
ME414 - Estatística para Experimentalistas

Resumo Teórico

30 de junho de 2021

Guilherme Nunes Trofino
217276

Conteúdo

1	Análise Descritiva	3
1.1	Estrutura dos Dados	3
1.2	Média	3
1.3	Mediana	3
1.4	Moda	3
1.5	Variância	3
1.6	Desvio Padrão	4
1.7	Coefficiente de Correlação	4
1.8	Análise Descritiva Univariada	4
1.9	Análise Descritiva Bivariada	5
2	Ferramentas Gráficas	6
2.1	Histograma	6
2.2	Ramo-e-Folhas	7
2.3	Quartis	7
2.4	Boxplot	7
3	Análise Combinatória	9
3.1	Regra da Adição	9
3.2	Regra da Multiplicação	9
3.3	Permutação	9
3.4	Arranjo	9
3.5	Combinação	9
3.6	Amostragem	10
4	Probabilidade	11
4.1	Experimento	11
4.2	Lei dos Grandes Números	11
4.3	Espaço Amostral	11
4.4	Probabilidade Elemento	11
4.5	Probabilidade Evento	12
4.6	Probabilidade União	12
4.7	Probabilidade Condicional	12
4.8	Teorema de Bayes	12
5	Variáveis Aleatórias Discretas	13
5.1	Função de Distribuição Acumulada	13
5.2	Esperança	13
5.3	Mediana	13
5.4	Moda	14
5.5	Variância	14
5.6	Variável Aleatória Discreta Uniforme	14
5.7	Modelo de Bernoulli	14
5.8	Modelo Binomial	14
5.9	Modelo Geométrico	15
5.10	Modelo Hipergeométrico	15
5.11	Modelo de Poisson	15

6	Variáveis Aleatórias Contínuas	16
6.1	Função de Distribuição Acumulada	16
6.2	Esperança	16
6.3	Variância	16
6.4	Variável Aleatória Contínua Uniforme	17
6.5	Modelo Exponencial	17
6.6	Modelo Normal	17
6.7	Modelo Normal Padrão	17
7	Distribuição Amostral	18
7.1	Teorema do Limite Central	18
8	Inferência	19
8.1	Intervalo de Confiança 1 População	19
8.2	Hipótese	20
8.3	Erros	21
8.4	Intervalo de Confiança 2 População	21

1. Análise Descritiva

Definição Métodos para resumir e sintetizar dados obtidos de uma amostra, podendo ser resumidos através de diferentes métricas, adequadas as variáveis analisadas.

1.1. Estrutura dos Dados

Definição Organização das condições de um elemento estudado de acordo com sua estrutura e atributos, classificados por sua semelhança como mostrados a seguir:

1. **Quantitativos:** Números;
 - (a) **Contínuos:** Valores possíveis em um intervalo, aberto ou fechado, real;
 - (b) **Discretos:** Valores possíveis em um intervalo, aberto ou fechado, natural;
2. **Qualitativos:** Categorias;
 - (a) **Nominais:** Não possui ordenação;
 - (b) **Ordinais:** Possui ordenação;

1.2. Média

Definição Medida de posição central em que se x_1, \dots, x_n são n observações, a média aritmética será dada pela seguinte equação:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.2.1)$$

1.3. Mediana

Definição Medida de posição central em que divide o conjunto de dados em dois grupos, cada um com 50% das observações, sendo dada pela seguinte equação:

$$Q_2 = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par;} \end{cases} \quad (1.3.1)$$

1.4. Moda

Definição Métrica com maior número de ocorrências na amostra;

1.5. Variância

Definição Medida de dispersão em que a média dos desvios ao quadrado será dada pela seguinte equação:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.5.1)$$

1. **Coefficiente de Variação:** Normaliza o desvio padrão com relação a média do conjunto de dados de tal forma que os desvios padrões de diferentes amostras possam ser comparadas, será dado pela seguinte equação:

$$C_V = \frac{s}{\bar{x}} \quad (1.5.2)$$

1.6. Desvio Padrão

Definição Medida de dispersão em que a raiz quadrada da variância será dada pela seguinte equação:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.6.1)$$

1.7. Coeficiente de Correlação

Definição Quantificação da relação linear entre duas variáveis quantitativas, avaliando como cada uma delas estará deslocada proporcionalmente da média da amostra com n observações. Formalmente descrita pela seguinte equação:

$$-1 \leq C_C(x, y) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \cdot \left(\frac{y_i - \bar{y}}{s_y} \right) \leq 1 \quad (1.7.1)$$

Onde:

1. **Desvio Padronizado:** Define-se que a operação realizada representa quão distante uma observação x_i , ou y_i , está afastada de sua respectiva média \bar{x} , ou \bar{y} , sendo descrita pela seguinte equação:

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x} \quad z_{y_i} = \frac{y_i - \bar{y}}{s_y} \quad (1.7.2)$$

Assim classifica-se a relação entre duas variáveis de acordo com o coeficiente de correlação:

1. **Positivamente Associadas:** Quando as variáveis apresentam associação linear crescente, ou seja, apresentam um coeficiente de correlação:

$$0 < C_C(x, y) \leq 1$$

2. **Negativamente Associadas:** Quando as variáveis apresentam associação linear decrescente, ou seja, apresentam um coeficiente de correlação:

$$-1 \leq C_C(x, y) < 0$$

3. **Não Associadas:** Quando as variáveis não apresentam associação linear, ou seja, apresentam um coeficiente de correlação:

$$C_C(x, y) = 0$$

1.8. Análise Descritiva Univariada

Descrição Análise individual de cada variável presente no estudo, buscando classificá-la e obter métricas adequadas para resumi-lá. Recomenda-se as seguintes métricas para cada categoria analisada:

1. **Quantitativos:** ;
 - (a) **Contínuos:** Recomenda-se agrupar valores, quando muitos, em intervalos;
 - i. Gráfico de Histograma;
 - ii. Gráfico de Ramo-e-Folhas;
 - (b) **Discretos:** Recomenda-se agrupar valores, quando muitos, em intervalos;
 - i. Tabela de Frequências: Absolutas e Relativas;
 - ii. Gráfico de Barras;
2. **Qualitativos:** ;
 - (a) **Nominais:** Ordem Arbitraria;
 - i. Tabela de Frequências: Absolutas e Relativas;
 - ii. Gráfico de Barras;

- iii. Gráfico de Setores;
- (b) **Ordinais:** Ordem Sequencial;
 - i. Tabela de Frequências: Absolutas e Relativas;
 - ii. Gráfico de Barras;
 - iii. Gráfico de Setores;

1.9. Análise Descritiva Bivariada

Definição Análise conjunta de dois conjuntos de variáveis presente no estudo, buscando classificá-las, relacioná-las e obter métricas para resumi-las. Recomenda-se as seguintes métrica para cada categoria analisada:

1. **Qualitativas:** Comparação dos dados independentes em relações proporcionais aos diferentes subconjuntos dos dados;
 - (a) Porcentagens Relativas;
 - (b) Gráficos de Barras;
2. **Quantitativas:** Compação relacional entre as variáveis procurando explorar como uma influencia a outra;
 - (a) Coeficiente de Correlação: Gráfico relacionando as diferente variáveis em diferentes eixos;
 - (b) Gráfico de Dispersão: Medida que representa a associação linear entre variáveis quantitativas;

2. Ferramentas Gráficas

2.1. Histograma

Definição Representação gráfica de uma variável contínua que agrupada os valores em intervalos, abertos em uma extremidade e fechados na outra, regulares de acordo com as frequências dos dados, construídos como mostrado no exemplo a seguir:

114	122	103	118	99	105	134	125
117	106	109	104	111	127	133	111
117	103	120	98	100	130	141	119
128	106	109	115	113	121	100	130

Tabela 1: Distribuição de Dados Inicial

1. **Etapas:** Organize os dados em ordem crescente;

98	99	100	100	103	103	104	105
106	106	109	109	111	111	113	114
115	117	117	118	119	120	121	122
125	127	128	130	130	133	134	141

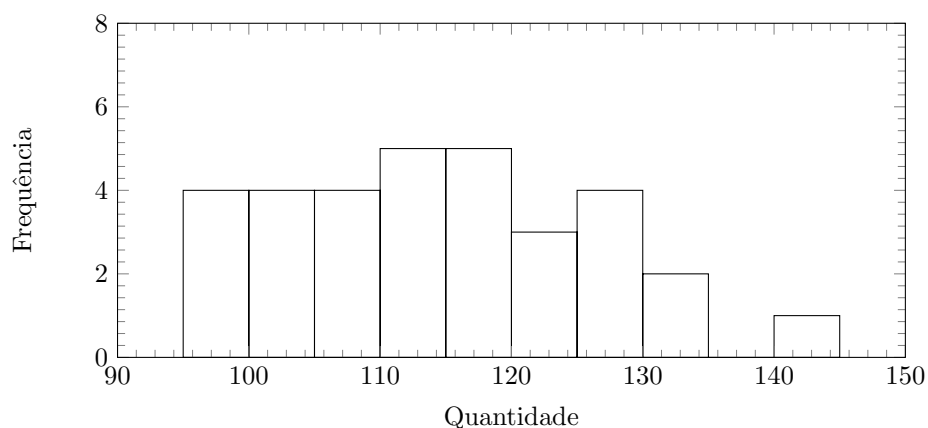
Tabela 2: Distribuição de Dados Ordenados

2. **Etapas:** Defina intervalos disjuntos, cada ocorrência em um único intervalo aberto à esquerda e fechado à direita;
3. **Etapas:** Construa uma tabela de frequências;

Intervalo	Frequência	Intervalo	Frequência
(95, 100]	4	(120, 125]	3
(100, 105]	4	(125, 130]	4
(105, 110]	4	(130, 135]	2
(110, 115]	5	(135, 140]	0
(115, 120]	5	(140, 145]	1

Tabela 3: Intervalos e Frequências

4. **Etapas:** Desenhe o gráfico com a frequência na ordenada e os intervalos na abscissas;



Estes gráficos possibilitam expressar dados de maneira eficiente, pois conclusões visuais auxiliam na análise dos dados apresentados. Há diferentes classificações de Histogramas, sendo apresentados abaixo:

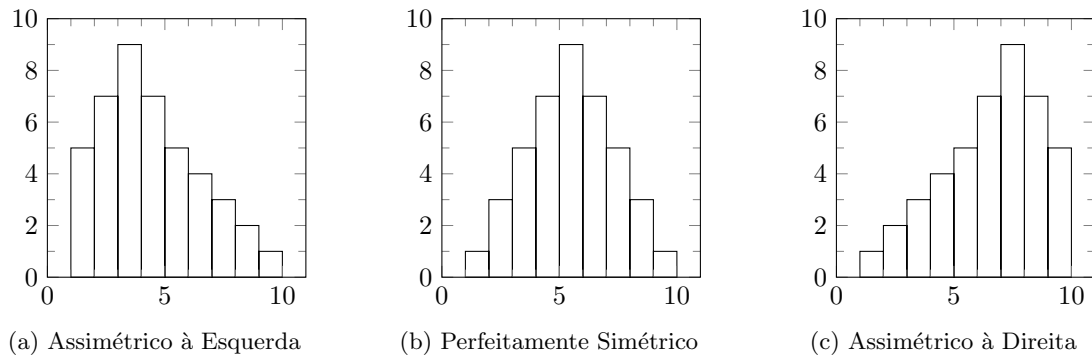


Figura 2.1: Simetria de Distribuição

2.2. Ramo-e-Folhas

Definição Representação Gráfica de uma contínua que agrupada os valores separando-os em duas partes: ramo, colocado a esquerda, e folhas, colocado a direita; mantendo as informações, construídos como mostrado no exemplo a seguir:

98	99	100	100	103	103	104	105
106	106	109	109	111	111	113	114
115	117	117	118	119	120	121	122
125	127	128	130	130	133	134	141

Tabela 4: Distribuição de Dados Ordenados

Em sequência separa-se os valores entre as dezenas e unidades, obtendo a tabela a seguir:

Ramo	Folha
9	8 9
10	0 0 3 3 4 5 6 6 9 9
11	1 1 3 4 5 7 7 8 9
12	0 1 2 5 7 8
13	0 0 3 4
14	1

Tabela 5: Distribuição de Ramos-e-Folhas

2.3. Quartis

Definição Representação gráfica que divide o conjunto de dados em 4 partes iguais: primeiro quartil, Q_1 ; segundo quartil, Q_2 ; e terceiro quartil, Q_3 . Obtidos através do seguinte procedimento:

1. **Ordenação:** Organizar os dados do conjunto amostral;
2. **Segundo Quartil:** Obter a mediana deste conjunto, Q_2 ;
3. **Primeiro Quartil:** Obter a mediana do conjunto inferior a Q_2 , Q_1 ;
4. **Terceiro Quartil:** Obter a mediana do conjunto superior a Q_2 , Q_3 ;

2.4. Boxplot

Definição Representação gráfica que permite resumir visualmente o esquema de 5 números, possibilitando analisar posição, dispersão, assimetria e outliers. Obtidos como representado abaixo:

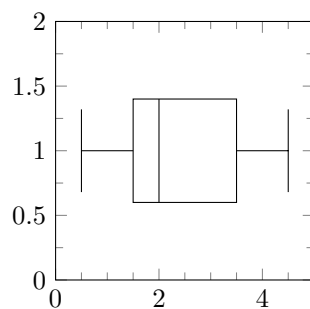


Figura 2.2: Representação de Boxplot

Neste diagrama temos, medidos da esquerda para direito, as seguintes informações sobre os dados apresentados:

1. **Limite Inferior:** Representa o valor mínimo da amostra, caso não hajam outliers, ou, caso hajam outliers, será obtido pela seguinte equação:

$$Q_1 - 1.5 \times IQ \quad (2.4.1)$$

2. Q_1 , **Primeiro Quartil**;
3. Q_2 , **Segundo Quartil**;
4. Q_3 , **Terceiro Quartil**;
5. IQ , **Intervalo Interquartilico:** Intervalo que compreende 50% dos dados da amostra, será obtido pela seguinte equação:

$$IQ = Q_3 - Q_1 \quad (2.4.2)$$

6. **Limite Superior:** Representa o valor máximo da amostra, caso não hajam outliers, ou, caso hajam outliers, será obtido pela seguinte equação:

$$Q_3 + 1.5 \times IQ \quad (2.4.3)$$

3. Análise Combinatória

3.1. Regra da Adição

Definição Suponha que hajam dois possíveis procedimentos, P_1 e P_2 , com diferentes formas de serem realizados, n_1 e n_2 , para executar uma mesma tarefa, então esta possuirá $n_1 + n_2$ formas de ser executada.

3.2. Regra da Multiplicação

Definição Suponha que hajam dois procedimentos necessários, P_1 e P_2 , com diferentes formas de serem realizados, n_1 e n_2 , para executar uma mesma tarefa, então esta possuirá $n_1 \times n_2$ formas de ser executada.

3.3. Permutação

Definição Suponha que tenhamos uma coleção $\Omega = \{\omega_1, \dots, \omega_n\}$ de n finitos objetos que devem ser dispostos em sequência, então:

Haverá n opções disponíveis para a primeira escolha, na sequência, haverá $n - 1$ opções disponíveis para a segunda escolha e assim suscetivamente até que reste 1 objeto da coleção. Isso ocorre, pois a **Permutação** permite trocar posições objetos de uma coleção para formação de uma sequência.

$$n! = n \times (n - 1) \times \dots \times 1 \quad (3.3.1)$$

Quando houverem r elementos com n_1, \dots, n_r repetições dentro uma coleção, respectivamente, então as organizações possíveis se reduzem:

$$\binom{n}{n_1, \dots, n_r} = \frac{n!}{n_1! \times \dots \times n_r!} \quad (3.3.2)$$

3.4. Arranjo

Definição Suponha que tenhamos uma coleção $\Omega = \{\omega_1, \dots, \omega_n\}$ de n finitos objetos que devem ser separadas em k grupos, então:

Haverá n opções disponíveis para a primeira escolha, na sequência, haverá $n - 1$ opções disponíveis para a segunda escolha e assim suscetivamente até que reste $n - k$ objetos da coleção. Isso ocorre, pois o **Arranjo** permite separar elementos de uma coleção em grupos considerando sua ordem.

$$A(n, k) = \frac{n!}{(n - k)!} = n \times (n - 1) \times \dots \times (n - k + 1) \quad (3.4.1)$$

3.5. Combinação

Definição Suponha que tenhamos uma coleção $\Omega = \{\omega_1, \dots, \omega_n\}$ de n finitos objetos que devem ser separadas em k grupos desconsiderando repetições, então:

Haverá n opções disponíveis para a primeira escolha, na sequência, haverá $n - 1$ opções disponíveis para a segunda escolha e assim suscetivamente até que reste $n - k$ objetos da coleção. Isso ocorre, pois a **Combinação** permite separar elementos de uma coleção em grupos desconsiderando sua ordem.

$$C(n, k) = \binom{n}{k} = \frac{n!}{k! \cdot (n - k)!} \quad (3.5.1)$$

3.6. Amostragem

Definição Seleção de elementos de um conjunto amostral que, de acordo com a seleção, terá melhores ou piores chances de representar o conjunto em sua completude, sendo classificados como descrito a seguir:

1. **Amostragem Aleatória Simples com Reposição:**

$$\boxed{N^n} \quad (3.6.1)$$

2. **Amostragem Aleatória Simples sem Reposição:**

(a) **Caso Não Ordenado:**

$$\boxed{\binom{N}{n} = \frac{N!}{(N-n)! \cdot n!}} \quad (3.6.2)$$

(b) **Caso Ordenado:**

$$\boxed{\frac{N!}{(N-n)!}} \quad (3.6.3)$$

4. Probabilidade

4.1. Experimento

Definição Processo que produza uma observação ou resultado que poderá ser reproduzido quantas vezes forem necessárias, sendo classificados como mostrado a seguir:

1. **Experimento Determinístico:** Aquele cujo o resultado obtido será conhecido;
2. **Experimento Aleatório:** Aquele cujo o resultado obtido será desconhecido;

4.2. Lei dos Grandes Números

Definição Segundo Jacob Bernoulli temos como mostrado a seguir:

Se um evento de probabilidade p é observado repetidamente em ocasiões independentes, a proporção da frequência observada deste evento em relação ao número total de repetições converge em direção a p à medida que o número de repetições se torna arbitrariamente grande.

4.3. Espaço Amostral

Definição Conjunto que agrega todas os possíveis resultados do experimento realidade, sendo denotado como mostrado a seguir:

$$\Omega = \{\omega_1, \dots, \omega_n\} \quad (4.3.1)$$

Onde:

1. **Elemento:** Acontecimento registrado, ω_i ;
2. **Espaço Amostral:** Conjunto de todos os eventos, Ω ;
3. **Evento:** Subconjunto do espaço amostra denotado por letras, A, B, \dots ;
 - (a) **Interseção de Eventos:** Elementos que pertencem simultaneamente ao evento A e ao evento B , $A \cap B$;
 - (b) **União de Eventos:** Elementos que pertencem ao evento A ou ao evento B , $A \cup B$;
 - i. **Eventos Disjuntos:** Nenhum elemento do evento A pertence ao evento B e vice-versa, $A \cap B = \emptyset$
 - (c) **Evento Complementar:** Se o evento A e B possuem $A \cap B = \emptyset$ e $A \cup B = \Omega$, então são complementares, $A = B^C$.

Nota-se que cada evento ou elemento possuirá uma probabilidade, uma chance, de ocorrer.

4.4. Probabilidade Elemento

Definição Seja ω_i um elemento amostral em Ω , então a probabilidade do elemento i , $P(\omega_i)$, ocorrer será:

$$\sum_{i=1}^n P(\omega_i) = 1 \quad 0 \leq P(\omega_i) \leq 1 \quad \begin{cases} P(\omega_i) = 1, & \text{Elemento Certo;} \\ P(\omega_i) = 0, & \text{Elemento Impossível;} \end{cases} \quad (4.4.1)$$

Onde:

1. **Equiprobabilidade:** Todos os elementos do espaço amostral possuem a mesma chance de ocorrer, formalmente descrito de acordo com a seguinte equação:

$$P(\omega_i) = \frac{1}{n}, \quad \forall i = 1, \dots, n \quad (4.4.2)$$

4.5. Probabilidade Evento

Definição Seja $A = \{\omega_1, \dots, \omega_m\}$ um evento em Ω com $m \leq n$ elementos amostrais, então a probabilidade do evento A , $P(A)$, ocorrer será:

$$P(A) = \frac{m}{n} \quad (4.5.1)$$

4.6. Probabilidade União

Definição Sejam A e B eventos de um espaço amostral Ω , então a probabilidade da união dos eventos A e B ocorrer será:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (4.6.1)$$

4.7. Probabilidade Condicional

Definição Sejam A e B eventos, então a probabilidade de que B ocorra dado que A ocorreu será:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (4.7.1)$$

Onde:

1. **Independência:** Quando informação sobre A não influencia a probabilidade do evento B , implicando:

$$P(A \cap B) = P(A) \cdot P(B) \quad (4.7.2)$$

4.8. Teorema de Bayes

Definição Seja $\{B_1, \dots, B_n\}$ uma partição de eventos de Ω e A mutuamente exclusivos cujo a união destes é Ω , então a probabilidade de A pode ser descrita pela seguinte expressão:

$$P(A) = \sum_{i=1}^n P(A \cap B_i)$$

Onde, aplicando probabilidade condicional, poderá ser reescrito como:

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

Assim, a probabilidade de um destes eventos da partição de A , B_i , ocorrer será dada pela seguinte equação:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{i=1}^n (P(A|B_i) \cdot P(B_i))} \quad (4.8.1)$$

5. Variáveis Aleatórias Discretas

Definição Experimento aleatório que resultada em uma quantidade numérica como resultado, ou seja, o evento em si não será prioridade. Caso uma função X seja utilizada para relacionar os elementos do espaço amostral a um conjunto enumerável de pontos da reta real então está será uma **Variável Aleatória Discreta**. Assim defini-se a probabilidade desta variável ocorrer como:

$$\boxed{\sum_{i=1}^n P(X = x_i) = 1} \quad \boxed{0 \leq P(X = x_i) \leq 1} \quad (5.0.1)$$

5.1. Função de Distribuição Acumulada

Definição Representação do somatório dos valores da probabilidade de uma variável aleatória X até um certo intervalos $\{x_1, \dots, x_n\}$, como representado na seguinte equação:

$$\boxed{F(x) = P(X \leq x), \quad x \in \mathbb{R}} \quad (5.1.1)$$

Pode-se representar essa variável, seguindo a seguinte notação:

$$\left\{ \begin{array}{ll} F(x_1) &= P(X = x_1) \\ \vdots & \vdots \\ F(x_i) &= P(X = x_1) + \dots + P(X = x_i) \\ \vdots & \vdots \\ F(x_n) &= P(X = x_1) + \dots + P(X = x_n) \end{array} \right. \quad F(x) = \left\{ \begin{array}{ll} F(x_1), & x \leq x_1 \\ \vdots & \\ F(x_i), & x_{i-1} \leq x \leq x_i \\ \vdots & \\ F(x_n), & x_{n-1} \leq x \leq x_n \end{array} \right. \quad (5.1.2)$$

5.2. Esperança

Definição Conhecido em inglês como **Expected Value**, representa qual seria o valor ponderado entre as probabilidades de todos os possíveis valores de X , podendo resultar em um valor impossível. Formalmente representada pela seguinte equação:

$$\boxed{E(X) = \mu = \sum_{i=1}^n x_i \cdot P(X = x_i)} \quad (5.2.1)$$

Obedecendo as seguintes propriedades:

1. Se X é uma variável aleatória e constantes a e b , então:

$$\boxed{E(aX + b) = aE(X) + b} \quad (5.2.2)$$

2. Se X_1, \dots, X_n são variáveis aleatórias:

$$\boxed{E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)} \quad (5.2.3)$$

5.3. Mediana

Definição Representa o valor de uma variável aleatória que a divide em dois conjuntos de igual probabilidade de ocorre, ou seja, atende a seguinte equação:

$$\boxed{P(X \geq Md) \geq \frac{1}{2}} \quad \text{e} \quad \boxed{P(X \leq Md) \geq \frac{1}{2}} \quad (5.3.1)$$

5.4. Moda

Definição Representa o valor da variável X com maior probabilidade de ocorrer, podendo ser mais do que um. Formalmente descrito pela seguinte equação:

$$P(X = Mo) = \max\{p_1, \dots, p_n\} \quad (5.4.1)$$

5.5. Variância

Definição Quantificação do quão distantes os valores de uma variável aleatória X estão de sua esperança, novamente uma métrica numérica que desconsidera o evento em si. Formalmente representada pela seguinte equação:

$$V(x) = \sigma^2 = E[(X - E(X))^2] = E(X^2) - (E(X))^2 \quad (5.5.1)$$

Obedecendo as seguintes propriedades:

1. Se X é uma variável aleatória e constantes a e b , então:

$$V(aX + b) = a^2 V(X) \quad (5.5.2)$$

2. Se X_1, \dots, X_n são variáveis aleatórias:

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) \quad (5.5.3)$$

5.6. Variável Aleatória Discreta Uniforme

Definição Quando uma variável aleatória discreta assume valores x_1, \dots, x_k possuam a mesma probabilidade de ocorrer, sendo descrita pela seguinte equação:

$$P(X = x_i) = p(x_i) = \frac{1}{k}, \quad \forall 1 \leq i \leq k \quad (5.6.1)$$

5.7. Modelo de Bernoulli

Definição Modelagem de experimentos aleatórios em que hajam apenas 2 resultados possíveis: **Sucesso** ou **Fracasso**. Denota-se assim, considerando que p seja a probabilidade de sucesso, a seguinte equação:

$$P(X = x) = \begin{cases} p, & \text{Sucesso se } x = 1 \\ 1 - p, & \text{Fracasso se } x = 0 \end{cases} \quad P(X = x) = p^x(1 - p)^{1-x} \quad (5.7.1)$$

Assim tem-se os seguintes resultados:

$$E(X) = p \quad V(X) = p(1 - p) \quad (5.7.2)$$

5.8. Modelo Binomial

Definição Modelo de experimentos de repetição de n ensaios X_i de Bernoulli independentes com mesma probabilidade p de sucesso. Assim a probabilidade de se observar x será dada pela seguinte equação:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (5.8.1)$$

Assim tem-se os seguintes resultados:

$$E(X) = np \quad V(X) = np(1 - p) \quad (5.8.2)$$

5.9. Modelo Geométrico

Definição Modelo de ensaios de Bernoulli independentes com mesma probabilidade p de sucesso que sejam repetidos até que o primeiro sucesso seja obtido. Assim a probabilidade de se observar x será dada pela seguinte equação:

$$P(X = x) = (1 - p)^{x-1}p \quad (5.9.1)$$

Assim tem-se os seguintes resultados:

$$E(X) = \frac{1}{p} \quad V(X) = \frac{1-p}{p^2} \quad (5.9.2)$$

5.10. Modelo Hipergeométrico

Definição Modelo em que a população de N objetos é dividida em duas características, r tem a característica A e $N - r$ tem a característica B , e realiza-se n extrações sem reposição. Assim a probabilidade de que neste grupo com n elementos possua x elementos da característica A será dada pela seguinte equação:

$$P(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \quad (5.10.1)$$

Assim tem-se os seguintes resultados:

$$E(X) = \frac{nr}{N} \quad V(X) = \frac{nr}{N} \left(1 - \frac{r}{N}\right) \frac{(N-n)}{(N-1)} \quad (5.10.2)$$

5.11. Modelo de Poisson

Definição Aproximação para distribuição binomial apresenta $n \rightarrow \infty$ e $p \rightarrow 0$, geralmente considera-se os seguintes critérios, $np \leq 7$ ou $n \geq 20$ e $p \leq 0.05$. Assim a probabilidade desta aproximação será dada pela seguinte equação:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \sim \frac{e^{-np} (np)^x}{x!} \quad (5.11.1)$$

Assim tem-se os seguintes resultados:

$$E(X) = V(X) = np = \lambda \quad (5.11.2)$$

6. Variáveis Aleatórias Contínuas

Definição Experimento aleatório que resultada em uma quantidade numérica como resultado, ou seja, o evento em si não será prioridade. Caso uma função X seja utilizada para relacionar os elementos do espaço amostral da reta real então está será uma **Variável Aleatória Contínua**. Assim defini-se a probabilidade desta variável ocorrer como:

$$\boxed{\int_{-\infty}^{+\infty} f_X(x)dx = 1} \quad \boxed{P(a \leq X \leq b) = \int_a^b f_X(x)dx \geq 0, \quad \forall x \in \mathbb{R}} \quad (6.0.1)$$

Onde:

$$\boxed{P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)} \quad (6.0.2)$$

6.1. Função de Distribuição Acumulada

Definição Representação a integral da função de probabilidade de uma variável aleatória X até um certo intervalos $\{x_1, \dots, x_n\}$, como representado na seguinte equação:

$$\boxed{F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u)du} \quad (6.1.1)$$

6.2. Esperança

Definição Conhecido em inglês como **Expected Value**, representa qual seria o valor ponderado entre as probabilidades de todos os possíveis valores de X , podendo resultar em um valor impossível. Formalmente representada pela seguinte equação:

$$\boxed{E(X) = \mu = \int_{-\infty}^{+\infty} x f_X(x)dx} \quad (6.2.1)$$

Obedecendo as seguintes propriedades:

1. Se X é uma variável aleatória, então o k -ésimo momento será dado por:

$$\boxed{E(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x)dx} \quad (6.2.2)$$

2. Se X é uma variável aleatória e constantes a e b , então:

$$\boxed{E(aX + b) = aE(X) + b} \quad (6.2.3)$$

3. Se X_1, \dots, X_n são variáveis aleatórias:

$$\boxed{E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)} \quad (6.2.4)$$

6.3. Variância

Definição Quantificação do quão distantes os valores de uma variável aleatória X estão de sua esperança, novamente uma métrica numérica que desconsidera o evento em si. Formalmente representada pela seguinte equação:

$$\boxed{V(X) = \sigma^2 = E[(X - E(X))^2] = E(X^2) - (E(X))^2} \quad (6.3.1)$$

Obedecendo as seguintes propriedades:

1. Se X é uma variável aleatória e constantes a e b , então:

$$\boxed{V(aX + b) = a^2 V(X)} \quad (6.3.2)$$

2. Se X_1, \dots, X_n são variáveis aleatórias:

$$\boxed{V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)} \quad (6.3.3)$$

6.4. Variável Aleatória Contínua Uniforme

Definição Quando uma variável aleatória contínua assume valores x_1, \dots, x_k possuam a mesma probabilidade de ocorrer, sendo descrita pela seguinte equação:

$$\boxed{f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x < a \text{ ou } x > b \end{cases}} \quad (6.4.1)$$

Assim tem-se os seguintes resultados:

$$\boxed{E(X) = \frac{b+a}{2}} \quad \boxed{V(X) = \frac{(b-a)^2}{12}} \quad (6.4.2)$$

6.5. Modelo Exponencial

Definição Quando uma variável aleatória, com média λ , contínua assume uma distribuição exponencial dos valores, sendo descrita pela seguinte equação:

$$\boxed{f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}} \quad \boxed{F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}} \quad (6.5.1)$$

Assim tem-se os seguintes resultados:

$$\boxed{E(X) = \frac{1}{\lambda}} \quad \boxed{V(X) = \frac{1}{\lambda^2}} \quad (6.5.2)$$

6.6. Modelo Normal

Definição Quando uma variável aleatória contínua assume uma distribuição normal dos valores, isto é, dependendo de sua média, μ , e variância, σ^2 , sendo descrita pela seguinte equação:

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \quad (6.6.1)$$

Assim tem-se os seguintes resultados:

$$\boxed{E(X) = \mu} \quad \boxed{V(X) = \sigma^2} \quad (6.6.2)$$

6.7. Modelo Normal Padrão

Definição Quando uma variável aleatória contínua assume uma distribuição normal padronizada dos valores, isto é, dependendo de sua média, μ , e variância, σ^2 , normalizadas, sendo descrita pela seguinte equação:

$$\boxed{Z(X) = \frac{X - \mu}{\sigma}} \quad \boxed{f_X(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}} \quad (6.7.1)$$

Assim tem-se os seguintes resultados:

$$\boxed{E(X) = 0} \quad \boxed{V(X) = 1} \quad (6.7.2)$$

7. Distribuição Amostral

Definição Descrição do comportamento de uma população através das características de uma amostra retirada aleatoriamente que poderá ser construída através de diferentes modelos.

Descrição Temos uma população com uma esperança, ou média, μ e variância σ^2 desconhecidas que deseja-se estimar a partir de uma amostra deste grupo. Retira-se então uma amostra aleatória de tamanho n e calcula-se sua esperança, ou média, amostral \bar{X} , ou \hat{p} , para estimar o parâmetro populacional desconhecido μ , implicando nos seguintes resultados:

$$\boxed{E(\bar{X}) = \mu} \quad \boxed{V(\bar{X}) = \frac{\sigma^2}{n}} \quad \boxed{EP(\bar{X}) = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}} \quad (7.0.1)$$

$$\boxed{E(\hat{p}) = p} \quad \boxed{V(\hat{p}) = \frac{p(1-p)}{n}} \quad \boxed{EP(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}} \quad (7.0.2)$$

Quando a amostral aumenta, tendendo ao tamanho da população, ao analisar os gráficos obtidos será possível aproximar as variáveis encontradas pela distribuição normal como mostrado nas seguintes equações:

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)} \quad \boxed{\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)} \quad (7.0.3)$$

Binomial Quando deseja-se obter o somatório de uma variável através de uma distribuição amostral pode-se realizar a aproximação pela **Distribuição Binomial** onde $S_n = X_1 + \dots + X_n$. Desta maneira, obtém-se o seguinte resultado:

$$\boxed{B(n, p) \sim N(np, np(1-p))} \quad (7.0.4)$$

7.1. Teorema do Limite Central

Definição Considera-se que as distribuições amostrais serão válidas pelo seguinte enunciado:

Considera-se uma amostra aleatória X_1, \dots, X_n coletada de uma população com média μ e variância σ^2 , então a distribuição amostral de \bar{X} aproxima-se de uma **Distribuição Normal** de média μ e variância $\frac{\sigma^2}{n}$ quando n for suficientemente grande:

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)} \quad \boxed{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0, 1)} \quad (7.1.1)$$

8. Inferência

Definição Descrição das características desconhecidas de uma população com base em uma amostral aleatória desta população. Qualquer análise realizada será realizada com um erro associado que deverá ser informado juntamente com os valores obtidos.

8.1. Intervalo de Confiança 1 População

Definição Representa um intervalo a partir de uma estimativa pontual $\hat{\theta}$ de uma variável θ que, a partir de uma margem de erro, poderá contemplar uma porcentagem do resultado real como descrito pela seguinte equação:

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right] \quad (8.1.1)$$

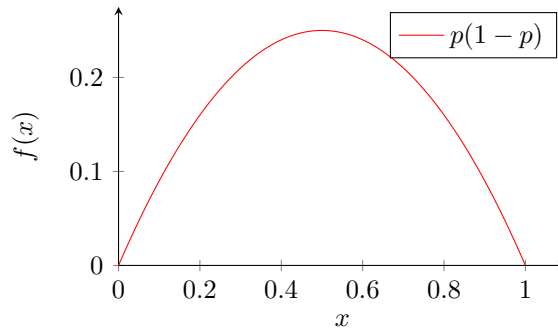
Onde:

1. p , **Proporção Populacional**: Parâmetro populacional desconhecido p que deseja-se encontrar;
2. \hat{p} , **Proporção Amostral**: Parâmetro amostral extraído para descrever o parâmetro p ;
3. $1 - \alpha$, **Intervalo de Confiança**: Representa a margem de erro aceitável para variável, descrito como:

$$1 - \alpha = \begin{cases} 90\%, & z_{\alpha/2} = 1.64 \\ 95\%, & z_{\alpha/2} = 1.96 \\ 99\%, & z_{\alpha/2} = 2.58 \end{cases} \quad (8.1.2)$$

Note que, quanto maior o intervalo de confiança aceitável, maior será a gama de valores aceitável. Este valores são obtidos pela análise da distribuição normal associada ao valores. Estes valores podem ser obtidos através de uma distribuição normal padrão.

Todavia, esta análise pressupõe que o parâmetro populacional p seja conhecido durante a análise o que não será sempre verdade. Desta maneira pode-se estudar qual seria o pior cenário desta distribuição, observando o seguinte gráfico:



Nota-se que esta função apresentará máximo, isto é, o pior resultado possível para o intervalo de confiança, com $p = \frac{1}{2}$ e portanto $p(1-p) = \frac{1}{4}$ desta maneira o cálculo se reduz para a seguinte expressão:

$$IC_C(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right] \quad (8.1.3)$$

Além dos resultados acima, aplicáveis para proporções, pode-se estender sua aplicação para variáveis que apresentem média amostral \bar{X} então as aproximações pelo **Teorema do Limite Central** serão válidos. Desta

maneira, considera-se as seguintes equações:

$$IC(\mu, 1 - \alpha) = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad (8.1.4)$$

Onde:

1. μ , **Média Populacional**: Parâmetro populacional conhecido;
2. σ^2 , **Variância Populacional**: Parâmetro populacional conhecido;
3. $1 - \alpha$, **Intervalo de Confiança**: Representa a margem de erro aceitável para variável, descrito como na Equação 8.1.2;

8.2. Hipótese

Definição Afirmações sobre um parâmetro de uma população que deverá ser testado. Desta maneira, será necessário estabelecer procedimentos a serem seguidos como explicados abaixo:

1. **Suposições**: Normalmente assume-se que os dados obtidos do experimento foram produzidos através de um processo de aleatorização;
2. **Hipóteses**: Afirmação sobre um parâmetro populacional desejado que deverá ser definido antes da realização do experimento;
 - (a) H_0 , **Hipótese Nula**: Parâmetro populacional assume um dado valor;
 - (b) H_A , **Hipótese Alternativa**: Parâmetro populacional assume valores diferentes de H_0 ;
3. **Estatísticas**: Comparação entre os valores encontradas na população após experimentação e os esperados pela **Hipótese**, obtido pela seguinte equação:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1) \quad \text{onde} \quad \begin{cases} np_0 \geq 10 \\ n(1-p_0) \geq 10 \end{cases}$$

Onde:

- (a) \hat{p} **Proporção Amostral**: Obtida pela seguinte equação:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

- (b) p_0 **Proporção Populacional**;

4. **p-valor**: Probabilidade que resume a possibilidade de que a **Hipótese Nula** seja verdadeira, obtido pela seguinte equação:

$$\mathbf{p\text{-}valor} = \begin{cases} P(|Z| \geq |z|) & \hat{p} \neq p_0 & \text{Bilateral} \\ P(Z \leq z) & \hat{p} < p_0 & \text{Unilateral à Esquerda} \\ P(Z \geq z) & \hat{p} > p_0 & \text{Unilateral à Direita} \end{cases} \quad (8.2.1)$$

5. **Conclusão**: Análise baseada no **p-valor** com base em um **Nível de Significância** $\alpha = 0.05$, representando uma constante usual que poderá ser interpretada da seguinte maneira:
 - (a) Se **p-valor** $\leq \alpha$: rejeita-se H_0 , há evidência contra a **Hipótese Nula**;
 - (b) Se **p-valor** $> \alpha$: não rejeita-se H_0 , não há evidência contra a **Hipótese Nula**;

8.3. Erros

Definição Quando realiza-se um teste de hipóteses haveram diferentes combinações entre o que de fato e verdadeiro e o que foi obtido através da análise. Alguns resultados possíveis são demonstrados na seguinte tabela:

Decisão	H_0 Verdadeira	H_0 Falsa
Rejeitar H_0	Erro do Tipo I	-
Aceitar H_0	-	Erro do Tipo II

Tabela 6: Classificação Erros

Onde:

1. **Erro do Tipo I:** Normalmente atribuído como **Falso Positivo**, apresentando a seguinte probabilidade de ocorrer, também nomeada **Nível de Significância**:

$$\alpha = P(\text{Erro do Tipo I}) \quad (8.3.1)$$

2. **Erro do Tipo II:** Normalmente atribuído como **Falso Negativo**, apresentando a seguinte probabilidade de ocorrer:

$$\beta = P(\text{Erro do Tipo II}) \quad (8.3.2)$$

Idealmente deseja-se que estas probabilidades sejam baixas, indicando hipóteses confiáveis. Entretanto, à medida que diminui-se α , β tende a aumentar. Desta maneira, estipula-se usualmente que $\alpha = 5\%$ para que os resultados sejam adequados.

8.4. Intervalo de Confiança 2 População

Definição Representa relação entre uma variável dentro de duas populações independentes, 1 e 2, que possuem respectivamente μ_1 e μ_2 como média populacional, σ_1^2 e σ_2^2 como variância e n e m de tamanho. Assim um intervalo a partir de uma margem de erro, poderá contemplar uma porcentagem do resultado real será descrito pela seguinte equação:

$$IC(\mu_1 - \mu_2, 1 - \alpha) = \left[(\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}; (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right] \quad (8.4.1)$$

Onde:

1. $\mu_1 - \mu_2$, **Diferença da Média:** Representa a relação entre a diferença as médias das populações;
2. $\bar{x} - \bar{y}$, **Diferença da Amostral:** Representa a relação entre a diferença as populações;

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$$

3. $1 - \alpha$, **Intervalo de Confiança:** Representa a margem de erro aceitável para variável, descrito como uma distribuição normal:

$$1 - \alpha = \begin{cases} 90\%, & z_{\alpha/2} = 1.64 \\ 95\%, & z_{\alpha/2} = 1.96 \\ 99\%, & z_{\alpha/2} = 2.58 \end{cases} \quad (8.4.2)$$

Note que quando as variâncias populacionais forem desconhecidas, será necessário realizar as seguintes contas:

$$IC(\mu_1 - \mu_2, 1 - \alpha) = \left[(\bar{x} - \bar{y}) - t \sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}; (\bar{x} - \bar{y}) + t \sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)} \right] \quad (8.4.3)$$

Onde:

1. S_p^2 , **Variância:** Representa a variância ponderada entre as populações, obtido pela seguinte equação:

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \quad (8.4.4)$$

2. $1 - \alpha$, **Intervalo de Confiança:** Representa a margem de erro aceitável para variável, descrito como uma distribuição t-student:

$$1 - \alpha = \begin{cases} 90\%, & t = t_{n+m-2, \alpha/2} = 1.64 \\ 95\%, & t = t_{n+m-2, \alpha/2} = 1.96 \\ 99\%, & t = t_{n+m-2, \alpha/2} = 2.58 \end{cases} \quad (8.4.5)$$