

MARCELO FRANCISCO NOGUEIRA

MÉTODOS QUANTITATIVOS

**Seminário apresentado na disciplina Métodos Quantitativos do programa de
Mestrado em Ciências Contábeis do Centro Universitário Álvares Penteado.**

**São Paulo
2006**

SUMÁRIO

LISTA DE TABELAS.....	3
2 INTRODUÇÃO.....	5
3 ESTATÍSTICA DESCRITIVA.....	7
3.1 POPULAÇÃO E AMOSTRA.....	7
3.1.1 Rol.....	8
3.2 MEDIDAS DE TENDÊNCIA CENTRAL.....	9
3.2.1 Média Simples.....	9
3.2.2 Média Ponderada.....	11
3.2.3 Mediana.....	12
3.2.4 Moda.....	12
3.3 MEDIDAS DA DISPERSÃO.....	13
3.3.1 Amplitude Total.....	14
3.3.2 Variância.....	14
3.3.3 Desvio Padrão.....	16
3.3.4 Coeficiente de Variação.....	16
3.6 DIAGRAMA DE FREQUÊNCIA.....	17
3.6.1 Intervalos de classe.....	18
3.6.2 Representação gráfica.....	19
3.7 DADOS AGRUPADOS.....	20
3.7.1 Média.....	21
3.7.2 Mediana.....	21
3.7.3 Moda.....	22
3.7.4 Variância e desvio padrão.....	22
4 PROBABILIDADE.....	24
4.1 EVENTOS ALEATÓRIOS.....	25
4.2 CÁLCULO DE PROBALIDADES.....	25
4.3 PROBABILIDADE DOS EVENTOS.....	29
4.3.1 Teorema da soma - eventos mutuamente excludentes.....	29
4.3.2 Teorema da soma - eventos não mutuamente excludentes.....	30
4.3.3 Teorema da multiplicação - eventos independentes.....	31
4.3.4 Teorema da multiplicação - eventos dependentes (probabilidade condicional).....	32

4.3.5 Teorema de Bayes - (teorema da probabilidade das causas ou partições)	33
4.3.6 Síntese	36
4.4 AMOSTRAGEM	37
4.4.1 Amostragem sem reposição	37
4.4.1 Amostragem com reposição	38
5 ANÁLISE COMBINATÓRIA	39
5.1 Arranjos simples	39
5.2 Combinações simples	40
5.3 PERMUTAÇÃO SIMPLES	41
6 CONCLUSÃO	43
7 BIBLIOGRAFIA	44

LISTA DE TABELAS

Tabela 1 – Dados Brutos	8
Tabela 2 – Rol.....	9
Tabela 3 – Média Ponderada.....	11
Tabela 5 – Amplitude Total	14
Tabela 6 – Variância Série B	15
Tabela 7 – Variância Série C	15
Tabela 8 – Desvio Padrão	16
Tabela 9 – Coeficiente de Variação.....	17
Tabela 10 – Distribuição de freqüência.....	17
Tabela 11 – Distribuição de freqüência AKI-SE-TRABALHA.....	19
Tabela 12 – Dados agrupados ajustados	21
Tabela 13 – Probalidades, alguns exemplos	26

1 RESUMO

A Estatística é ferramenta gerencial utilizada há séculos, porquanto o homem sempre teve o desejo de mensurar, descrever fenômenos numéricos e procurar conceber o que ocorreria no futuro.

A estatística descritiva, seja por meio da representação tabular ou da representação gráfica ajuda a entender melhor os dados colhidos em uma determinada pesquisa.

Já, a estatística indutiva, surgiu com a curiosidade humana a respeito dos jogos, sobretudo os jogos de azar, quando alguém se perguntava: qual a chance de ganhar?

Hoje, essas ferramentas constituem parte indissociável de qualquer processo decisório, servindo, no aspecto descritivo, como elemento de melhora da qualidade das informações e, no aspecto indutivo, como meio de economia e otimização de tempo, porquanto permite inferir sobre uma população a partir de uma amostra.

2 INTRODUÇÃO

Também chamada de métodos quantitativos a estatística encontra-se presente em nosso dia a dia em praticamente todos os instantes. Assim é quando alguém questiona quantos quilômetros determinado veículo percorre com um litro de combustível, qual a taxa de nascimento ou de mortalidade, qual a chance de ganhar na loteria ou mesmo de chover em determinado dia.

Não é diferente no meio empresarial, onde os métodos quantitativos são utilizados como verdadeiras ferramentas de gestão. Assim é que em todos os telejornais há indicação da variação dos índices das Bolsas de Valores e da cotação do Dólar, normalmente representadas por meio de gráficos e tabelas.

COSTA¹ esclarece que a primeira vez em que o verbete *statistics* (estatística) apareceu na Enciclopédia Britânica foi em 1797. MEDRI², complementa, aduzindo que o termo que deriva do latim *status* refere-se a Estado, ou, precisamente, às descrições e dados relativos aos Estados, tornando a Estatística um meio de administração para os governantes.

Segundo STEVENSON³ a estatística compreende a estatística descritiva, a teoria da probabilidade e amostragem. Quando se fala em taxas e índices como forma de resumo, organização e, em geral, simplificação de informações que podem ser muito complexas, isto é, a dificuldade na descrição dos dados obtidos, estamos diante da estatística descritiva. A análise de situações que envolvem o acaso, tais como jogos de dados e cartas ou a maioria dos jogos esportivos, faz com que nos encontremos com as probabilidades. Por fim, quando há a análise e interpretação de dados amostrais entramos na área da amostragem ou estatística indutiva.

O exemplo dado por STEVENSON (p. 2) é preciso: não é preciso comer o bolo inteiro para saber se é bom. Essa uma grande finalidade da estatística: economizar recursos e tornar mais ágeis e seguros os processos decisórios.

¹ COSTA, Sérgio Francisco. **Introdução Ilustrada à Estatística (com muito humor!)**. 2ª ed. São Paulo: Harbra, 1992, p. 6.

² MEDRI, Waldir. **Métodos quantitativos aplicados à Contabilidade**. Londrina: UEL, 2003, p. 1

³ STEVENSON, William J. **Estatística aplicada à administração**. 1ª ed. São Paulo: 1981, p.3

A Estatística é ferramenta indispensável de uma Pesquisa Científica, sobretudo no que pertine à apresentação e análise dos dados coletados, de sorte a justificar as conclusões obtidas.

3 ESTATÍSTICA DESCRITIVA

Descrever a informação advinda dos dados colhidos em uma pesquisa é a finalidade da Estatística Descritiva. SPIEGEL⁴ aduz que, em sentido mais restrito, o termo Estatística é usado para designar os próprios dados ou números deles derivados como, por exemplo, médias.

O primeiro passo para que se possa caminhar em direção à informação é organizar os dados, em conformidade com a população e/ou com a amostra em estudo.

MEDRI (p. 1) ensina que na pesquisa científica coleta-se as características de pessoas, animais, empresas, indústrias, sistemas de produção, fenômenos físicos ou químicos, com a finalidade de verificar as hipóteses lançadas sobre uma população. Essa coleta é feita com base em uma amostra, lembrando aqui o ensinamento de STEVENSON.

3.1 POPULAÇÃO E AMOSTRA

População ou universo, corresponde a todo o grupo passível de exame. SPIEGEL (p. 1) retrata que uma população pode ser *finita* ou *infinita*. Destaca como exemplo que a produção de parafusos em uma fábrica em determinado dia é uma população finita ao passo em que todos os resultados (cara ou coroa) em sucessivos lances de uma moeda constituem uma população infinita.

Na conceituação de COSTA (p. 25), população é qualquer conjunto de informações que tenham, entre si, uma característica comum. Importa, assim, a variável estudada, seja ela qual for.

Amostra, por seu turno, é uma parcela retirada da população para estudo, segundo uma técnica adequada, de sorte a caracterizar-se como representativa. COSTA (p. 26) chega a dizer que a amostra nada mais é que uma redução da população a dimensões menores, sem perda das características essenciais.

⁴ SPIEGEL, Murray. **Estatística**. 2ª ed. São Paulo: Mc Graw-Hill do Brasil, 1985, p. 1.

Exsurge daí o conceito de estatística indutiva posto que conclusões atinentes à população podem ser inferidas ou deduzidas a partir da amostra.

As inferências daí decorrentes não expressam, necessariamente, uma certeza absoluta, pelo que impõe analisar, também, o conceito de probabilidade.

3.1.1 Rol

No aspecto descritivo da estatística, os dados brutos, ou seja, aqueles obtidos diretamente da pesquisa, não permitem uma análise direta, daí porque necessitam ser ordenados na forma de um rol.

Tomemos como exemplo os depósitos bancários feitos pela empresa AKI-SE-TRABALHA, expressos em milhares de Reais, nos meses de fevereiro e março de 2003⁵

3,7	1,6	2,5	3,0	3,9	1,9	3,8	1,5	1,1
1,8	1,4	2,7	2,1	3,3	3,2	2,3	2,3	2,4
0,8	3,1	1,8	1,0	2,0	2,0	2,9	3,2	1,9
1,6	2,9	2,0	1,0	2,7	3,0	1,3	1,5	4,2
2,4	2,1	1,3	2,7	2,1	2,8	1,9		

Tabela 1 – Dados Brutos

Pode-se conceituar rol como o arranjo dos dados brutos em ordem crescente. A finalidade do rol é permitir, de maneira mais clara e concisa, a análise dos dados, de sorte a visualizar o maior e o menor valor, a amplitude dos dados e os elementos que se repetem.

Arranjando os dados da tabela anterior em ordem crescente, na forma de rol, obtemos a disposição constante da Tabela 2.

⁵ MEDRI, op. cit., p. 10.

0,8	1,0	1,0	1,1	1,3	1,3	1,4	1,5	1,5
1,6	1,6	1,8	1,8	1,9	1,9	1,9	2,0	2,0
2,0	2,1	2,1	2,1	2,3	2,3	2,4	2,4	2,5
2,7	2,7	2,7	2,8	2,9	2,9	3,0	3,0	3,1
3,2	3,2	3,3	3,7	3,8	3,9	4,2		

Tabela 2 – Rol

3.2 MEDIDAS DE TENDÊNCIA CENTRAL

A média, assim como a mediana e a moda são medidas de tendência central, ou seja, são usadas para indicar um valor que tende a tipificar, ou a representar melhor, um conjunto de números, conforme ensina STEVENSON (p. 19).

COSTA (p. 56), por seu turno, prefere conceituar as medidas de tendência central como estatísticas, cujos valores estão próximos do centro de um conjunto de dados.

MEDRI (p. 22) prefere afirmar que as medidas de tendência central são aquelas que produzem um valor em torno do qual os dados observados se distribuem, e que visam sintetizar em um único número o conjunto de dados.

Essa última definição, que remete ao aspecto da distribuição dos dados, parece ser a mais correta, porque as medidas de tendência, não necessariamente representam melhor o conjunto, ao final de uma análise e porque a noção de centro, não significa, necessariamente, proximidade.

3.2.1 Média Simples

Denomina-se como média simples ou média aritmética o resultado da divisão da soma de todos os n valores amostrados pelo número de elementos amostrados.

Em termos numéricos, pode-se representar a média da seguinte forma:

$$\text{m\u00e9dia de uma mostra} \quad \bar{x} = \frac{\sum x}{n}$$

$$\text{m\u00e9dia de uma popula\u00e7\u00e3o} \quad \bar{\mu} = \frac{\sum x}{N}$$

Utilizando-se os dados colhidos na empresa AKI-SE-TRABALHA⁶, pode-se calcular a m\u00e9dia da seguinte forma:

- Soma dos elementos = 98,70
- N\u00famero de elementos do rol = 43
- M\u00e9dia = 2,30

$$\bar{x} = \frac{98,70}{43} = 2,30$$

STEVENSON (p. 20), discorrendo sobre a m\u00e9dia apresenta suas propriedades, quais sejam:

1. A m\u00e9dia de um conjunto de n\u00fameros pode ser sempre calculada.
2. Para um dado conjunto de n\u00fameros, a m\u00e9dia \u00e9 \u00fanica.
3. A m\u00e9dia \u00e9 sens\u00edvel a (ou afetada por) todos os valores do conjunto. Assim, se um valor se modifica, a m\u00e9dia tamb\u00e9m se modifica.
4. Somando-se uma constante a cada valor do conjunto, a m\u00e9dia ficar\u00e1 aumentada do valor dessa constante. Assim, somando-se 4,5 a cada valor de um conjunto, a m\u00e9dia ficar\u00e1 aumentada de 4,5. Analogamente, subtraindo-se cada valor do conjunto uma constante, ou multiplicando-se ou dividindo-se por ela cada valor do conjunto, a m\u00e9dia fica reduzida dessa constante, ou multiplicada ou dividida por ela.
5. A soma dos desvios dos n\u00fameros dos n\u00fameros de um conjunto a contar da m\u00e9dia \u00e9 zero.

⁶ Conforme a tabela 2 do item 3.1

3.2.2 Média Ponderada

Na apuração da média simples, cada elemento observado e constante do rol tem a mesma importância, ou, fazendo uma analogia com a média ponderada, a importância é unitária, ou seja, todos os elementos tem importância igual a 1.

Ao contrário, quando os elementos constantes do rol tem importâncias diferentes há a necessidade de, primeiramente, ajustar os valores do rol, conforme suas respectivas importâncias. Essas importâncias diferentes são denominadas por SPIEGEL (p. 55) como *fatores de ponderação* ou *pesos*.

A média ponderada é representada com o uso da seguinte fórmula:

$$média = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

Exemplificando o uso da média ponderada, destacamos como exemplo um aluno que realizou quatro provas e um exame final, sendo que as provas 1 e 3 tem peso 1, as provas 2 e 4 possuem peso 2 e o exame final possui peso 3. Os dados estão representados na tabela a seguir:

Evento	Nota	Peso	Nota Ajustada
Prova 1	7,00	1	7,00
Prova 2	5,00	2	10,00
Prova 3	6,00	1	6,00
Prova 4	8,00	2	16,00
Exame Final	6,00	3	18,00
Soma		9	57,00
Média ponderada		6,33	

Tabela 3 – Média Ponderada

3.2.3 Mediana

A mediana é o ponto (ou elemento) que divide o rol em duas partes iguais. Caso o número de elementos do rol seja ímpar, a mediana será o elemento central do rol. Na hipótese do número de elementos ser par, a mediana terá dois valores centrais e corresponderá à média entre esses dois valores.

Suponhamos o seguinte rol, constituído de 10 elementos:

13,1 – 13,1 – 13,1 – 13,2 – **13,3** – **13,5** – 13,5 – 13,7 – 13,7 – 13,9

Os elementos centrais são os números 13,3 e 13,5, donde a mediana corresponde à média entre esses dois valores, ou seja, 13,4.

Em outra situação, o rol anterior é acrescido de um elemento, qual seja, o número 14,0, passando a ter 11 elementos e a compor-se da seguinte forma:

13,1 – 13,1 – 13,1 – 13,2 – 13,3 – **13,5** – 13,5 – 13,7 – 13,7 – 13,9 – 14,0

Verifica-se que, nesse caso, a mediana é, exatamente, o elemento central, no caso, 13,5.

São propriedades da mediana:

1. Existe somente uma mediana para um conjunto de dados.
2. A mediana não é afetada pelos valores extremos como a média aritmética, por isso, se diz que a mediana é uma medida robusta.
3. Sempre que o rol estiver constituído em progressão aritmética (PA), a média será equivalente à mediana.

3.2.4 Moda

Nesse aspecto, MEDRI (p. 23) e STEVENSON (p. 23) concordam na conceituação, definindo, simplesmente, que a moda é o valor que ocorre com maior frequência num conjunto. Conseqüência desse conceito é que se todos os elementos do conjunto forem diferentes entre si, não haverá moda.

A moda não é fruto de um cálculo, ela resulta de uma observação e, por essa razão, não se presta, diretamente, à análise matemática. Contudo, o valor da moda chama atenção sempre que estiver próximo ou coincidir com a média ou com a mediana, posto que reforçará a tendência central da apuração.

3.3 MEDIDAS DA DISPERSÃO

Apurado um valor médio para os elementos de um rol torna-se necessário examinar as medidas de dispersão dos demais elementos em relação à tendência central, como meio de definir a variabilidade que os dados apresentam entre si. COSTA (p. 78) prefere chamar essas medidas como Medidas de Variabilidade

Somente não haverá dispersão quando todos os elementos do rol forem iguais, como ensina MEDRI (p. 25). As medidas de dispersão, assim, apresentam o grau de agregação dos dados.

Tomemos um exemplo numérico proposto por MEDRI (p. 26) para destacar a importância da análise das medidas de dispersão ou de variabilidade.

Repetição	Série A	Série B	Série C
1	45	41	25
2	45	42	30
3	45	43	35
4	45	44	40
5	45	45	45
6	45	46	50
7	45	47	55
8	45	48	60
9	45	49	65
Média	45	45	45
Mediana	45	45	45

Tabela 4 – Média Ponderada

Pode-se observar que enquanto a série A não apresenta qualquer variabilidade entre os seus elementos, as séries B e C apresentam dispersão, sendo

que a da série C é maior que a da série B. Chamamos atenção para o fato de que, para as três séries, tanto a média quanto a mediana são iguais.

Contudo, nem todas as séries analisadas terão situação clara como a presente, pelo que torna-se necessário calcular outros elementos de apoio matemático para que se possa precisar, no mais das vezes, qual a série ou conjunto de dados mais estável, isto é, que apresenta a menor dispersão entre os seus elementos.

As medidas descritivas mais comumente utilizadas são: amplitude total; variância e desvio padrão.

3.3.1 Amplitude Total

De modo geral a amplitude total corresponde à diferença entre o maior e o menor elemento de um conjunto de observações ou de um rol. Como é lastreada apenas nos valores extremos, apresenta limitação em seu aspecto conclusivo, como alerta MEDRI (p. 26).

No caso dos dados agrupados na tabela 4, podemos calcular a Amplitude Total e sintetizar os dados como segue:

Ítem	Série A	Série B	Série C
Menor valor	45	41	25
Maior Valor	45	49	65
Amplitude	0	8	40

Tabela 5 – Amplitude Total

3.3.2 Variância

A variância tem como fundamento os desvios de cada elemento do rol em relação à média. Para evitar que a soma dos desvios seja igual a zero, posto que os desvios podem ser positivos ou negativos, a variância considera o quadrado de cada desvio, ou seja, $(x_i - \bar{x})^2$

A fórmula de cálculo da variância é a seguinte:

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

Tomando como exemplo os dados da tabela 4, para as séries B e C, temos:

x_i	x	$(x_i - x)$	$(x_i - x)^2$
41	45	-4	16
42	45	-3	9
43	45	-2	4
44	45	-1	1
45	45	0	0
46	45	1	1
47	45	2	4
48	45	3	9
49	45	4	16
9		0	60
Variância =		7,5	

Tabela 6 – Variância Série B

x_i	x	$(x_i - x)$	$(x_i - x)^2$
25	45	-20	400
30	45	-15	225
35	45	-10	100
40	45	-5	25
45	45	0	0
50	45	5	25
55	45	10	100
60	45	15	225
65	45	20	400
9		0	1500
Variância =		187,5	

Tabela 7 – Variância Série C

3.3.3 Desvio Padrão

De modo sintético, o desvio padrão é a raiz quadrada da variância, ou seja:

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

Para as séries A, B e C, indicadas na Tabela 4, teríamos os seguintes desvios padrões:

Ítem	Série A	Série B	Série C
Variância	0,0	7,5	187,5
Desvio Padrão	0,0	2,7	13,7

Tabela 8 – Desvio Padrão

3.3.4 Coeficiente de Variação

Após calcular o desvio padrão a pergunta que segue é a seguinte: o desvio é grande ou pequeno. Essa questão é relevante quando se quer saber a precisão do método indicado em uma pesquisa.

Contudo, em valores nominais a questão não pode ser respondida posto que depende da grandeza dos números envolvidos. Com efeito. Em uma base de observação com valores médios de 10.000 e moda e mediana nessa mesma faixa, um desvio de 10 é irrisório. Contudo, esse mesmo desvio para valores cuja observação média típica é 50 torna-se bastante elevado.

Para responder a indagação primitiva é necessário utilizar o Coeficiente de Variação, que, conforme MEDRI (p. 27), é um número adimensional, isto é, um número puro e é usualmente expresso em porcentagem. Quanto menor o coeficiente de variação, mais homogêneo é o conjunto analisado.

A fórmula de cálculo do Coeficiente de Variação é a seguinte:

$$CV = \frac{s}{x} \cdot 100$$

Com relação às séries estudadas (A, B e C), indicadas na Tabela 4, teríamos os seguintes coeficientes de variação:

Ítem	Série A	Série B	Série C
Média	45,0	45,0	45,0
Desvio Padrão	0,0	2,7	13,7
Coeficiente de Variação	0,00%	6,09%	30,43%

Tabela 9 – Coeficiente de Variação

3.6 DIAGRAMA DE FREQUÊNCIA

SPIEGEL (p. 33) define que "um arranjo tabular dos dados por classes, juntamente com as freqüências correspondentes, é denominado distribuição de freqüência ou tabela de freqüência". STEVENSON (p. 32) esclarece que "uma distribuição de freqüência é um método de agrupamento de dados em classes, ou intervalos, de tal forma que se possa determinar o número, ou a porcentagem (isto é, a freqüência) de cada classe".

Conquanto se possa adotar nessas situações em que se analisa um grande conjunto de dados, é imperioso observar, como lembra MEDRI (p. 28) que, "no agrupamento de dados acarreta alguma perda de informação. cada elemento perde sua identidade, por isso, sabe-se apenas quantos elementos há em cada classe".

Para exemplificar uma distribuição de freqüência, utilizaremos a tabela criada por SPIEGEL (p. 33), que retrata as alturas de 100 estudantes do sexo masculino da Universidade XYZ.

Altura (cm)	Número de estudantes
151-158	5
159-166	18
167-174	42
175-182	27
183-190	8
Total	100

Tabela 10 – Distribuição de freqüência

3.6.1 Intervalos de classe

Quando se trabalha com um grande volume de dados brutos e se quer utilizar o sistema de diagramas de freqüência é necessário agrupar os dados coletados em intervalos.

Esses intervalos são chamados de intervalos de classe e devem ser definidos com cuidado eis que, como recomenda MEDRI (p. 16) "poucos intervalos podem resultar em perda da informação. Por outro lado, muitos intervalos não resumem a informação".

Assim, a questão crucial em relação aos intervalos de classe é definir a amplitude dos mesmos, para o que devem ser observados os seguintes passos⁷:

- 1) Encontrar o menor e o maior valor do conjunto de dados e calcular a amplitude entre eles por: **$At = n^{\circ} \text{ do maior} - n^{\circ} \text{ do menor}$**
- 2) Não existindo um critério rígido para estabelecer o número ideal de intervalos, sugere-se que não se utilize menos de 6 e não mais de 15 intervalos. A experiência tem demonstrado que se pode fixar o número de intervalos como: **$K = \sqrt{n}$** ou **$K = 1 + 3,3 \cdot \log n$** , para uma amostra de

tamanho n

- 3) Uma vez determinado o número de intervalos, o tamanho destes é dado

por $C = \frac{At}{K}$

No exemplo destacado na Tabela 1, atinente à empresa AKI-SE-TRABALHA há um conjunto de 43 dados. Os intervalos de classe poderiam ser obtidos da seguinte forma:

$$At = n^{\circ} \text{ do maior} - n^{\circ} \text{ do menor} \therefore At = 4,2 - 0,8 = 3,43$$

$$K = \sqrt{n} \therefore K = \sqrt{43} = 6,56 \cong 7$$

$$C = \frac{At}{K} \therefore C = \frac{3,4}{7} \cong 5$$

⁷ MEDRI, op. cit., p. 17.

Segundo esses intervalos a distribuição de frequência dos dados colhidos na referida empresa ficaria desse modo:

Depósitos bancários (milhares R\$)	Frequência absoluta	Ponto médio	Frequência relativa %	Frequência acumulada
0,8 -- 1,3	4	1,05	9,30%	4
1,3 -- 1,8	7	1,55	16,28%	11
1,8 -- 2,3	11	2,05	25,58%	22
2,3 -- 2,8	8	2,55	18,60%	30
2,8 -- 3,3	8	3,05	18,60%	38
3,3 -- 3,8	2	3,55	4,65%	40
3,8 -- 4,3	3	4,05	6,98%	43
Total	43		100,00%	

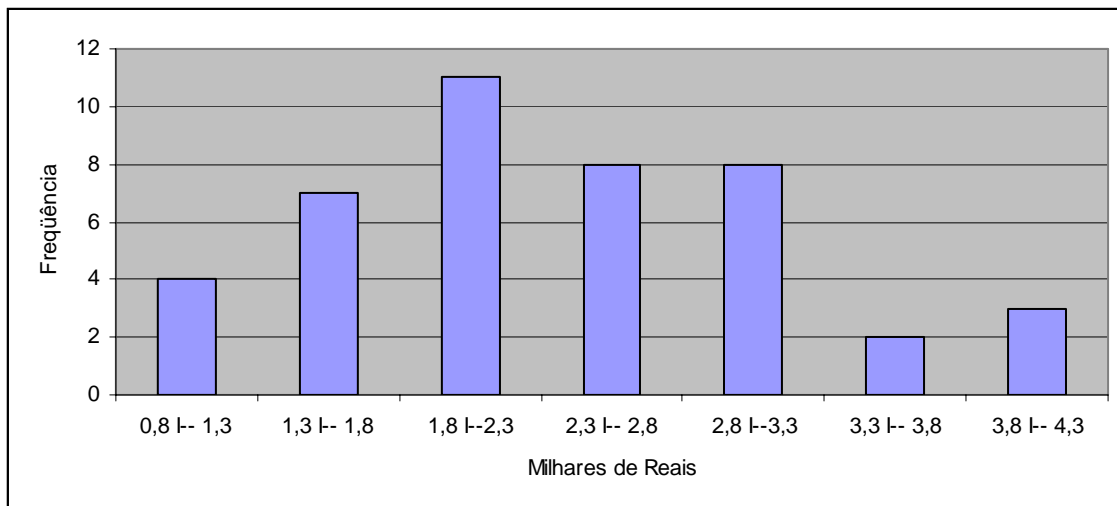
Tabela 11 – Distribuição de frequência AKI-SE-TRABALHA

3.6.2 Representação gráfica

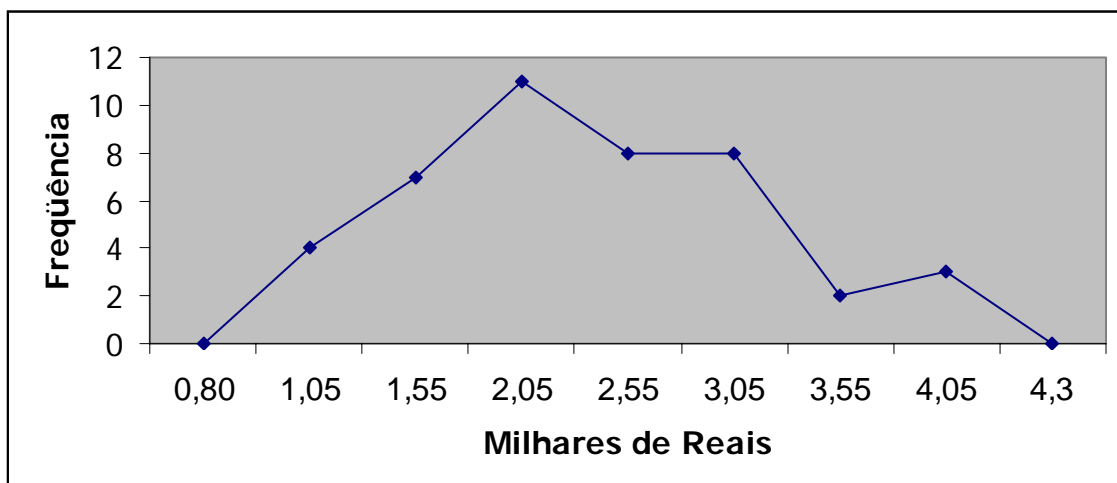
O diagrama mencionado no item anterior também poderia ser demonstrado em formato gráfico, sendo que as duas formas mais frequentes são o histograma e o polígono de frequência.

Temos:

- a) Histograma



b) Polígono de frequência



3.7 DADOS AGRUPADOS

Os dados agrupados também possibilitam o cálculo das medidas de tendência central e das medidas de dispersão tal qual ocorre com relação aos dados não agrupados em distribuição de frequência. A única diferença é que torna-se necessário observar a frequência de ocorrência de uma determinada classe de dados, como se fosse uma média ponderada, porém, em relação ao ponto médio da classe.

Nesse contexto, quando se trabalha com dados agrupados em distribuição de frequência o ponto de partida é ajustar a frequência em relação ao

ponto médio, tanto pelo valor simples, como pelo quadrado desse ponto médio, como feito com relação à variância dos dados não agrupados.

Considerando-se o exemplo da empresa AKI-SE-TRABALHA, tal qual proposto por MEDRI (p. 28), teríamos a seguinte tabela ajustada:

Depósitos bancários (milhares R\$)	Frequência absoluta (f_i)	Ponto médio (x_i)	Frequência acumulada (F_{ac})	x_i.f_i	x_i².f_i
0,8 -- 1,3	4	1,05	4	4,20	4,41
1,3 -- 1,8	7	1,55	11	10,85	16,82
1,8 -- 2,3	11	2,05	22	22,55	46,23
2,3 -- 2,8	8	2,55	30	20,40	52,02
2,8 -- 3,3	8	3,05	38	24,40	74,42
3,3 -- 3,8	2	3,55	40	7,10	25,21
3,8 -- 4,3	3	4,05	43	12,15	49,21
Total	43			101,65	268,31

Tabela 12 – Dados agrupados ajustados

A partir desse ajuste, todas as demais medidas de tendência central e dispersão poderiam ser calculadas.

3.7.1 Média

A fórmula para o cálculo da média de uma distribuição de frequências é a seguinte:

$$\bar{x} = \frac{\sum_{i=1}^n \bar{x}_i \cdot f_i}{n}, \text{ assim } \bar{x} = \frac{101,65}{43} = 2,36$$

3.7.2 Mediana

A expressão para determinar a mediana de uma distribuição de frequências é dada por:

$$Md = l_i + \frac{\frac{n}{2} - F_{ac-1}}{f_{MD}} a_c, \text{ assim } Md = 1,8 + \frac{21,5 - 11}{11} 0,5 = 2,28$$

Destaca-se que:

l_i = limite inferior da classe da mediana;

n = número de elementos;

a_c = amplitude da classe;

F_{ac-1} = frequência simples da classe Md;

f_{MD} = frequência simples da classe Md

3.7.3 Moda

Determina-se a moda de uma distribuição de frequências com a utilização da seguinte expressão:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_c, \text{ assim } Mo = 1,8 + \frac{4}{4 + 3} 0,5 = 2,09$$

Observamos que:

l_i = limite inferior da classe modal (de maior frequência);

Δ_1 = diferença entre a frequência da classe modal e a anterior;

Δ_2 = diferença entre a frequência da classe modal e a posterior;

3.7.4 Variância e desvio padrão

Do mesmo modo que em relação aos dados não agrupados, pode-se aferir medidas de dispersão dos dados acumulados, proporcionalizados por suas respectivas frequências em relação à tendência central média.

A primeira medição que se faz é a da variância, com o uso da seguinte expressão:

$$s^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \frac{(\sum_{i=1}^n x_i f_i)^2}{n}}{n-1}, \text{ assim, } s^2 = \frac{268,31 - \frac{(101,65)^2}{43}}{43-1} = 0,667$$

Na exata medida em que o desvio padrão corresponde à raiz quadrada da variância, temos:

$$s = \sqrt{s^2} \text{ e, } s = \sqrt{0,667} = 0,817$$

4 PROBABILIDADE

De modo geral, os autores concordam que as aplicações iniciais da matemática da probabilidade referiam-se quase todas aos jogos de azar. Assim é que:

O primeiro trabalho escrito de que se tem notícia e que envolve a noção de probabilidade data de 1477. Trata-se de um comentário feito á Divina Comédia (Dante), onde há referência às probabilidades associadas aos vários resultados decorrentes do jogo de 3 dados (COSTA, p. 90).

Pode-se recorrer ao cálculo da probabilidade sempre que, independente de qual seja a aplicação em particular, exista um elemento de acaso, ou de incerteza, quanto à ocorrência ou não de um evento futuro.

A probabilidade faz parte da Estatística Indutiva, ou seja, lastreia-se em processo de inferência através do qual, pelo correto exame de uma amostra pode-se inferir o que ocorrerá com a população. A maior utilidade da probabilidade, como menciona STEVENSON (p. 55) é auxiliar o desenvolvimento de estratégias, quantificando o quão provável é determinado evento.

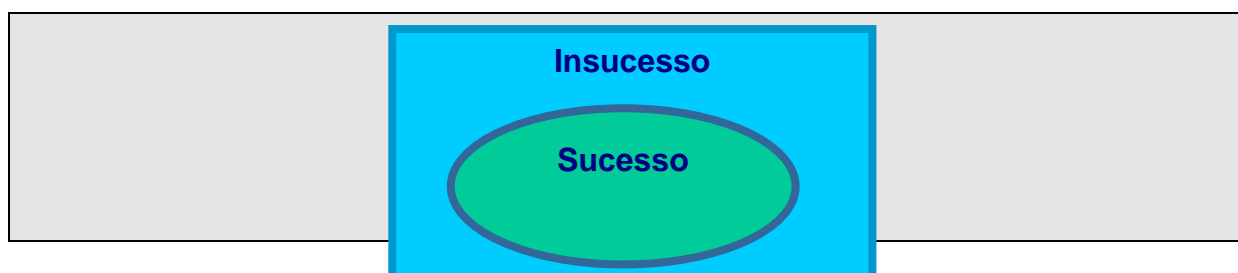
No campo contábil essa utilidade pode ser vista com facilidade no campo orçamentário e no campo da previsão de cenários, onde, baseado em dados históricos, devidamente mensurados e com apoio matemático para prever uma tendência (probabilidade) pode-se definir por este ou aquele cenário.

COSTA (p. 92) sintetiza que "probabilidade é o número que resulta da divisão do número de casos favoráveis a um evento pelo número total de casos possíveis".

Diante dessa definição, o número resultante desse cálculo somente pode estar compreendido entre 0 e 1, porque o maior número de casos favoráveis possível corresponde ao número total de casos possíveis e, em sentido inverso, o menor número possível de casos favoráveis é 0, ou seja, é quando não há possibilidade de ocorrer um evento favorável.

SPIEGEL (p. 127) diz que a definição quanto "a probabilidade da ocorrência do evento (denominado sucesso) é definida por: $p = \Pr\{E\} = \frac{h}{n}$ ". Por outro

lado, a probabilidade do chamado insucesso, é representada da seguinte forma: $1 - \Pr\{E\}$. Essa formulação pode ser representada através da seguinte figura:



4.1 EVENTOS ALEATÓRIOS

GUERRA e DONAIRE (p. 14) ensinam que, para que um evento possa ser considerado aleatório, devem estar presentes duas condições essenciais, quais sejam:

- a) deve ser sempre possível repetir a experiência indefinidamente, fixadas certas condições iniciais;
- b) mesmo mantendo as condições iniciais, deve ser impossível influenciar no resultado de uma particular repetição da experiência.

Como exemplos de eventos aleatórios podem ser considerados o lançamento de uma moeda honesta ou o lançamento de um dado não viciado.

4.2 CÁLCULO DE PROBALIDADES

Tomando como parâmetro a fórmula proposta por SPIEGEL (p. 127) pode-se concluir que o número de resultados associados ao evento (A) dividido pelo número total de resultados possíveis corresponde à probabilidade desse evento ocorrer.

Tomando como paradigma de exemplificação uma moeda, constata-se que a mesma tem duas faces (cara e coroa). Logo, quando se arremessa, a chance de obter cara corresponde a $\frac{1}{2}$ porque a moeda ó possui uma face denominada como cara, de modo que só há uma possibilidade de resultado que atenda ao requisito do lançamento (obter a face cara). De outro lado, como a moeda possui duas faces, há duas possibilidades de resultados (cara ou coroa).

No exemplo supra, as probabilidades são complementares, porque:

$P(cara) = \frac{1}{2}$ e $P(coroa) = \frac{1}{2}$, de modo que, somando-se esses dois eventos, obtém-se a totalidade dos eventos possíveis, ou seja: $\frac{1}{2} + \frac{1}{2} = 1 = 100\%$

se a totalidade dos eventos possíveis, ou seja: $\frac{1}{2} + \frac{1}{2} = 1 = 100\%$

Vários outros exemplos poderiam ser citados, pelo que destacamos os seguintes:

Evento esperado	Resultados associados ao evento esperado	Total resultados possíveis	de Formulação
obter cara no lançamento de uma moeda	1	2	$P(cara) = \frac{1 \text{ face}}{2 \text{ faces}} = \frac{1}{2} = 50,00\%$
obter a face 6 no lançamento de um dado	1	6	$P(cara) = \frac{1 \text{ face}}{6 \text{ faces}} = \frac{1}{6} = 16,67\%$
obter uma face impar qualquer no lançamento de um dado	3	6	$P(cara) = \frac{3 \text{ faces}}{6 \text{ faces}} = \frac{1}{2} = 50,00\%$
extrair uma carta de copas em um baralho completo	13	52	$P(cara) = \frac{13 \text{ cartas}}{52 \text{ cartas}} = \frac{1}{4} = 25,00\%$

Tabela 13 – Probalidades, alguns exemplos

Quando se fala em evento é necessário observar seus tipos e associações, para o que, apoiamo-nos na classificação e nos exemplos de GUERRA e DONAIRE (p. 16), quais sejam:

- a) **Evento simples:** é aquele formado por um único elemento do espaço amostral. Exemplo: no lançamento de moedas há dois eventos simples: ou cara ou coroa: $E_1 = \{K\}$ $E_2 = \{C\}$.

CARA | COROA

- b) **Evento composto:** é aquele formado por dois ou mais elementos do espaço de amostragem. Exemplo: no lançamento de dados podem ser considerados os números pares, ou os

ímpares, ou os dois primeiros números e, assim por diante: $E_1 = \{2,4,6\}$ $E_2 = \{1,3,5\}$ $E_3 = \{1,2\}$. $E_4 = \{1,2\}$

1	2
3	4
5	6

- c) **Evento certo:** é aquele que ocorre sempre, isto é, em todas as realizações da experiência. Exemplo: obter qualquer das faces de 1 a 6 no lançamento de dados.

1	2	RETIRAR	$1 \leq n \leq 6$
3	4		
5	6		

- d) **Evento impossível:** é aquele que nunca ocorre, isto é, em nenhuma realização de experiência. Exemplo: obter face maior que 6 no lançamento de um dado.

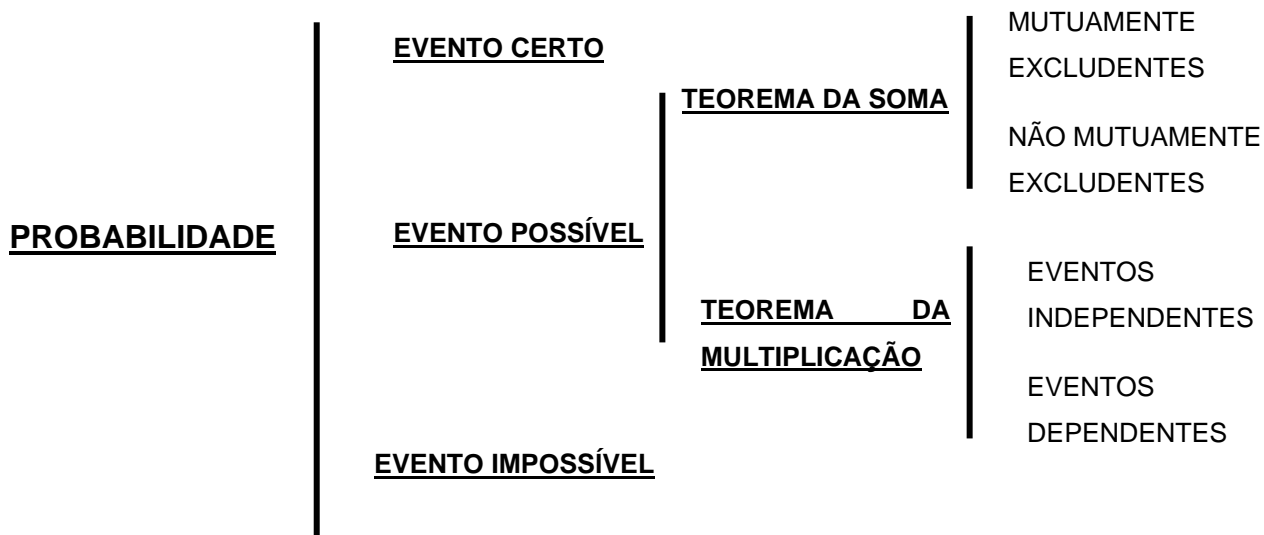
1	2	RETIRAR	$n > 6$
3	4		
5	6		

- e) **Evento soma (ou evento união):** é o evento que consiste na realização de pelo menos um dos eventos E_1 e E_2 . Exemplo: na retirada de uma carta do baralho quer-se que seja uma carta de ouro ou uma carta ás.
- f) **Evento produto (ou evento intersecção):** é o evento que consiste na realização de ambos os eventos E_1 e E_2 , isto é, eles devem ocorrer simultaneamente. Exemplo: na retirada das cartas de um baralho quer-se um ás de ouro.
- g) **Evento condicionado:** é o evento que consiste na realização do evento E_1 sob a condição de ter-se realizado o evento E_2 isto é, com a informação adicional de que o evento E_2 já ocorreu.

Exemplo: na retirada de uma carta do baralho, quer-se que ocorra uma carta às, sabendo-se que a carta é de ouro.

- h) **Eventos mutuamente exclusivos:** dois eventos, associados a uma experiência aleatória, são ditos mutuamente exclusivos se a ocorrência de um deles exclui a possibilidade de ocorrência do outro, isto é, não ocorrem simultaneamente. Exemplo: no lançamento de uma moeda, os eventos cara e coroa são mutuamente exclusivos.
- i) **Evento complementar (ou contrário):** define-se como evento complementar de um evento E , associado a uma experiência aleatória, e denota-se por \bar{E} , ao evento que só ocorre se E deixar de ocorrer, isto é, é o evento formado por todos os elementos do espaço amostral que não pertencem a E . Exemplos: no lançamento de uma moeda os eventos cara e coroa são mutuamente exclusivos e também são complementares. Já, no lançamento de um dado, os eventos 1 e 4 são mutuamente exclusivos mas não são complementares.
- j) **Eventos independentes:** dois eventos, associados a uma experiência aleatória, são ditos independentes quando a ocorrência de um deles não depende (ou não é condicionada ou não se vincula) à ocorrência do outro, isto é, a informação adicional de que um dos elementos já ocorreu em nada altera a possibilidade de ocorrência do outro. Exemplo: suponhamos que duas pessoas atirem numa caça; os eventos que consistem em que cada uma das pessoas acerte são independentes, pois o fato da primeira pessoa acertar em nada influencia no fato da outra também acertar.

Em termos esquemáticos, pode-se indicar:



4.3 PROBABILIDADE DOS EVENTOS

STEVENSON (p. 74) sintetiza as seguintes regras das propriedades:

4.3.1 Teorema da soma - eventos mutuamente excludentes

- P (A ou B), para eventos mutuamente excludentes:

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

1	2
3	4
5	6

RETIRAR

$n = 1$ ou $n = \text{par}$

Na situação proposta, observamos que:

- $P(A) = \text{obter a face } 1 = \frac{1}{6}$
- $P(B) = \text{obter um número par} = \frac{3}{6}$
- Aplicando a fórmula, obtemos: $P = P(A) + P(B) = \frac{1}{6} + \frac{3}{6} = \frac{4}{6}$

Pode-se destacar um outro exemplo: um jogo de dado em que se quer obter a face 1 ou a face 3. Observe-se que os eventos são excludentes porque, se sair a face 1 não sairá a face 3 e vice versa. Para as duas faces, a probabilidade de êxito é a mesma, ou seja, $\frac{1}{6}$ para cada uma. Conseqüentemente, a probabilidade de êxito nessa situação, em que duas de 6 faces atingem o objetivo é o resultado da seguinte soma: $\frac{1}{6} + \frac{1}{6} = \frac{2}{6}$.

4.3.2 Teorema da soma - eventos não mutuamente excludentes

- **P (A ou B), para eventos não mutuamente excludentes:**

$$P(\text{A ou B ou ambos ocorrerão}) \\ = P(A) + P(B) - P(A \text{ e } B)$$

1	2
3	4
5	6

RETIRAR

$$n = (1 \text{ ou } 2) \text{ ou} \\ n = (\text{par})$$

A aplicação da fórmula e o desenvolvimento do método, fazem observar que há um elemento no espaço amostral que atende tanto uma quanto outra condição, qual seja, o número 2. Chamando a primeira condição como P(A) e a segunda como P(B), temos:

- a) $P(A) = \frac{2}{6}$ referente à ocorrência do número 1 ou do número 2
- b) $P(b) = \frac{3}{6}$ atinente à ocorrência de um número par
- c) $P(A/B) = \frac{1}{6}$ referente à provável ocorrência do número 2
- d) Aplicando a fórmula, temos: $P = P(A) + P(B) - P(A/B)$

$$= \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6}$$

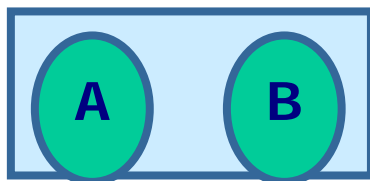
Outro exemplo: num jogo de cartas, quer-se uma carta 10 ou uma carta de copas. Observe-se que essa segunda opção não exclui a primeira posto que, mesmo a carta 10 pode ser de copas. Isoladamente, a probabilidade de sair a carta 10 corresponde a $\frac{4}{52}$ enquanto que a probabilidade de sair uma carta de copas a $\frac{13}{52}$. Aqui há um cuidado a ser tomado, que é o seguinte: pelo fato da carta 10

poder ser uma carta de copas, na probabilidade de sair uma carta desse naipe, já está computada a probabilidade de sair a carta 10, valendo dizer que a simples soma das duas probabilidades implicará em cômputo duplo dessa carta, sendo necessário um ajuste. Logo, nessa situação, somam-se as probabilidades e exclui-se o elemento comum, qual seja, a probabilidade de sair a carta 10 de copas ($1/52$). Numericamente: $4/52 + 13/52 - 1/52 = 16/52 = 4/13$.

4.3.3 Teorema da multiplicação - eventos independentes

- P (A e B), para eventos independentes:

$$P(\text{A e B ocorrem} - \text{ou não} - \text{sem interferência de um sobre o outro}) \\ = P(A) \cdot P(B)$$



Pode-se exemplificar esse tipo de cálculo com a seguinte situação: em 70% das ocasiões um motorista excede a velocidade máxima em uma determinada via pública. Sabe-se, também, que em 30% das ocasiões, o Departamento de Trânsito coloca um radar móvel nessa mesma via. O motorista deseja saber qual é a probabilidade de que ele exceda a velocidade máxima em um dia em que o radar estiver na via. Pois bem, chamando o excesso de velocidade de $P(A)$ e a presença do radar de $P(B)$, aplicamos a fórmula do seguinte modo:

$$P = P(A) \cdot P(B) = 0,70 \cdot 0,30 = 0,21 = 21,00\%$$

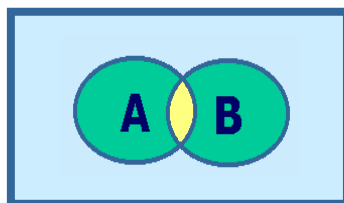
Admita-se um outro exemplo em que um determinado aluno chegue atrasado a aula em 25% das ocasiões e que, por outro lado, o professor falte em 10% das aulas. Nessa situação, qual é a probabilidade do aluno chegar atrasado e o professor ter faltado à aula. Observamos que um evento não depende e tampouco se relaciona com o outro, posto que ambos podem ocorrer, ambos podem não

ocorrer e pode ocorrer um sem que ocorra o outro, sendo, também importante, que um não afeta a probabilidade de ocorrência do outro. Para aferir a probabilidade de ocorrência simultânea desses eventos há que se efetuar a multiplicação das probabilidades, então: $0,25 \times 0,10 = 0,025$ ou 2,5%.

4.3.4 Teorema da multiplicação - eventos dependentes (probabilidade condicional)

- **P (A e B), para eventos dependentes:**

$$P(\text{A e B ocorrem} - \text{ou não} - \text{com interferência de um sobre o outro}) \\ = P(A) \cdot P(A|B) \text{ ou } P(B) \cdot P(B|A)$$



Essa é a chamada probabilidade condicional, em que os dois eventos devem ocorrer (ou não), havendo uma relação de dependência entre os mesmos, ou seja, a ocorrência (ou não) de um afeta de modo direto a probabilidade de ocorrência (ou não do outro). Citamos o seguinte exemplo: numa determinada urna existem 5 bolas, sendo 2 pretas e 3 brancas. Quer-se saber qual é a probabilidade de retirada de duas bolas, apenas, as quais deverão ser pretas, sendo que não haverá reposição da bola retirada. Pois bem: a probabilidade inicial de retirar uma bola preta corresponde a $\frac{2}{5}$, pois existem 2 bolas pretas dentre as 5 bolas que estão na urna. Contudo, na hipótese da primeira bola retirada ser preta, a probabilidade da segunda também ser preta corresponderá a $\frac{1}{4}$, posto que, nessa ocasião haverão apenas 4 bolas, das quais somente uma é preta. Assim, temos que a probabilidade, nessa situação, corresponderá a: $\frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10} = 10\%$.

Numa outra situação, uma empresa possui 10 veículos, sendo que 7 são do tipo passeio e 3 são utilitários. Essa empresa deseja fazer uma inspeção nos

veículos. Sem que exista reposição, qual é a probabilidade de que a empresa escolha, em duas amostras, dois veículos utilitários? Pois bem, chamando a probabilidade de retirar o primeiro veículo utilitário como $P(A)$ e a de escolher o segundo veículo utilitário como $P(A/B)$, temos:

$$P(A) = 3/10 \text{ pois há 3 utilitários em um universo de 10 veículos}$$

$P(A/B) = 2/9$ pois, após a escolha do 1º veículo permanecerão 9 veículos, dos quais 2 são utilitários.

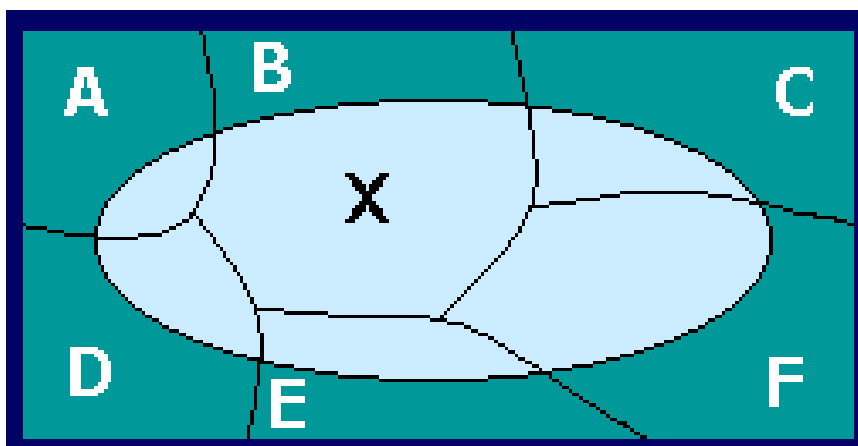
Aplicando a fórmula, temos:

$$P = P(A) \cdot P(A|B) = \frac{3}{10} \cdot \frac{2}{9} = \frac{6}{90} = \frac{1}{15} = 0,6666 = 6,66\%.$$

A diferença básica entre o cálculo de probabilidade para eventos dependentes e para eventos independentes, com o teorema da multiplicação é que, na primeira situação (eventos dependentes) a probabilidade de ocorrência (ou não) de determinado evento não se altera com a ocorrência do outro evento. Já, na probabilidade condicional (eventos dependentes) a ocorrência de um altera a probabilidade do outro. Nos nossos exemplos verifica-se que após o primeiro evento houve alteração do espaço amostral, modificando a probabilidade de ocorrência de qualquer outro evento.

4.3.5 Teorema de Bayes - (teorema da probabilidade das causas ou partições)

LIPSCHUTZ (p. 90) alude que o Teorema de Bayes, também conhecido como *partições* ou Teorema das Probabilidades das Causas tem como fundamento a suposição de que vários eventos formam uma partição num dado espaço amostral S .



Conforme a disposição da figura anterior, dentro do evento X o teorema (e por essa razão é chamado como Teorema da Probabilidade das Causas) indica qual é a probabilidade de que o citado evento tenha sido originado de A ou B ou D e E e, assim por diante.

OLIVEIRA (1999:96) esclarece que o teorema permite determinar as probabilidades dos vários eventos $A_1, A_2, A_3, \dots, A_n$, que podem ser a causa da ocorrência de B^8 . Acrescenta:

Sejam $A_1, A_2, A_3, \dots, A_n$, eventos mutuamente *excludentes* cuja união é o espaço amostral Ω , ou seja, um dos eventos *necessariamente* deve ocorrer.

LIPSCHUTZ (p. 90) apresenta a seguinte fórmula para o teorema:

$$P(A_1 \setminus B) = \frac{P(A_1)P(B \setminus A_1)}{P(A_1)P(B \setminus A_1) + P(A_2)P(B \setminus A_2) + \dots + P(A_N)P(B \setminus A_N)}$$

A fórmula supra é sintetizada por OLIVEIRA (1999:96) da seguinte maneira:

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

Aplicando o teorema, podemos utilizar o seguinte exemplo, destacado por GUERRA e DONAIRE (p. 28):

Numa fábrica existem três máquinas destinadas à produção de parafusos. A 1ª máquina produz diariamente 1.000 parafusos, a 2ª máquina 4.000 e a 3ª máquina 5.000. Sabendo-se que a 1ª máquina produz 4% de parafusos defeituosos, a 2ª máquina 3% e a terceira 1% e, tendo-se, ao final do dia, encontrado um parafuso defeituoso, qual a probabilidade dele ter sido produzido em cada uma dessas máquinas?

Resolução:

Sejam:

⁸ No nosso exemplo, $A_1, A_2, A_3, \dots, A_n$ são substituídos por A, B, C, D, E, F e G o espaço amostral B é substituído por X.

A = evento de ser parafuso produzido na 1ª máquina

B = evento de ser parafuso produzido na 2ª máquina

C = evento de ser parafuso produzido na 3ª máquina

X = evento de ser parafuso defeituoso.

Então:

$$P(A) = 1.000/10.000 = 0,10$$

$$P(B) = 4.000/10.000 = 0,40$$

$$P(C) = 5.000/10.000 = 0,50$$

$$P(X/A) = 0,04$$

$$P(X/B) = 0,03$$

$$P(X/C) = 0,01$$

Aplicando a fórmula temos:

EVENTO A

$$P(A \setminus X) = \frac{P(A) \cdot P(X \setminus A)}{P(A) \cdot P(X \setminus A) + P(B) \cdot P(X \setminus B) + P(C) \cdot P(X \setminus C)} =$$

$$P(A \setminus X) = \frac{0,10 \cdot 0,04}{0,10 \cdot 0,04 + 0,40 \cdot 0,03 + 0,50 \cdot 0,01} = 0,19 = 19 \%$$

EVENTO B

$$P(B \setminus X) = \frac{P(B) \cdot P(X \setminus B)}{P(A) \cdot P(X \setminus A) + P(B) \cdot P(X \setminus B) + P(C) \cdot P(X \setminus C)} =$$

$$P(B \setminus X) = \frac{0,40 \cdot 0,03}{0,10 \cdot 0,04 + 0,40 \cdot 0,03 + 0,50 \cdot 0,01} = 0,57 = 57\%$$

EVENTO C

$$P(C \setminus X) = \frac{P(C) \cdot P(X \setminus C)}{P(A) \cdot P(X \setminus A) + P(B) \cdot P(X \setminus B) + P(C) \cdot P(X \setminus C)} =$$

$$P(C \setminus X) = \frac{0,50 \cdot 0,01}{0,10 \cdot 0,04 + 0,40 \cdot 0,03 + 0,50 \cdot 0,01} = 0,24 = 24\%$$

Atendendo à questão formulada, pode-se observar que a probabilidade do parafuso defeituoso ter sido produzido pela máquina A corresponde a 19%, pela máquina B corresponde a 57% e, para a máquina C a probabilidade é de 24%. Somando-se essas três probabilidades obtém-se a totalidade do espaço amostral (19% + 57% + 24% = 100%).

4.3.6 Síntese

Genericamente quando a definição do problema permite uma opção, montando-se o esquema com a figura de um ou outro evento, estamos diante de um caso de **soma** de probabilidades. De outra parte, quando o problema impõe condições atingíveis ao mesmo tempo, encontramos uma situação de **multiplicação** das probabilidades.

4.4 AMOSTRAGEM

SPIEGEL (p. 175) define que “a teoria da amostragem é um estudo das relações existentes entre uma população e as amostras dela extraídas”. Por várias razões utilizam-se das técnicas de amostragem, destacando-se as de razão econômica e aquelas derivadas de outro exemplo curioso de STEVENSON (p. 158): “você não precisa beber todo o ponche para saber que gosto ele tem!”.

Exsurge daí a importância do processo de inferência que caracteriza a estatística indutiva, posto que o propósito da amostragem é permitir inferir generalizações a cerca da população.

As populações podem ser finitas (número de alunos de uma universidade, número de casa de um bairro, número de automóveis de uma frota) ou infinitas (jogadas de moeda – cara e coroa, nascimento de insetos, extração de bolas de uma urna com reposição) e, o conhecimento da natureza da população que se quer estudar interfere, de modo direto, na escolha do tipo de amostragem

De modo geral, existem duas formas de amostragem: com reposição e sem reposição.

4.4.1 Amostragem sem reposição

Amostragem sem reposição é aquela em que o item retirado para estudo e análise não retorna à base de dados ou à população, de sorte que não pode ser novamente retirado.

STEVENSON (p. 159) aponta as seguintes razões pelas quais deve ser utilizada a amostra sem reposição:

- 1) Os efeitos são desprezíveis quando a amostra é pequena em relação à população;
- 2) Se o teste tem caráter destrutivo, é impossível repor os itens examinados;
- 3) Na amostragem industrial, pode ser difícil convencer os inspetores não treinados em estatística a reporem na população os itens examinados, principalmente os itens defeituosos;
- 4) Quando se repõe um item examinado na população, há chance de ele ser novamente escolhido em extração futura. Assim, alguns itens são

examinados mais de uma vez. Se o processo de amostragem é dispendioso, é conveniente evitar o exame repetido de um ou mais itens.

4.4.1 Amostragem com reposição

A amostragem com reposição é aquela em que após o exame feito no item colhido na amostra, o mesmo retorna imediatamente à população, podendo ser novamente extraído, inclusive de forma sucessiva.

SPIEGEL (p. 176) destaca um fato interessante: a amostragem com reposição em uma população finita faz com que ela, no aspecto teórico, se transforme em infinita, visto que qualquer número de amostras pode ser extraído sem exaurir a população.

5 ANÁLISE COMBINATÓRIA

LEMOS, HIGUCHI e FRIDMAN (p. 115) esclarecem que:

Dado um conjunto, existem várias maneiras de **agruparmos** os seus elementos. Estes **agrupamentos** podem ser formados levando-se em conta a **quantidade**, a **ordem** e a **natureza** dos elementos.

Na análise combinatória podemos encontrar arranjos simples, combinações simples e permutações, os quais tem como premissa de cálculo o fatorial, que consiste em um número n qualquer, inteiro e maior que a unidade. Representa-se assim:

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

5.1 Arranjos simples

Quando se quer calcular a quantidade de agrupamentos possíveis em um determinado conjunto, onde um grupo é diferente do outro pela **ordem** e pela **natureza** dos elementos, estamos diante de uma situação de **arranjos simples**.

Os arranjos simples são representados através da seguinte fórmula:

$$A_{n,p} = \frac{n!}{(n-p)!}$$

Aplicando-se a fórmula, podemos destacar o seguinte exemplo⁹:
quantos números de três algarismos, sem repetição, obtém-se com os algarismos do conjunto {1,2,3,4,5,6,7,8,9}?

⁹ Extraído de LEMOS, HIGUCHI e FRIDMAN (p. 121)

Solução:

a) Toma-se uma resposta qualquer, por exemplo: 679

b) Inverte-se a ordem de dois elementos: 976

c) Obtivemos uma nova resposta, portanto, trata-se de um problema de **arranjos simples**.

d) $A_{n,p} = \frac{n!}{(n-p)!}$ onde $n=9$ e $p=3$

e) Substituindo os valores de **n** e **p** temos: $A_{9,3} = \frac{9!}{(9-3)!} = \frac{9!}{6!} = \frac{9 \cdot 8 \cdot 7 \cdot 6!}{6!} = 504$

Portanto, existem 504 números de três algarismos sem repetição, formados com os elementos do conjunto dado.

5.2 Combinações simples

Na combinação simples, um grupo é diferente do outro apenas pela sua **natureza**. Assim, a ordem dos elementos não importa porque o grupo continua a ser o mesmo.

LEMOS, FUGUCHI e FRIDMAN (p. 119) ensinam que, *para calcular a quantidade de agrupamentos chamados **combinações simples** de n elementos tomados p a p , utilizamos a fórmula:*

$$C_{n,p} = \frac{n!}{p!(n-p)!}$$

Tomando uma situação exemplificativa¹⁰, pergunta-se: quantas comissões de 4 membros são possíveis de se formar com 10 indivíduos?

Solução:

Seja {A,B,C,D,E,F,G,H,I,J} o conjunto de 10 pessoas.

a) Toma-se uma resposta qualquer, por exemplo ABCD.

b) Inverte-se a ordem de dois elementos: BACD.

c) Obtivemos a mesma resposta (é a mesma comissão); portanto, trata-se de um problema de **combinação simples**.

d) $C_{n,p} = \frac{n!}{p!(n-p)!}$ onde $n = 10$ e $p = 4$

e) Substituindo os valores de **n** e **p** temos: $C_{10,4} = \frac{10!}{4!(10-4)!} = \frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{4 \cdot 3 \cdot 2 \cdot 1 \cdot 6!} = 210$

Portanto, podemos formar 210 comissões de 4 membros com um grupo de 10 indivíduos.

5.3 PERMUTAÇÃO SIMPLES

Há situações, em problemas de **arranjos**, em que o número de elementos dos agrupamentos é igual ao número total dos elementos do grupo, como explicitam LEMOS, HIGUCHI e FRIDMAN (p. 121). Nessa situação diz-se que $p = n$ e, essa particular situação do arranjo simples é denominada como **permutação simples**, que pode ser calculada por:

$$P_n = n!$$

¹⁰ Extraído de LEMOS, HIGUCHI e FRIDMAN (p. 122)

Tal qual em relação aos itens anteriores da análise combinatória, vamos aplicar um exemplo¹¹: quantos números de 4 algarismos, sem repetição, obtém-se com os algarismos do conjunto {1,2,3,4}?

Solução:

- a) Observe que $n = p = 4$.
- b) Trata-se de um problema de arranjos simples, mas como $n = p$, é um problema que rotulamos de **permutações simples**.
- c) $P_n = n!$, donde $P_4 = 4! = 4.3.2.1. = 24$.

Portanto, existem 24 números de quatro algarismos sem repetição formados com os elementos do conjunto dado.

¹¹ Extraído de LEMOS, HIGUCHI e FRIDMAN (p. 122

6 CONCLUSÃO

A estatística é uma poderosa ferramenta de análise, devendo fazer parte de qualquer processo decisório. Propicia economia de recursos em seu sentido indutivo posto que permite inferir sobre a população a partir de uma amostra e, bem assim, propicia subsídio numérico à escolha de determinado cenário em detrimento de outro.

No aspecto descritivo torna mais facilmente compreendidos os resultados colhidos em amostras ou sob qualquer outra forma de dados brutos, destacando-se aqui as representações tabulares e gráficas.

7 BIBLIOGRAFIA

COSTA, Sérgio Francisco. **Introdução Ilustrada à Estatística (com muito humor!)**. 2ª ed. São Paulo: Harbra, 1992.

GUERRA, Mauri José; DONAIRE, Denis. **Estatística Indutiva - teoria e aplicações**. 3ª ed. São Paulo: LCTE, 1986.

LEMOS, Aluisio Andrade. HIGUCHI, Fidefico. FRIDMAN, Salomão. **Matemática**. São Paulo: Moderna, 1976.

LIPSCHUTZ, Seymour. **Probabilidade**. 4ª ed. rev. São Paulo: Makron Books, 1993.

MEDRI, Waldir. **Métodos quantitativos aplicados à Contabilidade**. Londrina: UEL, 2003.

OLIVEIRA, Francisco Estevam Martins de. **Estatística e Probabilidade - Exercícios resolvidos e propostos**. 1ª ed. São Paulo: Atlas, 1995

_____. **Estatística e Probabilidade**. 2ª ed. São Paulo: Atlas, 1999.

SPIEGEL, Murray. **Estatística**. 2ª ed. São Paulo: Mc Graw-Hill do Brasil, 1985.

STEVENSON, William J. **Estatística aplicada à administração**. 1ª ed. São Paulo: 1981.