

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ESTATÍSTICA - CURSO 1

Dra. Corina da Costa Freitas
MSc. Camilo Daleles Rennó
MSc. Manoel Araújo Sousa Júnior

Material de referência para o curso 1 de estatística.

INPE
São José dos Campos
Março de 2003

SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS	
LISTA DE TABELAS	
CAPÍTULO 1 – INTRODUÇÃO	11
1.1 – O que é Estatística?	11
1.2 – Conceitos de População e Amostra	12
1.3 – Tipos de Variáveis	12
1.4 – Distribuições de Freqüência	14
1.4.1 – Freqüências e freqüências relativas	14
1.5 – Distribuição de freqüências e sua representação gráfica para variáveis quantitativas discretas	16
1.6 – Distribuição de freqüências e sua representação gráfica para variáveis quantitativas contínuas	18
1.6.1 – Curvas de Freqüência	20
1.7 – Medidas de Tendência Central	20
1.7.1 – Média	21
1.7.2 – Mediana	21
1.7.3 – Moda	22
1.8 – Medidas de Dispersão	22
1.8.1 – Amplitude Total	22
1.8.2 – Desvio Médio	22
1.8.3 – Coeficiente de Variação	23
1.8.4 – Variância (σ^2)	23
1.8.5 – Desvio Padrão (σ)	24
1.9 – Momentos, Assimetria e Curtose	24
1.9.1 – Momentos	24
1.9.2 – Assimetria	25
1.9.3 – Curtose	25
CAPÍTULO 2 – PROBABILIDADE	27
2.1 – Definição de Probabilidade utilizando Freqüência Relativa	27

2.2 – Experimentos aleatórios	27
2.3 – Espaço Amostral	28
2.4 – Eventos	28
2.5 – Classe dos Eventos Aleatórios	30
2.6 – Definição Clássica de Probabilidade	30
2.7 – Operações com Eventos Aleatórios - Teoria dos Conjuntos	32
2.7.1 – União de conjuntos	32
2.7.2 – Intersecção de conjuntos	33
2.7.3 – Complemento	34
2.7.4 – Eventos Mutuamente Exclusivos ou Disjuntos	34
2.8 – Definição Axiomática de Probabilidade	35
2.9 – Eventos Independentes	37
2.10 –Análise Combinatoria	38
2.10.1 –Permutação	38
2.10.2 –Combinações	39
2.11 –Probabilidade Condicional	40
2.12 –Probabilidade Total	41
2.13 –Teorema de Bayes	43
2.13.1 –Teoria de Bayes	44
CAPÍTULO 3 – VARIÁVEIS ALEATÓRIAS E DISTRIBUIÇÕES	47
3.1 – Variável Aleatória	47
3.2 – Tipos de Variáveis Aleatórias, Função de Probabilidade e Função Densidade de Probabilidade	47
3.3 – Função de Distribuição Acumulada	50
3.4 – Distribuições Bivariadas	52
3.4.1 – Distribuições Conjuntas Discretas	52
3.4.2 – Distribuições Conjuntas Contínuas	54
3.5 – Distribuições Marginais	55
3.6 – Esperança de Uma Variável Aleatória	58
3.6.1 – Distribuições Discretas	58
3.6.2 – Distribuições Contínuas	59
3.6.3 – Propriedades da Esperança	59
3.7 – Variância de uma Variável Aleatória	61
3.7.1 – Variáveis Aleatórias Discretas	61

3.7.2 – Variáveis Aleatórias Contínuas	62
3.7.3 – Propriedades da Variância	62
3.8 – Momentos	63
3.8.1 – Momentos Centrais	64
3.8.2 – Função Geratriz de Momentos	64
3.8.3 – Propriedades das Funções Geradoras de Momentos	65
3.9 – Funções de uma Variável Aleatória	66
3.9.1 – Variável com Distribuição Discreta	66
3.9.2 – Variável com Distribuição Contínua	67
3.10 – Algumas Distribuições Discretas	69
3.10.1 – Distribuição Binomial	69
3.10.2 – Distribuição Hipergeométrica	74
3.10.3 – Distribuição Binomial Negativa	75
3.10.4 – Distribuição Geométrica	76
3.10.5 – Distribuição de Poisson	77
3.11 – Algumas Distribuições Contínuas Importantes	81
3.11.1 – Distribuição Normal	81
3.11.2 – Distribuição Normal Padrão	82
3.11.3 – Teorema do Limite Central	87
3.11.4 – Distribuição Uniforme	89
3.11.5 – Distribuição Exponencial	92
3.11.6 – Distribuição Gama	94
3.11.7 – Distribuição Normal Bivariada	94
3.11.8 – Distribuição Normal Multivariada	95
CAPÍTULO 4 – INFERÊNCIA ESTATÍSTICA	97
4.1 – Introdução	97
4.1.1 – Parâmetros de uma distribuição	97
4.1.2 – Estatística	97
4.1.3 – Estimação Pontual e por Intervalo	97
4.2 – Estimação Pontual	97
4.2.1 – Método dos Momentos	98
4.2.2 – Método da Máxima Verossimilhança	99
4.3 – Estimadores Não Tendenciosos	101
4.4 – A Distribuição χ^2	102

4.5 – A Distribuição <i>t-student</i>	104
4.5.1 – Distribuição da Média Amostral	105
4.5.2 – Distribuição da diferença de médias amostrais	106
4.6 – Distribuição <i>F</i>	107
4.6.1 – Distribuição da Razão entre duas Variâncias Amostrais	107
4.7 – Estimação por Intervalos - Intervalos de Confiança	108
4.7.1 – Intervalo de Confiança para a Média Populacional μ	108
4.7.2 – Intervalo de Confiança para a Variância Populacional σ^2	111
4.7.3 – Intervalo de Confiança para a diferença de médias de duas Populações	112
4.7.4 – Intervalo de Confiança para Razão das Variâncias σ_1^2/σ_2^2	114
4.7.5 – Intervalo de Confiança para uma Proporção	114
4.7.6 – Intervalo de Confiança para Diferença de Proporções	115
CAPÍTULO 5 – TESTES DE HIPÓTESES	117
5.1 – Hipótese Nula e Hipótese Alternativa	117
5.2 – Região Crítica do Teste	117
5.3 – Erros do Tipo I e Erros do tipo II	118
5.4 – Teste da hipótese de que a média populacional tem um valor específico	118
5.4.1 – σ conhecido	118
5.4.2 – σ desconhecido	119
5.5 – Controlando o erro tipo II (β)	122
5.6 – Teste da hipótese de que a variância populacional tem um valor específico	123
5.7 – Teste da razão de variâncias	125
5.8 – Teste da hipótese da igualdade de duas médias	127
5.8.1 – σ_1^2 e σ_2^2 conhecidas	127
5.8.2 – σ_1^2 e σ_2^2 desconhecidas, mas $\sigma_1^2 = \sigma_2^2$	127
5.8.3 – σ_1^2 e σ_2^2 desconhecidas, mas $\sigma_1^2 \neq \sigma_2^2$	128
5.9 – Teste para proporção	132
5.9.1 – Diferença entre proporções	133
5.10 – Teste χ^2 da independência	134
CAPÍTULO 6 – ANÁLISE DE VARIÂNCIA	139
6.1 – Introdução	139
6.2 – Análise de Variância de Um Fator	140
6.3 – Teste para Igualdade de Várias Variâncias	150

6.4 – Análise de Variância de Dois Fatores	152
6.5 – Análise de Variância de Dois Fatores - Várias observações por cela . . .	157
6.5.1 – Identidade da Soma de Quadrados	161

LISTA DE FIGURAS

	<u>Pág.</u>
1.1 Figura tirada de : Costa Neto, P. L. de O. - Estatística, Ed. Edgard Blücher Ltda - 1977	12
1.2 Histograma de frequência	16
1.3 Histograma de frequências relativas	17
1.4 Histograma de frequências relativas acumuladas	18
1.5 Histograma e Polígono de frequência	19
1.6 Histograma frequência acumulada	20
1.7 Histograma frequência acumulada	21
1.8 Exemplos de curtose	25
2.1 União dos conjuntos A e B	32
2.2 Intersecção dos conjuntos A e B	33
2.3 Complemento do conjunto A	34
2.4 A probabilidade a <i>priori</i> nas f.d.p. das classes	45
3.1 Gráfico de $P(x)$ para dois dados	48
3.2 Gráfico de $f dp$ de f	50
3.3 Gráfico F	52
3.4 Meyer, página 75.	53
3.5 Degroot, página 93.	54
3.6 Degroot, página 95.	55

3.7	Gráfico de $f(x)$ de uma função uniforme	90
3.8	Probabilidade de uma variável aleatória uniforme	90
6.1	Distribuições normais com mesma variância (σ^2) para todas as populações	142
6.2	Distribuições normais com mesma média (μ) para todas as populações	142
6.3	Efeitos de fertilizantes e variedades de trigo, sem interação	157
6.4	Efeitos de fertilizantes e variedades de trigo, sem interação	158

LISTA DE TABELAS

	<u>Pág.</u>
1.1 Número de filhos dos 50 funcionários da empresa Fictícia S.A.	15
1.2 Freqüências, freqüências relativas e freqüências relativas acumuladas . .	17
1.3 Altura (cm) dos 50 funcionários da empresa Fictícia S.A.	18
1.4 Altura (cm) dos funcionários da Fictícia S.A.	20
5.1 Representação do erros do tipo I e II	118
5.2 Candidatos selecionados	134
5.3 Valores da \hat{E}_{ij}	137

CAPÍTULO 1

INTRODUÇÃO

1.1 O que é Estatística?

Estatística é a ciência que investiga os processos de obtenção, organização e análise de dados sobre uma população, e os métodos de tirar conclusões ou fazer previsões com base nesses dados.

Este conceito tem um significado mais amplo do que aquele que usualmente se dá à palavra "estatística", isto é, o resultado de contagens sobre a ocorrência de determinados eventos e a sua representação através de gráficos e tabelas, como, por exemplo, as estatísticas de ocorrência de chuvas numa certa época do ano; as estatísticas sobre os ganhadores de prêmios de loteria; as estatísticas de renda média por região etc.

Em geral, este conceito mais popular de estatística corresponde somente à organização e descrição dos dados relativos a um determinado experimento ou situação e não trata da análise e interpretação desses dados. Ele está associado à parte da Estatística que denominamos de Estatística Descritiva. A *Estatística Descritiva*, portanto, é a parte da Estatística que se preocupa com a organização e descrição de dados experimentais.

Além da Estatística Descritiva há a *Estatística Indutiva* ou *Estatística Inferencial* que consiste, fundamentalmente, das técnicas de análise e interpretação dos dados. A partir de um conjunto restrito de dados, chamado de *amostra*, organizado e descrito pela Estatística Descritiva, a Estatística Indutiva procura fazer inferências ou, em outras palavras, tirar conclusões sobre a natureza desses dados e estender essas conclusões a conjuntos maiores de dados, chamados de *populações*.

É evidente que, para que a Estatística Indutiva possa deduzir conclusões válidas, é necessário que se tomem alguns cuidados para a escolha da amostra a ser utilizada. Esses cuidados, mais propriamente chamados de critérios, são estabelecidos por uma técnica chamada de *amostragem*.

Contudo, para permitir que a Estatística Indutiva proporcione conclusões válidas não basta utilizar as técnicas de organização e descrição dos dados da Estatística

Descritiva e as técnicas corretas de amostragem. Fica ainda faltando uma última ferramenta que é o *cálculo de probabilidades*. O cálculo de probabilidades é um conjunto de técnicas matemáticas que visa determinar as chances de ocorrência de eventos regidos pelas leis do acaso.

A Figura 1.1 abaixo interrelaciona os conceitos citados:

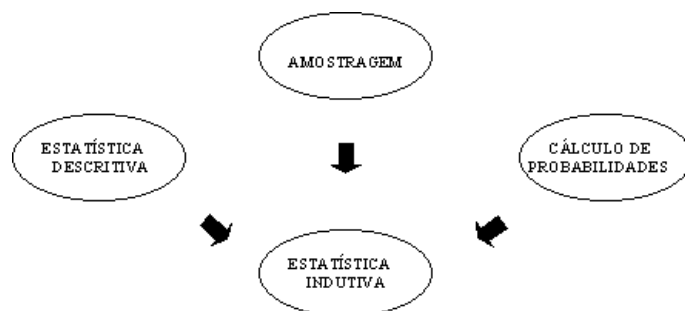


Fig. 1.1 – Figura tirada de : Costa Neto, P. L. de O. - Estatística, Ed. Edgard Blücher Ltda - 1977

1.2 Conceitos de População e Amostra

População ou **Universo** é a totalidade dos objetos concebíveis de uma certa classe em consideração.

Amostra são os objetos selecionados da população. Se esses objetos são selecionados de tal maneira que cada objeto tem a mesma chance de ser selecionado do que o outro, temos uma **amostra aleatória**

1.3 Tipos de Variáveis

É necessário, inicialmente, que se defina qual(is) a(s) características dos elementos que deverá(ão) ser verificada(s). Ou seja, não se trabalha estatisticamente com os elementos existentes, mas com alguma(s) característica(s) desses elementos. Por exemplo, os elementos a serem estudados podem ser a população de uma cidade, mas estaremos interessados em alguma característica como renda, idade, sexo, tipo de moradia, etc. Trabalha-se portanto com os valores de uma *variável* (que é a característica de interesse), e não com os elementos originalmente considerados. A escolha da variável (ou variáveis) de interesse dependerá dos objetivos do estudo

estatístico em questão. Esta característica (variável) poderá ser *qualitativa* ou *quantitativa*.

A variável será *qualitativa* quando resultar de uma classificação por tipos ou atributos, como nos seguintes exemplos:

- a) População: alunos de uma universidade
Variável: sexo (masculino ou feminino).
- b) População: moradores de uma cidade
Variável: tipo de habitação (casa, apartamento, barraco, etc.).
- c) População: peças produzidas por uma máquina
Variável: qualidade (perfeita ou defeituosa).
- d) Óbitos em um hospital, nos últimos cinco anos
Variável: causa mortis (moléstia cardiovasculares, cânceres, etc)
- e) População Brasileira
Variável: cor da pele (branca, preta, amarela, vermelha, parda).

A variável será *quantitativa* quando seus valores forem expressos em números. Pode ser subdivida em:

- 1 - **quantitativa discreta**: pode assumir apenas valores pertencentes a um conjunto enumerável;
- 2 - **quantitativa contínua**: pode assumir qualquer valor em um certo intervalo de variação.

Alguns exemplos de variáveis quantitativas discretas são:

- a) População: habitações de uma cidade.
Variável: número de banheiros.
- b) População: casais residentes em uma cidade.
Variável: número de filhos.

- c) População: aparelhos produzidos em uma linha de montagem.
Variável: número de defeitos por unidade.
- d) População: Bolsa de valores de São Paulo.
Variável: número de ações negociadas.

Alguns exemplos de variáveis quantitativas contínuas são:

- a) População: estação meteorológica de uma cidade.
Variável: precipitação pluviométrica durante um mês.
- b) População: pregos produzidos por uma máquina.
Variável: comprimento.
- c) População: propriedades agrícolas do Brasil.
Variável: produção de algodão.
- d) População: indústrias de uma cidade.
Variável: índice de liquidez.
- e) População: pessoas residentes em uma cidade.
Variável: idade.

Para atingir os objetivos da Estatística descritiva, os dados observados são muitas vezes sintetizados e apresentados em formas de tabelas ou gráficos, os quais irão fornecer rápidas e seguras informações a respeito das variáveis em estudo. Uma das tabelas mais utilizadas na estatística é a distribuição de freqüências. Os gráficos associados à ela são o gráfico de freqüências (denominado histograma, para o caso de variáveis quantitativas contínuas), o polígono de freqüências, o gráfico de freqüência acumulada e o polígono de freqüência acumulada.

1.4 Distribuições de Freqüência

Muitas vezes os gráficos são elaborados utilizando-se as freqüências dos valores da variável. Para tal, necessitamos definir alguns conceitos importantes.

1.4.1 Freqüências e freqüências relativas

Definimos freqüência de um valor de uma variável (qualitativa ou quantitativa) como sendo o número de vezes que aquele valor se repete no conjunto de dados

experimentais. Usaremos a notação f_i para representar a frequência do i -ésimo valor observado.

Seja n o número total de valores observados e k o número de diferentes valores obtidos, tem-se:

$$\sum_{i=0}^k f_i = n \quad (1.1)$$

Exemplo

Seja o conjunto de dados abaixo (Tabela 1.1), que representa o número filhos de funcionários da empresa Fictícia S.A..

TABELA 1.1 – Número de filhos dos 50 funcionários da empresa Fictícia S.A.

Num.de Filhos	Frequência	Freq. Relativa
0	15	0,30
1	10	0,20
2	13	0,26
3	6	0,12
4	3	0,06
5	3	0,06
Total	50	1,00

As frequências são:

- $f_0 = 15$ (corresponde ao valor 0)
- $f_1 = 10$ (corresponde ao valor 1)
- $f_2 = 13$ (corresponde ao valor 2)
- $f_3 = 6$ (corresponde ao valor 3)
- $f_4 = 3$ (corresponde ao valor 4)
- $f_5 = 3$ (corresponde ao valor 5)

Chamamos de *distribuição de freqüências* à associação das freqüências aos respectivos valores observados. Portanto, a representação acima caracteriza uma distribuição de freqüências. Do mesmo modo, podemos definir *freqüência relativa* de um valor observado como sendo a relação:

$$p'_i = \frac{f_i}{n} \quad (1.2)$$

Verifica-se facilmente que:

$$\sum_{i=0}^k p'_i = 1 \quad (1.3)$$

1.5 Distribuição de freqüências e sua representação gráfica para variáveis quantitativas discretas

A Tabela 1.1 representa a distribuição de frequências para a variável discreta "número de filhos". A representação gráfica de uma distribuição de freqüência de uma variável quantitativa discreta é denominada gráfico de freqüências (Figura 1.2). Utilizando o exemplo, temos o seguinte gráfico de frequências

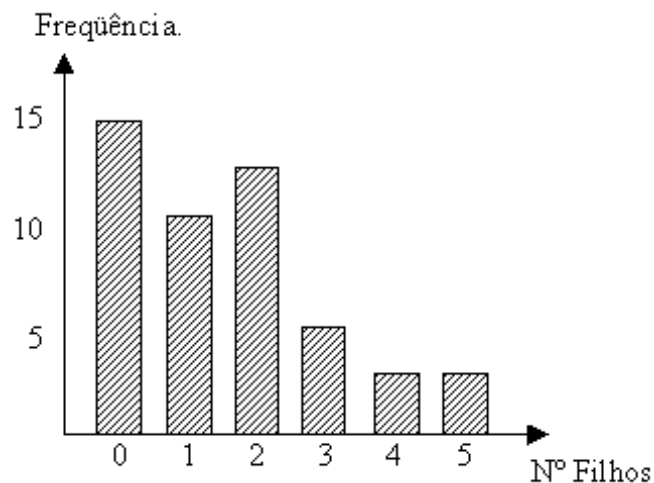


Fig. 1.2 – Histograma de freqüência

Uma outra representação utilizada é a do gráfico das freqüências acumuladas

e freqüências relativas acumuladas. Tomando-se os dados do exemplo anterior podemos calcular as freqüências, freqüências acumuladas e freqüências relativas acumuladas dos diversos valores. Esse cálculo está ilustrado na Tabela 1.2.

TABELA 1.2 – Freqüências, freqüências relativas e freqüências relativas acumuladas

Num.de Filhos	Freqüência	Freq. Relativa	Freq. Relativa Acumulada
0	15	0,30	0,30
1	10	0,20	0,50
2	13	0,26	0,76
3	6	0,12	0,88
4	3	0,06	0,94
5	3	0,06	1,00
Total	50	1,00	-

Com os dados acima podemos construir o gráfico de freqüências relativas (Figura 1.3) e freqüências relativas acumuladas (Figura 1.4).

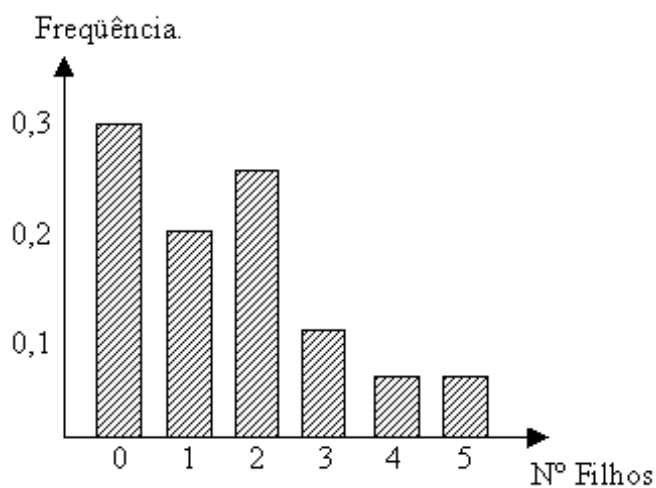


Fig. 1.3 – Histograma de freqüências relativas

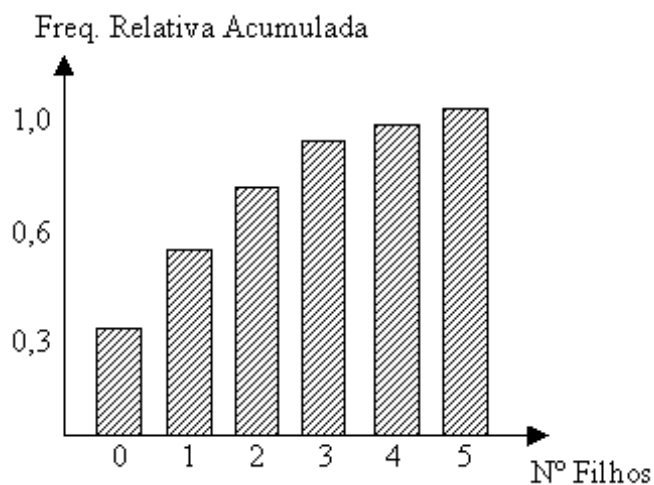


Fig. 1.4 – Histograma de frequências relativas acumuladas

1.6 Distribuição de frequências e sua representação gráfica para variáveis quantitativas contínuas

As variáveis quantitativas contínuas diferem um pouco das discretas na sua forma de representação gráfica. Para entender essa diferença temos que nos lembrar que as variáveis contínuas, por definição, têm os seus valores definidos num intervalo contínuo dos números reais. Portanto, não tem sentido falar em frequência de repetição de um determinado valor, pois os valores raramente se repetem. A Tabela 1.3 representa uma distribuição de frequência para a variável contínua "altura de funcionários".

TABELA 1.3 – Altura (cm) dos 50 funcionários da empresa Fictícia S.A.

Altura	Frequência	Freq. Relativa
151 - 159	2	0,04
159 - 167	11	0,22
167 - 175	18	0,36
175 - 183	10	0,20
183 - 191	8	0,16
191 - 199	1	0,02
Total	50	1,00

Intervalos de classes: O símbolo que define uma classe

Exemplo: 151 – 158

Limites da classe: Os números extremos de cada intervalo

Exemplo: o limite inferior da 1ª classe é 151 e o limite superior da 1ª classe é 158

Ponto médio de classe: ponto intermediário do intervalo de classe

Exemplo: Ponto médio da 1ª classe é 154,5

Amplitude do intervalo de classe: é a diferença entre os limites reais superior e inferior

Exemplo: amplitude da 1ª classe é 8

Com os dados da Tabela 1.1 podemos construir o gráfico de freqüências do mesmo modo que fizemos para as variáveis discretas. A diferença mais importante é que, agora, as freqüências são associadas a intervalos de valores (classes de freqüências) e não mais a valores individuais da variável em estudo.

Além disto, o gráfico, neste caso, é chamado de *histograma* (Figura 1.5) que consiste em um conjunto de retângulos, com centro no ponto médio e larguras iguais aos intervalos das classes, e áreas proporcionais às freqüências das classes.

A seguir está mostrado o histograma correspondente aos dados do exemplo acima, e o polígono de freqüências, que é o gráfico obtido unindo-se os pontos médios dos patamares do histograma.

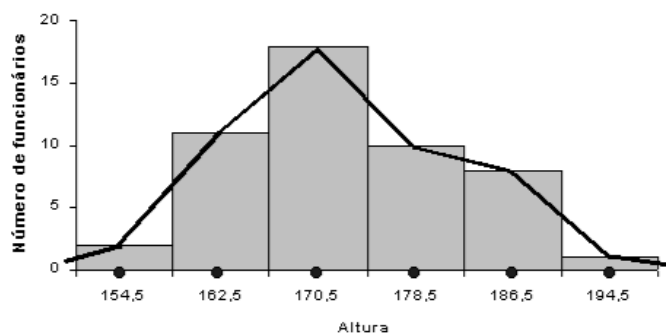


Fig. 1.5 – Histograma e Polígono de freqüência

O próximo gráfico é o polígono de *freqüências acumuladas*. Ele é construído unindo-se as freqüências acumuladas ao final de cada classe de freqüências (Tabela 1.4). Pode ser construído também com as freqüências relativas acumuladas e, neste caso, ele se chama polígono de freqüências relativas acumuladas. O primeiro está mostrado na Figura 1.6.

TABELA 1.4 – Altura (cm) dos funcionários da Fictícia S.A.

Altura	Num. de Funcionários
< 151	0
< 159	2
< 167	13
< 175	31
< 183	41
< 191	50

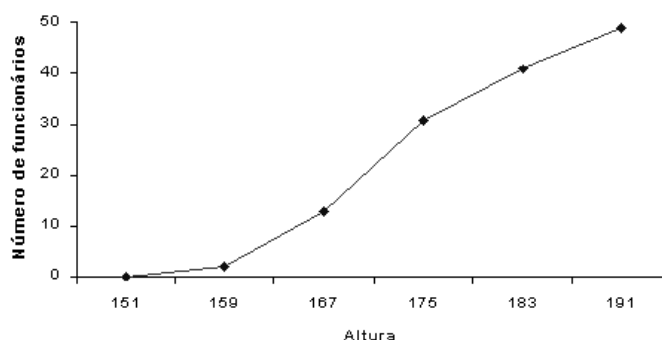


Fig. 1.6 – Histograma freqüência acumulada

1.6.1 Curvas de Freqüência

Tipos de curvas de freqüência: As curvas de freqüência aparecem, na prática, sob diversas formas características, como as indicadas na Figura 1.7

1.7 Medidas de Tendência Central

As medidas mais comuns são: média, mediana, moda

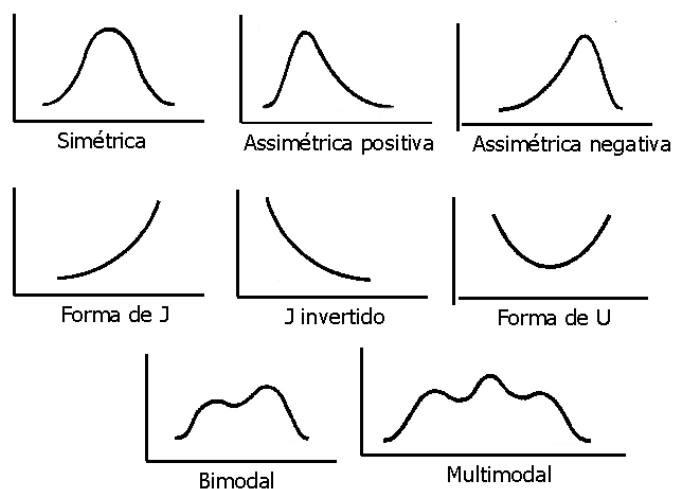


Fig. 1.7 – Histograma frequência

1.7.1 Média

A média de um conjunto de N números X_1, X_2, \dots, X_N é representada por \bar{X} e definida por:

$$\begin{aligned} \bar{X} &= \frac{X_1 + X_2 + \dots + X_N}{N} \\ &= \frac{\sum_{i=1}^N X_i}{N} \end{aligned} \quad (1.4)$$

De modo geral, para dados agrupados, temos:

$$\bar{X} \cong \sum_{i=1}^m x_i \left(\frac{f_i}{N} \right) \quad (1.5)$$

onde: x_i é o ponto médio da classe i ; $\frac{f_i}{N}$ é a frequência relativa da classe i ; e m é número de classes.

1.7.2 Mediana

A mediana de um conjunto de números, ordenados em ordem de grandeza, é o valor médio (N ímpar) ou a média aritmética dos dois valores centrais (N par).

Exemplos:

$\{3, 4, 4, 5, 6, 8, 8, 8, 10\}$ tem mediana 6

$\{5, 6, 7, 9, 11, 12, 13, 17\}$ tem mediana 10

1.7.3 Moda

A moda é o valor que ocorre com mais frequência. A moda pode não existir e, mesmo que exista, pode não ser única.

Exemplos:

$\{1, 1, 3, 3, 5, 7, 7, 7, 11, 13\}$ tem moda 7

$\{3, 5, 8, 11, 13, 18\}$ não tem moda

$\{3, 5, 5, 5, 6, 6, 7, 7, 7, 11, 12\}$ tem duas modas 5 e 7 (bimodal)

1.8 Medidas de Dispersão

O grau ao qual os dados numéricos tendem a dispersar-se em torno de um valor médio chama-se **variação** ou **dispersão** dos dados.

As medidas mais comuns são: amplitude total, desvio médio, desvio padrão e variância.

1.8.1 Amplitude Total

É a diferença entre o maior e o menor valor

Exemplos:

A amplitude total de $\{4, 7, 9, 11, 11, 15, 20\}$ é 16

1.8.2 Desvio Médio

O desvio médio de X_1, X_2, \dots, X_N pode ser obtido pela seguinte fórmula

$$DM = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N} \quad (1.6)$$

Exemplo:

O desvio médio de $\{2, 3, 6, 8, 11\}$ é 2.8

Para dados agrupados (em tabelas de frequência) o desvio padrão é computado por:

$$DM \cong \frac{1}{N} \sum_{i=1}^m |X_i - \bar{X}| f_i \quad (1.7)$$

Onde: x_i é o ponto médio da classe i , f_i é a frequência da classe i , e m é o número de classes.

1.8.3 Coeficiente de Variação

O efeito da variação ou dispersão em relação à média é medido pela dispersão relativa, que é definida por:

$$DR = \frac{\text{Disp. absoluta}}{\text{Media}}$$

Se a dispersão absoluta for o desvio padrão, a dispersão relativa é denominada coeficiente de variação (CV), que é pode ser representado por:

$$CV = \frac{s}{\bar{X}} \quad (1.8)$$

Obs: O Coeficiente de variação deixa de ser útil quando o \bar{X} está próximo de zero.

1.8.4 Variância (σ^2)

É o quadrado do desvio padrão

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N} \quad (1.9)$$

Observação: O desvio padrão corresponde aos dados de uma amostra é em geral calculado com o divisor $(N - 1)$ ao invés de N , para que se tornem estimadores *não tendenciosos*. Neste caso geralmente utiliza-se as letras s e s^2 para representar o desvio padrão e a variância, respectivamente.

1.8.5 Desvio Padrão (σ)

O desvio padrão de X_1, X_2, \dots, X_N é dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}} \quad (1.10)$$

Para dados agrupados o desvio padrão é computado por:

$$\sigma \cong \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{X})^2 f_i}{N}} \quad (1.11)$$

1.9 Momentos, Assimetria e Curtose

1.9.1 Momentos

Se X_1, X_2, \dots, X_N são os N valores assumidos pela variável X , define-se **momento de ordem r** por:

$$\begin{aligned} m'_r &= \frac{X_1^r + X_2^r + \dots + X_N^r}{N} \\ &= \frac{\sum_{i=1}^N X_i^r}{N} \end{aligned} \quad (1.12)$$

e o **momento de ordem r** centrado na média pela equação

$$m_r = \frac{\sum_{i=1}^N (X_i - \bar{X})^r}{N} \quad (1.13)$$

Observação:

$$\bar{X} = m'_1$$

$$\sigma^2 = m'_2 - (m'_1)^2$$

$$m_1 = 0$$

$$\sigma^2 = m_2$$

Para dados agrupados os momentos podem ser calculados por:

$$m_r = \frac{\sum_{i=1}^m (x_i - \bar{X})^r f_i}{N} \quad (1.14)$$

onde: x_i é o ponto médio da classe i , f_i é a frequência e m o número de classes.

1.9.2 Assimetria

É o grau de desvio, ou afastamento da simetria, de uma distribuição.

Para distribuições assimétricas, a média tende a situar-se do mesmo lado da moda que a cauda mais longa. Por isso, uma medida de assimetria é proporcionada pela diferença entre a média e a moda. Ela pode ser tomada sem dimensão através de uma divisão por uma medida de dispersão, como o desvio padrão, resultando em:

$$Assimetria = \frac{\bar{X} - moda}{s} \quad (1.15)$$

Uma outra medida importante é o **coeficiente do momento de assimetria**, que utiliza o 3º momento centrado, expresso sob a forma não dimensional, definida por

$$a_3 = \frac{m_3}{s^3} = \frac{m_3}{\sqrt{m_2^3}} \quad (1.16)$$

1.9.3 Curtose

É o grau de achatamento de uma distribuição, e é muitas vezes considerado em relação a uma distribuição normal como as indicadas na Figura 1.8



Fig. 1.8 – Exemplos de curtose

Uma medida de curtose é o **coeficiente do momento de curtose** definido por:

$$a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} \quad (1.17)$$

Para a distribuição normal, $a_4 = 3$. Por essa razão, a curtose é freqüentemente definida por $(a_4 - 3)$, que é positivo para uma distribuição leptocúrtica, e negativo para uma platicúrtica e nulo para uma normal.

CAPÍTULO 2

PROBABILIDADE

2.1 Definição de Probabilidade utilizando Frequência Relativa

Uma das definições de probabilidade utiliza a frequência relativa, já que as frequências relativas são estimativas de probabilidades. Podemos então definir a probabilidade como a proporção (ou frequência relativa) em uma seqüência muito grande de experimentos.

$$P(e_1) = \lim_{n \rightarrow \infty} \frac{n_1}{n} \quad (2.1)$$

Onde:

e_1 é o resultado;

n é o número total de vezes que se repete o experimento;

n_1 é o número de vezes que o resultado e_1 ocorre;

$\frac{n_1}{n}$ é portanto a frequência relativa de e_1 .

Observação: Se o experimento tiver N resultados possíveis e_1, e_2, \dots, e_N então:

$$\begin{aligned} 0 \leq P(e_i) \leq 1 & \quad i = 1, 2, \dots, N \\ P(e_1) + P(e_2) + \dots + P(e_N) &= 1 \end{aligned} \quad (2.2)$$

2.2 Experimentos aleatórios

Encontramos na natureza dois tipos de fenômenos: determinísticos e aleatórios.

Os fenômenos determinísticos são aqueles em que os resultados são sempre os mesmos, qualquer que seja o número de ocorrências verificadas.

Nos fenômenos aleatórios, os resultados não serão previsíveis, mesmo que haja um grande número de repetições do mesmo fenômeno. Por exemplo, se considerarmos um pomar com centenas de laranjeiras, as produções de cada planta serão diferentes e não previsíveis, mesmo que as condições de temperatura, pressão, umidade, solo, etc. sejam as mesmas para todas as árvores.

Experimentos ou fenômenos aleatórios são aqueles que, mesmo repetidos várias vezes sob condições semelhantes, apresentam resultados imprevisíveis.

Exemplos:

- a) Lançamento de uma moeda honesta;
- b) Lançamento de um dado;
- c) Lançamento de duas moedas;
- d) Retirada de uma carta de um baralho completo de 52 cartas;
- e) Determinação da vida útil de um componente eletrônico.

2.3 Espaço Amostral

Defini-se **espaço amostral** (S) ao conjunto de todos os resultados possíveis de um experimento.

Nos exemplos acima, os espaços amostrais são:

- a) $S = \{c, r\}$
- b) $S = \{1, 2, 3, 4, 5, 6\}$
- c) $S = \{(c, r), (c, c), (r, c), (r, r)\}$
- d) $S = \{A_o, \dots, K_o, A_p, \dots, K_p, A_c, \dots, K_c, A_e, \dots, K_e\}$
- e) $S = \{t \in R/t \geq 0\}$

Cada um dos elementos de S que corresponde a um resultado recebe o nome de **ponto amostral**.

2.4 Eventos

Chamamos de **evento** (E) a qualquer subconjunto do espaço amostral S de um experimento aleatório.

Assim, o evento aleatório pode ser um único ponto amostral ou uma reunião deles.

Qualquer que seja o evento E , se $E \subset S$, então E é um evento de S .

Se $E = S$, E é chamado evento certo.

Se $E \subset S$ e E é um conjunto unitário, E é chamado evento elementar.

Se $E = \emptyset$, E é chamado evento impossível.

Exemplos:

1. No lançamento de um dado, onde $S = \{1, 2, 3, 4, 5, 6\}$, temos:

- a) $A = \{2, 4, 6\} \subset S$; logo A é um evento de S;
- b) $B = \{1, 2, 3, 4, 5, 6\} \subset S$; logo B é um evento certo de S ($B = S$);
- c) $C = \{4\} \subset S$; logo C é um evento elementar de S;
- d) $D = \emptyset \subset S$; logo D é um evento impossível de S.

Um evento é sempre definido por uma sentença. Assim os eventos acima podem ser definidos pelas sentenças: a) "Obter um número par"; b) "Obter um número menor ou igual a seis"; c) "Obter um número maior que três e menor que cinco"; d) "Obter um número maior que seis".

2. Lançam-se dois dados. Enumerar os seguintes eventos:

- a) Saída de faces iguais;
- b) Saída de faces cuja soma seja igual a 10;
- c) Saída de faces cuja soma seja menor que 2;
- d) Saída de faces cuja soma seja menor que 15;
- e) Saída de faces onde uma face é o dobro da outra.

Neste caso, o espaço amostral pode ser representado por uma tabela de dupla entrada:

D_2	1	2	3	4	5	6
D_1						
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Os eventos pedidos são:

- a) $A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$;
- b) $B = \{(4, 6), (5, 5), (6, 4)\}$;
- c) $C = \emptyset$ (evento impossível);
- d) $D = S$ (evento certo);
- e) $E = \{(1, 2), (2, 1), (2, 4), (3, 6), (4, 2), (6, 3)\}$.

2.5 Classe dos Eventos Aleatórios

Definição: É o conjunto formado de todos os eventos (subconjuntos) do espaço amostral. Para efeito de exemplo, consideremos o espaço amostral finito: $S = \{e_1, e_2, e_3, e_4\}$. A classe dos eventos aleatórios é:

$$F(S) = \left\{ \begin{array}{l} \emptyset \\ \{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}, \\ \{e_1, e_2\}, \{e_1, e_3\}, \{e_1, e_4\}, \{e_2, e_3\}, \{e_2, e_4\}, \{e_3, e_4\}, \\ \{e_1, e_2, e_3\}, \{e_1, e_2, e_4\}, \{e_1, e_3, e_4\}, \{e_2, e_3, e_4\}, \\ \{e_1, e_2, e_3, e_4\} \end{array} \right\}$$

2.6 Definição Clássica de Probabilidade

Dado um experimento aleatório, sendo S o seu espaço amostral, vamos admitir que todos os elementos de S tenham a mesma chance de acontecer, ou seja, que S é um conjunto equiprovável.

Definimos probabilidade de um evento A ($A \subset S$) ao número real $P(A)$, tal que:

$$P(A) = \frac{n^\circ \text{ de resultados favoráveis a } A}{n^\circ \text{ de resultados possíveis}} = \frac{n(A)}{n(S)} \quad (2.3)$$

Exemplo:

Considerando o lançamento de um dado, pede-se:

- a) A probabilidade do evento A "obter um número par na face superior".
Temos:
 $S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$
 $A = \{2, 4, 6\} \Rightarrow n(A) = 3$.
Logo, $P(A) = \frac{3}{6} = \frac{1}{2}$.

- b) A probabilidade do evento B "obter um número menor ou igual a 6 na face superior".

Temos:

$$S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$$

$$B = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(B) = 6.$$

$$\text{Logo, } P(B) = \frac{6}{6} = 1.$$

- c) A probabilidade do evento C "obter um número 4 na face superior".

Temos:

$$S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$$

$$C = \{4\} \Rightarrow n(C) = 1.$$

$$\text{Logo, } P(C) = \frac{1}{6}.$$

- d) A probabilidade do evento D "obter um número maior que 6 na face superior".

Temos:

$$S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$$

$$D = \emptyset \Rightarrow n(D) = 0.$$

$$\text{Logo, } P(D) = \frac{0}{6} = 0.$$

Exercícios Complementares:

- 1 - No lançamento de dois dados, calcule a probabilidade de se obter soma igual a 5.
- 2 - Qual a probabilidade de sair uma figura quando retiramos uma carta de um baralho de 52 cartas?
- 3 - Retira-se uma carta de um baralho completo de 52 cartas.
 - a) Qual a probabilidade de sair uma carta de copas ou de ouros?
 - b) Qual a probabilidade de sair um rei ou uma carta de espadas?
- 4 - No lançamento de um dado, qual a probabilidade de se obter um número não inferior a 5?
- 5 - São dados dois baralhos de 52 cartas. Tiramos, ao mesmo tempo, uma carta do primeiro baralho e uma carta do segundo. Qual a probabilidade de tiramos uma dama e um rei, não necessariamente nessa ordem?

- 6 - Dois dados são lançados conjuntamente. Determine a probabilidade de a soma ser 10 ou maior que 10.

2.7 Operações com Eventos Aleatórios - Teoria dos Conjuntos

Consideremos um espaço amostral finito $S = \{e_1, e_2, e_3, \dots, e_n\}$. Sejam A e B dois eventos de $F(S)$. As seguintes operações são definidas:

2.7.1 União de conjuntos

Definição: $A \cup B = \{e_i \in S / e_i \in A \text{ ou } e_i \in B\}, i = 1, \dots, n$. Portanto, o evento união é formado pelos pontos amostrais que pertencem a pelo menos um dos conjuntos. A união pode ser vista na Figura 2.1.

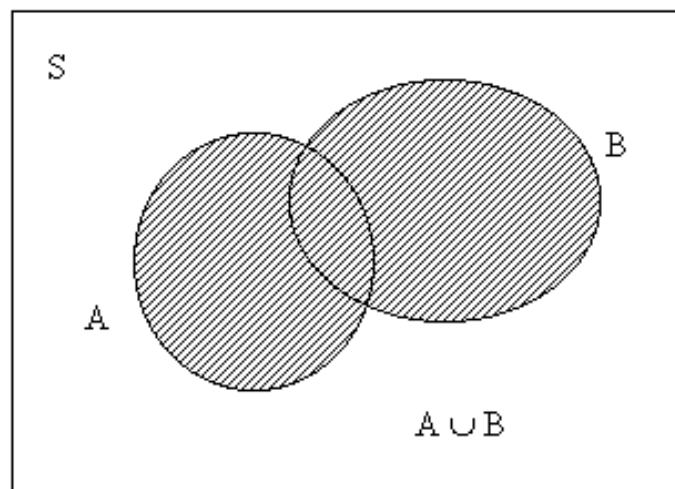


Fig. 2.1 – União dos conjuntos A e B

Observação:

- 1) $A \cup B = B \cup A$
- 2) $A \cup A = A$
- 3) $A \cup \phi = A$
- 4) Se $A \subset B \Rightarrow A \cup B = B$ (em particular $A \cup S = S$).

A representação da união de n eventos A_1, A_2, \dots, A_n ($A_1 \cup A_2 \cup \dots \cup A_n$) é dada por:

$$\bigcup_{i=1}^n A_i \quad (2.4)$$

2.7.2 Intersecção de conjuntos

Definição: $A \cap B = \{e_i \in S \mid e_i \in A \text{ e } e_i \in B\}, i = 1, \dots, n$. Portanto, o evento intersecção é formado pelos pontos amostrais que pertençam simultaneamente aos eventos A e B . A intersecção pode ser vista na Figura 2.2.

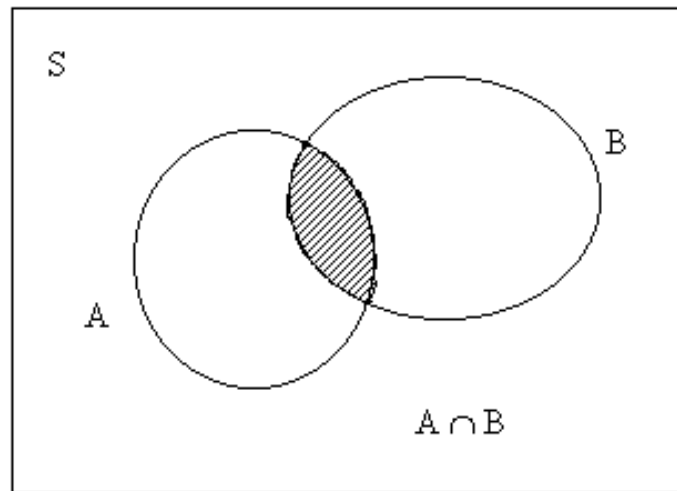


Fig. 2.2 – Intersecção dos conjuntos A e B

Observação:

- 1) $A \cap B = B \cap A$
- 2) $A \cap A = A$
- 3) $A \cap \phi = \phi$
- 4) Se $A \subset B \Rightarrow A \cap B = A$ (em particular $A \cap S = A$)
- 5) $(A \cap B) \cap C = A \cap (B \cap C)$.

A representação da intersecção de n eventos A_1, A_2, \dots, A_n ($A_1 \cap A_2 \cap \dots \cap A_n$) é dada

por:

$$\bigcap_{i=1}^n A_i \quad (2.5)$$

2.7.3 Complemento

Definição: $S - A = \bar{A} = A^c = \{e_i \in S / e_i \notin A\}, i = 1, \dots, n$. O complemento de um evento A é, portanto, o evento contendo todos os resultados no espaço amostral S que não pertençam a A . O complemento de A pode ser visto na Figura 2.3. Observação:

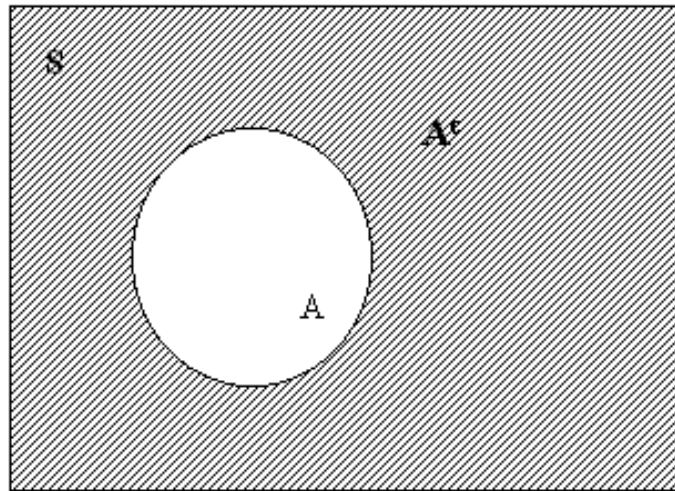


Fig. 2.3 – Complemento do conjunto A

- 1) $(A^c)^c = A$
- 2) $A \cup A^c = S$
- 3) $\phi^c = S$
- 4) $A \cap A^c = \phi$
- 5) $S^c = \phi$.

2.7.4 Eventos Mutuamente Exclusivos ou Disjuntos

Definição: Dois eventos são ditos *mutuamente exclusivos* ou *disjuntos* se A e B não puderem ocorrer juntos, ou seja a realização de um exclui a realização do outro. Segue que A e B são disjuntos se $A \cap B = \phi$.

Exemplo: Lançam-se duas moedas. Sejam os eventos

A : saída de faces iguais e

B : saída de cara na primeira moeda.

Determinar os eventos:

$A \cup B$, $A \cap B$, \bar{A} , \bar{B} , $\overline{(A \cup B)}$, $\overline{(A \cap B)}$, $(\bar{A} \cap \bar{B})$, $(\bar{A} \cup \bar{B})$, $B - A$, $A - B$, $\bar{A} \cap B$, e $\bar{B} \cap A$.

Resolução:

$$S = \{(c, c), (c, r), (r, c), (r, r)\}$$

$$A = \{(c, c), (r, r)\}$$

$$B = \{(c, c), (c, r)\}.$$

$$A \cup B = \{(c, c), (c, r), (r, r)\}$$

$$A \cap B = \{(c, c)\}$$

$$\bar{A} = \{(c, r), (r, c)\}$$

$$\bar{B} = \{(r, c), (r, r)\}$$

$$\overline{(A \cup B)} = \{(r, c)\}$$

$$\overline{(A \cap B)} = \{(c, r), (r, c), (r, r)\}$$

$$(\bar{A} \cap \bar{B}) = \{(r, c)\}$$

$$(\bar{A} \cup \bar{B}) = \{(c, r), (r, c), (r, r)\}$$

$$B - A = \{(c, r)\}$$

$$A - B = \{(r, r)\}$$

$$\bar{A} \cap B = \{(c, r)\}$$

$$\bar{B} \cap A = \{(r, r)\}.$$

Obs: Note que $\overline{(A \cup B)} = (\bar{A} \cap \bar{B})$ e $\overline{(A \cap B)} = (\bar{A} \cup \bar{B})$

2.8 Definição Axiomática de Probabilidade

Para um dado experimento, é necessário atribuir para cada evento A no espaço amostral S um número $P(A)$ que indica a probabilidade de A ocorrer. Para satisfazer a definição matemática de probabilidade, este número $P(A)$ deve satisfazer três axiomas específicos:

Axioma 1: Para qualquer evento A , $P(A) \geq 0$

Axioma 2: $P(S) = 1$

Axioma 3: Para qualquer seqüência infinita de eventos disjuntos A_1, A_2, \dots

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (2.6)$$

A definição matemática de probabilidade pode agora ser dada como segue:

A **distribuição de probabilidade**, ou simplesmente a probabilidade, no espaço amostral S é uma especificação de números $P(A)$ que satisfazem os axiomas 1, 2, e 3.

Teorema 1: $P(\phi) = 0$

Teorema 2: Para qualquer seqüência finita de eventos disjuntos A_1, A_2, \dots, A_n

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (2.7)$$

Teorema 3: Para qualquer evento A ,

$$P(A^c) = 1 - P(A) \quad (2.8)$$

Teorema 4: Para qualquer evento A , $0 \leq P(A) \leq 1$

Teorema 5: Se $A \subset B$, então $P(A) \leq P(B)$

Teorema 6: Para qualquer dois eventos A e B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.9)$$

Observação:

1.

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &= - [P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_1 \cap A_3)] \\ &= +P(A_1 \cap A_2 \cap A_3) \end{aligned} \quad (2.10)$$

2. Se dois eventos são mutuamente exclusivos, então $P(A \cup B) = P(A) + P(B)$.

2.9 Eventos Independentes

Suponha que dois eventos A e B ocorram independentes um do outro no sentido que a ocorrência ou não de um deles tenha nenhuma relação, e nenhuma influência na ocorrência ou na não ocorrência do outro. Nessas condições

$$P(A \cap B) = P(A) \cdot P(B) \quad (2.11)$$

Definição: Dois eventos são independentes se $P(A \cap B) = P(A) \cdot P(B)$

Teorema 1: Se dois eventos A e B são independentes, então os eventos A e B^c também são independentes.

Exercícios Complementares:

- 1 - De dois baralhos de 52 cartas retiram-se, simultaneamente, uma carta do primeiro baralho e uma carta do segundo. Qual a probabilidade de a carta do primeiro baralho ser um rei e a do segundo ser o 5 de paus?
- 2 - Uma urna A contém 3 bolas brancas, 4 pretas, 2 verdes; uma urna B contém 5 bolas brancas, 2 pretas, 1 verde; uma urna C contém 2 bolas brancas, 3 pretas, 4 verdes. Uma bola é retirada de cada urna. Qual a probabilidade de as três bolas retiradas da primeira, segunda e terceira urnas serem respectivamente, branca, preta e verde?
- 3 - De um baralho de 52 cartas retiram-se, ao acaso, duas cartas sem reposição. Qual a probabilidade de a primeira ser o ás de paus e a segunda ser o rei de paus?
- 4 - São dados dois baralhos de 52 cartas. Tiramos, ao mesmo tempo, uma carta do primeiro baralho e uma carta do segundo. Qual a probabilidade de tiramos uma dama e um rei, não necessariamente nessa ordem?
- 5 - Dois dados são lançados conjuntamente. Determine a probabilidade de a soma ser 10 ou maior que 10.
- 6 - A probabilidade de que um homem esteja vivo daqui a 30 anos é $2/5$; a

de sua mulher é de $2/3$. Determinar a probabilidade de que daqui a 30 anos:

- a) ambos estejam vivos;
- b) somente o homem esteja vivo;
- c) somente a mulher esteja viva;
- d) nenhum esteja vivo;
- e) pelo menos um esteja vivo.

7 - Sejam A e B eventos tais que $P(A) = 0,2$; $P(B) = p$; $P(A \cup B) = 0,6$. Calcular p considerando A e B:

- a) Mutuamente exclusivos
- b) Independentes.

2.10 Análise Combinatoria

Definição: Defini-se fatorial de n por $n!$

$$n! = n(n - 1)(n - 2)(n - 3)\dots 1$$

Por definição $0! = 1$

2.10.1 Permutação

Amostra sem reposição: Uma permutação de n objetos diferentes, tomados r de cada vez, é um arranjo de r dos n objetos, levando-se em consideração a ordem de sua disposição.

O número de permutações de n objetos, tomados r de cada vez é representado por P_n^r e calculado por

$$P_n^r = n(n - 1)(n - 2)\dots(n - r + 1) = \frac{n!}{(n - r)!} \quad (2.12)$$

Em particular $P_n^n = n!$

O número de permutações de n objetos distribuidos em grupos dos quais n_1 são

iguais, n_2 são iguais, ... n_k são iguais, é:

$$\frac{n!}{n_1!n_2!\dots n_k!} \quad \text{onde } n_1 + n_2 + \dots + n_k = n \quad (2.13)$$

Amostragem com Reposição: Considere uma amostragem com reposição, como por exemplo uma urna contendo n bolas e se selecione k bolas, uma de cada vez, repondo a bola na urna após o evento. O espaço amostral S será composto por n^k elementos.

Exemplo:

1. Suponha que um clube possua 25 membros, e que um presidente e um secretário serão escolhidos entre os membros. De quantas maneiras estes cargos podem ser preenchidos?

Solução: Como as posições podem ser preenchidas, escolhendo-se primeiro o presidente dentre os 25 membros, e depois o secretário dentre os 24 restantes, o número total de maneiras que os cargos poderão ser preenchidos será $P_{25}^2 = (25)(24) = 600$.

2. Suponha que seis livros diferentes serão arrumados em uma estante. O número de possíveis permutações dos livros é $6! = 720$.

3. O número de permutações da palavra *estatística* é $\frac{11!}{3!2!2!1!1!1!} = 831.600$.

2.10.2 Combinações

Uma combinação de n objetos tomados r de cada vez, é uma escolha dos n objetos, não se levando em consideração a ordem de sua posição. O número de combinações de n objetos, tomados k de cada vez, é representado por C_n^k ou $\binom{n}{k}$.

O cálculo (ou fórmula) para combinações pode ser obtido através de uma permutação que pode ser construída da seguinte maneira: Sabe-se que o número de permutações de n elementos tomados k de cada vez é P_n^k . Primeiro uma combinação particular de k elementos é selecionada. Cada diferente permutação desses k elementos levará a uma permutação na lista. Como há $k!$ permutações desses k elementos, esta combinação particular produzirá $k!$ permutações na lista. Quando uma combinação diferente de k elementos é selecionada, $k!$ outras permutações na lista são obtidas. Como cada combinação de k elementos produzirá $k!$ permutações, o número total

de permutações na lista será de $k! \cdot C_n^k$, isto é $P_n^k = k! \cdot C_n^k$. Portanto

$$C_n^k = \frac{P_n^k}{k!} = \frac{n!}{k!(n-k)!} \quad (2.14)$$

2.11 Probabilidade Condicional

Se A e B são dois eventos, a probabilidade de A ocorrer, depois de B ter acontecido, é representada por $P(A/B)$ (*probabilidade de A dado B*) e é denominada **probabilidade condicional** de A , depois de B ter ocorrido.

Neste caso, a probabilidade do evento A muda após se ter aprendido que o evento B ocorreu. Como se sabe que o evento B ocorreu, então sabemos que o resultado do evento A será um dos incluídos em B . Então, para calcular a probabilidade que A ocorrerá, devemos considerar o conjunto dos possíveis resultados de B que também resultariam na ocorrência de A . Este conjunto é precisamente $A \cap B$. É portanto natural definir-se a probabilidade condicional $P(A/B)$ como a proporção da probabilidade total $P(B)$ que é representada pela probabilidade $P(A \cap B)$. Portanto, tem-se a seguinte definição:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{dado } P(B) > 0 \quad (2.15)$$

Se $P(B) = 0$ a $P(A/B)$ não é definida.

Probabilidade Condicional para Eventos Independentes

Se A e B forem independentes, então $P(A \cap B) = P(A) \cdot P(B)$. Logo,

$$P(A/B) = \frac{P(A) \cdot P(B)}{P(B)} = P(A) \quad (2.16)$$

Da mesma forma,

$$P(B/A) = \frac{P(B) \cdot P(A)}{P(A)} = P(B) \quad (2.17)$$

Teorema: Suponha que A_1, A_2, \dots, A_n sejam quaisquer eventos tais que

$P(A_1) > 0, P(A_1 \cap A_2) > 0, \dots, P(A_1 \cap A_2 \cap A_3 \dots \cap A_{n-1}) > 0$. Então

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= \\ &= P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \dots \\ &\quad P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned} \quad (2.18)$$

Exemplo:

Suponha que 4 bolas sejam selecionadas, uma de cada vez, sem reposição, de uma urna contendo v bolas vermelhas e a azuis. ($v \geq 2, a \geq 2$). Qual a probabilidade de se obter uma sequência de resultados vermelho, azul, vermelho, azul?

v_j : uma bola vermelha é retirada na j -ésima vez

a_j : uma bola azul é retirada na j -ésima vez

onde $j=1,2,3,4$

$$\begin{aligned} P(v_1, a_2, v_3, a_4) &= \\ &= P(v_1) \cdot P(a_2/v_1) \cdot P(v_3/v_1 \cap a_2) \cdot P(a_4/v_1 \cap a_2 \cap v_3) \\ &= \frac{v}{v+a} \cdot \frac{a}{v+a-1} \cdot \frac{v-1}{v+a-2} \cdot \frac{a-1}{v+a-3} \end{aligned}$$

2.12 Probabilidade Total

Seja S o espaço amostral de um experimento, e considere k eventos A_1, A_2, \dots, A_k em S tal que A_1, A_2, \dots, A_k sejam disjuntos e $\bigcup_{i=1}^k A_i = S$. Diz-se, então, que estes eventos formam uma *partição de S*.

Se os eventos A_1, A_2, \dots, A_k formam uma partição de S , e B é qualquer outro evento em S , então:

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_k \cap B).$$

Como os k eventos do lado direito da equação acima são disjuntos:

$$P(B) = \sum_{j=1}^k P(A_j \cap B) \quad (2.19)$$

Mas $P(A_j \cap B) = P(A_j) \cdot P(B/A_j)$ onde $j = 1, 2, \dots, k$. Então

$$P(B) = \sum_{j=1}^k P(A_j) \cdot P(B/A_j) \quad (2.20)$$

De forma a apresentar a fórmula para a probabilidade total e a fórmula de Bayes, vamos supor que exista numa imagem duas classes de cobertura terrestre, denominadas A_1 e A_2 , e que os *pixels* desta imagem serão classificados como pertencentes a A_1 ou A_2 segundo o valor de alguma característica (por exemplo, nível de cinza). Suponha que a probabilidade de ocorrência da classe A_1 seja $P(A_1)$ e a probabilidade de ocorrência da classe A_2 seja $P(A_2)$. Obviamente *pixels* da classe A_1 podem ser erroneamente classificados como sendo da classe A_2 , e vice versa. Designemos estas probabilidades, $P(C_{A_2}/A_1)$ e $P(C_{A_1}/A_2)$ respectivamente.

Suponha que seja tomado aleatoriamente um *pixel* desta imagem e classificado, segundo o valor da característica estudada. Deseja-se saber a probabilidade deste *pixel* ser classificado como A_1 .

Aqui, temos que considerar o seguinte: o *pixel* pode ser classificado em $A_1(C_{A_1})$ e ser realmente de A_1 ou pode ser classificado em $A_1(C_{A_1})$ e ser realmente de A_2 . Então podemos escrever que:

$$C_{A_1} = (C_{A_1} \cap A_1) \cup (C_{A_1} \cap A_2)$$

Logo,

$$P(C_{A_1}) = P[(C_{A_1} \cap A_1) \cup (C_{A_1} \cap A_2)]$$

como os eventos são mutuamente exclusivos, temos:

$$P(C_{A_1}) = P(C_{A_1} \cap A_1) + P(C_{A_1} \cap A_2)$$

Uma vez que os eventos C_{A_1} e A_1 ($i = 1, 2$) não são independentes, vem:

$$P(C_{A_1}) = P(C_{A_1}/A_1)P(A_1) + P(C_{A_1}/A_2)P(A_2)$$

ou

$$P(C_{A_1}) = \sum_i P(C_{A_1} \cap A_i) = \sum_i P(C_{A_1}/A_i)P(A_i) \quad (2.21)$$

A expressão da equação 2.21 acima é normalmente chamada de fórmula para a probabilidade total.

2.13 Teorema de Bayes

Sejam os eventos A_1, A_2, \dots, A_k que formam uma partição do espaço S tal que $P(A_j) > 0$ para todo $j = 1, 2, \dots, k$, e seja B qualquer evento tal que $P(B) > 0$. Então, para $i = 1, 2, \dots, k$, temos:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{j=1}^k P(A_j)P(B/A_j)} \quad (2.22)$$

Prova: Pela definição de probabilidade condicional,

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} \quad (2.23)$$

O numerador da equação 2.22 é igual a $P(A_i \cap B)$ e o denominador é igual a $P(B)$ (pela fórmula para probabilidade total).

Imaginemos agora uma situação onde um *pixel* é classificado como A_1 . Quer-se saber qual a probabilidade dele ser realmente um *pixel* da classe A_1 . Note que aqui é fornecida uma informação sobre o resultado da classificação do *pixel*, isto é, é dito que ele foi classificado como pertencente à classe A_1 dado que o mesmo foi classificado em A_1 .

Em outras palavras, quer-se

$$P(A_1/C_{A_1}) = \frac{P(A_1 \cap C_{A_1})}{P(C_{A_1})} \quad (2.24)$$

Note que o denominador da expressão 2.24 acima é dado em 2.21, podendo-se

expressar

$$P(A_1/C_{A_1}) = \frac{P(C_{A_1}/A_1)P(A_1)}{\sum_i P(C_{A_1}/A_i)P(A_i)} \quad (2.25)$$

Esta formulação pode obviamente ser generalizada para uma situação onde os eventos A_1, A_2, \dots, A_n formam um sistema completo de resultados de alguma operação e onde K denota um resultado arbitrário desta operação. Neste caso tem-se que:

$$P(A_i/K) = \frac{P(A_i \cap K)}{P(K)} = \frac{P(K/A_i)P(A_i)}{\sum P(K/A_i)P(A_i)} \quad (2.26)$$

que é conhecida como a fórmula de Bayes e que tem aplicações diversas na área de sensoriamento remoto.

A seguir é dada uma aplicação específica utilizada em Classificação.

2.13.1 Teoria de Bayes

Suponha que seja medida alguma característica x de uma cena (por exemplo, o nível de cinza de cada *pixel*) e que tenha que se decidir a qual de duas classes (por exemplo, vegetação ou solo) um *pixel* pertence. Este é um problema unidimensional, de classificação em duas classes, no domínio de característica da imagem. Se um número grande de *pixels* está disponível, que pode ser considerado representativo de cada classe (isto é, dados de treinamento) podemos calcular um histograma da frequência relativa da característica para cada classe, conforme mostrado na Figura 2.4.

Considere-as como aproximações das funções densidade de probabilidade (f.d.p.) contínuas de uma amostra infinita de dados. Essas funções densidade de probabilidade condicionais, $P(x/A_1)$ e $P(x/A_2)$ tem área unitária e descrevem a probabilidade de um *pixel* ter valor x da característica, dado que ele está na classe A_1 ou na classe A_2 , respectivamente. Cada f.d.p. pode ser delineada pela probabilidade *a priori* $P(A_i)$ que a classe A_i ocorra na área de interesse na imagem.

Estas funções de probabilidade delineadas $P(x/A_i)P(A_i)$ representam a probabilidade que um *pixel* tenha o valor x na característica e está na classe A_i . Em

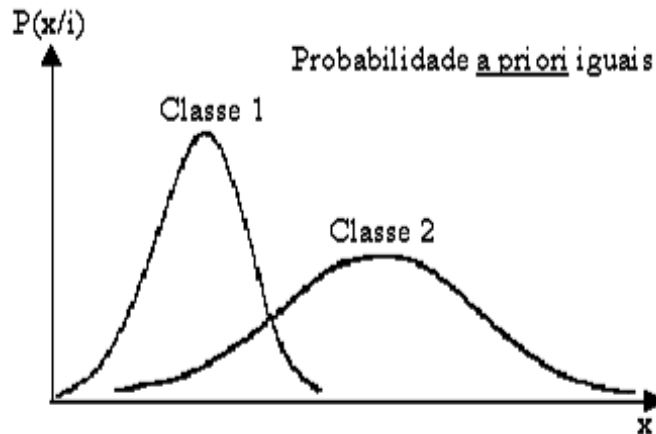


Fig. 2.4 – A probabilidade *a priori* nas f.d.p. das classes

sensoriamento remoto, estas probabilidades *a priori* podem ser estimadas através de fontes externas de informação sobre a cena, tais como pesquisas de campo, mapas ou dados históricos.

Para se tomar uma decisão de classificação para um *pixel*, precisamos conhecer as probabilidades *a posteriori* que o *pixel* pertença à cada uma das classes de treinamento, dado que o *pixel* tem valor x na característica. Esta probabilidade $P(A_i/x)$ pode ser calculada com a regra de Bayes

$$P(A_i/x) = \frac{P(x/A_i)P(A_i)}{P(x)} \quad (2.27)$$

Onde

$$P(x) = \sum_1^2 P(x/A_i)P(A_i) \quad (2.28)$$

Uma regra de decisão pode agora ser formada com as probabilidades *a posteriori* da equação 2.27. Se um *pixel* tem valor x na característica, uma abordagem intuitiva satisfatória é designar o *pixel* à classe A_1 se $P(A_1/x)$ é maior que $P(A_2/x)$. Semelhantemente, o *pixel* seria designado à classe A_2 se $P(A_2/x)$ é maior que $P(A_1/x)$. Sendo $P(x)$ igual para as duas classes na equação 2.28 ela pode ser ignorada numa comparação dos dois e podemos escrever a regra de decisão de Bayes

- um *pixel* pertence à *classe* A_1 se $P(x/A_1)P(A_1) > P(x/A_2)P(A_2)$

- um *pixel* pertence à *classe* A_2 se $P(x/A_2)P(A_2) > P(x/A_1)P(A_1)$

Numa situação atípica onde as duas probabilidades *a posteriori* são exatamente iguais, isto é

$$P(A_1/x) = P(A_2/x)$$

ou

$$P(x/A_1)P(A_1) = P(x/A_2)P(A_2)$$

uma decisão não pode ser tomada a partir das probabilidades de classe. Um processo de desempate deve ser empregado, tal como escolher aleatoriamente *classe 1* ou *classe 2*. Pode ser mostrado que a regra de decisão de Bayes minimiza a probabilidade média de erro sobre todo o conjunto de dados classificados, se todas as classes tem f.d.p. normal.

Na prática, as probabilidades $P(A_i)$ são difíceis de ser obtidas e conseqüentemente supõe-se que elas sejam iguais. Obviamente resultados mais exatos seriam obtidos se elas pudessem ser estimadas a partir de dados externos. Se, por exemplo, o objetivo é determinar a proporção dos tipos de cultura plantados numa estação particular, a partir de imagens do *Landsat* de uma área agrícola, podemos sensatamente estabelecer as probabilidades *a priori* iguais às estimativas históricas da porcentagem de cada cultura na área.

CAPÍTULO 3

VARIÁVEIS ALEATÓRIAS E DISTRIBUIÇÕES

3.1 Variável Aleatória

Definição: Considere um experimento para o qual o espaço amostral é denotado por S . Define-se variável aleatória como uma função que associa um valor real a cada elemento do espaço amostral.

$$X : S \longrightarrow \mathfrak{R}$$

Representamos as variáveis aleatórias por letras maiúsculas e suas ocorrências por letras minúsculas.

Exemplo:

Suponha o experimento "lançar três moedas". Seja X : número de ocorrências da face cara. O espaço amostral do experimento é:

$$S = \{(c, c, c), (c, c, r), (c, r, c), (c, r, r), (r, c, c), (r, c, r), (r, r, c), (r, r, r)\}.$$

Se X é o número de caras, X assume os valores 0, 1, 2 e 3. Podemos associar a esses números eventos que correspondem a nenhuma, uma, duas ou três caras respectivamente, como segue:

X	Evento Correspondente
0	$A_1 = \{(r, r, r)\}$
1	$A_2 = \{(c, r, r), (r, c, r), (r, r, c)\}$
2	$A_3 = \{(c, c, r), (c, r, c), (r, c, c)\}$
3	$A_4 = \{(c, c, c)\}$

3.2 Tipos de Variáveis Aleatórias, Função de Probabilidade e Função Densidade de Probabilidade

Definição: Seja X uma variável aleatória (*v.a.*). Se o número de valores possíveis de X (isto é o seu contradomínio), for finito ou infinito enumerável, denominaremos X de variável aleatória **discreta**. Isto é, existe um conjunto finito ou enumerável $\{x_1, x_2, \dots\} \subset \mathfrak{R}$ tal que $X(s) \in \{x_1, x_2, \dots\} \quad \forall s \in S$.

Definição: Seja X uma variável aleatória discreta. Portanto, o contradomínio de X será formado por um número finito ou enumerável de valores x_1, x_2, \dots . A cada

possível resultado x_i , associaremos um número $p(x_i) = P(X = x_i)$, $i = 1, 2, 3, \dots$, denominado probabilidade de x_i . Os números $p(x_i)$ devem satisfazer às seguintes condições:

- a) $p(x_i) \geq 0 \quad \forall i$,
- b) $\sum_{i=1}^{\infty} p(x_i) = 1$

A função p , definida acima, é denominada **função de probabilidade** da variável aleatória X . A coleção de pares $[x_i, p(x_i)]$ $i = 1, 2, \dots$, é denominada **distribuição de probabilidade** de X .

Exemplos:

- a) Lançam-se dois dados. Seja a v.a. X : soma das faces. Determinar a distribuição de probabilidade da variável aleatória X (Figura 3.1).

X	2	3	4	5	6	7	8	9	10	11	12
p(X)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

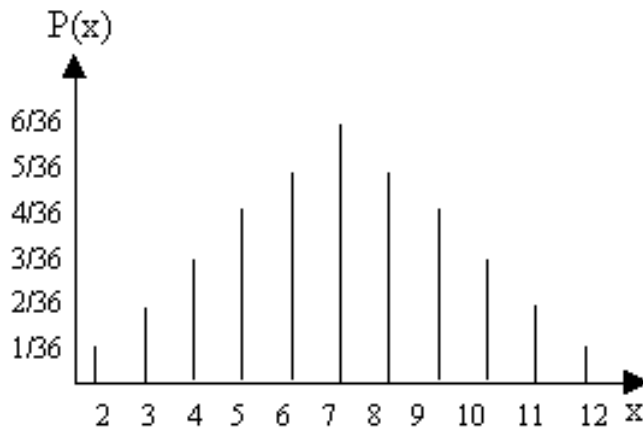


Fig. 3.1 – Gráfico de $P(x)$ para dois dados

- b) Considere o experimento no qual uma moeda é jogada dez vezes e seja X ser o número de caras que são obtidas. Neste experimento, os possíveis

valores de X são $0, 1, 2, \dots, 10$, e

$$P(X = x) = \binom{n}{x} \frac{1}{2^{10}}, \quad x = 0, 1, 2, \dots, 10. \quad (3.1)$$

Definição: Seja X uma variável aleatória. Suponha que \mathfrak{R}_X , o contradomínio de X , seja um intervalo ou uma coleção de intervalos. Então diremos que X é uma variável aleatória contínua.

Definição: Seja X uma variável aleatória contínua. A **função densidade de probabilidade** f , indicada abreviadamente por f.d.p., é uma função f que satisfaz às seguintes condições:

- a) $f(x) \geq 0 \quad x \in \mathfrak{R}_X$,
- b) $\int_{\mathfrak{R}_X} f(x) dx = 1$

Além disso, definimos, para qualquer $c < d$ (em \mathfrak{R}_X)

$$P(c < x < d) = \int_c^d f(x) dx$$

Obs:

- a) $P(c < x < d)$ representa a área sob a curva, como exemplificado no gráfico da Figura 3.2, da f.d.p. f , entre $x = c$ e $x = d$.
- b) Constitui uma consequência da descrição probabilística de X que, para qualquer valor especificado de X , digamos x_0 , teremos $P(X = x_0) = 0$, porque $P(X = x_0) = \int_{x_0}^{x_0} f(x) dx = 0$

Exemplo:

- a) Suponhamos que a v.a. X seja contínua. Seja a f.d.p. f dada por:

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

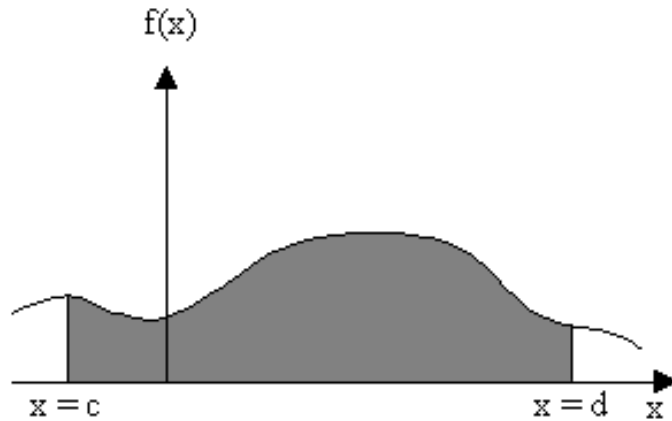


Fig. 3.2 – Gráfico de $f dp$ de f

Evidentemente, $f(x) \geq 0$ e $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 2x dx = 1$. Para calcular $P(X \leq 1/2)$ deve-se calcular $\int_0^{1/2} (2x)dx = \frac{1}{4}$.

- b) Seja X a duração de vida (em horas) de um certo tipo de lâmpada. Admitindo que X seja uma variável aleatória contínua, suponha que a f.d.p. f de X seja dada por:

$$f(x) = \begin{cases} \frac{a}{x^3} & 1500 \leq x \leq 2500, \\ 0 & \text{caso contrário.} \end{cases}$$

Para calcular a constante a , recorre-se à condição $\int_{-\infty}^{\infty} f(x)dx = 1$, que significa, neste caso $\int_{1500}^{2500} \frac{a}{x^3} dx = 1$, obtendo-se $a = 7.031.250$.

3.3 Função de Distribuição Acumulada

Definição: A *função de distribuição* da variável aleatória X , representada por F_X ou simplesmente por F , é definida por:

$$F_X(x) = P(X \leq x) \tag{3.2}$$

Observação:

- a) A função de distribuição de X é também frequentemente chamada de **função de distribuição acumulada** de X .

b) A função $F_X(x)$ é não-decrescente quando x aumenta, isto é, se $x_1 < x_2$, então $F_X(x_1) \leq F_X(x_2)$.

c) Para qualquer valor de x

$$P(X > x) = 1 - F_X(x) \quad (3.3)$$

d) Para quaisquer valores x_1 e x_2 , tais que $x_1 < x_2$,

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \quad (3.4)$$

Teoremas:

a) Se X for uma variável aleatória discreta,

$$F_X(x) = \sum_j p(x_j),$$

onde o somatório é estendido a todos os índices j que satisfaçam à condição $x_j \leq x$.

b) Se X for uma variável aleatória contínua com f.d.p. f ,

$$F_X(x) = \int_{-\infty}^x f(s) ds.$$

Exemplos:

a) Suponhamos que a v.a. X tome os três valores 0,1, e 2, com probabilidades $1/3$, $1/6$ e $1/2$, respectivamente. Então:

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

O gráfico de F está apresentado na Figura 3.3.

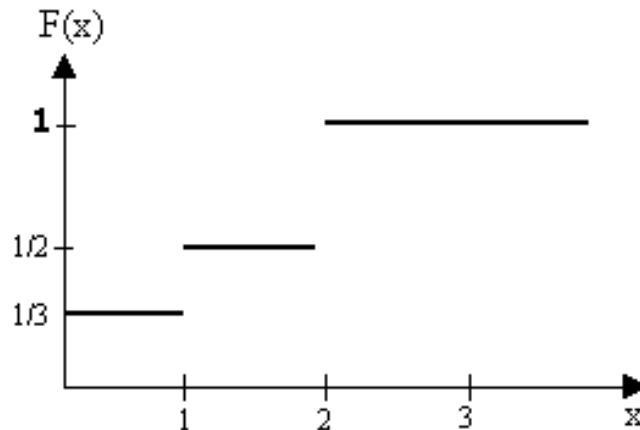


Fig. 3.3 – Gráfico F

b) Suponha que X seja uma variável contínua com f.d.p.

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

Portanto, a f.d. é dada por:

$$F(x) = \begin{cases} 0 & x \leq 0, \\ \int_0^x (2s) ds = x^2 & 0 < x \leq 1, \\ 1 & x > 1. \end{cases}$$

O gráfico está apresentado na Figura 3.4.

3.4 Distribuições Bivariadas

Em alguns experimentos, é necessário considerar as propriedades de 2 ou mais variáveis simultaneamente. A distribuição de probabilidade conjunta de duas *v.a.* é denominada uma **distribuição bivariada**.

3.4.1 Distribuições Conjuntas Discretas

Suponha que um certo experimento envolve duas *v.a.* X e Y , cada qual com uma distribuição discreta.

A **função de probabilidade conjunta** de X e Y é definida pela função p tal que

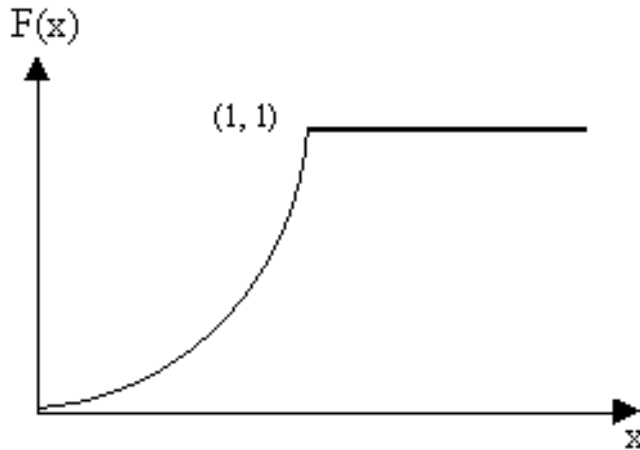


Fig. 3.4 – Meyer, página 75.

qualquer ponto (x, y) no plano xy ,

$$p(x, y) = P(X = x \text{ e } Y = y) \quad (3.5)$$

Observação:

- a) $\sum p(x_i, y_i) = 1.$
- b) Se X e Y forem independentes
 $p(x_i, y_i) = P(X = x_i) \cdot P(Y = y_i)$

Exemplo:

Suponha que a variável aleatória X possa assumir somente os valores 1, 2, e 3, que a variável aleatória Y possa assumir somente os valores 1, 2, 3 e 4, e que a função de probabilidade conjunta de X e Y seja dada pela tabela:

	Y	1	2	3	4
X					
1		0,1	0	0,1	0
2		0,3	0	0,1	0,2
3		0	0,2	0	0

A função de probabilidade conjunto é mostrada na Figura 3.5. Deseja-se determinar:

- a) $P(X \geq 2 \text{ e } Y \geq 2)$;
 b) $P(X = 1)$.

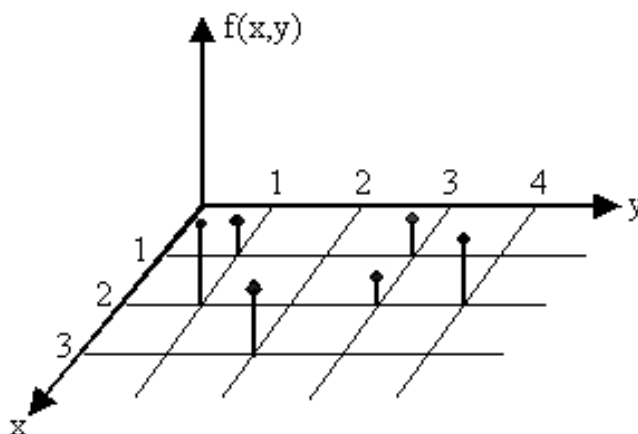


Fig. 3.5 – Degroot, página 93.

- a) Somando-se $p(x, y)$ para todos os valores de $x \geq 2$ e $y \geq 2$, tem-se:

$$P(X \geq 2 \text{ e } Y \geq 2) = p(2, 2) + p(2, 3) + p(2, 4) + p(3, 2) + p(3, 3) + p(3, 4) = 0,5$$
- b) $P(X = 1) = \sum_{y=1}^4 p(1, y) = 0,2$.

3.4.2 Distribuições Conjuntas Contínuas

É dito que duas *v.a.* X e Y possuem uma **distribuição conjunta contínua** se existe uma função f não negativa, definida sobre o plano xy , tal que para qualquer subconjunto A do plano

$$P[(x, y) \in A] = \int_A \int f(x, y) dx dy \quad (3.6)$$

A função f é denominada **função densidade de probabilidade conjunta** de X e Y . Esta função deve satisfazer

$$f(x, y) \geq 0 \quad \text{para} \quad -\infty < x < \infty \quad \text{e} \quad -\infty < y < \infty$$

e

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \quad (3.7)$$

A probabilidade que o par (X, Y) pertença a uma região do plano xy pode ser encontrada integrando a f.d.p. conjunta sobre esta região. A Figura 3.5 mostra um exemplo de f.d.p. conjunta.

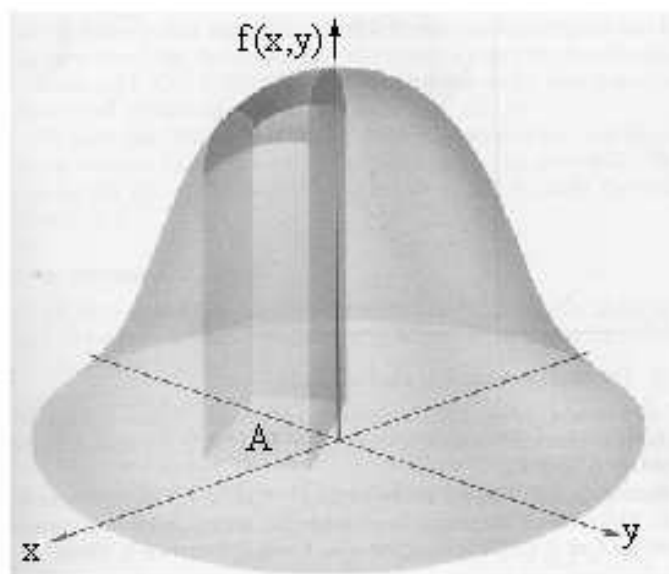


Fig. 3.6 – Degroot, página 95.

3.5 Distribuições Marginais

Denomina-se **função densidade de probabilidade marginal** de X à função densidade de probabilidade de X quando ela é obtida através da f.d.p. conjunta de X e Y .

Caso Discreto

Se X e Y são *v.a.* discretas com f.p. conjunta $p(x, y)$, então a f.p. marginal de X é obtida por:

$$P_X(x) = P(X = x) = \sum_y p(x, y) \quad (3.8)$$

Similarmente, a f.p. marginal de Y é:

$$P_Y(y) = P(Y = y) = \sum_x p(x, y) \quad (3.9)$$

Exemplo:

Suponha que a variável aleatória X possa assumir somente os valores 1, 2, e 3, que a variável aleatória Y possa assumir somente os valores 1, 2, 3 e 4, e que a função de probabilidade conjunta de X e Y seja dada pela tabela:

	Y	1	2	3	4	Marginal de X
X						
1		0,1	0	0,1	0	0,2
2		0,3	0	0,1	0,2	0,6
3		0	0,2	0	0	0,2
Marginal de Y		0,4	0,2	0,2	0,2	1,0

A f.p. marginal de X é determinada somando-se os valores de cada linha da tabela, e é dada na última coluna da tabela. Analogamente a f.p. marginal de Y é dada na última linha da tabela.

Caso Contínuo

Se X e Y possuem uma distribuição conjunta com f.d.p. conjunta $f(x, y)$, então a f.d.p. marginal $f_X(x)$ de X é obtida por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (3.10)$$

Similarmente, a f.d.p. marginal de Y é obtida por:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (3.11)$$

Observação

Se X e Y forem independentes

$$P(x, y) = P_X(x) \cdot P_Y(y) \quad (\text{Caso discreto})$$

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad (\text{Caso contínuo})$$

Exemplo:

Suponha que X e Y possuam uma distribuição contínua conjunta, cuja p.d.f. conjunta seja definida por:

$$f(x, y) = \begin{cases} \frac{3}{2}y^2 & 0 \leq x \leq 2 \text{ e } 0 \leq y \leq 1, \\ 0 & \text{caso contrário.} \end{cases}$$

- a) Determine as p.d.f.'s marginais de X e Y ;
- b) X e Y são independentes?
- c) Determine $P(X \geq \frac{1}{2})$ e $P(Y \geq \frac{1}{2})$.

Solução:

a) Tem-se que:

$$\int_0^1 \frac{3}{2}y^2 dy = \frac{y^3}{2} \Big|_0^1 = \frac{1}{2}.$$

Logo, a p.d.f. marginal de X é dada por:

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

Tem-se que:

$$\int_0^2 \frac{3}{2}y^2 dx = \frac{3}{2}y^2 x \Big|_0^2 = 3y^2$$

Logo, a p.d.f. marginal de Y é dada por:

$$f_Y(y) = \begin{cases} 3y^2 & 0 \leq y \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

b) Tem-se que $f(x, y) = f_X(x) \cdot f_Y(y)$. Logo X e Y são independentes.

c)

$$P(X \geq \frac{1}{2}) = \int_{\frac{1}{2}}^2 \frac{1}{2} dx = \frac{1}{2} x \Big|_{\frac{1}{2}}^2 = \frac{1}{2} \left(2 - \frac{1}{2} \right) = \frac{3}{4}.$$

$$P(Y \geq \frac{1}{2}) = \int_{1/2}^1 3y^2 dy = y^3 \Big|_{1/2}^1 = 1 - \frac{1}{8} = \frac{7}{8}.$$

3.6 Esperança de Uma Variável Aleatória

3.6.1 Distribuições Discretas

Suponha que uma variável aleatória (*v.a.*) X possua uma distribuição discreta cuja f.d.p. é $p(x)$. A **esperança** de X , denotada por $E(X)$, é um número definido por

$$\mu = E(X) = \sum_x xp(x) \tag{3.12}$$

Exemplo

Suponha que uma *v.a.* X possa assumir somente quatro valores: $-2, 0, 1, e 4$, e que $P(X = -2) = 0.1$; $P(X = 0) = 0.4$; $P(X = 1) = 0.3$; $P(X = 4) = 0.2$

Então

$$\begin{aligned} E(x) &= -2 \cdot (0.1) + 0 \cdot (0.4) + 1 \cdot (0.3) + 4 \cdot (0.2) \\ &= 0.9 \end{aligned}$$

3.6.2 Distribuições Contínuas

Se uma variável aleatória (*v.a.*) X possui uma distribuição contínua com f.d.p. $f(x)$, então a **esperança** $E(X)$ é definida por

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (3.13)$$

Exemplo

Suponha que a f.d.p. de uma *v.a.* X com uma distribuição contínua seja:

$$f(x) = \begin{cases} 2x & \text{para } 0 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Então

$$\begin{aligned} E(X) &= \int_0^1 x \cdot (2x) dx \\ &= \int_0^1 2x^2 dx \\ &= \frac{2x^3}{3} \Big|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

Observação

O número $E(X)$ é também denominado **valor esperado** de X , ou a **média** de X .

3.6.3 Propriedades da Esperança

P1.

Se a é uma constante qualquer

$$E(X \pm a) = E(X) \pm a$$

P2.

Se a é uma constante qualquer

$$E(aX) = a \cdot E(X)$$

P3.

Se X_1, X_2, \dots, X_n são n variáveis aleatórias tais que $E(X_i)$ existe ($i = 1, 2, \dots, n$), então

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

P4.

Se X_1, X_2, \dots, X_n são n variáveis aleatórias **independentes**, tais que $E(X_i)$ existe ($i = 1, 2, \dots, n$), então

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

Exemplos:

a) Suponha que $E(X) = 5$. Então:

$$E(3X - 5) = 3E(X) - 5 = 10 \quad \text{e}$$

$$E(-3X + 15) = -3E(X) + 15 = 0$$

b) Suponha que três v.a. X_1, X_2 e X_3 formem uma amostra aleatória de uma distribuição para o qual a média é 5. Determinar o valor de $E(2X_1 - 3X_2 + X_3 - 4)$

$$E(2X_1 - 3X_2 + X_3 - 4) = 2E(X_1) - 3E(X_2) + E(X_3) - 4$$

$$= 2(5) - 3(5) + 5 - 4$$

$$= 10 - 15 + 5 - 4 = -4$$

c) Suponha que X_1, X_2 e X_3 são v.a. independentes tais que $E(X_i) = 0$ e

$E(X_i^2 = 1)$, para $i = 1, 2, 3$. Determinar $E[X_1^2(X_2 - 4X_3)^2]$

$$\begin{aligned} E[X_1^2(X_2 - 4X_3)^2] &= E[X_1^2]E[(X_2 - 4X_3)^2] \\ &= E[X_1^2]E[(X_2^2 - 8E(X_2)E(X_3) + 16E(X_3^2))] \\ &= 1[1 - 8(0)(0) + 16(1)] = 17 \end{aligned}$$

3.7 Variância de uma Variável Aleatória

Definição

Suponha que X é uma *v.a.* com média $\mu = E(X)$. A **variância** de x , representada por $Var(X)$ é definida por

$$Var(X) = E[(x - \mu)^2], \quad \text{onde} \quad \mu = E(X) \quad (3.14)$$

3.7.1 Variáveis Aleatórias Discretas

Suponha que uma *v.a.* X possua uma distribuição discreta, cuja f.d.p. é $p(x)$. Então

$$\begin{aligned} Var(X) &= \sum_x (x - \mu)^2 \cdot p(x) \\ &= \sum_x x^2 \cdot p(x) - \mu^2 \end{aligned} \quad (3.15)$$

Exemplo:

Suponha que uma *v.a.* X possa assumir somente quatro valores: $-2, 0, 1$, e 4 , e que $P(X = -2) = 0,1$; $P(X = 0) = 0,4$; $P(X = 1) = 0,3$; $P(X = 4) = 0,2$

Como visto anteriormente, $E(X) = 0,9$. Então

$$\begin{aligned} Var(X) &= \sum_x (x - \mu)^2 \cdot p(x) \\ &= (-2 - 0,9)^2 \cdot (0,1) + (0 - 0,9)^2 \cdot (0,4) + (1 - 0,9)^2 \cdot (0,3) + (4 - 0,9)^2 \cdot (0,2) \\ &= 3,09 \end{aligned}$$

ou

$$\begin{aligned} \text{Var}(X) &= \sum_x x^2 \cdot p(x) - \mu^2 \\ &= (-2)^2 \cdot (0,1) + (0)^2 \cdot (0,4) + (1)^2 \cdot (0,3) + (4)^2 \cdot (0,2) - (0,9)^2 \\ &= 0,4 + 0,3 + 3,2 - 0,81 \\ &= 3,09 \end{aligned}$$

3.7.2 Variáveis Aleatórias Contínuas

Suponha que uma *v.a.* X possua uma distribuição contínua, cuja f.d.p. é $f(x)$. Então

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \mu^2 \end{aligned} \quad (3.16)$$

Exemplo

Suponha que a f.d.p. de uma *v.a.* X com uma distribuição contínua seja:

$$f(x) = \begin{cases} 2x & \text{para } 0 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Como visto anteriormente, $E(X) = \frac{2}{3}$. Então

$$\begin{aligned} \text{Var}(X) &= \int_0^1 x^2 \cdot (2x) dx - \left(\frac{2}{3}\right)^2 \\ &= \int_0^1 2x^3 dx - \left(\frac{2}{3}\right)^2 \\ &= \frac{2x^4}{4} \Big|_0^1 - \left(\frac{2}{3}\right)^2 \\ &= \frac{2}{4} - \frac{4}{9} = \frac{2}{36} \end{aligned}$$

3.7.3 Propriedades da Variância

P1.

$\text{Var}(X) = 0$ se e somente se existe uma constante c tal que $P(X = c) = 1$

P2.

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

P3.

$$\text{Var}(X + a) = \text{Var}(X)$$

P4.

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

P5.

Se X_1, X_2, \dots, X_n são v.a. **independentes**, então

$$\text{Var}(X_1 \pm X_2 \pm \dots \pm X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Exemplo:

Seja uma v.a. com média μ e desvio padrão σ . Calcular a média e variância de:

a) $Z = 3X - 7$

b) $Z = \frac{X-7}{2}$

c) $Z = \frac{X-\mu}{\sigma}$

a) $E(Z) = 3\mu - 7$
 $\text{Var}(Z) = 9\sigma^2$

b) $E(Z) = \frac{\mu}{2} - \frac{7}{2}$
 $\text{Var}(Z) = \frac{\sigma^2}{4}$

c) $E(Z) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$
 $\text{Var}(Z) = \frac{\sigma^2}{\sigma^2} = 1$

3.8 Momentos

Definição

Para qualquer variável aleatória (v.a.) X e qualquer inteiro positivo k , a esperança $E(X^k)$ é denominado **k-ésimo momento** de X , ou **momento de ordem k**

3.8.1 Momentos Centrais

Suponha que X seja uma *v.a.* com $E(X) = \mu$. Para qualquer inteiro positivo k , a esperança $E[(X - \mu)^k]$ é denominado **k-ésimo momento central** de X , ou **k-ésimo momento em torno da média**.

Observação

Se a distribuição de X é simétrica com respeito à sua média μ , e se o momento central $E[(X - \mu)^k]$ existe para um dado k ímpar, então o valor de $E[(X - \mu)^k]$ será igual a zero.

3.8.2 Função Geratriz de Momentos

Definição

Considere uma *v.a.* X , e para cada número real t , seja $M_x(t)$ a função

$$M_x(t) = E[e^{tx}] \quad (3.17)$$

Esta função é denominada **função geratriz de momentos** (*f.g.m.*)

A partir da *f.g.m.* pode-se gerar todos os momentos. Seja $M_x^k(t)$ a k -ésima derivada de $M_x(t)$. Então:

$$\begin{aligned} M_x^1(0) &= \left[\frac{d}{dt} E(e^{tx}) \right]_{t=0} \\ &= E \left[\left(\frac{d}{dt} e^{tx} \right)_{t=0} \right] \\ &= E[(X e^{tx})_{t=0}] \\ &= E(X) \end{aligned} \quad (3.18)$$

Analogamente, temos

$$\begin{aligned} M_x^k(0) &= \left[\frac{d^k}{dt^k} E(e^{tx}) \right]_{t=0} \\ &= E \left[\left(\frac{d^k}{dt^k} e^{tx} \right)_{t=0} \right] \\ &= E[(X^k e^{tx})_{t=0}] \\ &= E[X^k] \end{aligned} \quad (3.19)$$

Portanto:

$$M_x^1(0) = E(X), \quad M_x^2(0) = E(X^2), \quad M_x^3(0) = E(X^3), \dots$$

Exemplo:

Suponha que X seja uma v.a. com f.d.p. dada por

$$f(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ 0 & \text{caso contrario} \end{cases}$$

Determine a f.g.m. de X , e a $Var(X)$.

Solução:

$$M_x(t) = E[e^{tx}] = \int_0^{\infty} e^{tx} dx = \int_0^{\infty} e^{(t-1)x} dx = \frac{e^{(t-1)x}}{t-1} \Big|_0^{\infty}$$

A integral só será finita se e somente se $t < 1$. Neste caso,

$$\begin{aligned} M_X(t) &= \frac{1}{1-t} \\ M'_X(t) &= \frac{1}{(1-t)^2} \Rightarrow E(X) = 1 \\ M''_X(t) &= \frac{2}{(1-t)^3} \Rightarrow E(X^2) = 2 \end{aligned}$$

Logo, $Var(X) = 2 - 1^2 = 1$

3.8.3 Propriedades das Funções Geradoras de Momentos

P1.

Seja X uma v.a. com f.g.m. M_X , e seja $Y = aX + b$, onde a e b são constantes, e seja M_Y a f.g.m. de Y . Então, para qualquer valor de t tal que $M_X(at)$ exista,

$$M_Y(t) = e^{bt} M_X(at) \tag{3.20}$$

P2.

Suponha que X_1, X_2, \dots, X_n sejam n v.a. **independentes**, e que M_i ($i = 1, 2, \dots, n$) seja a f.g.m. de X_i . Seja $Y = X_1 + X_2 + \dots + X_n$, e seja M_Y a f.g.m. de Y . Então

para qualquer valor de t tal que $M_i(t)$ exista,

$$M_Y(t) = \prod_{i=1}^n M_i(t) \quad (3.21)$$

Exemplo:

Suponha que X e Y sejam independentes e identicamente distribuídas (i.i.d.) e que a f.g.m. de cada uma seja dada por:

$$M(t) = e^{t^2-3} \quad -\infty < t < \infty.$$

Encontre a f.g.m. de $Z = 3X - Y + 4$.

Solução:

Sejam

$$X_1 = 3X \Rightarrow M_1(t) = M(3t) = e^{9t^2-3}$$

$$X_2 = Y \Rightarrow M_2(t) = M(-t) = e^{t^2-3}$$

$$\begin{aligned} M_Z &= M(X_1 + X_2 + 4) = e^{4t} M_1(t) M_2(t) \\ &= e^{4t} (e^{9t^2-3}) (e^{t^2-3}) \\ &= e^{10t^2+4t-6} \end{aligned}$$

3.9 Funções de uma Variável Aleatória

3.9.1 Variável com Distribuição Discreta

Suponha que uma *v.a.* X possua uma distribuição discreta, cuja função de probabilidade (*f.p.*) seja $p(x)$, e que outra *v.a.* $Y = r(X)$ seja definida como uma certa função de X . Então a *f.p.* (g) de Y pode ser calculada de p de maneira direta:

$$\begin{aligned} g(y) &= P(Y = y) \\ &= P[r(X) = y] \\ &= \sum_{x:r(x)=y} p(x) \end{aligned} \quad (3.22)$$

Exemplo:

Suponha que X seja uma *v.a.* com *f.p.* dada por:

$$P(X = -2) = 0,2; P(X = -1) = 0,1; P(X = 0) = 0,3; P(X = 1) = 0,2; P(X = 2) = 0,1 \text{ e } P(X = 3) = 0,1.$$

Suponha que Y seja outra *v.a.* tal que $Y = 2X^2 - 3$. Qual a *f.p.* de Y ?

Solução:

Como X só pode assumir os valores $-2, 1, 0, 2$ e 3 , então $Y = 2X^2 - 3$ poderá assumir os valores -3 (quando $X = 0$), -1 (quando $X = -1$ ou 1), 5 (quando $X = -2$ ou 2), e 15 (quando $X = 3$). Logo, a f.p. de Y é dada por:

$$P(Y = -3) = P(X = 0) = 0,3; P(Y = -1) = P(X = -1) + P(X = 1) = 0,3; \\ P(Y = 5) = P(X = 2) + P(X = -2) = 0,3 \text{ e } P(Y = 15) = P(X = 3) = 0,1.$$

3.9.2 Variável com Distribuição Contínua

Suponha que X seja uma *v.a.* com distribuição contínua, cuja *f.d.p.* seja $f(x)$. Suponha que Y seja outra *v.a.* definida com uma função de X , isto é, $Y = r(X)$.

Para qualquer número y a **função de distribuição acumulada** $G(y)$ pode ser calculada por

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P[r(X) \leq y] \\ &= \int_{\{x:r(x) \leq y\}} f(x) dx. \end{aligned} \tag{3.23}$$

Se a variável Y também possuir uma distribuição contínua, sua **função densidade de probabilidade** $g(y)$ é também obtida por

$$g(y) = \frac{dG(y)}{dy} \tag{3.24}$$

Exemplo:

Suponha que X possua uma distribuição uniforme no intervalo $(-1,1)$, isto é:

$$f(x) = \begin{cases} \frac{1}{2} & \text{para } -1 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Determine a f.d.p. de $Y = X^2$.

Solução:

Como $Y = X^2 \Rightarrow 0 \leq Y < 1$. Então para qualquer número y tal que $0 \leq y < 1$,

$$\begin{aligned} G(y) &= P(Y \leq y) = P(X^2 \leq y) = \\ &= P(-\sqrt{y} \leq x \leq \sqrt{y}) = \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(x) dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx = \\ &= \sqrt{y} \end{aligned}$$

Logo,

$$\begin{aligned} g(y) &= \frac{d(\sqrt{y})}{dy} \\ &= \frac{1}{2\sqrt{y}} \quad \text{para } 0 \leq y < 1 \end{aligned}$$

Observação:

Em alguns casos, para algumas funções $r(X)$ a *f.d.p.* de Y pode ser calculada diretamente, sem ter que derivar primeiro a sua função de distribuição $G(y)$. Para tal, algumas condições devem ser satisfeitas, como enunciado abaixo:

- Seja uma *v.a.* com *f.d.p.* f e para qual $P(a < X < b) = 1$. Seja $Y = r(X)$, e suponha que $r(x)$ seja contínua e estritamente crescente ou estritamente decrescente para $a < x < b$. Suponha também que $a < X < b$ se e somente se $\alpha < Y < \beta$, e seja $X = s(Y)$ a função inversa para $\alpha < Y < \beta$

Então a *f.d.p.* de Y é dada por

$$g(x) = \begin{cases} f[s(y)] \cdot \left| \frac{ds(y)}{dy} \right| & \text{para } \alpha < y < \beta \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo:

Suponha que X seja uma *v.a.* com distribuição exponencial, isto é:

$$f(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & \text{para } x > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Determinar a *f.d.p.* de $Y = \sqrt{X}$.

Solução:

Como $Y = \sqrt{X}$, então $X = Y^2$. Logo, $\frac{dx}{dy} = 2y$. Portanto,

$$f(y) = \frac{1}{\mu} e^{-y^2/\mu} \cdot 2y.$$

Logo,

$$f(y) = \begin{cases} \frac{2y}{\mu} e^{-y^2/\mu} & \text{para } y > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Esta distribuição é denominada distribuição Rayleigh e é muito utilizada em dados de radar.

3.10 Algumas Distribuições Discretas

3.10.1 Distribuição Binomial

Suponha que certa máquina produza um item defeituoso com probabilidade p ($0 < p < 1$) e produza um item não defeituoso com probabilidade $q = 1 - p$. Suponha que n itens independentes produzidos por essa máquina sejam examinados, e faça X representar o número de itens que são defeituosos. A *v.a.* X terá uma distribuição discreta, e os possíveis valores de X serão $0, 1, 2, 3, \dots, n$.

A probabilidade de se obter uma sequência particular dos n itens contando exatamente x defeituosos e $n - x$ não defeituosos é $p^x \cdot q^{n-x}$. Como existem $\binom{n}{x}$ sequências diferentes desse tipo, segue que a **função de probabilidade de X** será:

$$p(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{para } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo:

- a) Suponha que uma válvula eletrônica, instalada em determinado circuito, tenha probabilidade 0,2 de funcionar mais do que 500 horas. Se ensaiarmos 20 válvulas, qual será a probabilidade de que delas, exatamente k , funcionem mais que 500 horas, $k = 0, 1, \dots, 20$?

Solução:

Seja X a *v.a.* que representa o número de válvulas que funcionam mais

de 500 horas. X possui uma distribuição binomial. Portanto,

$$P(X = k) = \binom{20}{k} (0,2)^k (0,8)^{20-k} \quad \text{para } k = 0, 1, 2, 3, \dots, 20$$

Estas probabilidades podem ser encontradas em tabelas da distribuição binomial. Os resultados são:

$$\begin{array}{lll} P(X = 0) = 0,012 & P(X = 4) = 0,218 & P(X = 8) = 0,022 \\ P(X = 1) = 0,058 & P(X = 5) = 0,175 & P(X = 9) = 0,007 \\ P(X = 2) = 0,137 & P(X = 6) = 0,109 & P(X = 10) = 0,002 \\ P(X = 3) = 0,205 & P(X = 7) = 0,055 & P(X = k) < 0,001 \quad (k \geq 11) \end{array}$$

b) Será extraída uma amostra de 5 eleitores de uma grande população, onde 60% votam no PSDB. Qual a probabilidade de:

- exatamente 3 dos eleitores escolhidos votarem no PSDB?
- pelo menos um dos eleitores votem no PSDB?
- ao menos 3 (uma maioria) votem no PSDB?

Solução: Se X é a v.a. que representa o número de eleitores que votam no PSDB, temos que X segue uma distribuição binomial, cuja probabilidade de "sucesso" (votar no PSDB) em cada tentativa é 0,60. Portanto,

$$P(X = 3) = \binom{5}{3} (0,6)^3 (0,4)^2 = 0,3456$$

A probabilidade que pelo menos um vote no PSDB é dada por

$$1 - P(X = 0) = 1 - \binom{5}{0} (0,4)^5 = 1 - 0,0102 = 0,9898$$

A probabilidade que a maioria votem no PSDB é dada por $P(X = 3) + P(X = 4) + P(X = 5)$, ou seja:

$$\binom{5}{3} (0,6)^3 (0,4)^2 + \binom{5}{4} (0,6)^4 (0,4)^1 + \binom{5}{5} (0,6)^5 = 0,6826$$

3.10.1.1 Média de uma *v.a.* com Distribuição Binomial

Cálculo da média para *distribuição binomial*

$$\begin{aligned}
 E(X) &= \sum_{x=0}^n x \cdot p(x) \\
 &= \sum_{x=1}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
 &= \sum_{x=1}^n x \frac{n(n-1)!}{x(x-1)!(n-x)!} p p^{x-1} q^{n-x} \\
 &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \tag{3.25}
 \end{aligned}$$

Onde $p(x) = \binom{n}{x} p^x q^{n-x}$

Fazendo $y = x - 1$ tem-se

$$E(X) = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y q^{n-1-y} \tag{3.26}$$

Mas, pela fórmula binomial (ver por exemplo Wonnacott, pag. 153)

$$\begin{aligned}
 (p+q)^k &= q^k + \binom{k}{1} p q^{k-1} + \binom{k}{2} p^2 q^{k-2} + \dots + p^k \\
 &= \sum_{i=0}^k \binom{k}{i} p^i q^{k-i} \tag{3.27}
 \end{aligned}$$

Portanto,

$$E(X) = np (p+q)^{n-1} \tag{3.28}$$

Como $p+q=1$, temos

$$E(X) = np \tag{3.29}$$

3.10.1.2 Variância de uma *v.a.* com Distribuição Binomial

Para o cálculo da variância fica mais fácil partir-se de

$$\text{Var}(X) = E(X^2) - \{E(X)^2\} \quad (3.30)$$

O valor de $E(X^2)$ pode ser encontrado através do cálculo da esperança de $E\{X(X-1)\}$ que é definida como:

$$\begin{aligned} E\{X(X-1)\} &= \sum x(x-1)f(x) \\ &= \sum x(x-1) \frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!} p^2 \cdot p^{x-2} q^{n-x} \\ &= n(n-1)p^2 \sum \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} \end{aligned} \quad (3.31)$$

Fazendo $y = x - 2$, a expressão acima se reduz a

$$\begin{aligned} &= n(n-1)p \sum \frac{(n-2)!}{y!(n-x)!} p^y q^{n-y-2} \\ &= n(n-1)p^2 \end{aligned} \quad (3.32)$$

Assim,

$$\begin{aligned} E\{X(X-1)\} &= E(X^2 - X) \\ &= E(X^2) - E(X) \\ &= n(n-1)p^2 \end{aligned} \quad (3.33)$$

Como:

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X) + E(X) - \{E(X)\}^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p) \\ &= npq \end{aligned} \tag{3.34}$$

De fato $f(x) = \binom{n}{x} p^x \cdot q^{n-x}$ é de fato uma função de probabilidade. Para tanto temos que ter

$$\sum \binom{n}{x} p^x \cdot q^{n-x} = 1 \tag{3.35}$$

Como a expansão deste somatório nada mais é que

$$\binom{n}{0} p^0 \cdot q^n + \binom{n}{1} p^1 \cdot q^{n-1} + \binom{n}{2} p^2 \cdot q^{n-2} + \dots + \binom{n}{n} p^n \cdot q^0$$

que por sua vez é igual a $(p+q)^n$, e como $p+q=1$ fica demonstrado que $f(x)$ é de fato uma função de probabilidade.

Exemplo

Em 100 lances de uma moeda *honest*a, a média do número de caras é

$$\mu = Np = 100 \cdot \frac{1}{2} = 50 \tag{3.36}$$

Este é o número esperado de caras em 100 lances da moeda. O desvio padrão é:

$$\sigma = \sqrt{Npq} = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5. \tag{3.37}$$

Observação

A função geratriz de momentos de uma distribuição binomial é

$$M_x(t) = (p \cdot e^t + q)^n \tag{3.38}$$

3.10.2 Distribuição Hipergeométrica

Suponha que uma urna contenha A bolas verdes e B bolas azuis. Suponha que se retire n bolas sem reposição. Seja X v.a. que indica o número de bolas verdes obtidas. Então

$$\max\{0, n - B\} \leq X \leq \min\{n, A\}.$$

A função de probabilidade de X será:

$$P(X = x) = \frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}} \quad (3.39)$$

Dizemos que X possui uma **distribuição hipergeométrica**. Pode-se provar que:

$$E(X) = \frac{n \cdot A}{A + B} \quad (3.40)$$

$$Var(X) = \frac{nAB}{(A + B)^2} \cdot \frac{A + B - n}{A + B - 1} \quad (3.41)$$

Exemplo:

Pequenos motores elétricos são expedidos em lotes de 50 unidades. Antes que uma remessa seja aprovada, um inspetor escolhe 5 desses motores e o inspeciona. Se nenhum dos motores inspecionados for defeituosos, o lote é aprovado. Se um ou mais forem verificados defeituosos, todos os motores da remessa são inspecionados. Suponha que existam, de fato, três motores defeituosos no lote. Qual é a probabilidade de que a inspeção 100 por cento seja necessária?

Solução:

Seja X a v.a. que representa o número de motores defeituosos encontrado. A inspeção de todo o lote será necessária se $X \geq 1$. Logo,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0} \binom{47}{5}}{\binom{50}{5}} = 0,28$$

3.10.3 Distribuição Binomial Negativa

Suponha que em uma sequência infinita de experimentos independentes, o resultado de cada experimento possa ser sucesso ou falha, e que a probabilidade de sucesso p onde $(0 < p < 1)$, e de falha $q = 1 - p$.

Seja X a *v.a.* que denota o número de falhas que ocorrem antes que exatamente r sucessos sejam obtidos. Então:

$$P(X = x) = \binom{r + x - 1}{x} p^r q^x \quad x = 0, 1, 2, \dots \quad (3.42)$$

Dizemos que X possui uma **distribuição binomial negativa**.

$$E(X) = \frac{rq}{p} \quad (3.43)$$

$$Var(X) = \frac{rq}{p^2} \quad (3.44)$$

Exemplo:

Suponha que a probabilidade de sair cara em uma moeda seja de $1/3$. Suponha também que esta moeda seja jogada até que apareçam 5 caras.

- a) Calcule a probabilidade de que a quinta cara apareça na décima segunda jogada.
- b) Qual o número esperado de coroas que aparecerão antes de se obter 5 caras?

Solução:

Seja X a *v.a.* que representa o número de coroas que aparecem antes de que a quinta cara apareça. X possui uma distribuição binomial negativa.

a)

$$P(X = 7) = \binom{11}{7} \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^7$$

b)

$$E(X) = \frac{5 \cdot \frac{2}{3}}{\frac{1}{3}} = 10$$

3.10.4 Distribuição Geométrica

Dizemos que a *v.a.* X possui **distribuição geométrica**, se X possui uma distribuição binomial negativa com $r = 1$

A função de probabilidade de X será:

$$P(X = x) = pq^x \quad (3.45)$$

Pode-se provar que:

$$E(X) = \frac{q}{p}; \quad (3.46)$$

$$Var(X) = \frac{q}{p^2} \quad (3.47)$$

Exemplos:

- a) Em determinada localidade, a probabilidade da ocorrência de uma tormenta em algum dia durante o verão (nos meses de dezembro e janeiro) é igual a 0,1. Admitindo independência de um dia para outro, qual a probabilidade da ocorrência da primeira estação de verão no dia 4 de janeiro?

Solução:

Chamando de X o número de dias (começando em 1o. de dezembro) até a primeira tormenta. Portanto, X possuirá uma distribuição geométrica, e a probabilidade desejada é:

$$P(X = 34) = (0,9)^{33}(0,1) = 0,003$$

- b) O custo de realização de um experimento é de R\$ 1500,00. Se o experimento não tiver sucesso, há um custo adicional de R\$ 400,00 para que sejam executadas as correções necessárias. Suponha que as provas sejam independentes, e que os experimentos sejam executados até que o primeiro sucesso ocorra. Sendo a probabilidade de sucesso em uma tentativa qualquer de 0,1, qual será o custo esperado do procedimento completo?

Solução:

Seja X a v.a. que representa o número de provas necessárias para alcançar o primeiro sucesso. Temos que X possuirá uma distribuição geométrica, com média

$$E(X) = \frac{q}{p} = \frac{0,9}{0,1} = 9.$$

Seja C o custo de realização do experimento:

$$C = 1500X + 400(X - 1) = 1900X - 400$$

Logo,

$$E(C) = 1900E(X) - 400 = (1900)9 - 400 = R\$16700,00$$

3.10.5 Distribuição de Poisson

Seja X uma v.a. com distribuição discreta, e suponha que X assuma valores inteiros não negativos. É dito que X possui uma **distribuição de Poisson** com média λ onde ($\lambda > 0$) se a função de probabilidade de X é dada por:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, 3, \dots \quad (3.48)$$

Observação

O símbolo e representa uma constante que é aproximadamente igual a 2,7183. O seu nome é uma homenagem ao matemático suíço *I. Euler*, e constitui a base do chamado logaritmo natural.

Para determinação da média e da variância de uma v.a. de *Poisson*, é preciso primeiro

determinar a sua função geradora de momento (*f.g.m.*),

$$\begin{aligned}
 M(t) &= E[e^{tX}] \\
 &= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \\
 &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
 &= e^{-\lambda} e^{\lambda e^t} \\
 &= e^{\lambda(e^t-1)}
 \end{aligned}$$

Diferenciando teremos

$$\begin{aligned}
 M'(t) &= \lambda e^t e^{\lambda(e^t-1)} \\
 M''(t) &= (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}
 \end{aligned}$$

Para $t = 0$ temos

$$E(X) = m'(0) = \lambda \tag{3.49}$$

$$\begin{aligned}
 Var(X) &= m''(0) - (E[X])^2 \\
 &= \lambda^2 + \lambda - \lambda^2 \\
 &= \lambda
 \end{aligned} \tag{3.50}$$

Portanto, a média e a variância de uma *v.a.* de *Poisson* são iguais ao parâmetro λ .

A *v.a.* de *Poisson* tem um amplo range de aplicações em uma grande variedade de áreas, porque se emprega como uma aproximação para uma *v.a.* binomial com parâmetros (n, p) quando n é grande e p é pequeno. Supondo que X é uma *v.a.*

binomial com parâmetros (n, p) e seja $\lambda = np$. Então

$$\begin{aligned}
 P(X = x) &= \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\
 &= \frac{n!}{(n-x)!x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \frac{n(n-1)\dots(n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}
 \end{aligned}
 \tag{3.51}$$

Agora, para n grande e p pequeno

$$\begin{aligned}
 \left(1 - \frac{\lambda}{n}\right)^n &\approx e^{-\lambda} \\
 \frac{n(n-1)\dots(n-x+1)}{n^x} &\approx 1 \\
 \left(1 - \frac{\lambda}{n}\right)^x &\approx 1
 \end{aligned}
 \tag{3.52}$$

Assim que, para n grande e p pequeno,

$$P(X = x) \approx e^{-\lambda} \frac{\lambda^x}{x!}
 \tag{3.53}$$

Quer dizer, ao se realizar n experimentos independentes, cada um dos quais tem como resultado êxito com probabilidade p , então quando n é grande e p pequeno, o número de êxitos que se apresentam é aproximadamente uma *v.a.* de *Poisson* com média $\lambda = np$.

Exemplos:

- a) Se a probabilidade de um indivíduo sofrer uma reação nociva, resultante de ter tomado um certo soro é 0,001, determinar a probabilidade de que, entre 2000 indivíduos:
 - 1) exatamente três sofrerem a reação;
 - 2) mais do que dois sofrerem a reação.

Solução:

Seja X a v.a. que representa o número de pessoas que sofrem a reação nociva após injerir o soro. Então,

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, 3, \dots,$$

onde $\lambda = 2000 \cdot 0,001 = 2$. Logo,

1)

$$P(X = 3) = e^{-2} \frac{2^3}{3!} = 0,18$$

2)

$$\begin{aligned} P(X \geq 3) &= 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} \\ &= 1 - \left\{ e^{-2} \frac{2^0}{0!} + e^{-2} \frac{2^1}{1!} + e^{-2} \frac{2^2}{2!} \right\} \\ &= 1 - \left\{ \frac{1}{e^2} + \frac{2}{e^2} + \frac{2}{e^2} \right\} = 1 - \frac{5}{e^2} \\ &= 0,323 \end{aligned}$$

OBS: Se usasse a distribuição binomial os cálculos se tornariam maiores. Por exemplo a primeira questão ficaria:

$$P(X = 3) = \binom{2000}{3} (0,001)^3 (0,999)^{1997}$$

- b) Suponha que em um certo final de semana, o número de acidentes num cruzamento possua uma distribuição de Poisson com média 0,7. Qual a probabilidade de que exista pelo menos três acidentes no cruzamento durante o final de semana?

Solução:

$$P(X = x) = \frac{e^{-0,7} 0,7^x}{x!} \quad x = 0, 1, 2, 3, \dots,$$

Logo,

$$\begin{aligned}P(X \geq 3) &= 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} \\&= 1 - \left\{e^{-0,7} + 0,7e^{0,7} + \frac{0,7^2 e^{-0,7}}{2}\right\} \\&= 1 - e^{-0,7}(1 + 0,7 + 0,245) \\&= 0,341\end{aligned}$$

c) Dez por cento das ferramentas produzidas por um certo processo de fabricação revelaram-se defeituosas. Determinar a probabilidade de, em uma amostra de 10 ferramentas escolhidas ao acaso, exatamente duas sejam defeituosas, mediante o emprego de:

- 1) distribuição binomial
- 2) distribuição de Poisson

Solução:

1)

$$P(X = 2) = \binom{10}{2} (0,1)^2 (0,9)^8 = 0,1937$$

2)

$$\lambda = Np = 10(0,1) = 1$$

$$P(X = 2) = \frac{1^2 e^{-1}}{2!} = \frac{1}{2e^{-1}} = 0,1839$$

3.11 Algumas Distribuições Contínuas Importantes

3.11.1 Distribuição Normal

Definição

Dizemos que uma *v.a.* X possui uma **distribuição Normal** (ou **Gaussiana**) com média μ e variância σ^2 ($-\infty < \mu < \infty$ e $\sigma > 0$) se X possuir uma distribuição contínua com *f.d.p.* dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty \quad (3.54)$$

Média

$$E(X) = \mu \quad (3.55)$$

Variância

$$Var(X) = \sigma^2 \quad (3.56)$$

Função Geratriz de Momentos

$$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \quad -\infty < x < \infty \quad (3.57)$$

Teorema 1

Se X possui uma distribuição normal com média μ e variância σ^2 , e se $Y = aX + b$, onde a e b são constantes, com $a \neq 0$, então Y terá uma distribuição normal com média $a\mu + b$ e variância $a^2\sigma^2$

Prova

A função geratriz de momentos de Y será

$$\begin{aligned} M_y(t) &= e^{bt} M_x(at) \\ &= e^{bt} \cdot e^{a\mu t + \sigma^2 a^2 \frac{t^2}{2}} \\ &= e^{(a\mu + b)t + a^2 \sigma^2 \frac{t^2}{2}}, \end{aligned} \quad (3.58)$$

que é a função geratriz de momentos de uma distribuição com média $a\mu + b$ e variância $a^2\sigma^2$.

3.11.2 Distribuição Normal Padrão

A distribuição normal com média zero ($\mu = 0$) e variância um ($\sigma^2 = 1$) é denominada **distribuição normal padrão** $\mathcal{N}(0, 1)$. A *f.d.p.* de uma distribuição normal padrão

é em geral representada por $\phi(x)$ e dada por

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad -\infty < x < \infty \quad (3.59)$$

Segue do Teorema 1 que se uma variável X tem uma distribuição normal com média μ e variância σ^2 , então a variável

$$Z = \frac{X - \mu}{\sigma} \quad (3.60)$$

terá uma distribuição normal padrão.

As probabilidades para uma distribuição normal com qualquer média e variância podem ser determinadas através de Tabelas de uma distribuição normal padrão.

Exemplos:

a) Supondo que a v.a. Z possua uma distribuição normal padrão, então:

$$i) P(Z > 1,0) = 1 - P(Z \leq 1) = 1 - 0,8413 = 0,1587$$

$$ii) P(Z < -1,0) = P(Z > 1,0) = 0,1587$$

$$\begin{aligned} iii) P(1,0 < Z \leq 1,5) &= P(Z \leq 1,5) - P(Z \leq 1,0) \\ &= 0,9332 - 0,8413 = 0,0919 \end{aligned}$$

$$\begin{aligned} iv) P(-1 \leq Z \leq 2) &= P(Z \leq 2) - P(Z < -1) \\ &= P(Z \leq 2) - P(Z > 1) \\ &= P(Z \leq 2) - (1 - P(Z \leq 1)) \\ &= P(Z \leq 2) + P(Z \leq 1) - 1 \\ &= 0,9773 + 0,8413 - 1 = 0,8186 \end{aligned}$$

$$\begin{aligned} v) P(|Z| \leq 2) &= P(-2 \leq Z \leq 2) \\ &= P(Z \leq 2) - P(Z < -2) \\ &= P(Z \leq 2) - P(Z > 2) \\ &= P(Z \leq 2) - (1 - P(Z \leq 2)) \\ &= 2 \cdot P(Z \leq 2) - 1 \\ &= 2 \cdot 0,9773 - 1 = 0,9546 \end{aligned}$$

- b) Suponha que X possua uma distribuição normal com média 5 e desvio padrão 2. Então:

$$\begin{aligned}
 P(1 \leq X \leq 8) &= P\left(\frac{1-5}{2} \leq \frac{X-\mu}{\sigma} \leq \frac{8-5}{2}\right) \\
 &= P(-2 \leq Z \leq 1,5) \\
 &= P(Z \leq 1,5) - P(Z < -2) \\
 &= P(Z \leq 1,5) - P(Z > 2) \\
 &= P(Z \leq 1,5) - (1 - P(Z \leq 2)) \\
 &= P(Z \leq 1,5) + P(Z \leq 2) - 1 \\
 &= 0,9332 + 0,9772 - 1 = 0,9104
 \end{aligned}$$

- c) Supor uma população em que o peso dos indivíduos seja distribuído normalmente com média 68 kg e desvio padrão 4 kg. Determinar a proporção de indivíduos
- i) abaixo de 66 kg
 - ii) acima de 72 kg
 - iii) entre 66 e 72 kg

$$\begin{aligned}
 i) P(X < 66) &= P\left(\frac{X-\mu}{\sigma} < \frac{66-68}{4}\right) = P(Z < -0,5) \\
 &= 1 - P(Z \leq 0,5) = 1 - 0,6915 = 0,3085
 \end{aligned}$$

$$ii) P(X > 72) = P\left(\frac{X-\mu}{\sigma} > \frac{72-68}{4}\right) = P(Z > 1) = 0,1587$$

$$\begin{aligned}
 iii) P(66 \leq X \leq 72) &= P\left(\frac{66-68}{4} \leq \frac{X-\mu}{\sigma} \leq \frac{72-68}{4}\right) \\
 &= P(-0,5 \leq Z \leq 1) \\
 &= P(Z \leq 1) - P(Z < -0,5) \\
 &= 0,8413 - 0,3085 = 0,5328
 \end{aligned}$$

Teorema 2

Se as *v.a.* X_1, X_2, \dots, X_k são independentes e se X_i é normalmente distribuída com média μ_i e variância σ_i^2 com $(i = 1, 2, 3, \dots, k)$, então a soma $X_1 + X_2 + \dots + X_k$ terá uma distribuição normal com média $\mu_1 + \mu_2 + \dots + \mu_k$ e variância $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2$.

Prova:

Seja $M_i(t)$ a *f.g.m.* de X_i para $(i = 1, 2, \dots, k)$ e seja $M(t)$ a *f.g.m.* de $X_1 + X_2 +$

... + X_k . Como as variáveis X_1, X_2, \dots, X_k são independentes, então

$$\begin{aligned} M(t) &= \prod_{i=1}^k M_i(t) \\ &= \prod_{i=1}^k e^{\left(\mu_i t + \frac{\sigma_i^2 t}{2}\right)} \\ &= e^{(\sum_{i=1}^k \mu_i)t + \frac{1}{2}(\sum_{i=1}^k \sigma_i^2)t} \quad \text{para } -\infty < x < \infty \end{aligned}$$

Esta *f.g.m.* pode ser identificada como a *f.g.m.* de uma distribuição normal com média $\left(\sum_{i=1}^k \mu_i\right)$ e variância $\left(\sum_{i=1}^k \sigma_i^2\right)$. Logo, é essa a distribuição de $X_1 + X_2 + \dots + X_k$

Exemplo:

Suponha que a altura, em inches, de mulheres de uma certa população segue uma distribuição normal com média 65 e desvio padrão 1, e que as alturas dos homens segue uma distribuição normal com média 68 e desvio padrão 2. Supor também que uma mulher seja selecionada aleatoriamente e independentemente um homem seja selecionado aleatoriamente. Determinar a probabilidade que a mulher seja mais alta que o homem.

Solução:

Sejam M e H as v.a. que representem as alturas da mulher e do homem, respectivamente. Então

$$M \sim N(65, 1^2)$$

$$H \sim N(68, 2^2)$$

Logo, $(M - H) \sim N(65 - 68, 1^2 + 2^2)$, isto é, $(M - H) \sim N(-3, 5)$.

Portanto $Z = \frac{M-H+3}{\sqrt{5}} \sim N(0, 1)$. Logo,

$$\begin{aligned} P(M > H) &= P(M - H > 0) = P\left(\frac{M-H+3}{\sqrt{5}} > \frac{3}{\sqrt{5}}\right) \\ &= P\left(Z > \frac{3}{\sqrt{5}}\right) = P(Z > 1,342) = 1 - P(Z < 1,342) \\ &= 1 - 0,9099 = 0,09 \end{aligned}$$

Corolário 1

Se as *v.a.* X_1, X_2, \dots, X_k são independentes, se X_i possui uma distribuição normal com média μ_i e variância σ_i^2 onde $(i = 1, 2, \dots, k)$, e se a_1, a_2, \dots, a_k e b são constantes para as quais pelo menos um dos valores de a_1, a_2, \dots, a_k é diferente de zero, então a variável $a_1X_1 + a_2X_2 + \dots + a_kX_k + b$ tem uma distribuição normal com média $a_1\mu_1 + a_2\mu_2 + \dots + a_k\mu_k + b$ e variância $a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_k^2\sigma_k^2$

Corolário 2

Suponha que as *v.a.* X_1, X_2, \dots, X_k formem uma amostra aleatória de uma distribuição com média μ e variância σ^2 , e seja \bar{X} a média amostral. Então \bar{X} terá uma distribuição normal com média μ e variância $\frac{\sigma^2}{n}$

Exemplo:

Suponha que em um certo exame de Matemática Avançada, os estudantes da Universidade A obtêm notas normalmente distribuídas com média 625 e variância de 100, e que os estudantes da Universidade B obtêm notas normalmente distribuídas com média 600 e variância de 150. Se dois estudantes da Universidade A e cinco estudantes da Universidade B fazem este exame, qual é a probabilidade que a nota média dos estudantes da Universidade A seja maior que a nota média dos estudantes da Universidade B?

Solução:

Temos que:

$$\begin{aligned}\mu_A &= 625 & \mu_B &= 600 \\ \sigma_A^2 &= 100 & \sigma_B^2 &= 150\end{aligned}$$

Portanto a média amostral de dois estudantes da Universidade A (\bar{x}_A) é normalmente distribuída com média 625 e variância $\frac{100}{2} = 50$. Analogamente, a média amostral de três estudantes da Universidade B (\bar{x}_B) é normalmente distribuída com média 600 e variância $\frac{150}{3} = 30$, isto é,

$$\bar{x}_A \sim N(625, 50)$$

$$\bar{x}_B \sim N(600, 30)$$

Logo, $(\bar{x}_A - \bar{x}_B) \sim N(625 - 600, 50 + 30)$, isto é, $(\bar{x}_A - \bar{x}_B) \sim N(25, 80)$.

Portanto,

$$\begin{aligned}P(\bar{x}_A > \bar{x}_B) &= P(\bar{x}_A - \bar{x}_B > 0) = P\left(\frac{\bar{x}_A - \bar{x}_B - 25}{\sqrt{80}} > \frac{0 - 25}{\sqrt{80}}\right) \\ &= P\left(Z > \frac{-25}{\sqrt{80}}\right) = P(Z > -2,975) = P(Z < 2,975) = 0,9986\end{aligned}$$

3.11.3 Teorema do Limite Central

Se as *v.a.* X_1, X_2, \dots, X_n formam uma amostra aleatória de tamanho n de uma certa distribuição com média μ e variância σ^2 onde $(0 < \sigma^2 < \infty)$, então para qualquer número fixado x

$$n \xrightarrow{\text{lim}} \infty P \left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x \right] = \phi(x), \quad (3.61)$$

onde $\phi(x)$ é a função distribuição da normal padrão.

A interpretação desse teorema é a seguinte: Se uma grande amostra aleatória é tomada de qualquer distribuição com média μ e variância σ^2 , não importando se a distribuição é discreta ou contínua, então a distribuição da *v.a.* $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ será aproximadamente a distribuição normal padrão. Portanto a distribuição de \bar{X} será aproximadamente uma distribuição normal com média μ e variância $\frac{\sigma^2}{n}$ ou, equivalentemente, a distribuição da soma $\sum_{i=1}^n X_i$ será aproximadamente uma normal com média $n\mu$ e variância $n\sigma^2$.

Exemplo:

Suponha que uma moeda honesta seja jogada 900 vezes. Qual a probabilidade de se obter mais de 495 caras?

Solução:

Para $i = 1, 2, \dots, 900$, seja $X_i = 1$ se cara é obtida na i -ésima jogada e $X_i = 0$ caso contrário (distribuição de Bernoulli).

Então $E(X_i) = \frac{1}{2}$ e $Var(X_i) = \frac{1}{4}$. Portanto os valores $X_1, X_2, X_3, \dots, X_{900}$ formam uma amostra aleatória de tamanho $n = 900$ de uma distribuição com média $\frac{1}{2}$ e variância $\frac{1}{4}$. Segue, pelo Teorema do Limite Central, que a distribuição do número de caras $H = \sum_{i=1}^{900} X_i$ será aproximadamente uma normal com média e variância dadas respectivamente por

$$E(H) = 900 \cdot \frac{1}{2} = 450$$

$$Var(H) = 900 \cdot \frac{1}{4} = 225 \text{ (e, consequentemente desvio padrão igual a 15).}$$

Portanto a variável $Z = \frac{H-450}{15}$ terá aproximadamente uma distribuição normal padrão. Logo,

$$\begin{aligned} P(H > 495) &= P\left(\frac{H-450}{15} > \frac{495-450}{15}\right) \\ &= P(Z > 3) = 0,0013 \end{aligned}$$

Distribuição Amostral das Proporções:

Admita-se que uma população é infinita e que a probabilidade de ocorrência de um evento (denominado sucesso) é p , enquanto a sua não ocorrência é $q = 1 - p$. Suponha que este experimento seja realizado n vezes e que para cada uma das tentativas i , seja definido a v.a. X_i , tal que $X_i = 1$ se ocorrer sucesso na i -ésima tentativa e $X_i = 0$ se ocorrer insucesso ($i = 1, 2, \dots, n$) (X_i é dito possuir uma distribuição de Bernoulli, com média $E(X_i) = p$ e variância $Var(X_i) = pq$). Seja P a proporção de sucessos, isto é:

$$P = \frac{\sum_{i=1}^n X_i}{n}.$$

Para grandes valores de n segue, pelo Teorema Central do Limite, que P será aproximadamente normalmente distribuído com média e variância dadas por:

$$E(P) = p$$

$$Var(P) = \frac{pq}{n}.$$

Exemplo:

Suponha que a proporção de itens defeituosos em um grande lote de peças seja 0,1. Qual é o menor número de itens que deve ser retirado do lote para que a probabilidade seja de pelo menos 0,99 que a proporção de itens defeituosos na amostra seja menor que 0,13.

Solução: Seja P a proporção de defeituosos na amostra. Para n grande tem-se que P será aproximadamente normalmente distribuído com média $\mu = 0,1$ e variância $\sigma^2 = \frac{(0,1)(0,9)}{n} = \frac{0,09}{n}$. Quer-se determinar n para que

$$P(P < 0,13) \geq 0,99$$

Portanto,

$$P\left(\frac{P - \mu}{\sigma} < \frac{0,13 - 0,1}{\sqrt{\frac{0,09}{n}}}\right) \geq 0,99$$

Ou seja,

$$P\left(Z < \frac{0,03}{\frac{0,3}{\sqrt{n}}}\right) = P(Z < 0,1\sqrt{n}) \geq 0,99$$

Logo, utilizando a tabela da distribuição normal padrão, temos que

$$0,1\sqrt{n} \geq 2,33 \quad \Rightarrow \quad n \geq 543$$

3.11.4 Distribuição Uniforme

Sejam α e β dois números reais tais que ($\alpha < \beta$), e considere um experimento no qual um ponto X é selecionado do intervalo $S = \{x : \alpha \leq x \leq \beta\}$ de tal maneira que a probabilidade de que X pertença a qualquer subintervalo de S é proporcional ao comprimento desse intervalo. A distribuição da *v.a.* X é denominada **distribuição uniforme no intervalo** (α, β) e é dada por

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{para } \alpha \leq x \leq \beta \\ 0 & \text{caso contrário} \end{cases}$$

Esta f.d.p. está representada pela Figura 3.7. Observe que a função anterior satisfaz os requisitos para ser uma *f.d.p.*, já que

$$\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} dx = 1 \tag{3.62}$$

A probabilidade de que X esteja em qualquer subintervalo de $[\alpha, \beta]$ é igual a comprimento do subintervalo dividido pelo comprimento do intervalo $[\alpha, \beta]$. Isso

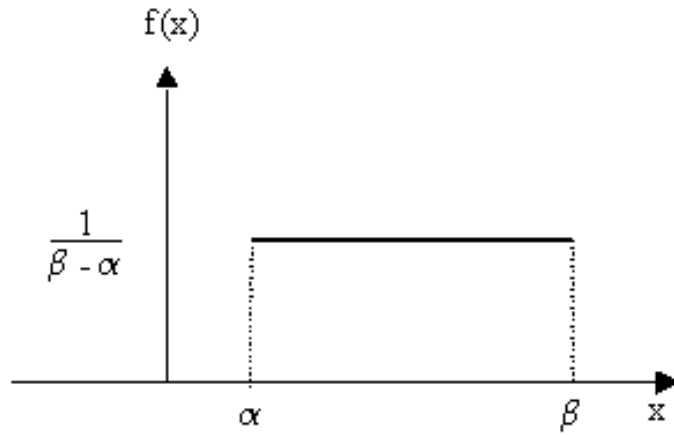


Fig. 3.7 – Gráfico de $f(x)$ de uma função uniforme

ocorre devido a que, quando $[a, b]$ é um subintervalo de $[\alpha, \beta]$ (ver Figura 3.8).

$$\begin{aligned}
 P\{a < X < b\} &= \frac{1}{\beta - \alpha} \int_a^b dx \\
 &= \frac{b - a}{\beta - \alpha}
 \end{aligned}
 \tag{3.63}$$

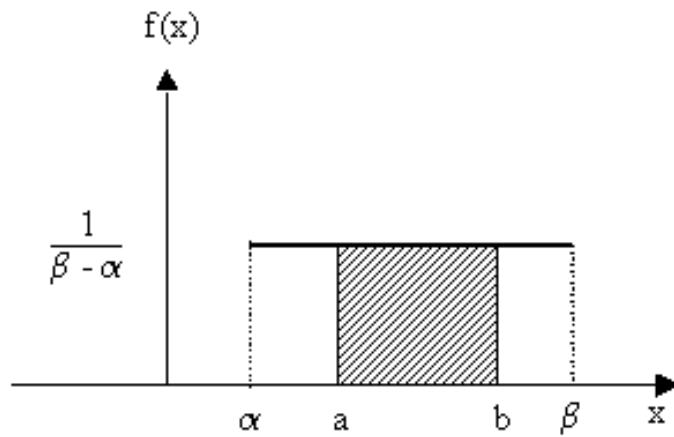


Fig. 3.8 – Probabilidade de uma variável aleatória uniforme

Exemplo:

Um ponto é escolhido ao acaso no segmento de reta $[0,2]$. Qual será a a probabilidade de que o ponto escolhido esteja entre 1 e 1,5?

Seja X a v.a. que representa a coordenada do ponto escolhido. Tem-se que a f.d.p. de X é dada por:

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

Portanto,

$$P(1 \leq X \leq 1,5) = \frac{0,5}{2} = \frac{1}{4}$$

Média

A média de uma v.a. uniforme $[\alpha, \beta]$ é

$$\begin{aligned} E(X) &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)} \end{aligned} \tag{3.64}$$

ou

$$E(X) = \frac{\alpha + \beta}{2} \tag{3.65}$$

Ou, em outras palavras, o valor esperado de uma v.a. uniforme $[\alpha, \beta]$ é igual ao ponto médio do intervalo $[\alpha, \beta]$.

Variância

O cálculo da variância é dado por:

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

mas

$$\begin{aligned} E[X^2] &= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{(\beta^2 + \alpha\beta + \alpha^2)}{3} \end{aligned} \tag{3.66}$$

Assim

$$\begin{aligned} Var[X] &= \frac{(\beta^2 + \alpha\beta + \alpha^2)}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 \\ &= \frac{(\beta^2 + \beta^2 - 2\alpha\beta)}{12} \\ &= \frac{(\beta - \alpha)^2}{12} \end{aligned} \tag{3.67}$$

3.11.5 Distribuição Exponencial

É dito que uma *v.a.* X possui uma **distribuição exponencial** com média μ onde ($\mu > 0$) se X possui uma distribuição contínua para a qual a *f.d.p.* $f(x)$ é dada por

$$f(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases}$$

A função distribuição acumulada da exponencial é dada por:

$$F(x) = P(X \leq x) = \int_0^x \frac{1}{\mu} e^{-\frac{x}{\mu}} = 1 - e^{-\frac{x}{\mu}} \quad \text{para } x \geq 0.$$

Portanto, $P(X > x) = e^{-\frac{x}{\mu}}$.

A média e variância de uma *v.a.* X com distribuição exponencial são dadas por:

$$E(X) = \mu \tag{3.68}$$

$$Var(X) = \mu^2 \tag{3.69}$$

Observação

Se X possui uma distribuição exponencial então a v.a. $Y = \sqrt{X}$ possui uma **distribuição Rayleigh**, cuja *f.d.p.* é dada por:

$$f(y) = \frac{2y}{\mu} e^{-\frac{y^2}{\mu}} \quad (3.70)$$

A qual tem média e variância representadas por:

$$E(X) = \frac{1}{2}\sqrt{\mu\pi} \quad (3.71)$$

$$Var(X) = \frac{4 - \pi}{4}\mu \quad (3.72)$$

Exemplo:

Suponha que um fusível tenha uma duração de vida X , a qual pode ser considerada uma v.a. contínua com uma distribuição exponencial. Existem dois processos pelos quais o fusível pode ser fabricado. O processo I apresenta uma duração de vida esperada de 100 horas, enquanto o processo II apresenta uma duração de vida esperada de 150 horas. Suponha que o processo II seja duas vezes mais custoso (por fusível) que o processo I, que custa C dólares por fusível. Admita-se, além disso, que se um fusível durar menos que 200 horas, uma multa de K dólares seja lançada sobre o fabricante. Qual processo deve ser empregado?

Solução:

O custo esperado por fusível para o processo I é dado por:

$$C_I = \begin{cases} C & \text{se } X > 200 \\ C + K & \text{se } X \leq 200 \end{cases}$$

Logo,

$$\begin{aligned} E(C_I) &= CP(X > 200) + (C + K)P(X \leq 200) \\ &= Ce^{-\frac{200}{100}} + (C + K)(1 - e^{-\frac{200}{100}}) \\ &= Ce^{-2} + (C + K)(1 - e^{-2}) \\ &= C + K(1 - e^{-2}) \end{aligned}$$

Analogamente, o custo esperado por fusível para o processo II é dado por:

$$E(C_{II}) = 2C + K(1 - e^{-\frac{4}{3}})$$

Portanto,

$$E(C_{II}) - E(C_I) = C + K(e^{-2} - e^{-\frac{4}{3}}) = C - 0,13K$$

Conseqüentemente, preferimos o processo I, visto que $C > 0,13K$.

3.11.6 Distribuição Gama

É dito que uma *v.a.* X possui uma distribuição *Gama* com parâmetro α e β se X possui uma distribuição contínua cuja *f.d.p.* é dada por:

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases}$$

A média e variância de uma *v.a.* X que possui uma distribuição *Gama* são dadas por:

$$E(X) = \frac{\alpha}{\beta} \quad (3.73)$$

$$Var(X) = \frac{\alpha}{\beta^2} \quad (3.74)$$

Observação 1

A distribuição exponencial é um caso particular da distribuição *Gama*, quando $\alpha = 1$.

Observação 2

Se as *v.a.* X_1, X_2, \dots, X_k são independentes e se X_i possui uma distribuição *Gama* com parâmetros α_i e β ($i = 1, 2, \dots, k$), então a soma de $X_1 + X_2 + \dots + X_k$ possui uma distribuição *Gama* com parâmetros $\alpha_1 + \alpha_2 + \dots + \alpha_k$ e β .

3.11.7 Distribuição Normal Bivariada

Diz-se que duas *v.a.* X_1 e X_2 possuem uma distribuição **Normal Bivariada** se a *f.d.p.* conjunta de X_1 e X_2 for dada por:

$$f(x_1, x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] \right\}} \quad (3.75)$$

onde:

$$E(X_1) = \mu_1 \quad Var(X_1) = \sigma_1^2$$

$$E(X_2) = \mu_2 \quad Var(X_2) = \sigma_2^2$$

e

$$\begin{aligned} \rho &= \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2} \\ &= Cor(X_1, X_2) \end{aligned}$$

Observação

X_1 e X_2 são independentes se e somente se forem não correlacionadas.

3.11.8 Distribuição Normal Multivariada

Diz-se que:

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

possue uma **distribuição normal multivariada** se a *f.d.p.* conjunta de X_1, X_2, \dots, X_n for dada por:

$$f(\underline{X}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{\left[-\frac{1}{2}(\underline{X}-\underline{\mu})^T \Sigma^{-1}(\underline{X}-\underline{\mu})\right]} \quad (3.76)$$

onde:

$$\begin{aligned} \underline{\mu} &\text{ é o vetor médias;} \\ \Sigma &\text{ é a matriz de covariância;} \\ \Sigma^{-1} &\text{ é a inversa de } \Sigma; \text{ e} \\ |\Sigma| &\text{ é o determinante de } \Sigma. \end{aligned}$$

e

$$\begin{aligned}\underline{\mu} &= E(\underline{X}) \\ \underline{\Sigma} &= E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T]\end{aligned}$$

ou seja:

$$\begin{aligned}\mu_i &= E(X_i) \\ \sigma_{ij} &= E[(X_i - \mu_i)(X_j - \mu_j)]\end{aligned}$$

CAPÍTULO 4

INFERÊNCIA ESTATÍSTICA

4.1 Introdução

Um problema de inferência estatística é um problema no qual os dados de uma certa população com uma distribuição de probabilidades desconhecida precisam ser analisados, e algum tipo de inferência sobre essa distribuição desconhecida precisa ser feito. Essa inferência é feita através dos dados de uma amostra.

4.1.1 Parâmetros de uma distribuição

Num problema de inferência estatística, qualquer característica da distribuição dos dados, tais como a média μ ou variância σ^2 , é denominada **parâmetro da distribuição**.

4.1.2 Estatística

Denomina-se **estatística** ao valor calculado inteiramente a partir da amostra. Exemplos de estatísticas são \bar{X} , σ^2 , mediana, moda, desvio médio, etc.

4.1.3 Estimação Pontual e por Intervalo

Suponha que alguma característica dos elementos de uma população seja representada pela v.a. X , cuja f.d.p. é $f(x; \theta)$, onde a forma dessa densidade seja conhecida, exceto pelo parâmetro θ desconhecido, que se deseja estimar.

Suponha que os valores x_1, x_2, \dots, x_k de uma amostra aleatória X_1, X_2, \dots, X_k possam ser observados. Com base nesses valores observados deseja-se estimar o valor desconhecido de θ . A estimação pode ser obtida de duas maneiras: **estimação pontual** e **estimação por intervalo**. Na **estimação pontual** determina-se uma estatística cujo valor representa, ou estima o parâmetro θ . Na **estimação por intervalo**, determina-se um intervalo para o qual a probabilidade que ele contenha o valor θ possa ser determinado.

4.2 Estimação Pontual

Apresentamos aqui dois métodos para se encontrar estimadores pontuais: **métodos dos momentos** e **método da máxima verossimilhança**

4.2.1 Método dos Momentos

Seja $\mu'_r = E(X^r)$ o r -ésimo momento de X . Em geral, μ'_r será uma função conhecida dos parâmetros desconhecidos $\theta_1, \theta_2, \dots, \theta_k$.

Seja m'_r o r -ésimo momento **amostral**, i.e.,

$$m'_r = \frac{1}{n} \sum_{i=1}^k x_i^r \quad (4.1)$$

O método dos momentos consiste em se resolver as k equações

$$m'_j = \mu'_j \quad j = 1, 2, \dots, k.$$

Obs: O primeiro momento amostral m'_1 é representado por \bar{x} .

Exemplos:

- a) Determinar o estimador pontual pelo método dos momentos do parâmetro μ da distribuição Rayleigh.

Solução:

Sabemos, pelo capítulo 3, que se uma v.a. X possui distribuição Rayleigh com parâmetro μ , sua f.d.p. é dada por:

$$f(x) = \frac{2x}{\mu} e^{-\frac{x^2}{\mu}} \quad \text{para } x > 0$$

A média de X é dada por:

$$E(X) = \frac{1}{2} \sqrt{\mu\pi}$$

Portanto, o estimador dos momentos de μ será obtido igualando-se o primeiro momento populacional ao primeiro momento amostral:

$$\frac{1}{2} \sqrt{\hat{\mu}\pi} = \bar{x}$$

Logo,

$$\hat{\mu} = \frac{4\bar{x}^2}{\pi}$$

- b) Determinar os estimadores dos momentos para os parâmetros μ e σ^2 da distribuição Normal.

Solução

Sabemos que se uma v.a. X possui distribuição Normal com parâmetros μ e σ^2 , então:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Como $Var(X) = E(X^2) - [E(X)]^2$, seu segundo momento será

$$E(X^2) = \sigma^2 + \mu^2$$

Igualando-se os dois primeiros momentos populacionais aos dois primeiros momentos amostrais tem-se:

$$\hat{\mu} = \bar{x}$$

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

- c) Determinar os estimadores dos parâmetros α e β da distribuição Gama pelo método dos momentos. (Fazer na série de exercícios)

4.2.2 Método da Máxima Verossimilhança

Definição 1: Função de Verossimilhança

A função de verossimilhança de n v.a. X_1, X_2, \dots, X_n é definida como sendo a densidade conjunta das n v.a., $f(x_1, x_2, \dots, x_n; \theta)$, que é considerada como sendo uma função de θ . Em particular, se X_1, X_2, \dots, X_n é uma amostra aleatória de uma densidade $f(x; \theta)$, então a função de verossimilhança é

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta)f(x_2; \theta)\dots f(x_n; \theta).$$

Definição 2: Estimador de Máxima Verossimilhança

Seja $L(\theta) = L(\theta; x_1, x_2, \dots, x_n)$ a função de verossimilhança das v.a. X_1, X_2, \dots, X_n . Se $\hat{\theta}$ é o valor de θ que maximiza $L(\theta)$, então $\hat{\theta}(X_1, X_2, \dots, X_n)$ é o **estimador de máxima verossimilhança de θ** , e $\hat{\theta}(x_1, x_2, \dots, x_n)$ é a estimativa de máxima verossimilhança de θ para a amostra x_1, x_2, \dots, x_n .

Os casos mais importantes que consideraremos são aqueles que X_1, X_2, \dots, X_n formam uma amostra aleatória de uma densidade $f(x; \theta)$, de tal modo que:

$$L(\theta) = f(x_1; \theta)f(x_2; \theta)\dots f(x_n; \theta).$$

Então, o estimador de máxima verossimilhança é a solução da equação

$$\frac{dL(\theta)}{d\theta} = 0 \tag{4.2}$$

Além disso, $L(\theta)$ e $\log L(\theta)$ possuem seus máximos para o mesmo valor de θ , e muitas vezes é mais fácil encontrar o máximo do logaritmo da função de verossimilhança.

Se a função de verossimilhança contém k parâmetros $\theta_1, \theta_2, \dots, \theta_k$, os estimadores de máxima verossimilhança serão a solução das k -equações

$$\begin{aligned} \frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_1} &= 0 \\ \frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_k} &= 0 \end{aligned}$$

Exemplo:

Determinar os estimadores dos parâmetros μ e σ^2 da distribuição Normal pelo método de Máxima Verossimilhança.

Solução:

$$\begin{aligned} L(\mu, \sigma^2) &= f(x_1; \mu, \sigma^2)f(x_2; \mu, \sigma^2)\dots f(x_n; \mu, \sigma^2). \\ &= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \dots \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\sum_{i=1}^n \left\{ -\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2 \right\}} \end{aligned}$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \sum_{i=1}^n \left\{ \frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}$$

$$\frac{\partial L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{\sum_{i=1}^n x_i - n\mu}{\sigma^2}$$

Igualando-se a zero a derivada acima, tem-se que o estimador de máxima verossimilhança de μ é,

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

Analogamente,

$$\frac{\partial L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} = \frac{-n\sigma^2 + \sum_{i=1}^n (x_i - n\mu)^2}{2\sigma^4} = 0$$

Igualando-se a zero a derivada acima, e substituindo-se o estimador de μ encontrado acima, tem-se que o estimador de máxima verossimilhança de σ^2 é

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

4.3 Estimadores Não Tendenciosos

Um estimador $\hat{\theta}$ é um estimador **não tendencioso** de um parâmetro θ , se $E(\hat{\theta}) = \theta$, para todo possível valor de θ . Em outras palavras, um estimador do parâmetro θ é não tendencioso se sua esperança (ou média) é igual ao verdadeiro valor (desconhecido) de θ

Exemplo:

Vimos que os estimadores dos momentos e de máxima verossimilhança de μ e σ^2 da distribuição normal são dados por:

$$\hat{\mu} = \bar{x}$$

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

Provar que \bar{x} é um estimador não tendencioso de μ , mas $\hat{\sigma}^2$ é um estimador tendencioso de σ^2 . Encontrar um estimador não tendencioso de σ^2 . (Fazer como série de exercícios).

4.4 A Distribuição χ^2

Uma distribuição *Gama* com $\alpha = \frac{n}{2}$ e $\beta = \frac{1}{2}$ é denominada **distribuição χ^2 com n graus de liberdade**, e sua *f.d.p.* é dada por

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases}$$

A média e variância de uma v.a. X que possui uma distribuição χ^2 são dadas por:

$$E(X) = n \quad (4.3)$$

$$Var(X) = 2n \quad (4.4)$$

Teorema 1:

Se as variáveis X_1, X_2, \dots, X_k são independentes e se X_i possui uma distribuição χ^2 com n_i graus de liberdade ($i = 1, 2, \dots, k$), então a soma $X_1 + X_2 + \dots + X_k$ possui uma distribuição χ^2 com n_1, n_2, \dots, n_k graus de liberdade.

Teorema 2:

Se as v.a. X_1, X_2, \dots, X_k são independentes e identicamente distribuídas, cada uma delas com distribuição normal padrão então a soma $X_1^2 + X_2^2 + \dots + X_k^2$ possui uma distribuição χ^2 com k graus de liberdade.

Exemplo:

Suponha que X_1, X_2, \dots, X_n formem uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 . Encontre a distribuição de

$$a) \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

$$b) \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}.$$

Solução

a) Sabemos que

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

Portanto,

$$\frac{(\bar{X} - \mu)^2}{\sigma^2/n} \sim \chi_1^2$$

b) Temos que

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1)$$

Portanto,

$$\frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_1^2$$

Somando estas n v.a., tem-se que

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2.$$

Observação:

1. Se μ for substituído por \bar{X} , então temos que

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

Ou seja,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \text{onde} \quad s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

2. Pela esperança e variância de uma v.a. com distribuição χ^2 , temos que:

$$E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1 \quad \Rightarrow \quad E(s^2) = \sigma^2$$

Ou seja, a variância amostral, com divisor $(n-1)$, é um estimador não tendencioso

da variância populacional. Além disto,

$$\text{Var}\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \frac{(n-1)^2}{\sigma^4} \text{Var}(s^2) = 2(n-1) \Rightarrow \text{Var}(s^2) = \frac{2\sigma^4}{n-1}$$

Exemplo:

Suponha que X_1, X_2, \dots, X_n formem uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 . Supondo que $n = 16$, determine os valores das seguintes probabilidades:

$$\text{a) } P\left(\frac{\sigma^2}{2} \leq \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \leq 2\sigma^2\right)$$

$$\text{b) } P\left(\frac{\sigma^2}{2} \leq \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \leq 2\sigma^2\right)$$

Solução:

$$\begin{aligned} \text{a) } P\left(\frac{\sigma^2}{2} \leq \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \leq 2\sigma^2\right) &= P\left(\frac{n}{2} \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \leq 2n\right) \\ &= P(8 \leq \chi_{16}^2 \leq 32) \\ &= 0,99 - 0,05 = 0,94 \end{aligned}$$

$$\begin{aligned} \text{b) } P\left(\frac{\sigma^2}{2} \leq \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \leq 2\sigma^2\right) &= P\left(\frac{n}{2} \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \leq 2n\right) \\ &= P(8 \leq \chi_{15}^2 \leq 32) \\ &= 0,995 - 0,10 = 0,895 \end{aligned}$$

4.5 A Distribuição *t-student*

Considere duas *v.a.* independentes Y e Z tais que Y possua uma distribuição normal padrão, e Z possua uma distribuição χ^2 com n graus de liberdade. Suponha que uma

v.a. X seja definida por

$$X = \frac{Y}{\left(\frac{Z}{n}\right)^{\frac{1}{2}}} \quad (4.5)$$

Então a distribuição de X é denominada **distribuição *t-student* com n graus de liberdade**.

4.5.1 Distribuição da Média Amostral

Suponha que X_1, X_2, \dots, X_n formem uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 . Sabemos que

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

e

$$\frac{\sum^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Logo,

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\left(\frac{\sum^n (X_i - \bar{X})^2}{\sigma^2(n-1)}\right)^{\frac{1}{2}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{s} \sim t_{n-1} \quad (4.6)$$

Exemplo:

Suponha que X_1, X_2, \dots, X_n formem uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 desconhecidos. Sejam \bar{X} e s^2 respectivamente a média e a variância amostral dos X_i 's. Para um tamanho de amostra $n = 16$, encontre o valor de k tal que:

$$P(\bar{X} > \mu + ks) = 0,05$$

$$P\left(\frac{(\bar{X} - \mu)\sqrt{n}}{s} > k\sqrt{n}\right) = 0,05$$

$$P(t_{n-1} > k\sqrt{n}) = P(t_{15} > 4k) = 0,05$$

Logo, $4k = 1,753$. Ou seja, $k = 0,438$.

4.5.2 Distribuição da diferença de médias amostrais

Considere duas amostras aleatórias

$$X_{A_1}, X_{A_2}, \dots, X_{A_{n_A}}$$

$$X_{B_1}, X_{B_2}, \dots, X_{B_{n_B}}$$

com n_A e n_B elementos, obtidas independentemente de duas populações A e B, normalmente distribuídas com médias μ_A e μ_B e variâncias $\sigma_A^2 = \sigma_B^2 = \sigma^2$, respectivamente. Sejam:

$$\bar{X}_A = \frac{\sum_{i=1}^{n_A} X_{A_i}}{n_A} \quad \text{e} \quad \bar{X}_B = \frac{\sum_{i=1}^{n_B} X_{B_i}}{n_B}$$

as médias respectivas das duas amostras.

Temos então que:

$$\bar{X}_A - \bar{X}_B \sim N\left(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$$

Portanto,

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0, 1)$$

Temos também que:

$$\frac{(n_A - 1)s_A^2}{\sigma_A^2} \sim \chi_{n_A-1}^2, \quad \text{e} \quad \frac{(n_B - 1)s_B^2}{\sigma_B^2} \sim \chi_{n_B-1}^2$$

Logo,

$$\frac{(n_A - 1)s_A^2}{\sigma_A^2} + \frac{(n_B - 1)s_B^2}{\sigma_B^2} \sim \chi_{n_A+n_B-2}^2$$

Pela definição da distribuição t-student, e considerando $\sigma_A^2 = \sigma_B^2 = \sigma^2$, temos que:

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t_{n_A + n_B - 2}$$

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}} \sim t_{n_A + n_B - 2}$$

4.6 Distribuição F

Considere duas *v.a.* independentes Y e Z tais que Y possua uma distribuição χ^2 com m graus de liberdade, e Z possua uma distribuição χ^2 com n graus de liberdade. Suponha que uma *v.a.* X seja definida por

$$X = \frac{Y/m}{Z/n} = \frac{nY}{mZ} \quad (4.7)$$

Então a distribuição de X é denominada **distribuição F com m e n graus de liberdade**.

Propriedade da distribuição F:

Se uma variável aleatória X possui uma distribuição F com m e n graus de liberdade, então $1/X$ possui uma distribuição F com n e m graus de liberdade.

4.6.1 Distribuição da Razão entre duas Variâncias Amostrais

Suponha que X_1, X_2, \dots, X_n formem uma amostra aleatória de m observações de uma distribuição normal com média μ_1 e variância σ_1^2 desconhecidos, e suponha que Y_1, Y_2, \dots, Y_n formem uma amostra aleatória de n observações de uma distribuição normal com média μ_2 e variância σ_2^2 desconhecidos. Sabemos que

$$\frac{\sum_{i=1}^m (X_i - \bar{X})^2}{\sigma_1^2} = \frac{(m-1)s_1^2}{\sigma_1^2} \sim \chi_{m-1}^2$$

e

$$\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma_2^2} = \frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi_{n-1}^2$$

Logo,

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{m-1, n-1} \quad (4.8)$$

Se $\sigma_1^2 = \sigma_2^2$, então

$$\frac{s_1^2}{s_2^2} \sim F_{m-1, n-1} \quad (4.9)$$

Exemplo:

Suponha que uma amostra de tamanho 6 seja retirada de uma população normalmente distribuída com média μ_1 e variância 30, e que uma amostra de tamanho 3 seja retirada de uma outra população normalmente distribuída com média μ_2 e variância 76. Qual é a probabilidade de $s_1^2 > s_2^2$?

Solução:

$$\begin{aligned} P(s_1^2 > s_2^2) &= P\left(\frac{s_1^2}{s_2^2} > 1\right) = P\left(\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} > \frac{\sigma_2^2}{\sigma_1^2}\right) \\ &= P\left(F_{5,29} > \frac{76}{30}\right) = P(F_{5,29} > 2,5) = 0,05 \end{aligned}$$

4.7 Estimação por Intervalos - Intervalos de Confiança

4.7.1 Intervalo de Confiança para a Média Populacional μ

1 - Desvio Padrão (σ) conhecido

Como

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1), \quad (4.10)$$

podemos determinar, pela tabela da distribuição normal padrão, o número $z_{\frac{\alpha}{2}}$ tal que $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, e portanto, :

$$P\left[-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}\right] = 1 - \alpha,$$

ou seja

$$P \left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha. \quad (4.11)$$

Chamando de L_1 o limite inferior e L_2 o limite superior, isto é:

$$L_1 = \bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad (4.12)$$

$$L_2 = \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad (4.13)$$

dizemos que o intervalo (L_1, L_2) é um intervalo de confiança para μ com coeficiente de confiança $(1 - \alpha)$, ou em outras palavras, que μ cai no intervalo (L_1, L_2) com confiança α .

Exemplo:

Suponha que se extraia uma amostra de tamanho 35 de uma população com média μ e desvio padrão conhecido e igual a 3,90. Suponha que a média amostral seja 44,8. Determinar um intervalo com 95% de confiança para μ .

Solução:

Temos que:

$$L_1 = 44,8 - 1,96 \cdot \frac{3,90}{\sqrt{35}} = 43,51 \quad (4.14)$$

$$L_2 = 44,8 + 1,96 \cdot \frac{3,90}{\sqrt{35}} = 46,09. \quad (4.15)$$

Logo, o intervalo com 95% de confiança para μ é $[43,51; 46,09]$

2 - Desvio Padrão (σ) desconhecido

Como

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}, \quad (4.16)$$

podemos determinar, pela tabela da distribuição t-student com $n - 1$ g.l., o número $t_{\frac{\alpha}{2},(n-1)}$ tal que $P(T > t_{\frac{\alpha}{2},(n-1)}) = \frac{\alpha}{2}$, e portanto, :

$$P \left[-t_{\frac{\alpha}{2},(n-1)} \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2},(n-1)} \right] = 1 - \alpha,$$

ou seja

$$P \left[\bar{X} - t_{\frac{\alpha}{2}(n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}(n-1)} \frac{s}{\sqrt{n}} \right] = 1 - \alpha. \quad (4.17)$$

Portanto o intervalo

$$\left(\bar{X} - t_{\frac{\alpha}{2}(n-1)} \frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}(n-1)} \frac{s}{\sqrt{n}} \right)$$

é um intervalo de confiança para μ com coeficiente de confiança $(1 - \alpha)$.

Exemplo:

Suponha que se extraia uma amostra de tamanho 25 de uma população com média μ e desvio padrão desconhecido. Suponha que a média amostral seja 4,004 e o desvio padrão amostral seja 0,366. Determinar intervalos com 95% e 99% de confiança para μ .

Solução:

Temos que $t_{0,025; 24} = 2,064$:

$$L_1 = 4,004 - 2,064 \cdot \frac{0,366}{\sqrt{25}} = 3,853 \quad (4.18)$$

$$L_2 = 4,004 + 2,064 \cdot \frac{0,366}{\sqrt{25}} = 4,155 \quad (4.19)$$

$$(4.20)$$

Logo, o intervalo com 95% de confiança para μ é $[3,853; 4,155]$.

Analogamente, temos que $t_{0,005; 24} = 2,797$:

$$L_1 = 4,004 - 2,797 \cdot \frac{0,366}{\sqrt{25}} = 3,799 \quad (4.21)$$

$$L_2 = 4,004 + 2,797 \cdot \frac{0,366}{\sqrt{25}} = 4,209 \quad (4.22)$$

$$(4.23)$$

Logo, o intervalo com 95% de confiança para μ é $[3,799; 4,209]$.

4.7.2 Intervalo de Confiança para a Variância Populacional σ^2

Vimos que

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2. \quad (4.24)$$

Logo, podemos determinar pela tabela da distribuição χ^2 com $n-1$ g.l., os números $\chi_{(1-\frac{\alpha}{2}),[n-1]}^2$ e $\chi_{\frac{\alpha}{2},[n-1]}^2$ tal que:

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

ou seja

$$P \left[\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2},[n-1]}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{(1-\frac{\alpha}{2}),[n-1]}^2} \right] = 1 - \alpha. \quad (4.25)$$

Portanto o intervalo

$$\left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2},[n-1]}^2}, \frac{(n-1)s^2}{\chi_{(1-\frac{\alpha}{2}),[n-1]}^2} \right)$$

é um intervalo de confiança para σ^2 com coeficiente de confiança $(1 - \alpha)$.

Exemplo:

Suponha que seja retirada uma amostra de tamanho cinco de uma população normalmente distribuída, e que se tenha encontrado uma variância amostral de 13,52. Construa um intervalo com 95% de confiança para a variância populacional.

Solução:

Temos que $\chi_{0,975;4}^2 = 0,484$ e $\chi_{0,025;4}^2 = 11,143$. Portanto, os limites inferior e superior do I.C. de 95% para σ^2 são:

$$L_1 = \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2},[n-1]}^2} = \frac{4(13,52)}{11,143} = 4,85$$

$$L_2 = \frac{(n-1)s^2}{\chi_{(1-\frac{\alpha}{2}), [n-1]}^2} = \frac{4(13,52)}{0,484} = 111,74$$

Portanto, $P(4,85 \leq \sigma^2 \leq 111,74) = 0,95$.

4.7.3 Intervalo de Confiança para a diferença de médias de duas Populações

1 - Variâncias σ_1^2 e σ_2^2 Conhecidas

Como

$$\begin{aligned} \frac{\bar{X}_1 - \mu_1}{\frac{\sigma_1}{\sqrt{n_1}}} &\sim \mathfrak{N}(0,1) \\ &e \\ \frac{\bar{X}_2 - \mu_2}{\frac{\sigma_2}{\sqrt{n_2}}} &\sim \mathfrak{N}(0,1). \end{aligned} \quad (4.26)$$

Logo

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathfrak{N}(0,1) \quad (4.27)$$

Portanto o intervalo de confiança para $\mu_1 - \mu_2$ com coeficiente de confiança $1 - \alpha$ é dado por

$$\left((\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

2 - σ_1^2 e σ_2^2 Desconhecidas mas $\sigma_1^2 = \sigma_2^2$

Vimos que

$$\frac{\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{1/n_1 + 1/n_2}}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1 + n_2 - 2} \quad (4.28)$$

Logo, o intervalo de confiança para $\mu_1 - \mu_2$, quando $\sigma_1^2 = \sigma_2^2$, com coeficiente de

confiança $1 - \alpha$ é dado por

$$\left((\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, [n_1+n_2-2]} \cdot \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Exemplo:

Uma amostra de 10 lâmpadas elétricas, da marca A, apresentou a vida média de 1400 horas e desvio padrão de 120 horas. Uma amostra de 20 lâmpadas elétricas, da marca B, apresentou a vida média de 1200 horas e o desvio padrão de 100 horas. Supondo que $\sigma_A = \sigma_B$, determinar os limites de confiança de a) 95% e b) 99% para a diferença entre as vidas médias das populações das marcas A e B.

Solução:

a) Para $1 - \alpha = 0,95$ tem-se que $t_{0,025,28} = 2,048$. Portanto, o I.C. de 95% para $\mu_A - \mu_B$ será:

$$\left((1400 - 1200) \pm 2,048 \cdot \sqrt{\frac{9(120)^2 + 19(100)^2}{28}} \cdot \sqrt{\frac{1}{10} + \frac{1}{20}} \right) = (200 \pm 67,77)$$

Ou seja,

$$P(132,23 \leq \mu_A - \mu_B \leq 267,77) = 0,95$$

b) Analogamente, para $1 - \alpha = 0,99$ tem-se que $t_{0,005,28} = 2,763$. Portanto, o I.C. de 99% para $\mu_A - \mu_B$ será:

$$\left((1400 - 1200) \pm 2,763 \cdot \sqrt{\frac{9(120)^2 + 19(100)^2}{28}} \cdot \sqrt{\frac{1}{10} + \frac{1}{20}} \right) = (200 \pm 91,43)$$

Ou seja,

$$P(108,57 \leq \mu_A - \mu_B \leq 291,43) = 0,99$$

4.7.4 Intervalo de Confiança para Razão das Variâncias σ_1^2/σ_2^2

Vimos que

$$\frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} \sim F_{n_2-1, n_1-1} \quad (4.29)$$

Logo, podemos determinar pela tabela da distribuição F com $n_2 - 1$ e $n_1 - 1$ g.l., os números $F_{(1-\alpha/2), [n_2-1], [n_1-1]}$ e $F_{\alpha/2, [n_2-1], [n_1-1]}$ tal que:

$$P \left[F_{(1-\alpha/2), [n_2-1], [n_1-1]} \leq \frac{s_2^2 \sigma_1^2}{s_1^2 \sigma_2^2} \leq F_{\alpha/2, [n_2-1], [n_1-1]} \right] \sim 1 - \alpha \quad (4.30)$$

ou

$$P \left[\frac{1}{F_{\alpha/2, [n_1-1], [n_2-1]}} \cdot \frac{s_1^2}{s_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq F_{\alpha/2, [n_2-1], [n_1-1]} \cdot \frac{s_1^2}{s_2^2} \right] \sim 1 - \alpha \quad (4.31)$$

Exemplo:

Duas máquinas A e B produzem parafusos com o mesmo tamanho médio. Duas amostras de tamanho $n_A = 61$ e $n_B = 41$ dos parafusos de A e B foram analisadas e os desvios padrões amostrais foram $s_A^2 = 3,5$ mm e $s_B^2 = 4,5$ mm. Determine um intervalo de 95% de confiança para $\frac{\sigma_A^2}{\sigma_B^2}$.

Solução:

Tem-se que $F_{0,975,40,60} = 1/F_{0,025,60,40} = 1/1,80 = 0,556$ e $F_{0,025,40,60} = 1,74$. Logo,

$$P \left[0,556 \cdot \frac{3,5}{4,5} \leq \frac{\sigma_A^2}{\sigma_B^2} \leq 1,74 \cdot \frac{3,5}{4,5} \right] = 0,95$$

$$P \left[0,432 \leq \frac{\sigma_A^2}{\sigma_B^2} \leq 1,353 \right] = 0,95$$

4.7.5 Intervalo de Confiança para uma Proporção

Admita-se que uma população é infinita e que a probabilidade de ocorrência de um evento (denominado de sucesso) seja p . Considerem-se todas as amostras possíveis de tamanho n extraídas da população e, para cada amostra, determinaremos a proporção \hat{p} de sucessos.

Vimos que:

$$E(\hat{p}) = p \quad (4.32)$$

$$Var(\hat{p}) = \frac{p(1-p)}{n} \quad (4.33)$$

Vimos também que para grandes valores de n , a distribuição de \hat{p} é uma normal, isto é

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \quad (4.34)$$

Portanto, o intervalo de confiança para p , com coeficiente $1 - \alpha$ é dado por:

$$\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right) \quad (4.35)$$

Para n grande, em geral substitui-se p por \hat{p} , resultando em:

$$P\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

4.7.6 Intervalo de Confiança para Diferença de Proporções

Sejam duas proporções p_1 e p_2 , e suas respectivas proporções amostrais \hat{p}_1 e \hat{p}_2 , baseadas em amostras de tamanhos n_1 e n_2 . Para grandes tamanhos de amostra tem-se que:

$$\hat{p}_1 - \hat{p}_2 \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right) \quad (4.36)$$

Portanto, o intervalo de confiança para $p_1 - p_2$, com coeficiente de confiança $1 - \alpha$ é dado por:

$$\left((\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

CAPÍTULO 5

TESTES DE HIPÓTESES

Consideraremos aqui problemas estatísticos envolvendo um parâmetro θ cujo valor é desconhecido mas deve cair dentro de um certo domínio Ω (isto é, Ω é o conjunto de todos os possíveis valores de θ). Vamos supor que Ω possa ser particionado em 2 (dois) subconjuntos distintos Ω_0 e Ω_1 , e que o estatístico deva decidir se o valor desconhecido de θ cai em Ω_0 ou em Ω_1 .

5.1 Hipótese Nula e Hipótese Alternativa

Seja H_0 a hipótese de que $\theta \in \Omega_0$ e H_1 a hipótese de que $\theta \in \Omega_1$, isto é:

$$H_0 : \theta \in \Omega_0$$

$$H_1 : \theta \in \Omega_1$$

Como Ω_0 e Ω_1 são disjuntos ($\Omega_0 \cup \Omega_1 = \Omega$), somente umas das hipóteses são verdadeiras. O estatístico deve decidir se aceita H_0 ou se aceita H_1 . Um problema desse tipo é chamado um problema de **teste de hipóteses**.

- H_0 é denominada **hipótese nula**, e
- H_1 é denominada **hipótese alternativa**

5.2 Região Crítica do Teste

O procedimento adotado para decidir se ele aceita H_0 ou aceita H_1 é denominado **procedimento do teste**, ou simplesmente **teste**.

Suponha que antes de decidir se aceita ou não a hipótese nula, ele observa uma amostra aleatória X_1, X_2, \dots, X_n . Seja S o espaço amostral do vetor $\vec{X} = (X_1, X_2, \dots, X_n)$, isto é, S é o conjunto de todos os possíveis resultados da amostra.

Num problema desse tipo o estatístico especifica um **procedimento de teste** que consiste em dividir o espaço amostral em dois subconjuntos: um deles consiste dos valores da amostra para o qual ele aceita H_0 , e o outro contem os valores para o qual ele rejeita H_0 .

O subconjunto para o qual H_0 será rejeitada é chamada **região crítica do teste**. O complemento da região crítica contém portanto todos os possíveis valores para o qual H_0 será aceita.

5.3 Erros do Tipo I e Erros do tipo II

Quando estabelecemos um procedimento do teste, podemos incorrer em dois tipos de erros:

- a) O de rejeitar H_0 quando ela é de fato verdadeira. Este erro é denominado **erro do tipo I**. A probabilidade (α) deste tipo de erro ocorrer é controlada pelo estatístico e é denominada **nível de significância do teste**.
- b) O de aceitar H_0 quando H_1 é verdadeira. Este erro é denominado **erro do tipo II**. A probabilidade deste erro ocorrer é representada por β .

A Tabela 5.1 mostra os dois tipos de erros.

TABELA 5.1 – Representação do erros do tipo I e II

	H_0 é verdadeira	H_0 é falsa
aceita H_0	$1 - \alpha$ (Coef. de confiança)	β
rejeita H_0	α (nível de significância)	$1 - \beta$ (poder do Teste)

5.4 Teste da hipótese de que a média populacional tem um valor específico

5.4.1 σ conhecido

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

1. Retira-se uma amostra de tamanho \underline{n} e calcula-se \bar{X} .

2. Calcula-se o valor da estatística

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

3. Sob a hipótese nula, tem-se que Z possui uma distribuição normal padrão. Portanto,

Rejeita-se H_0 se $|Z| > Z_{\alpha/2}$ (isto é, se $Z < -Z_{\alpha/2}$ ou $Z > Z_{\alpha/2}$)

Aceita-se H_0 se $|Z| < Z_{\alpha/2}$ (isto é, $-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$),

onde α é o nível de significância do teste.

5.4.2 σ desconhecido

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Calcula-se a estatística

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Sob a hipótese nula, tem-se que t possui uma distribuição t-Student com $n - 1$ graus de liberdade. Portanto,

Rejeita-se H_0 se $|t| > t_{\alpha/2, [n-1]}$

Aceita-se H_0 se $|t| \leq t_{\alpha/2, [n-1]}$

Observação

Se os testes das seções (5.4.1 e 5.4.2) tiverem uma hipótese alternativa unilateral (i.e. se $H_1 : \mu > \mu_0$, ou $H_1 : \mu < \mu_0$) o teste deverá rejeitar unilateralmente (i.e. se $t > t_{\alpha, [n-1]}$, ou $t < -t_{\alpha, [n-1]}$, respectivamente.)

Exemplos:

1. Uma máquina automática para encher pacotes de café enche-os segundo uma distribuição normal, com média μ e variância sempre igual a 400 g^2 . A máquina foi regulada para $\mu = 500\text{g}$. Colhe-se, periodicamente uma amostra de 16 pacotes para verificar se a produção está sob controle, isto é, se $\mu = 500\text{g}$ ou não. Se uma dessas amostras apresentasse uma média amostral de 492 g, você pararia ou não a produção para regular máquina, considerando o nível de significância de 1%? Para quais valores de média amostral a máquina será regulada?

Solução:

As hipóteses são:

$$H_0 : \mu = 500$$

$$H_1 : \mu \neq 500$$

Pelos dados do problema a variância é sempre a mesma e igual a $\sigma^2 = 400$. A estatística a ser calculada é:

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{492 - 500}{\sqrt{400/16}} = -1,6$$

Ao nível de significância de 0,01, a regra de decisão é:

$$\text{Rejeita-se } H_0 \text{ se } |z| > z_{0,005} = 2,58$$

$$\text{Aceita-se } H_0 \text{ se } |z| < 2,58$$

Portanto, aceita-se H_0 , e a máquina não necessita ser parada. A máquina só sera parada se

$$\bar{x} < \mu_0 - z_{0,005} \frac{\sigma}{\sqrt{n}} = 500 - 2,58 \frac{20}{4}, \text{ isto é, } \bar{x} < 487,1, \text{ ou}$$

$$\bar{x} > \mu_0 + z_{0,005} \frac{\sigma}{\sqrt{n}} = 500 + 2,58 \frac{20}{4}, \text{ isto é, } \bar{x} > 512,9$$

2. A tensão de ruptura dos cabos produzidos por um fabricante apresenta média de 1800 kg e desvio padrão de 100 kg. Mediante nova técnica no processo de fabricação, proclama-se que a tensão de ruptura pode ter aumentado. Para testar esta declaração, ensaiou-se uma amostra de 50 cabos, tendo-se determinado a tensão média de ruptura de 1850 kg. Pode-se confirmar a declaração ao nível de significância de 0,01?

Solução:

As hipóteses são:

$$H_0 : \mu = 1800 \quad (\text{não houve modificação da tensão de ruptura})$$

$$H_1 : \mu > 1800 \quad (\text{houve modificação da tensão de ruptura})$$

Como supõe-se que o desvio padrão não se tenha modificado, temos que $\sigma = 100$. A estatística a ser calculada é:

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1850 - 1800}{100/\sqrt{50}} = 3,55$$

Ao nível de significância de 0,01, a regra de decisão é:

$$\text{Rejeita-se } H_0 \text{ se } z > z_{0,01} = 2,33$$

$$\text{Aceita-se } H_0 \text{ se } z < 2,33$$

Portanto, rejeita-se H_0 , e confirma-se a declaração.

3. Um fabricante afirma que seus cigarros contém não mais que 30 mg de nicotina. Uma amostra de 25 cigarros fornece média de 31,5 mg e desvio padrão de 3 mg. Ao nível de 5%, os dados refutam ou não a afirmação do fabricante?

Solução:

Neste caso, as hipóteses são:

$$H_0 : \mu = 30$$

$$H_1 : \mu > 30$$

Como não se conhece a variância populacional, e esta foi estimada pela amostra, devemos utilizar a estatística t :

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{31,5 - 30}{3/\sqrt{25}} = 2,5$$

A regra de decisão é dada por

$$\text{Rejeita-se } H_0 \text{ se } t > t_{\alpha, [n-1]} = t_{0,05, 24} = 1,711$$

$$\text{Aceita-se } H_0 \text{ se } t < 1,711$$

Portanto, rejeita-se H_0 , ou seja, há evidências de que os cigarros contenham mais de 30 g de nicotina.

5.5 Controlando o erro tipo II (β)

Vimos que o erro tipo I representa o erro de se rejeitar H_0 quando ela é de fato verdadeira. A probabilidade deste erro é α e é fixada e portanto controlada pelo estatístico.

Temos também o erro tipo II (*beta*) que representa o erro de aceitar H_0 quando ela é falsa.

Quando rejeitamos H_0 , automaticamente estamos aceitando H_1 , isto é, estamos aceitando que o parâmetro pertença ao espaço definido pela hipótese H_1 . O erro tipo II dependerá do verdadeiro valor do parâmetro. Quanto mais afastado o verdadeiro valor do parâmetro estiver do valor especificado em H_0 , menor será o erro tipo II. Portanto, para calcular β , temos que especificar este valor em H_1 , isto é, ter-se as hipóteses definidas por:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1$$

Para exemplificar isto, considere o exemplo 2 dado anteriormente. Suponha que as hipóteses tenham sido definidas da seguinte maneira:

$$H_0 : \mu = 1800$$

$$H_1 : \mu = 1850$$

Vimos que o erro tipo I foi fixado em 0,01, supondo que H_0 fosse verdadeira, isto é,

$$\begin{aligned} 0,01 &= P(\text{erro I}) = P(Z \geq 2,33) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq 2,33 \mid \mu = 1800\right) = \\ &= P\left(\frac{\bar{X} - 1800}{100/\sqrt{50}} \geq 2,33\right) = P\left(\bar{X} \geq 1800 + 2,33 \frac{100}{\sqrt{50}}\right) = \\ &= P(\bar{X} \geq 1832,95) \end{aligned}$$

Portanto, a probabilidade do erro tipo II será:

$$\begin{aligned}\beta &= P(\text{erro II}) = P(\bar{X} < 1832,95 \mid \mu = 1850) = \\ &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{1832,95 - 1850}{100/\sqrt{50}}\right) = P(Z < -1,206) \cong 0,1131\end{aligned}$$

O cálculo acima pode ser efetuado para vários valores de μ . Considerando-se $\beta(\mu)$ a probabilidade de aceitar H_0 como função de μ , isto é,

$$\beta(\mu) = P(\text{aceitar } H_0 \mid \mu) = P(\bar{X} < 1832,95 \mid \mu),$$

pode-se calcular a função

$$\pi(\mu) = 1 - \beta(\mu).$$

Esta função é denominada **função poder do teste**.

5.6 Teste da hipótese de que a variância populacional tem um valor específico

Suponha que uma variável seja normalmente distribuída com uma variância desconhecida e se deseje efetuar o seguinte teste de hipóteses:

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Calcula-se a estatística

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Rejeita-se H_0 se $X^2 < \chi_{1-\alpha/2, [n-1]}^2$ ou $X^2 > \chi_{\alpha/2, [n-1]}^2$

Aceita-se H_0 se $\chi_{1-\alpha/2, [n-1]}^2 \leq X^2 \leq \chi_{\alpha/2, [n-1]}^2$

Observação

1. Se a hipótese alternativa fosse

$$H_1 : \sigma^2 > \sigma_0^2,$$

H_0 seria rejeitada se $X^2 > \chi_{\alpha, [n-1]}^2$.

2. Se a hipótese alternativa fosse

$$H_1 : \sigma^2 < \sigma_0^2,$$

H_0 seria rejeitada se $X^2 < \chi_{1-\alpha, [n-1]}^2$.

Exemplo:

Uma das maneiras de manter sob controle a qualidade de um produto é controlar a sua variabilidade. Uma máquina de encher pacotes de café está regulada para enchê-los com média de 500 g e desvio padrão de 10 g. Colheu-se uma amostra de 16 pacotes e observou-se uma variância $s^2 = 169g^2$. Supondo que o peso de cada pacote segue uma distribuição normal, você diria que a máquina está desregulada com relação à variância?

Solução:

Deseja-se testar:

$$H_0 : \sigma^2 = 100$$

$$H_1 : \sigma^2 \neq 100$$

A estatística a ser calculada é:

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(15)(169)}{100} = 25,35$$

e o procedimento do teste é: Aceita-se H_0 se $\chi_{1-\alpha/2, [n-1]}^2 \leq X^2 \leq \chi_{\alpha/2, [n-1]}^2$, isto é,

Aceita-se H_0 se $6,262 \leq X^2 \leq 27,488$,

e

Rejeita-se H_0 se $X^2 < 6,262$ ou $X^2 > 27,488$

Portanto, aceita-se H_0 , e concluímos que a máquina não está desregulada quanto à variância.

5.7 Teste da razão de variâncias

Suponha que se deseje testar:

$$\begin{aligned}H_0 &: \sigma_1^2 = \sigma_2^2 \\H_1 &: \sigma_1^2 \neq \sigma_2^2\end{aligned}\tag{5.1}$$

ou, equivalentemente,

$$\begin{aligned}H_0 &: \frac{\sigma_1^2}{\sigma_2^2} = 1 \\H_1 &: \frac{\sigma_1^2}{\sigma_2^2} \neq 1\end{aligned}\tag{5.2}$$

O procedimento do teste é:

Calcula-se a estatística

$$f = \frac{s_1^2}{s_2^2}\tag{5.3}$$

Vimos que, sob a hipótese H_0 , a estatística f possui uma distribuição F com $n_1 - 1$ e $n_2 - 1$ graus de liberdade. Portanto,

Aceita-se H_0 ao nível de significância α se

$$\frac{1}{F_{\alpha/2, [n_2-1], [n_1-1]}} \leq f \leq F_{\alpha/2, [n_1-1], [n_2-1]}\tag{5.4}$$

Rejeita-se H_0 ao nível de significância de α se

$$f < \frac{1}{F_{\alpha/2, [n_2-1], [n_1-1]}}\tag{5.5}$$

ou

$$f > F_{\alpha/2, [n_1-1], [n_2-1]}\tag{5.6}$$

Exemplo:

Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é por meio do desvio padrão de seus salários. A fábrica A diz ser mais coerente na política salarial do que a fábrica B. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de A, e 15 de B, obtendo-se os desvios padrões $s_A = 1000$ reais e $s_B = 1600$ reais. Qual seria a sua conclusão?

Solução:

A hipótese a ser testada é:

$$\begin{aligned}H_0 &: \sigma_A^2 = \sigma_B^2 \\H_1 &: \sigma_A^2 \neq \sigma_B^2\end{aligned}\tag{5.7}$$

Temos que:

$$f = \frac{s_A^2}{s_B^2} = \frac{1000}{1500} = 0,667$$

Devemos aceitar H_0 ao nível de significância $\alpha = 0,05$ se

$$\frac{1}{F_{0,025,[14],[9]}} \leq f \leq F_{0,025,[9],[14]}\tag{5.8}$$

ou seja, se

$$\begin{aligned}\frac{1}{3,77} &\leq f \leq 3,12 \\0,27 &\leq f \leq 3,12\end{aligned}\tag{5.9}$$

Como este é o caso, aceitamos H_0 ao nível de significância de 0,05, e concluímos que as duas fábricas são igualmente homogêneas.

5.8 Teste da hipótese da igualdade de duas médias

Suponha que se tenha

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

5.8.1 σ_1^2 e σ_2^2 conhecidas

Calcula-se a estatística

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (5.10)$$

Sabemos que, sob a hipótese H_0 , a variável Z possui uma distribuição normal padrão. Portanto, o procedimento do teste consiste em:

Rejeita-se H_0 se $|Z| > Z_{\alpha/2}$

Aceita-se H_0 se $|Z| \leq Z_{\alpha/2}$

5.8.2 σ_1^2 e σ_2^2 desconhecidas, mas $\sigma_1^2 = \sigma_2^2$

Suponha que a hipótese de igualdade de variâncias não seja rejeitada. Então podemos supor que $\sigma_1^2 = \sigma_2^2$, mas esta variância comum não é conhecida. Para efetuar o teste de igualdade de médias, neste caso, procedemos da seguinte maneira:

Calcula-se a estatística

$$t = \frac{\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}} \quad (5.11)$$

Como vimos anteriormente, esta estatística possui uma distribuição t-Student com $n_1 + n_2 - 2$ graus de liberdade. Portanto,

Rejeita-se H_0 se $|t| > t_{\alpha/2; n_1+n_2-2}$

Aceita-se H_0 se $|t| \leq t_{\alpha/2; n_1+n_2-2}$

5.8.3 σ_1^2 e σ_2^2 desconhecidas, mas $\sigma_1^2 \neq \sigma_2^2$

Suponha que a hipótese de igualdade de variâncias tenha sido rejeitada. Neste caso, devemos calcular a estatística

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (5.12)$$

Pode-se provar que, sob a hipótese que H_0 é verdadeira, a estatística acima aproxima-se de uma distribuição t de Student com graus de liberdade dado aproximadamente por:

$$\nu = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}},$$

onde

$$A = \frac{s_1^2}{n_1}, \quad \text{e} \quad B = \frac{s_2^2}{n_2}.$$

Sendo este valor geralmente fracionário, costuma-se arredondar para o inteiro mais próximo para obter o número de graus de liberdade.

O procedimento do teste é então

Rejeita-se H_0 se $|t| > t_{\alpha/2, \nu}$

Aceita-se H_0 se $|t| \leq t_{\alpha/2, \nu}$

Exemplos:

1. Uma amostra de 10 lâmpadas elétricas, da marca A, apresentou a vida média de 1400 horas e uma amostra de 20 lâmpadas elétricas, da marca B, apresentou a vida média de 1200 horas. Suponha que os desvios padrões populacionais dos tempos de vida das lâmpadas das duas marcas sejam conhecidos e iguais a 120 e 100, respectivamente. Teste, ao nível de significância de 99%, a hipótese que as duas marcas produzem lâmpadas com o mesmo tempo médio de vida.

Solução:

Queremos testar a hipótese:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

Como estamos supondo que as variâncias são conhecidas, podemos usar a estatística

$$Z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{1400 - 1200}{\sqrt{\frac{120^2}{10} + \frac{100^2}{20}}} = 4,54 \quad (5.13)$$

O procedimento do teste consiste em:

Rejeita-se H_0 se $|Z| > Z_{0,005} = 2,58$

Aceita-se H_0 se $|Z| \leq 2,58$

Como o valor da estatística pertence à região de rejeição, concluímos que H_0 é rejeitada e que as lâmpadas não possuem o mesmo tempo médio de vida.

2. Duas técnicas de vendas são aplicadas por dois grupos de vendedores: a técnica A, por 12 vendedores, e a técnica B, por 15 vendedores. Espera-se que a técnica B produza melhores resultados que a técnica A. No final de um mês, os vendedores de A venderam uma média de 68 itens, com uma variância de 50, enquanto que os vendedores de B venderam uma média de 76 itens com uma variância de 75. Testar, ao nível de significância de 5%, se a técnica B é realmente melhor que a técnica A.

Solução:

Supondo que as vendas sejam normalmente distribuídas, vamos inicialmente testar a hipótese de que as variâncias são iguais:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

Temos que:

$$f = \frac{s_A^2}{s_B^2} = \frac{50}{75} = 0,667$$

Devemos aceitar H_0 ao nível de significância $\alpha = 0,05$ se

$$\frac{1}{F_{0,025,[14],[11]}} \leq f \leq F_{0,025,[11],[14]} \quad (5.14)$$

ou seja, se aproximadamente

$$\begin{aligned} \frac{1}{3,52} &\leq f \leq 3,06 \\ 0,28 &\leq f \leq 3,06 \end{aligned}$$

Logo, aceitamos H_0 ao nível de significância de 0,05, e concluímos que as variâncias são iguais.

Portanto, agora podemos testar a hipótese de igualdade de médias, sob a suposição que as vendas pelas duas técnicas possuem uma variância comum, mas desconhecida:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Devemos calcular a estatística

$$t = \frac{\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}}{\sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}} = \frac{\frac{68 - 76}{\sqrt{\frac{1}{12} + \frac{1}{15}}}}{\sqrt{\frac{(11)(50) + (14)(75)}{25}}} = -2,56$$

O procedimento do teste será:

Rejeita-se H_0 se $t < -t_{0,05,[25]} = -1,708$

Aceita-se H_0 se $t \geq -1,708$

Como o valor encontrado pertence à região de rejeição, rejeitamos H_0 e concluímos que a técnica B produz melhores resultados que a técnica A.

3. Queremos testar as resistências de dois tipos de vigas de aço, A e B. Tomando-se $n_A = 15$ vigas do tipo A e $n_B = 20$ vigas do tipo B, obtemos os valores da tabela a seguir. Testar a hipótese que as resistências médias dos dois tipos de vigas são iguais, ao nível de significância de 5%.

Tipo	Média	Variância
A	70,5	71,6
B	84,3	169,5

Solução:

Vamos inicialmente testar a hipótese de que as variâncias são iguais:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

Temos que:

$$f = \frac{s_A^2}{s_B^2} = \frac{71,6}{169,5} = 0,42$$

Devemos aceitar H_0 ao nível de significância $\alpha = 0,10$ se

$$\frac{1}{F_{0,05,[19],[14]}} \leq f \leq F_{0,05,[14],[19]} \quad (5.15)$$

ou seja, se aproximadamente

$$\begin{aligned} \frac{1}{2,33} &\leq f \leq 2,20 \\ 0,43 &\leq f \leq 2,20 \end{aligned}$$

Logo, rejeitamos H_0 ao nível de significância de 0,10, e concluímos que as variâncias não são iguais.

Portanto, agora podemos testar a hipótese de igualdade de médias:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

A estatística a ser utilizada deve ser:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{70,5 - 84,3}{\sqrt{\frac{71,6}{15} + \frac{169,5}{20}}} = -3,79 \quad (5.16)$$

O valor crítico deve ser encontrado pela tabela t-Student com graus de liberdade dado por:

$$\nu = \frac{175,51}{1,627 + 3,780} = 32,46 \cong 32,$$

O procedimento do teste é então

Rejeita-se H_0 se $|t| > t_{0,025,32} \cong 2,042$

Aceita-se H_0 se $|t| \leq 2,042$

Portanto, rejeitamos H_0 , e concluímos que há evidências de que os dois tipos de vigas possuem resistências médias diferentes.

5.9 Teste para proporção

Suponha que se deseje testar a hipótese:

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

Calcula-se a estatística

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (5.17)$$

Rejeita-se H_0 se $|Z| > Z_{\alpha/2}$

Aceita-se H_0 se $|Z| \leq Z_{\alpha/2}$

Exemplo:

Em uma experiência sobre percepção extra sensorial (PES), um indivíduo em uma sala é solicitado a declarar a cor vermelha ou preta de uma carta escolhida, de um baralho de 50 cartas, por outro indivíduo colocado em outra sala. O indivíduo desconhece quantas cartas vermelhas ou pretas há no baralho. Se o sujeito identifica corretamente 32 cartas, determinar se os resultados são significativos, ao nível de significância de 5% e 1%.

Solução:

Queremos testar a hipótese:

$$H_0 : p = 0,50 \quad \text{o indivíduo não tem PES}$$

$$H_1 : p > 0,50 \quad \text{o indivíduo tem PES}$$

Temos que $\hat{p} = \frac{32}{50} = 0,64$. Calculamos a estatística

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,64 - 0,50}{\sqrt{\frac{0,50^2}{50}}} = 1,98 \quad (5.18)$$

Ao nível de significância de 0,05 o procedimento do teste será:

Rejeita-se H_0 se $Z > Z_{0,05} = 1,645$

Aceita-se H_0 se $Z \leq 1,645$

Portanto, ao nível de significância de 0,05 rejeitamos H_0 , e concluímos que o indivíduo tem faculdades de PES.

Para o nível de significância de 0,01 o valor crítico será 2,33. Como o valor calculado é menor que o valor crítico, aceitamos H_0 , e concluímos que os resultados são devidos ao acaso e que o indivíduo não tem faculdades de PES.

5.9.1 Diferença entre proporções

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

Como

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0 \quad (\text{sob } H_0) \quad (5.19)$$

e

$$\sigma_{\hat{p}_A - \hat{p}_B}^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \quad (\text{sob } H_0) \quad (5.20)$$

em que

$$P = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \quad (5.21)$$

é adotado como estimativa de p . Calcula-se

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}_1 - \hat{p}_2}} \quad (5.22)$$

e aceita-se H_0 se $|Z| \leq Z_{\alpha/2}$

Exemplo:

Dois grupos, A e B, são formados, cada um por 100 pessoas que têm a mesma enfermidade. É ministrado um soro ao grupo A, mas não ao B (denominado grupo de controle); a todos os outros respeitos, os dois grupos são tratados de modo idêntico. Determinou-se que 75 e 65 pessoas dos grupos A e B, respectivamente, curaram-se da enfermidade. Testar a hipótese de que o soro auxilia a cura da enfermidade, adotado o nível de significância de 0,01.

Solução:

Denominando de p_A e p_B as proporções populacionais curadas mediante o uso do

soro e sem o uso do soro, respectivamente, queremos testar a hipótese

$$H_0 : p_A = p_B \quad \text{o soro não é eficaz}$$

$$H_1 : p_A > p_B \quad \text{o soro é eficaz}$$

Temos que $\hat{p}_A = 0,75$ $\hat{p}_B = 0,65$, e

$$P = \frac{75 + 65}{200} = 0,70 \quad (5.23)$$

$$\sigma_{\hat{p}_1 - \hat{p}_2}^2 = (0,7)(0,3)\left(\frac{1}{100} + \frac{1}{100}\right) = 0,0042 \quad (5.24)$$

A estatística a ser calculada é:

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sigma_{\hat{p}_A - \hat{p}_B}} = \frac{0,75 - 0,65}{\sqrt{0,0042}} = 1,543 \quad (5.25)$$

Devemos aceitar H_0 se $Z < Z_{0,01} = 2,33$. Portanto, aceitamos H_0 , e concluímos que os resultados são devidos ao acaso, e que o soro não é eficaz.

5.10 Teste χ^2 da independência

Uma tabela na qual cada observação é classificada em dois ou mais modos é denominada **tabela de contingência**.

Exemplo

Suponha que 200 estudantes sejam selecionados aleatoriamente em uma universidade e que cada estudante seja classificado de acordo com a sua área de estudo, e com sua preferência entre dois candidatos para uma próxima eleição (ver Tabela 5.2).

TABELA 5.2 – Candidatos selecionados

Área de Estudo	Candidato			Totais
	A	B	Indeciso	
Engenharia	24	23	12	59
Humanas	24	14	10	48
Artes	17	8	13	38
Administração	27	19	9	55
Totais	92	64	44	200

Quer-se tentar a hipótese de que as diferentes classificações são independentes, isto é, que a preferência a uma certo candidato é independente da área de estudo (i.e. a probabilidade de estar na área de estudo i e preferir o candidato j é igual a probabilidade de estar em i vezes a probabilidade de preferir j).

Em geral, conderamos uma tabela de contingência contendo L linhas e C colunas. Para $i = 1, 2, 3, \dots, L$ e $j = 1, 2, 3, \dots, C$, seja P_{ij} a probabilidade de que um individuo selecionado aleatoriamente de uma dada população seja classificado na i -ésima linha e na j -ésima coluna. Além disso, seja $P_{i\bullet}$ a probabilidade marginal que o individuo seja classificado na i -ésima linha, e seja $P_{\bullet j}$ a probabilidade marginal que o individuo seja classificado na j -ésima coluna.

$$P_i = \sum_{j=1}^C P_{ij} \quad e \quad P_j = \sum_{i=1}^L P_{ij} \quad (5.26)$$

Observe que:

$$\sum_{i=1}^L \sum_{j=1}^C P_{ij} = \sum_{i=1}^L P_{i\bullet} = \sum_{j=1}^C P_{\bullet j} = 1 \quad (5.27)$$

Suponha agora que uma amostra de n individuos seja retirada da população. Para $i = 1, 2, 3, \dots, L$, e $j = 1, 2, 3, \dots, C$, seja N_{ij} o número de individuos classificados na i -ésima linha e j -ésima coluna. Além disso, seja $N_{i\bullet}$ o número de individuos classificados na i -ésima linha e $N_{\bullet j}$ o número total de individuos classificados na j -ésima coluna.

$$N_{i\bullet} = \sum_{j=1}^C N_{ij} \quad e \quad N_{\bullet j} = \sum_{i=1}^L N_{ij} \quad (5.28)$$

Observe que

$$\sum_{i=1}^L \sum_{j=1}^C N_{ij} = \sum_{i=1}^L N_{i\bullet} = \sum_{j=1}^C N_{\bullet j} = n \quad (5.29)$$

Com base nessas observações, as seguintes hipóteses serão testadas:

$$\begin{aligned} H_0 & : P_{ij} = P_{i\bullet}P_{\bullet j} \\ H_1 & : \text{A hipótese } H_0 \text{ não é verdadeira.} \end{aligned}$$

O teste χ^2 pode ser usado para testar essa hipótese. Cada indivíduo na população deve pertencer a uma das $L \cdot C$ células da tabela de contingência. Sob a hipótese H_0 , as probabilidades desconhecidas P_{ij} dessas células foram expressas em função dos parâmetros desconhecidos $P_{i\bullet}$ e $P_{\bullet j}$. Como $\sum_{i=1}^L P_{i\bullet} = 1$ e $\sum_{j=1}^C P_{\bullet j} = 1$, o número de parâmetros desconhecidos a serem estimados quando H_0 é verdadeiro é $(L - 1) + (C - 1)$, ou $L + C - 2$.

Para $i = 1, 2, 3, \dots, L$, e $j = 1, 2, 3, \dots, C$, seja \hat{E}_{ij} o estimador quando H_0 é verdadeira, do número esperado de observações classificadas na i -ésima linha e j -ésima coluna. Portanto, a estatística Q definida no teste χ^2 de aderência é dada por

$$Q = \sum_{i=1}^L \sum_{j=1}^C \frac{(N_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \quad (5.30)$$

Como a tabela de contingência tem $L \cdot C$ células, e como $L + C - 2$ parâmetros terão que ser estimados quando H_0 for verdadeiro, segue-se que quando H_0 for verdadeiro e $n \rightarrow \infty$, a distribuição de Q converge para uma χ^2 com $L \cdot C - 1 - (L + C - 2) = (L - 1)(C - 1)$ graus de liberdade.

Consideraremos agora o estimador \hat{E}_{ij} . O número esperado de observações na i -ésima linha e j -ésima coluna é nP_{ij} . Quando H_0 é verdadeiro $P_{ij} = P_{i\bullet}P_{\bullet j}$. Portanto, se $\hat{P}_{i\bullet}$ e $\hat{P}_{\bullet j}$ são estimadores de $P_{i\bullet}$ e $P_{\bullet j}$, segue-se que $\hat{E}_{ij} = n\hat{P}_{i\bullet}\hat{P}_{\bullet j}$. Como $P_{i\bullet}$ é a probabilidade que uma observação seja classificada na i -ésima linha, $\hat{P}_{i\bullet}$ é a proporção de observações na amostra que são classificadas na i -ésima linha, isto é, $\hat{P}_{i\bullet} = \frac{N_{i\bullet}}{n}$. Da mesma maneira, $\hat{P}_{\bullet j} = \frac{N_{\bullet j}}{n}$. Portanto,

$$\hat{E}_{ij} = n \left(\frac{N_{i\bullet}}{n} \right) \left(\frac{N_{\bullet j}}{n} \right) = \frac{N_{i\bullet}N_{\bullet j}}{n} \quad (5.31)$$

A hipótese nula será rejeitada quando $Q > c_\alpha$, onde c_α é obtido na tabela χ^2 com $(L - 1)(C - 1)$ graus de liberdade.

Para o exemplo dado, $N_{1\bullet} = 59$; $N_{2\bullet} = 48$, $N_{3\bullet} = 38$; $N_{4\bullet} = 55$ e $N_{\bullet 1} = 92$, $N_{\bullet 2} = 64$, $N_{\bullet 3} = 44$.

Como $n = 200$, os valores (ver Tabela 5.3) \hat{E}_{ij} são dados por

TABELA 5.3 – Valores da \hat{E}_{ij}

27,14	18,88	12,98
22,08	15,36	10,56
17,48	12,16	8,36
25,30	16,60	12,10

O valor de Q será $Q = 6,68$. Como $L = 4$ e $C = 3$, o número de graus de liberdade é $(L - 1)(C - 1) = 6$.

Para $\alpha = 0,005$, tem-se que $c_\alpha = 12,59$, e portanto não há nenhuma evidência que H_0 não seja verdadeira.

CAPÍTULO 6

ANÁLISE DE VARIÂNCIA

6.1 Introdução

A análise de variância é um método de se dividir a variação total dos dados em componentes significativos que medem diferentes fontes de variação. Assim, considerando que nosso interesse esteja voltado para testar se diferentes variedades de trigo produzem, em média quantidades iguais, isto é, se:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$$H_1 : \text{pelo menos duas médias são diferentes}$$

obtem-se duas componentes: uma devida ao erro experimental (incontrolável) e outra medindo a variação devida ao erro experimental mais qualquer variação devida às diferentes variedades de trigo. Se a hipótese nula for verdadeira, as k espécies de trigo produzem igualmente, em média, então ambos componentes fornecem estimativas independentes do erro experimental.

A análise de variância é um método para testar a hipótese H_0 acima, por meio da análise das variâncias das diversas amostras. Este método estende o método visto no capítulo 5, onde a comparação envolvia somente duas médias.

É importante ressaltar que num experimento desta natureza é importante se ter controladas outras possíveis fontes de variação: tipos diferentes de solo onde se plantarão as variedades conduzirão a uma estimativa tendenciosa do erro experimental e conseqüentemente aumentará a probabilidade de se cometer um erro do tipo II. Desta forma, uma variedade A, de pior qualidade que as outras variedades sendo testadas, pode, se plantada em uma área de melhor composição, fornecer, em média, valores que não a tornem distinta das outras variedades. Assim, a hipótese nula, de igualdade de todas as médias das k variedades, pode ser aceita mesmo não sendo verdadeira, isto é, mesmo que uma das variedades sendo testada seja inferior que as outras. Aqui, um fator não controlado (composição do solo) mascara o verdadeiro resultado que seria obtido caso houvesse um controle mais rigoroso. Portanto, em uma análise de variância estamos supondo que estamos variando somente uma (ou algumas) variável(eis) de interesse, e que todos os outros fatores permaneçam constantes.

Vamos estudar experimentos de um e de dois fatores (poderiam até ser mais do que dois), em função do número de variáveis envolvidas no problema. Se além das diferentes variedades de trigo sendo testadas estivéssemos também interessados em investigar o efeito de diferentes tipos de fertilizantes na produtividade, então deixaríamos de ter um experimento de um só fator e passaríamos a estar envolvidos com um experimento de dois fatores.

Em resumo, a classificação de observações com base em um único critério, tal como variedade de trigo, é chamada *de análise de variância de um só fator*, enquanto que se as observações forem classificadas de acordo com dois critérios, tal como variedades de trigo e tipo de fertilizante, temos a chamada *análise de variância de dois fatores*.

6.2 Análise de Variância de Um Fator

Como dito anteriormente, desejamos testar a hipótese:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$$H_1 : \text{pelo menos duas médias são diferentes}$$

Para tal, amostras aleatórias de tamanho n são selecionadas de cada uma das k populações. Supõe-se que as k populações *são independentes e normalmente distribuídas com médias $\mu_1, \mu_2, \mu_3, \dots, \mu_k$ e com mesma variância σ^2* .

Seja y_{ij} a j -ésima observação na i -ésima população ($i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n$). Estes dados estão representados na tabela abaixo.

Tabela de arranjos dos dados em um experimento de um fator

	População						
	1	2	3	...	i	...	k
	y_{11}	y_{21}	y_{31}	...	y_{i1}	...	y_{k1}
	y_{12}	y_{22}	y_{32}	...	y_{i2}	...	y_{k2}
	y_{13}	y_{23}	y_{33}	...	y_{i3}	...	y_{k3}

	y_{1n}	y_{2n}	y_{3n}	...	y_{in}	...	y_{kn}
Total	$T_1.$	$T_2.$	$T_3.$...	$T_i.$...	$T_k.$
Média	$\bar{y}_1.$	$\bar{y}_2.$	$\bar{y}_3.$...	$\bar{y}_i.$...	$\bar{y}_k.$

onde

$$T_{i\cdot} = \sum_{j=1}^n y_{ij}$$

$$\bar{y}_{i\cdot} = \frac{\sum_{j=1}^n y_{ij}}{n}.$$

$T_{i\cdot}$ e $\bar{y}_{i\cdot}$ representam, respectivamente, o total e a média, de todas as observações na amostra da i -ésima população. Além disto, podemos definir o total e a média das $n \cdot k$ observações por:

$$T_{\cdot\cdot} = \sum_{i=1}^k \sum_{j=1}^n y_{ij} = \sum_{i=1}^k T_{i\cdot}$$

$$\bar{y}_{\cdot\cdot} = \frac{\sum_{i=1}^k \sum_{j=1}^n y_{ij}}{nk} = \frac{\sum_{i=1}^k \bar{y}_{i\cdot}}{k}.$$

Cada observação y_{ij} pode ser escrita da seguinte forma:

$$y_{ij} = \mu_i + \epsilon_{ij} \tag{6.1}$$

onde ϵ_{ij} representa o desvio da observação y_{ij} da média populacional correspondente μ_i (ou seja, ϵ_{ij} é o efeito aleatório, não controlado, da observação j da população i). Além disto, supomos que os ϵ_{ij} são v.a. independentes, de média zero e mesma variância σ^2 . Note que estamos supondo que as variâncias residuais das diferentes populações são iguais, isto é:

$$Var(\epsilon_{1j}) = Var(\epsilon_{2j}) = \dots = Var(\epsilon_{kj}) = \sigma^2$$

Esta propriedade é denominada de **homocedasticidade**. As figuras a seguir ilustram isto para o caso em que as médias são diferentes (Fig. 6.1) e para o caso em que as médias são iguais (Fig. 6.2). Note que $E(y_{ij}) = \mu_i$, e que com esta suposição de homocedasticidade, estamos também supondo que as observações possuem variância iguais, já que:

$$Var(y_{ij}) = Var(\mu_i + \epsilon_{ij}) = Var(\epsilon_{ij}) = \sigma^2$$

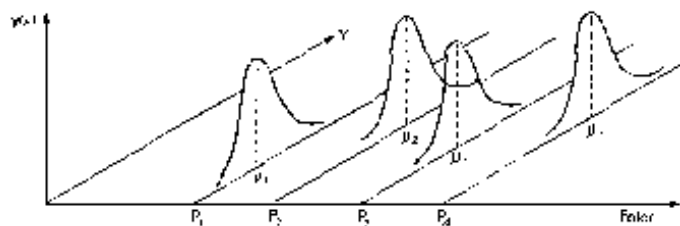


Fig. 6.1 – Distribuições normais com mesma variância (σ^2) para todas as populações

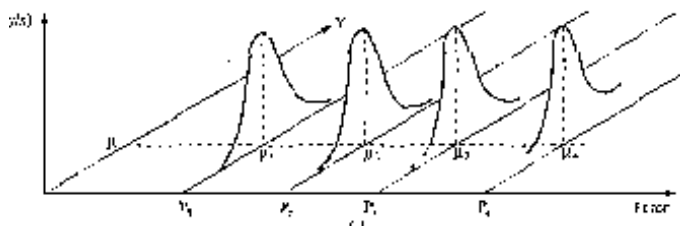


Fig. 6.2 – Distribuições normais com mesma média (μ) para todas as populações

Uma forma alternativa para escrever a equação (6.1) é obtida substituindo-se μ_i por $\mu + \alpha_i$, onde μ é definida como sendo a média de todos os μ_i .

Assim, podemos escrever y_{ij} como:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (6.2)$$

sujeito à restrição de que $\sum_{i=1}^k \alpha_i = 0$. Refere-se à α_i como sendo o *efeito da i-ésima população*.

Uma forma alternativa, então, de se expressar a hipótese nula e a hipótese alternativa, faz uso dos α_i . Caso não haja efeito da i -ésima população ($i=1,2, \dots, k$), as médias das mesmas são iguais (um tratamento não é superior ao outro), e as hipóteses ficam expressas como:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_k = 0$$

$$H_1 : \text{pelo menos um dos } \alpha_i \text{ não é zero.}$$

Nosso teste se baseará na comparação de duas estimativas independentes da variância populacional comum, σ^2 . Essas estimativas serão obtidas dividindo-se a variabilidade total dos dados em duas componentes.

A variância de todas as observações agrupadas em uma única amostra de tamanho

$n.k$ é dado pela fórmula:

$$s^2 = \sum_{i=1}^k \sum_{j=1}^n \frac{(y_{ij} - \bar{y}_{..})^2}{nk - 1}$$

O numerador de s^2 é chamado de soma total dos quadrados (*total sum of squares*) e mede a variabilidade total dos dados. O importante é demonstrar que esta soma total de quadrados pode ser particionada em duas componentes, por meio da seguinte identidade:

Teorema:

$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

É conveniente identificar os termos da soma total de quadrados pela seguinte notação:

$$\text{SST} = \text{soma total de quadrados} = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

$$\text{SSC} = \text{soma dos quadrados para as médias das colunas} = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\text{SSE} = \text{soma dos quadrados dos erros} = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2.$$

Portanto:

$$\text{SST} = \text{SSC} + \text{SSE}$$

Muitos autores referem-se à soma dos quadrados para as colunas como soma dos quadrados dos tratamentos. Esta terminologia é derivada do fato que as k diferentes populações são frequentemente classificadas de acordo com diferentes tratamentos.

Uma estimativa de σ^2 , baseada em $(k - 1)$ graus de liberdade é dada por:

$$s_1^2 = \frac{\text{SSC}}{k - 1}$$

Se H_0 for verdadeira, será demonstrado que s_1^2 é um estimador não tendencioso de σ^2 . Entretanto, se H_1 for verdadeira, então s_1^2 superestima σ^2 .

Uma segunda estimativa de σ^2 , baseada em $k(n-1)$ graus de liberdade é dada por:

$$s_2^2 = \frac{\text{SSE}}{k(n-1)}$$

Mostraremos que s_2^2 é sempre um estimador não tendencioso, independente de H_0 ser verdadeira ou não.

Vimos também que a variância dos dados agrupados, com $nk-1$ graus de liberdade, é:

$$s^2 = \frac{\text{SST}}{nk-1}$$

É interessante notar que a identidade da soma dos quadrados não somente particiona a variabilidade total dos dados, mas também o número de graus de liberdade:

$$nk-1 = (k-1) + k(n-1)$$

Vamos encontrar agora, o valor esperado das variações expressas pelo Teorema. Tomando a **variação dentro das colunas** (SSE), temos:

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^k (n-1) \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{n-1}$$

Mas $\frac{\sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{n-1}$ representa a variância amostral do tratamento i (s_i^2), que é um estimador não tendencioso de σ_i^2 . Assim,

$$SSE = \sum_{i=1}^k (n-1) s_i^2$$

e o valor esperado será

$$E(SSE) = E\left(\sum_{i=1}^k (n-1) s_i^2\right) = (n-1) \sum_{i=1}^k E(s_i^2) = (n-1) \sum_{i=1}^k \sigma_i^2$$

Entretanto, como se supõe que as variâncias dos k tratamentos são iguais, então:

$$E(SSE) = (n-1)k\sigma^2$$

Ou seja,

$$E\left(\frac{SSE}{(n-1)k}\right) = E(s_2^2) = \sigma^2$$

Assim sendo, $s_2^2 = \frac{SSE}{(n-1)k}$ é um estimador não tendencioso de σ^2 .

Vamos analisar agora o valor esperado da *variabilidade entre tratamentos* (SSC)

$$\begin{aligned} SSC &= n \sum_{i=1}^k (\bar{y}_i - \bar{y}_{..})^2 = n \sum_{i=1}^k \bar{y}_i^2 - 2n\bar{y}_{..} \sum_{i=1}^k \bar{y}_i + n \sum_{i=1}^k \bar{y}_{..}^2 \\ &= n \sum_{i=1}^k \bar{y}_i^2 - 2n\bar{y}_{..} \sum_{i=1}^k \bar{y}_i + nk\bar{y}_{..}^2 = n \sum_{i=1}^k \bar{y}_i^2 - 2n\bar{y}_{..}k\bar{y}_{..} + nk\bar{y}_{..}^2 \\ &= n \sum_{i=1}^k \bar{y}_i^2 - nk\bar{y}_{..}^2 \end{aligned}$$

Portanto,

$$E(SSC) = E\left(n \sum_{i=1}^k \bar{y}_i^2 - nk\bar{y}_{..}^2\right) = n \sum_{i=1}^k E(\bar{y}_i^2) - nkE(\bar{y}_{..}^2)$$

Como tem-se que, para qualquer v.a. X ,

$$Var(X) = E(X^2) - E^2(X) \implies E(X^2) = Var(X) + E^2(X)$$

podemos escrever que:

$$E(SSC) = n \sum_{i=1}^k [Var(\bar{y}_i) + E^2(\bar{y}_i)] - nk [Var(\bar{y}_{..}) + E^2(\bar{y}_{..})]$$

Sabemos também, da teoria de amostragem que

$$Var(\bar{y}_i) = \frac{\sigma_i^2}{n} = \frac{\sigma^2}{n} \quad \text{pois } \sigma_i^2 = \sigma^2 \text{ para } i = 1, 2, \dots, k.$$

Temos também que:

$$E(\bar{y}_i) = \mu_i = \mu + \alpha_i$$

$$\text{Var}(\bar{y}_{..}) = \frac{\sigma^2}{nk}$$

$$E(\bar{y}_{..}) = \mu$$

Então

$$\begin{aligned} E(SSC) &= n \sum_{i=1}^k \left[\frac{\sigma^2}{n} + (\mu + \alpha_i)^2 \right] - nk \left[\frac{\sigma^2}{nk} + \mu^2 \right] \\ &= n \sum_{i=1}^k \left[\frac{\sigma^2}{n} + \mu^2 + 2\mu\alpha_i + \alpha_i^2 \right] - \sigma^2 - nk\mu^2 \\ &= k\sigma^2 + nk\mu^2 + 2n\mu \sum_{i=1}^k \alpha_i + n \sum_{i=1}^k \alpha_i^2 - \sigma^2 - nk\mu^2 \\ &= \sigma^2(k-1) + n \sum_{i=1}^k \alpha_i^2 \end{aligned}$$

Assim,

$$E\left(\frac{SSC}{k-1}\right) = E(s_1^2) = \sigma^2 + n(k-1) \sum_{i=1}^k \alpha_i^2$$

Ou seja, temos que $s_1^2 = \frac{SSC}{k-1}$ será um estimador tendencioso de σ^2 , superestimando σ^2 , a não ser que a hipótese nula seja verdadeira, isto é, se todos os $\alpha_i = 0$.

Resumindo, s_1^2 é um estimador tendencioso de σ^2 se H_0 não for verdadeira, e s_2^2 é um estimador não tendencioso de σ^2 independentemente de H_0 ser ou não verdadeira.

Se H_0 for verdadeira, a razão

$$f = \frac{s_1^2}{s_2^2}$$

possui uma distribuição F com $(k-1)$ e $k(n-1)$ graus de liberdade. Uma vez que s_1^2 superestima quando H_0 é falsa, teremos um teste unilateral com a região crítica inteiramente na cauda direita da distribuição. A hipótese nula é rejeitada ao nível de significância α quando

$$f > F_{\alpha, (k-1), k(n-1)}$$

Em geral, calcula-se SST e SSC primeiro e daí, fazendo uso da identidade da soma dos quadrados obtem-se $SSE = SST - SSC$.

As fórmulas definidas anteriormente para o cômputo de SST e SSC não são as mais simples para se utilizar. Fórmulas alternativas preferenciais à elas são:

$$SST = \sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{T_{..}^2}{nk}$$

$$SSC = \frac{\sum_{i=1}^k T_i^2}{n} - \frac{T_{..}^2}{nk}$$

Os cálculos para um problema de análise de variância de um fator são geralmente sumarizados em forma de uma tabela, chamada Tabela ANOVA, como mostrado abaixo:

TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Tratamento	SSC	$k - 1$	$s_1^2 = \frac{SSC}{k-1}$	$f = \frac{s_1^2}{s_2^2}$
Resíduo	SSE	$k(n - 1)$	$s_2^2 = \frac{SSE}{k(n-1)}$	
Total	SST	$nk - 1$		

Exemplo:

Os dados da tabela abaixo representam 5 amostras, cada uma de tamanho $n=5$, tiradas de distribuições normais independentes com médias $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ e variância comum σ^2 . Testar a hipótese de que as médias são iguais, ao nível de significância de 5%.

	População					
	A	B	C	D	E	
	5	9	3	2	7	
	4	7	5	3	6	
	8	8	2	4	9	
	6	6	3	1	4	
	3	9	7	4	7	
Total	26	39	20	14	33	132
Média	5,2	7,8	4,0	2,8	6,6	5,28

Solução:

Deseja-se testar a hipótese:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1 : \text{pelo menos duas médias são diferentes}$$

Tem-se que:

$$\begin{aligned}
 SST &= \sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{T_{..}^2}{nk} \\
 &= 5^2 + 4^2 + \dots + 4^2 + 7^2 - \frac{132^2}{25} \\
 &= 834 - 696,96 = 137,040
 \end{aligned}
 \tag{6.3}$$

$$\begin{aligned}
 SSC &= \frac{\sum_{i=1}^k T_i^2}{n} - \frac{T_{..}^2}{nk} \\
 &= \frac{26^2 + 39^2 + 20^2 + 14^2 + 33^2}{5} - \frac{132^2}{25} \\
 &= 776,400 - 696,960 = 79,440
 \end{aligned}
 \tag{6.4}$$

$$SSE = SST - SSC = 137,040 - 79,440 = 57,600$$

A Tabela ANOVA é dada por:

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Tratamento	79,440	4	19,860	
Resíduo	57,600	20	3,880	6,90
Total	137,040	24		

A região crítica é dada por $f > F_{0,05,4,20} = 2,87$

Portanto, rejeita-se H_0 e conclui-se que as amostras provêm de diferentes populações.

Observação:

Quando os tamanhos de amostras são diferentes, isto é, quando se tem k amostras aleatórias com tamanhos n_1, n_2, \dots, n_k , respectivamente e $N = \sum_{i=1}^k n_i$, as fórmulas computacionais para SST e SSC são dadas por:

$$\begin{aligned}
 SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \frac{T_{..}^2}{N} \\
 SSC &= \sum_{i=1}^k \frac{T_i^2}{n_i} - \frac{T_{..}^2}{N}
 \end{aligned}$$

O valor de SSE é encontrado por subtração. Os graus de liberdade são particionados da mesma maneira: $N - 1$ g.l. para SST, $k - 1$ para SSC e $N - k$ para SSE.

6.3 Teste para Igualdade de Várias Variâncias

Como a análise de variância supõe que as variâncias das populações são iguais (suposição de homocedasticidade) pode-se, antes de efetuar o teste de igualdade de médias, testar a hipótese:

$$\begin{aligned} H_0 & : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2 \\ H_1 & : \text{as variâncias não são todas iguais} \end{aligned}$$

Um dos testes mais utilizados é o teste de Bartlett, baseado em uma estatística cuja distribuição amostral é aproximadamente χ^2 quando as k amostras aleatórias são retiradas de populações nomais independentes.

Primeiro calcula-se as k variâncias amostrais, $s_1^2, s_2^2, s_3^2, \dots, s_k^2$ das amostras de tamanho $n_1, n_2, n_3, \dots, n_k$, com $\sum_{i=1}^k n_i = N$. Depois disto combinam-se as variâncias amostrais para fornecer a estimativa:

$$s_p^2 = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{N - k}$$

Calcula-se também

$$b = 2,3026 \frac{q}{h}$$

onde

$$q = (N - k) \log s_p^2 - \sum_{i=1}^k (n_i - 1) \log s_i^2$$

$$h = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N - k} \right]$$

b é um valor da variável aleatória B que possui uma distribuição χ^2 com $k - 1$ graus de liberdade. A quantidade q será grande quando as variâncias amostrais diferem significativamente, e será igual a zero quando todas as variâncias amostrais forem iguais. Assim, rejeita-se H_0 ao nível de significância α quando

$$b > \chi_{\alpha, k-1}^2$$

Exemplo:

Use o teste de Bartlett para testar a hipótese de que as variâncias das três populações abaixo são iguais:

	Amostra			
	A	B	C	
	4	5	8	
	7	1	6	
	6	3	8	
	6	5	9	
		3	5	
		4		
Total	23	21	36	80

Solução:

Deseja-se testar a hipótese:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$$H_1 : \text{as variâncias não são iguais}$$

Tem-se que: $n_1 = 4, n_2 = 6, n_3 = 5, N = 15, k = 3$.

$$s_1^2 = 1,583; \quad s_2^2 = 2,300; \quad s_3^2 = 2,700$$

Assim,

$$s_p^2 = \frac{3(1,583) + 5(2,300) + 4(2,700)}{12} = 2,254. \tag{6.5}$$

$$\begin{aligned} q &= 12 \log 2,254 - (3 \log 1,583 + 5 \log 2,300 + 4 \log 2,700) \\ &= 12(0,3530) - (3(0,1995) + 5(0,3617) + 4(0,4314)) = 0,1034 \end{aligned}$$

$$h = 1 + \frac{1}{6} \left(\frac{1}{3} + \frac{1}{5} + \frac{1}{4} - \frac{1}{12} \right) = 1,1167 \tag{6.6}$$

$$b = \frac{(2,3026)(0,1034)}{1,1167} = 0,213$$

Ao nível de significância de 0,05, região crítica é dada por $B > 5,991$.

Portanto, aceita-se H_0 e conclui-se que as variâncias das três populações são iguais.

6.4 Análise de Variância de Dois Fatores

A análise de variância de dois fatores trata de problemas onde se investiga o efeito de dois fatores. Analogamente à análise de variância de um fator, fazemos as suposições de independência, normalidade e homocedasticidade das populações.

Sob a suposição de aditividade dos efeitos dos fatores, isto é, supondo que cada média μ_{ij} pode ser obtida pela adição dos respectivos efeitos dos fatores A e B à média global μ , o modelo para a análise de dois fatores se torna:

$$\mu_{ij} = \mu + \alpha_i + \beta_j \quad (6.7)$$

onde i se refere ao nível do fator A ($i = 1, \dots, r$), j se refere ao nível do fator B ($j = 1, \dots, c$), sujeito às restrições de que $\sum_{i=1}^r \alpha_i = 0$ e $\sum_{j=1}^c \beta_j = 0$. Refere-se à α_i como sendo o *efeito do i -ésimo nível do fator A*, e à β_j como sendo o *efeito do j -ésimo nível do fator B*.

Quando as médias de todos os tratamentos pode ser decomposta na forma da expressão (6.7), dizemos que *os fatores não interagem*, ou que *não existe interação entre os fatores*. O caso em que se supõe interação entre os fatores é visto na próxima seção.

Assim como no caso da análise de variância de um fator, onde a soma total de quadrados foi particionada em componentes, pode-se também efetuar este particionamento para o caso da análise de variância de dois fatores, obtendo-se:

$$\sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_{..})^2 = c \sum_{i=1}^r (\bar{y}_{i.} - \bar{y}_{..})^2 + r \sum_{j=1}^c (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

A demonstração disto parte do princípio de que o termos à esquerda da igualdade acima pode ser escrito da seguinte forma:

$$\sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^r \sum_{j=1}^c [(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})]^2$$

Pela expansão desta expressão chega-se à identidade da soma de quadrados. A identidade de quadrados pode ser representada simbolicamente por:

$$\text{SST} = \text{SSR} + \text{SSC} + \text{SSE}$$

onde

$$SST = \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_{..})^2 = \text{soma total de quadrados}$$

$$SSR = c \sum_{i=1}^r (\bar{y}_{i.} - \bar{y}_{..})^2 = \text{soma dos quadrados para as médias das linhas (rows)}$$

$$SSC = r \sum_{j=1}^c (\bar{y}_{.j} - \bar{y}_{..})^2 = \text{soma dos quadrados para as médias das colunas (columns)}$$

$$SSE = \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \text{soma dos quadrados dos erros.}$$

Um estimador de σ^2 , baseado em $r - 1$ graus de liberdade é dado por:

$$s_1^2 = \frac{SSR}{r - 1}$$

Se não houver efeito do fator A (efeitos das linhas), isto é, se

$$\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$$

então s_1^2 é um estimador não tendencioso de σ^2 . Entretanto se os efeitos das linhas não são todos iguais a zero, SSR terá um valor numérico maior e s_1^2 superestima σ^2 .

Um segundo estimador de σ^2 , baseado em $c - 1$ graus de liberdade é dado por:

$$s_2^2 = \frac{SSC}{c - 1}$$

Este estimador é não tendencioso se os efeitos das colunas (fator B) forem iguais a zero, isto é, se:

$$\beta_1 = \beta_2 = \dots = \beta_c = 0$$

Caso contrário, SSC também terá um valor numérico maior e s_2^2 superestima σ^2 .

Um terceiro estimador de σ^2 , baseado em $(r - 1)(c - 1)$ graus de liberdade e independente de s_1^2 e s_2^2 é dado por:

$$s_3^2 = \frac{SSE}{(r - 1)(c - 1)}$$

Este estimador é sempre não tendencioso, independente de haver ou não efeito das linhas ou colunas.

Para testar a hipótese nula de que os efeitos das linhas são todos zeros, isto é:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$$

$$H_1 : \text{pelo menos um dos } \alpha_i \text{ não é zero,}$$

calculamos a estatística

$$f_1 = \frac{s_1^2}{s_3^2},$$

a qual é uma v.a. possuindo uma distribuição F com $(r - 1)$ e $(r - 1)(c - 1)$ g.l., quando H_0 for verdadeira. A hipótese nula é rejeitada ao nível de significância α se

$$f_1 > F_{\alpha; (r-1); (r-1)(c-1)}.$$

Similarmente, para testar a hipótese nula de que os efeitos das colunas são todos zeros, isto é:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_r = 0$$

$$H_1 : \text{pelo menos um dos } \beta_j \text{ não é zero,}$$

calculamos a estatística

$$f_2 = \frac{s_2^2}{s_3^2},$$

a qual é uma v.a. possuindo uma distribuição F com $(c - 1)$ e $(r - 1)(c - 1)$ g.l., quando H_0 for verdadeira. A hipótese nula é rejeitada ao nível de significância α se

$$f_2 > F_{\alpha; (c-1); (r-1)(c-1)}.$$

Para efetuar os cálculos de uma análise de variância, em geral calculamos SST, SSR e SSC, e obtemos SSE por subtração:

$$\text{SSE} = \text{SST} - \text{SSR} - \text{SSC}$$

É interessante observar que os graus de liberdade também são particionados, fornecendo:

$$(r - 1)(c - 1) = (rc - 1) - (r - 1) - (c - 1).$$

Fórmulas alternativas para as quantidades SST, SSR e SSC são:

$$SST = \sum_{i=1}^r \sum_{j=1}^c y_{ij}^2 - \frac{T_{..}^2}{rc}$$

$$SSR = \frac{\sum_{i=1}^r T_{i.}^2}{c} - \frac{T_{..}^2}{rc}$$

$$SSC = \frac{\sum_{j=1}^c T_{.j}^2}{r} - \frac{T_{..}^2}{rc}$$

Os cálculos para um problema de análise de variância de dois fatores, com apenas uma observação por cela são geralmente sumarizados em forma de uma tabela, chamada Tabela ANOVA, como mostrado abaixo:

TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Devido às linhas	SSR	$r - 1$	$s_1^2 = \frac{SSR}{r-1}$	$f_1 = \frac{s_1^2}{s_3^2}$
Devido às colunas	SSC	$c - 1$	$s_2^2 = \frac{SSC}{c-1}$	$f_2 = \frac{s_2^2}{s_3^2}$
Erro	SSE	$(r - 1)(c - 1)$	$s_3^2 = \frac{SSE}{(r-1)(c-1)}$	
Total	SST	$rc - 1$		

Exemplo:

A tabela abaixo apresenta os resultados da safra média de trigo, para três variedades de trigo e quatro tipos de fertilizantes. Teste a hipótese H_0' de que não há diferença na safra média de trigo quando diferentes tipos de fertilizantes são utilizados. Teste também a hipótese H_0'' de que não há diferença na safra média de trigo quando diferentes variedades de trigo são utilizadas.

Fertilizante	Variedades de Trigo			Total
	V_1	V_2	V_3	
F_1	64	72	74	210
F_2	55	57	47	159
F_3	59	66	58	183
F_4	58	57	53	168
Total	236	252	232	720

Solução: Para testar se os fertilizantes têm efeitos diferentes sobre a safra média de trigo, precisa-se testar a hipótese de que:

$$H'_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

$$H'_1 : \text{pelo menos um dos } \alpha_i \text{ não é zero,}$$

Para testar se as variedades de trigo têm efeitos diferentes sobre a safra média, precisa-se testar a hipótese de que:

$$H''_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H''_1 : \text{pelo menos um dos } \beta_j \text{ não é zero,}$$

Os cálculos da soma de quadrados são:

$$SST = \sum_{i=1}^r \sum_{j=1}^c y_{ij}^2 - \frac{T_{..}^2}{rc} = 64^2 + 55^2 + \dots + 53^2 - \frac{720^2}{12} = 662$$

$$SSR = \frac{\sum_{i=1}^r T_i^2}{c} - \frac{T_{..}^2}{rc} = \frac{210^2 + 159^2 + 183^2 + 168^2}{3} - \frac{720^2}{12} = 498$$

$$SSC = \frac{\sum_{j=1}^c T_j^2}{r} - \frac{T_{..}^2}{rc} = \frac{236^2 + 252^2 + 232^2}{4} - \frac{720^2}{12} = 56$$

$$SSE = 662 - 498 - 56 = 108$$

TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Fertilizantes	498	3	166	9,22
Var. de trigo	56	2	28	1,56
Erro	108	6	18	
Total	662	11		

As regiões críticas para H'_0 é $f_1 > 4,76$ e para H''_0 é $f_2 > 5,14$.

Portanto, concluímos que:

1. rejeitamos H'_0 e concluímos que há diferença na safra média de trigo quando diferentes tipos de fertilizantes são utilizados.
2. aceitamos H''_0 e concluímos que não há diferença na safra média das três variedades de trigo.

6.5 Análise de Variância de Dois Fatores - Várias observações por cela

Na seção anterior foi suposto que os efeitos das linhas e colunas eram aditivos. Isto é equivalente a dizer que

$$\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'}$$

ou

$$\mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'}$$

para qualquer valor i, i', j, j' . Isto é, a diferença entre as médias populacionais das colunas j e j' é a mesma para cada linha e a diferença entre as médias populacionais para as linhas colunas i e i' é a mesma para cada coluna. Referindo-se à tabela do exemplo anterior, isto implica que se a variedade V_2 produz em média 5 toneladas de trigo por acre a mais que a variedade V_1 quando o fertilizante F_1 é usado, então V_2 produzirá em média 5 toneladas a mais que V_1 se os fertilizantes F_2, F_3 ou F_4 forem usados. Da mesma forma, se V_1 produz em média 3 toneladas a mais por acre, quando o fertilizante F_4 é utilizado ao invés de F_2 , então V_2 ou V_3 produzirão em média 3 toneladas a mais por acre usando o fertilizante F_4 ao invés de F_2 . Isto é exemplificado nas Figuras 6.3 e 6.4 abaixo, quando notamos que as curvas são paralelas.

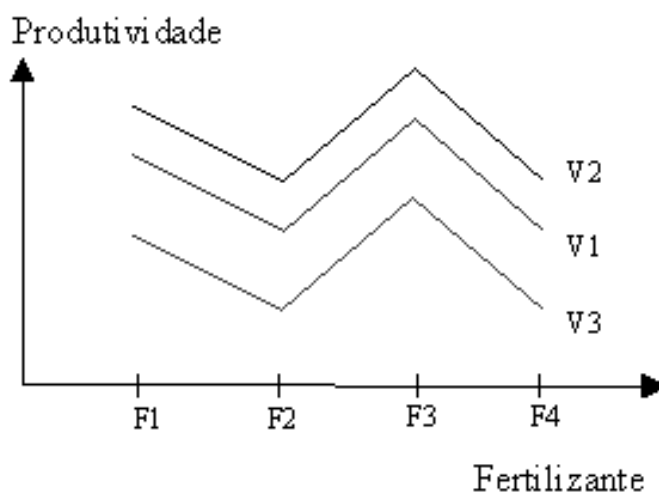


Fig. 6.3 – Efeitos de fertilizantes e variedades de trigo, sem interação

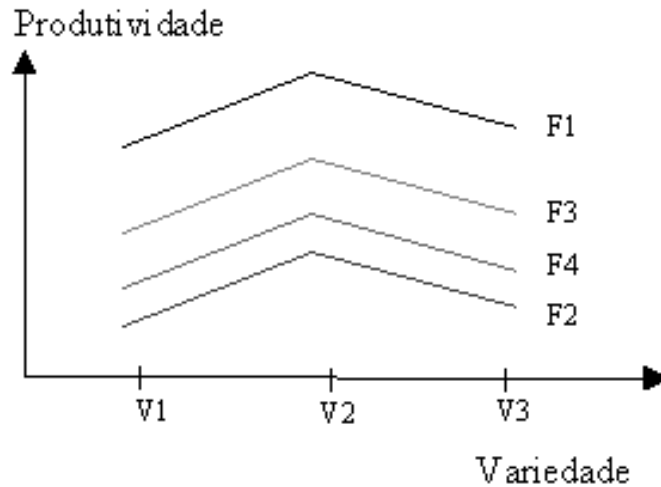


Fig. 6.4 – Efeitos de fertilizantes e variedades de trigo, sem interação

Em muitos experimentos, a hipótese de aditividade não é válida e a análise feita anteriormente nos levará a conclusões errôneas. Suponha, por exemplo, que a variedade V_2 produza em média, 5 toneladas a mais de trigo por acre do que a variedade V_1 quando F_1 , mas produza uma média de 2 toneladas por acre menos que V_1 quando F_2 é utilizado. As variedades de trigo e os tipos de fertilizantes são ditos a *interagir*.

Obviamente, quando analisamos os dados de uma tabela, os gráficos não são perfeitamente paralelos. Isto pode ser devido a uma interação real, ou pode ser simplesmente ao erro experimental. A análise do exemplo anterior partiu do pressuposto de que era simplesmente devido ao erro experimental.

Para testar as diferenças entre as médias das linhas e colunas quando a interação é um fator importante, consideramos a variação de medidas tomadas sob situações semelhantes, ou seja consideramos as replicações dos experimentos.

Para apresentar as fórmulas gerais para a análise de variância usando observações repetidas (ou várias observações por cela), vamos considerar o caso de n replicações. Como antes, consideraremos uma matriz regular consistindo de r linhas e c colunas. Assim teremos rc celas, mas agora cada cela possui n observações. Denotamos a k -ésima observação da i -ésima linha na j -ésima coluna por y_{ijk} . Isto é exemplificado na tabela abaixo:

Tabela dos dados em um experimento com dois fatores, e replicações

Linhas	Colunas				Total	Média
	1	2	...	c		
1	y_{111}	y_{121}	...	y_{1c1}	$T_{1..}$	$\bar{y}_{1..}$
	y_{112}	y_{122}	...	y_{1c2}		
	\vdots	\vdots	\vdots	\vdots		
	y_{11n}	y_{12n}	...	y_{1cn}		
		
		
		
r	y_{r11}	y_{r21}	...	y_{rc1}	$T_{r..}$	$\bar{y}_{r..}$
	y_{r12}	y_{r22}	...	y_{rc2}		
	\vdots	\vdots	\vdots	\vdots		
	y_{r1n}	y_{r2n}	...	y_{rcn}		
Total	$T_{.1.}$	$T_{.2.}$...	$T_{.r.}$	$T_{...}$	

As observações na cela (i, j) constituem uma amostra aleatória de tamanho n de uma população que se supõe ser normalmente distribuída com média μ_{ij} e variância σ^2 . Supõe-se também que todas as rc populações possuem a mesma variância σ^2 . A seguinte notação é utilizada:

$$T_{ij.} = \sum_{k=1}^n y_{ijk} \quad (\text{soma das observações na cela } (i, j))$$

$$T_{i..} = \sum_{j=1}^c \sum_{k=1}^n y_{ijk} \quad (\text{soma das observações na } i\text{-ésima linha})$$

$$T_{.j.} = \sum_{i=1}^r \sum_{k=1}^n y_{ijk} \quad (\text{soma das observações na } j\text{-ésima coluna})$$

$$T_{...} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijk} \quad (\text{soma de todas as } rcn)$$

$$\bar{y}_{ij\cdot} = \frac{\sum_{k=1}^n y_{ijk}}{n} = \frac{T_{ij\cdot}}{n} \quad (\text{m\u00e9dia das observa\u00e7\u00f5es na cela } (i, j))$$

$$\bar{y}_{i\cdot\cdot} = \frac{\sum_{j=1}^c \sum_{k=1}^n y_{ijk}}{cn} = \frac{T_{i\cdot\cdot}}{cn} \quad (\text{m\u00e9dia das observa\u00e7\u00f5es na } i\text{-\u00e9sima linha})$$

$$\bar{y}_{\cdot j\cdot} = \frac{\sum_{i=1}^r \sum_{k=1}^n y_{ijk}}{rn} = \frac{T_{\cdot j\cdot}}{rn} \quad (\text{m\u00e9dia das observa\u00e7\u00f5es na } j\text{-\u00e9sima coluna})$$

$$\bar{y}_{\cdot\cdot j} = \frac{\sum \sum}{rn} = \frac{T_{\cdot\cdot j}}{rn} \quad (\text{m\u00e9dia das observa\u00e7\u00f5es na } j\text{-\u00e9sima coluna})$$

$$\bar{y}_{\dots} = \frac{\sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijk}}{rcn} = \frac{T_{\dots}}{rcn} \quad (\text{m\u00e9dia de todas as } rcn)$$

Cada observa\u00e7\u00e3o y_{ijk} pode ser escrita da seguinte forma:

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad (6.8)$$

onde ϵ_{ijk} representa o desvio da observa\u00e7\u00e3o y_{ijk} da m\u00e9dia populacional correspondente μ_{ij} . Al\u00e9m disto, supomos que os ϵ_{ij} s\u00e3o v.a. independentes, de m\u00e9dia zero e mesma vari\u00e2ncia σ^2 .

Se denotarmos por γ_{ij} o efeito de intera\u00e7\u00e3o da i -\u00e9sima linha e j -\u00e9sima coluna, por α_i o efeito da i -\u00e9sima linha, por β_j o efeito da j -\u00e9sima coluna, e por μ a m\u00e9dia global, podemos escrever:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

e ent\u00e3o,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

no qual impomos as seguintes restri\u00e7\u00f5es:

$$\sum_{i=1}^r \alpha_i = 0; \quad \sum_{j=1}^c \beta_j = 0; \quad \sum_{i=1}^r \gamma_{ij} = 0; \quad \sum_{j=1}^c \gamma_{ij} = 0;$$

As três hipóteses a serem testadas são:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$$

$$H_1 : \text{pelo menos um dos } \alpha_i \text{ não é zero.}$$

$$H'_0 : \beta_1 = \beta_2 = \dots = \beta_c = 0$$

$$H'_1 : \text{pelo menos um dos } \beta_j \text{ não é zero.}$$

$$H''_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{rc} = 0$$

$$H''_1 : \text{pelo menos um dos } \gamma_{ij} \text{ não é zero.}$$

6.5.1 Identidade da Soma de Quadrados

Neste caso, a soma de quadrados pode ser particionada da seguinte maneira:

$$\begin{aligned} \sum_{i=1}^r \sum_{j=1}^c (y_{ijk} - \bar{y}_{...})^2 &= cn \sum_{i=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 + rn \sum_{j=1}^c (\bar{y}_{.j.} - \bar{y}_{...})^2 + \\ &+ n \sum_{i=1}^r \sum_{j=1}^c (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

Simbolicamente podemos escrever esta identidade como:

$$SST = SSR + SSC + SS(RC) + SSE$$

onde

$$SSR = cn \sum_{i=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 = \text{soma de quadrados das médias das linhas}$$

$$SSC = rn \sum_{j=1}^c (\bar{y}_{.j.} - \bar{y}_{...})^2 = \text{soma de quadrados das médias das colunas}$$

$$SS(RC) = n \sum_{i=1}^r \sum_{j=1}^c (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = \text{soma de quadrados para a interação de linhas e colunas}$$

$$SSE = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 = \text{soma de quadrados dos erros}$$

Os graus de liberdade são particionados segundo a relação:

$$rcn - 1 = (r - 1) + (c - 1) + (r - 1)(c - 1) + rc(n - 1)$$

Da mesma forma que antes, através da divisão das somas de quadrados pelos graus de liberdade correspondentes obtém-se quatro estimativas independentes de σ^2 , todas não tendenciosas desde que as hipóteses H_0 , H'_0 , e H''_0 sejam verdadeiras.

Estas estimativas são:

$$s_1^2 = \frac{SSR}{r - 1}; \quad s_2^2 = \frac{SSC}{c - 1}; \quad s_3^2 = \frac{SS(RC)}{(r - 1)(c - 1)}; \quad s_4^2 = \frac{SSE}{rc(n - 1)}.$$

Para testar as hipóteses H_0 , H'_0 , e H''_0 calculam-se as seguintes razões:

$$f_1 = \frac{s_1^2}{s_4^2} \text{ para } H_0; \quad f_2 = \frac{s_2^2}{s_4^2} \text{ para } H'_0; \quad f_3 = \frac{s_3^2}{s_4^2} \text{ para } H''_0;$$

e comparam-se com os respectivos valores de uma distribuição F, isto é:

$$\text{rejeita-se } H_0 \text{ se } f_1 > F_{\alpha; (r-1); rc(n-1)}$$

$$\text{rejeita-se } H'_0 \text{ se } f_2 > F_{\alpha; (c-1); rc(n-1)}$$

$$\text{rejeita-se } H''_0 \text{ se } f_3 > F_{\alpha; (r-1)(c-1); rc(n-1)}$$

OBS: Fórmulas alternativas para as quantidades SST, SSR e SSC são:

$$\begin{aligned} SST &= \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijk}^2 - \frac{T_{...}^2}{rcn} \\ SSR &= \frac{\sum_{i=1}^r T_{i..}^2}{cn} - \frac{T_{...}^2}{rcn} \\ SSC &= \frac{\sum_{j=1}^c T_{.j.}^2}{rn} - \frac{T_{...}^2}{rcn} \\ SS(RC) &= \frac{\sum_{i=1}^r \sum_{j=1}^c T_{ij.}^2}{n} - \frac{\sum_{i=1}^r T_{i..}^2}{cn} - \frac{\sum_{j=1}^c T_{.j.}^2}{rn} + \frac{T_{...}^2}{rcn} \\ SSE &= SST - SSR - SSC - SS(RC) \end{aligned}$$

Os cálculos para um problema de análise de variância com várias observações por

celas podem ser resumidas em uma tabela ANOVA, da seguinte maneira:

TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Devido às linhas	SSR	$r - 1$	$s_1^2 = \frac{SSR}{r-1}$	$f_1 = \frac{s_1^2}{s_4^2}$
Devido às colunas	SSC	$c - 1$	$s_2^2 = \frac{SSC}{c-1}$	$f_2 = \frac{s_2^2}{s_4^2}$
Devido à interação	SS(RC)	$(r - 1)(c - 1)$	$s_3^2 = \frac{SS(RC)}{(r-1)(c-1)}$	$f_3 = \frac{s_3^2}{s_4^2}$
Erro	SSE	$rc(n - 1)$	$s_4^2 = \frac{SSE}{rc(n-1)}$	
Total	SST	$rcn - 1$		

Exemplo:

Utilizando os dados a seguir, teste as hipóteses abaixo, utilizando nível de significância de 5%:

Exemplo:

Fertilizantes	Variedades de Trigo		
	V_1	V_2	V_3
F_1	64	72	74
	66	81	51
	70	64	65
F_2	65	57	47
	63	43	58
	58	52	67
F_3	59	66	58
	68	71	39
	65	59	42
F_4	58	57	53
	41	61	59
	46	53	38

Solução:

As três hipóteses a serem testadas são:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

$$H_1 : \text{pelo menos um dos } \alpha_i \text{ não é zero.}$$

$$H'_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H'_1 : \text{pelo menos um dos } \beta_j \text{ não é zero.}$$

$$H''_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{43} = 0$$

$$H''_1 : \text{pelo menos um dos } \gamma_{ij} \text{ não é zero.}$$

Pelos dados da Tabela, podemos construir o resumo abaixo, contendo os totais:

Fertilizantes	Variedades de Trigo			Total
	V_1	V_2	V_3	
F_1	200	217	190	607
F_2	186	152	172	510
F_3	192	196	139	527
F_4	145	171	150	466
Total	723	736	651	2110

$$SST = 64^2 + 66^2 + \dots + 38^2 - \frac{2110^2}{36} = 3779$$

$$SSR = \frac{607^2 + 510^2 + 527^2 + 466^2}{9} - \frac{2110^2}{36} = 1157$$

$$SSC = \frac{723^2 + 736^2 + 651^2}{12} - \frac{2110^2}{36} = 350$$

$$SS(RC) = \frac{200^2 + 186^2 + \dots + 150^2}{3} - 124826 - 124019 - 123669 = 771$$

$$SSE = 3779 - 1157 - 350 - 771 = 1501$$

Estes resultados são resumidos na tabela ANOVA:

TABELA ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	f calculado
Devido aos fertilizantes	1157	3	385,67	6,17
Devido às variedades de trigo	350	2	175,00	2,80
Devido à interação	771	6	128,50	2,05
Erro	1501	24	62,54	
Total	3779	35		

Os procedimentos dos testes são dados por:

rejeita-se H_0 se $f_1 > F_{0,05; 3; 24} = 3,01$

rejeita-se H'_0 se $f_2 > F_{0,05; 2; 24} = 3,40$

rejeita-se H''_0 se $f_3 > F_{0,05; 6; 24} = 2,51$.

Portanto, a conclusão é:

- rejeitar H_0 e concluir que existe uma diferença na safra média de trigo quando diferentes tipos de fertilizantes são utilizados.
- aceitar H'_0 e concluir que não há diferença entre as diferentes variedades de trigo.
- aceitar H''_0 e concluir que não há interação entre as diferentes variedades de trigo e os diferentes tipos de fertilizantes.