBUIÇÕES CONTÍNUAS DE PROBABILIDADE

DISTRIBUIÇÃO NORMAL

No Excel utilize a função “DIST.NORM.N” para encontrar uma probabilidade e a função “INV.NORM.N” para encontrar um quantil. É claro que neste caso você deve especificar a média e o desvio padrão da população.

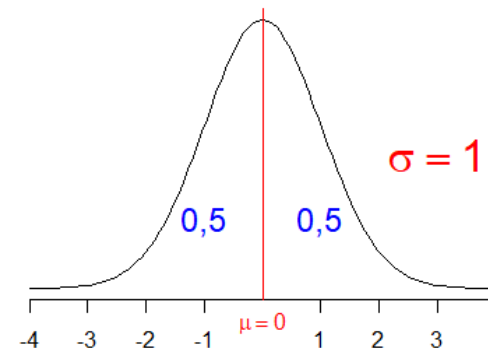
- Quando agrupamos uma variável quantitativa (na sua maioria contínua) em intervalos de classe (tabela de distribuição de frequência, lembra?), esta variável irá apresentar um formato, padrão, comportamento que apelidamos de normal, pois a maioria dos fenômenos quantitativos costumam ter tal padrão.
- Dado a descoberta de Gauss e a formalização de tal padrão em função matemática, podemos até os dias de hoje calcular probabilidades de um fenômeno quantitativo e vivenciar a revolução científica e tecnológica.



DISTRIBUIÇÃO NORMAL PADRÃO

No Excel utilize a função “DIST.NORMP.N” para encontrar uma probabilidade e a função “INV.NORMP.N” para encontrar um quantil. Neste caso lembrem-se que a média é zero e o desvio padrão é 1.

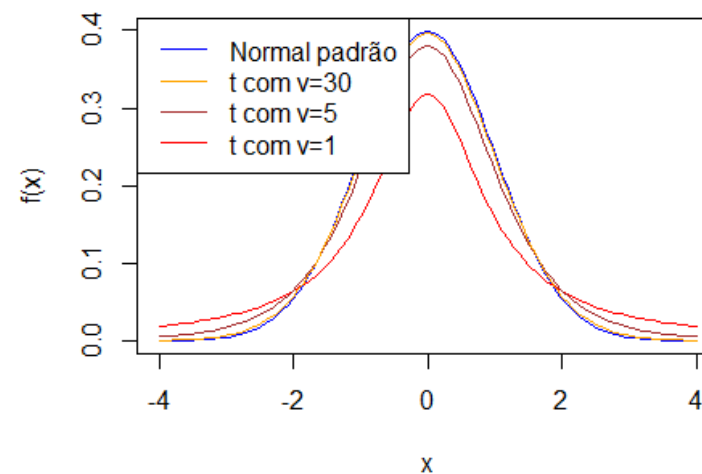
- Surgiu com o intuito de facilitar cálculos que só podiam ser feitos por profissionais com boa habilidade matemática.
- Posteriormente descobriu-se que tal distribuição dava origem a outras distribuições de probabilidade como a **qui-quadrado** e a **t de student** e, muitas teorias estatísticas puderam ser provadas com o auxílio desta distribuição (teorema do limite central).
- Se você aprendeu a utilizar tabelas para encontrar o quantil de uma distribuição esqueça!!!! Isso fez parte do século passado. Com os adventos da tecnologia, podemos utilizar softwares computacionais, aplicativos de smartphone (*Probability distribution app*), planilhas eletrônicas (Excel por exemplo), dentre outros.



DISTRIBUIÇÃO t DE STUDENT

No Excel utilize a função “DIST.T” para encontrar uma probabilidade e a função “INV.T” para encontrar um quantil.

- Elaborada por William Gosset, tal distribuição revolucionou o século XX, pois possibilitou estudos científicos (mais especificamente a média) nos quais não se tinham qualquer conhecimento acerca das características da população.
- Essa distribuição tem características semelhantes a distribuição normal padrão, ou seja, a média é zero e é simétrica.
- Outra característica importante é quando o tamanho da amostra tende ao infinito a distribuição t se aproxima da distribuição normal padrão.





APLICAÇÕES



- 1) Considere os [dados](#) da Pelican Stores, uma divisão da National Clothing, que é uma cadeia de lojas de roupas femininas que opera em todos os Estados Unidos. Elabore um relatório administrativo.
- 2) Utilizando o arquivo excel disponível [aqui](#) encontre os quantis e probabilidade das distribuições normal, normal padrão e t de student