

Cálculo das Probabilidades e Estatística I

Departamento de Estatística

Versão - 2013

Sumário

1	Introdução à Estatística	1
1.1	Conceitos básicos de amostragem	2
1.1.1	Tipos de variáveis	3
1.1.2	Níveis de Mensuração	4
1.1.3	Tipos de estudos	4
1.1.4	Tipos de Amostragem	5
1.2	Principais planos de amostragem probabilística	5
1.2.1	Erros de amostragem	7
2	Estatística Descritiva	9
2.1	Tabela de distribuição de freqüências	9
2.2	Distribuição de freqüências por valores	10
2.3	Distribuição de Freqüências por classes ou intervalos	11
2.3.1	Regras para elaboração da Tabela de Distribuição de Freqüências	11
2.4	Elementos em uma tabela de Distribuição de Freqüências	12
2.5	Medidas de Tendência Central	14
2.5.1	Média Aritmética	15
2.5.2	Moda	16
2.5.3	Mediana	17
2.5.4	Assimetria	18
2.6	Medidas de Dispersão	18
3	Introdução a Probabilidade	23
3.1	Probabilidade Condicional	27
4	Variáveis aleatórias	31
4.1	Conceitos e definições	31
4.2	Classificação das variáveis aleatórias	32
4.3	Esperança de uma variável aleatória	35
4.3.1	Propriedades da Esperança	36
4.4	Variância de uma variável aleatória	37
4.4.1	Propriedades da Variância	37
5	Modelos Probabilísticos para variáveis aleatórias	39
5.1	Modelos Probabilísticos para variáveis aleatórias discretas	39
5.1.1	Distribuição de Bernoulli	39

5.1.2	Distribuição Binomial	40
5.1.3	Distribuição de Poisson	42
5.2	Modelos Probabilísticos para variáveis aleatórias contínuas	44
5.2.1	Distribuição Normal	44
5.2.2	Distribuição t-Student	46
6	Distribuições Amostrais	47
6.1	Distribuição Amostral da Média	48
6.2	Distribuição Amostral da Proporção	49
7	Inferência Estatística	51
7.1	Estimação Pontual	51
7.1.1	Propriedades de um estimador	52
7.2	Intervalo de Confiança	53
7.3	Intervalo de Confiança para a Média	54
7.3.1	Caso 1: X possui distribuição normal com Variância conhecida.	54
7.3.2	Caso 2: X possui distribuição normal com Variância desconhecida.	54
7.3.3	Caso 3: Grandes Amostras: $n \geq 30$	54
7.4	Intervalo de Confiança para a proporção	55
7.5	Teste de Hipótese	56
7.6	Procedimento Geral do Teste de Hipótese - Uma Amostra	58
7.7	Teste de hipótese para a média	60
7.7.1	Caso 1: X possui distribuição normal com Variância conhecida.	60
7.7.2	Caso 2: X possui distribuição normal com Variância desconhecida.	60
7.7.3	Caso 3: Grandes Amostras: $n \geq 30$	61
7.8	Teste de hipótese para a proporção	61
8	Correlação e Regressão Linear Simples	63
8.1	Coefficiente de Correlação Linear(ρ)	63
8.1.1	Interpretação geométrica	64
8.1.2	Teste de hipótese para o Coeficiente de Correlação	65
8.2	Regressão Linear Simples	65
8.3	Estimação dos parâmetros	65
8.3.1	Coeficiente de Determinação (R^2)	66
	Bibliografia	66

Capítulo 1

Introdução à Estatística

Ainda hoje no conceito popular, a palavra estatística evoca dados numéricos apresentados em quadros ou gráficos, publicados por agências governamentais, referentes a fatos demográficos ou econômicos. A palavra estatística é derivada da palavra latina status, que significa estado, usada aqui para designar a coleta e a apresentação de dados quantitativos de interesse do Estado. Entretanto, a mera coleta de dados assim apresentados está longe de ser o que entendemos, hoje, por Estatística. Na verdade, sua feição essencial é a de ser um conjunto de métodos estatísticos, especialmente apropriados ao tratamento de dados afetados por uma multiplicidade de causas. Esses métodos fazem uso da Matemática, particularmente do cálculo de probabilidades, na coleta, apresentação, análise e interpretação dos dados.

Essa prática tem sido continuada nos tempos modernos, por meio dos recenseamentos, dos quais temos um exemplo naquele que se efetua a cada decênio, em nosso País, pela Fundação IBGE, órgão responsável por nossas estatísticas (dados estatísticos) oficiais.

A primeira tentativa para se tirar conclusões a partir de dados numéricos foi feita somente no século 17, na Inglaterra em [GRAUNT \(1662\)](#). Graunt baseou sua análise sobre razões e proporções de fatos vitais, nos quais ele observou uma regularidade estatística em um grande número de dados. Graunt colocou os dados em tabelas e através de cálculos básicos produziu alguns comentários sobre os resultados obtidos, analisando a confiabilidade dos dados e comparando o número dos nascimentos e das mortes masculinas e femininas. No capítulo XI, [GRAUNT \(1662\)](#), Graunt produz uma tabela de vida primitiva, estas tabelas transformam-se mais tarde em uma das principais ferramentas da demografia e do seguro.

Entretanto, a estatística só começou realmente a existir como disciplina autônoma no início do século 20, o verdadeiro início da estatística moderna. Fisher (1890-1962) foi um dos mais influentes estatísticos do século 20. Em [FISHER \(1922\)](#) são apresentados os fundamentos matemáticos da Teoria Estatística, neste trabalho Fisher introduz os termos estimação e estimativa. Neste mesmo trabalho é apresentado três critérios de estimação, mais precisamente, propriedades que os estimadores devem ter: Consistência, eficiência e suficiência. Um marco importante na Teoria da Estatística moderna foi a formalização da teoria das probabilidades feita por Kolmogorov em 1933, [KOLMOGOROV \(1956\)](#), .

Uma definição moderna para estatística poderia ser: uma coleção de métodos para planejar experimentos, obter e organizar dados, resumí-los, analisá-los, interpretá-los e deles extrair conclusões. Pode-se afirmar portanto, que o foco da estatística é estudar os fenômenos coletivos.

Sendo assim, a primeira etapa em uma pesquisa que envolverá procedimentos estatísticos é

a coleta de dados. A coleta de dados é uma fase crucial na estatística pois se os mesmos não forem colhidos de maneira adequada as outras fases do processo de análise estatística estarão definitivamente comprometidos. A seguir serão dados alguns conceitos e informações básicas sobre o planejamento e coleta de dados.

1.1 Conceitos básicos de amostragem

Amostragem é o procedimento utilizado na obtenção da amostra, que deve ser de tal forma que a amostra obtida seja representativa da população de interesse.

Todos nós em nosso dia a dia temos contato com a amostragem, por exemplo, quando alguém está adoçando uma xícara de café ele primeiro coloca um pouco de açúcar, mistura bem e depois prova (coleta uma amostra) para verificar se precisa ou não mais açúcar. Note que, o processo de mexer bem antes de provar é um procedimento (plano) amostral intuitivo. Entretanto, neste caso, a amostra poderia não ser representativa do todo se a pessoa não mexesse bem, e por conseguinte, poderia-se colocar mais açúcar quando na verdade não precisaria de mais ou, não colocar mais açúcar quando de fato precisaria de mais. Logo, um procedimento amostral mal elaborado ou mal executado pode levar a uma conclusão errônea devido a um viés de interpretação do resultado. Portanto, planos (planejamentos) amostrais que produzam amostras representativas e conseqüentemente resultados confiáveis e livres de possíveis vieses é o objetivo principal do pesquisador. A seguir, serão apresentados alguns conceitos e termos técnicos que são utilizados na Teoria da Amostragem.

Definição 1.1 (População ou População alvo). *É o conjunto de todos os seres, objetos ou informações que estão sob investigação.*

Notação: *Um população de tamanho N será denotada por $\underline{U} = (1, \dots, N)$.*

Exemplo 1.1. *Um grupo de pesquisadores desejam analisar a influência de fatores sociodemográficos, físicos e mentais sobre a mobilidade de idosos, pessoas com 60 anos ou mais, residentes no município de Santa Cruz, Rio Grande do Norte. Neste caso a população são todas as pessoas com 60 anos ou mais residentes no município de Santa Cruz.*

Definição 1.2 (População de estudo). *É o conjunto de todos os seres, objetos ou informações que poderiam ser incluídas no estudo. Teoricamente, o mesmo que a população alvo, porém muitas vezes diferente.*

Exemplo 1.2. *No Exemplo 1.1 suponha que a pesquisa tenha sido realizada durante um determinado mês do ano, e que neste mês possivelmente algumas das pessoas desta população poderiam não estar na cidade e deste modo não poderiam ser incluídas na pesquisa. Deste modo, neste caso, a população alvo é diferente da população de estudo.*

Definição 1.3 (Censo). *É o levantamento de informações de toda uma população.*

Definição 1.4 (Amostra). *É o conjunto dos elementos selecionados de uma população.*

Notação: *Uma amostra de tamanho n será denotada por $\underline{s} = (k_1, \dots, k_n)$ para $k_i \in \underline{U}$.*

Definição 1.5 (Unidade amostral ou elementar). *São os elementos alvo da pesquisa. Podem ser pessoas, animais, objetos, domicílios, empresas, etc. Deve ser definida no início da investigação de acordo com o interesse do estudo. É muito importante que a unidade elementar seja claramente definida, para que o processo de coleta e análise tenha sempre um significado preciso e uniforme.*

Definição 1.6 (Variáveis). *É uma característica qualitativa ou quantitativa que observamos em cada unidade amostral. Ex.: altura, sexo, peso, idade, classe social, etc.*

Notação: As variáveis são usualmente denotadas pelas letras maiúsculas X, Y, Z, W .

- Em uma população $\underline{U} = (1, \dots, N)$, o conjunto de valores que essas variáveis assumem são denotadas por $\underline{x} = (x_1, x_2, \dots, x_N)$;
- Em uma amostra $\underline{s} = (k_1, \dots, k_n)$, os valores que essas variáveis podem assumir são denotadas por $\underline{X} = (X_1, X_2, \dots, X_n)$ em que cada X_i pode assumir qualquer valor x_u , para $u \in \underline{U}$ e $x_u \in \underline{x}$.

Definição 1.7 (Parâmetro). *Uma medida numérica que descreve alguma característica de uma população, por exemplo, peso médio ao nascer de crianças na cidade de João Pessoa, proporção de peças defeituosas produzidas em um dia em uma linha de produção.*

Notação: Utiliza-se usualmente letras gregas, μ, σ^2, τ para se denotar parâmetros. Entretanto, existem exceções, por exemplo, para o parâmetro proporção utiliza-se p .

Definição 1.8 (Estimador). *É qualquer função dos elementos X_1, \dots, X_n da amostra \underline{X} , que assume valores em Θ (espaço paramétrico), em que Θ é o conjunto de todos os valores que o parâmetro θ pode assumir.*

Notação: Usualmente utiliza-se $\hat{\mu}, \hat{\sigma}^2, \hat{p}$ para se denotar parâmetros. Entretanto, existem exceções, por exemplo, para o parâmetro μ utiliza-se \bar{X} .

Exemplo 1.3. Seja $\underline{X} = (X_1, X_2, \dots, X_n)$, então um estimador para a média populacional μ para essa amostra é dada por:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

Definição 1.9 (Estimativa). *É o valor observado de um estimador após a amostra ser coletada.*

Exemplo 1.4. Considere a seguinte amostra da variável X , $\underline{X} = (5, 3, 4, 2, 6)$, então

$$\bar{X} = \frac{5 + 3 + 4 + 2 + 6}{5} = 4.$$

Definição 1.10 (Cadastro amostral). *Lista das unidades da população de pesquisa de onde a amostra será extraída. Nem sempre aplicável.*

1.1.1 Tipos de variáveis

As variáveis podem ser: qualitativas ou quantitativas.

- Qualitativas: são variáveis categóricas.
 - Nominal: Não existe nenhuma relação entre as categorias. Ex.: sexo (masc, fem), curso (fisioterapia, Enfermagem, etc), procedência.
 - Ordinal: Existe uma ordenação natural entre as categorias. Ex.: Grau de instrução (1o grau, 2o grau, superior), nível socio-econômico (A, B, C, D).
- Quantitativas: são variáveis numéricas.

- Discreta: Admitem somente números inteiros. Ex: batimentos cardíacos, número de filhos.
- contínuas: os valores podem ser qualquer número real. Ex.: tempo de coagulação, peso, altura.

1.1.2 Níveis de Mensuração

1. Escala Nominal: as unidades amostrais são classificadas em categorias segundo uma característica. Ex.: sexo(Masc, Fem), Hábito de fumar(fumante, não fumante), sobrepeso(sim, não).

Observação 1.1. *Note que, não existe ordem entre as categorias e suas representações, se numéricas são destituídas de significado numérico. Ex.: sexo masculino=1, sexo feminino=2*

2. Escala Ordinal: as unidades amostrais são classificadas em categorias que possuem algum tipo inerente de ordem. Ex.: nível socio-econômico(A,B,C,D), nível de retinol sérico(alto, aceitável, baixo, deficiente).

Observação 1.2. *Embora exista ordem entre as categorias, a diferença entre as categorias adjacentes não tem o mesmo significado em toda a escala.*

3. Escala intervalar: Neste nível de mensuração podemos quantificar as diferenças entre as categorias. Entretanto, o zero nesta escala é arbitrário. Ex.: Temperatura (graus Celsius, Fahrenheit).

Observação 1.3. *Nesta escala, embora pode-se quantificar as diferenças entre as categorias, essas diferenças não são absolutas. Por exemplo, $50^{\circ}C$ embora seja o dobro de $25^{\circ}C$, não implica que é duas vezes mais quente, pois se mudarmos a unidade de medida para Fahrenheit teremos $50^{\circ}C = 50 \times 1,8 + 32 = 122^{\circ}F$ e $25^{\circ}C = 25 \times 1,8 + 32 = 77^{\circ}F$ o que implica que nesta unidade a razão entre as duas temperaturas é 1,58.*

4. Escala das razões: Nesta escala o zero é absoluto, isto implica que a razão entre duas medidas é igual independentemente da unidade que está sendo utilizada. Ex.: Altura(cm, m), peso(g, Kg).

1.1.3 Tipos de estudos

Em princípio, pode-se dizer que os estudos científicos se dividem em dois grupos: estudos observacionais e os estudos experimentais.

Estudos Observacionais. Se caracterizam pela não intervenção do pesquisador sobre os dados do estudo. De um modo geral, esses estudos efetuam descrições a respeito de um determinado problema, como, por exemplo: a estimativa da proporção de peças defeituosas em uma linha de produção, ou a estimativa do número médio de chamadas atendidas em central PABX. Em resumo, em um estudo observacional, o pesquisador observa e mede, mas não modifica;

Estudos Experimentais. Nos estudos experimentais, o pesquisador intervém sobre os elementos pesquisados, mediante a adoção de algum tratamento ou mediante a alteração da situação. Nesses casos, pretende-se comparar os resultados obtidos nas diversas situações ou tratamentos com a finalidade de detectar diferenças nos dados.

Exemplo 1.5. *Um engenheiro precisa saber se a quantidade de corrosão em uma tubulação utilizada pela sua empresa depende do tipo do revestimento usado ou do tipo de solo em que se encontra a tubulação. Deste modo, planejou-se o seguinte experimento: utilizou-se quatro diferentes revestimentos e três diferentes tipo de solo no experimento. Selecionaram-se 12 peças de tubulação e cada uma é revestida com um dos quatro revestimentos e enterrada em um dos três tipos de solo durante um período fixo de tempo, após o qual se determina a quantidade de corrosão.*

1.1.4 Tipos de Amostragem

Quando se realiza um estudo observacional, a coleta dos dados pode ser feita através de uma amostragem probabilística ou de uma amostragem não probabilística.

Definição 1.11 (Amostragem Probabilística). *É o procedimento pelo qual se utilizam mecanismos aleatórios de seleção dos elementos de uma amostra, atribuindo a cada elemento uma probabilidade de pertencer a amostra.*

Definição 1.12 (Amostragem não Probabilística). *É o procedimento pelo qual se não utilizam mecanismos aleatórios de seleção dos elementos de uma amostra, tais como: amostras intencionais, nas quais os elementos são escolhidos com o auxílio de especialistas; e amostras de voluntários, como ocorre em alguns experimentos sobre novos medicamentos e vacinas.*

Observação 1.4. *A grande vantagem da amostra probabilística é medir a precisão da amostra obtida, baseando-se apenas no resultado contido na própria amostra.*

1.2 Principais planos de amostragem probabilística

Serão apresentados os planos de amostragem, para os casos mais comuns na prática. Estes casos satisfazem os seguintes pressupostos: População finita e amostragem sem reposição.

Amostragem aleatória(AA): Procedimento pelo qual cada elemento da população tem a mesma chance(probabilidade) de ser selecionada.

Amostragem aleatória simples(AAS): Procedimento pelo qual uma amostra de tamanho n é selecionada de tal forma que cada amostra possível de tamanho n tem a mesma chance(probabilidade) de ser selecionada. Esse plano amostral subdivide-se ainda em dois outros: Amostragem aleatória simples com reposição(AASCR) e Amostragem aleatória simples sem reposição(AASSR).

Exemplo 1.6 (Diferença entre AA e AAS). *Imagine uma sala com 48 alunos, distribuídos em 8 fileiras. Suponha que o professor deseja selecionar uma amostra de 8 alunos. Assim, coloca-se em uma urna 8 bolas numeradas de 1 a 8. Seleciona-se ao acaso uma bola e verifica-se seu número. A amostra será a fileira selecionada. A amostra selecionada é uma amostra aleatória(AA)? É uma amostra aleatória simples(AAS)?*

Exemplo 1.7. Considere uma população $U = (1, 2, 3, 4, 5)$ e uma amostra $\underline{s} = (k_1, k_2, k_3)$. Determine todas as amostra possíveis para um plano amostral com reposição, $S(A)$, e para um plano amostral sem reposição, $S(B)$, de tamanho 3.

- Para o plano amostral com reposição tem-se que:

$$S(A) = \left\{ \begin{aligned} \underline{s}_1 &= (1, 1, 1), \underline{s}_2 = (1, 1, 2), \underline{s}_3 = (1, 1, 3), \underline{s}_4 = (1, 1, 4), \underline{s}_5 = (1, 1, 5), \\ \underline{s}_6 &= (1, 2, 1), \underline{s}_7 = (1, 2, 2), \underline{s}_8 = (1, 2, 3), \underline{s}_9 = (1, 2, 4), \underline{s}_{10} = (1, 2, 5), \\ &\vdots \\ \underline{s}_{121} &= (5, 5, 1), \underline{s}_{122} = (5, 5, 2), \underline{s}_{123} = (5, 5, 3), \underline{s}_{124} = (5, 5, 4), \underline{s}_{125} = (5, 5, 5) \end{aligned} \right\}$$

- Para o plano amostral sem reposição tem-se que:

$$S(B) = \left\{ \begin{aligned} \underline{s}_1 &= (1, 2, 3), \underline{s}_2 = (1, 2, 4), \underline{s}_3 = (1, 2, 5), \underline{s}_4 = (1, 3, 2), \underline{s}_5 = (1, 3, 4), \underline{s}_6 = (1, 3, 5) \\ &\vdots \\ \underline{s}_{55} &= (5, 3, 1), \underline{s}_{56} = (5, 3, 2), \underline{s}_{57} = (5, 3, 4), \underline{s}_{58} = (5, 4, 1), \underline{s}_{59} = (5, 4, 2), \underline{s}_{60} = (5, 4, 3) \end{aligned} \right\}$$

Amostragem sistemática: É realizada quando os elementos da população estão ordenados e a seleção dos elementos da amostra é feita periodicamente ou sistematicamente.

Exemplo 1.8. Deseja-se selecionar uma amostra de tamanho 30 de um cadastro amostral com 500 elementos. Seja,

$$k = \frac{500}{30} = 16,7.$$

Então, como k não é inteiro arredondamos para o maior inteiro menor igual a 16,7. Assim, $k = 16$. Agora selecionamos ao acaso um número entre 1 e k , para isso utilize um gerador de números aleatórios, por exemplo. Suponha que o número sorteado seja 9. Assim os elementos da amostra serão 9, 25, 41, ..., 473.

Amostragem estratificada: Esse procedimento consiste em dividir a população em sub-populações (estratos). Estratos são divisões de acordo com algum critério, por exemplo: sexo, faixa etária, estado civil, assim dentro de cada estrato teremos uma maior homogeneidade. Dessa forma, para uma população com N unidades amostrais e d estratos com tamanhos N_1, \dots, N_d , tem-se que $\sum_{i=1}^d N_i = N$, portanto teremos os seguinte coeficiente de proporcionalidade $c_i = \frac{N_i}{N}$. Deste modo, para uma amostra de tamanho n devemos selecionar uma AAS de tamanho $n_i = c_i \times n$ de cada estrato.

Exemplo 1.9. Suponha no exemplo anterior que tenhamos dois estratos (masculino, feminino), em que $N_1 = 290$ e $N_2 = 210$. Assim, $c_1 = \frac{290}{500} = 0,58$ e $c_2 = \frac{210}{500} = 0,42$. Logo, $0,58 \times 30 = 17,4$ e $0,42 \times 30 = 12,6$. Portanto, $n_1 = 17$ e $n_2 = 13$. O próxima etapa é coletar uma amostra utilizando o plano AASCR ou AASSR de tamanho $n_1 = 17$ para o 1o. estrato e $n_2 = 13$ para o 2o. estrato.

Observação 1.5. *Este é o tipo de amostragem que produz o menor erro.*

Amostragem por conglomerado: Neste procedimento cada unidade amostral é um grupo (conglomerado) de elementos. Conglomerados são partes representativas da população, por exemplo, dividimos um bairro em quarteirões. Assim cada quarteirão é uma unidade amostral. Deste modo, selecionamos uma AAS dos quarteirões para depois proceder-se o levantamento dos dados de todos os elementos do Conglomerado.

Observação 1.6. *Este é o tipo de amostragem que produz o maior erro entre os procedimentos apresentados.*

1.2.1 Erros de amostragem

Sempre que coletarmos uma amostra e a partir dela procuramos estimar certos parâmetros populacionais de interesse, estaremos sujeitos a cometer algum erro, não importa o quão bem planejando tenha sido a coleta dos dados. Pode-se classificar os erros de amostragem em dois tipos: **Erro amostral** e **erro não amostral**.

Erro amostral(E). é a diferença entre o resultado amostral e o verdadeiro resultado da população. Tais erros resultam das flutuações amostrais devidas ao acaso.

Exemplo 1.10. *Seja $\underline{x} = (x_1, \dots, x_{100})$ os valores de uma certa característica em uma população. Seja $\underline{X} = (X_1, \dots, X_5)$ uma amostra de tamanho 5 do vetor de característica da população \underline{x} . Suponha que $X_1 = x_3, X_2 = x_{52}, X_3 = x_{11}, X_4 = x_{77}, X_5 = x_{31}$. Então, por exemplo, a média populacional é dada por,*

$$\mu = \frac{x_1 + x_2 + \dots + x_{100}}{100}$$

e a média amostral,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_5}{5} = \frac{x_3 + x_{52} + x_{11} + x_{77} + x_{31}}{5}$$

Assim o erro amostral neste caso será $E = \bar{X} - \mu$.

Erro não amostral. ocorre quando os dados amostrais são coletados ou registrados incorretamente. Exemplos de erros não amostrais: seleção de uma amostra por conveniência, uso de um instrumento de medida defeituoso, digitação incorreta dos dados, etc.

Observação 1.7. *Se coletarmos uma amostra de maneira apropriada de modo que ela seja representativa da população, poderemos utilizar os métodos estatísticos para avaliar o erro amostral.*

Capítulo 2

Estatística Descritiva

Após coletado os dados, a próxima etapa no processo de análise estatística consiste em descrever os dados coletados, isto é, resumir e descrever suas características mais importantes. Esta etapa do processo é denominada de estatística descritiva. Três métodos básicos da estatística descritiva são: construção de tabelas de frequência, construção de gráficos e cálculo de medidas resumo. Quando se descreve um conjunto de dados, que pode ser composto de uma única variável, caso univariado, ou por um conjunto de variáveis, caso multivariado, algumas características devem ser observadas:

- 1. Centro:** O centro de um conjunto de dados é um valor que seja representativo do todo, isto é, uma medida que possa representar o conjunto de dados;
- 2. Dispersão ou variação:** Uma medida de dispersão ou variação é uma medida que resume a variabilidade presente num conjunto de dados;
- 3. Valores discrepantes ou outliers:** Elementos da amostra que se encontram muito distantes da grande maioria dos dados;
- 4. Distribuição:** A distribuição de frequências dos dados fornece informação sobre a forma da distribuição de probabilidade dos dados, por exemplo, a distribuição pode ser simétrica, assimétrica, pode ter a forma de um sino ou pode ser achatada.

2.1 Tabela de distribuição de frequências

É uma tabela em que se colocam as frequências observadas de cada categoria ou classe. Um dos objetivos de se construir uma tabela de distribuição de frequências é obter informações sobre a forma da distribuição de probabilidade dos dados. Esta informação ajudará na escolha de um modelo probabilístico. Mais adiante, iremos estudar alguns modelos probabilísticos comuns na prática.

Existem dois tipos de tabela de distribuição de frequências.

1. tabela de distribuição de frequências por valores: esta tabela é adequada para variáveis qualitativas, ou quantitativas dicretas que não possuam muitos valores diferentes;
2. tabela de distribuição de frequências por classes ou intervalos: esta tabela é adequada para variáveis quantitativas contínuas, ou dicretas que possuam muitos valores diferentes;

2.2 Distribuição de frequências por valores

É construída considerando-se todos os diferentes valores ou categorias, levando-se em consideração suas respectivas repetições (frequências).

Tabela 2.1: *Procedência dos alunos*

Procedência	Nº de alunos	%
Capital	10	33,33
Interior	12	40,00
O. região	8	26,67
Total	30	100

Fonte: Pesquisa em classe

Tabela 2.2: *Número de disciplinas matriculadas*

Número de disciplinas	Nº de alunos	%
3	3	10,00
4	5	16,67
5	8	26,67
6	10	33,33
7	4	13,33
Total	30	100

Fonte: Pesquisa em classe

2.3 Distribuição de Freqüências por classes ou intervalos

Utiliza-se a Distribuição de Freqüências por intervalos ou classes quando temos uma grande variabilidade no dados, isto é tem-se muitos valores diferentes.

2.3.1 Regras para elaboração da Tabela de Distribuição de Freqüências

1. Efetua-se um rol estatístico nos dados brutos, isto é, ordenar os dados em ordem crescente;

Tabela 2.3: *Idade dos alunos*

27	18	21	23	23
21	21	27	21	19
18	18	22	19	19
18	19	23	19	18
18	20	19	24	20
21	20	21	20	21

Tabela 2.4: *Rol Estatístico*

18	18	18	18	18
18	19	19	19	19
19	19	20	20	20
20	21	21	21	21
21	21	21	22	23
23	23	24	27	27

2. Determina-se a amplitude total(AT) dos dados: $AT = X_{max} - X_{min}$ em que X_{min} e X_{max} são os valores mínimo e máximo do conjunto de dados respectivamente. Para o nosso exemplo te-se que $AT = 27 - 18 = 9$;
3. Escolhe-se convenientemente o número de classes(K). Geralmente, entre 5 e 20 classes são satisfatórios. Um maneira prática de determinar o número de classes é utilizar $k \approx \sqrt{n}$.
No exemplo temos que: $\sqrt{30} = 5,47$, portanto $K = 5$.
4. Determinar a amplitude de classe: $h \approx \frac{AT}{K}$. Assim,

$$\frac{AT}{K} = \frac{9}{5} = 1,8 \Rightarrow h = 2$$

Obs.: Deve-se ter sempre $h \times K \geq AT$.

5. Efetua-se o agrupamento em classes e elabora-se a tabela de Distribuição de Freqüências.

Tabela 2.5: *Tabela de distribuição de freqüências*

Idades (anos)	Número de alunos	%
18-20	12	40
20-22	11	36,67
22-24	4	13,33
24-26	1	3,33
26-28	2	6,67
Total	30	100

Fonte: Pesquisa em classe

2.4 Elementos em uma tabela de Distribuição de Freqüências

Limites de classe: $L_{inf_i} \vdash L_{sup_i} = [L_{inf_i}, L_{sup_i})$;

Amplitude de classe: $h_i = L_{sup_i} - L_{inf_i}$;

Ponto médio da classe:

$$X_i = \frac{L_{sup_i} + L_{inf_i}}{2}.$$

Observação 2.1. No caso em que todas as classes possuem a mesma amplitude tem-se que $X_{i+1} = X_i + h$ ou $X_i = X_1 + (i - 1) \times h$, para $i \geq 2$;

Freqüência simples ou absoluta: f_i : freq. simples da i -ésima classes, isto é, o número de elementos da classe. Portanto, $\sum_{i=1}^n f_i = n$;

Freqüência relativa: $f_{r_i} = \frac{f_i}{n}$;

Freqüência percentual: $f_i\% = f_{r_i} \times 100$;

Freqüência simples acumulada: $F_i = \sum_{j=1}^i f_j = f_1 + f_2 + \dots + f_i$;

Freqüência relativa acumulada: $F_{r_i} = \sum_{j=1}^i f_{r_j} = f_{r_1} + f_{r_2} + \dots + f_{r_i}$;

Freqüência percentual acumulada: $F_i\% = \sum_{j=1}^i f_j\% = f_1\% + f_2\% + \dots + f_i\%$;

Exemplo 2.1. Continuação do exemplo anterior.

Tabela 2.6: Procedência dos alunos

Procedência	f_i	f_{r_i}	$f_i\%$	F_i	F_{r_i}	$F_i\%$
Capital	10	0,3333	33,33	10	0,3333	33,33
Interior	12	0,4000	40,00	22	0,7333	73,33
Outras regiões	8	0,2667	26,67	30	1	100
Total	30	1	100	—	—	—

Fonte: Pesquisa em classe

Figura 2.1: Gráfico em setores

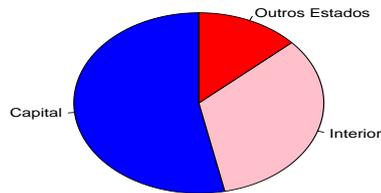
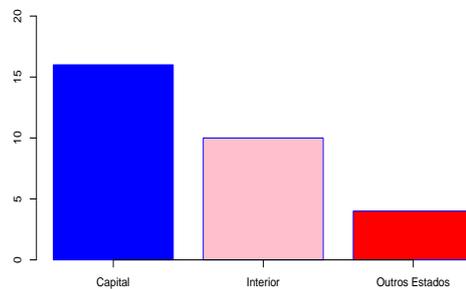


Figura 2.2: Gráfico de barras



Para a variável quantitativa tem-se que:

Tabela 2.7: Tabela de distribuição de frequências

Idade(Anos)	X_i	f_i	f_{r_i}	$f_i\%$	F_i	F_{r_i}	$F_i\%$
[18,20)	19	12	0,4000	40,00	12	0,4000	40,00
[20,22)	21	11	0,3667	36,67	23	0,7667	76,67
[22,24)	23	4	0,1333	13,33	27	0,9000	90,00
[24,26)	25	1	0,0333	3,33	28	0,9333	93,33
[26,28)	27	2	0,0667	6,67	30	1	100
Total	—	30	1	100	—	—	—

Fonte: Pesquisa em classe

Questões:

1. Qual a proporção de alunos com idade mínima de 22 anos ?

Resposta: $\frac{7}{30} \times 100 = 100 - 76,67 = 23,33\%$.

2. Qual a proporção de alunos com idade inferior a 24 anos mas que tenham no mínimo 20 anos?

Resposta: $\frac{15}{30} \times 100 = 90 - 40 = 50\%$.

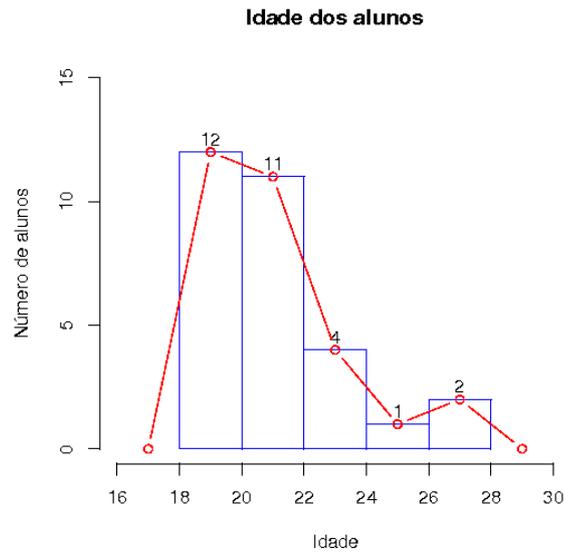


Figura 2.3: Histograma e Polígono de frequências

Exemplo 2.2. Uma amostra aleatória de 36 alunos da disciplina de Estatística Vital foi selecionada e o curso de cada aluno foi anotada.

Tabela 2.8: Curso

11	29	22	22	22	8
22	22	9	22	22	22
22	22	9	22	11	11
22	11	29	22	22	11
22	11	22	11	22	11
22	29	11	9	22	11

Tabela 2.9: Rol Estatístico

8	9	9	9	11	11
11	11	11	11	11	11
11	11	22	22	22	22
22	22	22	22	22	22
22	22	22	22	22	22
22	22	22	29	29	29

- *Elabore uma tabela de distribuição de frequências adequada para os dados;*
- *Faça um gráfico baseado nos valores da tabela acima.*

2.5 Medidas de Tendência Central

Dentre as medidas de tendência central, destacamos:

1. Média Aritmética ou Média;
2. Moda;
3. Mediana

Tabela 2.10: Amostra de 36 alunos da variável Curso

Curso	f_i	f_{r_i}	$f_i\%$	F_i	F_{r_i}	$F_i\%$
08	1	0,0278	2,78	1	0,0278	2,78
09	3	0,0833	8,33	4	0,1111	11,11
11	10	0,2778	27,78	14	0,3889	38,89
22	19	0,5278	52,78	33	0,9167	91,67
29	3	0,0833	8,33	36	1	100
Total	36	1	100	—	—	—

Fonte: Pesquisa em classe

2.5.1 Média Aritmética

A) Dada uma população $\underline{x} = (x_1, \dots, x_N)$, então a média, chamada de média populacional e denotada por μ , é dada por,

$$\mu = \frac{x_1 + \dots + x_N}{N};$$

B) Dada uma amostra $\underline{X} = (X_1, \dots, X_n)$, então a média, chamada de média amostral e denotada por \bar{X} , é dada por,

$$\bar{X} = \frac{X_1 + \dots + X_n}{n};$$

Exemplo 2.3. Para a amostra de idades de 30 alunos do exemplo calcular a idade média dos alunos.

$$\bar{X} = \frac{27 + 18 + 21 + 23 + 23 + \dots + 20 + 21}{30} = 20,6;$$

C) Para uma tabela de distribuição de freqüências, a média é dada por,

$$\bar{X} = \frac{X_1 \times f_1 + \dots + X_n \times f_n}{n}.$$

em que, f_i é a freqüência simples da i -ésima classe e X_i é o ponto médio da i -ésima classe.

Exemplo 2.4. Para a tabela de distribuição de freqüências da amostra de 30 alunos calcular a idade média dos alunos.

$$\bar{X} = \frac{19 \times 12 + 21 \times 11 + 23 \times 4 + 25 \times 1 + 27 \times 2}{30} = 21;$$

DESvantagens da Média

- É uma medida de tendência central que por uniformizar os valores de um conjunto de dados, não representa bem os conjuntos que revelam tendências extremas. Ou seja, é grandemente influenciada pelos valores extremos (grandes) do conjunto;

Exemplo 2.5. Considere os salários de 6 empregados de uma determinada empresa:

R\$1180	R\$1230	R\$1250	R\$1240	R\$1220	R\$2940
---------	---------	---------	---------	---------	---------

Portanto o salário médio pago pela empresa é:

$$\bar{X} = \frac{1180 + 1230 + 1250 + 1240 + 1220 + 2940}{6} = 1510$$

- Não pode ser calculada para distribuições de frequências com limites indeterminados (indefinidos);

VANTAGENS

- É a medida de tendência central mais conhecido e de maior emprego;
- É facilmente calculável;
- Tem propriedades interessantes;
- Depende de todos os valores do conjunto de dados.

Propriedades

1. A soma dos desvios em relação a média é zero, isto é,

$$(X_1 - \bar{X}) + (X_2 - \bar{X}) + \dots + (X_n - \bar{X}) = 0$$

Para as propriedades seguintes considere as duas amostras (X_1, \dots, X_n) e (Y_1, \dots, Y_n) das variáveis X e Y respectivamente e $c > 0$ uma constante arbitrária.

2. Se $Y = X + c$ então $\bar{Y} = \bar{X} + c$. Se $Y = X - c$ então $\bar{Y} = \bar{X} - c$;
3. $Y = c \times X$ então $\bar{Y} = c \times \bar{X}$. $Y = \frac{X}{c}$ então $\bar{Y} = \frac{\bar{X}}{c}$.
4. Seja $Z = X + Y$ uma outra variável então $\bar{Z} = \bar{X} + \bar{Y}$.

2.5.2 Moda

- A) Dada uma amostra de n elementos (X_1, \dots, X_n) , então a moda, denotada por Mo , será o valor mais freqüente na amostra, isto é, de maior freqüência simples f .

Exemplo 2.6. Para a amostra de idades da Tabela 2.3 a moda é $Mo = 21$ pois é o valor com maior freqüência simples $f = 7$;

- B) Para uma tabela de distribuição de frequências por classes ou intervalos, a moda será um valor dentro da classe modal (classe com maior freqüência simples f_i). Portanto, para dados agrupados, a moda será dada por:

$$Mo = L_{inf_{Mo}} + \frac{\Delta_1 \times h_{Mo}}{\Delta_1 + \Delta_2}$$

em que,

$L_{inf_{Mo}}$ = Limite inferior da classe modal;

h_{Mo} = Amplitude da classe modal;

$\Delta_1 = f_{Mo} - f_A$;

$\Delta_2 = f_{Mo} - f_P$;

f_A = frequência simples da classe anterior a classe modal;

f_P = frequência simples da classe posterior a classe modal;

Exemplo 2.7. Para a Tabela 2.5 de distribuição de frequências das idades dos alunos, a moda é dado por:

- A classe modal é 18+ 20;
- $L_{inf_{Mo}} = 18$; $h_{Mo} = 2$; $f_{Mo} = 12$ $f_A = 0$, $f_P = 11$
- $\Delta_1 = 12 - 0 = 12$; $\Delta_2 = 12 - 11 = 1$;
- Portanto,

$$Mo = 18 + \frac{12 \times 2}{12 + 1} = 18 + 1,8 = 19,8$$

DESVANTAGENS DA MODA

- Nem sempre é única e nem sempre existe;
- Seu valor não depende de todos os valores da amostra;

VANTAGENS DA MODA

- Não é influenciada por extremos;
- Pode ser calculada na maioria das vezes para distribuições de frequência com limites indeterminados.

2.5.3 Mediana

A) Dada uma amostra de n elementos (X_1, \dots, X_n) , então a mediana, denotada por Me , será o valor que divide a amostra em duas partes iguais.

Se o tamanho da amostra for ímpar então $Me = X_{(\frac{n+1}{2})}$. Ex: 2,9,5,8,3,13,11 então $Me = X_{(\frac{7+1}{2})} = X_{(4)} = 8$.

Se o tamanho da amostra for par então $Me = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$. Ex: 2,9,5,8,3,13,11,13 então $Me = \frac{X_{(4)} + X_{(5)}}{2} = \frac{8+9}{2} = 8,5$.

Exemplo 2.8. Para a amostra de idades da Tabela ?? a mediana é $Me = \frac{X_{(15)} + X_{(16)}}{2} = \frac{20+20}{2} = 20$;

B) Para uma tabela de distribuição de freqüências por classes ou intervalos, a mediana será um valor dentro da classe da mediana. A classe da mediana será a classe que F_i satisfizer: O menor F_i maior ou igual a $\frac{n}{2}$. Portanto, para os dados agrupados a mediana será dado por:

$$Me = L_{inf_{Me}} + \frac{\left(\frac{n}{2} - F_A\right) \times h_{Me}}{f_{Me}}$$

em que:

$L_{inf_{Me}}$ = Limite inferior da classe da mediana;

h_{Me} = Amplitude da classe da mediana;

f_{Me} = freqüência simples da classe da mediana;

F_A = freqüência simples acumulada da classe anterior a classe da mediana;

Exemplo 2.9. Para a Tabela 2.5 de distribuição de freqüências das idades dos alunos, a mediana é dado por:

- A classe modal é 20┆22;
- $L_{inf_{Me}} = 20$; $h_{Me} = 2$; $f_{Me} = 11$ $F_A = 12$
- $\Delta_1 = 12 - 0 = 12$; $\Delta_2 = 12 - 11 = 1$;
- Portanto,

$$Me = 20 + \frac{\left(\frac{30}{2} - 12\right) \times 2}{11} = 20 + \frac{6}{11} = 20,5$$

2.5.4 Assimetria

Para verificar se uma distribuição é simétrica ou assimétrica é usual utilizar a relação entre a média e a mediana, Assim:

- Se média \approx Mediana a distribuição é simétrica;
- Se média $>$ Mediana a distribuição é assimétrica à direita;
- Se média $<$ Mediana a distribuição é assimétrica à esquerda;

2.6 Medidas de Dispersão

Uma medida de dispersão que nos fornece informação sobre a variabilidade de um conjunto de dados, conseqüentemente da população. Diz-se que conjunto de dados é mais homogêneo que outro quando possui variabilidade menor que o do outro conjunto de dados.

1. **Amplitude Total:** $AT = X_{max} - X_{min}$

Exemplo 2.10. Seja $\underline{X} = (3, 5, 5, 7)$ e $\underline{Y} = (4, 5, 5, 6, 5)$ duas amostras aleatórias das variáveis X e Y respectivamente. Então,

$$\bar{X} = \frac{3+5+5+7}{4} = 5 \quad e \quad \bar{Y} = \frac{4+5+5+6+5}{5} = 5$$

$$AT_X = 7 - 3 = 4 \quad e \quad AT_Y = 6 - 4 = 2$$

2. Variância e Desvio padrão:

(A) Variância(σ^2) e Desvio padrão(σ) populacional: Para uma população de tamanho N , tem-se que

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad e \quad \sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

(B) Variância(S^2) e Desvio padrão(S) amostral: Para uma amostra de tamanho N , tem-se que

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad e \quad S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

Observação 2.2. Como S^2 é um estimador viciado para σ , isto é, o valor esperado de S^2 , $E(S^2)$, não é σ^2 , ou seja, $E(S^2) \neq \sigma^2$. Assim, na prática utilizamos,

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}.$$

que é não viciado, isto é, $E(S^2) = \sigma^2$. Apesar disso,

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

é um estimador viciado para σ .

Exemplo 2.11. Para a amostra do exemplo anterior temos que:

$$S_x^2 = \frac{(3-5)^2 + (5-5)^2 + (5-5)^2 + (7-5)^2}{4-1} = 2,67 \quad e \quad S_x = \sqrt{2,67} = 1,63$$

$$S_y^2 = \frac{(4-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2 + (5-5)^2}{5-1} = 0,5 \quad e \quad S_y = \sqrt{0,5} = 0,71$$

(C) Variância(S^2) e Desvio padrão(S) amostral para dados agrupados:

$$S^2 = \frac{\sum_{i=1}^n m_i \times (X_i - \bar{X})^2}{n - 1}$$

em que X_i é o ponto médio da classes ou intervalo, f_i é a frequência simples da classes ou intervalo.

Exemplo 2.12. Para os dados sobre a idade dos alunos temos que:

$$S^2 = \frac{12 \times (19 - 21)^2 + 11 \times (21 - 21)^2 + 4 \times (23 - 21)^2 + 1 \times (25 - 21)^2 + 2 \times (27 - 21)^2}{30 - 1}$$

$$= \frac{152}{29} = 5,24 \quad e \quad S = \sqrt{5,24} = 2,29$$

3. **Coefficiente de variação:** Utilizamos esta medida quando desejamos comparar conjuntos de dados diferentes em escala ou em unidade.

$$CV = \frac{S}{\bar{X}} \times 100$$

Exemplo 2.13. Para $\tilde{X} = (3, 5, 5, 7)$ e $\tilde{Y} = (4, 5, 5, 6, 5)$ tem-se que:

$$CV_X = \frac{S_X}{\bar{X}} \times 100 = \frac{1,63}{5} \times 100 = 32,6 \quad e \quad CV_Y = \frac{S_Y}{\bar{Y}} \times 100 = \frac{0,71}{5} \times 100 = 14,2$$

Exemplo 2.14. Os dados da tabela abaixo são de 11 pessoas do sexo masculino, aparentemente normais, com idades variando entre 14 e 24 anos: Verificar qual variável apresenta:

Nível de colesterol no sangue ^a	162	158	157	155	156	154	169	181	175	180	174
Peso(Kg)	51	53	56	57	58	60	58	61	59	56	61
Pressão sistólica sanguínea ^b	108	111	115	116	117	120	124	127	122	121	125

^aem mg/100cc

^bem mm de Hg

(a) *Maior variabilidade;*

(b) *Menor variabilidade.*

Solução: Denotando $X = \text{Nível de colesterol no sangue}$, $Y = \text{Peso}$ e $Z = \text{Pressão sistólica sanguínea}$, tem-se que:

$$\bar{X} = \frac{162 + 158 + 157 + 155 + 156 + 154 + 169 + 181 + 175 + 180 + 174}{11} = 165,5$$

$$\bar{Y} = \frac{51 + 53 + 56 + 57 + 58 + 60 + 58 + 61 + 59 + 56 + 61}{11} = 57,3$$

$$\bar{Z} = \frac{108 + 111 + 115 + 116 + 117 + 120 + 124 + 127 + 122 + 121 + 125}{11} = 118,7$$

$$S_X^2 = \frac{(162 - 165,5)^2 + (158 - 165,5)^2 + \dots + (174 - 165,5)^2}{10} = 109,9$$

$$S_Y^2 = \frac{(51 - 57,3)^2 + (53 - 57,3)^2 + \dots + (61 - 57,3)^2}{10} = 10$$

$$S_Z^2 = \frac{(108 - 118,7)^2 + (111 - 118,7)^2 + \dots + (125 - 118,7)^2}{10} = 35$$

$$CV_X = \frac{S_X}{\bar{X}} \times 100 = \frac{\sqrt{109,9}}{165,5} \times 100 = 6,33\%, \quad CV_Y = \frac{S_Y}{\bar{Y}} \times 100 = \frac{\sqrt{10}}{57,3} \times 100 = 5,52\%$$

$$CV_Z = \frac{S_Z}{\bar{Z}} \times 100 = \frac{\sqrt{35}}{118,7} \times 100 = 4,98\%$$

Portanto, o nível de colesterol no sangue é o que tem maior variabilidade e a pressão sistólica sanguínea é a que tem menor variabilidade.

Capítulo 3

Introdução a Probabilidade

Objetivo: O objetivo da teoria da Probabilidade é criar modelos teóricos que reproduzam de maneira razoável a distribuição de frequências de fenômenos(experimentos) aleatórios de interesse. Tais modelos são chamados modelos probabilísticos.

Definição 3.1 (Experimento aleatório). *Um experimento que pode fornecer diferentes resultados, muito embora seja repetido toda vez da mesma maneira, é chamado **experimento aleatório**.*

Características essencial de um experimento aleatório: Imprevisibilidade: o resultado do experimento não pode ser conhecido a priori;

Exemplos de experimentos aleatórios

- (E1) Lançar uma moeda uma vez. Anota-se o resultado;
- (E2) Lançar uma moeda duas vezes. Anota-se a seqüência obtida;
- (E3) Lançar uma moeda duas vezes. Anota-se o número de caras obtido;
- (E4) Numa linha de produção conta-se o número de peças defeituosas num dia de trabalho;
- (E5) Uma urna contém duas bolas brancas e três bolas vermelhas. Retira-se uma bola ao acaso da urna. Se for branca, lança-se uma moeda; se for vermelha, ela é devolvida à urna e retira-se outra bola. Anota-se o resultado obtido.

Definição 3.2 (Espaço amostral). *É o conjunto de todos os resultados de um experimento aleatório.*

Notação: Ω

Cada resultado possível é denominado ponto ou elemento de Ω e denotado genericamente por ω . Assim, escrevemos $\omega \in \Omega$ para indicar que o elemento ω está em Ω .

Exemplos de espaço amostral:

- (E1) $\Omega = \{c, r\}$, em que c=cara e r=coroa;
- (E2) $\Omega = \{(c, c), (c, r), (r, c), (r, r)\}$;
- (E3) $\Omega = \{0, 1, 2\}$;

(E4) $\Omega = \{0, 1, 2, \dots\}$;

(E5) $\Omega = \{(B, c), (B, r), (V, B), (V, V)\}$, em que B=bola branca, V=bola vermelha;

Definição 3.3. *Sejam A e B dois conjuntos. Então diz-se que A é um subconjunto de B se, e somente se $\omega \in A$ implicar $\omega \in B$. Notação: $A \subset B$.*

Observação 3.1. *Da Definição 3.3 segue que $A \subset A$, pois $\omega \in A$ implicar $\omega \in A$.*

Observação 3.2. *Se A não é um subconjunto de B, então existe pelo menos um $\omega \in A$ tal que $\omega \notin B$. Notação: $A \not\subset B$.*

Definição 3.4 (Igualdade de conjuntos). *Sejam A e B dois conjuntos. Então diz-se que $A = B$ se, e somente se, $A \subset B$ e $B \subset A$, isto é, $\omega \in A$ implicar $\omega \in B$ e $\omega \in B$ implicar $\omega \in A$.*

Observação 3.3. *Se A não é igual a B, então existe pelo menos um $\omega \in A$ tal que $\omega \notin B$ ou um $\omega \in B$ tal que $\omega \notin A$.*

Definição 3.5 (Evento). *É um subconjunto do espaço amostral Ω .*

Os subconjuntos de Ω serão denotados por letras latinas maiúsculas (A,B,C,...). Se A é um subconjunto de Ω então denotamos $A \subset \Omega$.

Exemplo 3.1. *Considere o experimento aleatório (E2). Seja A=Obtenção de faces iguais. Portanto, $A = \{(c, c), (r, r)\}$;*

Observação 3.4. *Diz-se que "ocorre o evento A" quando o resultado do experimento aleatório for um elemento de A.*

Observação 3.5. *O espaço amostral Ω e o conjunto vazio \emptyset também são eventos, em que Ω é o evento certo e \emptyset é o evento impossível.*

Operações básicas entre conjuntos

Sejam $A \subset \Omega$ e $B \subset \Omega$, então:

- **Complementar:** $A^c = \{\omega \in \Omega : \omega \notin A\}$;
- **Interseção:** $A \cap B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}$;
- **União:** $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\} = \{\omega \in \Omega : \omega \text{ a pelo menos um dos eventos}\}$;
- **Diferença:** $A - B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \notin B\}$, deste modo segue que $A - B = A \cap B^c$;
- **Diferença simétrica:** $A \Delta B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \notin B \text{ ou } \omega \notin A \text{ e } \omega \in B\}$, deste modo segue que $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$ ou também $A \Delta B = (A \cup B) - (A \cap B)$.

Definição 3.6 (Eventos disjuntos). *Dois eventos são disjuntos se e somente se $A \cap B = \emptyset$.*

Observação 3.6. *Da Definição 3.6 segue que o conjunto vazio é disjunto de qualquer outro evento, pois para todo evento A tem-se que $A \cap \emptyset = \emptyset$.*

Definição 3.7 (Partição de um evento). *Seja A um subconjunto de Ω . Então A_1, \dots, A_n formam uma partição de A se e somente se $A_i \cap A_j = \emptyset$ para todo $i \neq j$ e $\cup_{i=1}^n A_i = A$.*

Deste modo, se $A = \Omega$ então A_1, \dots, A_n formam uma partição de Ω se e somente se $A_i \cap A_j = \emptyset$ para todo $i \neq j$ e $\cup_{i=1}^n A_i = \Omega$.

Proposição 3.1 (Leis de De Morgan). *Sejam A_1, \dots, A_n tal que $A_i \subset \Omega$ para todo i , então:*

(i) $(\cup_{i=1}^n A_i)^c = \cap_{i=1}^n A_i^c$ **Interpretação:** *o complementar da ocorrência de pelo menos um dos eventos é a não ocorrência de todos os eventos;*

(ii) $(\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c$. **Interpretação:** *o complementar da ocorrência de todos os eventos é a não ocorrência de pelo menos um dos eventos.*

Definição 3.8 (σ -Álgebra). *Uma classe \mathcal{F} de subconjuntos de Ω é denominada uma σ -álgebra se ela satisfaz:*

(F1) $\Omega \in \mathcal{F}$;

(F2) Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$;

(F3) Se $A_i \in \mathcal{F}$ para todo $i \geq 1$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;

Exemplo 3.2. *Exemplos de σ -álgebras:*

1. $\mathcal{F} = \{\emptyset, \Omega\}$, esta é a σ -álgebra trivial;

2. $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$, para $\Omega = A \cup A^c$;

3. Considere o experimento (E3), assim $\Omega = \{0, 1, 2\}$. Portanto, $\mathcal{F} = \{\emptyset, \Omega, 0, 1, 2, \{0, 1\}, \{0, 2\}, \{1, 2\}\}$ é uma σ -álgebra de subconjuntos de Ω . Neste caso, \mathcal{P} é chamado de σ -álgebra das partes de Ω e é denotado por \mathcal{P} .

Definição 3.9. *Seja Ω finito enumerável, um espaço de eventos equiprováveis. Assim, para todo $A \in \mathcal{F}$ segue que a probabilidade de A é dada por,*

$$P(A) = \frac{\#A}{\#\Omega}$$

em que $\#$ é o número de elementos do conjunto. Esta definição é também conhecida como regra de Laplace.

A Definição 3.9 é a definição clássica de probabilidade.

Exemplo 3.3. *Considere o experimento aleatório (E2). Seja $A = \text{Obtenção de faces iguais}$. Portanto, $A = \{(c, c), (r, r)\}$. Deste modo,*

$$P(A) = \frac{2}{4} = 0,5.$$

Definição 3.10. Seja Ω um espaço amostral de um experimento aleatório. Seja n repetições independentes de um experimento aleatório e n_A o número de ocorrências do evento $A \in \mathcal{F}$ nas n repetições independentes do experimento. Então, a probabilidade de A é dada por,

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \rightarrow p_A$$

em que $0 \leq p_A \leq 1$. Esta convergência é garantida pelas Lei dos Grandes Numeros.

Definição 3.11. Seja (Ω, \mathcal{F}) um espaço mensurável. Então uma função $P: \mathcal{F} \rightarrow [0, 1]$ é uma probabilidade se,

(P1) $P(\Omega) = 1$;

(P2) Para todo $A \in \mathcal{F}$ tem-se $P(A) \geq 0$;

(P3) P é σ -aditiva, isto é, se A_1, A_2, \dots , são dois a dois disjuntos então,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Esta é a definição axiomática devida a Kolmogorov. A trinca (Ω, \mathcal{F}, P) é chamada de espaço de probabilidade.

Propriedades de uma medida de probabilidade

(C1) Para todo $A \in \mathcal{F}$ tem-se $P(A^c) = 1 - P(A)$. De fato, como $\Omega = A \cup A^c$ e $A \cap A^c = \emptyset$ segue que,

$$P(\Omega) = 1 = P(A \cup A^c) = P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A) \geq 0;$$

(C2) $P(\emptyset) = 0$, pois $\Omega = \emptyset^c$ logo por (C1)

$$P(\emptyset) = 1 - P(\Omega) = 1 - 1 = 0;$$

(C3) P é uma função não decrescente, isto é, para todo $A, B \in \mathcal{F}$ tal que $A \subseteq B$ tem-se que $P(A) \leq P(B)$. Para ver isso, basta notar que $B = A \cup (B - A)$ e $A \cap (B - A) = \emptyset$, portanto,

$$P(B) = P(A \cup (B - A)) = P(A) + P(B - A)$$

pela condição (P2) segue que $P(B - A) \geq 0$ portanto $P(B) \geq P(A)$;

(C4) Para todo $A, B \in \mathcal{F}$ tal que $A \subseteq B$ tem-se que $P(B - A) = P(B) - P(A)$; Este resultado segue diretamente do anterior;

(C5) Para todo $A, B \in \mathcal{F}$ arbitrários tem-se que:

$$P(A - B) = P(A) - P(A \cap B) \text{ e } P(B - A) = P(B) - P(A \cap B).$$

De fato,

$$\begin{aligned} P(A - B) &= P(A \cap B^c) = P(A \cap (\Omega - B)) = P(A \cap \Omega - A \cap B) \\ &= P(A - A \cap B) = P(A) - P(A \cap B), \text{ por (C4), pois } A \cap B \subseteq A \end{aligned}$$

e

$$\begin{aligned} P(B - A) &= P(A^c \cap B) = P((\Omega - A) \cap B) = P(\Omega \cap B - A \cap B) \\ &= P(B - A \cap B) = P(B) - P(A \cap B), \text{ por (C4), pois } A \cap B \subseteq B \end{aligned}$$

(C6) Para todo $A \in \mathcal{F}$ tem-se que $0 \leq P(A) \leq 1$. Este resultado segue de (P1), (P2) e (C3) e do fato que $A \subseteq \Omega$;

3.1 Probabilidade Condicional

Seja (Ω, \mathcal{F}, P) o espaço de probabilidade para um determinado experimento aleatório. Suponha que tenhamos a priori alguma informação a respeito do resultado do experimento aleatório. Por exemplo, suponha que saibamos que um determinado evento $B \in \mathcal{F}$ ocorreu. Isto elimina qualquer incerteza que tínhamos a respeito da ocorrência ou não do evento B. Além do mais, esta nova informação a respeito do experimento aleatório pode mudar as incertezas a respeito de outros eventos em \mathcal{F} e portanto uma nova medida de probabilidade deve ser considerada. Esta nova medida de probabilidade é também uma medida no espaço mensurável (Ω, \mathcal{F}) , será chamada de Probabilidade condicional.

Exemplo 3.4. Seja $\Omega = \{1, 2, 3, 4, 5, 6\}$ um espaço de eventos equiprováveis. Seja $\mathcal{F} = \mathcal{P}$ a σ -álgebra das partes de Ω e P a medida de probabilidade definida em (Ω, \mathcal{F}) assim,

$$P(A) = \frac{\#A}{\#\Omega}$$

para todo $A \in \mathcal{F}$. Considere os seguintes eventos,

$$A = \{1, 2, 6\} \text{ e } B = \{2, 3, 5\}.$$

Deste modo, tem-se que

$$P(A) = \frac{3}{6} = \frac{1}{2} \text{ e } P(B) = \frac{3}{6} = \frac{1}{2}.$$

Suponha agora que tenhamos a informação que o evento B ocorreu. Essa informação poderá alterar a probabilidade atribuída aos eventos em \mathcal{F} . A nova medida de probabilidade será denotada por $P(\cdot|B)$. Observe que podemos considerar que temos um novo espaço amostral $\Omega_B = B$ e uma nova σ -álgebra

$$\mathcal{F}_B = \{C \subset B : C = A \cap B, \text{ para algum } A \in \mathcal{F}\}.$$

Desta maneira, tem-se que $\mathcal{F}_B \subset \mathcal{F}$, por este motivo \mathcal{F}_B é denominada uma restrição de \mathcal{F} ao evento B . Assim, o novo espaço de probabilidade seria $(B, \mathcal{F}_B, P(\cdot|B))$. Para o exemplo acima, dado que o evento B ocorreu, então o evento A só irá ocorrer se o evento $C = \{1\} = A \cap B$ ocorrer, assim

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)} = \frac{1}{3}.$$

Entretanto, não é necessária a construção deste novo espaço de probabilidade, pois pode-se considerar apenas uma nova medida de probabilidade para o mesmo espaço mensurável (Ω, \mathcal{F}) . Para fazer isso, basta que a nova medida de probabilidade $P(\cdot|B)$ seja válida para todo $A \in \mathcal{F}$ e não apenas para $A \in \mathcal{F}_B$. Deste modo, para um dado evento $B \in \mathcal{F}$ tem-se

$$\begin{aligned} P(A) &= \frac{P(A)}{P(\Omega)} = \frac{P(A \cap (B \cup B^c))}{P(B \cup B^c)} \\ &= \frac{P(A \cap B) + P(A \cap B^c)}{P(B) + P(B^c)} \end{aligned}$$

Nestas condições segue que, dado que o evento B ocorreu, tem-se que

$$P(B^c) = 0 \text{ e } P(A \cap B^c) = 0$$

logo pode-se definir $P(\cdot|B)$ para todo $A \in \mathcal{F}$, com segue,

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

Para o exemplo assim tem-se que,

$$P(A|B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Definição 3.12 (Probabilidade Condicional). *Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Seja $B \in \mathcal{F}$ um evento tal que $P(B) > 0$. Então a probabilidade condicional, dado o evento B , é uma função denotada por $P(\cdot|B)$ e definida para todo $A \in \mathcal{F}$ como segue,*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (3.1)$$

em que $P(A|B)$ é chamada a probabilidade condicional de A dado B .

Teorema 3.1 (Regra do Produto). *Seja $\{A_i \in \mathcal{F}, i = 1, \dots, n\}$ eventos tais que,*

$$P\left(\bigcap_{i=1}^{n-1} A_n\right) > 0$$

então,

$$P\left(\bigcap_{i=1}^n A_n\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Demonstração. Fazer por indução. Para $n = 2$ tem-se que $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$ pela Definição 3.12 de probabilidade condicional. Supor que é valido para $n = k$, isto é,

$$P(A_1 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1) \times \cdots \times P\left(A_k \mid \bigcap_{i=1}^{k-1} A_i\right)$$

e então mostrar que vale para $n = k + 1$. Assim, seja $B_k = \bigcap_{i=1}^k A_i$, logo

$$P(B_k \cap A_{k+1}) = P(B_k)P(A_{k+1}|B_k).$$

□

Exemplo 3.5. *Uma urna contém 2 bolas brancas, 3 pretas e 4 verdes. Duas bolas são retiradas ao acaso sem reposição. Qual a probabilidade de que:*

(a) *Ambas sejam verdes?*

(b) *Ambas sejam da mesma cor?*

Solução: Tem-se que $\Omega = \{(b, b), (b, p), (b, v), (p, b), (p, p), (p, v), (v, b), (v, p), (v, v)\}$, em que b =bola branca; p =bola preta e v =bola verde, assim:

(a) sejam os eventos A =retirar verde no 1o. sorteio e B =retirar verde no 2o. sorteio, logo,

$$A = \{(v, b), (v, p), (v, v)\}; B = \{(b, v), (p, v), (v, v)\} \text{ e } A \cap B = \{(v, v)\}.$$

Assim,

$$P(A \cap B) = P(\{(v, v)\}) = P(A)P(B|A) = \frac{4}{9} \times \frac{3}{8} = \frac{1}{6}.$$

(b) Seja C =Ambas sejam da mesma cor, logo, $C = \{(b, b), (p, p), (v, v)\}$. Portanto, de modo semelhante ao item (a) tem-se que:

$$\begin{aligned} P(C) &= P(\{(b, b)\}) + P(\{(p, p)\}) + P(\{(v, v)\}) \\ &= \frac{2}{9} \times \frac{1}{8} + \frac{3}{9} \times \frac{2}{8} + \frac{4}{9} \times \frac{3}{8} \\ &= \frac{20}{72} = \frac{5}{18} \end{aligned}$$

Teorema 3.2 (Probabilidade Total). Seja $\{A_i \in \mathcal{F}, i = 1, \dots, n\}$ uma partição de Ω com $P(A_i) > 0$ para todo $i = 1, \dots, n$. Então, para todo $B \in \mathcal{F}$ tem-se que,

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

Demonstração. De fato, pois

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^n A_i\right),$$

assim,

$$\begin{aligned} P(B) &= P\left(B \cap \left(\bigcup_{i=1}^n A_i\right)\right) \\ &= P\left(\bigcup_{i=1}^n (A_i \cap B)\right) \\ &= \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i)P(B|A_i). \end{aligned}$$

□

Exemplo 3.6. Seja U_1 e U_2 duas urnas. A urna U_1 contém 3 bolas pretas e 2 vermelhas e a urna U_2 contém 4 bolas pretas e 2 vermelhas. Escolhe-se ao acaso uma urna e dela retira-se ao acaso uma bola. Qual a probabilidade de que a bola seja preta?

Solução: Seja A =bola preta. Note que U_1 e U_2 formam uma partição de Ω , assim

$$\begin{aligned} P(A) &= P(U_1)P(A|U_1) + P(U_2)P(A|U_2) = \frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{4}{6} \\ &= \frac{3}{10} + \frac{1}{3} = \frac{19}{30} \end{aligned}$$

Teorema 3.3 (Fórmula de Bayes). *Seja $\{A_i, i = 1, \dots, n\}$ uma partição de Ω com $P(A_i) > 0$ para todo $i = 1, \dots, n$. Então, para todo $B \in \mathcal{F}$ para o qual $P(B) > 0$ tem-se que,*

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

Exemplo 3.7. *Do exemplo anterior, calcule a probabilidade de que dado uma bola preta tenha sido sorteada, ela seja da urna U_1 ?*

Solução: *Primeiro note que U_1 e U_2 são uma partição de Ω . Assim,*

$$\begin{aligned} P(U_1|P) &= \frac{P(U_1)P(P|U_1)}{P(U_1)P(P|U_1) + P(U_2)P(P|U_2)} \\ &= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{4}{6}} \\ &= \frac{\frac{3}{10}}{\frac{19}{30}} = \frac{9}{19} \end{aligned}$$

Definição 3.13. *Sejam $A, B \in \mathcal{F}$ e (Ω, \mathcal{F}, P) um espaço de probabilidade. Então A e B são independentes se*

$$P(A|B) = P(A) \text{ com } P(B) > 0 \text{ ou se } P(B|A) = P(B) \text{ com } P(A) > 0.$$

Isto é, o fato de um dos eventos ocorrer não altera a probabilidade do outro ocorrer, assim da Definição 3.12 de probabilidade condicional segue que,

$$P(A \cap B) = P(A)P(A|B) = P(A|B)P(B) = P(A)P(B).$$

Observação 3.7. *Note que os conceitos de eventos disjuntos e eventos independentes não iguais. De um modo geral, dois eventos disjuntos não são independentes e vice-versa. A única possibilidade para que dois eventos disjuntos sejam independentes é se um deles tiver probabilidade zero. Por outro lado, dois eventos independentes só serão disjuntos se um dos eventos for o conjunto vazio.*

Exemplo 3.8. *Sejam dois eventos A e B tal que $P(A) = \frac{1}{2}$, $P(A|B) = \frac{1}{2}$ e $P(B|A) = \frac{1}{4}$. Esses eventos são independentes?*

Solução: *Os eventos A e B são independentes, pois,*

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8} \\ P(A \cap B) &= P(B)P(A|B) \Rightarrow P(B) = \frac{P(A \cap B)}{P(A|B)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4} \\ P(A)P(B) &= \frac{1}{2} \times \frac{1}{4} = \frac{1}{8} = P(A \cap B) \end{aligned}$$

Definição 3.14. *Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e A_1, \dots, A_n eventos em \mathcal{F} . Então os n eventos são independentes se,*

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \times P(A_n)$$

Entretanto, os eventos são completamente independentes se,

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k})$$

para $k = 2, \dots, n$ e $i_1, \dots, i_k = 1, \dots, n$ tal que $1 \leq i_1 < \dots < i_k \leq n$.

Capítulo 4

Variáveis aleatórias

Neste capítulo serão estudados o conceito de variável aleatória, sua classificação: discreta e contínua; os tipos de distribuição de probabilidade: função de probabilidade, função de distribuição e densidade de probabilidade; e os conceitos de esperança e variância para cada tipo de variável aleatória apresentada.

4.1 Conceitos e definições

Definição 4.1 (Variável aleatória). *Seja \mathcal{E} um experimento aleatório e (Ω, \mathcal{F}, P) o espaço de probabilidade associado a \mathcal{E} . Então uma função $X : \Omega \rightarrow \mathbb{R}$, que associa a cada elemento de $\omega \in \Omega$ um número real é uma variável aleatória se,*

$$X^{-1}(B) = \{ \omega \in \Omega : X(\omega) \in B \} \in \mathcal{F}.$$

Observação 4.1. *A função X deve ser unívoca, isto é, para cada $\omega \in \Omega$ deve haver apenas um $X(\omega)$ associado. Entretanto, diferentes valores de ω podem levar a um mesmo valor de X .*

Exemplo 4.1. *Considere o seguinte experimento: selecionar uma peça em uma linha de produção e observar se a peça é boa ou ruim. Nestas condições, segue que $\Omega = \{b, r\}$ em que b =boa e r =ruim. Consideremos a seguinte variável aleatória,*

$$X(\omega) = \begin{cases} 0 & \text{se } \omega = b, \\ 1 & \text{se } \omega = r, \end{cases}$$

Assim, considerando a σ -álgebra das partes de Ω , isto é, $\mathcal{F} = \{ \emptyset, \Omega, b, r \}$ tem-se que para todo intervalo $I \subset \mathbb{R}$ tal que:

- $\{0, 1\} \notin I$, por exemplo $I = (-5, 0)$, assim $X^{-1}(I) = \emptyset \in \mathcal{F}$;
- $0 \in I$ e $1 \notin I$, por exemplo $I = [0, \frac{1}{2}]$, assim

$$X^{-1}(I) = X^{-1} \left((0, 0] \cup \left(0, \frac{1}{2} \right] \right) = X^{-1}(0) \cup X^{-1} \left(\left(0, \frac{1}{2} \right] \right) = \{b\} \cup \emptyset = \{b\} \in \mathcal{F};$$

- $0 \notin I$ e $1 \in I$, por exemplo $I = [1, 2]$, assim

$$X^{-1}(I) = X^{-1}((1, 1] \cup (1, 2]) = X^{-1}((1, 1]) \cup X^{-1}((1, 2]) = \{r\} \cup \emptyset = \{r\} \in \mathcal{F};$$

- $\{0, 1\} \in I$, por exemplo $I = [0, 1]$, assim

$$X^{-1}(I) = X^{-1}((0, 0] \cup (1, 1] \cup (0, 1)) = X^{-1}(0) \cup X^{-1}(1) \cup X^{-1}((0, 1)) = \{b, r\} \cup \emptyset = \Omega \in \mathcal{F};$$

Portanto X como definido é uma variável aleatória.

Definição 4.2. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e X uma variável aleatória. Então P_X será a medida de probabilidade induzida por X no espaço $(\mathbb{R}, \mathcal{F}_X)$, tal que para todo $A = X^{-1}(I) \in \mathcal{F}$ tem-se que

$$P_X(I) = P(X^{-1}(I)) = P(A).$$

Portanto $(\mathbb{R}, \mathcal{F}_X, P_X)$ será o espaço de probabilidade induzido pela variável aleatória X .

Observação 4.2. De um modo geral, sempre que estamos trabalhando com a medida de probabilidade induzida por uma variável aleatória X utilizamos a notação P_X , entretanto para que a notação não fique muito carregada e desde que não possa haver confusão não utilizaremos o subscrito.

Definição 4.3 (Função de Distribuição). Seja X uma variável aleatória então sua função de distribuição é definida como,

$$F(x) = P(X \leq x) = P(X(\omega) \in (-\infty, x]) = P\left(X^{-1}\left((-\infty, x]\right)\right),$$

para todo $x \in \mathbb{R}$. F é também conhecida como função de distribuição acumulada de X .

Propriedades:

(P1) $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$;

(P2) F é contínua à direita, isto é, para $x_n \downarrow x$ tem-se que $\lim_{x_n \rightarrow x} F(x_n) = F(x^+) = F(x)$;

(P3) F é não decrescente, isto é, $F(x) \leq F(y)$ para todo $x, y \in \mathbb{R}$ tal que $x \leq y$.

4.2 Classificação das variáveis aleatórias

As variáveis aleatórias podem ser discretas ou contínuas. A classe das variáveis aleatórias contínuas ainda pode ser subdividida em três: absolutamente contínua, singular e mista. Neste livro será abordada apenas a variável aleatória discreta e a absolutamente contínua.

Definição 4.4 (Variável aleatória discreta). Uma variável aleatória X é discreta se o número de valores que X possa assumir for enumerável.

Definição 4.5 (Função de Probabilidade). *A função de probabilidade de uma variável aleatória X discreta com $\Omega_X = \{x_1, \dots, x_n\}$, é uma função que atribui probabilidade a cada um dos possíveis valores x_i de X , isto é,*

$$P_X(x_i) = P(X = x_i) = P(\{\omega \in \Omega : X(\omega) = x_i\}) = p(x_i)$$

para todo $i \in \{1, 2, \dots\}$ e satisfaz as seguintes condições,

(i) para todo $i \in \{1, 2, \dots\}$ tem-se que $0 \leq p(x_i) \leq 1$;

(ii) $\sum_{i \in \{1, 2, \dots\}} p(x_i) = 1$.

Exemplo 4.2. *Seja \mathcal{E} =lançamento de duas moeda, seja X =número de caras ocorridos. Assim, $\Omega = \{(c, c), (c, r), (r, c), (r, r)\}$ $\Omega_X = \{0, 1, 2\}$ e*

$$\begin{aligned} P_X(0) &= P(X^{-1}(0)) = P(r, r) = \frac{1}{4}, \\ P_X(1) &= P(X^{-1}(1)) = P(\{(c, r), (r, c)\}) = P(c, r) + P(r, c) = \frac{2}{4} = \frac{1}{2}, \\ P_X(2) &= P(X^{-1}(2)) = P(c, c) = \frac{1}{4}. \end{aligned}$$

Deste modo, a função de probabilidade de X é dada por

X	$P_X(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

Definição 4.6. *Para uma variável aleatória discreta X com $\Omega_X = \{x_1, \dots, x_n\}$, a função de distribuição é dada por,*

$$F(x) = \sum_{x_i \in \Omega_X : x_i \leq x} P_X(x_i) = \sum_{x_i \in \Omega_X : x_i \leq x} P(X = x_i).$$

para todo $x \in \mathbb{R}$.

Exemplo 4.3. *Do exemplo anterior tem-se que,*

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{4} & \text{se } 0 \leq x < 1 \\ \frac{3}{4} & \text{se } 1 \leq x < 2 \\ 1 & \text{se } x \geq 2 \end{cases}$$

Da definição acima segue que

$$P(X(\omega) \in (a, b]) = P(a < X \leq b) = F(b) - F(a).$$

De fato,

$$\begin{aligned}
 F(b) - F(a) &= \sum_{x_i \leq b} P(X = x_i) - \sum_{x_i \leq a} P(X = x_i) \\
 &= \left[\sum_{x_i \leq a} P(X = x_i) + \sum_{a < x_i \leq b} P(X = x_i) \right] - \sum_{x_i \leq a} P(X = x_i) \\
 &= \sum_{a < x_i \leq b} P(X = x_i) = P(a < X \leq b).
 \end{aligned}$$

Definição 4.7 (Variável aleatória contínua). *Uma variável aleatória X é contínua se o número de valores que X possa assumir for não enumerável.*

Definição 4.8 (Variável aleatória absolutamente contínua). *Uma variável aleatória X é absolutamente contínua se existir uma função não negativa f tal que para todo $x \in \mathbb{R}$,*

$$F(x) = \int_{-\infty}^x f(t) dt.$$

em que F é a função de distribuição da variável aleatória X .

Observação 4.3. *Note que toda variável aleatória absolutamente contínua é uma variável aleatória contínua mas nem toda variável aleatória contínua é uma variável aleatória absolutamente contínua.*

A função f da Definição 4.8 é chamada de função densidade de probabilidade, e da Definição 4.8, segue que,

$$f(x) = \frac{dF(x)}{dx}$$

para todo $x \in \mathbb{R}$ aonde F for derivável.

Propriedades da função densidade:

(i) $f(x) \geq 0$ para todo $x \in \mathbb{R}$;

(ii) $\int_{-\infty}^{\infty} f(x) dx = 1$.

As propriedades (i) e (ii) são condições necessárias e suficientes para que a função f seja uma densidade de probabilidade.

A partir da definição de função de distribuição para uma variável aleatória contínua tem-se que,

$$\begin{aligned}
 P(X(\omega) \in (a, b]) &= P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) \\
 &= \int_a^b f(x) dx = F(b) - F(a).
 \end{aligned}$$

para todo $a, b \in \mathbb{R}$ tal que $a < b$.

Exemplo 4.4. *Seja X uma variável aleatória contínua. Seja f uma função como segue,*

$$f(x) = \begin{cases} 2x & \text{se } 0 \leq x \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

(a) Mostre que f é uma função densidade de probabilidade;

(b) Calcule $P(X \leq 0,75)$?

(c) Calcule $P(X \leq 0,75|0,5 \leq X \leq 1)$?

Solução:

(a) f é uma fdp pois,

(i) $f(x) > 0$ para todo $x \in \mathbb{R}$;

(ii)

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} 2x dx = x^2 \Big|_0^1 = 1$$

Portanto,

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ x^2 & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1. \end{cases}$$

(b) Do ítem anterior tem-se que,

$$P(X \leq 0,75) = F(0,75) = \left(\frac{3}{4}\right)^2 = \frac{9}{16}.$$

(c) temos que,

$$\begin{aligned} P(X \leq 0,75|0,5 \leq X \leq 1) &= \frac{P((X \leq 0,75) \cap (0,5 \leq X \leq 1))}{P(0,5 \leq X \leq 1)} = \frac{P(0,5 \leq X \leq 0,75)}{P(0,5 \leq X \leq 1)} \\ &= \frac{F(0,75) - F(0,5)}{F(1) - F(0,5)} = \frac{\frac{9}{16} - \frac{1}{4}}{1 - \frac{1}{4}} = \frac{5}{12} \end{aligned}$$

4.3 Esperança de uma variável aleatória

A esperança de uma variável aleatória nada mais é que o valor médio esperado da variável. Por este motivo a esperança é usualmente denominado de valor esperado.

Definição 4.9 (Valor Esperado). Sejam (Ω, \mathcal{F}, P) um espaço de probabilidade, X variável aleatória neste espaço e $(\Omega_X, \mathcal{F}_X, P_X)$ o espaço de probabilidade induzido por X . Então o valor esperado de X , denotado por $E(X)$, é para o caso em que X é discreta,

$$E(X) = \sum_{i=1}^n x_i P(x_i) = \sum_{x \in \Omega_X} x P(x)$$

se $\sum_{i=1}^n |x_i| P_X(x_i) < \infty$, em que n é o número de elementos de Ω_X , e para o caso em que X é contínua

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

se $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

Exemplo 4.5. Seja X uma variável aleatória discreta com função de probabilidade,

X	P_X
0	0,5
1	0,25
2	0,25

então,

$$E(x) = 0 \times 0,5 + 1 \times 0,25 + 2 \times 0,25 = 0,75$$

Considere agora que X é contínua e tem densidade de probabilidade dada por

$$f(x) = \begin{cases} 2x & \text{se } 0 \leq x \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

então,

$$E(X) = \int_{-\infty}^0 x \times 0 dx + \int_0^1 x \times 2x dx + \int_1^{\infty} x \times 0 dx = \frac{2}{3}.$$

4.3.1 Propriedades da Esperança

1. Seja $c \in \mathbb{R}$ uma constante, então $E(c) = c$;
2. Seja h uma função real, então

$$E(h(X)) = \sum_{i=1}^n h(x_i)P(x_i) = \sum_{x \in \Omega_X} h(x)P(x)$$

para o caso em que X é discreta e

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

para o caso em que X é contínua.

3. Sejam X_1, \dots, X_n n variáveis aleatórias então

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

Exemplo 4.6 (lista 3, exerc. 6).

4.4 Variância de uma variável aleatória

Definição 4.10. *Seja X uma variável aleatória definida no espaço de probabilidade (Ω, \mathcal{F}, P) . Então a variância da variável aleatória X , denotado por $Var(X)$ é,*

$$Var(X) = E\left((X - E(X))^2\right) = E(X^2) - [E(X)]^2.$$

Se X é uma variável aleatória discreta, então

$$Var(X) = \sum_{i=1}^n (x_i - \mu_X)^2 P_X(x_i),$$

em que $\mu_X = E(X)$, ou

$$Var(X) = \sum_{i=1}^n x_i^2 P_X(x_i) - \left[\sum_{i=1}^n x_i P_X(x_i) \right]^2$$

Se X é uma variável aleatória contínua, então

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx$$

ou

$$Var(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - \left[\int_{-\infty}^{\infty} x f(x) dx \right]^2.$$

4.4.1 Propriedades da Variância

1. $Var(c) = 0$, pois $E((c - E(c))^2) = 0$;

2. $Var(X \pm c) = Var(X)$, pois

$$\begin{aligned} Var(X + c) &= E\left((X + c - E(X + c))^2\right) = E\left((X + c - E(X) - c)^2\right) \\ &= E\left((X - E(X))^2\right) = Var(X). \end{aligned}$$

3. $Var(cX) = c^2 Var(X)$, pois

$$\begin{aligned} Var(cX) &= E((cX)^2) - [E(cX)]^2 = E(c^2 X^2) - [cE(X)]^2 \\ &= c^2 E(X^2) - c^2 [E(X)]^2 = c^2 [E(X^2) - [E(X)]^2] = c^2 Var(X). \end{aligned}$$

Um resultado interessante que relaciona a probabilidade de um evento e a variância é dada pela desigualdade de Chebyshev.

Teorema 4.1 (Desigualdade de Chebyshev). *Seja X uma variável aleatória tal que a variância exista. Então para qualquer constante $c > 0$, tem-se que*

$$P\left(|X - E(X)| \geq c\right) \leq \frac{Var(X)}{c^2}.$$

Capítulo 5

Modelos Probabilísticos para variáveis aleatórias

Neste capítulo vamos apresentar alguns dos modelos mais usuais para variáveis aleatórias. O capítulo está dividido em duas partes: uma para variáveis aleatórias discretas e outra para variáveis aleatórias contínuas.

5.1 Modelos Probabilísticos para variáveis aleatórias discretas

Nesta seção serão apresentados três modelos para variáveis aleatórias discretas: Bernoulli, Binomial e Poisson. Existem ainda muitos outros modelos, entretanto, estes são os mais básicos a partir dos quais vários outros modelos podem ser derivados.

5.1.1 Distribuição de Bernoulli

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e X uma variável aleatória neste espaço tal que para um dado evento de interesse $A \in \mathcal{F}$ tem-se,

$$X(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \in A^c \end{cases}$$

Nestas condições segue que a função distribuição de X é dado por,

$$P_X(x) = \begin{cases} p^x \times (1-p)^{1-x} & \text{se } x = \{0, 1\} \\ 0 & \text{caso contrário} \end{cases}$$

Uma função de probabilidade assim definida é chamada de distribuição de Bernoulli. Agora observe que $p = P(A)$. De fato, para $x = 1$ tem-se,

$$P_X(1) = p = P(X^{-1}(1)) = P(A).$$

Os experimentos que originam uma variável aleatória com distribuição de Bernoulli são chamados de experimentos de Bernoulli.

Note o espaço amostral Ω pode ser enumerável ou não enumerável. Deste modo, o evento de interesse A pode conter:

- um único elemento, por exemplo, $\Omega = \{cara, coroa\}$, $A = \{coroa\}$ e $A^c = \{cara\}$;
- vários elementos, por exemplo, $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$ e $A^c = \{1, 3, 5\}$;
- um número não enumerável de elementos, por exemplo, $\Omega = \{x \in \mathbb{R} : 0 < x \leq 3\}$, $A = \{x \in \mathbb{R} : x > 1, 65\}$ e $A^c = \{x \in \mathbb{R} : x \leq 1, 65\}$;

Esperança e variância

Para $X \sim ber(p)$ tem-se que,

$$E(X) = \sum_{x=0}^1 xP(x) = 0 \times (1-p) + 1 \times p = p.$$

e

$$Var(X) = E(X^2) - [E(X)]^2 = \sum_{x=0}^1 x^2P(x) - p^2 = 0^2 \times (1-p) + 1^2 \times p - p^2 = p(1-p).$$

Exemplo 5.1. Considere \mathcal{E} = lançamento de um dado. Considere que o evento de interesse seja a ocorrência de um número par, portanto $A = \{2, 4, 6\}$ assim

$$X(\omega) = \begin{cases} 1 & \text{se } \omega \in \{2, 4, 6\} \\ 0 & \text{se } \omega \in \{1, 3, 5\} \end{cases}$$

e $p = P(A) = \frac{3}{6} = 0,5$. Portanto, a função de probabilidade de X é dada por,

$$P(x) = \begin{cases} 0,5^x \times 0,5^{1-x} & \text{se } x = \{0, 1\} \\ 0 & \text{caso contrário} \end{cases}$$

Notação: $X \sim ber(p)$

5.1.2 Distribuição Binomial

Considere agora que um experimento de Bernoulli é repetido n vezes de maneira independente, isto é, o espaço de probabilidade (Ω, \mathcal{F}, P) associado ao experimento na i -ésima repetição é o mesmo em todas as n repetições do experimento. Seja $(\Omega^n, \mathcal{F}_n, P_n)$ o espaço de probabilidade ao experimento composto das n repetições e X uma variável aleatória neste espaço tal que para um dado evento de interesse $A \in \mathcal{F}$, a variável aleatória X conta o número de vezes que o evento A ocorre em n repetições independentes de um experimento de Bernoulli. Nestas condições segue que a função distribuição de X é dado por,

$$P_X(x) = \begin{cases} \binom{n}{x} p^x \times (1-p)^{n-x} & \text{se } x = \{0, 1, 2, \dots, n\} \\ 0 & \text{caso contrário} \end{cases}$$

Em que $p = P(A)$. De fato, considere $x = n$, assim

$$P_X(n) = \binom{n}{n} p^n \times (1-p)^{n-n} = p^n = P(X^{-1}(n)) = P(\cap_{i=1}^{20} A) = \prod_{i=1}^{20} P(A) = P(A)^{20}$$

Portanto, $p = P(A)$. **Notação:** $X \sim bin(n, p)$.

Note que $\binom{n}{x}$ representa o número de arranjos possíveis para a ocorrência do evento A x vezes em um seqüência de tamanho n . Por exemplo, seja $n = 3$ e $x = 2$ então os arranjos possíveis são:

$$\{(A, A, A^c), (A, A^c, A), (A^c, A, A)\}$$

portanto 3 arranjos possíveis. Então a probabilidade do evento A ocorrer duas vezes em três repetições independentes do experimento é,

$$\begin{aligned} P(\{(A, A, A^c), (A, A^c, A), (A^c, A, A)\}) &= P(A, A, A^c) + P(A, A^c, A) + P(A^c, A, A) \\ &= P(A \cap A \cap A^c) + P(A \cap A^c \cap A) + P(A^c \cap A \cap A) \\ &= P(A)P(A)P(A^c) + P(A)P(A^c)P(A) + P(A^c)P(A)P(A) \\ &= 3 \times P(A)P(A)P(A^c) = 3 \times p^2(1 - p). \end{aligned}$$

Pela fórmula tem-se

$$P_X(x) = \binom{3}{2} \times p^2(1 - p)^{3-2} = \frac{3!}{2!1!} p^2(1 - p)^{3-2} = 3p^2(1 - p).$$

Exemplo 5.2. Considere \mathcal{E} = lançamento de uma moeda 10 vezes. Considere que o evento de interesse seja a ocorrência da face cara. Deste modo, qual a probabilidade de ocorrer 4 caras? Tem-se que $p = P(A) = P(c) = 0,5$. Da definição segue que,

$$P_X(4) = \binom{10}{4} 0,5^4 \times (1 - 0,5)^{10-4} = \frac{10!}{4!6!} 0,5^4 \times 0,5^6 = 0,2051.$$

Esperança e variância

Para $X \sim b(n, p)$ tem-se que,

$$\begin{aligned} E(X) &= \sum_{x=0}^n x P_X(x) = \sum_{x=1}^n x \binom{n}{x} p^x \times (1 - p)^{n-x} = \sum_{x=1}^n x \frac{n}{x} \binom{n-1}{x-1} p^x \times (1 - p)^{n-x} \\ &= \sum_{x=1}^n n \binom{n-1}{x-1} p \times p^{x-1} \times (1 - p)^{n-x} = np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} \times (1 - p)^{n-x-1+1} \\ &= np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} \times (1 - p)^{(n-1)-(x-1)} = np \times (p + 1 - p)^{n-1} = np \end{aligned}$$

e

$$\begin{aligned}
 E(X^2) &= \sum_{x=1}^n x^2 \binom{n}{x} p^x \times (1-p)^{n-x} = \sum_{x=1}^n x(x-1+1) \binom{n}{x} p^x \times (1-p)^{n-x} \\
 &= \sum_{x=1}^n x(x-1) \binom{n}{x} p^x \times (1-p)^{n-x} + \sum_{x=1}^n x \binom{n}{x} p^x \times (1-p)^{n-x} \\
 &= \sum_{x=1}^n x(x-1) \frac{n(n-1)}{x(x-1)} \binom{n-2}{x-2} p^x \times (1-p)^{n-x} + E(X) \\
 &= n(n-1) \sum_{x=1}^n \binom{n-2}{x-2} p^2 \times p^{x-2} \times (1-p)^{n-x-2+2} + np \\
 &= n(n-1)p^2 \sum_{x=2}^n \binom{n-2}{x-2} p^{x-2} \times (1-p)^{(n-2)-(x-2)} + np \\
 &= n(n-1)p^2 \times (p+1-p) + np.
 \end{aligned}$$

portanto, $Var(X) = n(n-1)p^2 \times (p+1-p) + np - [np]^2 = np - np^2 = np(1-p)$.

5.1.3 Distribuição de Poisson

Seja X uma variável aleatória que conta o número de ocorrência de um determinado evento A por unidade (tempo, comprimento, área, volume, etc), então a função de probabilidade de X é dada por,

$$P(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{se } x = \{0, 1, \dots, \} \\ 0 & \text{caso contrário} \end{cases}$$

Esta função é chamada de distribuição de Poisson. **Notação:** $X \sim \mathcal{P}(\lambda)$

Esperança e Variança:

$$\begin{aligned}
 E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda \lambda^{x-1}}{(x-1)!} \\
 &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \\
 &= \lambda e^{-\lambda} e^{\lambda} = \lambda
 \end{aligned}$$

e, agora note que $x^2 = x(x - 1) + x$, assim

$$\begin{aligned}
 E(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=0}^{\infty} [x(x-1) + x] \frac{e^{-\lambda} \lambda^x}{x!} \\
 &= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} + \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\
 &= \sum_{x=2}^{\infty} \frac{e^{-\lambda} \lambda^2 \lambda^{x-2}}{(x-2)!} + \lambda \\
 &= \lambda^2 e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} + \lambda = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda \\
 &= \lambda^2 + \lambda
 \end{aligned}$$

Portanto,

$$\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Exemplo 5.3. Num livro de 800 páginas há 800 erros de impressão. Qual a probabilidade de que uma página escolhida ao acaso contenha pelo menos 3 erros?

Solução: Seja X a variável aleatória que conta o número de erros por página, assim

$$\lambda = E(X) = \frac{800}{800} = 1$$

Portanto,

$$\begin{aligned}
 P_X(X \geq 3) &= 1 - P_X(X < 3) = 1 - [P_X(0) + P_X(1) + P_X(2)] \\
 &= 1 - \left[\frac{e^{-1} 1^0}{0!} + \frac{e^{-1} 1^1}{1!} + \frac{e^{-1} 1^2}{2!} \right] \\
 &= 1 - e^{-1} \left[1 + 1 + \frac{1}{2} \right] = 1 - 2,5e^{-1} \\
 &= 0,0803
 \end{aligned}$$

Exemplo 5.4. Numa central telefônica chegam 300 telefonemas por hora. Qual a probabilidade de que:

(a) Num minuto não haja nenhum chamado?

(b) Em 2 minutos haja 2 chamadas?

(c) Em t minutos, não haja chamadas?

Solução:

(a) Seja X a variável aleatória que conta o número de chamadas por minuto. Assim,

$$\lambda = E(X) = \frac{300}{60} = 5$$

Portanto,

$$P_X(0) = \frac{e^{-5}5^0}{0!} = e^{-5} = 0,0067$$

(b) Seja X_2 a variável aleatória que conta o número de chamadas por cada 2 minuto. Assim,

$$\lambda_2 = E(X_2) = \frac{300}{30} = 10$$

Portanto,

$$P_{X_2}(2) = \frac{e^{-10}10^2}{2!} = 50e^{-10} = 0,0023$$

(c) Seja X_t a variável aleatória que conta o número de chamadas por cada t minuto. Assim,

$$\lambda_t = E(X_t) = \frac{300}{\frac{60}{t}} = 5t$$

Portanto,

$$P_{X_t}(0) = \frac{e^{-5t}(5t)^0}{0!} = e^{-5t}.$$

Observação 5.1. Do exemplo anterior pode-se concluir que a probabilidade de ocorrência de um determinado evento A em t unidades é dada por,

$$P_{X_t}(x) = \begin{cases} \frac{e^{-\lambda t}(\lambda t)^x}{x!} & \text{se } x = \{0, 1, \dots, \} \\ 0 & \text{caso contrário} \end{cases}$$

5.2 Modelos Probabilísticos para variáveis aleatórias contínuas

Nesta seção serão apresentados os dois modelos contínuos que serão necessários para o desenvolvimento do restante deste livro, a saber: Distribuição normal e a distribuição t-Student.

5.2.1 Distribuição Normal

Dizemos que uma v.a. X tem distribuição normal com média μ e variância σ^2 se sua função densidade de probabilidade é dada por

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

para todo $x \in \mathbb{R}$, em que $E(X) = \mu$ e $Var(X) = \sigma^2$. **Notação:** $X \sim N(\mu, \sigma^2)$.

Principais características:

1. A moda, mediana e a média são iguais a μ ;
2. A função tem dois pontos de inflexão, um em $x = \mu - \sigma$ e outro em $x = \mu + \sigma$, em que σ é o desvio padrão de X ;
3. A curva é simétrica em torno de $x = \mu$, isto implica que dado um $a \in \mathbb{R}$ tem-se que $f(\mu - a) = f(\mu + a)$, logo $F(\mu - a) = P_X(X \leq \mu - a) = P_X(X \geq \mu + a) = 1 - F(\mu + a)$ se $\mu = 0$ então $F(-a) = 1 - F(a)$.

Problema: Dificuldade no cálculo de P_X . Existem tabelas apenas para $X \sim N(0, 1)$.

Solução: Fazendo a transformação,

$$Z = \frac{X - \mu}{\sigma} \Rightarrow Z^{-1}(z) = \sigma z + \mu$$

segue que,

$$P_Z(Z \leq z) = P_Z((-\infty, z]) = P_X(Z^{-1}((-\infty, z])) = P_X(X \leq Z^{-1}(z)) = P_X(X \leq \sigma z + \mu).$$

Portanto a variável aleatória Z também tem distribuição normal. Falta determinar com que parâmetros. Agora note que,

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = E\left(\frac{X}{\sigma}\right) - E\left(\frac{\mu}{\sigma}\right) = \frac{E(X)}{\sigma} - \frac{\mu}{\sigma} = 0$$

e

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \text{Var}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \text{Var}\left(\frac{X}{\sigma}\right) = \frac{\text{Var}(X)}{\sigma^2} = 1.$$

Portanto $Z \sim N(0, 1)$. Nestas condições segue que para $x = \sigma z + \mu$ tem-se que

$$z = \frac{x - \mu}{\sigma}$$

portanto,

$$P_X(X \leq x) = P_Z(Z \leq z) = P_Z\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

Deste modo pode-se utilizar as tabelas para distribuições normais com média zero e variância 1 para calcular a probabilidade de variáveis com distribuições normais com média μ e variância σ^2 .

Exemplo 5.5. Seja $X \sim N(100, 25)$, calcular:

(a) $P_X(100 \leq X \leq 106)$;

(b) $P_X(X \geq 108)$;

(b) $P_X(X \geq x) = 0,025$;

Solução:

(a) Tem-se que $\sigma = \sqrt{25} = 5$ e Portanto,

$$P_X(100 \leq X \leq 106) = P_Z\left(\frac{100 - 100}{5} \leq Z \leq \frac{106 - 100}{5}\right) = P_Z(0 \leq Z \leq 1,2) = F(1,2) - F(0).$$

Da tabela: $F(0) = 0,5$ e $F(1,2) = 0,8849$, assim $P_X(100 \leq X \leq 106) = 0,8849 - 0,5 = 0,3849$.

(b) $P_X(X \geq 108) = 1 - P_X(X \leq 108) = 1 - P_Z\left(Z \leq \frac{108 - 100}{5}\right) = 1 - P_Z(Z \leq 1,6) = 1 - F(1,6) = 1 - 0,9452 = 0,0548$.

(c) $P_X(X \geq x) = 1 - P_X(X \leq x) = 1 - P_Z\left(Z \leq \frac{x - 100}{5}\right) = 1 - P_Z(Z \leq z) = 0,025$ portanto $P_Z(Z \leq z) = 0,975$ da tabela tem-se que: $x = 1,96$.

5.2.2 Distribuição t-Student

Dizemos que uma v.a. X tem distribuição t com ν graus de liberdade se sua função densidade de probabilidade é dada por

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

para todo $x \in \mathbb{R}$. **Notação:** $X \sim t_\nu$. Tem-se ainda que $E(X) = 0$ para $\nu > 1$ e

$$Var(X) = \frac{\nu}{\nu - 2}$$

para $\nu > 2$.

Principais características:

1. A moda, mediana e a média são iguais a 0;
2. A curva é simétrica em torno do 0, isto implica que dado um $a \in \mathbb{R}$ tem-se que $f(-a) = f(+a)$, logo $P_X(\leq -a) = P_X(\geq a)$;
3. quando os graus de liberdade aumentam a distribuição t_ν se aproxima da distribuição normal com média zero e variância 1.

Exemplo 5.6. Seja $X \sim t_5$, calcular:

(a)

$$\begin{aligned} P_X(-2,57 \leq X \leq 2,57) &= P(X \leq 2,57) - P(X \leq -2,57) = [1 - P(X > 2,57)] - P(X > 2,57) \\ &= 1 - 2 \times P(X > 2,57) = 1 - 2 \times 0,025 = 0,95; \end{aligned}$$

(b) $P_X(X \geq x) = 0,01$ isto implica $x = 3,365$.

Capítulo 6

Distribuições Amostrais

A inferência estatística está interessada em tomar decisões sobre uma população, baseando-se apenas na informação contida em uma amostra aleatória da população de interesse. Por exemplo, o engenheiro de uma fábrica de refrigerantes pode estar interessado no volume médio de enchimento de uma lata de refrigerante que espera-se ser de 300 ml. Deste modo, para verificar se a máquina que faz o enchimento está regulada, o engenheiro coleta uma amostra aleatória de 25 latas e calcula o volume médio amostral obtendo $\bar{x} = 298$ ml. A próxima pergunta que o engenheiro desejará responder é qual a probabilidade do volume médio de enchimento de uma lata de refrigerante seja maior que 305 ml e menor que 295 ml dado que o valor observado da média amostral foi $\bar{x} = 298$ ml? Para responder a esta questão, em primeiro lugar, note que a média amostral $\bar{X} = \sum_{i \geq 1} X_i$ é uma função de variáveis aleatórias, portanto é também uma variável aleatória, logo \bar{X} possui uma distribuição de probabilidade associada.

Definição 6.1. *Uma amostra aleatória de tamanho n de uma variável aleatória X com função distribuição F , é um vetor $\underline{X} = (X_1, X_2, \dots, X_n)$ em que as componentes X_i são independentes e possuem distribuição F .*

Da Definição 6.1 pode-se concluir que dada uma amostra aleatória $\underline{X} = (X_1, X_2, \dots, X_n)$ de uma variável X com média μ e variância σ^2 então $E(X_i) = \mu$ e $Var(X_i) = \sigma^2$ para todo $i = \{1, 2, \dots, n\}$.

Definição 6.2. *A distribuição de probabilidade de um estimador é chamada de **distribuição amostral**.*

Por exemplo, a distribuição de probabilidade de \bar{X} é chamada de distribuição amostral da média. Portanto, dado que \bar{X} em uma distribuição de probabilidade pode-se calcular $P(295 < \bar{X} < 305)$ bastando para isso conhecer a distribuição de probabilidade de \bar{X} .

Observação 6.1. *A distribuição amostral de um estimador depende da distribuição de probabilidade da população da qual a amostra foi selecionada, do tamanho da amostra e do método de seleção da amostra.*

Um resultado importante muito utilizado em inferência é o Teorema Central do Limite, que fornece uma importante conclusão a respeito da distribuição da soma de variáveis aleatórias independentes.

Teorema 6.1. *Seja $\{X_n, n \geq 1\}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância $\sigma^2 < \infty$. Então, para $S_n = \sum_{i=1}^n X_n$, tem-se*

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1)$$

6.1 Distribuição Amostral da Média

Seja X uma variável aleatória com média μ e variância σ^2 . Então,

(i) Se $X \sim N(\mu, \sigma^2)$ tem-se que, $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$, em que

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n},$$

para X_1, \dots, X_n uma amostra aleatória da variável X . De fato, pode-se provar que a soma de variáveis aleatórias independentes, cada uma com distribuição normal com média μ e variância σ^2 também terá um distribuição normal, com média

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu,$$

e variância

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right)$$

Resultado:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n Cov(X_i, X_j)$$

em que

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

se X_i e X_j forem independentes então $E(X_i X_j) = E(X_i)E(X_j)$, logo $Cov(X_i, X_j) = 0$ e portanto para X_1, \dots, X_n independentes, segue que

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i).$$

Deste modo, segue que,

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Portanto,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Entretanto, se o valor da variância não for conhecido utilizaremos o estimador,

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1},$$

para σ^2 . Deste modo temos que,

$$T = \frac{\bar{X} - \mu}{S^2} \sim t_{n-1}$$

Se X não tiver distribuição normal então pelo Teorema Central do Limite segue que a distribuição da média amostral será aproximadamente normal com média μ e variância $\frac{\sigma^2}{n}$.

Observação 6.2. *A qualidade da aproximação normal para a distribuição amostral da média dependerá do tamanho da amostra e da distribuição da população de onde foi retirada a amostra. Em muito casos de interesse prático, se $n \geq 30$ a aproximação normal será satisfatória, independente da distribuição da população.*

Exemplo 6.1. *Uma fábrica produz resistores que têm uma resistência média de 100Ω com desvio padrão de 10Ω . Supondo que a distribuição das resistências seja normal, encontre a probabilidade de uma amostra aleatória de 25 resistores ter uma média menor que 95Ω .*

Solução:

$$P(\bar{X} < 95) = P\left(Z < \frac{95 - 100}{\frac{10}{\sqrt{25}}}\right) = P(Z < -2,5) = 0,0062$$

6.2 Distribuição Amostral da Proporção

Seja $X \sim ber(p)$. Retirada uma amostra aleatória (X_1, \dots, X_n) da variável X , tem-se que, $Y = X_1 + \dots + X_n \sim b(n, p)$, pois Y conta o número de vezes que um certo evento de interesse A aparece na amostra. Lembrando que $E(Y) = np$, isto é, $E(Y)$ é o número médio de vezes que o evento de interesse aparece em uma amostra de tamanho n . Assim, $p = \frac{E(Y)}{n}$, logo p é a proporção de vezes que o evento de interesse aparece em uma amostra de tamanho n . Portanto, dada amostra aleatória (X_1, \dots, X_n) , um estimador para o parâmetro p é dado por,

$$\hat{p} = \frac{X_1 + \dots + X_n}{n}.$$

Agora note que, para $0 \leq k \leq n$, tem-se,

$$P\left(\hat{p} = \frac{k}{n}\right) = P\left(\frac{X_1 + \dots + X_n}{n} = \frac{k}{n}\right) = P(Y = k)$$

Portanto, podemos obter a distribuição de probabilidade de \hat{p} a partir da distribuição de probabilidade de Y . Foi anteriormente visto que a distribuição da média amostral pode ser aproximada pela distribuição normal para n grande. Assim note que,

$$\hat{p} = \frac{X_1 + \dots + X_n}{n} = \bar{X}.$$

Logo, do Teorema Central do Limite, segue que \hat{p} terá distribuição aproximadamente normal com média,

$$E(\hat{p}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n} np = p$$

e variância,

$$Var(\hat{p}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + \dots + X_n) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}.$$

Portanto, $\hat{p} \stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$. Deste modo,

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1).$$

Exemplo 6.2. Tem-se que $p = 0,47$ logo

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \stackrel{a}{\sim} N(0, 1)$$

portanto,

$$P_{\hat{p}}(\hat{p} > 0,5) = P_Z\left(Z > \frac{0,5 - 0,47}{\sqrt{\frac{0,47 \times 0,53}{500}}}\right) = P_Z(Z > 1,34) = 0,09$$

Exemplo 6.3. Tem-se que $X \sim N(180, 40^2)$ logo para uma amostra de 16 elementos tem-se que $\bar{X} \sim N\left(180, \frac{40^2}{16}\right)$, portanto:

(a)

$$P(X > 168, X < 192) = 1 - P(168 \leq X \leq 192) = 1 - P(-1,2 \leq Z \leq 1,2) = 0,2301;$$

(b) $36 \times P(X > 175) = P(Z > -0,125) \approx 20$;

(c) Do problema tem-se que $p = 0,2$ e $P_{\hat{p}}(\hat{p} \leq 0,1) = 0,05$ isto implica que,

$$\frac{0,1 - 0,2}{\sqrt{\frac{0,2 \times 0,8}{n}}} = -1,64 \text{ logo } \sqrt{\frac{0,2 \times 0,8}{n}} = \frac{1}{16,4}$$

Deste modo, segue que

$$P_{\hat{p}}(\hat{p} > 0,25) = P_Z(Z > 0,82) = 0,2061.$$

Capítulo 7

Inferência Estatística

Objetivo: Produzir afirmações a respeito de uma determinada população de interesse, usualmente sobre características desta população, a partir de uma amostra desta população.

Exemplo 7.1. Para investigar se um determinado processo está produzindo peças dentro das especificações técnicas exigidas, neste caso diâmetro nominal de projeto é 15 mm, realizou-se o seguinte experimento: coletou-se uma amostra aleatória de 50 peças e mediu-se o diâmetro de cada uma, obtendo-se um diâmetro médio de $\bar{X} = 16,5$ mm. Esta é uma estimativa pontual da verdadeira média populacional μ .

A próxima questão é: Qual a margem de erro (E) desta estimativa? Ou de outra maneira, para qual intervalo de valores possíveis para μ ,

$$(\bar{X} - E ; \bar{X} + E)$$

posso ter uma confiança $100(1 - \alpha)\%$ de que este intervalo conterá o verdadeiro valor μ ?

Uma outra questão de interesse é: Será que o valor de \bar{X} mostra evidências que $\mu = 15$ mm?

Descrevemos neste exemplo, os três problemas básicos da Inferência Estatística:

- (i) Estimação pontual;
- (ii) Intervalo de confiança;
- (iii) Teste de hipótese.

7.1 Estimação Pontual

Objetivo: Encontrar estimadores que possuam boas propriedades, para que a partir deles se possa encontrar estimativas para os parâmetros populacionais de interesse.

Definição 7.1 (Estimador). É uma função da amostra, logo é também uma variável aleatória. Ex.: Dada uma amostra aleatória $\tilde{X} = (X_1, \dots, X_n)$ da variável X tem-se que um estimador para a média é dado por:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

Definição 7.2 (Estimativa). *É um particular valor numérico assumido por um estimador. Ex.: Dado a amostra $\tilde{X} = (5, 4, 6)$ tem-se que,*

$$\bar{X} = \frac{5+4+6}{3} = 5$$

é uma estimativa para μ .

Notação: $\begin{cases} \theta : & \text{parâmetro populacional de interesse} \\ \hat{\theta} : & \text{Estimador para } \theta \end{cases}$

7.1.1 Propriedades de um estimador

Que propriedades deveríamos esperar de um bom estimador? É importante que a distribuição seja o mais concentrada possível em torno do verdadeiro valor do parâmetro θ . Se tal ocorrer, então quase toda a vez que for extraída uma amostra, a estimativa resultante $\hat{\theta}$ estará próxima do verdadeiro valor θ .

Não viciado. Um estimador $\hat{\theta}$ é não viciado para θ se, $E(\hat{\theta}) = \theta$, para todo $\theta \in \Theta$ (espaço paramétrico) e para todo n (tamanho da amostra). Portanto o vício de um estimador é dado por,

$$b(\theta) = E(\hat{\theta}) - \theta.$$

Exemplo 7.2. Para $\hat{\mu} = \bar{X}$ temos que,

$$E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{E(X_1) + \dots + E(X_n)}{n}$$

Suposição: $\tilde{X} = (X_1, \dots, X_n)$ é uma amostra aleatória da variável X que tem média μ e variância σ^2 , portanto $E(X_1) = E(X_2) = \dots = E(X_n) = E(X) = \mu$ e $Var(X_1) = Var(X_2) = \dots = Var(X_n) = Var(X) = \sigma^2$. Logo,

$$E(\bar{X}) = \frac{n \times \mu}{n} = \mu.$$

Portanto \bar{X} é um estimador não viciado para μ .

Consistência. Um estimador $\hat{\theta}$ é consistente se ele for assintoticamente não viciado, isto é,

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

e se sua variância tende a zero quando n aumenta, isto é,

$$\lim_{n \rightarrow \infty} Var(\hat{\theta}) = 0$$

7.2 Intervalo de Confiança

Em muitas situações, uma estimativa pontual não fornece informação suficiente sobre um parâmetro. No exemplo sobre o processo produtivo de uma peça em que o diâmetro nominal de projeto era 15 mm e a partir de uma amostra aleatória de 50 peças, verificou-se um diâmetro médio de $\bar{X} = 16,5$ mm. Entretanto, é improvável que a verdadeira média μ seja exatamente igual a 16,5. Assim, é necessário que se saiba o quão preciso foi a estimativa pontual obtida. Uma maneira de se fazer isso é através de uma estimativa intervalar do parâmetro denominado intervalo de confiança.

Um intervalo de confiança é um intervalo de valores utilizado para estimar o verdadeiro valor de parâmetro populacional. De um modo geral, estamos interessados em encontrar um intervalo da forma $(\hat{\theta} - E; \hat{\theta} + E)$, em que $\hat{\theta}$ é o estimador de um parâmetro de interesse θ e E é a margem de erro ou erro de precisão.

Definição 7.3 (Margem de Erro). *Seja $\epsilon = \hat{\theta} - \theta$ o erro amostral, então, a margem de erro é definido como a diferença máxima provável, com probabilidade $1 - \alpha$, entre o estimador $\hat{\theta}$ e o parâmetro θ , isto é,*

$$P(|\hat{\theta} - \theta| \leq E) = 1 - \alpha$$

Para [BUSSAB E MORETTIN \(2005\)](#) a margem de erro é denominada erro amostral máximo, enquanto que [TRIOLO \(2005\)](#) afirma que a margem de erro é também conhecida como erro máximo de estimativa.

Da definição de margem de erro, percebe-se que todo intervalo de confiança está associado a um nível de confiança $100(1 - \alpha)\%$ que é a probabilidade de que o intervalo contenha o verdadeiro valor do parâmetro, isto é,

$$P(\hat{\theta} - E < \theta < \hat{\theta} + E) = 1 - \alpha, \quad 0 < \alpha < 1$$

Logo, α será a probabilidade de que o intervalo não contenha o verdadeiro valor do parâmetro.

A margem de erro E deverá ser tal que,

$$P(|\hat{\theta} - \theta| \leq E).$$

Deste modo, considerando que $\hat{\theta} \sim D(\theta, \sigma_{\hat{\theta}}^2)$, segue que,

$$|\hat{\theta} - \theta| = \begin{cases} \hat{\theta} - \theta & \text{se } \hat{\theta} \geq \theta \\ -(\hat{\theta} - \theta) & \text{se } \hat{\theta} < \theta. \end{cases}$$

Assim,

$$\{|\hat{\theta} - \theta| \leq E\} = \{\hat{\theta} - \theta \leq E\} \cap \{-(\hat{\theta} - \theta) \geq E\} = \{-E \leq \hat{\theta} - \theta \leq E\}$$

Portanto,

$$\begin{aligned} P(|\hat{\theta} - \theta| \leq E) &= P(-E \leq \hat{\theta} - \theta \leq E) = P\left(-\frac{E}{\sigma_{\hat{\theta}}} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq \frac{E}{\sigma_{\hat{\theta}}}\right) \\ &= P\left(-\frac{E}{\sigma_{\hat{\theta}}} \leq W \leq \frac{E}{\sigma_{\hat{\theta}}}\right) = P(-w_{\alpha_1} \leq W \leq w_{\alpha_2}) = 1 - \alpha \end{aligned}$$

em que $\alpha_1 + \alpha_2 = \alpha$. Se a distribuição de $\hat{\theta}$ for simétrica, então $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. Logo, considerando a simetria tem-se que

$$E = w_{\frac{\alpha}{2}} \sigma_{\hat{\theta}}$$

Notação: $IC(\theta; (1 - \alpha)\%) = (\hat{\theta} - E; \hat{\theta} + E)$

7.3 Intervalo de Confiança para a Média

Seja $\mathbf{X} = (X_1, \dots, X_n)$ uma amostra iid (independente e identicamente distribuída).

7.3.1 Caso 1: X possui distribuição normal com Variância conhecida.

Tem-se que, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ e

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

assim,

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow IC(\mu; (1 - \alpha)\%) = \left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

7.3.2 Caso 2: X possui distribuição normal com Variância desconhecida.

Quando a Variância σ^2 é desconhecida, substituímos σ^2 por S^2 , assim,

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

portanto,

$$E = t_{(n-1, \frac{\alpha}{2})} \frac{S}{\sqrt{n}} \Rightarrow IC(\mu; (1 - \alpha)\%) = \left(\bar{X} - t_{(n-1, \frac{\alpha}{2})} \frac{S}{\sqrt{n}}; \bar{X} + t_{(n-1, \frac{\alpha}{2})} \frac{S}{\sqrt{n}} \right)$$

7.3.3 Caso 3: Grandes Amostras: $n \geq 30$.

Se a Variância σ^2 for desconhecida,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \stackrel{a}{\sim} N(0, 1)$$

logo,

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow IC(\mu; (1 - \alpha)\%) = \left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Se a Variância σ^2 for desconhecida, substituímos σ^2 por S^2 ,

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \stackrel{a}{\sim} N(0, 1)$$

logo,

$$E = z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \Rightarrow IC(\mu; (1 - \alpha)\%) = \left(\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}; \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

Exemplo 7.3. Em uma amostra aleatória de 25 mulheres observou-se uma taxa média de hemoglobina de 16g/100ml. Supondo que a taxa de hemoglobina em mulheres é uma variável aleatória com distribuição normal com desvio padrão $\sigma = 1$ g/100ml de sangue. Determine um intervalo de confiança com um nível de confiança de 95% para a média μ . Se a taxa média de hemoglobina em mulheres normais fosse de 17g/100ml, o que você pode concluir a partir do IC acima?

Solução: Do problema tem-se que $\bar{X} = 16$ e $\sigma = 1$. Tem-se ainda que $\alpha = 0,05$ portanto $z_{\frac{\alpha}{2}} = 1,96$. Assim,

$$IC(\mu; 95\%) = \left(16 - 1,96 \frac{1}{\sqrt{25}}; 16 + 1,96 \frac{1}{\sqrt{25}} \right) = (15,6; 16,4)$$

7.4 Intervalo de Confiança para a proporção

Seja $X \sim ber(p)$. Retirada uma amostra aleatória (X_1, \dots, X_n) da variável X , tem-se que, $Y = X_1 + \dots + X_n \sim b(n, p)$, pois Y conta o número de vezes que um certo evento de interesse A aparece na amostra. Um estimador para o parâmetro p é dado por,

$$\hat{p} = \frac{X_1 + \dots + X_n}{n}.$$

Do Teorema Central do Limite, segue que,

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1)$$

e portanto, para n grande ($n \times \min(p, 1 - p) > 10$),

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

Como p não é conhecido a variância do estimador \hat{p} também não é conhecida e portanto deveremos utilizar o próprio estimador \hat{p} para estimá-la. Nestas condições, segue que,

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \stackrel{a}{\sim} N(0, 1)$$

Intervalo de Confiança para a proporção

Suposições:

- A amostra é aleatória simples;
- As condições para a distribuição binomial são satisfeitas.

- A distribuição normal pode ser utilizada para aproximar a distribuição das proporções amostrais se $(n \times \min(p, 1 - p) > 10)$ são satisfeitos.

Um intervalo de confiança com nível de confiança de $(1 - \alpha)\%$ é dado por:

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \Rightarrow IC(p; (1 - \alpha)\%) = \left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

Exemplo 7.4. Quando Mendel realizou seus famosos experimentos em genética com ervilhas, uma amostra das descendentes consistia de 428 ervilhas verdes e 152 ervilhas amarelas.

- Determine um intervalo de confiança com nível de confiança de 95% para a porcentagem de ervilhas amarelas;
- Com base na teoria da genética, Mendel esperava que 25% das ervilhas descendentes fossem amarelas. Dado que a porcentagem das ervilhas amarelas não é 25%, os resultados contradizem a teoria de Mendel?

Solução:

(a) Dada a amostra de 580 ervilhas, temos que uma estimativa para a proporção de ervilhas amarelas é

$$\hat{p} = \frac{152}{580} = 0,262$$

portanto, $np = 152 > 5$ e $n(1 - p) > 5$, assim,

$$\begin{aligned} IC(p; 95\%) &= \left(0,262 - 1,96 \times \sqrt{\frac{0,262(1 - 0,262)}{580}}; 0,262 + 1,96 \times \sqrt{\frac{0,262(1 - 0,262)}{580}} \right) \\ &= (0,262 - 0,036; 0,262 + 0,036) = (0,226; 0,298) \end{aligned}$$

7.5 Teste de Hipótese

Um hipótese é uma suposição a respeito de um determinado problema, por exemplo:

Um lote de parafusos, de origem desconhecida, será leiloada a um preço muito convidativo. Um indústria está interessada em adquirir um lote desses parafusos, entretanto, ela precisa saber se os parafusos satisfazem as especificações técnicas relacionadas a resistência a tração. O edital do leilão diz que, pouco antes do início do leilão será divulgada a resistência média de uma amostra de 25 parafusos. Qual a regra de decisão deve ser utilizada pela indústria?

Estas suposições podem ser formuladas através de um teste de hipótese estatístico, que é um processo de decisão para avaliar as hipóteses feitas a respeito de uma determinada população. Desta forma, testar uma hipótese, significa verificar se um pressuposto é verdadeiro ou não. Esta verificação é feita através de uma amostra coletada da população em estudo; no exemplo anterior a população era o lote de parafusos.

Portanto, o objetivo de um teste de hipótese é fornecer uma metodologia(procedimento) que nos permita verificar se os dados amostrais trazem evidências que apóiem ou não uma hipótese estatística formulada.

Assim sendo, a formulação de um teste de hipótese estatístico inicia-se com a afirmação de uma hipótese estatística.

Definição 7.4 (Hipótese Estatística). *É usualmente uma conjectura a respeito de um parâmetro populacional.*

No exemplo dos parafusos, a indústria deseja saber se a resistência média à tração é superior a 145 Kg, isto é, $\mu > 145$.

Para cada situação existem dois tipos de hipótese estatística: a hipótese nula denotada por H_0 e a hipótese alternativa denotada por H_1

Existem basicamente 3 tipos de formulações para os testes de hipótese:

Situação A. Uma máquina automática para encher pacotes de café foi regulada para colocar em média 500 g de café com uma variância de 400 g². Após algum tempo de trabalho, deseja-se verificar se a média do processo está sob controle, as hipóteses para esta situação são:

$$\begin{cases} H_0: \mu = 500 \\ H_1: \mu \neq 500 \end{cases}$$

Este teste é denominado teste bilateral;

Situação B. O dono de uma fábrica de confecção de tapetes está desconfiado que está havendo um gasto excessivo de tinta em uma das etapas do processo. Sabe-se que a quantidade média de tinta gasta no processo é de 1,6 l, as hipóteses para esta situação são:

$$\begin{cases} H_0: \mu = 1,6 \quad \text{ou} \quad \mu \leq 1,6 \\ H_1: \mu > 1,6 \end{cases}$$

Este teste é denominado teste unilateral à direita;

Situação C. Uma companhia farmacêutica desconfia que o tempo de duração do efeito de um medicamento da companhia concorrente é menor que o anunciado por ela que é 225 minutos, as hipóteses para esta situação são:

$$\begin{cases} H_0: \mu = 225 \quad \text{ou} \quad \mu \geq 225 \\ H_1: \mu < 225 \end{cases}$$

Este teste é denominado teste unilateral à esquerda.

Em um teste de hipótese, existem apenas quatro resultados possíveis:

	H_0 é verdadeira	H_0 é falsa
Rejeitar H_0	Erro tipo I	Decisão correta
Não Rejeitar H_0	Decisão correta	Erro tipo II

Elementos de um teste de hipótese

Nível de significância: É a probabilidade de se cometer o erro tipo I, é denotado por α , isto é,

$$P(\text{Erro tipo I}) = \alpha = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}).$$

Beta do teste: É a probabilidade de se cometer o erro tipo II, é denotado por β , e é dado por,

$$P(\text{Erro tipo II}) = \beta = P(\text{Não Rejeitar } H_0 | H_0 \text{ é falsa}).$$

Região Crítica(RC): É o conjunto de valores de $\hat{\theta}$ para o qual a hipótese deve ser rejeitada, também chamada de região de rejeição.

Nível descritivo ou p -valor do teste: É a probabilidade de ocorrer valores do estimador $\hat{\theta}$, mais extremos que o valor observado $\hat{\theta}(\omega) = x$, isto é, que a estimativa obtida, sob a hipótese que H_0 é verdadeira, isto é,

- Se $H_1 : \theta > \theta_0$ então $x - \theta_0 > 0$, assim

$$p\text{-valor} = P(\hat{\theta} > x | H_0 \text{ é verdadeira}) = P\left(W > \frac{x - \theta_0}{\sigma_{\hat{\theta}}}\right);$$

- Se $H_1 : \theta < \theta_0$ então $x - \theta_0 < 0$, assim

$$p\text{-valor} = P(\hat{\theta} < x | H_0 \text{ é verdadeira}) = P\left(W < \frac{x - \theta_0}{\sigma_{\hat{\theta}}}\right).$$

Logo, em qualquer uma dessas situações tem-se que

$$p\text{-valor} = P\left(W > \frac{|x - \theta_0|}{\sigma_{\hat{\theta}}}\right)$$

- Se $H_1 : \theta \neq \theta_0$ então, $x - \theta_0 > 0$ ou $x - \theta_0 < 0$, assim

$$p\text{-valor} = 2 \times P\left(W > \frac{|x - \theta_0|}{\sigma_{\hat{\theta}}}\right)$$

Observe que quanto menor for o p -valor, mais forte será a evidência de que a hipótese H_0 não é verdadeira. Portanto, o p -valor mede a força da evidência contra H_0 . Em outras palavras, quanto menor o p -valor menor será a probabilidade de H_0 ser verdadeira.

Observação 7.1. Sempre que acontecer $\hat{\theta}(\omega) = x = \theta_0$ então não rejeita-se a hipótese H_0 .

7.6 Procedimento Geral do Teste de Hipótese - Uma Amostra

1. Formulação das hipóteses:

$$\text{Situação A: } \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \quad \text{Situação B: } \begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{Situação C: } \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

2. p -valor:

- Nas situações B e C,

$$\text{p-valor} = P\left(W > \frac{|x - \theta_0|}{\sigma_{\hat{\theta}}}\right)$$

- Na situação A,

$$\text{p-valor} = 2 \times P\left(W > \frac{|x - \theta_0|}{\sigma_{\hat{\theta}}}\right)$$

3. Região crítica:

$$\text{Situação A: } RC =] -\infty, -w_{c_1}] \cup [w_{c_2}, \infty[$$

$$\text{Situação B: } RC =] -\infty, -w_c]$$

$$\text{Situação C: } RC = [w_c, \infty[$$

em que w_{c_1} , w_{c_2} e w_c satisfaz as seguintes condições:

$$P(W \leq w_{c_1}) = \alpha_1$$

$$P(W \geq w_{c_2}) = \alpha_2.$$

$$P(W \geq w_c) = P(W \leq -w_c) = \alpha.$$

em que $\alpha_1 + \alpha_2 = \alpha$. Se a distribuição de W for simétrica então, $w_{c_2} = -w_{c_1}$ e nesse caso $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. A variável transformada W é chamada de **estatística do teste**, e nesse caso como a distribuição de W não depende de nenhum parâmetro desconhecido, denominamos de quantidade pivotal.

Se $\sigma_{\hat{\theta}}$ não for conhecido então substitui-se pelo respectivo estimador $\hat{\sigma}_{\hat{\theta}}$.

4. Decisões e Conclusões possíveis:

Pelo método do p-valor:

- rejeitar H_0 se $\text{p-valor} \leq \alpha$.

Conclusão: Como $\text{p-valor} \leq \alpha$ rejeitamos H_0 ao nível de significância de $100\alpha\%$. Logo, existem evidências de que a hipótese H_1 é verdadeira;

- não rejeitar H_0 caso contrário.

Conclusão: Como $\text{p-valor} > \alpha$ não rejeitamos H_0 ao nível de significância de $100\alpha\%$. Logo, não existem evidências de que a hipótese H_1 é verdadeira.

Pelo método da região crítica:

- rejeitar H_0 se $W_{cal} \in RC$.

Conclusão: Como $W_{cal} \in RC$ rejeitamos H_0 ao nível de significância de $100\alpha\%$. Logo, existem evidências de que a hipótese H_1 é verdadeira;

- não rejeitar H_0 se $W_{cal} \notin RC$.

Conclusão: Como $W_{cal} \notin RC$ não rejeitamos H_0 ao nível de significância de $100\alpha\%$. Logo, não existem evidências de que a hipótese H_1 é verdadeira.

7.7 Teste de hipótese para a média

Seja X uma variável aleatória com média μ e desvio padrão σ . Seja $\hat{\mu} = \bar{X}$ um estimador para μ e $\sigma_{\bar{X}}$ o desvio padrão deste estimador.

7.7.1 Caso 1: X possui distribuição normal com Variância conhecida.

Estatística do teste:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Região crítica:

$$\text{Situação A: } RC = \{x \in \mathbb{R} : x \leq -z_{\frac{\alpha}{2}} \text{ ou } x \geq z_{\frac{\alpha}{2}}\}$$

$$\text{Situação B: } RC = \{x \in \mathbb{R} : x \geq z_{\alpha}\}$$

$$\text{Situação C: } RC = \{x \in \mathbb{R} : x \leq -z_{\alpha}\}$$

p-valor do teste:

- Para as situações B e C tem-se que $\text{p-valor} = P\left(Z > \frac{|x - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$

- Para a situação A tem-se que $\text{p-valor} = 2 \times P\left(Z > \frac{|x - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$

7.7.2 Caso 2: X possui distribuição normal com Variância desconhecida.

Estatística do teste:

$$T = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1};$$

Região crítica:

$$\text{Situação A: } RC = \{x \in \mathbb{R} : x \leq -t_{(n-1, \frac{\alpha}{2})} \text{ ou } x \geq t_{(n-1, \frac{\alpha}{2})}\}$$

$$\text{Situação B: } RC = \{x \in \mathbb{R} : x \geq t_{(n-1, \alpha)}\}$$

$$\text{Situação C: } RC = \{x \in \mathbb{R} : x \leq -t_{(n-1, \alpha)}\}$$

p-valor do teste: $\text{p-valor} = P\left(T > \frac{|x - \mu_0|}{\frac{S}{\sqrt{n}}}\right)$ **p-valor do teste:**

- Para as situações B e C tem-se que $\text{p-valor} = P\left(T > \frac{|x - \mu_0|}{\frac{S}{\sqrt{n}}}\right)$

- Para a situação A tem-se que $\text{p-valor} = 2 \times P\left(T > \frac{|x - \mu_0|}{\frac{S}{\sqrt{n}}}\right)$

7.7.3 Caso 3: Grandes Amostras: $n \geq 30$.

Estatística do teste:

- Se a variância for conhecida:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \stackrel{a}{\sim} N(0, 1)$$

- Se a variância for desconhecida:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \stackrel{a}{\sim} N(0, 1)$$

Região crítica:

$$\text{Situação A: } RC = \{x \in \mathbb{R} : x \leq -z_{\frac{\alpha}{2}} \text{ ou } x \geq z_{\frac{\alpha}{2}}\}$$

$$\text{Situação B: } RC = \{x \in \mathbb{R} : x \geq z_{\alpha}\}$$

$$\text{Situação C: } RC = \{x \in \mathbb{R} : x \leq -z_{\alpha}\}$$

p-valor do teste:

- Para as situações B e C tem-se que $p\text{-valor} = P\left(Z > \frac{|x - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$
- Para a situação A tem-se que $p\text{-valor} = 2 \times P\left(Z > \frac{|x - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$

7.8 Teste de hipótese para a proporção

Seja X uma variável aleatória com distribuição $X \sim \text{ber}(p)$. Seja $\underline{X} = (X_1, \dots, X_n)$ uma amostra i.i.d. de X , então um estimador para o parâmetro p é dado por

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \frac{k}{n}$$

em que k é o número de vezes que o evento de interesse aparece na amostra \underline{X} .

Estatística do teste: pelo Teorema Central do Limite, tem-se para n grande que a estatística do teste é dada por

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{a}{\sim} N(0, 1)$$

Região crítica:

$$\text{Situação A: } RC = \{x \in \mathbb{R} : x \leq -z_{\frac{\alpha}{2}} \text{ ou } x \geq z_{\frac{\alpha}{2}}\}$$

$$\text{Situação B: } RC = \{x \in \mathbb{R} : x \geq z_{\alpha}\}$$

$$\text{Situação C: } RC = \{x \in \mathbb{R} : x \leq -z_{\alpha}\}$$

p-valor do teste:

- Para as situações B e C tem-se que p-valor = $P\left(Z > \frac{|p-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$
- Para a situação A tem-se que p-valor = $2 \times P\left(Z > \frac{|p-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$

Capítulo 8

Correlação e Regressão Linear Simples

Nesse capítulo iremos estudar a Correlação e a Regressão Linear Simples. Na primeira seção iremos tratar sobre coeficiente de correlação linear que é um coeficiente que mede a intensidade da relação linear entre duas variáveis. Na segunda seção trataremos da regressão linear simples. Na análise de regressão o objetivo é investigar a relação entre as variáveis e prever o valor de uma em função da outra.

8.1 Coeficiente de Correlação Linear(ρ)

O coeficiente de correlação linear é utilizado quando se deseja verificar se duas variáveis estão relacionadas. Mais especificamente, se duas variáveis possuem relação linear entre elas. Esse coeficiente é também denominado correlação de Pearson.

Definição 8.1 (Coeficiente de Correlação Linear). *Sejam X e Y duas variáveis aleatórias com média μ_X e μ_Y e desvio padrão σ_X e σ_Y respectivamente, então o Coeficiente de Correlação Linear é definido como,*

$$\rho_{X,Y} = \rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

Propriedades:

1. O coeficiente de correlação linear independe da unidade de medida das variáveis. Trata-se de um número adimensional;
2. O coeficiente de correlação linear é invariante sobre transformações lineares, isto é, se $U = aX + b$ e $V = cY + d$ então, $\rho_{U,V} = \rho_{X,Y}$;
3. O coeficiente de correlação linear é um valor entre -1 e 1, em que:
 - (a) Se $\rho < 0$ temos uma relação negativa, isto é, uma relação linear inversa;
 - (b) Se $\rho > 0$ temos uma relação positiva, isto é, uma relação linear direta;
 - (c) Se $\rho = 0$ temos uma ausência relação linear;

(d) Se $|\rho| = 1$ temos uma relação linear perfeita.

Definição 8.2 (Coeficiente de Correlação Linear amostral). *Dada uma amostra i.i.d das variáveis X e Y, $((X_1, Y_1), \dots, (X_n, Y_n))$, então um estimador ($\hat{\rho}$) para o Coeficiente de Correlação Linear é dado por,*

$$\hat{\rho} = r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) \sqrt{\left(\sum_{i=1}^n (Y_i - \bar{Y})^2\right)}} = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{\sqrt{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i\right)^2}}$$

8.1.1 Interpretação geométrica

O produto escalar de dois vetores $\mathbf{A} = (a_1, a_2, \dots, a_n)$ e $\mathbf{B} = (b_1, b_2, \dots, b_n)$ é o resultado do produto do comprimento (também chamado de norma ou módulo) de \mathbf{A} pela [[projeção escalar]] de \mathbf{B} em \mathbf{A} , isto é,

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \alpha$$

Onde α é o ângulo formado pelos vetores e $\|\mathbf{A}\|$ e $\|\mathbf{B}\|$ são seus comprimentos, dados por,

$$\|\mathbf{A}\| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

e

$$\|\mathbf{B}\| = \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}$$

O produto escalar entre dois vetores também pode ser visto como,

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

Deste modo o cosseno do angulo entre os dois vetores (α) é dado por:

$$\cos(\alpha) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \times \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}} \quad (8.1)$$

Considere duas amostras i.i.d. das variáveis X e Y, (X_1, \dots, X_n) de X e (Y_1, \dots, Y_n) . Essas amostras podem ser consideradas como vetores em um espaço de n dimensões. Assim, subtraindo cada valor de sua respectiva média, tem-se $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ e $(Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})$. Assim, da equação 8.1 o cosseno do ângulo α entre estes vetores é dado por:

$$\cos(\alpha) = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Logo, $\cos(\alpha) = \rho$. Sendo assim:

- Se $\rho = 1$, o ângulo $\alpha = 0^\circ$, os dois vetores são colineares (paralelos);
- Se $\rho = 0$, o ângulo $\alpha = 90^\circ$, os dois vetores são ortogonais;
- Se $\rho = -1$, o ângulo $\alpha = 180^\circ$, os dois vetores são colineares com sentidos opostos;

8.1.2 Teste de hipótese para o Coeficiente de Correlação

Hipótese:

$$\begin{cases} H_0: \rho = \rho_0 \\ H_1: \rho \neq \rho_0 \end{cases}$$

Estatística do Teste:

$$T = |r| \sqrt{\frac{n-2}{1-r^2}}$$

Região crítica: $RC = \{x \in [0, \infty) : x \geq t_{\frac{\alpha}{2}}\}$

Decisão: rejeitar H_0 se $T_{cal} \in RC$

8.2 Regressão Linear Simples

Tem por objetivo encontrar qual a relação linear entre as variáveis aleatórias, se a mesma existir.

Relação linear simples: $Y = b_0 + b_1X + e$. Em que, e é erro aleatório. Dada uma amostra $\underline{X} = ((X_1, Y_1), \dots, (X_n, Y_n))$ tem-se que,

$$Y_i = b_0 + b_1X_i + e_i$$

onde e_i é suposto ter distribuição normal com média zero e variância σ^2 com (e_1, \dots, e_n) independentes e identicamente distribuídos.

Nestas condições deseja-se estimar b_0 e b_1 obtendo-se assim a reta estimada $\hat{Y}_i = \hat{b}_0 + \hat{b}_1X_i$, para a partir dela podermos fazer predições de Y a partir de valores conhecidos de X .

Observação 8.1. *A variável X é denominada variável independente ou explicativa e a variável Y de variável dependente ou resposta.*

8.3 Estimação dos parâmetros

O método de mínimos quadrados é usado para estimar os parâmetros do modelo (b_0 e b_1) e consiste em fazer com que a soma dos erros quadráticos seja menor possível, ou seja, este método consiste em obter os valores de b_0 e b_1 que minimizam a expressão:

$$f(b_0, b_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - (b_0 + b_1X_i))^2$$

Aplicando-se derivadas parciais à expressão acima, e igualando-se a zero, acharemos as seguintes estimativas para b_0 e b_1 , as quais chamaremos de \hat{b}_0 e \hat{b}_1 , respectivamente:

$$\hat{b}_0 = \frac{\sum_{i=1}^n Y_i - \hat{b}_1 \sum_{i=1}^n X_i}{n}$$

$$\hat{b}_1 = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$

A chamada equação (reta) de regressão é dada por

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i.$$

A diferença entre os valores observados e os preditos é chamada de resíduo (\hat{e}_i):

$$\hat{e}_i = Y_i - \hat{Y}_i$$

O resíduo relativo à i -ésima observação (\hat{e}_i) pode ser considerado uma estimativa do erro aleatório (e_i) desta observação.

8.3.1 Coeficiente de Determinação (R^2)

O coeficiente de determinação é uma medida descritiva da proporção da variação de Y que pode ser explicada por variações em X , segundo o modelo de regressão especificado. Ele é dado pela seguinte razão:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} = 1 - \frac{\sum_{i=1}^n Y_i^2 - \hat{b}_0 \sum_{i=1}^n Y_i - \hat{b}_1 \sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 / n}$$

Referências Bibliográficas

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**, 5ª Edição, São Paulo: Saraiva, 2005.

FISHER, R. A. On the Mathematical Foundations of Theoretical Statistics. **Philosophical Transactions of the Royal Society**, A, v.222, p.309-368, 1922.

FISHER, R. A. **Statistical methods for research workers**. Edinburgh: Oliver and Boyd, 1925. (Biological monographs and manuals, n.5)

GRAUNT, J. (1662). **Bills of Mortality**. London. Disponível em <<http://www.ac.wvu.edu/~stephan/Graunt/bills.html>>. Acesso em: 5 de novembro de 2007.

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Normas de apresentação tabular**. 3. ed. Rio de Janeiro, 1993. 63p.

KOLMOGOROV, A. N. **Foundations of the Theory of Probability**. 2. ed., New York: Chelsea Publishing Company, 1956. 84p. Original publicado em 1933 em Alemão como “Grundbegriffe der Wahrscheinlichkeitrechnung”. Disponível em <<http://www.kolmogorov.com/Foundations.html>>. Acesso em: 5 de novembro de 2007.

TRIOLA, M. F. **Introdução à Estatística**, Tradução da 9ª Edição, Rio de Janeiro: LTC, 2005.