

Estatística

Heyder Diniz Silva

Sumário

1	Introdução	7
2	Estatística descritiva	9
2.1	Organização e apresentação de dados.	9
2.1.1	Apresentação tabular	11
2.1.1.1	Distribuições de frequências	11
2.1.1.2	Distribuições de frequências acumuladas	16
2.1.2	Apresentação gráfica	19
2.1.2.1	Histogramas	19
2.1.2.2	Polígonos de frequência	20
2.1.2.3	Gráfico de setores (pizza)	21
2.1.2.4	Ogivas	22
2.2	Medidas de Posição e dispersão	25
2.2.1	Medidas de Posição	25
2.2.1.1	Média	28
2.2.1.2	Mediana	34
2.2.1.3	Moda	36
2.2.1.4	Separatrizes (Quantis)	38
2.2.2	Medidas de dispersão	40
2.2.2.1	Amplitude Total (A)	41
2.2.2.2	Variância e desvio padrão	42
2.2.2.3	Coefficiente de variação	47

		3
	2.2.2.4 Erro padrão da média	48
	2.2.2.5 Momentos, Assimetria e Curtose	48
3	Probabilidades	54
3.1	Processo aleatório:	54
3.2	Espaço amostral e Eventos	55
3.3	Probabilidade	56
3.4	Probabilidade Condicionada	58
3.5	Independência de eventos.	59
3.6	Teorema de Bayes	61
4	Variáveis aleatórias	63
4.1	Variáveis Aleatórias Unidimensionais.	63
	4.1.1 Distribuições de Probabilidades	63
	4.1.2 Função Repartição ou Distribuição Acumulada $F(x)$	64
4.2	Variáveis Aleatórias Bidimensionais	66
	4.2.1 Distribuição Conjunta de duas variáveis aleatórias	66
	4.2.2 Distribuição Marginal	67
	4.2.3 Variáveis Aleatórias Independentes	68
	4.2.4 Esperança Matemática	69
	4.2.4.1 Propriedades da Esperança Matemática	70
	4.2.5 Variância de uma variável aleatória.	70
	4.2.5.1 Propriedades da variância	72
	4.2.6 Covariância	72
4.3	Distribuições de variáveis aleatórias discretas	73
	4.3.1 Distribuição Uniforme Discreta	73
	4.3.2 Distribuição de Bernoulli	73
	4.3.3 Distribuição Binomial.	75
	4.3.4 Distribuição de Poisson	78
	4.3.4.1 Aproximação da distribuição Binomial a Poisson.	81

	4
4.3.5	Distribuição Geométrica 82
4.3.6	Distribuição Pascal (Binomial Negativa) 84
4.3.7	Distribuição Hipergeométrica 86
4.3.8	Distribuição Multinomial 87
4.4	Distribuições de variáveis aleatórias contínuas 88
4.4.1	Distribuição Uniforme 88
4.4.2	Distribuição Normal 90
4.4.2.1	Distribuição Normal Reduzida ou Padronizada. 91
4.4.3	Distribuição Exponencial 94
4.4.4	Distribuição Qui-Quadrado 96
4.4.5	Distribuição t de Student 97
4.4.6	Distribuição F de Snedcor 98
4.4.7	Aproximação da Distribuição Binomial à Normal 99
5	Amostragem 101
5.1	Introdução. 101
5.1.1	Definições 101
5.1.2	Importância do uso de amostras. 102
5.1.3	Vantagens do processo de amostragem em relação ao censo. 103
5.2	Técnicas de amostragem. 104
5.2.1	Principais técnicas de amostragem probabilísticas. 105
5.2.1.1	Amostragem Simples ao Acaso 105
5.2.1.2	Amostragem Sistemática 105
5.2.1.3	Amostragem por Conglomerados 105
5.2.1.4	Amostragem Estratificada 106
5.2.2	Principais técnicas de amostragem não probabilísticas. 106
5.2.2.1	Inacessibilidade a toda população 106
5.2.2.2	Amostragem sem norma (a esmo) 107
5.2.2.3	População formada por material contínuo. 107
5.2.2.4	Intencional 107

5.3	Distribuições Amostrais	108
5.3.1	Distribuição amostral da média	108
5.3.1.1	Amostragem com reposição	109
5.3.1.2	Amostragem sem reposição	111
6	Inferência	113
6.1	Teoria da estimação	113
6.1.1	Definições	113
6.1.2	Propriedades dos Estimadores	114
6.1.2.1	Não tendenciosidade	114
6.1.2.2	Consistência.	116
6.1.2.3	Eficiência	116
6.1.3	Intervalos de confiança	117
6.1.3.1	Intervalo de confiança para a média μ	117
6.1.3.1.1	Variância conhecida	117
6.1.3.1.2	Variância desconhecida	119
6.1.3.2	Diferença entre duas média ($\mu_a - \mu_b$)	120
6.1.3.2.1	Variâncias Conhecidas:	120
6.1.3.2.2	Variâncias Desconhecidas:	122
6.1.3.3	Intervalo de confiança para proporção	123
6.1.3.3.1	Amostras grandes ($n > 30$)	123
6.1.3.3.2	Amostras pequenas ($n \leq 30$)	123
6.1.3.4	Intervalo de confiança para a diferença entre proporções	124
6.1.3.4.1	Amostras grandes ($n > 30$)	124
6.1.3.4.2	Amostras pequenas ($n \leq 30$)	124
6.1.3.5	Intervalo de confiança para a variância (σ^2)	124
6.2	Teoria da decisão (Testes de Hipóteses)	125
6.2.1	Metodologia de um teste de hipótese	125
6.2.2	Tipos de erros	127
6.2.3	Tipos de testes	128

	6
6.2.4 Algoritmo para realização de um teste de hipótese	129
6.2.5 Estatística apropriadas para os testes de hipóteses	130
6.2.6 Teste de Qui-Quadrado (χ^2)	133
6.2.6.1 Teste de aderência	134
6.2.6.2 Teste de independência	136
7 Regressão e Correlação linear	139

Capítulo 1

Introdução

O cidadão comum pensa que a estatística se resume apenas a apresentar tabelas de números em colunas esportivas e ou econômicas de jornais e revistas, ilustradas com gráficos, pilhas de moedas, etc. ou quando muito associam a estatística à previsão de resultados eleitorais. Mas estatístico de hoje não se limita a compilar tabelas de dados e os ilustrar graficamente. Pois à partir de 1925, com os trabalhos de Fisher, a estatística iniciou-se como método científico, então, o trabalho do estatístico passou a ser o de ajudar a planejar experimentos, interpretar e analisar os dados experimentais e apresentar os resultados de maneira a facilitar a tomada de decisões razoáveis. Deste modo, podemos então definir estatística como sendo a ciência que se preocupa com a coleta, organização, apresentação, análise e interpretação de dados. Didaticamente podemos dividir a estatística em duas partes a estatística descritiva e a inferência estatística. A estatística descritiva se refere a maneira de apresentar um conjunto de dados em tabelas e gráficos, e ao modo de resumir as informações contidas nestes dados a algumas medidas. Já a inferência estatística baseia-se na teoria das probabilidades para estabelecer conclusões sobre todo um grupo (chamado população), quando se observou apenas uma parte (amostra) desta população. É necessário ter em mente que a estatística é uma ferramenta para o pesquisador, nas respostas dos “por quês” de seus problemas. E que para ela ser bem usada é necessário conhecer os seus fundamentos e princípios, e acima de tudo que o pesquisador desenvolva um espírito crítico e jamais deixe de pensar. Pois ”em ciência é fácil mentir usando a

estatística, o difícil é falar a verdade sem usar a estatística”.

Capítulo 2

Estatística descritiva

2.1 Organização e apresentação de dados.

As observações são o material básico com que o pesquisador trabalha. Estas observações podem ser por exemplo a produtividade de uma planta, a velocidade de processamento de um computador, a resistência à ruptura de determinado cabo, suscetibilidade ou não de um indivíduo a determinada doença, cor de uma flor, sexo do primeiro filho de um casal, opinião dos alunos quanto a didática de um professor, etc. Estas observações apresentam uma característica em comum que é a variação ou variabilidade, ou seja assumem diferentes valores de indivíduo para indivíduo.

Uma característica que pode assumir diferentes valores de indivíduo para indivíduo é denominada variável. Caso contrário é denominada constante. As variáveis são classificadas em:

- **QUALITATIVAS:** São aquelas para as quais uma medição numérica não é possível.
Ex: Cor de flor, sexo, sabor.
- **QUANTITATIVAS:** São aquelas para as quais é possível se realizar-se uma medição numérica, subdivide-se em:
 - **DISCRETAS:** Próprias de dados de contagem, ou seja só assumem valores inteiros. Ex: número de filhos, número de acidentes de trânsito ocorridos num

certo período, etc.

- CONTÍNUAS: São aquelas originárias de medições, deste modo, podem assumir qualquer valor real entre dois extremos. Ex: Peso corporal, altura, resistência a ruptura, volume, etc.

Os dados coletados no campo e trazidos para o laboratório (escritório), na forma em que se encontram, como os apresentados na Tabela 2.1, são denominados dados brutos. Normalmente este tipo de dados trás pouca ou nenhuma informação ao leitor, sendo necessário uma elaboração (organização) destes dados, afim de aumentar sua capacidade de informação.

Tabela 2.1: Diâmetro à altura do peito (DAP), em mm de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG

104	122	129	144	183
108	142	138	151	138
138	106	122	146	115
101	201	161	82	179
163	169	167	137	142
141	120	189	132	111
90	210	132	172	140
154	98	127	87	136

A mais simples organização numérica é ordenação dados em ordem crescente ou decrescente (ROL). Como pode-se observar na Tabela 2, a simples organização dos dados em um Rol, aumenta muito a capacidade de informação destes. Pois enquanto a Tabela 2.1 nos informava apenas que tínhamos 40 pés de Eucalipto, e alguns D.A.P., na Tabela 2.2, verificamos que o menor diâmetro observado foi 82 mm e o maior 210 mm, o que nos fornece uma amplitude total de variação da ordem de 128 mm.

Amplitude total

$$A = \text{maior valor observado} - \text{menor valor observado} \quad (2.1)$$

$$A = 210\text{mm} - 82\text{mm} = 128\text{mm}$$

Pode-se observar ainda que alguns diâmetros como 122 mm, 132 mm 138 mm e 142 mm são mais comuns.

Tabela 2.2: Tabela 2. Diâmetro à altura do peito (DAP), em mm de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG

82	111	132	142	167
87	115	136	142	169
90	120	137	144	172
98	122	138	146	179
101	122	138	151	183
104	127	138	154	189
106	129	140	161	201
108	132	141	163	210

2.1.1 Apresentação tabular

2.1.1.1 Distribuições de freqüências

Após esta primeira organização dos dados, podemos ainda agrupa-los em classes de menor tamanho, afim de aumentar sua a capacidade de informação.

Distribuindo-se os dados observados em classes e contando-se o numero de indivíduos contidos em cada classe, obtém-se a freqüência de classe. A disposição tabular dos dados agrupados em classes, juntamente com as freqüências correspondentes denomina-se distribuição de freqüência.

Para identificar uma classe, deve-se conhecer os valores dos limites inferior e superior da classe, que delimitam o intervalo de classe. Por exemplo, para o caso dos

DAP dos eucaliptos, pode-se desejar incluir em uma única classe todos os indivíduos que possuam DAP entre 120 e 138 mm assim, o intervalo de classe seria de 120 mm a 138 mm.

Neste ponto surge uma dúvida fundamental. Indivíduos que apresentem DAP exatamente iguais a 120 mm ou a 138 mm pertencem ou não a esta classe? Deste modo surge a necessidade de definir a natureza do intervalo de classe, se é aberto ou fechado. Quando o intervalo de classe é aberto, os limites da classe não pertencem a ela, e quando o intervalo é fechado, os limite de classe pertencem a classe em questão. Notação:

- Intervalos abertos

$]128\text{mm} - 138\text{mm}[$

- Intervalos fechados

$[128\text{mm} - 138\text{mm}]$

$(128\text{mm} - 138\text{mm})$

- Intervalos mistos

São fechados em um extremo e abertos no outro, como por exemplo:

$[128\text{mm} - 138\text{mm}[$

para este tipo de intervalo pode-se utilizar ainda a seguinte notação:

$128\text{mm} \vdash 138\text{mm}$

Construção de uma distribuição de frequência

Para montar uma distribuição de frequência é necessário que primeiramente se determine o número de classes (k) em que os dados serão agrupados. Por questões de ordem prática e estética sugere-se utilizar de 5 a 20 classes. Uma indicação do número de classes a ser utilizado, em função do número de dados (n) é:

1. $n \leq 100$

$$k = \sqrt{n}. \quad (2.2)$$

2. $n > 100$

$$k = 5\log(n). \quad (2.3)$$

3. Critério de SCOTT (1979), baseado na normalidade dos dados:

$$k = \frac{An^{\frac{1}{3}}}{3.49s} \quad (2.4)$$

em que:

A é a amplitude total;

s é o desvio padrão;

n é o número de observações.

Após determinado o número de classes (k) em que os dados serão agrupados, deve-se então determinar o intervalo de classe (c), que é dado pela seguinte expressão:

$$c = \frac{A}{k - 1} \quad (2.5)$$

em que:

c é amplitude de classe;

A é a amplitude total;

k é o número de classes.

Conhecida a amplitude de classes, determina-se então os intervalos de classe. Os limites inferior e superior das classes devem ser escolhidos de modo que o menor valor observado esteja localizado no ponto médio da primeira classe, que é dado por:

$$PM = \frac{L_{inf} - L_{sup}}{2} \quad (2.6)$$

em que:

L_{inf} é o limite inferior da classe;

L_{sup} é o limite superior da classe;

Assim, o limite inferior da primeira classe será:

$$L_{inf1} = menorvalor - \frac{c}{2} \quad (2.7)$$

E os demais limites são obtidos somando-se c ao limite anterior.

A título de ilustração agruparemos os dados referentes ao DAP de eucaliptos em classes

1º Amplitude total (A)

$A = \text{maior valor observado} - \text{menor valor observado} = 210 - 82 = 128 \text{ mm.}$

2º Determinar o número de classes (k)

$n = 40$

$$K = \sqrt{40} = 6,32$$

, como o número de classes é inteiro usaremos 6 classes.

3º Determinar a amplitude de classe (c)

$$c = \frac{128}{6 - 1} = 25,6 \text{ mm}$$

4º Determinar o limite inferior da primeira classe (Li)

$$L_{inf1} = \text{menor valor} - \frac{c}{2}$$

$$L_{inf1} = 82 - \frac{25,6}{2} = 69,2$$

5º Determinar os intervalos de classe

69,2 † 94,8

94,8 † 120,4

120,4 † 146,0

146,0 † 171,6

171,6 † 197,2

197,2 † 222,8

6º Montar a distribuição de freqüência.

Para montar a distribuição de freqüência, basta apresentar as classes obtidas na forma tabular e contar quantos indivíduos existem em classe.

Apresentando os dados na forma de distribuição de freqüência, sintetiza-se a informação contida nos mesmos, além de facilitar sua visualização. Pois pode-se verificar claramente na Tabela 2.3 que os DAP dos 40 pés de *Eucalyptus citriodora* em questão estão concentrados entorno dos valores centrais, decrescendo em direção aos valores extremos. A apresentação dos dados em forma de distribuição de freqüência facilita ainda o cálculo

Tabela 2.3: Distribuição de frequências dos DAP de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG.

Classes (mm)	Tabulação	Frequência
69,2 † 94,8		3
94,8 † 120,4		8
120,4 † 146,0		16
146,0 † 171,6		7
171,6 † 197,2		4
197,2 † 222,8		2
Total		40

de várias medidas estatísticas de interesse, além de permitir a apresentação gráfica dos mesmos.

Frequência Absoluta e Frequência Relativa.

As frequências apresentadas na Tabela 2.3 são denominadas frequências absolutas, pois indicam o número de observações pertencentes a cada classe, este tipo de frequência apresenta a limitação de sua interpretação ser dependente do número total de observações. Para solucionar este problema utiliza-se a frequência relativa, que expressa a quantidade de observações pertencentes a uma classe, em relação ao número total de observações, e é dada por:

$$Fr_i = \frac{F_i}{\sum_{i=1}^k F_i} \quad (2.8)$$

e que,

Fr_i é a frequência relativa da i -ésima classe;

F_i é a frequência absoluta da i -ésima classe;

Assim a Tabela 2.3 pode ser apresentada do seguinte modo:

Tabela 2.4: Distribuição de frequências dos DAP de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG.

Classes (mm)	Frequência Absoluta	Frequência Relativa	Frequência Relativa (%)
69,2 – 94,8	3	0,075	7,5
94,8 – 120,4	8	0,200	20,0
120,4 – 146,0	16	0,400	40,0
146,0 – 171,6	7	0,175	17,5
171,6 – 197,2	4	0,100	10,0
197,2 – 222,8	2	0,050	5,0
Total	40	1,000	100,0

2.1.1.2 Distribuições de frequências acumuladas

Muitas vezes pode-se estar interessado não em saber a quantidade de observações que existe numa determinada classe, mas sim a quantidade de observações acima ou abaixo de um determinado ponto na distribuição.

Deste modo, a soma das frequências de todos os valores abaixo do limite superior de uma determinada classe é definida como frequência acumulada para baixo deste ponto, assim como a soma das frequências de todos os valores acima do limite inferior de uma classe é denominada frequência acumulada para cima.

A título de ilustração, estão apresentadas nas Tabelas 2.5 e 2.6, respectivamente, as frequências acumuladas para cima e para baixo dos DAP dos 40 pés de *Eucalyptus citriodora* o em questão.

Aplicações das distribuições de frequências acumuladas

Para verificar qual a porcentagem de pés de *Eucalyptus citriodora* que possuem DAP inferior a 146 mm basta consultar diretamente a Tabela 2.5 e verificar a frequência acumulada abaixo deste valor (6,75%), pois o valor 146 mm é um dos limites de classe apresentados nesta tabela. Mas como proceder para obter as frequências acumuladas para valores intermediários aos apresentados na tabela? Como por exemplo a

Tabela 2.5: Distribuição de frequência acumulada para baixo dos DAP de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG.

Diâmetro (mm)	Frequência Acumulada	
	Absoluta	Relativa
Abaixo de 69,2	0	0,000
Abaixo de 94,8	3	0,075
Abaixo de 120,4	11	0,275
Abaixo de 146,0	27	0,675
Abaixo de 171,6	34	0,850
Abaixo de 197,2	38	0,950
Abaixo de 222,8	40	1,000

frequência acumulada abaixo de 150 mm?

Para este tipo de cálculo, pressupõe-se que os diâmetros estejam uniformemente distribuídos dentro das classes, e procede-se do seguinte modo:

Freq. acumulada abaixo, da classe imediatamente inferior a 150 (abaixo de 146)= 0,675;

Freq. acumulada abaixo, da classe imediatamente superior a 150 (abaixo de 171,6) = 0,850;

$$\text{Freq. abaixo de 146,0 mm} = 0,675$$

$$\text{Freq. abaixo de 171,6 mm} = 0,850$$

Assim,

$$\text{Freq. entre 146,0 e 171,6 mm} = 0,850 - 0,675 = 0,175$$

de 146,0 a 171,6 mm são 25,6 mm de 146,0 a 150,0 mm são 4,0 mm

então,

para uma diferença de 25,6 mm existem 0,175 dos DAP;

para uma diferença de 4,0 mm existirão x dos DAP;

ou seja

Tabela 2.6: Distribuição de frequência acumulada para cima dos DAP de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras - MG.

Diâmetro (mm)	Frequência Acumulada	
	Absoluta	Relativa
Acima de 69,2	40	1,000
Acima de 94,8	37	0,925
Acima de 120,4	29	0,725
Acima de 146,0	13	0,325
Acima de 171,6	6	0,150
Acima de 197,2	2	0,050
Acima de 222,8	0	0,000

$$25,6 \text{ mm} \rightarrow 0,175$$

$$4,0 \rightarrow x$$

de onde verifica-se que $x = 0,0273$.

Como abaixo de 140,0 mm existem 0,675 dos DAP, e entre 140,0 e 150 mm existem 0,0273, conclui-se que abaixo de 150 mm existam $0,675 + 0,0273 = 0,7023$ dos DAP.

Obs. Para variáveis qualitativas não se faz necessário a distribuição dos dados em classes pois cada “valor” da variável já representa uma classe distinta como pode ser observado na tabela 7, o mesmo ocorrendo com as variáveis quantitativas discretas (Exemplo 2).

Exemplo 2.: Realizou-se num grande hospital um estudo referente ao número de vitimas fatais de acidentes de transito, atendidas diariamente, durante um certo mês, obtendo os seguintes resultados:

0 2 1 5 3 2 1 2 2 3 0 3 0 0 1 2 3 5 1 0 0 2 0 4 0 3 4 0 2 1

Dispondo estes dados em um Rol tem-se:

0 0 0 0 0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 4 4 5 5

Tabela 2.7: Distribuição de Freqüências do Número de Funcionários da Empresa Tabajara Classificado Quanto ao Sexo em 1996.

SEXO	Fa	Fr
MASCULINO	20	0,40
FEMININO	30	0,60
TOTAL	50	1,00

Tabela 2.8: Número de vítimas fatais de acidentes de trânsito atendidas diariamente em um grande hospital, durante um certo mês

Vitimas fatais (X/dia)	Número de dias
0	9
1	5
2	7
3	5
4	2
5	2
Total	30

2.1.2 Apresentação gráfica

As mesmas informações fornecidas pelas distribuições de freqüências podem ser obtidas, e mais facilmente visualizadas através de gráficos, tais como histogramas, polígonos de freqüência, ogivas, gráficos de setores, pictogramas e outros.

2.1.2.1 Histogramas

Os histogramas são constituídos por um conjunto de retângulos, com as bases assentadas sobre um eixo horizontal, tendo o centro da mesma no ponto médio da classe que representa, e cuja altura é proporcional à freqüência da classe. Se as amplitudes

de classe forem todas iguais, as alturas serão numericamente iguais as frequências das classes. Porém, se os intervalos de classe não tiverem todos a mesma amplitude, as alturas dos retângulos deverão ser convenientemente ajustadas, afim de que as áreas dos mesmos sejam proporcionais às frequências das classes.

Figura 2.1: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

2.1.2.2 Polígonos de frequência

Polígono de frequência é um gráfico de análise no qual as frequências das classes são localizadas sobre perpendiculares levantadas nos ponto médios das classes. E pode ser obtido pela simples união dos pontos médios dos topos dos retângulos de um histograma. Completa-se o polígono unindo-se as extremidades da linha que une os pontos representativos das frequências de classe aos pontos médios das classes imediatamente anterior e posterior as classes extremas, que têm frequência nula.

Figura 2.2: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

Figura 2.3: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

2.1.2.3 Gráfico de setores (pizza)

O processo de construção de um gráfico de setores é simples pois sabe-se que setor de circunferência formado por um ângulo de 360° equivale a 100% da área da circunferência, assim para obter-se o setor cuja área representa uma determinada frequência, basta resolver uma regra de três simples, como a apresentada a seguir:

$$\begin{array}{rcl} 360^\circ & \rightarrow & 100\% \\ x^\circ & \rightarrow & Fr \end{array}$$

Figura 2.4: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

2.1.2.4 Ogivas

Ogiva é o nome dado a um polígono de frequências acumuladas, nas quais as frequências acumuladas são localizadas sobre perpendiculares levantadas nos limites inferiores ou superiores das classes, dependendo se a ogiva representar as frequências acumuladas abaixo ou acima, respectivamente.

Figura 2.5: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

Histogramas com classes de diferentes tamanhos

Como foi comentado, as áreas dos retângulos de um histograma são proporcionais à frequência da classe que representam. Deste modo, quando as classes apresentam

Figura 2.6: Diâmetro à altura do peito de 40 pés de *Eucalyptus citriodora* aos 6 anos de idade em Lavras-MG.

diferentes amplitudes, torna-se necessário fazer um ajuste na altura dos retângulos, afim de que suas áreas permaneçam fieis à sua frequência.

Para realizar esta correção utilizamos como altura dos retângulos a dfr (densidade de frequência relativa), dada por

$$dfr = \frac{Fr}{c} \quad (2.9)$$

em que,

Fr = frequência relativa da classe

c = amplitude de classe

Para ilustrar o uso da dfr na construção de histogramas utilizaremos os dados da tabela 2.9.

Tabela 2.9: Distribuição de Frequências da Áreas em ha de 900 Propriedades Rurais em uma Região de MG.

Área	Fa	Fr
0 † 50	500	0,5000
50 † 150	250	0,250
150 † 400	250	0,250
Total	1000	1,0000

Caso construíssemos um histograma utilizando as Fr obteríamos a figura 2.7. Como pode-se observar, na tabela 2.9 nas classes 50 † 150 e 150 † 400 existem o mesmo número de propriedades. Porém ao observar-se a figura 2.7, nota-se que área do histograma, que deveria representar a % de propriedades com área entre 150 † 400 ha,

é 2,5 vezes maior que a área referente à classe 50 † 150. Quando deveriam ser iguais. Construindo o histograma, tendo como coordenadas do eixo Y as dfr , apresentadas na tabela 2.10, esta distorção é corrigida, como pode ser verificado na figura 2.8.

Figura 2.7: Distribuição do Número de Propriedades Rurais Segundo suas Áreas, Num Certo Município de MG.

Tabela 2.10: Distribuição de Frequências do Áreas, em ha de 900 Propriedades Rurais em uma Região de MG.

Área	Fr	c	dfr
0 † 50	0,500	50	0,0100
50 † 150	0,250	100	0,0025
150 † 400	0,250	250	0,0010
Total	1,000		

Figura 2.8: Distribuição de Frequências do Áreas, em ha de 900 Propriedades Rurais em uma Região de MG.

Os histogramas e polígonos de frequências podem indicar ainda qual é o tipo de distribuição que os dados seguem, como pode ser visto a seguir:

Figura 2.9: Distribuição simétrica.

Figura 2.10: Distribuição assimétrica a esquerda.

Figura 2.11: Distribuição assimétrica a direita.

Figura 2.12: Distribuição em jota.

Figura 2.13: Distribuição em jota invertido.

Figura 2.14: Distribuição distribuição bimodal.

Figura 2.15: Distribuição multimodal.

2.2 Medidas de Posição e dispersão

2.2.1 Medidas de Posição

As medidas de posição ou de tendência central constituem uma forma mais sintética de apresentar os resultados contidos nos dados observados, pois representam um valor central, em torno do qual os dados se concentram. As medidas de posição mais empregadas são a média, a mediana e a moda.

2.2.1.1 Média

É a mais usada das três medidas de posição mencionadas, por ser a mais comum e compreensível delas, bem como pela relativa simplicidade do seu cálculo, além de prestar-se bem ao tratamento algébrico.

- Média Aritmética:

A média aritmética ou simplesmente média de um conjunto de n observações, x_1, x_2, \dots, x_n é definida como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.10)$$

Ex.: Dados os pesos ao nascer (kg) de cinco leitões da raça Large White; 1,350; 1,500; 1,800; 1,800; 1,900.

Temos que o peso médio ao nascer destes leitões é

$$\bar{x} = \frac{1,350 + 1,500 + 1,800 + 1,800 + 1,900}{5} = 1,670kg$$

Interpretação: O peso médio ao nascer dos cinco leitões foi de 1,670kg, isto quer dizer que alguns leitões pesaram menos de 1,670kg, outros pesaram mais, mas em média, o peso ao nascer dos leitões foi 1,670kg. Ou seja 1,670kg é um valor em torno do qual os pesos ao nascer destes cinco leitões se concentra.

Se os dados estiverem agrupados na forma de uma distribuição de frequência, lança-se mão da Hipótese tabular básica, para o cálculo da média.

Hipótese Tabular Básica

“Todas as observações contidas numa classe são consideradas iguais ao ponto médio da classe.”

Que é realizado através da seguinte expressão:

$$\bar{x} = \frac{\sum_{i=1}^k x_i F a_i}{\sum_{i=1}^k F a_i} = \sum_{i=1}^k x_i F r_i \quad (2.11)$$

em que,

x_i = ponto médio da classe i;

Fa_i = frequência absoluta da classe i;

Fr_i = frequência relativa da classe i.

Ex.: Calcular a média dos dados apresentados a seguir:

Tabela 2.11: Durabilidade (horas) de 400 válvulas ensaiadas na Tabajara Inc.

Duração (horas)	Fa	Fr
300 † 400	14	0,035
400 † 500	46	0,115
500 † 600	58	0,145
600 † 700	76	0,190
700 † 800	68	0,170
800 † 900	62	0,155
900 † 1000	48	0,120
1000 † 1100	22	0,055
1100 † 1200	6	0,015
TOTAL	400	1,000

$$\bar{x} = \frac{(350)(14) + (450)(46) + \dots + (1150)(6)}{14 + 46 + \dots + 6} = \frac{286200}{400} = 715,5 \text{ horas}$$

Propriedades da Média

- i. A soma dos desvios de um conjunto de dados em relação a sua média é nula

Ex.: Dados ,1,2,3;

$$\bar{x} = 2$$

$$(1 - 2) + (2 - 2) + (3 - 2) = 0$$

Prova:

$$\begin{aligned}
 \sum_{i=1}^n [x_i - \bar{x}] &= \sum_{i=1}^n x_i - n\bar{x} \\
 &= \sum_{i=1}^n x_i - n \frac{\sum_{i=1}^n x_i}{n} \\
 &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i \\
 &= 0
 \end{aligned}$$

- ii. A soma dos quadrados dos desvios de um conjunto de dados em relação a uma constante k é mínima quando k for a média.

Ex.: Dados 1,2,3, $\bar{x} = 2$

k	Soma dos quadrados dos desvios
1,0	5,00
1,5	2,75
2,0	2,00
2,5	2,75
3,0	5,00

Prova:

Seja

$$S = \sum_{i=1}^n [x_i - \theta]^2,$$

Então o valor de θ que minimiza S é obtido solucionando-se o sistema:

$$\frac{dS}{d\theta} = 0$$

Assim

$$\begin{aligned}
2 \sum_{i=1}^n [x_i - \theta] &= 0 \\
\sum_{i=1}^n x_i - n\theta &= 0 \\
\theta &= \frac{\sum_{i=1}^n x_i}{n} \\
\theta &= \bar{x}
\end{aligned}$$

iii. A média de um conjunto de dados acrescidos ou subtraídos em cada elemento de uma constante k é igual á média original somada ou subtraída desta constante.

Ex.: Dados 1, 2, 3 $\bar{x} = 2$

$k=2$ novos dados: 3, 4, 5 $\bar{x}^* = 4 = 2 + 2 = \bar{x} + k$

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

fazendo $x_i^* = (x_i \pm k)$

tem-se:

$$\begin{aligned}
\bar{x}^* &= \frac{\sum_{i=1}^n x_i^*}{n} \\
&= \frac{\sum_{i=1}^n (x_i \pm k)}{n} \\
&= \frac{\sum_{i=1}^n x_i \pm nk}{n} \\
&= \frac{\sum_{i=1}^n x_i}{n} \pm k \\
\bar{x}^* &= \bar{x} \pm k
\end{aligned}$$

iv. Multiplicando-se todos os dados por uma constante k , a nova média fica multiplicada por k .

Ex.: Dados: 1, 2, 3 $\bar{x} = 2$

$k=3$ novos dados: 3, 6, 9 $\bar{x}^* = 6 = (3)(2) = k\bar{x}$

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

fazendo $x_i^* = (kx_i)$

tem-se:

$$\begin{aligned}\bar{x}^* &= \frac{\sum_{i=1}^n x_i^*}{n} \\ &= \frac{\sum_{i=1}^n (kx_i)}{n} \\ &= \frac{k \sum_{i=1}^n x_i}{n} \\ \bar{x}^* &= k\bar{x}\end{aligned}$$

Características e importância:

- i. É muito influenciada pelos valores extremos da distribuição;
- ii. Localiza-se, em geral, na classe de maior frequência;
- iii. Na sua determinação são considerados todos os dados da distribuição;
- iv. A sua precisão está na razão direta do número de observações com que é calculada;
- v. É única para um conjunto de dados.
- vi. Não pode ser calculada para dados agrupados que apresentam classes extremas abertas.

Ex.:

Classe	Fa
0 a 500	5
mais de 500	6

- Média Ponderada

Às vezes associa-se às observações x_1, x_2, \dots, x_n determinadas ponderações ou pesos w_1, w_2, \dots, w_n que dependem da importância atribuída a cada uma das observações, neste caso a média é dada por:

$$\bar{x}_p = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} \quad (2.12)$$

Ex.: Se o exame final de um curso tem peso 3, e as provas correntes peso 1. Qual a nota média de um aluno que obteve 85 no exame final e 70,90 nas provas correntes?

Aplicando-se a equação 2.12 tem-se:

$$\bar{x}_p = \frac{(3)(85) + (1)(70) + (1)(90)}{3 + 1 + 1} = \frac{415}{5} = 85 \text{ pontos}$$

- Média Geométrica

A média geométrica de um conjunto de n observações, x_1, x_2, \dots, x_n , é dada pela raiz de ordem n do produto dessas observações, ou seja:

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad (2.13)$$

ou ainda:

$$\ln \bar{x}_G = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (2.14)$$

A média geométrica é utilizada para representar variáveis assimétricas a direita, pois, nestes casos, média aritmética, por ser muito influenciada pelos valores extremos, não representa bem a variável. Como exemplos de variáveis, para as quais a média geométrica é um melhor localizador do que a média aritmética pode situar-se a distribuição de renda da população brasileira, a condutividade hidráulica de um solo e o diâmetro de torrões de solo.

- Média Harmônica

A média harmônica de um conjunto de n observações, x_1, x_2, \dots, x_n , é a recíproca da média aritmética dos recíprocos das observações:

$$\bar{x}_H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (2.15)$$

Este tipo de média é utilizado para variáveis que apresentem periodicidade, ou seja uma variação harmônica, como por exemplo ondas de rádio, variação de preços de produtos agrícolas no decorrer do ano (safra/entre safra), sinais de TV, etc.

2.2.1.2 Mediana

Para um conjunto de dados ordenados (Rol) a mediana é o valor que é precedido e seguido pelo mesmo número de dados (observações). Isto é 50% dos dados são superiores à mediana e 50% são inferiores.

Cálculo da mediana

- i. Quando o número de dados (n) for ímpar, a mediana é dada por:

$$Md = x_{\left(\frac{n+1}{2}\right)} \quad (2.16)$$

Ex.: 0, 1, 2, 3, 4 $\Rightarrow n = 5$ (*ímpar*)

$$Md = x_{\left(\frac{5+1}{2}\right)} = x_{(3)} = 2$$

- ii. Quando o número de dados for par, a mediana será dada por:

$$Md = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2} \quad (2.17)$$

Ex.: 0, 1, 2, 3 $\Rightarrow n = 4$ (*par*)

$$Md = \frac{x_{\left(\frac{4}{2}\right)} + x_{\left(\frac{4+2}{2}\right)}}{2} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{1+2}{2} = 1,5$$

- iii. Dados agrupados:

$$Md = L_i + \left[\frac{\frac{n}{2} - F_a}{F_{Md}} + \right] c, \quad (2.18)$$

em que,

L_i = e o limite inferior da classe mediana;

F_a = é a frequência acumulada das classes anteriores a classe mediana;

F_{Md} é a frequência da classe mediana; e

c é amplitude da classe mediana.

Ex.: Para os dados da Tabela 2.11 (Durabilidade das válvulas) temos:

$n = 400$ observações. o valor da mediana encontra-se entre a posição 200 e 201 $x_{\frac{n}{2}}$ e $x_{\frac{n+2}{2}}$, que pertencem à 5ª classe $[700 - 800[$ Aplicando-se a fórmula da mediana vem:

$$\begin{aligned} Md &= L_i + \left[\frac{\frac{n}{2} - F_a}{F_{Md}} + \right] c \\ &= 700 + \left[\frac{\frac{400}{2} - 194}{68} + \right] 100 \\ &= 708,82 \text{ horas} \end{aligned}$$

Interpretação: A mediana igual a 708,82 horas indica que 50% das válvulas duram menos que 708,82 horas e 50% duram mais que 708,82 horas.

Propriedades da Mediana:

- i. A soma dos módulos dos desvios dos dados em relação à mediana é mínima.

$$\sum_{i=1}^n |x_i - Md| = \text{mínimo}$$

- ii. Somando-se ou subtraindo-se uma constante (k) a todas as observações, a mediana fica somada ou subtraída desta constante (k).

$$x^* = X \pm k \Rightarrow Md^* = Md \pm k$$

- iii. Multiplicando-se todas as observações por uma constante (k), a mediana fica multiplicada por esta constante (k).

$$x^* = kx \Rightarrow Md^* = kMd$$

Características e Importância:

- i. Pode ser obtida em distribuições de frequências que apresentem classes com limites indefinidos;
- ii. É muito empregada em pesquisas nas quais os valores extremos têm pouca importância;

- iii. Não é influenciada por valores extremos e sim pelo número de observações;
- iv. É mais realista do que a média para representar certas variáveis, como o nível salarial de uma empresa.

2.2.1.3 Moda

A moda de um conjunto de dados é o valor que ocorre com maior frequência, isto é, o valor mais comum. Para um conjunto de dados a moda pode não ser única, bem como pode não existir.

Ex.:

2, 3, 4, 5, 7, 7, 7, 8, 9 $Mo = 7$;

1, 2, 3, 4, 7, 9, 10, 13, 20 não possui moda;

1, 2, 3, 4, 4, 8, 10, 10 13 $Mo = 4$ e $Mo = 10$.

Dados Agrupados

Quando os dados estão agrupados, na forma de uma distribuição de frequências, a moda é o ponto do eixo x, correspondente à ordenada máxima da distribuição. O processo para cálculo da moda em dados agrupados é o geométrico, a partir do histograma de frequências (Método de Czuber). Este método é baseado na influência que as classes adjacentes exercem sobre a moda, deslocando-a no sentido da classe de maior frequência.

Figura 2.16: Esquema para obtenção da moda pelo método de Czuber

No histograma acima, marca-se, na classe modal, os vértices A, B, C e D. Traça-se as retas \overline{AC} e \overline{BD} . No ponto de intersecção destas retas (E) traça-se uma perpendicular ao eixo das classes, localizando o ponto Mo , valor da moda. O ponto Mo divide o intervalo da classe modal (c) em duas partes, cujos comprimentos são proporcionais a

Δ_1 e Δ_2 . Sendo Δ_1 a diferença entre a frequência da classe modal e da classe imediatamente anterior, e Δ_2 a diferença entre as frequências da classe modal e da imediatamente posterior.

Por E traça-se a reta $\overline{FF'}$, paralela ao eixo das classes, obtendo assim, os segmentos \overline{EF} e $\overline{EF'}$, que representam as alturas dos triângulos ABE e CDE .

Sendo Li o limite inferior da classe modal, Ls o limite superior e x a distância entre Li e a moda (Mo), verificasse na figura 2.16 que:

$$Mo = Li + x \quad (2.19)$$

Sendo os triângulos ABE e CDE semelhantes (pois possuem dois ângulos iguais) tem-se que:

$$\begin{aligned} \frac{\overline{EF}}{\overline{EF'}} &= \frac{\overline{AB}}{\overline{CD}} \\ \frac{x}{c-x} &= \frac{\Delta_1}{\Delta_2} \\ x\Delta_2 &= c\Delta_1 - x\Delta_1 \\ x &= \frac{\Delta_1}{\Delta_1 + \Delta_2}c \end{aligned} \quad (2.20)$$

Substituindo 2.20 em 2.19 tem-se:

$$Mo = Li + \frac{\Delta_1}{\Delta_1 + \Delta_2}c, \quad (2.21)$$

em que:

L_i é o limite inferior da classe modal;

Δ_1 é a diferença entre a frequência da classe modal e da imediatamente anterior;

Δ_2 é a diferença entre a frequência da classe modal e da imediatamente anterior;

c é a amplitude da classe modal.

Características e Importância

- i. Não é afetada por valores extremos, a não ser que estes constituam a classe modal;
- ii. É uma medida bastante utilizada em Estatística Econômica;
- iii. Não apresenta boas propriedades algébricas; d) Maximiza o número de desvios iguais a zero.

Propriedades da Moda

- i. Somando-se ou subtraindo uma constante a todos os dados, a moda fica somada ou subtraída da mesma constante.

$$x^* = x \pm k \Rightarrow Mo^* = Mo \pm k$$

- ii. Multiplicando-se todos os dados por uma constante k , a moda fica multiplicada por esta constante.

$$x^* = kx \Rightarrow Mo^* = kMo$$

2.2.1.4 Separatrizes (Quantis)

Quartis

Os quartis separam um conjunto de dados ordenados (Rol) em quatro partes iguais.

Assim:

Q_1 é o 1º quartil, deixa 25% dos elementos abaixo dele;

$Q_2 = 2^\circ$ quartil, coincide com a mediana, deixa 50% dos elementos abaixo dele;

$Q_3 = 3^\circ$ quartil, deixa 75% dos elementos abaixo dele.

Determinação de Q_1

$$Q_1 = Li_{Q_1} + \left[\frac{\frac{n}{4} - Fa_{Q_1}}{F_{Q_1}} \right] c \quad (2.22)$$

em que:

Li_{Q_1} é o limite inferior da classe que contém Q_1 ;

Fa_{Q_1} é a frequência acumulada das classes anteriores à classe que contém Q_1 ;

F_{Q_1} é a frequência da classe que contém Q_1 e

c é a amplitude da classe que contém Q_1 .

Determinação de Q_3

$$Q_3 = Li_{Q_3} + \left[\frac{\frac{3n}{4} - Fa_{Q_3}}{F_{Q_3}} \right] c \quad (2.23)$$

em que:

Li_{Q_3} é o limite inferior da classe que contém Q_3 ;

Fa_{Q_3} é a frequência acumulada das classes anteriores à classe que contém Q_3 ;

F_{Q_3} é a frequência da classe que contém Q_3 e

c é a amplitude da classe que contém Q_3 .

Decis

São valores que dividem uma série de dados ordenados em dez partes iguais. O i - *ésimo* decil, ($i = 1, 2, \dots, 10$), de um conjunto de observações organizadas na forma de uma distribuição de frequências pode ser obtido por:

$$D_i = Li_{D_i} + \left[\frac{\frac{in}{10} - Fa_{D_i}}{F_{D_i}} \right] c \quad (2.24)$$

em que:

Li_{D_i} é o limite inferior da classe que contém D_i ;

Fa_{D_i} é a frequência acumulada das classes anteriores à classe que contém D_i ;

F_{D_i} é a frequência da classe que contém D_i e

c é a amplitude da classe que contém D_i .

Percentis

São valores que dividem uma série de dados ordenados em 100 partes iguais. Dada uma distribuição de frequências, o valor do i – *ésimo* percentil, ($i = 1, 2, \dots, 10$) é obtido por:

$$P_i = Li_{P_i} + \left[\frac{\frac{in}{100} - Fa_{P_i}}{F_{P_i}} \right] c \quad (2.25)$$

em que:

Li_{P_i} é o limite inferior da classe que contém P_i ;

Fa_{P_i} é a frequência acumulada das classes anteriores à classe que contém P_i ;

F_{P_i} é a frequência da classe que contém P_i e

c é a amplitude da classe que contém P_i .

Relações empíricas entre média, mediana, moda e as distribuições de dados:

Distribuição	Relação
Simétrica	$\bar{x} = Md = Mo$
Assimétrica a direita (assimétrica positiva)	$\bar{x} > Md > Mo$
Assimétrica a esquerda (assimétrica negativa)	$\bar{x} < Md < Mo$

2.2.2 Medidas de dispersão

A utilização de uma medida de posição para substituir um conjunto de dados é insuficiente para sintetizar a informação nele contida, como pode ser observado a seguir:

$$A = 10, 10, 10, 10, 10, 10, 10$$

$$B = 1, 8, 10, 10, 11, 12, 18$$

$$C = 1, 2, 10, 10, 10, 13, 24$$

Calculando a média (eq 2.10), mediana (eq 2.16 e moda desses três conjuntos tem-se:

$$\bar{x}_A = \bar{x}_B = \bar{x}_C = 10 \text{ unidades}$$

$$Md_A = Md_B = Md_C = 10 \text{ unidades}$$

$$Mo_A = Mo_B = Mo_C = 10 \text{ unidades}$$

Assim, verifica-se que os três conjuntos (A,B,C) apresentam médias, medianas e modas iguais a 10 *unidades*, porém observando-os, percebe-se que eles são bem diferentes entre si, pois enquanto no conjunto A os dados são todos iguais, os demais apresentam uma certa variação, sendo que esta variação é maior no conjunto C. Deste modo, para sintetizarmos eficientemente a informação de um conjunto de dados temos que associar à medida de posição utilizada, uma medida de dispersão, que vai informar como estes dados se comportam em torno da medida de posição em questão.

2.2.2.1 Amplitude Total (A)

A amplitude total é a diferença entre o maior e o menor valor observado

$$A = MVO - mvo, \quad (2.26)$$

em que:

MVO é o maior valor observado, e

mvo é o menor valor observado.

Para os conjuntos A,B e C tem-se:

$$A_A = 10 - 10 = 0 \text{ unidades}$$

$$A_B = 18 - 1 = 17 \text{ unidades e}$$

$$A_C = 24 - 1 = 23 \text{ unidades.}$$

Nota-se, então, que a amplitude do conjunto *C* é bem maior que nos demais. A amplitude é uma medida de dispersão fácil de ser calculada e é certamente a maneira mais natural e comumente utilizada para descrever a variabilidade de um conjunto de dados. Porém sua interpretação depende do número de observações, mas, no seu cálculo não são consideradas todas as observações, pois só utiliza os valores extremos.

2.2.2.2 Variância e desvio padrão

Uma boa medida de dispersão deve basear-se em todos os dados, ser facilmente calculável e compreensível, além de prestar-se bem ao tratamento algébrico. Uma medida com todas estas características é obtida considerando-se os desvios de cada observação em relação a média (chamados erros) :

$$e_i = x_i - \bar{x} \quad (2.27)$$

Para obter um único número que represente a dispersão dos dados, pensou-se inicialmente em obter-se a média destes desvios, mas deve-se lembrar que a soma dos desvios de um conjunto de dados em relação a sua média é nula. Então, optou-se por utilizar a soma dos quadrados dos desvios, pois elevando-se cada desvio ao quadrado elimina-se o sinal negativo, que estava trazendo complicações, e dividindo-se a soma dos quadrados dos desvios pelo número de observações obtém-se a variância populacional que é uma medida quantitativa da dispersão de um conjunto de dados entorno da sua média, além do fato, de esta soma de quadrados de desvios ser mínima, como já foi visto em propriedades da média.

$$V(x) = \sigma^2 = \frac{SQD}{N} = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.28)$$

Para os exemplos anteriores tem-se:

$$\begin{aligned} \sigma_A^2 &= \frac{(10 - 10)^2 + (10 - 10)^2 + \dots + (10 - 10)^2}{7} = 0 \text{ unidades}^2 \\ \sigma_B^2 &= \frac{(1 - 10)^2 + (8 - 10)^2 + \dots + (18 - 10)^2}{7} = 22 \text{ unidades}^2 \\ \sigma_A^2 &= \frac{(1 - 10)^2 + (2 - 10)^2 + \dots + (24 - 10)^2}{7} = 50 \text{ unidades}^2 \end{aligned}$$

Obs. Quando estiver trabalhando com amostras, a variância é dada pela soma dos quadrados dos desvios dividida por $n - 1$ (número de observações menos um) que é denominado graus de liberdade. Assim:

$$s^2 = \frac{SQD}{n - 1} = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.29)$$

Formulas computacionais para o cálculo da variância

$$\sigma^2 = \frac{1}{N} \left[\sum_{i=1}^n x^2 - \frac{(\sum_{i=1}^n x)^2}{N} \right] \quad (2.30)$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x^2 - \frac{(\sum_{i=1}^n x)^2}{n} \right] \quad (2.31)$$

prova:

$$\begin{aligned} SQD &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2 \frac{\sum_{i=1}^n x_i}{n} \sum_{i=1}^n x_i + n \left[\frac{\sum_{i=1}^n x_i}{n} \right]^2 \\ &= \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \end{aligned}$$

Para dados tabulados o cálculo da variância é realizado por meio das seguintes expressões:

$$\sigma^2 = \frac{1}{N} \left[\sum_{i=1}^n x_i^2 F_i - \frac{(\sum_{i=1}^n x_i F_i)^2}{N} \right] \quad (2.32)$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 F_i - \frac{(\sum_{i=1}^n x_i F_i)^2}{n} \right] \quad (2.33)$$

em que:

x_i é o ponto médio da i -ésima classe, e

F_i é a frequência absoluta da i -ésima classe.

Propriedades da variância

i. A variância de uma constante k é nula

$$V(k) = 0, k = \text{constante}.$$

prova

Dado que:

$$x_1 = x_2 = \dots = x_n = k \Rightarrow \bar{x} = k, \text{ portanto:}$$

$$\begin{aligned} V(x) &= \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{N} \sum_{i=1}^n (k - k)^2 \\ &= 0 \end{aligned}$$

ii. A variância de uma soma ou diferença entre variáveis é a soma das variâncias das variáveis se estas forem independentes.

$$V(X \pm Y) = V(X) + V(Y) \text{ se } X \text{ e } Y \text{ forem independentes}$$

Prova: $w = x \pm y$

$$\begin{aligned} V(W) &= \frac{1}{N} \sum_{i=1}^n (w_i - \bar{W})^2 \\ &= \frac{1}{N} \sum_{i=1}^n [(x_i - y_i) - (\bar{X} - \bar{Y})]^2 \\ &= \frac{1}{N} \sum_{i=1}^n [(x_i - y_i)^2 - 2(x_i - y_i)(\bar{X} - \bar{Y}) + (\bar{X} - \bar{Y})^2]^2 \\ &= \frac{1}{N} \sum_{i=1}^n [x_i^2 - 2x_i y_i + y_i^2 - 2(x_i \bar{X} - x_i \bar{Y} - y_i \bar{X} + y_i \bar{Y}) + (\bar{X}^2 - 2\bar{X}\bar{Y} + \bar{Y}^2)]^2 \\ &= \frac{1}{N} \sum_{i=1}^n [(x_i^2 - 2x_i \bar{x} + \bar{x}^2) + (y_i^2 - 2y_i \bar{y} + \bar{y}^2) - 2(x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y})]^2 \\ &= \frac{1}{N} \sum_{i=1}^n [(x_i - \bar{X})^2 + (y_i - \bar{Y})^2 - 2(x_i - \bar{X})(y_i - \bar{Y})]^2 \\ &= V(X) + V(Y) - 2cov(X, Y) \end{aligned}$$

para $W = X + Y$ tem-se: $V(W) = V(X) + V(Y) + 2cov(X, Y)$

- iii. Somando-se ou subtraindo-se uma constante k a todos os dados a variância não se altera.

$$X^* = X \pm k \Rightarrow V(X^*) = V(X)$$

Prova:

$$\begin{aligned} V(X^*) &= \frac{1}{N} \sum_{i=1}^n (x_i^* - \bar{X}^*)^2 \\ &= \frac{1}{N} \sum_{i=1}^n [(x_i \pm k) - \bar{X} \pm k]^2 \\ &= \frac{1}{N} \sum_{i=1}^n (x_i - \bar{X})^2 \\ &= V(X) \end{aligned}$$

- iv. Multiplicando-se todos os dados por uma constante k , a variância fica multiplicada por k^2

$$X^* = X.k \Rightarrow V(X^*) = k^2V(X)$$

Prova:

$$\begin{aligned} V(X^*) &= \frac{1}{N} \sum_{i=1}^n (x_i^* - \bar{X}^*)^2 \\ &= \frac{1}{N} \sum_{i=1}^n (kx_i - k\bar{x})^2 \\ &= \frac{1}{N} \sum_{i=1}^n (k^2x_i^2 - 2k^2x_i\bar{x} + k^2\bar{x}) \\ &= \frac{1}{N} \sum_{i=1}^n k^2(x_i^2 - 2x_i\bar{x} + \bar{x}) \\ &= k^2V(X) \end{aligned}$$

Desvio Padrão

Um inconveniente da variância é que ela é expressa em unidades ao quadrado, ou seja, caso esteja-se trabalhando com o peso corporal de indivíduos, tomados em kg, a variância destes pesos é expressa em kg^2 , o que causa algumas dificuldades de interpretação. No intuito de resolver este problema trabalha-se com o desvio padrão que é definido como a raiz quadrada positiva da variância, o qual é expresso na mesma unidade em que os dados foram coletados.

$$\sigma = \sqrt{\sigma^2} \quad (2.34)$$

$$s = \sqrt{s^2} \quad (2.35)$$

Para o exemplo em questão tem-se:

$$\sigma_A = \sqrt{0} = 0 \text{ unidades}$$

$$\sigma_B = \sqrt{22} = 4,6 \text{ unidades}$$

$$\sigma_A = \sqrt{50} = 7,07 \text{ unidades}$$

Propriedades do desvio padrão

- i. Somando-se ou subtraindo-se uma constante k a todos os dados o desvio padrão não se altera.

$$X^* = X \pm k \Rightarrow s(X^*) = s(X)$$

- ii. Multiplicando-se todos os dados por uma constante k , o desvio padrão fica multiplicada por k

$$X^* = kX \Rightarrow s(X^*) = ks(X)$$

2.2.2.3 Coeficiente de variação

A variância e o desvio padrão são medidas de dispersão absolutas, deste modo só podem ser utilizadas para comparar a variabilidade de dois ou mais conjuntos de dados quando estes apresentarem a mesma média, mesmo número de observações e estiverem expressos nas mesmas unidades. Então para comparar qualquer conjunto de dados quanto à sua variabilidade deve-se lançar mão de uma medida de dispersão relativa como o coeficiente de variação (CV), que expressa a variabilidade dos dados em relação à sua média.

$$CV = \frac{\sigma}{\bar{X}} 100 \quad (2.36)$$

Ex.1:

$$\begin{aligned} CV_A &= \frac{0}{10} 100 = 0\% \\ CV_B &= \frac{4,69}{10} 100 = 46,9\% \\ CV_A &= \frac{7,07}{10} 100 = 70,7\% \end{aligned}$$

Ex.2.

Estatística	Temperatura ($^{\circ}C$)	Precipitação (mm)
Média	22	800
s	5	100
$CV\%$	22,7	12,5

Os dados acima referem-se à temperatura ambiental e precipitação de uma certa região. Caso deseje-se saber qual das duas variáveis possui maior dispersão, e para tanto utilizar-se o desvio padrão (S), concluiria-se erroneamente que a precipitação é

mais variável que a temperatura, além de cometer o disparate de comparar numericamente duas variáveis que estão expressas em unidades diferentes.

2.2.2.4 Erro padrão da média

É uma medida que fornece uma idéia da precisão com que a média foi estimada.

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} \quad (2.37)$$

2.2.2.5 Momentos, Assimetria e Curtose

Momentos

Se x_1, x_2, \dots, x_n são os n valores assumidos pela variável X , define-se a quantidade

$$\bar{X}^r = \frac{x_1^r + x_2^r + \dots + x_n^r}{n} = \frac{\sum_{i=1}^n x_i^r}{n}, \quad (2.38)$$

como o momento de ordem r em relação a origem. Nota-se que o primeiro momento em relação a origem (\bar{X}^1) é a média de X .

O momento de ordem r em relação a uma origem k , qualquer, é dado por:

$$M'_r(k) = \frac{\sum_{i=1}^n (x_i - k)^r}{n}, \quad (2.39)$$

O momento de ordem r em relação a média \bar{X} é dado por:

$$M'_r(\bar{X}) = \frac{\sum_{i=1}^n (x_i - \bar{X})^r}{n}, \quad (2.40)$$

Nota-se que o segundo momento em relação a média é a variância.

Para o caso dos dados encontrarem-se agrupados, na forma de uma distribuição de freqüências, as expressões para o cálculo dos momentos serão:

$$M'_r(k) = \frac{\sum_{i=1}^n (x_i - k)^r F_i}{\sum_{i=1}^n F_i}, \quad (2.41)$$

em que:

x_i é o ponto médio da i -ésima classe, e

F_i = frequência absoluta da i -ésima.

Coeficiente de Assimetria (Cs).

Assimetria é o grau de desvio, ou afastamento da simetria, de uma distribuição. Se a curva de frequência (polígono de frequência suavizado) de uma distribuição tem uma “cauda” mais longa à direita da ordenada máxima do que à esquerda, diz-se que a distribuição é assimétrica à direita ou assimétrica positiva. Se o inverso ocorre, diz-se que ela é assimétrica à esquerda ou negativa.

O coeficiente de assimetria (Cs) é dado por:

$$Cs = \frac{M'_3}{(\sigma^2)^{1,5}} \quad (2.42)$$

Classificação das distribuições quanto a assimetria:

$Cs = 0$ distribuição é simétrica perfeita.

$Cs > 0$ a distribuição é assimétrica à direita.

$Cs < 0$ a distribuição é assimétrica à esquerda.

Existem ainda o primeiro e segundo coeficientes de assimetria de Pearson dados respectivamente por:

$$Cs = \frac{\bar{X} - Mo}{s} \quad (2.43)$$

e

$$Cs = \frac{3(\bar{X} - Md)}{s} \quad (2.44)$$

Tipos de distribuições quanto a assimetria

Figura 2.17: Distribuição simétrica.

Coeficiente de Curtose.

Figura 2.18: Distribuição assimétrica a esquerda.

Figura 2.19: Distribuição assimétrica a direita.

Curtose é o grau de achatamento de uma distribuição, considerado usualmente em relação a distribuição normal. A distribuição que tem um pico relativamente alto é chamada leptocúrtica, enquanto a distribuição que possui o topo achatado é denominada platicúrtica e a distribuição que não é muito pontiaguda, nem muito achatada, como acontece com a distribuição normal é denominada mesocúrtica. O coeficiente de curtose é dado por:

$$Ck = \frac{M'_4}{(\sigma^2)^2} \quad (2.45)$$

Tipos de distribuição quanto à curtose:

Figura 2.20: Distribuição leptocúrtica.

Figura 2.21: Distribuição mesocúrtica.

Figura 2.22: Distribuição platicúrtica.

Ex.: Os dados a seguir referem-se ao número de partos/dia ocorrido num determinado hospital durante o mês de março de 1997

X (partos/dia)	F (número de dias)
12	1
14	4
16	6
18	10
20	7
22	2
Total	30

a) Obter o primeiro momento em relação a origem

$$M'_1 = \frac{\sum_{i=1}^n (x_i) F_i}{\sum_{i=1}^n F_i} = \frac{12.1 + 14.4 + \dots + 22.2}{30} = 17,6$$

b) segundo, terceiro e quarto momentos em relação a média

$$\begin{aligned} M'_2(\bar{X}) &= \frac{\sum_{i=1}^n (x_i - \bar{X})^2 F_i}{\sum_{i=1}^n F_i} \\ &= \frac{(12 - 17,6)^2 1 + (14 - 17,6)^2 4 + \dots + (22 - 17,6)^2 2}{30} = 5,97 \end{aligned}$$

$$\begin{aligned} M'_3(\bar{X}) &= \frac{\sum_{i=1}^n (x_i - \bar{X})^3 F_i}{\sum_{i=1}^n F_i} \\ &= \frac{(12 - 17,6)^3 1 + (14 - 17,6)^3 4 + \dots + (22 - 17,6)^3 2}{30} = -3,96 \end{aligned}$$

$$\begin{aligned} M'_4(\bar{X}) &= \frac{\sum_{i=1}^n (x_i - \bar{X})^4 F_i}{\sum_{i=1}^n F_i} \\ &= \frac{(12 - 17,6)^4 1 + (14 - 17,6)^4 4 + \dots + (22 - 17,6)^4 2}{30} = 89,22 \end{aligned}$$

c) Obter os coeficientes de assimetria e curtose

$$C_s = \frac{M'_3}{(\sigma^2)^{1,5}} = \frac{-3,96}{5,97^{1,5}} = -0,27$$

$$C_k = \frac{M'_4}{(\sigma^2)^2} = \frac{89,22}{5,97^2} = 2,50$$

Capítulo 3

Probabilidades

3.1 Processo aleatório:

Qualquer fenômeno que gere resultado incerto ou casual.

Ex.:

- Jogar uma moeda e observar a sua face superior;
- Sexo do primeiro filho de um casal;
- Número de chips defeituosos encontrados num lote de 100 chips;
- Peso de uma pessoa.

Características de um processo (experimento) aleatório

- Cada experimento pode ser repetido indefinidamente sob as mesmas condições;
- Não se conhece a priori o resultado do experimento, mas pode-se descrever todos os possíveis resultados;
- Quando o experimento for repetido um grande número de vezes, surgirá uma regularidade do resultado, isto é, haverá uma estabilidade da fração $\frac{r}{n}$ (frequência relativa) da ocorrência de um particular resultado.

Figura 3.1: Freqüência relativa de sucessos em função do número de repetições (r) para um experimento com $p = 0,5$

3.2 Espaço amostral e Eventos

O conjunto formado por todos os possíveis resultados de um processo aleatório é denominado **espaço amostral** (Ω).

Ex1.: Processo aleatório: Conformação dos dos filhos de um casal com três filhos quanto ao sexo,

$$\Omega = MMM, MMF, MFM, FMM, MFF, FMF, FFM, FFF,$$

em que , F = sexo feminino e M = sexo masculino.

Ex2.: Processo aleatório: Verificar a idade de uma pessoa,

$$\Omega = x \in \mathbb{R} : 0 \leq x \leq 120$$

Ex3.: Processo aleatório: Verificar a cor das flores de uma planta fe feijoeiro,

$$\Omega = \{branca, roxa, amarela\}.$$

Qualquer sub-conjunto do espaço amostral (Ω) é denominado **evento**. Como exemplo tem-se:

$$E_1: \text{Três filhos do sexo feminino } E_1 = \{FFF\}.$$

$$E_2: \text{Dois filhos do sexo feminino } E_2 = \{FFM, FMF, MFF\}.$$

$$E_3: \text{Uma pessoa ter entre 20 e 25 anos } E_3 = \{20 \leq x \leq 25\}.$$

Eventos especiais:

Evento certo $E_4 = \Omega$;

Evento impossível ou vazio $E_5 = \phi$.

Eventos mutuamente exclusivos.

Dois eventos, A e B , são mutuamente exclusivos se eles não puderem ocorrer simultaneamente.

Ex.: A = primeiro filho ser do sexo feminino.

B = primeiro filho ser do sexo masculino.

3.3 Probabilidade

A chamada definição clássica de probabilidade é: Dado um conjunto de N eventos equiprováveis, a probabilidade de ocorrência de um determinado evento A , é dada pela razão:

$$P(A) = \frac{n}{N},$$

em que:

n é o número de eventos de interesse, e

N o número total de eventos.

Exemplos:

$$P(E_1) = \frac{1}{8} = 0,125$$

$$P(E_1) = \frac{3}{8} = 0,375$$

Deste modo, a probabilidade pode ser vista como uma medida da possibilidade de ocorrência de um particular evento. Um problema da definição clássica de probabilidade reside no fato de em sua definição lançar-se mão do conceito de eventos equiprováveis, e portanto da expressão a qual deseja-se definir.

Modernamente adota-se uma definição axiomática, dada pelos seguintes axiomas, atribuídos à Kolmogorov:

Axiomas de Kolmogorov.

- $A1 : P(A) \geq 0$;
- $A2 : P(\Omega) = 1$;
- $A3 : P(A \cup B) = P(A) + P(B)$ se os eventos A e B forem mutuamente exclusivos.

A partir dos axiomas $A1$, $A2$ e $A3$, pode-se enunciar e demonstrar os seguintes teoremas, que serão muito úteis no cálculo de probabilidades

T1. $P(\phi) = 0$ Prova:

$$\begin{aligned}\Omega &= \Omega \cup \phi \\ P(\Omega) &= P(\Omega \cup \phi) \leftarrow A3 \\ P(\Omega) &= P(\Omega) + P(\phi) \leftarrow A2 \\ 1 &= 1 + P(\phi) \\ P(\phi) &= 0\end{aligned}$$

T2. $P(A^c) = 1 - P(A)$, $A^c = A$ complementar. Prova:

$$\begin{aligned}\Omega &= A \cup A^c \\ P(\Omega) &= P(A \cup A^c) \leftarrow A3 \\ P(\Omega) &= P(A) + P(A^c) \leftarrow A2 \\ 1 &= P(A) + P(A^c) \\ P(A^c) &= 1 - P(A)\end{aligned}$$

T3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Prova: Os eventos A e $A^c \cup B$ são mutuamente exclusivos.

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

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

Logo, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

T4. $0 \leq P(A) \leq 1$

3.4 Probabilidade Condicionada

Ex: Seja o processo aleatório de se lançar um dado de seis faces, e observar o valor numérico da face superior deste.

$$\Omega = 1, 2, 3, 4, 5, 6$$

Sejam ainda os seguintes eventos:

Evento A: (face superior par) $A = 2, 4, 6$;

Evento B: (face superior maior ou igual a 4) $B = 4, 5, 6$;

Tem-se que: $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{3}$.

Qual a probabilidade de que a face superior do dado seja maior ou igual 4, sabendo-se que ela é par? Neste caso quer saber-se qual a probabilidade do evento B , sabendo-se que o evento A já ocorreu, $P(B/A)$. Se já sabemos que o evento A ocorreu (que a face superior é par), isto implica numa redução ou restrição no espaço amostral Ω , que passa a ser: $\omega^* = \{2, 4, 6\}$ e evento em questão, (face superior maior ou igual a 4) passa a ser: $B^* = \{4, 6\}$. Então $P(B^*) = \frac{2}{3}$.

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Definição:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \text{ e } P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (3.1)$$

3.5 Independência de eventos.

Dois eventos, A e B são considerados independentes (a ocorrência de um, não altera a probabilidade de ocorrência do outro) se e somente se:

$$P(A/B) = P(A) \text{ ou } P(B/A) = P(B)$$

Deste modo, se A e B forem independentes tem-se:

$$\begin{aligned} P(A/B) &= \frac{P(A \cap B)}{P(B)} \\ P(A \cap B) &= P(A/B)P(B) \therefore \\ P(A \cap B) &= P(A)P(B) \end{aligned}$$

Ex. Considere o quadro a seguir, representativo da distribuição dos alunos matriculados num determinado Instituto de matemática:

Curso	sexo		total
	Masculino	Feminino	
Mat. pura	70	40	110
Mat. aplicada	15	15	30
Estatística	10	20	30
Computação	20	10	30
Total	115	85	200

Observando-se os dados acima verifica-se que a probabilidade de um aluno aleatoriamente escolhido ser:

a. do sexo masculino é

$$P(M) = \frac{115}{200} = 0,575$$

b. do sexo feminino é

$$P(F) = \frac{85}{200} = 0,425$$

c. do curso de Mat. Pura é

$$P(P) = \frac{110}{200} = 0,550$$

d. do sexo feminino e fazer Mat. Pura é

$$P(F \cap P) = \frac{40}{200} = 0,200$$

e. do curso de Mat. Pura, dado que é do sexo feminino é

$$P(P/F) = \frac{40}{85} = 0,4706 \text{ ou}$$

$$P(P/F) = \frac{P(F \cap P)}{P(F)} = \frac{0,200}{0,425} = 0,4706$$

Como $P(P) \neq P(P/F)$ conclui-se que os eventos sexo e curso no qual o aluno esta matriculada não são independentes.

3.6 Teorema de Bayes

Ex. Considere 5 urnas exatamente iguais, cada uma com 6 bolas. Duas destas urnas, (tipo C_1) têm 3 bolas brancas, duas outras, (tipo C_2) têm 2 bolas brancas e a última, (tipo C_3) tem 6 bolas brancas. Escolhe-se uma urna ao acaso, e retira-se uma bola desta. Qual a probabilidade de que a urna escolhida seja do tipo C_3 , sabendo-se que a bola retirada era branca? Esquema:

Figura 3.2: Representação esquemática das urnas contendo 6 bolas cada

Quer se obter $P(C_3/B)$, sabendo-se que:

$$P(C_1) = \frac{2}{5}; P(B/C_1) = \frac{1}{2}; P(C_2) = \frac{2}{5}; P(B/C_2) = \frac{1}{3}; P(C_3) = \frac{1}{5}; P(B/C_3) = 1;$$

Da definição de probabilidade condicionada, tem-se:

$$P(C_3/B) = \frac{P(C_3 \cap B)}{P(B)} = \frac{P(C_3)P(B/C_3)}{P(B)} \quad (1)$$

Como o numerador é conhecido, necessita-se conhecer $P(B)$. E como C_1 , C_2 e C_3 são eventos mutuamente exclusivos, e reunidos formam o espaço amostral completo, pode-se decompor o evento B em 3 eventos mutuamente exclusivos:

$$B = (C_1 \cap B) \cup (C_2 \cap B) \cup (C_3 \cap B),$$

então,

$$\begin{aligned} P(B) &= P(C_1 \cap B) + P(C_2 \cap B) + P(C_3 \cap B) \\ &= P(C_1)P(B/C_1) + P(C_2)P(B/C_2) + P(C_3)P(B/C_3) \\ &= \left(\frac{2}{5}\right) \left(\frac{1}{2}\right) + \left(\frac{2}{5}\right) \left(\frac{1}{2}\right) + \left(\frac{1}{5}\right) (1) \\ &= 8/15(2) \end{aligned}$$

Substituindo (2) em (1)

$$P(C_3/B) = \frac{P(C_3)P(B/C_3)}{P(B)} = \frac{\frac{1}{5}(1)}{\frac{8}{5}} = \frac{3}{8}$$

Generalizando este resultado, tem-se o teorema de Bayes:

$$P(C_i/A) = \frac{P(C_i)P(A/C_i)}{\sum_{i=1}^n P(C_i)P(A/C_i)} \quad (3.2)$$

Ex. Uma companhia produz circuitos integrados em três fábricas I, II e III. A fábrica I produz 40% dos circuitos, enquanto que as fábricas II e III produzem 30% cada. A probabilidade de que um circuito produzido por estas fábricas não funcione é de 0,01, 0,04 e 0,03, respectivamente. Pegando-se um circuito, ao acaso, da produção conjunta desta companhia, e verificando-se que ele não funciona, qual a probabilidade

dele ter sido produzido na fábrica I?

$$\begin{aligned}P(I/defeito) &= \frac{P(I)P(defeito/I)}{P(I)P(defeito/I) + P(II)P(defeito/II) + P(III)P(defeito/III)} \\ &= \frac{(0,40)(0,01)}{(0,40)(0,01) + (0,30)(0,04) + (0,30)(0,03)} \\ &= 0,16\end{aligned}$$

Capítulo 4

Variáveis aleatórias

4.1 Variáveis Aleatórias Unidimensionais.

Uma função que associa valores reais aos eventos de um espaço amostral é definida como uma variável aleatória, que pode ser discreta ou contínua.

Ex1. Se um experimento consiste em verificar o número de componentes defeituosos, num sistema formado por três componentes, a função:

$X =$ “número de componentes defeituosos”, define uma variável aleatória discreta, que pode assumir os valores 0, 1, 2 ou 3.

Ex2. Se um experimento consiste em verificar as alturas de 30 universitários, a função:

$Y =$ “Altura de um universitário”, define uma variável aleatória contínua, que pode assumir quaisquer valores entre 130 e 220cm.

4.1.1 Distribuições de Probabilidades

Se uma variável aleatória X pode assumir os valores x_1, x_2, \dots, x_n com probabilidades respectivamente iguais a p_1, p_2, \dots, p_n , tais que $\sum_{i=1}^n p_i = 1$, tem-se definida uma **distribuição de probabilidade**.

Se a variável X em questão, for discreta, sua distribuição é caracterizada por uma **função de probabilidade** $P(X = x)$, que associa probabilidades não nulas aos possíveis valores da variável aleatória.

Ex. Para o exemplo do número de componentes defeituosos em um sistema composto por três componentes tem-se:

X	0	1	2	3	
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\sum = 1,00$

Se, a variável X for contínua, somente haverá interesse na probabilidade de que a variável assumira valores dentro de determinados intervalos, sendo sua distribuição de probabilidades caracterizada por uma **função densidade probabilidade** (f.d.p.), $f(x)$, a qual deverá possuir as seguintes propriedades:

- i. $f(x) \geq 0, \forall x \in \mathfrak{R}$.
- ii. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Ex. Para o caso as alturas dos universitários tem-se

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}},$$

que é a distribuição normal.

4.1.2 Função Repartição ou Distribuição Acumulada $F(x)$

$$F(x) = \sum_{i=1}^x P(X = x), \quad (4.1)$$

para variáveis aleatórias discretas, e

$$F(x) = \int_{-\infty}^x f(x)dx, \quad (4.2)$$

para variáveis aleatórias contínuas.

Exemplos:

Seja a seguinte variável aleatória contínua, definida pela função densidade de probabilidade (f.d.p):

$$\begin{cases} f(x) = 0 \text{ para } x < 0 \\ f(x) = kx \text{ para } 0 \leq x \leq 2 \\ f(x) = 0 \text{ para } x > 2 \end{cases}$$

graficamente tem-se:

Figura 4.1: Representação gráfica da função $F(x)$ anteriormente definida

a) Obter o valor de k .

Como $f(x)$ é uma fdp tem-se que:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= 1, \text{ portanto:} \\ \int_{-\infty}^0 0dx + \int_0^2 kx dx + \int_2^{\infty} 0dx &= 1, \\ 0 + k \int_0^2 x dx + 0 &= 1, \\ k &= \frac{1}{2} \end{aligned}$$

b) calcular $F(1)$

$$\begin{aligned} F(1) &= P(X \leq 1) = \int_{-\infty}^1 f(x)dx \\ &= \int_{-\infty}^1 \frac{x}{2} dx = \frac{1}{4} \end{aligned}$$

4.2 Variáveis Aleatórias Bidimensionais

Para o estudo de variáveis aleatórias, até este ponto, considerou-se que o resultado do experimento em questão seria registrado como um único valor x . Todavia, existem casos em que há interesse por dois resultados simultâneos, como por exemplo observar

o peso e altura de uma pessoa, o sexo e peso de um recém-nascido, etc. Para tanto, faz-se necessário a seguinte definição:

Sejam E um experimento aleatório, e S o espaço amostral associado a E . Sejam X e Y duas variáveis aleatórias. Então (X, Y) define uma variável aleatória bidimensional, que pode ser discreta, contínua ou mista.

4.2.1 Distribuição Conjunta de duas variáveis aleatórias

Se (X, Y) é uma variável aleatória bidimensional discreta, sua função de probabilidade, representada por $P(X = x_i; Y = y_i)$ que associa um valor $p(x_i, y_i)$ a cada valor do par (X, Y) deve satisfazer as seguintes condições:

- i. $P(x_i, y_i) \geq 0 \forall (x_i, y_i)$.
- ii. $\sum \sum P(x_i, y_i) = 1$.

Exemplo: Seja o experimento de se lançar simultaneamente um dado e uma moeda, observando o resultado da face superior de ambos. Teremos então a seguinte função de probabilidade, onde :

X= face superior do dado, e Y= face superior da moeda

$X \backslash Y$	Cara	Coroa
1	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$
4	$\frac{1}{12}$	$\frac{1}{12}$
5	$\frac{1}{12}$	$\frac{1}{12}$
6	$\frac{1}{12}$	$\frac{1}{12}$
	1	

Se (X, Y) for uma variável aleatória bidimensional contínua, diz-se que $f(x, y)$ é uma função densidade de probabilidade conjunta se:

- i. $f(x, y) \geq 0, \forall (x, y) \in \mathfrak{R}$.
- ii. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

4.2.2 Distribuição Marginal

Dada uma variável aleatória bidimensional, e sua distribuição de probabilidade conjunta, pode-se obter a distribuição da variável X , sem considerar Y ou vice-versa, que são denominadas distribuições marginais de X e Y respectivamente.

Distribuição marginal de X

$$P(X = x_i) = \sum_j P(x_i, y_j), \quad (4.3)$$

para variáveis aleatórias discretas e

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad (4.4)$$

para variáveis aleatórias contínuas.

Distribuição marginal de Y

$$P(Y = y_j) = \sum_i P(x_i, y_j), \quad (4.5)$$

para variáveis aleatórias discretas e

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx, \quad (4.6)$$

para variáveis aleatórias contínuas.

Exemplo no exemplo do lançamento simultâneo de um dado e uma moeda teremos:

X = face superior do dado, e Y = face superior da moeda

X = face superior do dado, e Y = face superior da moeda

$X \setminus Y$	Cara	Coroa	$P(X = x_i)$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
2	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
4	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
5	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
6	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
$P(Y = y_j)$	$\frac{1}{2}$	$\frac{1}{2}$	1

4.2.3 Variáveis Aleatórias Independentes

Seja (X, Y) uma variável aleatória bidimensional, então as variáveis X e Y são independentes se e somente se

$$P(x_i, y_j) = P(x_i) \cdot P(y_j) \forall i \text{ e } j,$$

para variáveis aleatórias discretas, ou

$$f(x, y) = g(x) \cdot h(y) \forall i \text{ e } j,$$

para variáveis aleatórias contínuas

4.2.4 Esperança Matemática

A esperança matemática de uma variável aleatória ou valor esperado, $E(X)$, e definida por:

$$E(X) = \mu = \sum_i x_i P(X = x_i), \quad (4.7)$$

para variáveis aleatórias discretas, e

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx, \quad (4.8)$$

para variáveis aleatórias contínuas

Ex1: Para a variável aleatória contínua definida por:

$$\begin{cases} f(x) = 0 \text{ para } x < 0 \\ f(x) = \frac{x}{2} \text{ para } 0 \leq x \leq 2 \\ f(x) = 0 \text{ para } x > 2 \end{cases}$$

tem-se

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^0 0dx + \int_0^2 \frac{x^2}{2}dx + \int_2^{\infty} 0dx \\ &= \frac{4}{3} \text{ unidade} \end{aligned}$$

Ex2: Para a variável aleatória discreta número de componentes pifados tem-se:

X	0	1	2	3	
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\sum = 1, 0$

$$\begin{aligned} E(X) &= \sum_i x_i P(X = x_i) \\ &= 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} \\ &= 1,5 \text{ componentes pifados por sistema.} \end{aligned}$$

4.2.4.1 Propriedades da Esperança Matemática

- i. $E(k) = k$, $k = \text{constante}$;
- ii. $E(kX) = k.E(X)$;

- iii. $E(X \pm Y) = E(X) \pm E(Y)$;
- iv. $E(X \pm k) = E(X) \pm k$;
- v. $E(XY) = E(X)E(Y)$ se X e Y forem independentes.

4.2.5 Variância de uma variável aleatória.

$$V(X) = E[X - E(X)]^2 = E[X - \mu]^2. \quad (4.9)$$

Aplicando-se a definição de esperança matemática (eq 4.7 e 4.8 verifica-se que a variância pode ser então definida como:

$$V(X) = \sum_i (x_i - \mu)^2 P(X = x_i), \quad (4.10)$$

para variáveis aleatórias discretas, e

$$V(X) = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x) dx, \quad (4.11)$$

para variáveis aleatórias contínuas.

Obs.

$$\begin{aligned} V(X) &= E[X - E(X)]^2 \\ &= E\{X^2 - 2XE(X) + [E(X)]^2\} \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

com,

$$E(X^2) = \sum_i x_i^2 P(X = x_i),$$

no caso discreto, e

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

no caso contínuo.

Ex1. Para a variável aleatória contínua definida por:

$$\begin{cases} f(x) = 0 \text{ para } x < 0 \\ f(x) = \frac{x}{2} \text{ para } 0 \leq x \leq 2 \\ f(x) = 0 \text{ para } x > 2 \end{cases}$$

tem-se:

$$V(X) = E(X^2) - [E(X)]^2$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_{-\infty}^0 0 dx + \int_0^2 \frac{x^3}{2} dx + \int_2^{\infty} 0 dx \\ &= 2 \text{ unidade} \end{aligned}$$

$$V(X) = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9} \text{ unidade}^2$$

Ex2: Para a variável aleatória discreta número de componentes pifados tem-se:

X	0	1	2	3	
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\Sigma = 1,0$

$$V(X) = E(X^2) - [E(X)]^2$$

$$\begin{aligned} E(X^2) &= \sum_i x_i^2 P(X = x_i) \\ &= (0^2)\frac{1}{8} + (1^2)\frac{3}{8} + (2^2)\frac{3}{8} + (3^2)\frac{1}{8} \\ &= 3 \end{aligned}$$

$$V(X) = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} \text{ (componentes pifados por sistema)}^2$$

4.2.5.1 Propriedades da variância

- i. $V(k) = 0$, $k = \text{constante}$;
- ii. $V(k.X) = k^2V(X)$;
- iii. $V(k \pm X) = V(X)$;
- iv. $V(X \pm Y) = V(X) \pm V(Y)$ se X e Y forem independentes;

4.2.6 Covariância

A covariância mede o grau de dispersão conjunta de duas variáveis aleatórias.

$$Cov(X, Y) = E \{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y), \quad (4.12)$$

com,

$$E(XY) = \sum_i \sum_j x_i y_j P(X = x_i)(Y = y_j),$$

para variáveis aleatórias discretas, e

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(xy) dx dy,$$

para variáveis aleatórias contínuas.

Obs. Para duas variáveis aleatórias quaisquer tem-se:

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y).$$

Se X e Y forem independentes, $Cov(X, Y) = 0$, voltando-se a propriedade iv. das variâncias. Porém o fato de $Cov(X, Y) = 0$ não implica que X e Y sejam independentes.

4.3 Distribuições de variáveis aleatórias discretas

4.3.1 Distribuição Uniforme Discreta

Enquadram-se aqui as distribuições em que os possíveis valores da variável aleatória tenham todos a mesma probabilidade de ocorrência. Logo, se existem n valores possíveis, cada um terá probabilidade igual a $\frac{1}{n}$.

Ex. Seja o lançamento de um dado e a variável aleatória $X =$ “face superior do dado”, tem-se que:

X	1	2	3	4	5	6	
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\Sigma = 1$

ou $P(X = x) = 1/6$

4.3.2 Distribuição de Bernoulli

Seja um experimento onde só podem ocorrer dois possíveis resultados, “sucesso” e “fracasso”, como por exemplo:

- Um jogador de basquete converter ou não converter um arremesso,
- Um indivíduo portador de certa doença morrer ou não,
- Uma peça produzida por uma Cia. Ser perfeita ou defeituosa,
- O sexo do primeiro filho de um casal ser masculino ou feminino,
- Um consumidor que entra numa loja comprar ou não comprar um produto.

Associando-se uma variável aleatória X aos possíveis resultados do experimento, de forma que:

$$\begin{aligned}
 X &= 1 && \text{se o resultado for “sucesso” e} \\
 X &= 0 && \text{se o resultado for “fracasso”}.
 \end{aligned}$$

Então, a variável aleatória X , assim definida tem distribuição Bernoulli, com p sendo a probabilidade de ocorrer “sucesso”, e $q = (1-p)$ a probabilidade de ocorrer “fracasso”.

Função de probabilidade

A função de probabilidade da Distribuição de Bernoulli é dada por:

$$P(X = x) = \begin{cases} q = (1 - p) & \text{para } x = 0 \\ p & \text{para } x = 1 \\ 0 & \text{para outros valores de } x. \end{cases} \quad (4.13)$$

Parâmetros característicos

$$E(X) = p$$

Prova:

$$\begin{aligned} E(X) &= \sum x_i P(X = x_i) \\ &= 1p + 0(1 - P) \\ &= p \end{aligned}$$

$$V(X) = pq$$

Prova:

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ E(X^2) &= \sum x_i^2 P(X = x_i) \\ &= p \\ &\therefore \\ V(X) &= p - p^2 \\ &= P(1 - p) \\ &= pq \end{aligned}$$

4.3.3 Distribuição Binomial.

É a mais importante das distribuições de probabilidades discretas. Sendo que, para um experimento se enquadrar na distribuição Binomial, deve-se atender as seguintes condições:

- i. São realizadas n provas (tentativas) independente;
- ii. Cada tentativa é uma prova de Bernoulli (só podem ocorrer dois possíveis resultados);
- iii. A probabilidade p de sucesso em cada prova é constante.

Se um experimento atende a todas as condições acima, então a variável aleatória $X =$ número de sucessos obtidos nas n tentativas terá uma distribuição Binomial, com n tentativas e p (probabilidade de sucesso). Simbolicamente : $X \sim B(n, p)$

Função de Probabilidade

$$P(X = x) = C_n^x p^x q^{n-x}, \quad (4.14)$$

com

$$C_n^x = \frac{n!}{x!(n-x)!};$$

$p =$ probabilidade de “sucesso”;

$q = 1 - p =$ probabilidade de “fracasso”

Parâmetros característicos

$$E(X) = np$$

$$V(X) = npq$$

prova:

$$\begin{aligned}
E(X) &= \sum_{x=0}^n xP(X=x) \\
&= \sum_{x=0}^n xC_n^x p^x q^{n-x} \\
&= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} \text{ fazendo } s = x - 1 \\
&= \sum_{s=0}^{n-1} n \frac{(n-1)!}{s!(n-1-s)!} p^{(s+1)} q^{n-(s+1)} \\
&= n \sum_{s=0}^{n-1} C_{(n-1)}^s p^{(s+1)} q^{n-(s+1)} \\
&= np \sum_{s=0}^{n-1} C_{(n-1)}^s p^s q^{(n-1)-s} \\
&= np
\end{aligned}$$

Exemplos:

Ex1. Sabendo-se que a probabilidade de um determinado casal gerar um filho com olhos azuis é de $\frac{1}{4}$, qual a probabilidade de que dentre três filhos deste casal,

- a) Nenhum tenha olhos azuis.
- b) Um tenha olhos azuis.
- c) Dois tenham olhos azuis.
- d) Os Três tenham olhos azuis.

Serão considerados dois métodos para resolução deste exemplo:

1º Método - pela definição de probabilidades:

Espaço amostral:

$$\Omega = \left\{ \begin{array}{ll} EEE & EAA \\ EEA & AEA \\ EAE & AAE \\ AEE & AAA \end{array} \right\}$$

Uma vez que a cor dos olhos de um filho independe da cor dos olhos dos demais (são eventos independentes), a probabilidade de cada um dos eventos do espaço amostral é:

$$\begin{aligned} P(EEE) &= \frac{27}{64} & P(EAA) &= \frac{3}{64} \\ P(EEA) &= \frac{9}{64} & P(AEA) &= \frac{3}{64} \\ P(EAE) &= \frac{9}{64} & P(AAE) &= \frac{3}{64} \\ P(AEE) &= \frac{9}{64} & P(AAA) &= \frac{1}{64} \end{aligned}$$

Assim,

- $P(\text{Nenhum com olhos azuis}) = P(EEE) = \frac{27}{64};$
- $P(\text{um com olhos azuis}) = P(EEA \cup EAE \cup AEE) = \frac{9}{64} + \frac{9}{64} + \frac{9}{64} = \frac{27}{64};$
- $P(\text{dois com olhos azuis}) = P(EAA \cup AAE \cup AEA) = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{9}{64};$
- $P(\text{três com olhos azuis}) = P(AAA) = \frac{1}{64};$

2º Método - utilizando a função de probabilidade binomial:

$$X \sim B(3, \frac{1}{4})$$

- $P(\text{Nenhum com olhos azuis}) = P(X = 0) = C_3^0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^3 = \frac{27}{64};$
- $P(\text{um com olhos azuis}) = P(X = 1) = C_3^1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^2 = \frac{27}{64};$
- $P(\text{dois com olhos azuis}) = P(X = 2) = C_3^2 \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1 = \frac{9}{64};$
- $P(\text{três com olhos azuis}) = P(X = 3) = C_3^3 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^0 = \frac{1}{64};$

Deste modo, verifica-se que a probabilidade total é dada por: $C_3^0 p^0 q^3 + C_3^1 p^1 q^2 + C_3^2 p^2 q^1 + C_3^3 p^3 q^0$ que corresponde a expansão do binômio $(p + q)^3$ daí o nome distribuição binomial.

Ex2. Num determinado processo de fabricação, 10% das peças produzidas são consideradas defeituosas. As peças são acondicionadas em caixas com 5 unidades cada uma.

a) Qual a probabilidade de haverem exatamente 3 peças defeituosas numa caixa?

$$X \sim B(5, 0, 1)$$

$$P(X = 3) = C_5^3(0, 1)^3(0, 9)^2 = 0, 0081$$

b) Qual a probabilidade de haverem duas ou mais peças defeituosas em uma caixa?

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 1 - [P(X = 0) + P(X = 1)] = 0, 0815$$

c) Qual a probabilidade de uma caixa não apresentar nenhuma peça defeituosa?

$$P(X = 0) = C_5^0(0, 1)^0(0, 9)^5 = 0, 5905$$

d) Supondo que a empresa pague uma multa de R\$10,00 por caixa que apresente peças defeituosas, qual o valor esperado desta multa em um lote de 1000 caixas?

$$P(\text{uma caixa ter peça defeituosa}) = 1 - P(X = 0) = 0, 4095.$$

O número de caixas com peças defeituosas em um lote de 1000 caixas segue uma distribuição binomial com $n = 1000$ e $p = 0, 4095$. Assim,

$$E(Y) = np = 1000 \cdot 0, 4095 = 409, 5 \text{ caixas. e o valor esperado da multa:}$$

$$E(\text{Multa}) = 10E(Y) = (10)(409, 5) = R\$4095, 00$$

4.3.4 Distribuição de Poisson

A distribuição de Poisson é empregada em experimentos nos quais não se está interessado no número de sucessos obtido em n tentativas, como ocorre no caso da distribuição Binomial, mas sim no número de sucessos ocorridos durante um intervalo contínuo, que pode ser um intervalo de tempo, espaço, etc. Como por exemplo:

- O número de suicídios ocorridos em uma cidade durante um ano;
- O número de acidentes automobilísticos ocorridos numa rodovia em um mês;

- O número de defeitos encontrados em um rolo de arame ovalado, de 500m;

Note que nos exemplos acima, não há como determinar-se a probabilidade de ocorrência de um sucesso, mas sim a frequência média de sua ocorrência, como por exemplo dois suicídios por ano, a qual será que denominada λ . Em um experimento com estas características, e assumindo-se que os sucessos sejam independentes, a variável aleatória

X = número de sucessos em um intervalo,

terá uma distribuição Poisson, com parâmetro λ . Simbolicamente : $X \sim P(\lambda)$

Função de Probabilidade

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad (4.15)$$

com

$e = 2,7182$ (base dos logaritmos neperianos).

Parâmetros Característicos

$$E(X) = V(X) = \lambda$$

Prova:

$$\begin{aligned} E(X) &= \sum_{x=0}^n xP(X = x) \\ &= \sum_{x=0}^n x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=0}^n \frac{e^{-\lambda} \lambda^x}{(x-1)!} \text{ fazendo } s=x-1 \text{ tem-se:} \\ &= \sum_{x=0}^n \frac{e^{-\lambda} \lambda^{s+1}}{s!} \\ &= \lambda \sum_{x=0}^n \frac{e^{-\lambda} \lambda^s}{s!} \\ &= \lambda \end{aligned}$$

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^n x^2 P(X = x) \\
&= \sum_{x=0}^n x^2 \frac{e^{-\lambda} \lambda^x}{x!} \\
&= \sum_{x=0}^n x \frac{e^{-\lambda} \lambda^x}{(x-1)!} \text{ fazendo } s=x-1 \text{ tem-se:} \\
&= \sum_{x=0}^n (s+1) \frac{e^{-\lambda} \lambda^{s+1}}{s!} \\
&= \lambda \sum_{x=0}^n (s+1) \frac{e^{-\lambda} \lambda^s}{s!} \\
&= \lambda \left[\sum_{x=0}^n s \frac{e^{-\lambda} \lambda^s}{s!} + \sum_{x=0}^n \frac{e^{-\lambda} \lambda^s}{s!} \right] \\
&= \lambda [\lambda + 1] \\
&= \lambda^2 + \lambda
\end{aligned}$$

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

Exemplo: O Corpo de Bombeiros de uma determinada cidade recebe, em média, 3 chamadas por dia. Qual a probabilidade de receber:

a) 4 chamadas num dia X $P(3)$

$$P(X = 4) = \frac{e^{-3} 3^4}{4!} = 0,1680.$$

b) Nenhuma chamada em um dia

$$P(X = 0) = \frac{e^{-3} 3^0}{0!} = 0,0498.$$

c) 20 chamadas em uma semana. X = número de chamadas por dia

Y = número de chamadas por semana

$E(X) = \lambda = 3$ chamadas por dia $\rightarrow E(Y) = \lambda^* = 7E(X) = 21$ chamadas por semana

$$P(Y = 20) = \frac{e^{-21}21^{20}}{20!} = 0,0867.$$

4.3.4.1 Aproximação da distribuição Binomial a Poisson.

Pode-se demonstrar que uma distribuição Binomial, cujo evento de interesse é raro (p muito pequeno e n muito grande), tende para uma distribuição de Poisson. Na prática, a aproximação é considerada boa quando $n \geq 50$ e $p \leq 0,10$.

Aproximação: Sabe-se que se $X \sim B(n, p)$, $E(X) = np$, então $\lambda = E(X) = np$

Exemplo: A probabilidade de um indivíduo sofrer uma reação alérgica, resultante da injeção de determinado soro é de 0,01. Determinar a probabilidade de entre 200 indivíduos, submetidos a este soro, nenhum sofrer esta reação alérgica.

$$X \sim B(200, 0,01) \rightarrow E(X) = n.p = 200 \times 0,01 = 2 = \lambda$$

$$P(X = 2) \simeq \frac{e^{-2}2^0}{0!} = 0,1353.$$

4.3.5 Distribuição Geométrica

Suponha-se um experimento, no qual esteja-se interessado apenas na ocorrência ou não de um determinado evento, como por exemplo o sexo do filho de uma determinada mulher ser feminino. E, assim como na distribuição binomial, que esse experimento seja repetido um número n de vezes, que em cada repetição seja independente das demais e que a probabilidade de sucesso p em cada repetição seja constante. Suponha-se que o experimento seja repetido até que ocorra o primeiro sucesso (o sexo do filho seja feminino).

Então a variável aleatória: $X =$ número de tentativas até que se obtenha o primeiro sucesso, seguirá uma distribuição geométrica, com parâmetro p (probabilidade de sucesso) . Simbolicamente $X \sim G(p)$.

Função de Probabilidade

Como o experimento será repetido até que se obtenha o primeiro sucesso, e considerando que esse ocorra na k -ésima repetição, deverão ocorrer $k - 1$ fracassos antes que o experimento seja encerrado. Assim, a a probabilidade de que a variável aleatória $X =$ número de repetições até se obter o primeiro sucesso é

$$P(X = x) = pq^{x-1}, \quad (4.16)$$

com

$p =$ probabilidade de “sucesso”;

$q = 1 - p =$ probabilidade de “fracasso”

Parâmetros característicos

$$E(X) = \frac{1}{p}$$

$$V(X) = \frac{q}{p^2}$$

Prova:

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xP(X = x) \\ &= \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1} \\ &= p \sum_{x=1}^{\infty} \frac{d}{dq} q^x = p \frac{d}{dq} \sum_{x=1}^{\infty} q^x \\ &= p \frac{d}{dq} (q^1 + q^2 + \dots) = p \frac{d}{dq} \left[\frac{q}{1-q} \right] \\ &= p \frac{dq(1-q) - d(1-q)q}{(1-q)^2} = p \frac{1(1-q) - (-1)q}{(1-q)^2} \\ &= p \frac{1}{(1-q)^2} = p \frac{1}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

Obs: a permuta da derivação e do somatório é válida aqui, porque a série converge quando $|q| < 1$.

$$\begin{aligned}
 E(X^2) &= \sum_{x=1}^{\infty} x^2 P(X = x) \\
 &= \sum_{x=1}^{\infty} x^2 pq^{x-1} = p \sum_{x=1}^{\infty} x^2 q^{x-1} \\
 &= p \sum_{x=1}^{\infty} \frac{d}{dq} q^x = p \frac{d}{dq} \sum_{x=1}^{\infty} q^x \\
 &= p \frac{d}{dq} (q^1 + q^2 + \dots) = p \frac{d}{dq} \left[\frac{q}{1-q} \right] \\
 &= p \frac{dq(1-q) - d\partial(1-q)q}{(1-q)^2} = p \frac{1(1-q) - (-1)q}{(1-q)^2} \\
 &= p \frac{1}{(1-q)^2} = p \frac{1}{p^2} \\
 &= \frac{1}{p}
 \end{aligned}$$

Exemplo:

Um casal com problemas para engravidar, recorreu a uma técnica de inseminação artificial no intuito de conseguir o primeiro filho. A eficiência da referida técnica é de 0,20 e o custo de cada inseminação U\$2000,00.

a) Qual a probabilidade de que o casal obtenha êxito na terceira tentativa?

$$\begin{aligned}
 P(X = k) &= pq^{k-1} \\
 &= (0,2)(0,8)^2 \\
 &= 0,128
 \end{aligned}$$

b) Qual o custo esperado deste casal para obter o primeiro filho?

$$\begin{aligned}
 E(X) &= \frac{1}{p} \\
 &= \frac{1}{0,2} \\
 &= 5
 \end{aligned}$$

Custo esperado = $5 * 2000,00 = U\$10000,00$

4.3.6 Distribuição Pascal (Binomial Negativa)

Nas mesmas condições em que foi definida a distribuição geométrica, e considerando que o experimento será repetido até que se obtenha o *r-ésimo* sucesso, então a variável $X = \text{número de tentativas até se obter o } r\text{-ésimo sucesso}$ seguirá a distribuição de Pascal.

Função de Probabilidade:

Para que o *r-ésimo* sucesso ocorra na *k-ésima* tentativa, é necessário que ocorra um sucesso nesta tentativa (repetição do experimento) e que tenham ocorrido $r-1$ sucessos nas $k-1$ repetições anteriores. Dado que a probabilidade de ocorrência de sucesso, numa dada repetição do experimento é dada por p e a probabilidade de ocorrerem $r-1$ sucessos em $k-1$ repetições, e sendo estes dois eventos independentes, a probabilidade de que o *r-ésimo* sucesso ocorra na *k-ésima* repetição do experimento é dada por:

$$\begin{aligned} P(X = k) &= pC_{(k-1)}^{(r-1)}p^{r-1}q^{(k-1)-(r-1)}; \\ &= C_{(k-1)}^{(r-1)}p^r q^{k-r}, k \geq r; \end{aligned} \quad (4.17)$$

em que:

p = probabilidade de “sucesso”;

$q = 1 - p$ = probabilidade de “fracasso”

Parâmetros característicos

$$E(X) = \frac{r}{p}$$

$$V(X) = \frac{rq}{p^2}$$

Prova:

De acordo com Meyer (1969) a esperança e variância da distribuição Pascal podem ser obtidas do seguinte modo:

Sejam as variáveis:

Z_1 = número de repetições necessárias até o primeiro sucesso

Z_2 = número de repetições necessárias entre o primeiro sucesso
e o segundo, inclusive

\vdots \vdots

Z_r = número de repetições necessárias entre o $(r-1)$ sucesso e o r -ésimo sucesso,
inclusive

É imediato verificar que todas as variáveis Z_i são independentes e possuem distribuição geométrica. Assim,

$$\begin{aligned} E(Y) &= E(Z_1 + Z_2 + \cdots + Z_r) \\ &= E(Z_1) + E(Z_2) + \cdots + E(Z_r) \\ &= \frac{1}{p} + \frac{1}{p} + \cdots + \frac{1}{p} \\ &= \frac{r}{p} \end{aligned}$$

de modo análogo:

$$\begin{aligned} V(Y) &= V(Z_1 + Z_2 + \cdots + Z_r) \\ &= V(Z_1) + V(Z_2) + \cdots + V(Z_r) \\ &= \frac{q}{p^2} + \frac{q}{p^2} + \cdots + \frac{q}{p^2} \\ &= \frac{rq}{p^2} \end{aligned}$$

4.3.7 Distribuição Hipergeométrica

Considere um conjunto de N elementos, r dos quais têm uma determinada característica ($r \leq N$), e que destes N elementos serão extraídos n elementos sem reposição ($n \leq N$). A variável aleatória $X = \text{número de elementos com a referida característica, que estarão entre os } n \text{ retirados}$, segue uma distribuição hipergeométrica, cuja função de probabilidade é derivada diretamente da definição clássica de probabilidade.

Função de probabilidade:

$$P(X = x) = \frac{C_r^x C_{N-r}^{n-x}}{C_N^n} \quad (4.18)$$

Parâmetros característicos:

Fazendo $\frac{r}{N} = p$ e $\frac{N-r}{N} = q$ tem-se:

$$E(X) = np \quad (4.19)$$

$$V(X) = npq \frac{N-n}{N-1} \quad (4.20)$$

Exemplo:

No fichário de um hospital, estão arquivados os prontuários dos de 20 pacientes, que deram entrada no PS apresentando algum problema cardíaco. Destes 5 sofreram infarto. Retirando-se uma amostra ao acaso de 3 destes prontuários, qual a probabilidade de que dois deles sejam de pacientes que sofreram infarto?

$$\begin{aligned} P(X = 2) &= \frac{C_5^2 C_{20-5}^{3-2}}{C_{20}^3} \\ &= \frac{C_5^2 C_{15}^1}{C_{20}^3} \\ &= \frac{(10)(15)}{1140} \\ &= 0,1315 \end{aligned}$$

4.3.8 Distribuição Multinomial

Considere um experimento com as seguintes características:

- i. São realizadas n provas independentes;
- ii. Cada prova admite um único resultado entre r possíveis;
- iii. As probabilidades p_i de ocorrer um determinado resultado são constantes para todas as repetições do experimento.

Associando a este experimento r variáveis aleatórias (X_1, X_2, \dots, X_r) cada uma indicando o número de vezes que ocorreu o resultado nas n repetições. Então, a distribuição da variável multi dimensional (X_1, X_2, \dots, X_r) é chamada distribuição multinomial.

Função de probabilidade

$$P(X_1 = x_1; X_2 = x_2; \dots; X_r = x_r) = \frac{n!}{x_1!x_2! \dots x_n!} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n} \quad (4.21)$$

Exemplo:

Em um determinado cruzamento entre duas plantas de milho, a probabilidade de se obter uma planta com genótipo MM é igual a 0,25, com genótipo Mm , 0,50 e com genótipo mm 0,25. De 10 descendentes deste cruzamento, qual a probabilidade de que se obtenham respectivamente 2, 5 e 3 indivíduos com genótipos MM , Mm e mm ?

$$\begin{aligned} P(MM = 2; Mm = 5; mm = 3) &= \frac{10!}{2!5!3!} (0,25)^2 (0,50)^5 (0,25)^3 \\ &= 0,0769 \end{aligned}$$

4.4 Distribuições de variáveis aleatórias contínuas

4.4.1 Distribuição Uniforme

A função densidade probabilidade da distribuição uniforme contínua é dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para outros valores de } x \text{ (pov)} \end{cases} \quad (4.22)$$

É fácil verificar que a equação 4.22 é uma função densidade probabilidade pois:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^a 0dx + \int_a^b \frac{1}{b-a}dx + \int_b^{\infty} 0dx \\ &= 0 + \frac{1}{b-a}x \Big|_a^b + 0 \\ &= \frac{1}{b-a}(b-a) = 1 \end{aligned}$$

Parâmetros característicos

$$E(X) = \frac{a+b}{2} \quad (4.23)$$

$$V(X) = \frac{(b-a)^2}{12} \quad (4.24)$$

Prova:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^a x0dx + \int_a^b x\frac{1}{b-a}dx + \int_b^{\infty} x0dx \\ &= 0 + \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b + 0 \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\
&= \int_{-\infty}^a x^2 0 dx + \int_a^b x^2 \frac{1}{b-a} dx + \int_b^{\infty} x^2 0 dx \\
&= 0 + \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b + 0 \\
&= \frac{b^3 - a^3}{3(b-a)}
\end{aligned}$$

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{b^3 - a^3}{3(b-a)} - \left[\frac{b+a}{2} \right]^2 \\
&= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{4(b^3 - a^3) - 3(b+a)^2}{12(b-a)} \\
&= \frac{b^3 - 3ab^2 + 3a^2b - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

4.4.2 Distribuição Normal

É a mais importante das distribuições de probabilidades contínuas, tendo grande aplicação em pesquisas científicas e tecnológicas. Pois, a maioria das variáveis contínuas de interesse prático, seguem esta distribuição, aliado ao fato da facilidade e boa precisão que é obtida na aproximação de outras distribuições, como a Binomial, para esta, e o Teorema do Limite Central (TLC) que é a base das estimativas e testes de hipóteses, realizados sobre a média de uma população qualquer, que garante que a distribuição amostral das médias segue uma distribuição normal, independentemente da distribuição da variável em estudo, como será visto mais adiante.

Função Densidade Probabilidade

A função densidade probabilidade normal é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (4.25)$$

em que:

μ e σ são os parâmetros média e desvio padrão respectivamente,

π e e são as constantes 3,1415 e 2,7182 respectivamente.

Gráfico.

O gráfico da função normal é dado por:

Figura 4.2: Distribuição normal.

Propriedades.

- i. É simétrica em relação ao ponto $x = \mu$;
- ii. Tem forma campanular (sino);
- iii. As três medidas de posição, média, mediana e moda se confundem no ponto de máximo da curva ($x = \mu$);
- iv. Fica perfeitamente definida conhecendo-se a média e o desvio padrão;
- v. Tem dois pontos de inflexão em $x = \mu \pm \sigma$;
- vi. É assintótica em relação ao eixo das abscissas.

Sendo a função 4.25 uma função densidade de probabilidade (*fdp*), área compreendida entre a curva e eixo x é igual a 1, ou seja $\int_{-\infty}^{\infty} f(x)dx = 1$.

Portanto, a área sob a curva entre os pontos a e b , em que $a < b$, dada por $\int_a^b f(x)dx = 1$ representa a probabilidade da variável X assumir um valor entre a e b .

Deste modo, é imediato verificar que probabilidade de um ponto qualquer é nula, pois $\int_a^a f(x)dx = 0$.

Notação

$$X \sim N(\mu, \sigma^2)$$

4.4.2.1 Distribuição Normal Reduzida ou Padronizada.

Como pode-se notar, o cálculo de probabilidades via distribuição normal envolve a solução de integrais que não são nada triviais. Em virtude da grande aplicação da distribuição normal, procurou-se tabelar os valores de probabilidade, que seriam obtidos por meio da integração da função densidade probabilidade normal num determinado intervalo. A dificuldade para se processar esse tabelamento se prendeu na infinidade de valores que μ e σ poderiam assumir. Nestas condições teria que se dispor de uma tabela para cada uma das infinitas combinações de μ e σ . Procurou-se, por isso, obter uma nova forma para a distribuição normal, que não sofresse a influencia destes parâmetros (μ e σ). O problema foi solucionado mediante o emprego de uma nova variável, z definida por: $z = \frac{x-\mu}{\sigma}$, que transforma todas as distribuições normais, em uma distribuição normal reduzida, ou padronizada, de média zero e desvio padrão um, $z \sim N(0, 1)$. Assim, utilizamos apenas uma tabela para o cálculo de probabilidades, para qualquer que seja a curva correspondente a uma distribuição normal. Desta forma, para um valor de $x = \mu$ numa distribuição normal qualquer, corresponde o valor: $z = 0$, na distribuição normal reduzida. Para $x = \mu + \sigma$ tem-se $z = 1$, e assim por diante.

Exemplo:

(a) A duração de um certo tipo de pneu, em quilômetros rodados, é uma variável normal com duração média $60000Km$ e desvio padrão $10000Km$.

a) Qual a probabilidade de um pneu aleatoriamente escolhido durar mais de $75000Km$?

Sabe-se que $X \sim N(60000; 10000^2)$ e deseja-se obter: $P(X \geq 75000) = ?$

Figura 4.3:

Utilizando-se a transformação:

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

tem-se:

$$\text{que o valor } x = 75000 \text{ equivale a } z = \frac{75000 - 60000}{10000} = \frac{15000}{10000} = 1,5,$$

portanto,

$$P(X \geq 75000) = P(z \geq 1,5) = 0,5 - 0,4332 = 0,0668$$

- b) Qual a probabilidade de um pneu aleatoriamente escolhido durar entre 50000km e 70000km? $P(50000 \leq X \leq 70000) = ?$

Figura 4.4:

$$P(50000 \leq X \leq 70000) = P(-1 \leq z \leq 1) = 0,3413 + 0,3413 = 0,6826$$

- c) Qual a probabilidade de um pneu aleatoriamente escolhido durar entre 63000km e 70000km?

Figura 4.5:

$$P(63000 \leq X \leq 70000) = P(0,30 \leq z \leq 1) = 0,3413 + 0,1179 = 0,2234$$

- d) Qual a probabilidade de um pneu aleatoriamente escolhido durar exatamente 70000km?

$$P(X = 70000) = P(z = 0) = 0$$

- e) O fabricante deseja fixar prazo de garantia, em quilômetros, de tal modo que, se a duração do pneu for inferior à garantia, o pneu seja trocado. De quantos quilômetros deve ser este prazo, para que somente 1% dos pneus sejam trocados?

$$x : P(X \leq x) = 0,01$$

$$z : P(Z \leq z) = 0,01 = -2,33$$

$$z = \frac{x - \mu}{\sigma} \Rightarrow -2,33 = \frac{x - 60000}{10000} \Rightarrow x = 36700 \text{ km}$$

Figura 4.6:

4.4.3 Distribuição Exponencial

Em um processo de Poisson, com parâmetro λ (isto é, tal que o número de sucessos em um determinado intervalo t segue uma distribuição de Poisson com média $\mu = \lambda t$), como por exemplo contar o número de carros que passam por um determinado ponto de uma estrada, num certo período de tempo. A distribuição da variável T , que representa o intervalo decorrido entre dois sucessos consecutivos, é conhecida como Distribuição Exponencial. Cujas função densidade probabilidade é dada por:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \quad (4.26)$$

Figura 4.7: Distribuição exponencial.

Prova:

Por definição a variável T representa o tempo decorrido entre dois sucessos em um processo de *Poisson*. Então para que T seja maior que um t qualquer é preciso que o próximo sucesso demore mais do que t para ocorrer. Assim,

$$P(t > t) = P(0 \text{ sucessos em } t) = e^{-\lambda t}$$

a função partição no ponto t será:

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t}$$

Derivando-se a função partição em relação a t , obtém-se a função densidade probabilidade:

$$\begin{aligned} f(t) = \frac{dF(t)}{dt} &= \lambda e^{-\lambda t} \text{ para } t \geq 0 \\ &= 0 \text{ para } t < 0 \end{aligned}$$

Parâmetros característicos:

Média:

$$E(t) = \frac{1}{\lambda} \quad (4.27)$$

Variância

$$V(t) = \frac{1}{\lambda^2} \quad (4.28)$$

Prova:

$$\begin{aligned} E(t) &= \int_{-\infty}^{\infty} t f(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt \\ &= \frac{1}{\lambda} \end{aligned}$$

Exemplo:

Certo tipo de fusível tem duração de vida que segue uma distribuição exponencial com vida média de 100 horas. Cada fusível tem um custo de R\$10,00, e se durar menos de 200 horas, existe um custo adicional de R\$8,00. a) Qual a probabilidade de um fusível, aleatoriamente escolhido, dura mais de 150 horas?

l=? Média :m(t)=

b) Qual o custo esperado dos fusíveis custo=

$$P(T > 200) = 1 - E(\text{custo}) = 10.0,1353 + 18.0,8647 = R\$ 16,92$$

4.4.4 Distribuição Qui-Quadrado

A distribuição de χ^2 (lê-se qui-quadrado) é um caso particular da distribuição gama, sendo muito empregada em estatística não paramétrica, uma vez que a estatística χ^2 , utilizada para verificação od ajuste de modelos probabilísticos teóricos a um conjunto de dados observados segue tal distribuição.

A função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{2^{(v/2)}\Gamma(v/2)} x^{(v/2-1)} e^{-(x/2)} \quad (4.29)$$

em que:

v são os graus de liberdade;

$\Gamma(n)$ é a função gama. Para n inteiro positivo, $\Gamma(n) = (n - 1)!$

Figura 4.8: Distribuições Qui-Quadrado com 1, 5 e 10 graus de liberdade

Esperança:

$$E(\chi_v) = v;$$

Variância:

$$V(\chi_v) = 2v$$

4.4.5 Distribuição t de Student

Viu-se que a variável $z = \frac{x-\mu}{\sigma} \sim N(0, 1)$. De modo semelhante, pode-se demonstrar que:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad (4.30)$$

Suponha-se que o parâmetro σ em 4.30 seja substituído por seu estimador não tendencioso

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

Assim a eq.4.30 ficará:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad (4.31)$$

Pode-se demonstrar que a variável t , 4.31 segue uma distribuição t de student com $v = n - 1$ graus de liberdade, cuja função densidade probabilidade é:

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} \quad (4.32)$$

em que:

v são os graus de liberdade;

$\Gamma()$ é a função Gama.

Esperança:

$$E(t) = 0;$$

Variância:

$$V(t) = \frac{v}{v-2}$$

Características:

- i. é simétrica em relação ao ponto $x = 0$ (média)
- ii. se v tende para infinito, t tende para z , como pode ser observado na figura 4.9

$$\lim_{v \rightarrow \infty} f(t) = z$$

4.4.6 Distribuição F de Snedcor

$$f(x) = \frac{\Gamma\left(\frac{v_1+2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{\left(\frac{v_1}{2}\right)} \frac{x^{\left(\frac{v_1-2}{2}\right)}}{\left[1 + \left(\frac{v_1}{v_2}\right)y\right]^{\left(\frac{v_1+v_2}{2}\right)}} \quad (4.33)$$

Figura 4.9: Distribuições t de student com 5 e 30 graus de liberdade e distribuição normal padronizada.

Esperança:

$$E(F) = \frac{v_2}{v_2 - 2};$$

Variância:

$$V(F) = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 4)(v_2)}$$

Figura 4.10: Distribuição F, com 10 graus de liberdade para o numerador e 20 para o denominador.

4.4.7 Aproximação da Distribuição Binomial à Normal

Os problemas relacionados com a distribuição Binomial são fáceis de serem resolvidos desde que o número de repetições (n) não seja grande, pois, quando n for grande, tais cálculos tornam-se demorados e tedioso e uma boa aproximação torna-se útil. Quando se utiliza a aproximação da distribuição Binomial à Normal, o erro cometido será tanto menor quanto maior for n e mais próximo de $\frac{1}{2}$ for p (probabilidade de sucesso). Alguns autores afirmam que a aproximação é considerada boa quando $np \geq 5$. Como a Distribuição Binomial é discreta, e a Normal contínua, ao realizar-se a aproximação deve-se fazer uma correção, chamada correção para descontinuidade da curva, que consiste em supor distribuída entre $x_i + 0,5$ e $x_i - 0,5$ a probabilidade concentrada em x_i . Assim,

$P(X = x_i)$ pela Binomial é aproximada para $P(x_i - 0,5 \leq X \leq x_i + 0,5)$ na Distribuição Normal

Exemplo.

Em um determinado processo de produção de chips para computador, 5% dos chips produzidos são considerados defeituosos. Sabendo-se que a produção diária da fábrica em questão é de 1000 chips, qual a probabilidade de que em um dia sejam produzidos:

a) 50 chips defeituosos?

$$X \sim B(1000, 0,05)$$

$$\text{Média} = E(X) = n \cdot p = (1000)(0,05) = 50 \text{ chips}$$

$$\text{Variância} = V(X) = npq = (1000)(0,05)(0,95) = 47,5 \text{ chip}^2$$

$$P(X = 50) \simeq P(49,5 \leq X \leq 50,5) = P(-0,07 \leq z \leq 0,07) = 0,0558$$

b) menos que 50 chips defeituosos

$$P(X \leq 50) \simeq P(X \leq 50,5) = P(z \leq 0,07) = 0,5279$$

Capítulo 5

Amostragem

5.1 Introdução.

5.1.1 Definições

- i. **População:** conjunto de indivíduos com pelo menos uma característica observável em comum.
- ii. **Amostra:** porção ou fração da população, retirada segundo algumas técnicas específicas, que matem as mesmas características de interesse da população.
- iii. **Parâmetro:** é uma medida associada à uma característica populacional Ex: Média (μ), variância (σ^2), etc.
- iv. **Estatística:** é uma medida associada à uma característica amostral. Ex: Média (\bar{x}), variância (s^2).

Um dos principais problemas apresentados na estatística é o de se fazer afirmações sobre os parâmetros populacionais (geralmente desconhecidos), como por exemplo saber qual o tempo necessário para o organismo humano degradar certo composto químico, qual a produção total de grãos de um país num determinado ano, qual a altura média da população brasileira, afirmar se um novo composto é carcinogênico ou não. E para respondermos a estas questões, muitas das vezes, temos que lançar mão

do processo de amostragem, que consiste em estudar apenas uma fração da população (a amostra) e a partir desta fazer inferências sobre a população. Esquemáticamente tem-se:

Figura 5.1: Representação esquemática do processo de amostragem e inferência.

Para que o processo anteriormente descrito seja confiável, é necessário que a amostra utilizada seja representativa da população, e para isso, ela deve ser retirada segundo determinadas técnicas de amostragem. De posse de uma amostra, representativa da população, para fazermos a inferência sobre os parâmetros populacionais, a partir desta amostra, é necessário o conhecimento das relações existentes entre as estimativas obtidas e os valores dos parâmetros populacionais, ou seja, é necessário conhecer a distribuição amostral do estimador utilizado, para que se possa fazer uma inferência segura sobre um parâmetro qualquer.

5.1.2 Importância do uso de amostras.

- i. Conveniente no estudo de populações grandes.
- ii. Indispensável no estudo de populações infinitas.
- iii. Indispensável em estudos nos quais a coleta de dados implica na destruição do material utilizado.

5.1.3 Vantagens do processo de amostragem em relação ao censo.

As principais vantagens do processo de amostragem, em relação a observação de todos os indivíduos da população (censo) são:

- i. **Custo reduzido:** Sendo os dados obtidos apenas de uma fração da população, as despesas são menores do que as oriundas de um censo. Tratando-se de grandes populações, pode-se obter resultados suficientemente precisos, para serem úteis, de amostras que representam apenas uma pequena fração da população. Segundo COCHRAN (1977), nos Estados Unidos, os mais importantes levantamentos periódicos, realizados pelo governo, usavam amostras de cerca de 100.000 pessoas, ou, aproximadamente uma pessoa em cada 1800.
- ii. **Maior rapidez:** Os dados podem ser apurados e sintetizados mais rapidamente em uma amostragem do que em uma contagem completa. Este é um fator primordial, quando se necessita urgentemente das informações. O objetivo de uma investigação, é o de conhecer a situação de um determinado fenômeno, no momento da coleta da informação, para que de acordo com a informação obtida, se possa tomar as medidas possíveis para resolver algum problema. Se o resultado dessa pesquisa for conhecida muito tempo depois, é bem possível que a situação que se pretendia resolver, seja nesse momento, completamente diferente da que existia no momento da coleta dos dados.
- iii. **Maior amplitude e flexibilidade:** Em certos tipos de investigação, tem-se que utilizar pessoal bem treinado e equipamento altamente especializado, cuja disponibilidade é limitada para a obtenção de dados. O censo completo torna-se impraticável e resta a escolha entre obter as informações por meio de uma amostra, ou não consegui-las de todo. Dessa forma, os levantamentos que se fundamentam na amostragem tem maior amplitude e flexibilidade, relativamente as informações que podem ser obtidas.
- iv. **Maior exatidão:** Em virtude de se poder empregar pessoal de melhor qualidade e intensivamente treinado, e por se tornar exequível a supervisão mais cuidadosa do campo de trabalho e do processamento de dados, dada a redução no volume de trabalho, uma amostragem pode, na realidade, proporcionar resultados mais exato que o censo.

5.2 Técnicas de amostragem.

Ao coletarmos uma amostra podemos fazê-la com reposição ou sem reposição, caso a amostragem seja realizada com reposição, um mesmo indivíduo tem chance de pertencer mais de uma vez a amostra, o que não acontece, no caso da amostragem ser sem reposição. Independentemente da maneira como a amostra é coletada (com ou sem reposição) o importante é que os indivíduos que compõem a amostra deverão ser selecionados através de um processo aleatório qualquer (sorteio), pois, somente nestas condições, podemos aplicar os modelos probabilísticos da estatística a esta amostra, o que vai garantir a validade dos testes estatísticos que serão realizados com base nos resultados destas amostras. Os principais tipos de amostragem são:

- Probabilísticas ou aleatória: Quando todos os indivíduos da população têm probabilidade conhecida e não nula de pertencer a amostra, dentre estas se destacam:
 - Amostragem simples ao acaso (ASA).
 - Amostragem sistemática (AS).
 - Amostragem por conglomerados.
 - Amostragem estratificada (AE).
- Não probabilística, não aleatória, escolha racional ou escolha justificada: Quando alguns indivíduos da população têm probabilidade desconhecida ou nula de pertencer a amostra, as principais são:
 - Inacessibilidade a toda a população.
 - Amostragem sem norma (a esmo).
 - População formada por material contínuo.
 - Intencional

5.2.1 Principais técnicas de amostragem probabilísticas.

5.2.1.1 Amostragem Simples ao Acaso

Esta técnica só pode ser aplicada em populações homogêneas e de tamanho conhecido. Técnica: Enumera-se todos indivíduos da população e sorteia-se (por meio de um dispositivo aleatório qualquer), os indivíduos que comporão a amostra. Neste tipo de amostragem podem ser retiradas N^n amostras diferentes com reposição ou C_N^n amostras diferentes sem reposição.

5.2.1.2 Amostragem Sistemática

É uma simplificação do processo anterior. Neste caso, apenas o primeiro elemento da amostra será sorteado, e os demais serão retirados em uma progressão aritmética, com razão k , em que:

$$k = \frac{N}{n},$$

com N = tamanho da população e n = tamanho da amostra até se completar o tamanho da amostra desejado.

5.2.1.3 Amostragem por Conglomerados

Quando uma população apresenta uma subdivisão natural em grupos menores (denominados conglomerados), sorteia-se um número suficiente desses grupos (conglomerados) e todos os elementos destes vão compor a amostra.

5.2.1.4 Amostragem Estratificada

É uma técnica utilizada quando a população a ser estudada é heterogênea, deste modo, subdivide-se a população em estratos (sub-populações) que sejam homogêneos dentro de si, e heterogêneos entre si, e aplica-se uma das técnicas de amostragens anteriormente descritas, para retirar-se sub-amostras dentro de cada estrato, de modo

que a amostra final seja representativa da população, como um todo (contenha indivíduos de todos os estratos). Quanto ao tamanho das sub-amostras retiradas (n_i), é classificada em:

i. Uniforme

Quando de K estratos, retiram-se amostras de mesmo tamanho n , independentemente do tamanho do estrato.

ii. Proporcional

Quando o tamanho da amostra retirado em cada estrato (n_i) é proporcional ao tamanho do estrato.

Ex. Para exemplificar os dois tipos de amostragem estratificada descritos, consideremos um estudo realizados em propriedades rurais de um município, composto por 1000 propriedades rurais, distribuídas, quanto a sua área, conforme a Tabela 1 e que neste município sejam amostradas 50 propriedades:

Tabela 5.1: Distribuição do número de propriedades rurais de um município qualquer, quanto a área e número de propriedades a serem amostradas por estrato (classes)

Área (ha)	Número de Propriedades	Amostra estratificada (N=50)	
		Uniforme	Proporcional
0- 20	500	10	25
20- 50	320	10	16
50- 100	100	10	5
100-200	50	10	3
200- 400	30	10	1
Total	1000	50	50

5.2.2 Principais técnicas de amostragem não probabilísticas.

5.2.2.1 Inacessibilidade a toda população

A amostragem é realizada na parte da população que é acessível.

Ex. Controle de qualidade numa linha de produção de cigarros. Só tem-se acesso aos cigarros que já estão prontos, embora os que ainda serão produzidos fazem parte da população de cigarros produzidos por aquela linha produção.

5.2.2.2 Amostragem sem norma (a esmo)

Não se utiliza nenhum sorteio, para identificar a amostra, muito embora o amostrador procure ser aleatório.

Ex. Amostrar 80 frangos num galpão com 3000 frangos, amostrar peixes em um lago, pessoas em uma praça, etc.

5.2.2.3 População formada por material contínuo.

Processo utilizado para se amostrar líquidos, gases ou sólidos. Homogeniza-se o material a ser amostrado e em seguida colhe-se a amostra.

5.2.2.4 Intencional

O pesquisador escolhe deliberadamente certos elementos da população para formar a amostra, baseado num pré-julgamento.

Ex. Pesquisa de mercado para lançar uma nova marca de leite longa vida tipo A. O pesquisador selecionará indivíduos com poder aquisitivo médio/alto, que são os principais consumidores deste produto (público alvo), embora toda a população independentemente do poder aquisitivo possa ser consumidora deste produto.

5.3 Distribuições Amostrais

Figura 5.2: Representação esquemática da distribuição amostral de um estimador.

5.3.1 Distribuição amostral da média

Considere-se, a título de exemplo, uma população hipotética, formada por três indivíduos, para os quais a variável de interesse (X), seja a nota final destes indivíduos na disciplina estatística, a qual segue uma distribuição uniforme discreta como apresentado a seguir:

X	8	9	10	
$P(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\Sigma = 1$

Figura 5.3: Distribuição das notas de três alunos.

Neste caso tem-se:

$$N = 3;$$

$$E(X) = \mu = 9;$$

$$V(X) = \sigma^2 = \frac{2}{3}.$$

5.3.1.1 Amostragem com reposição

Retirando-se todas as possíveis amostras com reposição, de tamanho $n = 2$, tem-se um total de $3^2 = 9$ possíveis amostras, as quais estão apresentadas a seguir:

Amostra	Indivíduos	Notas	\bar{x}
1	1;1	8 e 8	8
2	1;2	8 e 9	8,5
3	1;3	8 e 10	9
4	2;1	9 e 8	8,5
5	2;2	9 e 9	9
6	2;3	9 e 10	9,5
7	3;1	10 e 8	9
8	3;2	10 e 9	9,5
9	3;3	10 e 10	10

A distribuição amostral de \bar{x} será:

\bar{x}	8	8,5	9	9,5	10
$P(\bar{x} = \bar{x}_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Figura 5.4: Distribuição amostral de \bar{x} .

Em que:

$$E(\bar{x}) = \mu_{\bar{x}} = \sum_{i=1}^n \bar{x}_i P(\bar{x} = \bar{x}_i) = 9,0$$

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \sum_{i=1}^n [\bar{x}_i - E(\bar{x})]^2 P(\bar{x} = \bar{x}_i) = \frac{1}{3}$$

Assim, verifica-se que:

$$E(\bar{x}) = 9,0 = \mu \text{ e } V(\bar{x}) = \frac{1}{3} = \frac{\sigma^2}{n}$$

Prova:

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

$$\begin{aligned}
E(\bar{x}) &= E\left[\frac{1}{n}\sum_{i=1}^n x_i\right] \\
&= \frac{1}{n}E[x_1 + x_2 + \cdots + x_n] \\
&= \frac{1}{n}[E(x_1) + E(x_2) + \cdots + E(x_n)] \\
&= \frac{1}{n}[\mu + \mu + \cdots + \mu] \\
&= \frac{1}{n}n\mu \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
V(\bar{x}) &= V\left[\frac{1}{n}\sum_{i=1}^n x_i\right] \\
&= \frac{1}{n^2}V[x_1 + x_2 + \cdots + x_n] \\
&= \frac{1}{n^2}[V(x_1) + V(x_2) + \cdots + V(x_n)] \\
&= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \cdots + \mu] \\
&= \frac{1}{n^2}n\sigma^2 \\
&= \sigma^2
\end{aligned}$$

Então tem-se que: $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

5.3.1.2 Amostragem sem reposição

Retirando-se todas as possíveis amostras sem reposição, de tamanho $n = 2$, tem-se um total de 6 possíveis amostras, as quais estão apresentadas a seguir:

Amostra	Indivíduos	Notas	\bar{x}
1	1;2	8 e 9	8,5
2	1;3	8 e 10	9
3	2;1	9 e 8	8,5
4	2;3	9 e 10	9,5
5	3;1	10 e 8	9
6	3;2	10 e 9	9,5

A distribuição amostral de \bar{x} será:

\bar{x}	8,5	9	9,5
$P(\bar{x} = \bar{x}_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Em que:

$$E(\bar{x}) = \mu_{\bar{x}} = \sum_{i=1}^n \bar{x}_i P(\bar{x} = \bar{x}_i) = 9,0$$

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \sum_{i=1}^n [\bar{x}_i - E(\bar{x})]^2 P(\bar{x} = \bar{x}_i) = \frac{1}{6}$$

Assim, verifica-se que:

$$E(\bar{x}) = 9,0 = \mu \text{ e } V(\bar{x}) = \frac{1}{6} = \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

Deste modo, se amostragem for sem reposição, $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n} \frac{N-n}{N-1}\right)$.

O termo $\frac{N-n}{N-1}$ é conhecido como fator de correção para amostragem sem reposição em populações finitas (ASRPF). Uma população é considerada finita quando $\frac{n}{N} > 0,05$ ou seja a amostra representar mais de 5% do tamanho da população. Quando tal critério não for satisfeito, o fator de correção torna-se desprezível, podendo, portanto ser eliminado.

Capítulo 6

Inferência

6.1 Teoria da estimação

6.1.1 Definições

Estimador

Consideremos uma amostra $(x_1, x_2, x_3, \dots, x_n)$ de uma variável aleatória que deve descrever uma característica de interesse da população. Seja θ um parâmetro que desejamos estimar, como por exemplo a média $\mu = E(x)$ ou a variância $\sigma^2 = V(x)$. Um estimador, $\hat{\theta}$, do parâmetro θ é uma variável aleatória, que é função das observações $x_1, x_2, x_3, \dots, x_n$.

Assim,

$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ é um estimador da média populacional μ ,

$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ é um estimador da variância populacional σ^2

Estimativa

Estimativa é o valor numérico assumido pelo estimador quando os valores observados $x_1, x_2, x_3, \dots, x_n$ são considerados.

Assim,

$\bar{x} = 70kg$ é uma estimativa da média populacional μ ,

$s^2 = 9kg^2$ é uma estimativa da variância populacional σ^2

Estimação por ponto e por intervalo.

Quando a estimativa de um parâmetro populacional é dada por um único valor, tem-se uma estimativa pontual do parâmetro populacional, desconhecido, como por exemplo a altura média de uma amostra de 500 universitários é $\bar{x} = 1,68m$, é uma estimativa pontual da verdadeira altura média da população de universitários. Porém sabe-se que $\bar{x} \sim N(\mu; \frac{\sigma^2}{n})$, assim sendo, para cada amostra retirada da população, poderá se obter uma diferente estimativa para μ . Deste modo, torna-se mais interessante obter-se, a partir, de uma determinada amostra, um intervalo que apresente uma probabilidade conhecida de conter o verdadeiro parâmetro populacional, ou seja obter uma estimativa por intervalo para o parâmetro em questão, como por exemplo $P(1,60 \leq \mu \leq 1,76) = 0,95$, ou seja existe 0,95 de probabilidade de que a verdadeira média populacional esteja entre 1,60 e 1,76 metros, ou ainda existe 95% de confiança em se afirmar que a verdadeira média populacional esteja entre 1,60 e 1,76 metros. Apesar disto, o uso de estimativas pontuais é imprescindível, haja vistas, serem necessárias para a obtenção das estimativas por intervalo. Deste modo desejável que estas estimativas sejam bastantes confiáveis, e para isso é necessário que os estimadores que as fornecerão apresentem boas propriedades, aliado ao fato de serem obtidas a partir de amostras representativas.

6.1.2 Propriedades dos Estimadores

6.1.2.1 Não tendenciosidade

Um estimador $\hat{\theta}$ é dito um estimador não tendencioso do parâmetro θ se

$$E(\hat{\theta}) = \theta$$

obs. Os termos não tendencioso, não viciado, não viesado e imparcial são sinônimos.

Ex1.: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ é um estimador não tendencioso da média populacional μ

prova:

$$\begin{aligned}
 E(\bar{x}) &= E\left[\frac{\sum_{i=1}^n x_i}{n}\right] \\
 &= \frac{1}{n} E\left[\sum_{i=1}^n x_i\right] \\
 &= \frac{1}{n} E[x_1 + x_2 + \cdots + x_n] \\
 &= \frac{1}{n} [E(x_1) + E(x_2) + \cdots + E(x_n)] \\
 &= \frac{1}{n} [\mu + \mu + \cdots + \mu] \\
 &= \frac{1}{n} n\mu \\
 &= \mu
 \end{aligned}$$

Ex2.: $s^{2*} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ é um estimador tendencioso da variância populacional σ^2 .

prova:

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i - \mu + \mu - \bar{x})^2 \\
 &= \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2 \\
 &= \sum_{i=1}^n (x_i - \mu)^2 - 2 \sum_{i=1}^n (x_i - \mu)(\bar{x} - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \\
 &= \text{como } (\bar{x} - \mu) \text{ é uma constante e } \sum_{i=1}^n (x_i - \mu) = n(\bar{x} - \mu), \text{ tem-se:} \\
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2
 \end{aligned}$$

Portanto,

$$\begin{aligned}
 E [s^{2*}] &= E \left[\frac{\sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2}{n} \right] \\
 &= \frac{1}{n} \left\{ \sum_{i=1}^n E [(x_i - \mu)^2] - nE [(\bar{x} - \mu)^2] \right\} \\
 &= \frac{1}{n} \{nV(X) - nV(\bar{x})\} \\
 &= \frac{1}{n} \left\{ n\sigma^2 - n\frac{\sigma^2}{n} \right\} \\
 &= \frac{n-1}{n}\sigma^2
 \end{aligned}$$

Deste modo, verifica-se que s^{2*} é um estimador tendencioso de σ^2 . UM estimador não tendencioso é facilmente obtido por:

$$s^2 = \frac{n}{n-1}s^{2*} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (6.1)$$

6.1.2.2 Consistência.

Um estimador $\hat{\theta}$ é um estimador consistente do parâmetro θ se:

- i. $\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta$;
- ii. $\lim_{n \rightarrow \infty} V(\hat{\theta}) = 0$.

$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ é um estimador consistente da média populacional μ , pois

- i. $E(\bar{x}) = \mu$
- ii. $\lim_{n \rightarrow \infty} V(\hat{x}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$.

6.1.2.3 Eficiência

Se θ_1 e θ_2 são dois estimadores não tendenciosos de θ , então, θ_1 é mais eficiente que θ_2 se:

$$V(\theta_1) < V(\theta_2)$$

Eficiência relativa

A eficiência relativa do estimador θ_1 , em relação ao estimador θ_2 é dada por:

$$Ef_{\theta_1, \theta_2} = \frac{V(\theta_2)}{V(\theta_1)} \quad (6.2)$$

6.1.3 Intervalos de confiança

Conhecendo-se a distribuição amostral do estimador, de um parâmetro θ , pode-se facilmente determinar um intervalo que apresente uma confiança $1 - \alpha$ para θ , como será visto a seguir.

6.1.3.1 Intervalo de confiança para a média μ

6.1.3.1.1 Variância conhecida Sabe-se que $\bar{x} \sim N(\mu; \frac{\sigma^2}{n})$, assim a variável $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ terá distribuição $N(0; 1)$. Fixando-se um nível de confiança $(1 - \alpha)$ virá:

$$\begin{aligned} P(-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(-z_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \\ P(-\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \\ P(\bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \text{ reorganizando vem} \\ P(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \end{aligned}$$

E o intervalo de confiança para μ , com uma confiança $1 - \alpha$ pode ser então escrito como:

$$IC(\mu)_{1-\alpha} = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (6.3)$$

em que

n é o tamanho da amostra.

Obs. Se ocorrer amostragem sem reposição em população finita (ASRPF) o intervalo de confiança para a média será:

$$IC(\mu)_{1-\alpha} = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (6.4)$$

onde:

N é o tamanho da população;

n é o tamanho da amostra.

Ex.: Uma máquina produz rolamentos que apresentam desvio padrão de 0,042 *polegadas* em seu diâmetro. Desejando-se conhecer o diâmetro médio dos rolamentos produzidos por esta máquina, extraiu-se uma amostra de 100 rolamentos, observando-se uma média igual a 0,824 *polegadas*. Obter o intervalo com 0,90 de confiança para o verdadeiro diâmetro médio dos rolamentos.

Solução:

Tem-se $\bar{x} = 0,824$ $\sigma = 0,042$ $n = 100$ $1 - \alpha = 0,90$ substituindo esses valores em 6.3 vem:

$$\begin{aligned} IC(\mu)_{0,90} &= 0,824 \pm z_{0,05} \frac{0,042}{\sqrt{100}} \\ &= 0,824 \pm 1,65 \frac{0,042}{\sqrt{100}} \\ &= 0,824 \pm 0,007 \end{aligned}$$

Interpretação: Como μ é um parâmetro e não uma variável aleatória, a interpretação correta do intervalo de confiança é: Construídos todos os intervalos do tipo $\bar{x} \pm 1,65 \frac{\sigma}{\sqrt{n}}$, 90% deles conterão o parâmetro μ . Na prática, apenas um único intervalo é construído, no presente exemplo tal intervalo foi [0,817; 0,831]. Esse intervalo é então comumente chamado intervalo de confiança de 90% para μ . Isto é tem-se 90% de confiança de que esse intervalo contenha o valor μ , no sentido de que 90% dos intervalos assim construídos conteriam μ .

É obviamente incorreto, do ponto de vista da estatística clássica ou freqüentista, dizer que a probabilidade do intervalo $[0,817; 0,831]$ conter o valor μ é 0,90. Pois essa probabilidade é 0 ou 1, dependendo de μ pertencer ou não ao intervalo ao intervalo fixo.

6.1.3.1.2 Variância desconhecida Quando não se conhece σ^2 e conseqüentemente σ , mas sim sua estimativa s , o intervalo de confiança para a média será dado por:

Amostras Pequenas ($n \leq 30$)

$$IC(\mu)_{1-\alpha} = \bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \quad (6.5)$$

$t_{\frac{\alpha}{2}}$ com $n - 1$ graus de liberdade,

em que:

n é o tamanho da amostra.

Obs. Se ocorrer amostragem sem reposição em população finita (ASRPF) o intervalo de confiança para a média será:

$$IC(\mu)_{1-\alpha} = \bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \quad (6.6)$$

$t_{\frac{\alpha}{2}}$ com $n - 1$ graus de liberdade,

onde:

N é o tamanho da população;

n é o tamanho da amostra.

Amostras Grandes ($n > 30$)

Foi visto que à medida que aumenta-se o tamanho da amostra, a distribuição t se Student se aproxima da distribuição normal, deste modo, quando se estiver trabalhando com amostras grandes ($n > 30$) pode-se utilizar a distribuição normal padronizada,

z , em lugar da t na obtenção dos intervalos de confiança, mesmo que σ^2 seja desconhecida.

Ex.: Um Cia adquiriu 500 cabos. Uma amostra de 30 deles selecionados ao acaso apresentou tensão de ruptura media igual a 2400 *kg* com desvio padrão de 150 *kg*. Obter o intervalo com 95% de confiança para a verdadeira tensão media de ruptura destes cabos.

solução:

Tem-se: $N = 500$ $n = 30$ $\bar{x} = 2400$ $s = 150$ $1 - \alpha = 0,95$

$\frac{n}{N} = \frac{30}{500} = 0,06 > 0,05 \rightarrow$ ocorreu ASRPF.

$$\begin{aligned} IC(\mu)_{0,95} &= 2400 \pm t_{0,025} \frac{150}{\sqrt{30}} \sqrt{\frac{500-30}{500-1}} \\ &= 2400 \pm (2,045)(27,38)(0,97) \\ &= 2400 \pm 54,31 \end{aligned}$$

Interpretação: Existe 95% de confiança em se dizer que a verdadeira tensão media de ruptura dos cabos está entre 2345,69 e 2454,31kg.

6.1.3.2 Diferença entre duas média ($\mu_a - \mu_b$)

.

6.1.3.2.1 Variâncias Conhecidas:

$$IC(\mu_a - \mu_b)_{1-\alpha} = \bar{x}_a - \bar{x}_b \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} \quad (6.7)$$

em que:

\bar{x}_a e \bar{x}_b são as estimativas pontuais das médias das populações a e b , respectivamente;

σ_a^2 e σ_b^2 as variâncias das populações a e b , respectivamente e

n_a e n_b os tamanhos das amostras das populações a e b , respectivamente.

Obs: Se ocorrer ASRPF deve-se multiplicar a variância da população na qual ocorreu ASRPF pelo fator de correção $\frac{N-n}{N-1}$.

Ex.: As empresas A e B produzem tubos para esgoto com a variâncias em seus diâmetros iguais a $8mm^2$ e $10mm^2$, respectivamente. Uma amostra de 48 tubos da empresa A apresentou diâmetro médio igual a $40mm$, e uma amostra de 36 tubos da empresa B apresentou diâmetro médio de $42mm$. Verifique, por meio de um intervalo de confiança com 0,95 de probabilidade, se existe diferença entre os diâmetros médios dos tubos das marcas A e B.

Solução:

Pop. A	Pop. B
$\sigma_A^2 = 8$	$\sigma_B^2 = 10$
$\bar{x}_A = 40$	$\bar{x}_B = 42$
$n_A = 48$	$n_B = 36$

$$\begin{aligned}
 IC(\mu_a - \mu_b)_{0,95} &= \bar{x}_a - \bar{x}_b \pm z_{0,025} \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} \\
 &= 40 - 42 \pm 1,96 \sqrt{\frac{8}{40} + \frac{10}{36}} \\
 &= -2 \pm 1,2973
 \end{aligned}$$

Conclusão: Pode-se afirmar com 95% de confiança que a verdadeira diferença entre os diâmetros médios dos tubos produzidos pelas empresas A e B está entre $-2 \pm 1,2973mm$, isto é entre $-3,2973$ e $-0,7027 mm$. Como esse intervalo não compreende o valor 0 (zero) Tem-se 95% de confiança em afirmar que os diâmetros médios dos tubos produzidos por estas empresas não são iguais.

6.1.3.2.2 Variâncias Desconhecidas: Quando desconhece-se as variâncias populacionais (σ_a^2 e σ_b^2) torna-se necessário a substituição de seus valores paramétricos por suas estimativas amostrais (s_a^2 e s_b^2). Neste caso, deve-se utilizar a distribuição t de Student, em lugar da normal. Além desta alteração deve-se considerar ainda se as duas populações são homocedásticas ou heterocedásticas, isto é, se as variâncias populacionais (desconhecidas) são iguais ou diferentes, o que pode ser aferido por meio de um teste de hipótese para homogeneidade das variâncias (Cap 7).

Populações homocedásticas

Sendo as populações homocedásticas ($\sigma_a^2 = \sigma_b^2 = \sigma^2$), assim, s_a^2 e s_b^2 são duas estimativas para um mesmo parâmetro (σ^2) então o intervalo de confiança para a diferença entre duas médias é dado por:

$$IC(\mu_a - \mu_b)_{1-\alpha} = \bar{x}_a - \bar{x}_b \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}, \quad (6.8)$$

$t_{\frac{\alpha}{2}}$ com $n_a + n_b - 2$ graus de liberdade.

em que:

$$s_p = \sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}}$$

Populações heterocedásticas

Sendo as populações heterocedásticas ($\sigma_a^2 \neq \sigma_b^2$), assim, s_a^2 e s_b^2 são estimativas de diferentes parâmetros, não podendo, pois serem combinadas em um único valor. Então o intervalo de confiança para a diferença entre duas médias é dado por:

$$IC(\mu_a - \mu_b)_{1-\alpha} = \bar{x}_a - \bar{x}_b \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}} \quad (6.9)$$

$t_{\frac{\alpha}{2}}$ com v graus de liberdade.

em que:

$$v = \frac{\left(\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}\right)^2}{\frac{\left(\frac{s_a^2}{n_a}\right)^2}{n_a-1} + \frac{\left(\frac{s_b^2}{n_b}\right)^2}{n_b-1}}$$

6.1.3.3 Intervalo de confiança para proporção

6.1.3.3.1 Amostras grandes ($n > 30$) O intervalo de confiança para a proporção é dado por:

$$IC(P)_{1-\alpha} = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (6.10)$$

em que:

\hat{p} é a proporção estimada na amostra;

$\hat{q} = 1 - \hat{p}$ e;

n é o tamanho da amostra.

Obs: Se ocorrer ASRPF, o intervalo de confiança para proporção é dado por:

$$IC(P)_{1-\alpha} = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \sqrt{\frac{N-n}{N-1}} \quad (6.11)$$

6.1.3.3.2 Amostras pequenas ($n \leq 30$) Quando a amostra for pequena deve-se utilizar a distribuição t de Student, em lugar da normal e o intervalo de confiança para a proporção será dado então por:

$$IC(P)_{1-\alpha} = \hat{p} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \quad (6.12)$$

$t_{\frac{\alpha}{2}}$ com $n - 1$ graus de liberdade

Obs: Se ocorrer ASRPF, o intervalo de confiança para proporção é dado por:

$$IC(P)_{1-\alpha} = \hat{p} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \sqrt{\frac{N-n}{N-1}} \quad (6.13)$$

$t_{\frac{\alpha}{2}}$ com $n - 1$ graus de liberdade

6.1.3.4 Intervalo de confiança para a diferença entre proporções

Dadas duas amostras independentes, de populações diferentes, o intervalo de confiança para a diferença entre as proporções nestas populações é dado por:

6.1.3.4.1 Amostras grandes ($n > 30$)

$$IC(P_a - P_b)_{1-\alpha} = (\hat{p}_a - \hat{p}_b) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_a \hat{q}_a}{n_a} + \frac{\hat{p}_b \hat{q}_b}{n_b}} \quad (6.14)$$

em que:

\hat{p}_a é a proporção estimada na amostra;

$$\hat{q}_a = 1 - \hat{p}_a;$$

$$\hat{q}_b = 1 - \hat{p}_b;$$

n_a e n_b são os tamanhos das amostras a e b, respectivamente

Obs: Se ocorrer ASRPF, deve-se multiplicar o componente da variância, referente à população na qual ocorreu ASRPF pelo fator de correção $\frac{N-n}{N-1}$.

6.1.3.4.2 Amostras pequenas ($n \leq 30$)

$$IC(P_a - P_b)_{1-\alpha} = (\hat{p}_a - \hat{p}_b) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_a \hat{q}_a}{n_a} + \frac{\hat{p}_b \hat{q}_b}{n_b}} \quad (6.15)$$

$t_{\frac{\alpha}{2}}$ com $n_a + n_b - 2$ graus de liberdade

Obs: Se ocorrer ASRPF, deve-se multiplicar o componente da variância, referente à população na qual ocorreu ASRPF pelo fator de correção $\frac{N-n}{N-1}$.

6.1.3.5 Intervalo de confiança para a variância (σ^2)

O intervalo de confiança para a variância populacional é dado por:

Sabe-se que

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Então,

$$P \left[\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2} \right] = 1 - \alpha$$

E o intervalo de confiança para a variância será:

$$IC(\sigma^2)_{1-\alpha} = \left[\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2}; \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2} \right] \quad (6.16)$$

6.2 Teoria da decisão (Testes de Hipóteses)

Uma hipótese científica é qualquer afirmação que possa ser refutada, caso contrário pertencerá a outro ramo do conhecimento humano, como por exemplo a religião. Assim sendo, a hipótese: “Os motores da marca x são mais econômicos que os da marca y” é uma hipótese científica, pois qualquer pessoa que duvide, ou queira comprová-la, pode montar um experimento e averiguar sua veracidade. Por outro lado, a hipótese: “Deus existe”, não pode ser avaliada, não sendo, portanto, científica. Uma determinada hipótese é tida como verdadeira, se em sua avaliação não forem encontrados indícios que a desaprovem, permanecendo assim até que se prove o contrário. Para que uma hipótese científica seja testada, ela deve ser convertida em uma hipótese estatística, que é uma afirmação sobre um parâmetro populacional. Um teste de hipótese, fundamenta-se em um conjunto de regras, que permitem, a partir dos resultados experimentais (amostrais) rejeitar ou não tal hipótese, associando a esta decisão uma determinada confiança.

6.2.1 Metodologia de um teste de hipótese

Para a realização de um teste de hipóteses, deve-se formular duas hipóteses estatísticas, a saber:

- Hipótese de nulidade (H_0) é a hipótese que será testada, sendo geralmente formulada com o intuito de ser rejeitada.

- Hipótese alternativa (H_a) é qualquer hipótese que contrarie H_0 .

Suponha que esteja-se interessado em verificar se a verdadeira performance (*km/litro de combustível*) dos veículos, de determinada marca, equipados com motores 1.6 c.c. seja de 14km/l , como afirma o fabricante, ou se este é inferior a 14km/l . Então deve-se formular as seguintes hipótese estatísticas:

$$\begin{cases} H_0 : \mu = 14\text{km/l} \\ H_a : \mu < 14\text{km/l} \end{cases}$$

Para verificar a veracidade da hipótese H_0 , deve-se conduzir um experimento (coletar uma amostra), no qual será medida a performance de vários carros, que fornecerão uma estimativa da performance média, e sua variância, a partir das quais, verifica-se a veracidade da hipótese H_0 . Suponha que no experimento acima tenham sido avaliados 9 carros, e que estes tenham apresentado uma performance média de 13 km/l , com variância $4(\text{Km/l})^2$. Pelo simples fato desta amostra de 9 carros ter apresentado uma performance média inferior a informada pelo fabricante (14 km/l), não se pode concluir que esta afirmativa seja falsa, pois como já é sabido, esta estimativa está sujeita uma distribuição amostral. Deste modo, para verifica a veracidade de H_0 , assume-se que esta hipótese seja verdadeira, isto é $\mu = 14\text{ km/l}$. e calcula-se a probabilidade de uma amostra, com tamanho $n = 9$, retirada desta população, fornecer uma estimativa inferior a estimativa obtida (13 km/l). Caso esta probabilidade seja alta, não haverá nenhuma razão para rejeitar a hipótese H_0 (isto é duvidar de sua veracidade), sendo esta tida como verdadeira. Nesta situação disse que a diferença observada entre a média amostral (13 km/l) e a populacional (14 km/l) não é significativa, daí a terminologia usual de que “o teste foi não significativo”, usada para dizer que a hipótese H_0 não foi rejeitada. Por outro lado, se a probabilidade de se obter esta estimativa for pequena ($p < 0,05$) há razões para acreditar que a verdadeira média populacional seja menor do que se imaginava, ou seja a verdadeira performance deve ser menor que 14 km/l . Nesta situação, diz-se que a diferença foi significativa, portanto a hipótese H_0 deve ser rejeitada (o teste foi significativo). Obs: Não existe nenhum argumento

científico para se fixar o nível de probabilidade limite de um teste em 0,05. Este é apenas um valor usual, devido a facilidade de sua obtenção em tabelas. No nosso exemplos temos:

$$\begin{cases} H_0 : \mu = 14km/l \\ H_a : \mu < 14km/l \end{cases}$$

na amostra de $n = 9$ carros obteve-se $\bar{x} = 13 km/l$ e $s^2 = 4 (km/l)^2$; sabendo-se que \bar{x} sin $N\left(\mu, \frac{\sigma^2}{n}\right)$, assumido $\mu = 14 km/l$, e como não se conhece σ^2 , mas sim s^2 , tem-se:

$$\bar{x} \text{ sin } t_{(8)}\left(14, \frac{4}{9}\right)$$

gráfico

$$t_c = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{13 - 14}{\frac{2}{\sqrt{9}}} = -1,5$$

Então,

$$P|_{H_0}(\bar{x} \leq 13) = P(t \leq -1,5) = 0,1720$$

como esta probabilidade é alta, não há razões para acreditar que a verdadeira performance média seja inferior a 14 km/l

6.2.2 Tipos de erros

Ao realizar-se um teste de hipótese, pode-se incorrer em dois tipos de erros, que serão discutidos a seguir. Suponha que a hipótese H_0 formulada, no exemplo anterior seja verdadeira, isto é a performance média dos carros realmente é de 14 km/l, isto é ($\mu = 14 km/l$), e por efeito de acaso obtenha-se, na amostra, uma estimativa de performance, cuja probabilidade de ocorrência seja muito baixa, o que levaria a rejeição da hipótese $H_0 : \mu = 14 km/l$, que é verdadeira. Então ter-se-a cometido um erro denominado erro Tipo I (rejeitar uma hipótese H_0) verdadeira. A probabilidade de

se cometer este erro é denominada nível de significância (α) sendo esta, determinada (fixada) pelo pesquisador. Por outro lado, a hipótese formulada pode ser falsa, isto é na verdade $\mu \neq 14 \text{ km/l}$, e por efeito de acaso obter uma estimativa, que nos leve a não rejeição da hipótese $H_0 : \mu = 14 \text{ km/l}$. Nesta situação ter-se-a cometido o erro Tipo II (aceitar H_0 falsa). A probabilidade de cometer este erro é (β), sendo esta uma função de α , H_0 e do tamanho amostral. As probabilidades de se cometer os erros Tipo I e Tipo II, (α e β) são inversamente proporcionais, como pode ser observado na fig?, sendo que, a única maneira de se diminuir simultaneamente α e β é aumentando o tamanho amostral (n).

Figura 6.1: Erros Tipo I e Tipo II.

Figura tipos de erros

Os tipos de erros que podem ser cometidos em um teste de hipóteses, bem como suas probabilidades estão resumidos na tabela 6.1

Tabela 6.1: Tipos de erros passíveis de serem cometidos ao se testar uma hipótese

Realidade	Decisão	
	Rejeita H_0	Não Rejeita H_0
H_0 verdadeira	α (erro Tipo I)	$1 - \alpha$ Decisão correta
H_0 falsa	$1 - \beta$ Decisão correta	β (erro Tipo II)

6.2.3 Tipos de testes

De acordo com o tipo de hipótese formulada pode-se ter os seguintes tipos de testes de hipóteses:

- i. Teste Bilateral: Apresenta duas regiões de rejeição de da hipótese H_0 , situadas nos extremos da distribuição amostral, é utilizado para testar as hipóteses do tipo:

$$\begin{cases} H_0 : \mu = \theta \\ H_a : \mu \neq \theta \end{cases}$$

- ii. Teste Unilateral a Direita: Apresenta uma única região de rejeição da hipótese H_0 , situada no extremo superior da distribuição amostral, é utilizado para testar as hipóteses do tipo:

$$\begin{cases} H_0 : \mu = \theta \\ H_a : \mu > \theta \end{cases}$$

- iii. Teste Unilateral a Esquerda Apresenta: uma região de rejeição da hipótese H_0 , situada no extremo inferior da distribuição amostral, é utilizado para testar as hipóteses do tipo:

$$\begin{cases} H_0 : \mu = \theta \\ H_a : \mu < \theta \end{cases}$$

6.2.4 Algoritmo para realização de um teste de hipótese

- i. Formular as hipótese H_0 e H_a ;
- ii. Fixar o valor de α ;
- iii. Construir a regra de decisão (regiões de rejeição e não rejeição de H_0);
- iv. Calcular a estatística adequada para o teste;
- v. Tomar a decisão;
- vi. Conclusão.

Exemplo. Aplicando-se este algoritmo ao exemplo da performance média dos carros tem-se:

i.

$$\begin{cases} H_0 : \mu = 14km/l \\ H_a : \mu < 14km/l \end{cases}$$

ii. $\alpha = 0,05$ iii. $t_{0,05}(8) = 1,860$ Regra: rejeitar H_0 se $t_{calc} \leq -1,860$

iv.

$$t_{calc} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{13 - 14}{\frac{2}{\sqrt{9}}} = -1,5$$

v. Como $t_{calc} = -1,5 > t_{tab} = -1,860$, não rejeita-se a hipótese H_0 , pois o valor da estatística teste (t_{calc}) encontra-se na região de não rejeição de H_0 .

vi. Conclui-se pelo teste t de Student, ao nível de 0,05 de probabilidade que a verdadeira performance média destes carros não é inferior a 14 km/l.

6.2.5 Estatística apropriadas para os testes de hipóteses

· Média:

Amostras grandes

$$z_{calc} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \quad (6.17)$$

Amostras pequenas

$$t_{calc} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \quad (6.18)$$

, $v=n-1$ graus de liberdade

· Diferença entre duas médias (amostras independentes)

Amostras grandes

$$z_{calc} = \frac{\bar{x}_a - \bar{x}_b - \mu_a - \mu_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}} \quad (6.19)$$

Amostras pequenas

Variâncias diferentes ($\sigma_a^2 \neq \sigma_b^2$)

$$t_{calc} = \frac{\bar{x}_a - \bar{x}_b - \mu_a - \mu_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}} \quad (6.20)$$

; com

$$v = \frac{\left(\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}\right)^2}{\frac{\left(\frac{s_a^2}{n_a}\right)^2}{n_a-1} + \frac{\left(\frac{s_b^2}{n_b}\right)^2}{n_b-1}}$$

graus de liberdade, sendo v o maior inteiro que não exceda ao valor calculado.

Variâncias iguais ($\sigma_a^2 = \sigma_b^2$)

$$t_{calc} = \frac{\bar{x}_a - \bar{x}_b - \mu_a - \mu_b}{s_p \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}} \quad (6.21)$$

$v = n_a + n_b - 2$ com graus de liberdade

$$s_p = \frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}$$

Diferença entre duas médias (amostras dependentes, dados pareados, “antes e depois”)

Amostras grandes

$$z_{calc} = \frac{\bar{d} - \mu_0}{\frac{s_d}{\sqrt{n}}} \quad (6.22)$$

em que:

$$d_i = x_{i_{antes}} - x_{i_{depois}}$$

Amostras pequenas

$$t_{calc} = \frac{\bar{d} - \mu_0}{\frac{s_d}{\sqrt{n}}} \quad (6.23)$$

, em que:

$$d_i = x_{i_{antes}} - x_{i_{depois}}$$

$v = n - 1$ graus de liberdade

· Proporção

Amostras grandes

$$z_{calc} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \quad (6.24)$$

em que:

$$\hat{q} = 1 - \hat{p}$$

Amostras pequenas

$$t_{calc} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \quad (6.25)$$

em que:

$$\hat{q} = 1 - \hat{p}$$

$v = n - 1$ graus de liberdade

· Diferença entre duas proporções

Amostras grandes

$$z_{calc} = \frac{\hat{p}_a - \hat{p}_b - p_a - p_b}{\sqrt{\frac{\hat{p}_a\hat{q}_a}{n_a} + \frac{\hat{p}_b\hat{q}_b}{n_b}}} \quad (6.26)$$

em que:

$$\hat{q}_a = 1 - \hat{p}_a$$

e

$$\hat{q}_b = 1 - \hat{p}_b$$

Amostras pequenas

$$z_{calc} = \frac{\hat{p}_a - \hat{p}_b - p_a - p_b}{\sqrt{\frac{\hat{p}_a\hat{q}_a}{n_a} + \frac{\hat{p}_b\hat{q}_b}{n_b}}} \quad (6.27)$$

em que:

$$\hat{q}_a = 1 - \hat{p}_a$$

e

$$\hat{q}_b = 1 - \hat{p}_b$$

com

$$v = n_a + n_b - 2$$

· Razão entre variâncias

$$f = \frac{\sigma_b^2 s_a^2}{\sigma_a^2 s_b^2} \quad (6.28)$$

$v_1 = n_a - 1$ graus de liberdade para o numerador

$v_2 = n_b - 1$ graus de liberdade para o denominador

6.2.6 Teste de Qui-Quadrado (χ^2)

O teste de Qui-Quadrado é utilizado para comparação entre as frequências observadas as esperadas segundo um modelo probabilístico qualquer.

Uma medida da discrepância entre as frequências observadas e esperadas é dada por:

$$\chi_{calc}^2 = \sum_{i=1}^k \frac{(F_{o_i} - F_{e_i})^2}{F_{e_i}}; \quad (6.29)$$

em que:

F_{o_i} é a frequência observada;

F_{e_i} é a frequência esperada.

A expressão 6.29 fornece um valor sempre positivo, e pode-se demonstrar que $\chi_{calc}^2 \sim \chi_v^2$. Em que χ_v^2 é uma distribuição Qui-Quadrado com v graus de liberdade.

Assim, a estatística 6.29 pode ser utilizada tanto para verificar a aderência das frequências observadas a um modelo, (teste Qui-Quadrado de aderência), como para verificar a independência entre duas variáveis.

6.2.6.1 Teste de aderência

É utilizado para verificar o ajustamento de um modelo de probabilidade aos dados observados, ou seja, verificar se as diferenças entre as frequências observadas e esperadas são estatisticamente significativas.

Neste caso o número de graus de liberdade (v) será:

$$v = k - 1 - m,$$

em que:

k é o número de classes, e

m o número de parâmetros estimados para se obter as freq. esperadas.

Exemplo1. Em seus experimentos com ervilhas, Mendel, ao cruzar plantas de sementes amarelas lisas com plantas de sementes verdes enrugadas, observou a seguinte descendência na geração F_2 : 315 plantas com sementes amarelas lisas, 108 com sementes amarelas enrugadas, 101 com sementes verdes lisas e 32 com sementes verdes enrugadas. De acordo com os postulados de Mendel a segregação esperada nesta geração deveria seguir a proporção de 9:3:3:1 Verificar se a teoria da segregação independente dos genes explica a segregação observada.

Solução:

Hipóteses a serem testadas:

$$\begin{cases} \text{proporção} = 9 : 3 : 3 : 1 \\ \text{proporção} \neq 9 : 3 : 3 : 1 \end{cases} \Rightarrow \begin{cases} \chi^2 = 0 \\ \chi^2 > 0 \end{cases}$$

Obter as frequência esperadas

lh Número de gols por partida marcados pelo Cruzeiro Esporte Clube durante o campeonato brasileiro de 2002.

Número de gols	0	1	2	3	4
Número de partidas	8	9	4	2	3

classes	F_o	F_e
Amarelas lisas	315	312,75
Amarelas enrugadas	108	104,25
Verdes lisas	101	104,25
Verdes enrugadas	32	34,75
Total	56	556

Obter a estatística χ_{calc}^2

$$\chi_{calc}^2 = \frac{(315 - 312,75)^2}{312,75} + \dots + \frac{(32 - 34,75)^2}{34,75} = 0,470$$

como nenhum parâmetro foi estimado, o número de graus de liberdade será: $v = 4 - 1 - 0 = 3$

Verifica-se na tabela de χ^2 que $\chi_{(0,01)}^2(3) = 11,345$

Como $\chi_{calc}^2 < \chi_{tab}^2$ o teste foi não significativo.

Exemplo2 A distribuição do número de gols/partida, realizadas pelo Cruzeiro, durante o Brasileirão 2001, foi:

Verificar se o número de gols por partida pode ser modelado segundo uma distribuição de Poisson

Solução:

Para obtenção das frequências esperadas pela distribuição torna-se necessário estimar o número médio de gols: $\bar{x} = 1,35$ gols por partida, em seguida obter a distribuição de probabilidade do número de gols/partidas:

Assim a frequência esperada pela distribuição de Poisson será dada pelo produto da probabilidade do cruzeiro realizar um determinado número de gols em uma partida

lh Probailidades estimadas via modelo de Poisson do número de gols por partida marcados pelo Cruzeiro Esporte Clube durante o campeonato brasileiro de 2002.

número de gols	0	1	2	3	4
Prob.	0,26	0,35	0,23	0,10	0,04

lh Frequências esperadas do número de gols por partida marcados pelo Cruzeiro Esporte Clube durante o campeonato brasileiro de 2002, estimadas pelo modelo Poisson.

Número de gols	0	1	2	3	4
Número de partidas (Fo)	8	9	4	2	3
Fe	6,76	9,10	6,24	2,86	1,04

pelo número de partidas realizadas:

O valor da estatística χ^2 será:

$$\chi_{calc}^2 = \frac{(8 - 6,76)^2}{6,76} + \frac{(9 - 9,10)^2}{9,10} + \dots + \frac{(3 - 1,04)^2}{1,04} = 4,98$$

Comparando esse valor com o de $\chi_{Tabela}^2 =$ com 3 graus de liberdade (5-1-1) tem-se que o pvalor=0.1732, portanto não rejeita-se H_0 .

6.2.6.2 Teste de independencia

O teste χ^2 de independência é aplicado a tabelas de contingência, as quais são construídas no intuito de estudar a relação entre duas variáveis categóricas. Considere-se como exemplo a tabela 6.3 na qual estão apresentados os número de alunos matriculados nos colégios A e B, em relação à sua classe econômica (alta, média ou baixa).

A estatística utilizada para o teste é

$$\chi_{calc}^2 = \sum_{j=1}^k \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}} \quad (6.30)$$

em que:

Fo_{ij} é a frequência observada na casela ij ;

Tabela 6.2: Número de alunos matriculados em dois colégios em relação à classe social dos mesmos

colégio	Classe social			Total
	Alta	Média	Baixa	
A	20	40	40	100
B	50	40	30	120
Total	70	80	70	220

Tabela 6.3: Número de alunos matriculados em dois colégios em relação à classe social dos mesmos

colégio	Classe social			Total
	Alta	Média	Baixa	
A	20(31,82)	40(36,36)	40(31,82)	100
B	50(31,18)	40(43,64)	30(38,18)	120
Total	70	80	70	220

() Frequência esperada

Fe_{ij} é a frequência esperada na casela ij , a qual é dada por:

$$Fe_{ij} = \frac{(Total\ da\ linha\ i)(total\ da\ coluna\ j)}{totalgeral}$$

Sob H_0 , a estatística 6.30 tem distribuição de χ^2 com

$$v = (h - 1)(k - 1) - p$$

graus de liberdade, sendo p o número de parâmetros estimados.

No exemplo tem-se:

substituindo esses resultados em 6.30

$$\chi_{calc}^2 = \frac{(20 - 31,82)^2}{31,82} + \frac{(40 - 36,36)^2}{336,36} + \dots + \frac{(30 - 30,18)^2}{30,18} = 20,27$$

Verifica-se na tabela de χ^2 que o valor de $\chi_{0,05}^2(2) = 5,99$. Como o valor de χ_{calc}^2 é maior que o de χ_{tab}^2 , este se encontra na região de rejeição de H_0 portanto, rejeita-se a hipótese de independência entre os colégios e a classe social dos alunos. Ou seja pode-se afirmar, ao nível de 0,05 que a classe social e o colégio no qual os alunos estudam não são independentes.

Capítulo 7

Regressão e Correlação linear

Estimação dos parâmetros do modelo de regressão pelo método dos mínimos quadrados:

Seja o modelo:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (7.1)$$

em que:

y_i é o valor observado da variável resposta (dependente);

β_0 é o intercepto do modelo;

β_1 é coeficiente angular;

x_i é o valor da variável preditora e

e_i é o erro aleatório associado a observação y_i .

Ajustar um modelo de regressão, via método de mínimos quadrados, implica procurar os valores ($\hat{\beta}_i$) tais que os valores estimados (preditos) de y_i , $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ sejam os mais próximos possíveis dos valores observados. Isto é os erros sejam mínimos

Partindo-se do modelo 7.1 tem-se que o erro cometido ao se estimar uma observação é

$$e_i = y_i - \beta_0 - \beta_1 x_i.$$

Definindo a função

$$S(\beta_0, \beta_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (7.2)$$

Os estimadores de mínimos quadrados de β_0 e β_1 , $\hat{\beta}_0$ e $\hat{\beta}_1$ são aqueles que minimizam a função 7.2. Assim, estes estimadores são obtidos solucionando-se o sistema:

$$\begin{cases} \frac{\partial S}{\partial \beta_0} = 0 \\ \frac{\partial S}{\partial \beta_1} = 0 \end{cases}$$

$$\begin{cases} 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0 \\ 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i = 0 & (a) \\ \sum_{i=1}^n y_i x_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 = 0 & (b) \end{cases}$$

de (a) tem-se:

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \quad (7.3)$$

de (b) tem-se:

$$\begin{aligned}
\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \\
\left(\frac{\sum_{i=1}^n y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} \right) \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \\
\frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i^2}{n} + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \\
\frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} + \hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n x_i^2}{n} \right) &= \sum_{i=1}^n x_i y_i \\
\hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n x_i^2}{n} \right) &= \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n x_i^2}{n}} \quad (7.4) \\
\hat{\beta}_1 &= \frac{SPXY}{SQDX}
\end{aligned}$$

Uma medida da qualidade do ajuste, do modelo obtido, aos dados é dada pelo coeficiente de determinação (R^2),

$$r^2 = \frac{SPXY^2}{SQDY} \quad (7.5)$$

Exemplo: Os dados a seguir referem-se ao número de CDs vendidos por uma determinada gravadora, em milhares de unidades, em 10 semanas consecutivas após o lançamento do mesmo. Ajustar um modelo de regressão linear simples que descreva a quantidade de CDs vendidos em função do tempo de lançamento.

Semanas	1	2	3	4	5	6	7	8	9	10
CDs (<i>Milunid</i>)	5,0	6,7	6,0	8,7	6,2	8,6	11,0	11,9	10,6	10,8

Tem-se que: $n = 10$ $\sum_{i=1}^{10} X_i = 55$ $\sum_{i=1}^{10} X_i^2 = 385$

$\sum_{i=1}^{10} Y_i = 85,5$ $\sum_{i=1}^{10} X_i Y_i = 529,4$

Substituindo esses valores em 7.4 tem-se:

$$\hat{\beta}_1 = \frac{529,4 - \frac{(55)(85,5)}{10}}{385 - \frac{55^2}{10}} = 0,72$$

e em 7.3:

$$\hat{\beta}_0 = 8,55 - (0,72)(5,5) = 4,59$$

Portanto a equação de regressão que descreve o número de CDs vendidos em função do número de semanas após o lançamento é:

$$y = 4,59 + 0,72x$$

Cujo coeficiente de determinação é:

$$r^2 = \frac{59,15^2}{82,5} = 0,77$$