

Excel INTERMEDIÁRIO

Estatística



Prof. Cassiano Isler
2017.1 - Turma 3

Agenda

Agenda

Bibliografia

Estatística Descritiva

Medidas de Posição
Medidas de Dispersão

Histograma

Distribuições Estatísticas

Padronização de Distribuição Normal
Cálculo de Probabilidade

Exercícios

- Estatística Descritiva

Medidas de Posição

Medidas de Dispersão

- Histograma

- Distribuições Estatísticas

Padronização de Distribuição Normal

Cálculo de Probabilidade

Bibliografia

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição
Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal
Cálculo de
Probabilidade

Exercícios

COSTA NETO, P. L. O. **Estatística**. 2. ed. São Paulo: Edgard Blücher (2002).

GÓMEZ, Luis Alberto. **Excel para engenheiros**. Visual Books, 2009. **Capítulos 6 e 7**. Disponível na biblioteca UFSC-Joinville.

DIXON, Helen. **Excel 2007: Beyond the Manual**. Berkeley, 2007. **Capítulo 10** disponível em :

<http://dx.doi.org/10.1007/978-1-4302-0389-6>

Slides das aulas e material complementar disponíveis em:

[Curso Intermediário Excel](#)

Estatística Descritiva

Agenda

Bibliografia

Estatística Descritiva

Medidas de Posição

Medidas de Dispersão

Histograma

Distribuições Estatísticas

Padronização de Distribuição Normal

Cálculo de Probabilidade

Exercícios

A estatística descritiva é utilizada para descrever dados através de gráficos, distribuições de frequência ou medidas associadas a essas distribuições.

São utilizadas variáveis unidimensionais para tratamento matemático de um problema quando apenas uma característica de interesse está associada a cada elemento do conjunto examinado.

As variáveis de interesse podem ser qualitativas ou quantitativas e essas últimas podem ser classificadas em variáveis discretas ou contínuas.

Medidas de Posição

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

As medidas de posição são utilizadas para caracterizar os dados contidos em uma amostra.

- **MÍNIMO(núm1;[núm2];...):** retorna o menor valor observado em um conjunto de números.
- **MÁXIMO(núm1;[núm2];...):** retorna o maior valor observado em um conjunto de números.
- **MAIOR(matriz; k):** retorna o k -ésimo maior valor observado em uma amostra de uma matriz ou vetor.
- **MENOR(matriz; k):** retorna o k -ésimo menor valor observado em uma amostra de uma matriz ou vetor.

Medidas de Posição

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **MED(núm1;[núm2];...)**: retorna a mediana, a observação que separa a metade superior da metade inferior de uma amostra.
- **MOD0(núm1;[núm2];...)**: retorna a moda, a observação mais frequente da amostra.
- **MÉDIA(núm1;[núm2];...)**: retorna a média aritmética das observações.
- **MÉDIASE(intervalo;critério;[intervalo média])**: retorna a média de um conjunto de observações selecionadas a partir de uma condição específica.

As funções “MÉDIA.GEOMÉTRICA” e “MÉDIASES” são extensões das apresentadas.

Medidas de Dispersão

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

As medidas de dispersão representam a variabilidade dos dados de uma amostra em torno das medidas de tendência central, mais especificamente da média.

- **[MÁXIMO() - MÍNIMO()]**: retorna a amplitude, diferença entre os valores máximo e mínimo observados .
- **QUARTIL(matriz; quarto)**: retorna um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais, segundo o valor de “quarto” .

Primeiro quartil (quartil inferior): valor até 25% da amostra ordenada (25º percentil).

Segundo quartil (mediana): valor até 50% da amostra ordenada (50º percentil).

Terceiro quartil (quartil superior): valor até 75% da amostra ordenada (75º percentil).

Medidas de Dispersão

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **PERCENTIL(matriz; k)**: retorna o k -ésimo percentil de uma amostra, medida que divide a amostra ordenada em 100 partes.
- **VAR.A(núm1;[núm2];...)**: retorna a variância de um conjunto de observações.

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

- **DESVPAD.A(núm1;[núm2];...)** retorna o desvio padrão de um conjunto de observações.

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

Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

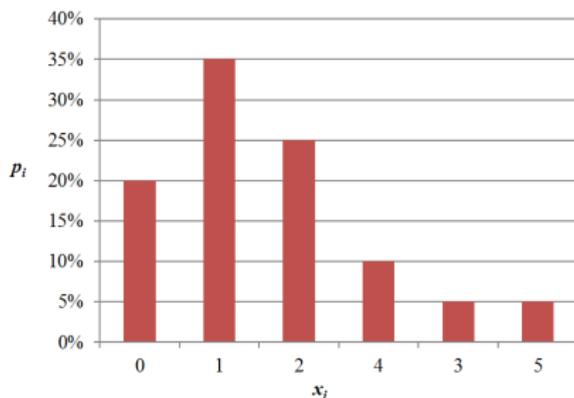
Cálculo de
Probabilidade

Exercícios

O histograma é uma ferramenta para representação da ocorrência de uma “variável aleatória” dado o seu comportamento estocástico.

No caso das variáveis quantitativas discretas o histograma é um gráfico de colunas que representa a contagem de ocorrências (frequência) de cada valor observado da variável.

x_i	f_i	p_i
0	4	20%
1	7	35%
2	5	25%
4	2	10%
3	1	5%
5	1	5%
Total	20	100%



Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

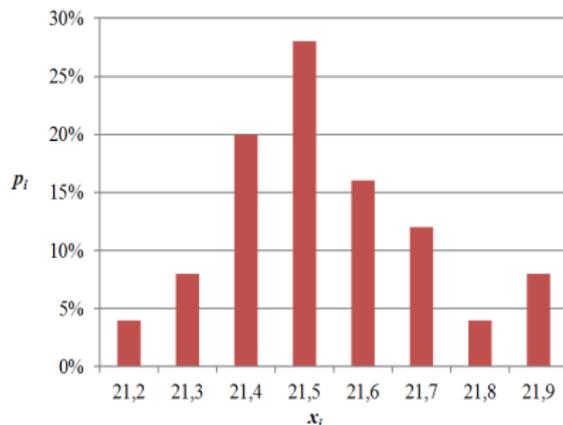
Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

No caso de variáveis quantitativas contínuas, o histograma é um gráfico de colunas que representa a contagem de ocorrências (frequência) em um intervalo de classe.

x_i	f_i	F_i	p_i	P_i
21,2	1	1	4%	4%
21,3	2	3	8%	12%
21,4	5	8	20%	32%
21,5	7	15	28%	60%
21,6	4	19	16%	76%
21,7	3	22	12%	88%
21,8	1	23	4%	92%
21,9	2	25	8%	100%
Total	25		100%	



Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

Existem diferentes maneiras de construir um histograma a partir de dados das células em uma planilha do Excel.

O software possui um suplemento que constrói o gráfico automaticamente, o qual será apresentado posteriormente.

Nesta aula será apresentada uma técnica de construção de histogramas que tem a vantagem de atualização automática do gráfico a cada alteração do conteúdo das células.

Pela própria definição, o histograma requer a contagem do número de ocorrências de valores uma variável discreta ou do número de ocorrências de valores em um intervalo (classe) no caso de variáveis aleatórias contínuas.

Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

O procedimento para construção de um histograma no Excel pode ser definido pelos seguintes passos para uma variável aleatória contínua:

- **PASSO 1:** Identifique os valores máximo e mínimo dos dados usando as funções **MÁXIMO()** e **MÍNIMO()**;
- **PASSO 2:** Defina o número de classes (colunas do histograma);
- **PASSO 3:** Calcule a amplitude da classe, igual à amplitude do intervalo de dados dividida pelo número de classes $[(\text{máximo} - \text{mínimo})/\text{classes}]$;
- **PASSO 4:** Insira o número de identificador de cada classe em uma coluna denominada “Classe”;

Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição
Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **PASSO 5:** Na primeira linha de uma coluna de “Limite Inferior da Classe”, insira o valor mínimo da série de dados (utilize uma referência relativa pelo símbolo “=”).
- **PASSO 6:** Na primeira linha de uma coluna de “Limite Superior da Classe”, insira o valor mínimo da série de dados adicionado ao valor da amplitude da classe;
- **PASSO 7:** Defina o valor da segunda linha da coluna de “Limite Inferior da Classe” igual ao da primeira linha da coluna de “Limite Superior da Classe” (utilize uma referência relativa pelo símbolo “=”).
- **PASSO 8:** Replique as fórmulas da segunda linha da coluna de “Limite Inferior da Classe” e da primeira linha da coluna de “Limite Superior da Classe” para as linhas das demais classes;

Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **PASSO 9:** Em uma coluna “Frequência Acumulada”, calcule a frequência de valores menores que o limite superior em cada linha pela função **CONT.SE()**;
- **PASSO 10:** Em uma coluna “Frequência Relativa”, na linha da primeira classe insira o número de ocorrências calculado pela “Frequência Acumulada”;
- **PASSO 11:** Para as demais linhas, calcule a frequência relativa subtraindo a frequência acumulada da classe anterior da frequência acumulada da classe corrente;

Histograma

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **PASSO 12:** Para estabelecer o histograma com valores percentuais de frequência dividida a frequência calculada de cada classe (relativa ou acumulada) pelo total de observações, dado pela função **CONT.VALORES()**;
- **PASSO 13:** Construa um gráfico de colunas em que os rótulos do eixo horizontal são o limite inferior ou superior das classes e os valores da série são as frequências relativas ou acumuladas (absolutas ou percentuais).

Para construir um histograma de frequência acumulada ou relativa de uma variável aleatória discreta, substitua os limites superior e inferior das classes pelos possíveis valores observados da variável aleatória.

Distribuições Estatísticas

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

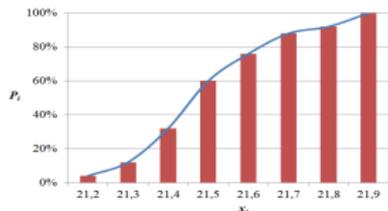
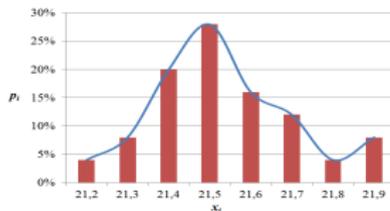
Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

O conceito de distribuição de uma variável aleatória é utilizado para caracterizar o comportamento dessa variável sob a forma de uma função matemática.

Uma distribuição de probabilidade pode ser caracterizada como uma função contínua correspondente à aproximação dos pontos centrais dos intervalos de frequência de um histograma para uma linha.



A terminologia adequada é que variáveis discretas são representadas por “funções de probabilidade” e variáveis contínuas por “funções de densidade de probabilidade”.

Funções de Distribuição de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição
Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

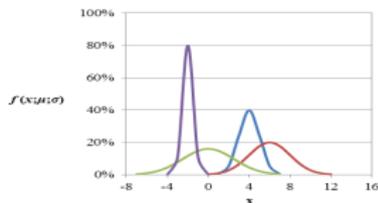
Cálculo de
Probabilidade

Exercícios

Na prática, as frequências de ocorrência apresentadas no histograma de valores observados para a variável aleatória são aproximadas para valores de probabilidade de ocorrência obtidos a partir de funções contínuas (equações) de formato conhecido.

Algumas distribuições de probabilidades são mais relevantes na modelagem de demanda de transporte.

Distribuição Normal



- $\mu = -4; \sigma = 1$
- $\mu = 6; \sigma = 2$
- $\mu = 0; \sigma = 2,5$
- $\mu = -2; \sigma = 0,5$

$$f(x; \mu; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

Funções de Distribuição de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição
Medidas de
Dispersão

Histograma

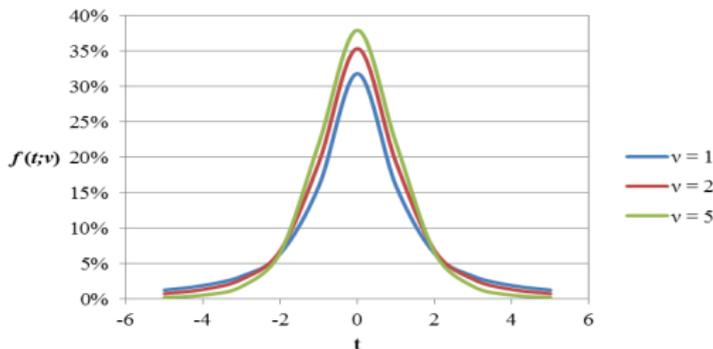
Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

Distribuição t-Student



$$f(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu \cdot \pi} \cdot \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

em que em que $\Gamma(n+1) = n!$ é denominada função Gama.

Funções de Distribuição de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

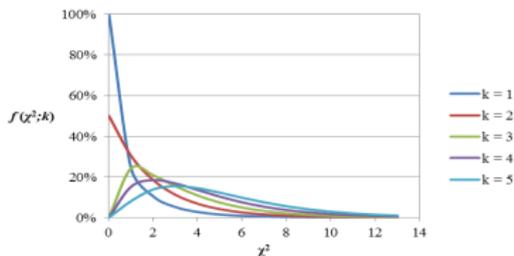
Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

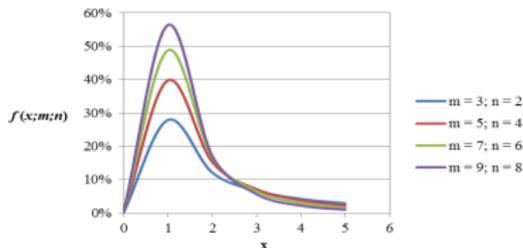
Exercícios

Distribuição χ^2 (Qui-quadrado)



$$f(\chi^2; k) = \frac{1}{2^{k/2} \Gamma(k/2)} (\chi_k^2)^{k/2-1} \cdot e^{-\chi_k^2/2}$$

Distribuição F-Fisher



$$f(x; m; n) = \frac{\Gamma\left(\frac{m+n}{2}\right) \cdot \left(\frac{m}{n}\right)^{m/2} \cdot x^{m/2-1}}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right) \cdot \left[\frac{m \cdot x}{n} + 1\right]^{\frac{m+n}{2}}}$$

Padronização de Distribuição Normal

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

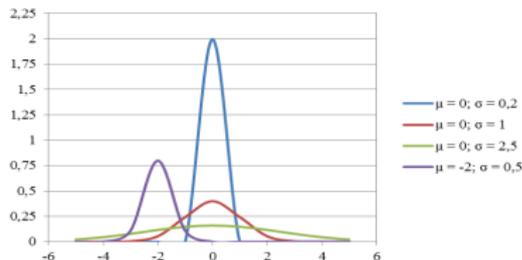
Cálculo de
Probabilidade

Exercícios

A distribuição Normal de uma população qualquer com média \bar{x} e desvio padrão s conhecidos pode ser reduzida para uma “Distribuição Normal Padrão” com média ($\mu = 0$) e desvio padrão unitário ($\sigma = 1$) através do procedimento de padronização da distribuição.

A padronização corresponde à conversão de cada valor observado da variável pela subtração da média da população e divisão pelo respectivo desvio padrão.

$$z_i = \frac{x_i - \mu}{\sigma}$$



Padronização de Distribuição Normal

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

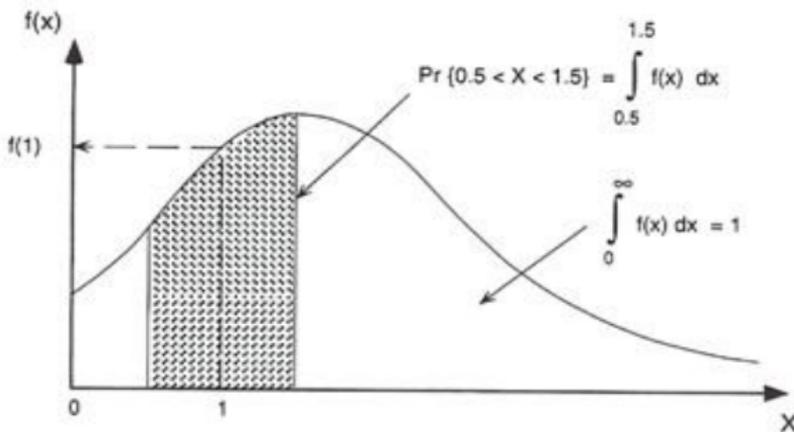
O processo de padronização foi extensivamente utilizado na prática pois, dada a Distribuição Normal Padrão foram construídas tabelas de probabilidade de uma variável aleatória assumir determinados valores para uma amostra com média \bar{x} e desvio padrão s .

Atualmente diferentes softwares são capazes de calcular o valor de probabilidade de ocorrência de uma variável aleatória através de cálculos matemáticos suprimida necessidade de tabelas.

Um recurso interessante desses softwares é identificar o valor de uma variável aleatória dada uma probabilidade ou seja executar o caminho inverso ao cálculo da probabilidade.

Cálculo de Probabilidade

Dado que a função de distribuição de probabilidade de uma variável aleatória é conhecida, é possível calcular a probabilidade de ocorrência de um valor ou intervalo de valores dessa variável.



Cálculo de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

O Excel possui funções para cálculo da probabilidade de diferentes funções de probabilidade e funções de distribuição de probabilidade dados os seus parâmetros.

- **DIST.NORM.N(x,média,desv_padrão,cumulativo)**: retorna a probabilidade da variável “x” para uma **Distribuição Normal** com média e desvio padrão conhecidos, acumulada (cumulativo=VERDADEIRO) ou relativa (cumulativo=FALSO).
- **DIST.NORMP.N(z,cumulativo)**: retorna a probabilidade da variável “z” (ver slide 20) para uma **Distribuição Normal Padrão** com média e desvio padrão conhecidos, acumulada (cumulativo=VERDADEIRO) ou relativa (cumulativo=FALSO).

Cálculo de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **DIST.T(x, graus_liberdade, cumulativa)**: retorna a probabilidade da variável “x” para uma **Distribuição t-Student** com “graus_liberdade” conhecido, acumulada (cumulativo=VERDADEIRO) ou relativa (cumulativo=FALSO).
- **DIST.QUIQUA(x, graus_liberdade, cumulativa)**: retorna a probabilidade da variável “x” para uma **Distribuição Qui-quadrado** com “graus_liberdade” conhecido, acumulada (cumulativo=VERDADEIRO) ou relativa (cumulativo=FALSO).

Cálculo de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **DIST.F(x, graus_liberdade1, graus_liberdade2, cumulativo)**: retorna a probabilidade da variável “x” para uma **Distribuição F-Fisher** com “graus_liberdade1” e “graus_liberdade2” conhecidos, acumulada (cumulativo=VERDADEIRO) ou relativa (cumulativo=FALSO).

Essas funções permitem a construção de gráficos das distribuições para diferentes valores de variável aleatória.

Para calcular a probabilidade de ocorrência de uma variável aleatória em um intervalo (área sob o gráfico da distribuição), subtraia a probabilidade acumulada da variável aleatória no limite superior do intervalo da probabilidade acumulada da variável no limite inferior do intervalo.

Cálculo de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição
Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

O Excel possui ainda funções para cálculo do valor de uma variável aleatória dada a probabilidade de sua ocorrência (ou seja, a inversa) segundo a função de distribuição de probabilidade a que se refere.

- **INV.NORM.N(probabilidade,média,desv_padrão):** retorna o valor da variável “x” para uma **Distribuição Normal** com média e desvio padrão conhecidos dada a “probabilidade” de sua ocorrência.
- **INV.NORMP.N(probabilidade):** retorna o valor da variável “x” para uma **Distribuição Normal Padrão** dada a “probabilidade” de sua ocorrência.

Cálculo de Probabilidade

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- **INV.T(probabilidade, graus_liberdade)**: retorna o valor da variável “x” para uma **Distribuição t-Student** com “graus_liberdade” conhecido dada a “probabilidade” de sua ocorrência.
- **INV.QUIQUA(probabilidade, graus_liberdade)**: retorna o valor da variável “x” para uma **Distribuição Qui-quadrado** com “graus_liberdade” conhecido dada a “probabilidade” de sua ocorrência.
- **INV.F(probabilidade,graus`liberdade1,graus`liberdade2)**: retorna o valor da variável “x” para uma **Distribuição F-Fisher** com “graus_liberdade1” e “graus_liberdade2” conhecidos.

Exercícios

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

- (1) Considere o conjunto de observações de uma variável aleatória discreta indicado no arquivo disponibilizado no site do curso.
 - Calcule as medidas de estatística descritiva da amostra.
 - Construa os gráficos de frequência absoluta e relativa dessa variável.
 - Verifique a função de distribuição de probabilidade (contínua) que melhor representa as observações.
 - Identifique os parâmetros necessários para caracterização da função de distribuição de probabilidade.

Agenda

Bibliografia

Estatística
Descritiva

Medidas de
Posição

Medidas de
Dispersão

Histograma

Distribuições
Estatísticas

Padronização de
Distribuição
Normal

Cálculo de
Probabilidade

Exercícios

(1) Continuação...

- Calcule as probabilidades da variável aleatória ser igual a 5, maior ou igual a 3, menor ou igual a 2, e de estar no intervalo entre 1 e 4.