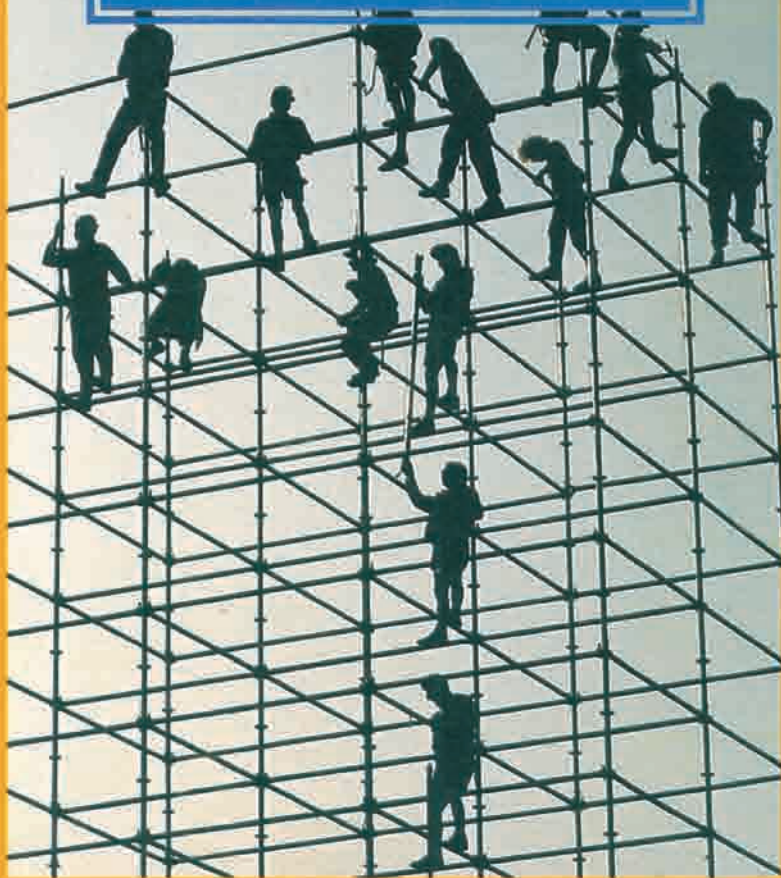


ESTATÍSTICA APLICADA À
SAÚDE
OCUPACIONAL



FUNDACENTRO
FUNDACENTRO
FUNDACENTRO


Brasil
EM AÇÃO

ESTATÍSTICA APLICADA À SAÚDE OCUPACIONAL

Presidente da República

Fernando Henrique Cardoso

Ministro do Trabalho

Paulo Paiva

FUNDACENTRO

Presidência

Humberto Carlos Parro

Diretoria Executiva

José Gaspar Ferraz de Campos

Diretoria Técnica

Sonia Maria José Bombardi

Diretoria de Administração e Finanças

Marco Antonio Seabra de Abreu Rocha

Assessoria de Comunicação Social

José Carlos Crozera

ESTATÍSTICA APLICADA À SAÚDE OCUPACIONAL

MARCO ANTONIO BUSSACOS



MINISTÉRIO DO TRABALHO

FUNDACENTRO

FUNDAÇÃO JORGE DUPRAT FIGUEIREDO
DE SEGURANÇA E MEDICINA DO TRABALHO

**São Paulo
1997**

Rua Capote Valente, 710
05409-002 - São Paulo-SP

CATALOGAÇÃO NA FONTE: DDB - FUNDACENTRO

Qyf A

B986e Bussacos, Marco Antonio

v. 1 Estatística aplicada à saúde ocupacional / Marco Antonio
Bussacos - São Paulo: FUNDACENTRO, 1997.

v. 1 : il.

103pp.

Inclui tabelas estatísticas e cálculos matemáticos. Referências
Bibliográficas no final de cada capítulo.

1. Métodos estatísticos - Medicina e higiene do trabalho
I. FUNDACENTRO II. Título.

CDU 311(001.8)

CIS Qyf A

Índice para o catálogo sistemático

1. Métodos estatísticos - Medicina e higiene
do trabalho

311(001.8):616-057*
Qyf A**

2. Medicina e higiene do trabalho - Métodos
estatísticos

616-057:311(001.8)*
A Qyf**

* Classificação Decimal Universal

** Classificação do "Centre International d'Informations de Sécurité et d'Hygiène du Travail"

Há mais de 10 anos em contato com os pesquisadores da FUNDACENTRO e de outras instituições, assessorando-os, principalmente na área de estatística, verifiquei a dificuldade que alguns pesquisadores têm em compreender a linguagem estatística, e a ausência de uma obra referente a Saúde Ocupacional, fez com que surgisse a idéia deste livro básico de estatística.

Os exemplos apresentados foram retirados de trabalhos desenvolvidos pelos pesquisadores da FUNDACENTRO e/ou de artigos publicados na Revista Brasileira de Saúde Ocupacional (RBSO), e adaptados, quando necessário. Evitei ser "muito" matemático, mas espero que também não tenha tornado este livro um "receituário". Para efetuar os cálculos apresentados é necessário apenas uma calculadora de bolso.

Um outro volume está sendo elaborado, com apresentação de técnicas estatísticas mais complexas, tais como, análise de regressão múltipla, regressão logística, clusters, análise fatorial e outras, mas com o mesmo objetivo, isto é, aplicar os conceitos estatísticos em exemplos extraídos da área de Saúde Ocupacional.

Finalmente, quero agradecer aos pesquisadores da FUNDACENTRO que muito me ajudaram e contribuíram para a elaboração deste livro, em especial aos mestres e críticos Dra. Leda Leal Ferreira e Dr. José Tarcísio P. Buschinelli.

Marco Antonio Bussacos
Agosto, 1994

SUMÁRIO

APRESENTAÇÃO

10 _____
11 _____
11 _____

16 _____

20 _____

23 _____

29 _____

30 _____

31 _____

35 _____

35 _____

36 _____

37 _____

39 _____

24 _____

53 _____

1 BÁSICO

1.1 Variável

1.2 Análise descritiva

1.2.1 Distribuição de frequência

1.2.2 Representação gráfica

1.3 Medidas de tendência central

1.3.1 Moda

1.3.2 Mediana

1.3.3 Média aritmética

1.4 Medida de dispersão

1.4.1 Variância

1.4.2 Desvio padrão

1.4.3 Coeficiente de variação

1.5 Intervalo de classe

1.5.1 Método de Sturges

1.5.2 Média

1.5.3 Moda

1.5.4 Mediana

1.5.5 Variância

1.5.6 Desvio padrão

1.5.7 Coeficiente de variação

Referências bibliográficas

2 DISTRIBUIÇÃO DE PROBABILIDADE

2.1 Distribuição normal

2.2 Distribuição t'Student

Referências bibliográficas

3 TESTE DE HIPÓTESES

3.1 Introdução

3.2 Teste de uma média

3.3 Comparação de duas médias

3.3.1 Dados não emparelhados

3.3.2 Dados emparelhados

3.4 Comparação de várias médias

3.4.1 Amostras de mesmo tamanho

3.4.2 Amostras de tamanhos diferentes

Referências bibliográficas

4 CORRELAÇÃO	54
4.1 Relação entre duas variáveis quantitativas	55
4.1.1 Diagrama de dispersão	
4.1.2 Coeficiente de correlação linear de Pearson	
Referências bibliográficas	59
5 ANÁLISE DE TABELA 2X2	60
5.1 Introdução	61
5.2 Qui-Quadrado (Prova de independência)	61
5.2.1 Qui-Quadrado (sem correção)	
5.2.2 Qui-Quadrado (com correção)	
5.2.3 Qui-Quadrado de Mantel Haenszel	63
5.3 Outras medidas de associação	
5.3.1 Coeficiente Phi	
5.3.2 Risco relativo	
5.3.3 Odds ratio	
Referências bibliográficas	67
6 TEORIA JÁ QUE	68
Referências bibliográficas	75
7 EXEMPLO DE APLICAÇÃO ESTATÍSTICA NA ÁREA DE PNEUMOLOGIA	76
7.1 Análise da curva de peak-flow	77
7.1.1 Análise gráfica	
7.1.2 Comparação de duas médias: período trabalho x afastado	
Referências bibliográficas	79
8 EXEMPLO DE PROGRAMA INTERLABORATORIAL DE CONTROLE DE QUALIDADE EM ANÁLISES TOXICOLÓGICAS	80
8.1 Introdução	81
8.2 Identificação de "outliers"	83
8.3 Estimativa da concentração média	85
8.4 Construção da carta de controle tipo Shewhart	86
Referências bibliográficas	88
9 USO DA DISTRIBUIÇÃO DE POISSON NO CONTROLE DE ANÁLISE DE FIBRAS INORGÂNICAS POR MICROSCOPIA ÓTICA	90
9.1 Introdução	91
9.2 Procedimentos para avaliação de fibras inorgânicas	91
9.3 Distribuição de Poisson	92
9.3.1 Exemplos de aplicação da distribuição de Poisson	
9.4 Comentário final	96
Referências bibliográficas	97

98	Anexos
99	Anexo I - Tabela da Distribuição Normal Reduzida
100	Anexo II - Tabela da Distribuição t'Student
101	Anexo III - Tabela da Distribuição de Qui-Quadrado
102	Anexo IV - Tabela da Distribuição de F'Snedecor

LISTA DE TABELAS

Tabela 1 - Distribuição dos pilotos segundo a alteração da escala dos últimos 15 dias

Tabela 2 - Distribuição percentual dos pilotos segundo a alteração da escala dos últimos 15 dias

Tabela 3 - Distribuição dos motoristas de ônibus segundo local de refeição, São Paulo, 1991

Tabela 4 - Distribuição dos trabalhadores acidentados segundo idade, Empresa X, 1994 (dados hipotéticos)

Tabela 5 - Distribuição dos trabalhadores segundo ramo de atividade, Cidade Y, 1994 (dados hipotéticos)

Tabela 6 - Distribuição dos trabalhadores segundo idade, da empresa X, 1994 (dados hipotéticos)

Tabela 7 - Distribuição dos trabalhadores segundo renda, em salário mínimo, da empresa Y, 1994

Tabela 8 - Distribuição dos trabalhadores segundo renda, em salário mínimo, da empresa Y, 1994

Tabela 9 - Distribuição dos trabalhadores segundo o tempo na função da empresa W

Tabela 10 - Distribuição de trabalhadores com saturnismo segundo tempo de exposição a poeiras e fumos de chumbo

Tabela 11 - Distribuição da jornada de trabalho (em horas) dos condutores, segundo tipo de empresa

Tabela 12 - Distribuição dos trabalhadores segundo o tipo de turno

Tabela 13 - Distribuição dos trabalhadores segundo os valores de ALA-U e Pbs

Tabela 14 - Distribuição dos trabalhadores segundo a exposição à vibração de máquina agrícola e queixa gástrica

Tabela 15 - Distribuição de casos e controles segundo tabagismo, Região Metropolitana de São Paulo, 1990/91

Tabela 16 - Distribuição dos resultados (duplicatas) da amostra X segundo os laboratórios participantes do programa

LISTA DE FIGURAS

Figura 1 - Distribuição percentual dos condutores segundo a variável "Sindicalização", São Paulo, 1991

Figura 2 - Distribuição das doenças profissionais registradas no Brasil, entre os anos de 1981 e 1990, por regiões geográficas

Figura 3 - Distribuição percentual de trabalhadores portadores de LER segundo faixa etária, atendidos no CRST-Lapa, São Paulo, out. 1990 a out. 1992

Figura 4 - Distribuição percentual de trabalhadores portadores de LER segundo faixa etária, atendidos no CRST-Lapa, São Paulo, out. 1990 a out. 1992

Figura 5 - Diagrama de dispersão

Figura 6 - Distribuição dos resultados de ALA-U e Pbs, segundo trabalhador

Figura 7 - Número de tabelas geradas segundo o número de variáveis

Figura 8 - Curva de Peak-Flow

Figura 9 - Carta de controle

1.1 Variável

Um estudo sobre os acidentes do trabalho será realizado por um pesquisador e um problema básico que se coloca numa pesquisa é definir quais variáveis irão participar do estudo. Entenda-se por **variável**, o conjunto de resultados possíveis de uma característica analisada. Assim, em um estudo sobre os acidentes do trabalho, uma variável a ser considerada seria, por exemplo, a “hora do acidente”.

A aplicação de técnicas e métodos estatísticos está condicionada aos tipos de variável: qualitativa não ordenável, qualitativa ordenável, quantitativa discreta e quantitativa contínua.

As variáveis **qualitativas** geralmente recebem o nome de atributos que representam um conjunto de categorias ou modalidades enquanto que as variáveis **quantitativas** representam um conjunto de números.

A variável sexo (masculino, feminino) é um exemplo de variável qualitativa não ordenável. Escolaridade (analfabeto, primário, secundário, colegial, universitário) é uma variável qualitativa ordenável. Número de Filhos (0, 1, 2, 3, ...) é uma variável quantitativa discreta enquanto que a idade (28 anos, 6 meses, 3 semanas, etc.) é uma variável quantitativa contínua.

Não existe fórmula matemática para definir quantas variáveis um estudo deve ter, mas, usando-se o bom senso e lembrando-se de que alguns softwares estatísticos definem o número máximo de variáveis, a pesquisa deve ter somente aquelas que realmente estão relacionadas ao objetivo do estudo. De nada adianta criar um instrumento de coleta de dados com 80, 100 ou 200 variáveis, pois, além de diminuir a qualidade das respostas, no final haverá uma grande quantidade de papéis para rascunho. (Ver Capítulo 6: Teoria Já Que)

Assim, ao definir o objetivo do estudo com bastante clareza, deve-se ter em mente quais variáveis o contemplam.

1.2 Análise descritiva

1.2.1 Distribuição de freqüências:

Em qualquer pesquisa é fundamental para o pesquisador ter uma visão da distribuição de freqüência das variáveis consideradas no estudo. E é por meio da análise descritiva, algumas vezes menosprezada, que se irá obter esta “visão”.

A **freqüência absoluta** de uma variável nada mais é que a

contagem de cada categoria (item) de uma variável, e a **freqüência relativa** é a porcentagem de cada categoria em relação ao total de elementos analisados.

No estudo de FERREIRA, utilizando uma enquete junto aos pilotos de uma empresa aérea, a tabela abaixo é um exemplo de uma distribuição de freqüência absoluta:

Tabela 1. Distribuição dos pilotos segundo a alteração da escala dos últimos 15 dias.

Função	Sim	Não	Total
Comandante	20	29	49
Co-Piloto	20	17	37
1° Oficial	4	7	11
Total	44	53	97

Fonte: FERREIRA, L.L.e alli.- Voando com os Pilotos, APVAR, 1992

Para a mesma tabela, porém com a distribuição relativa, tendo-se como base os 97 pilotos, ou seja, 100% da freqüência relativa:

Tabela 2. Distribuição percentual dos pilotos segundo a alteração da escala dos últimos 15 dias.

Função	Sim	Não	Total
Comandante	20,6	29,9	50,5
Co-Piloto	20,6	17,5	38,1
1° Oficial	4,1	7,2	11,4
Total	45,4	54,6	100,0

Fonte: FERREIRA, L.L.e alli.- Voando com os Pilotos, APVAR, 1992

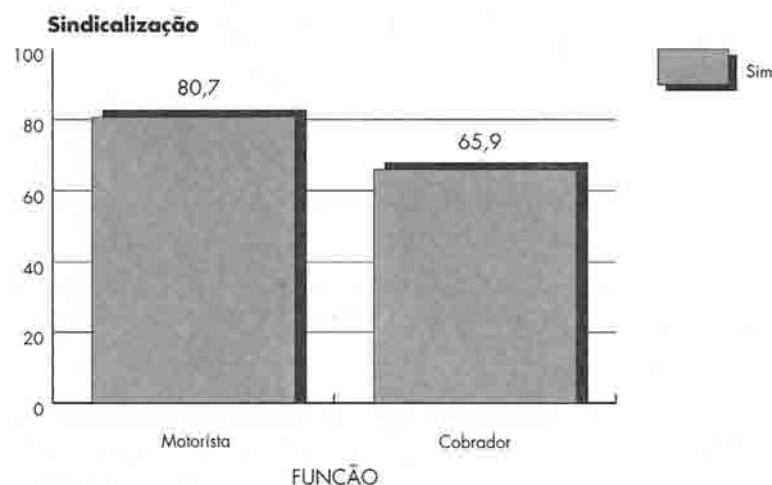
Neste exemplo dado, verificamos que quase a metade dos pilotos (45,4%) tiveram sua escala alterada nos últimos 15 dias, "índice que nos pareceu bastante elevado" (FERREIRA). Ou seja, esta simples análise descritiva apontou aos pesquisadores um indicio de como é a organização do trabalho desses profissionais.

1.2.2 Representação gráfica:

É uma das maneiras de apresentar os resultados da pesquisa. Porém, deve-se ter o cuidado de verificar que tipo de variável está sendo analisada. Quando a variável é qualitativa, os gráficos mais comumente utilizados são: diagrama de barras separadas e diagrama de círculos.

Diagrama de barras separadas:

No estudo realizado por FREIRE e col., analisando as condições de saúde e trabalho dos motoristas e cobradores urbanos de São Paulo, destacamos a variável "Sindicalização" dessa categoria como exemplo de um gráfico de barras separadas. A variável em questão é qualitativa, com duas categorias: Sim/Não.



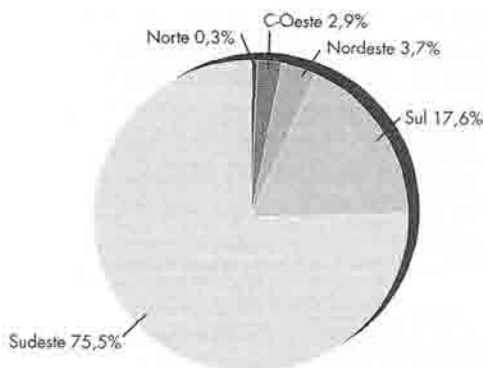
Fonte: FREIRE, N. B. e col. - Estudo das condições de saúde e trabalho, 1991

Figura 1. Distribuição percentual dos condutores segundo a variável Sindicalização, São Paulo, 1991.

Diagrama de círculos: ("pizza")

Na dissertação de mestrado de BUSCHINELLI, sobre a epidemiologia das doenças ocupacionais registradas no Brasil na década de 1980, destacamos a distribuição dos casos de doenças profissionais registradas segundo regiões geográficas no período analisado, na qual a região Sudeste, concentrou 75,5% dos casos, seguida da região Sul com 17,6%, como pode ser

verificado no gráfico abaixo:



Fonte: BUSCHINELLI, J. T. P., 1993

Figura 2. Distribuição das doenças profissionais registradas no Brasil, entre os anos de 1981 e 1990, por regiões geográficas.

Para as variáveis **quantitativas discretas**, basicamente se utilizam os mesmos tipos de gráficos das variáveis qualitativas. BERQUÓ apresenta em seu livro um capítulo dedicado aos diversos tipos de representações gráficas que um pesquisador pode aplicar.

Para as variáveis **quantitativas contínuas**, destacamos os seguintes modelos de gráficos mais frequentemente utilizados: histograma e gráfico de linhas.

Histograma

No estudo de SATO e col., sobre a dimensão psicossocial da LER, a partir de atividade em grupo com portadores, desenvolvida em serviços públicos de saúde do trabalhador, destacamos a distribuição da faixa etária dos portadores de LER, atendidos no período de 1990/92.

Fonte: SATO, L. e col., 1993

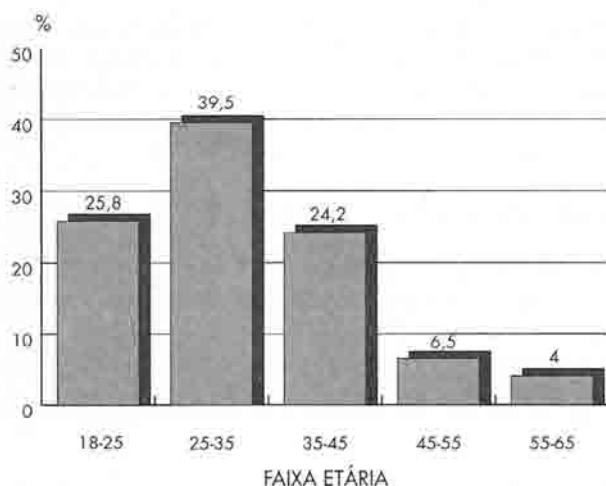
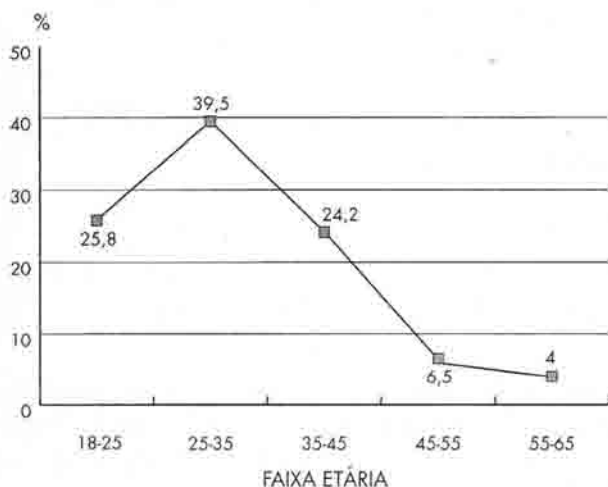


Figura 3. Distribuição percentual de trabalhadores portadores de LER segundo faixa etária, atendidos no CRST-Lapa, São Paulo, outubro de 1990 a outubro de 1992.

Para o mesmo exemplo, em vez de utilizar o histograma, poderíamos fazer um gráfico de linhas:



Fonte: SATO, L. e col., 1993

Figura 4. Distribuição percentual de trabalhadores portadores de LER segundo faixa etária, atendidos no CRST-Lapa, São Paulo, outubro de 1990 a outubro de 1992.

Com acesso aos diversos softwares gráficos nos microcomputadores, o pesquisador que não estiver atento ao tipo da variável poderá incorrer no erro de misturar os diversos tipos de gráficos, sem considerar a categoria da variável, criando um carnaval de figuras ou cores.

1.3 Medidas de tendência central

O objetivo de utilizar as medidas de tendência central é o de caracterizar o centro da distribuição de uma variável. As principais medidas utilizadas são a Moda, Mediana e Média.

O mesmo cuidado que se tem na representação gráfica, deve-se ter também nas medidas de tendência central em relação ao tipo da variável.

1.3.1 Moda: Mo

Quando a variável for **qualitativa**, a única medida que se pode utilizar é a moda. Esta medida, como o próprio nome diz, é a categoria da variável mais freqüente numa distribuição, ou seja, é o valor da variável mais comum.

A moda pode ser **unimodal**, como se pode verificar no exemplo de FREIRE e col.:

Tabela 3. Distribuição dos motoristas de ônibus segundo local de refeição, São Paulo, 1991.

Local	Nº
No bar	169
No ônibus	125
Em casa	78
Não comeu	64
Outros	28
Total	464

Fonte: FREIRE, N.B e col-Estudo das condições de saúde e trabalho, 1991

Neste exemplo, a moda do local de refeição é "No bar", ou seja, é a categoria da variável que apresentou a maior freqüência ($f=169$), indicando que o mais comum é os motoristas fazerem suas refeições no bar.

A moda pode ser **bimodal**:

Tabela 4. Distribuição dos trabalhadores acidentados segundo idade, empresa X, 1994 (dados hipotéticos).

Idade	Nº
18	5
20	30
24	30
25	8
Total	73

Neste exemplo, há duas idades com a maior frequência ($f = 30$), que são 20 e 24 anos, assim, esta distribuição é bimodal.

A moda pode ser poli-modal (três ou mais) ou pode ocorrer de não haver moda; neste caso, a distribuição é **amodal**, ou seja, quando todos os valores da frequência das categorias da variável são iguais, como se pode verificar no seguinte exemplo:

Tabela 5. Distribuição dos trabalhadores segundo ramo de atividade, cidade Y, 1994 (dados hipotéticos).

Ramo	Nº
Metalúrgico	20
Bancário	20
Serviços	20
Comércio	20
Total	80

Observa-se que os quatro ramos de atividades apresentam a mesma distribuição, assim, esta distribuição é **amodal**.

Pelos exemplos apresentados para moda, verifica-se que esta medida pode ser utilizada tanto para variável qualitativa como para variável quantitativa.

1.3.2 Mediana: Md

Uma outra medida de tendência central é a mediana cuja finalidade é a de dividir a distribuição de frequência em duas partes, em que 50% da frequência está abaixo do valor mediano e 50% acima, ou seja, o valor mediano divide ao meio a distribuição de uma variável, que deve ser **quantitativa**.

Para realizar o cálculo da mediana, o seguinte procedimento deve ser seguido:

a) verificar se os valores da variável estão em ordem crescente,

caso contrário, ordená-los.

b) construir uma frequência acumulada da distribuição (Fac).

c) encontrar a posição da mediana:

se **n**, total da frequência, for um número **ímpar**, a posição

$$\text{será: } p = \frac{(n + 1)}{2}$$

se **n** for um número **par**, há duas posições:

$$p_1 = \frac{n}{2} \text{ e } p_2 = \frac{(n + 2)}{2}$$

d) localizar a posição na frequência acumulada e, assim, obter o valor mediano da variável, como se pode verificar nos exemplos a seguir.

Tabela 6. Distribuição dos trabalhadores segundo idade, da empresa X, 1994 (dados hipotéticos).

Idade	Nº	Fac
18	5	5
20	20	25
25 ← Md	30	55
30	8	63
40	4	67
Total	67	

Os valores da variável estão em ordem crescente. Constrói-se a frequência acumulada (que é a soma acumulativa da frequência simples) e para encontrar a posição da mediana, verifica-se que **n**, total da frequência, é um número ímpar ($n = 67$), assim a posição será:

$$p_2 = \frac{(n + 1)}{2} = \frac{(67 + 1)}{2} = 34$$

Para localizar essa posição na frequência acumulada, verifica-se em que linha dessa frequência se encontra a posição 34. Nesse exemplo, está na 3ª linha, que inclui do 26º ao 55º. Buscando a correspondência na coluna idade, o valor mediano será igual a 25 anos.

Tabela 7. Distribuição dos trabalhadores segundo renda (em salário mínimo), da empresa Y, 1994.

S.M.	Nº	Fac
1	15	15
2 ←	19	34 ←
3 ←	21	55 ←
4	9	64
5	4	68
Total	68	

Neste exemplo, o total da frequência é um número par, assim, para encontrar o valor da posição da mediana na frequência acumulada, tem-se:

$$p = \frac{n}{2} = \frac{68}{2} = 34 \text{ e } p_2 = \frac{(n+2)}{2} = \frac{(68+2)}{2} = 35$$

Para a primeira posição, encontramos o valor da variável de 2 s.m. e o valor da variável da segunda posição é igual a 3 s.m., assim, o valor mediano desta distribuição será a soma dos dois valores da variável dividido por 2, isto é,:

$$Md = \frac{2 + 3}{2} = 2,5 \text{ s. m.}$$

1.3.3 Média aritmética: (\bar{x})

Uma outra medida de tendência central é a média, um dos parâmetros mais utilizados no dia a dia. Esta medida sempre existe e quando calculada admite um único valor, porém sofre influência de valor aberrante, ou seja, se numa distribuição houver um valor da variável discrepante dos demais, com baixa frequência, a média será atraída para este valor. Deve ser utilizada em variáveis **quantitativas**.

Para calcular o valor da média, aplica-se a seguinte fórmula:

$$\bar{x} = \frac{\sum x \cdot f}{n}$$

Para o seguinte exemplo, iremos calcular o valor do salário médio (\bar{x}):

Tabela 8. Distribuição dos trabalhadores segundo renda (em salário mínimo) da empresa Y, 1994.

S.M. (x)	No (f)	x • f
1	15	15
2	19	38
3	21	63
4	9	36
5	4	20
Total	68	172

Assim, o salário médio será:

$$\bar{x} = \frac{172}{68} = 2,53 \text{ s. m.}$$

1.4 Medidas de dispersão

Só com o valor da média não se pode ter uma idéia de como se comporta a distribuição de uma variável. Por exemplo, duas empresas A e B com a mesma renda média por operário. Este simples fato de igualdade das duas médias permite concluir que a situação econômica dos operários das duas empresas são iguais? Evidentemente que não, pois esta igualdade poderia existir mesmo que a empresa A fosse perfeitamente estabilizada no sentido de que todos os operários tivessem praticamente o mesmo salário e na empresa B houvesse uns poucos operários com salários muito altos e a maioria recebesse salários muito baixos.

Isso significa que é necessário conhecer a variabilidade da distribuição da variável, o que nos permite concluir que na empresa A não há variação do salário de um operário para outro enquanto que na empresa B, há uma grande variação.

Para obter esta variação tem-se as medidas de dispersão. As medidas mais comumente utilizadas são: variância, desvio padrão e coeficiente de variação.

1.4.1 Variância:

Esta medida mede a dispersão em torno da média. Se todos os valores da distribuição forem iguais, não existirá variabilidade, e o valor da variância será igual a zero. A **variância**

populacional (σ^2) é utilizada quando se trabalha com todos os elementos da população. Basicamente é pouco utilizada, já que frequentemente as pesquisas trabalham com amostras (subconjunto da população). A **variância amostral** (s^2) é a melhor estimativa da variância populacional, sendo a que se utiliza nas pesquisas.

Fórmula para variância populacional:

$$\sigma^2 = \frac{\sum x^2 \times f - \frac{(\sum x \times f)^2}{n}}{n}$$

Fórmula para a variância amostral:

$$S^2 = \frac{\sum x^2 \times f - \frac{(\sum x \times f)^2}{n}}{n - 1}$$

1.4.2 Desvio padrão:

O problema na utilização da variância é que no seu resultado, a unidade da variável (por exemplo, anos, cm, kg, etc) é dada ao quadrado. Assim, uma comparação do resultado da variância com o valor da média não é direta. Para contornar este problema é que aplicamos o desvio padrão (s), que nada mais é do que a raiz quadrada do resultado obtido pela variância.

O desvio padrão é um parâmetro bastante utilizado e todas as vezes em que é calculada a média de uma distribuição, esta deve estar acompanhada do desvio padrão, pois esta simples informação já indica se a distribuição é pouco ou bastante dispersa.

1.4.3 Coeficiente de variação:

O coeficiente de variação (**cv**) é outra medida de dispersão cujo objetivo é o de classificar a variabilidade da distribuição em termos percentuais (%), e cuja fórmula é:

$$cv = \frac{s}{\bar{x}} \times 100, \text{ onde } s = \text{desvio padrão} \quad \bar{x} = \text{média}$$

Este coeficiente classifica a distribuição em três categorias:

a) se o resultado do coeficiente de variação for menor que 10% (**cv < 10%**), isto significa que há pouca variabilidade na

distribuição, ou seja, todos os valores da variável estão bem próximos do valor da média. Neste caso, a distribuição é **homogênea**.

b) se o coeficiente de variação estiver entre 10% e 30%, ($10\% < cv < 30\%$), significa que há uma **média dispersão** da distribuição da variável.

c) se o coeficiente de variação for maior que 30% ($cv > 30\%$), existe uma grande variabilidade, ou seja, os resultados da distribuição estão distantes da média e, nesta situação, tem-se uma distribuição **heterogênea**.

Para os exemplos hipotéticos apresentados a seguir, iremos apresentar os cálculos para variância, desvio padrão e coeficiente de variação:

Exemplo 1

Distribuição dos trabalhadores segundo renda (em salário mínimo) da empresa Y, 1994.

S.M.(x)	f	x • f	x ² • f
1	15	15	15
2	19	38	76
3	21	63	189
4	9	36	144
5	4	20	100
Total	68	172	524

a média encontrada nesta distribuição é :

$$\bar{x} = \frac{172}{68} = 2,53 \text{ s. m.}$$

Variância:

$$s^2 = \frac{524 - \frac{(172)^2}{68}}{67} = 1,33 \text{ s. m.}^2$$

Desvio padrão:

$$s = \sqrt{1,33} = 1,15 \text{ s. m.}$$

Coeficiente de variação:

$$cv = \frac{1,15}{2,53} \times 100 = 45,4\%$$

Exemplo 2

Para este exemplo, analisando a idade dos trabalhadores da empresa Z, obteve-se:

Idade(x)	f	x • f	x ² • f
30	100	3.000	90.000
Total	100	3.000	90.000

verifica-se que nesta distribuição não há variação dos valores da idade, ou seja, todos os trabalhadores tem a mesma idade, assim,

$$\text{média: } \bar{x} = \frac{3.000}{100} = 30 \text{ anos} \quad \text{desvio padrão } s = \sqrt{0} = 0$$

$$\text{variância: } s^2 = \frac{90.000 - \frac{(3.000)^2}{100}}{99} = 0$$

coeficiente de variação:

$$cv = \frac{0}{30} \times 100 = 0\%$$

1.5 Intervalo de classe:

Nos exemplos anteriores para as medidas de tendência central e dispersão, as variáveis eram do tipo quantitativas discretas, ou seja, os resultados eram pontuais. Porém, é pouco freqüente que o pesquisador encontre esta situação a não ser em amostras pequenas.

É usual trabalhar com os resultados em intervalos de classes, principalmente quando há grande quantidade de elementos na distribuição analisada. Para a construção de uma distribuição de freqüência de uma variável em intervalo de classe, deve-se procurar trabalhar com intervalos do mesmo tamanho (amplitude), e com um número reduzido de intervalos de classes (no máximo 8 intervalos).

Um estudo realizado na empresa (dados hipotéticos), obteve-se os seguintes valores sobre o tempo na função (em meses) de um grupo de trabalhadores da empresa W:

10 10 15 15 15 18 18 20 20 20
 25 25 25 28 28 29 29 30 30 32
 35 35 36 36 37 38 39 39 40 40

Se fosse construída uma tabela de frequência, com os valores pontuais, ter-se-ia uma tabela "bem comprida" e de difícil visualização. Neste caso, é melhor apresentar a distribuição em intervalo de classe. Mas como fazê-lo?

Há várias normas para se construir uma distribuição em intervalo de classe. Por exemplo, quando se trabalha com a variável renda, é importante utilizar a classificação do IBGE, pois esta facilita a comparação dos resultados com as regiões e/ou cidades do país. Um outro método para a construção desse intervalo é fornecido por Sturges (SPIEGEL):

1.5.1 Método de Sturges:

Este método determina o número de classes k (número de linhas) que a distribuição terá, utilizando-se a seguinte fórmula:

$$k = 1 + 3,3 \times \log N$$

onde k = número de classes a ser determinado e
 N = número total de elementos

Por exemplo, $N = 30$ e $k = 1 + 3,3 \times \log 30 = 5,87 = 6$ classes. Uma vez definido o número de classes, será determinada a amplitude (h), ou seja, o tamanho de cada classe: $h = (\text{Maior valor} - \text{Menor valor}) / k$

No estudo sobre o tempo na função, observamos que o menor valor foi de 10 meses enquanto que o maior valor foi de 40 meses. Assim, a amplitude do intervalo será:

$$h = \frac{(40 - 10)}{6} = 5$$

Para construir a distribuição em intervalo de classe, inicia-se com o menor valor (10 meses) ao qual é somada a amplitude ($h = 5$) e assim sucessivamente. Para cada intervalo de classe, a frequência é construída contando-se o número de elementos dentro de cada classe, ou seja,

Tabela 9. Distribuição dos trabalhadores segundo o tempo na função da empresa W.

Meses (x)	Nº (f)
10- 15	2
15- 20	5
20- 25	3
25- 30	7
30- 35	3
35- 40	10
Total	30

Um estudo realizado, por exemplo, com prontuários médicos de trabalhadores com diagnóstico de Saturnismo, apresentou a seguinte distribuição do tempo de exposição (em anos) a poeiras e fumos de chumbo: (dados hipotéticos)

Tabela 10. Distribuição de trabalhadores com saturnismo segundo tempo de exposição a poeiras e fumos de chumbo.

Anos (x)	f
0- 2	8
2- 4	15
4- 6	7
6- 8	4
Total	34

Qual seria o tempo médio de exposição desses trabalhadores? E a dispersão dessa distribuição? Quando o conjunto de dados está apresentado numa distribuição de intervalos de classes, os cálculos para as medidas de tendência central e dispersão são:

1.5.2 Média: (\bar{x})

Para calcular a média, utiliza-se um artifício que é concentrar a frequência de classe em valor do intervalo e o melhor valor é o ponto médio (pm) (BERQUÓ), definido pela soma dos limites do intervalo (inferior + superior) dividido por 2:

Os valores do ponto médio do exemplo sobre saturnismo são:

Primeiro intervalo: $(0 + 2) \div 2 = 1$

Segundo intervalo: $(2 + 4) \div 2 = 3$

Terceiro intervalo: $(4 + 6) \div 2 = 5$

Último intervalo: $(6 + 8) \div 2 = 7$

A fórmula da média em intervalo de classe é: $\bar{x} = \frac{\sum pm \times f}{n}$

Assim, tem-se:

Anos (x)	f	pm	pm • f
0 - 2	8	1	8
2 - 4	15	3	45
4 - 6	7	5	35
6 - 8	4	7	28
Total	34		116

O tempo médio de exposição a poeiras e fumos de chumbo para esses trabalhadores é

$$\bar{x} = \frac{116}{34} = 3,4 \text{ anos}$$

1.5.3 Moda (Mo):

Para calcular o **tempo modal** (Mo) de exposição, a única restrição é que a distribuição de frequência da variável seja unimodal. Primeiramente, identifica-se qual o intervalo que apresenta a maior frequência e utiliza-se a seguinte fórmula;

$$Mo = Li + \frac{D1}{(D1 + D2)} \times Hmo$$

onde **Li** = limite inferior da classe modal, **D1** = diferença entre o valor da frequência da classe modal com o valor da frequência da classe anterior, **D2** = diferença entre o valor da classe modal com o valor da frequência da classe posterior e **Hmo** = amplitude da classe modal.

No exemplo sobre tempo de exposição a poeiras e fumos de chumbo, encontramos o seguinte valor da moda:

Anos (x)	f
0 - 2	8
2 - 4	15
4 - 6	7
6 - 8	4
Total	34

A maior frequência desta distribuição é 15, que corresponde ao intervalo de 2 a 4 anos de exposição, ou seja, este é o intervalo modal e para conhecer o valor pontual da moda, aplica-se a fórmula:

$$\begin{aligned}Li &= 2 \\D1 &= 15 - 8 = 7 \\D2 &= 15 - 7 = 8 \\Hmo &= 4 - 2 = 2\end{aligned}$$

Assim

$$Mo = 2 + \frac{7}{(7 + 8)} \times 2 = 2,9 \text{ anos}$$

1.5.4 Mediana: Md

O tempo mediano de exposição será calculado utilizando-se uma fórmula, porém com os seguintes procedimentos:

- verificar se os intervalos estão em ordem crescente.
- construir a frequência acumulada (Fac)
- encontrar a posição da mediana:

se n for **par**, a posição será: $p = \frac{n}{2}$

se n for **ímpar**, $p = \frac{(n + 1)}{2}$

- localizar a posição na frequência acumulada e aplicar a seguinte fórmula, baseando-se nos valores da classe mediana:

$$Md = Li + \frac{\left(\frac{n}{2} - Fa\right) \times h}{Fmd} \quad \text{onde:}$$

Li = limite inferior da classe mediana

n = número total de elementos

Fa = valor da frequência acumulada antes da classe mediana

h = amplitude da classe mediana

Fmd = frequência simples da classe da mediana.

Ainda sobre o exemplo do tempo de exposição a poeiras e fumos de chumbo, tem-se:

Anos(x)	f	Fac
0 - 2	8	8
2 - 4	15	23 ←
4 - 6	7	30
6 - 8	4	34
Total	34	

Posição da Mediana: $n = 34$ (par), então $p = \frac{34}{2} = 17$ e o intervalo mediano é o segundo, pois antes deste na frequência acumulada, há 8 elementos da distribuição e o valor 17 está contido no intervalo que vai de 2 a 4 anos, assim, para utilizar a fórmula da mediana, trabalhamos com os valores deste intervalo:

$$Li = 2$$

$$Fa = 8$$

$$h = 4 - 2 = 2$$

$$Fmd = 15$$

$$Md = 2 + \frac{(\frac{34}{2} - 8) \times 2}{15} = 3,2 \text{ anos}$$

1.5.5 Variância: $\{s^2\}$

Para encontrar a variabilidade desta distribuição, aplicaremos a variância cuja alteração na fórmula é o ponto médio:

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

Assim,

Anos(x)	f	PM	PM x f	PM ² x f
0 - 2	8	1	8	8
2 - 4	15	3	45	135
4 - 6	7	5	35	175
6 - 8	4	7	38	196
Total	34		116	514

logo:

$$\frac{514 - \frac{(116)^2}{34}}{33} = 3,58 \text{ anos}^2$$

1.5.6 Desvio Padrão (s):

Sendo este parâmetro definido como a raiz quadrada do valor da variância, tem-se:

$$s = \sqrt{3,58} = 1,89 \text{ anos}$$

1.5.7 Coeficiente de variação (cv):

Fornecerá a variabilidade em termos percentuais e, neste exemplo, será:

$$cv = \frac{1,89}{3,4} \times 100 = 55,58\%$$

Referências bibliográficas

BERQUÓ, E.S., SOUZA, J.M.P., GOTLIEB, S.L.D. Bioestatística. São Paulo : EPU, 1980.

BUSCHINELLI, J.T.P. Epidemiologia das doenças profissionais registradas no Brasil na década de 80. Dissertação (Mestrado em Saúde Ocupacional) - Faculdade de Saúde Pública, Universidade de São Paulo, 1993.

FERREIRA, L., BUSSACOS, M.A., SCHLITER, C.R.S. et al. Voando com os pilotos: condições de trabalho dos pilotos de uma empresa de aviação comercial. São Paulo: Associação dos Pilotos da VARIG : FUNDACENTRO, 1992. 89p.

FREIRE, B., SOUZA, M.F.M., UEHARA, M.K. et al. Estudo das condições de saúde e trabalho dos motoristas e cobradores urbanos da cidade de São Paulo. São Paulo : FUNDACENTRO, 1991.

SATO, L., ARAÚJO, M.D., UDIHARA, M.L. et al. Atividade em grupo com portadores de LER e achados sobre a dimensão psicossocial. Revista Brasileira de Saúde Ocupacional, v.21, n.79, p.49-62, jul./set. 1993.

SPIEGEL, M.R. Estatística. 3.ed. São Paulo : Makrow Books, 1993.

2 DISTRIBUIÇÃO DE PROBABILIDADE

2.1 Distribuição normal: $N(\mu, \sigma)$

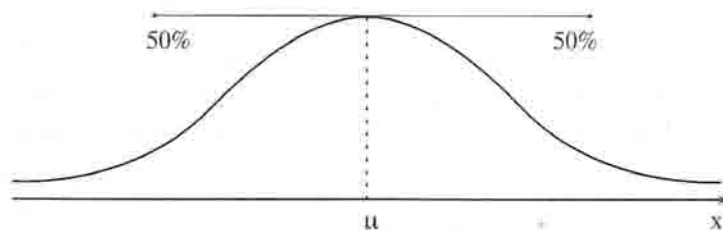
É uma das mais importantes distribuições de probabilidade, e vários testes estatísticos são aplicados quando o conjunto de dados analisados tem distribuição Normal.

A característica fundamental desta distribuição é ser simétrica, ou seja, a média, a moda e a mediana são iguais e os parâmetros que a identificam são a média (μ) e o desvio padrão (σ).

A distribuição ou curva normal (ou de Gauss) é definida pela seguinte função;

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

e a curva correspondente é:



Sob a curva há 100% dos elementos da distribuição. Para se obter a probabilidade, calcula-se a integral desta função ($f(x)$) ou se utiliza uma maneira mais prática, que é trabalhar com a distribuição da Normal reduzida, na qual a média é igual a zero e o desvio padrão é igual a 1. Após a transformação da Normal para a Normal reduzida, utiliza-se a tabela (anexo I) para encontrar as probabilidades solicitadas.

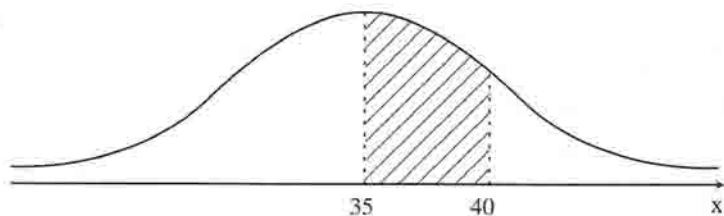
Fórmula de transformação:

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

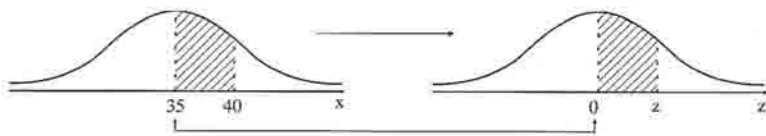
Tomando-se como exemplo um estudo realizado com 100 trabalhadores intoxicados por praguicidas derivados de compostos fosforados obteve-se, entre diversas variáveis, uma idade média dos trabalhadores de 35 anos com desvio padrão de 8 anos. Qual seria a porcentagem de trabalhadores com idade entre 35 e 40

anos, supondo-se uma distribuição Normal? Em termos probabilísticos, têm-se: $P(35 < X < 40)$?

Graficamente, temos a seguinte situação:



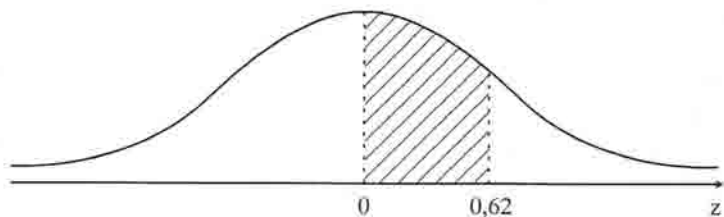
Para encontrar esta probabilidade (área), é necessário transformar o conjunto de dados para a Normal reduzida, ou seja,



O valor da média (35 anos), na Normal reduzida sempre valerá zero e para se encontrar o valor 40 na Normal reduzida, aplica-se a fórmula de transformação;

$$z = \frac{x - \mu}{\sigma} = \frac{40 - 35}{8} = 0,62$$

Graficamente, tem-se :

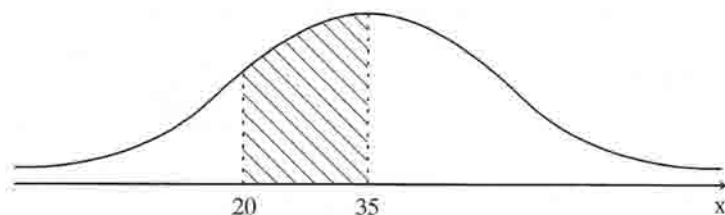


Para se encontrar o valor desta área, utiliza-se a tabela da Normal (Anexo I), em que se verifica, na primeira coluna a parte inteira e a primeira decimal de z (no exemplo, 0,6) e procura-se a segunda decimal (no exemplo, 2) e no cruzamento dos valores, obtém-se a área procurada:

			↓ 2
→	0,6		23237

Assim, o valor obtido na tabela : 0,23237 é a probabilidade, ou seja, há 23,2% de trabalhadores com idade entre 35 a 40 anos. ($P(35 < X < 40) = 0,23237$).

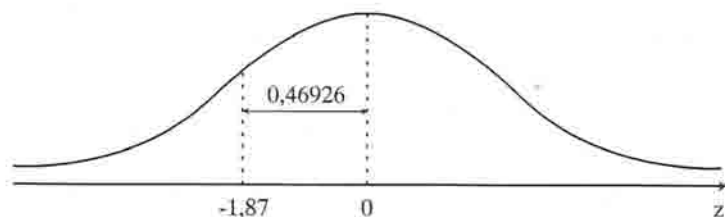
Entre 20 a 35 anos? $P(20 < X < 35)$?



Utilizando a fórmula da transformação da Normal reduzida para $X=20$ tem-se:

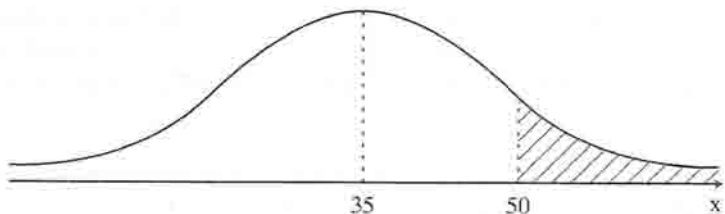
$$z = \frac{20 - 35}{8} = -1,87$$

cuja área vale 0,46926, graficamente assim representada:



Sendo assim, há 46,9% de trabalhadores com idade entre 20 a 35 anos. E se fosse perguntado qual a percentagem de trabalhadores com idade igual ou maior que 50 anos? $P(X \geq 50) = ?$

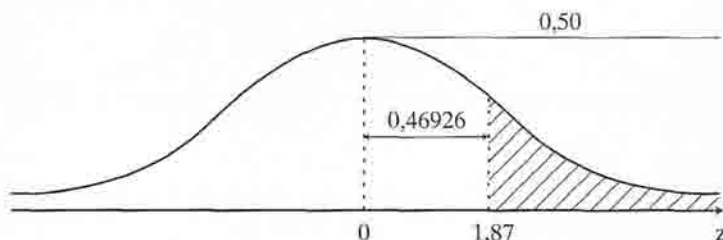
Graficamente



Para $X=50$ tem-se na Normal reduzida $z = \frac{50 - 35}{8} = 1,87$,
cuja área vale: 0,46926

Mas, como a área de interesse é a extremidade da direita, e a probabilidade fornecida pela tabela é a área entre a média e o valor, deve-se então subtrair o valor 0,5

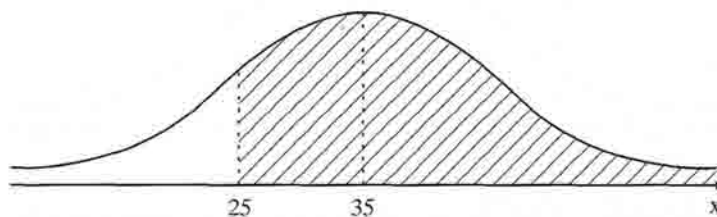
Graficamente



$P(X > 50) = 0,50 - 0,46926 = 0,0307$ ou 3,07% de trabalhadores tem idade acima de 50 anos.

Finalmente, qual é a porcentagem de trabalhadores com idade acima ou igual a 25 anos? $P(X \geq 25) = ?$

Graficamente,



Para $X = 25$ tem-se o seu valor na Normal reduzida,

$$z = \frac{25 - 35}{8} = -1,25$$

cuja área fornecida pela tabela é de 0,39435. Porém, pelo gráfico acima, verifica-se que está incluída a área à direita da média, que vale 0,5 ou 50%, assim o valor da probabilidade será acrescido de 0,50.

$P(X \geq 25) = 0,39435 + 0,50 = 0,89435$ ou 89,4% dos trabalhadores tem idade igual ou acima de 25 anos.

2.2 Distribuição t'Student

É outra distribuição de probabilidade importante, principalmente para os testes de hipóteses de comparação de médias populacionais. (ver capítulo 3 - Teste de Hipóteses). Quando o tamanho da amostra aumenta, esta distribuição se aproxima da Normal. Para consultar a tabela de t'Student (Anexo II), deve-se definir o parâmetro graus de liberdade (ν) e o nível de significância. (Conforme o tipo de teste estatístico o valor do grau de liberdade é definido de uma maneira, como será visto no capítulo 3).

Referências bibliográficas

BUSSAB, W.O., MORETTIN, P.A. Estatística básica. 3.ed. São Paulo : Atual, 1986.

3 TESTE DE HIPÓTESES

3.1 Introdução

Ao realizar um estudo, o pesquisador está constantemente formulando hipóteses sobre os fatos que ocorrem na natureza ou sobre as possíveis relações existentes entre eles. Por exemplo:

1) a pesquisa desenvolvida por MENDONÇA, teve como objetivo conhecer a prevalência de alterações funcionais pulmonares e hiper-reatividade brônquica inespecífica em trabalhadores expostos a produtos de exaustão de motores diesel de um terminal de ônibus, e compará-los a um grupo de trabalhadores não-expostos a esses produtos. 2) No estudo "Voando com os pilotos" de FERREIRA e col., entre as diversas análises, havia o objetivo de verificar se o tempo fora da base (período em que o piloto ficou fora de sua base, incluindo o tempo de viagem e o repouso) era igual ou diferente segundo os tipos de vôos: nacionais e internacionais. 3) No trabalho realizado por RODRIGUES, no qual analisou 1.606 exames de monitoramento biológico, como fenol urinário, de trabalhadores com risco de exposição ao benzeno, um dos objetivos era o de comparar a média do indicador biológico com o limite de tolerância biológico. Nos três exemplos citados, o objetivo é por à prova algumas hipóteses, entre elas, as hipóteses estatísticas.

Hipótese estatística é uma afirmação qualquer sobre a forma ou sobre os parâmetros de uma distribuição. Assim, no primeiro exemplo, compararam-se as médias da curva dose-resposta de carbacol, a fim de se verificar se os dois grupos (exposto/não-exposto) eram iguais ou diferentes. No segundo exemplo, foi observado se o tempo médio fora da base dos vôos nacionais era igual ou diferente do vôos internacionais. E no terceiro exemplo, comparou-se o valor médio encontrado do indicador biológico com o valor do limite de tolerância, para verificar se houve igualdade ou diferença.

Note que em todos os exemplos citados aparecem duas hipóteses contrárias e complementares. Uma afirma a existência de um certo tipo de relação, enquanto que a outra nega a existência dessa relação. Ambas são necessárias porque, uma delas vai ser posta à prova e a outra servirá de alternativa no caso de a primeira ser rejeitada.

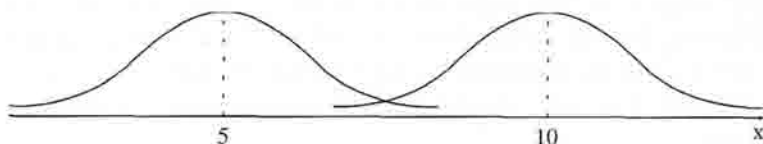
Toda hipótese que afirma uma relação de igualdade, chama-se

hipótese nula (H_0). A hipótese nula é a que vai ser posta à prova. E toda hipótese que estabelece uma relação de desigualdade, chama-se **hipótese alternativa (H_a).** Normalmente é a hipótese nula que é testada, e caso seja rejeitada, o pesquisador usará a hipótese alternativa. Mas como decidir na opção das hipóteses?

Suponha que a variável tempo na função (em anos) de um setor da indústria esteja sendo analisada e há dúvidas com relação ao valor do tempo médio na função. Alguns afirmam que é próximo de 5 anos, enquanto que outros afirmam que a média está próxima de 10 anos. Pode-se construir duas hipóteses: (supondo-se uma distribuição Normal)

$$\begin{aligned} H_0: \mu &= 5 \text{ anos} \\ H_a: \mu &= 10 \text{ anos} \end{aligned}$$

Graficamente, tem-se a seguinte situação:



Sendo a hipótese nula a "verdadeira", é válido o gráfico da esquerda, e se a hipótese alternativa for a "verdadeira", o gráfico correto é o da direita. Mas como tomar a decisão?

A única solução possível é estabelecer um critério probabilístico, pois qualquer valor da variável escolhido ao acaso pode pertencer à duas curvas. E supondo-se a hipótese nula verdadeira, estabelece-se um valor crítico, com uma probabilidade pequena de ocorrer (α), se o valor da variável, sorteado ao acaso, for maior que este valor crítico, rejeita-se a hipótese nula. A probabilidade fixada α chama-se nível de significância. Os níveis de significância usuais são de 5%, 1% e 0,1%.

Porém, podem-se cometer erros na rejeição ou não das hipóteses. Se rejeitar a hipótese nula sendo ela a "verdadeira", tem-se o erro tipo I e a probabilidade α , nível de significância, que é a probabilidade de cometer este erro. Por outro lado, se a hipótese nula não for rejeitada, sendo ela "falsa", tem-se o erro tipo II, representado pela probabilidade β . À probabilidade $(1-\beta)$ dá-se o nome de poder do teste.

A situação ideal seria que as duas probabilidades de cometer os erros fossem pequenas porém essa situação é impossível. Assim, fixa-se um valor de α e procura-se tornar máximo o poder do teste, aumentando o tamanho da amostra.

A seguir, serão apresentados alguns testes de hipóteses, mais comumente utilizados em saúde ocupacional:

3.2 Teste de uma média

Ao se comparar uma média obtida na amostra com um valor prefixado, é comum não se conhecer o desvio padrão populacional (σ). A distribuição de probabilidade a ser utilizada no teste depende do conhecimento ou não do desvio padrão populacional (σ). Assim, se o desvio padrão é conhecido, situação bastante rara, a distribuição de probabilidade utilizada é a Normal. Caso não o seja, situação bastante comum, a distribuição utilizada é a t'Student.

Por exemplo, um pesquisador irá verificar se um grupo de cirurgiões-dentistas apresentam teores de mercúrio urinário superior ao valor de referência de normalidade ($Hg > 10 \mu g/l$).

Representado o objetivo em hipótese estatística:

$$H_0: \mu = 10 \mu g/l$$

$$H_a: \mu > 10 \mu g/l$$

Analisando-se os teores de mercúrio urinário de 10 cirurgiões-dentistas, obtiveram-se os seguintes resultados: (em $\mu g/l$)

16,4 11,8 24,0 16,0 8,5 13,8 13,4 14,9 10,7 10,0

Fixou-se o nível de significância a 5% , ou seja, o erro de rejeitar a hipótese nula sendo ela a verdadeira.

Como não existe nenhuma informação referente ao desvio padrão populacional (σ), a estimativa deste parâmetro (s) deverá ser utilizada e conseqüentemente o teste estatístico será o t'Student.

Calcula-se a média (\bar{x}) e o desvio padrão (s) do conjunto de dados obtidos na amostra:

$$\bar{x} = 13,95 \mu g/l \quad s = 4,38 \mu g/l \quad e \quad n = 10$$

Fórmula do teste:

$$t_n - 1 = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

onde \bar{x} = média obtida na amostra, m = média populacional, s = desvio padrão da amostra e n = nº de elementos da amostra.

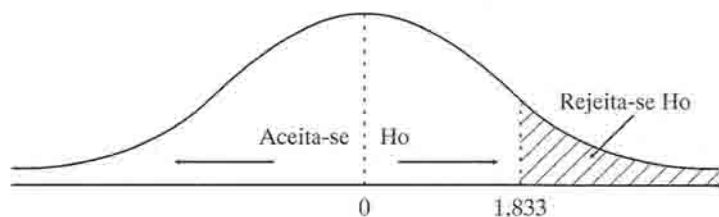
Assim:

$$t_9 = \frac{13,95 - 10}{\frac{4,38}{\sqrt{10}}} = 2,84$$

Mas, como decidir qual das duas hipóteses, H_0 ou H_a , será a escolhida? Se o valor obtido na fórmula do teste (t_{obs} = valor observado) for superior ao valor crítico (t_c), rejeita-se a hipótese nula e, para obter o valor crítico, utiliza-se a tabela de t-Student (Anexo II) com os seguintes parâmetros: graus de liberdade ($n-1$) e o nível de significância (α).

Neste exemplo, $gl = 10-1=9$ e $\alpha = 5\%$, $t_c = 1,833$.

Graficamente tem-se a seguinte situação:



Se o valor encontrado no teste ($t_{obs} = 2,84$) for inferior ao valor crítico ($t_c = 1,833$) não se rejeita a hipótese nula; se for maior, rejeita-se esta hipótese. No exemplo, $t_{obs} > t_c$, rejeita-se a hipótese nula e conclui-se que ao nível de significância de 5% a média do teor de mercúrio deste grupo de cirurgiões-dentistas é superior estatisticamente ao valor de normalidade.

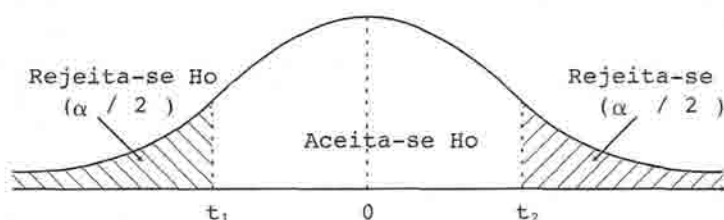
Suponhamos que o pesquisador apenas quisesse verificar se o

teor de mercúrio urinário era igual ou diferente ao valor de normalidade, ou seja,

$$H_0: u = 10 \text{ ug/l}$$

$$H_a: u \neq 10 \text{ ug/l}$$

Nesta situação, haveria duas regiões críticas:

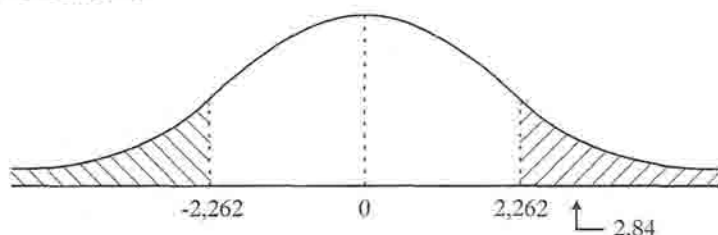


Para obter o valor t_c se utilizaria $n-1$ graus de liberdade e o nível de significância será $\alpha/2$.

No exemplo, $gl = 9$ e $\alpha/2 = 2,5\%$ ($t_c = 2,262$).

O valor do t_{obs} será o mesmo ($t_{obs} = 2,84$) uma vez que a fórmula não é modificada.

Graficamente:



Conclusão: Rejeita-se a hipótese nula.

Na primeira situação, tem-se o teste **UNICAUDAL** e na segunda, o teste **BICAUDAL**.

Quadro do procedimento do teste de uma média com desvio padrão desconhecido.

Hipóteses	Rejeita-se H_0 se
$H_0: u = u$	$t_{obs} < -t_c$
$H_a: u < u$	
$H_0: u = u$	$t_{obs} > t_c$
$H_a: u > u$	
$H_0: u = u$	$t_{obs} > t_c$ ou
$H_a: u \neq u$	$t_{obs} < -t_c$

3.3 Comparação de duas médias

3.3.1 Dados não emparelhados

No estudo "Voando com os Pilotos" de FERREIRA e col., um dos parâmetros analisados foi o tempo de voo ampliado (TVA). Este parâmetro é definido como sendo o período de tempo compreendido entre a partida e a chegada de um voo, acrescido do tempo de apresentação do piloto e de 30 minutos após o final de voo. O TVA permitiu uma quantificação do tempo em que o piloto ficou à disposição da companhia, seja em voo seja nos aeroportos, durante uma viagem.

Um dos objetivos era comparar se os TVA de voos nacionais eram iguais ou diferentes dos voos internacionais. Estatisticamente, tem-se as seguintes hipóteses:

$$H_0: \mu_n = \mu_i$$

$$H_a: \mu_n \neq \mu_i$$

A distribuição dos valores dos TVA (em horas) para os voos nacionais no período de análise de 15 dias foi:

48,42	61,08	64,00	69,75	71,25	12,75	14,42
36,50	52,75	65,67	43,25	46,25	43,75	57,33
34,25	45,75	72,83	66,42	48,70	68,25	46,75

Obteve-se uma média (\bar{x}) igual a 50,96 horas, desvio padrão(s) igual a 17,01 horas e com 21 pilotos analisados ($n = 21$).

A distribuição dos valores do TVA para os voos internacionais foi:

41,25	46,25	62,15	39,25	45,10	46,00	46,40	59,20	60,00	61,45
126,45	31,10	33,35	44,35	46,40	47,00	47,30	47,35	50,05	50,30
51,30	52,20	53,50	54,20	62,20	64,00	65,20	73,10		

Obteve-se uma média (\bar{x}) igual a 53,80 horas, com desvio padrão (s) igual a 17,23 e com 28 pilotos ($n=28$).

A distribuição de probabilidade utilizada é a t'Student, uma vez

que não são conhecidos os desvios padrões populacionais e a fórmula do teste é:

$$t_{\text{obs}} = \frac{\bar{x}_1 - \bar{x}_2}{Sp \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

onde:

\bar{x}_1 = média amostral do 1º grupo

\bar{x}_2 = média amostral do 2º grupo

n_1 = tamanho da amostra do 1º grupo

n_2 = tamanho da amostra do 2º grupo

Sp = desvio padrão ponderado

onde:

$$Sp = \sqrt{\frac{(n_1 - 1) \times S_1 + (n_2 - 1) \times S_2}{n_1 + n_2 - 1}}$$

onde S_1 = variância do 1º grupo e S_2 = variância do 2º grupo.

O valor do desvio padrão ponderado neste exemplo será:

$$Sp = \sqrt{\frac{(21 - 1) \times 289,34 + (28 - 1) \times 296,87}{21 + 28 - 2}} = 17,14$$

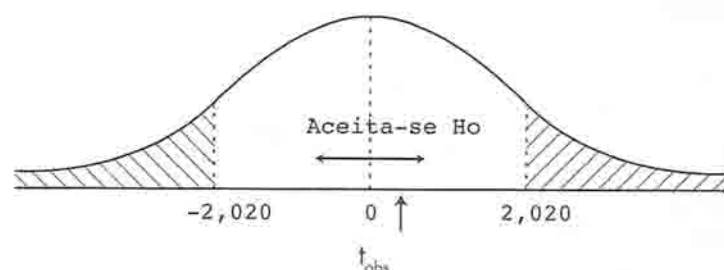
obs: a variância do vôo internacional é: $(17,23)^2 = 296,87$

Substituindo-se os valores encontrados na fórmula :

$$t_{\text{obs}} = \frac{53,80 - 50,96}{17,14 \times \sqrt{\frac{1}{21} + \frac{1}{28}}} = 0,57$$

Sendo esse um teste bicaudal, pois $H_a: u_n \neq u_1$, a representação gráfica com o valor crítico (t_c) com 47 graus de liberdade (21+28-2) e com nível de significância $\alpha/2$ (2,5%) obtido na tabela de

t' Student é igual a $t_c = 2,020$.



Conclusão: O valor t_{obs} pertence à região de aceitação de H_0 , ou seja, não há diferença estatística entre as médias do TVA dos vôos nacionais e internacionais.

3.3.2 Comparação de duas médias - dados emparelhados

Os resultados de duas amostras constituem dados emparelhados quando estão relacionados dois a dois segundo algum critério que introduz uma influência marcante entre os diversos pares (COSTA NETO).

A situação típica de dados emparelhados é "Antes-Depois", em que cada valor da primeira amostra está perfeitamente associado ao respectivo valor da segunda amostra.

Por exemplo, um pesquisador pretende avaliar os efeitos de um contaminante (tolueno) em setor de uma indústria gráfica, utilizando o monitoramento biológico mediante a dosagem de ácido hipúrico (g/l), amostrada antes e depois da jornada de trabalho.

Para isto, sorteou 10 operários deste setor e obteve resultados expressos em g/l:

operário	antes do início da jornada	depois da jornada
1	0,5	2,3
2	0,8	3,0
3	1,0	2,5
4	0,9	2,6
5	0,5	2,8
6	0,4	2,9
7	0,3	2,9
8	0,8	3,1
9	0,9	3,0
10	0,7	2,8

Se houver presença desse contaminante no ambiente de trabalho, é esperado que os valores obtidos após a jornada de trabalho sejam superiores aos do início da jornada. Caso contrário, as duas distribuições serão bastante semelhantes.

Na análise de dados emparelhados são calculadas as diferenças d_i correspondentes a cada par de valores (fixando-se o sentido da diferença: antes-depois ou depois-antes) e, se não houver diferença entre os pares, é esperada uma média destas diferenças igual a zero.

No exemplo fornecido, as hipóteses estatísticas formuladas são:

$$\begin{aligned} H_0: d &= 0 \\ H_a: d &> 0 \quad \text{sendo } d_i = \text{Depois} - \text{Antes} \end{aligned}$$

A distribuição de probabilidade utilizada é a t -Student com $n-1$ graus de liberdade e o nível de confiança.

$$t_{\text{obs}} = \frac{\bar{d}}{\frac{Sd}{\sqrt{n}}}$$

onde \bar{d} = médias das diferenças
 Sd = desvio padrão das diferenças
 n = número de pares

Assim,

Operário	Antes	Depois	$d = \text{Depois} - \text{Antes}$
1	0,5	2,3	1,8
2	0,8	3,0	2,2
3	1,0	2,5	1,5
4	0,9	2,6	1,7
5	0,5	2,8	2,3
6	0,4	2,9	2,5
7	0,3	2,9	2,6
8	0,8	3,1	2,3
9	0,9	3,0	2,1
10	0,7	2,8	2,1

$$\text{assim, } \bar{d} = \frac{\sum d_i}{n} = \frac{21,1}{10} = 2,11s \quad = \quad \frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n - 1} = 0,35$$

Logo, o valor do t_{obs} será:

$$t_{obs} = \frac{2,11}{\frac{0,35}{\sqrt{10}}} = 1,92$$

Ao nível de 5%, por exemplo, o valor do t_c obtido na tabela da distribuição t'Student, com 9 graus de liberdade e $\alpha = 5\%$ é igual a 1,833.

A conclusão é de se rejeitar a hipótese nula uma vez que o valor $t_{obs} > t_c$, ou seja, os trabalhadores estão expostos a esse contaminante. Sugerimos a leitura sobre o monitoramento biológico por ar exalado apresentada por BUSCHINELLI, em que são apresentados processos, métodos e critérios de avaliação e monitoramento biológico por ar exalado, mediante exemplos, com o objetivo de introduzir a discussão do assunto.

Quadro do procedimento do teste de duas médias com desvio padrão desconhecido.

Hipóteses	Rejeita-se H_0 se
$H_0: \mu = \mu$	$t_{obs} < -t_c$
$H: \mu < \mu$	
$H_0: \mu = \mu$	$t_{obs} > t_c$
$H: \mu > \mu$	
$H_0: \mu = \mu$	$t_{obs} < -t_c$ ou
$H: \mu \neq \mu$	$t_{obs} > t_c$

3.4 Comparação de várias médias

3.4.1 Amostras do mesmo tamanho:

No "Estudo das condições de saúde e trabalho dos motoristas e cobradores urbanos da cidade de São Paulo" de FREIRE e cols. entre as diversas análises, há uma referente a jornada de trabalho.

A jornada de trabalho dessa categoria é estipulada legalmente em 6h40min. diárias, porém o estudo constatou que era pouco respeitada pelas empresas de ônibus. Apenas 28% dos cobradores e 32% dos motoristas tinham uma jornada de até 7 horas. Estas porcentagens diminuíam quando analisadas segundo o tipo de empresa (estatal/privada). Nas particulares as porcentagens eram de 17% para os cobradores e de 14% para os motoristas.

A análise descritiva realizada nesse estudo possibilitou encontrarem-se as diferenças entre as duas categorias e o tipo de

empresa. Porém, poderia verificar também se a jornada média de trabalho era igual ou diferente nos grupos. Ou seja, comparar as jornadas médias do cobrador de empresa estatal (a), a jornada média do cobrador particular (b), a jornada média do motorista estatal (c) e a jornada média do motorista particular (d).

Nessa situação, há 4 médias sendo analisadas e para se poder compará-las, deve-se utilizar a distribuição de probabilidade de F'Snedecor, o teste de Análise de Variância (ANOVA) com um fator, pois há uma variável sendo testada (jornada de trabalho). Para efeito de cálculo, serão usados dados hipotéticos com amostras menores:

Tabela 11. Distribuição da jornada de trabalho (em horas) dos condutores, segundo tipo de empresa.

Cobrador		Motorista	
Estatal	Particular	Estatal	Particular
7	9	8	8
7	9	8	7
8	10	7	10
9	11	7	13
10	13	8	12
7	12	8	11
7	7	7	10
8	7	9	12
9	10	10	7
7	13	8	8

O objetivo é verificar se as jornadas médias dos condutores de empresas estatais e particulares são iguais ou diferentes, quanto ao nível de significância, por exemplo, igual a 5%:

$$H_0 : \mu_a = \mu_b = \mu_c = \mu_d$$

$$H_a : \text{pelo menos uma média difere}$$

Ao fazer a Análise de Variância, deve-se dispor os cálculos no seguinte quadro (COSTA NETO).

Análise de Variância- ANOVA

Fonte de Variação	Soma de Quadrados(SQ)	Graus de liberdade(GL)	Quadrado Médio(QM)	F
Entre amostras	SQE	K-1	QME=SQE/k-1	$F_{obs} = QME/QMR$
Resíduo	SQR	K.(n-1)	QMR=SQR/K.(n-1)	
Total	SQT	n.k-1		

onde K = número de amostras ou grupos, no exemplo, $k = 4$
 n = tamanho da amostra, no exemplo, $n = 10$ elementos

A obtenção do valor crítico de F é feita através da tabela de F 'Snedecor (Anexo IV) com os parâmetros: $K - 1$, $K \cdot (n - 1)$ e o nível de significância. No exemplo, $F_{3,36,5\%}$. E a hipótese nula será rejeitada se $F_{obs} > F_{crítico}$.

Os cálculos necessários para preencher o quadro da ANOVA, serão dispostos de uma maneira facilitada, necessitando-se apenas de uma calculadora:

Cobrador				Motorista			
Estatal		Particular		Estatal		Particular	
X1	X_1^2	X2	X_2^2	X3	X_3^2	X4	X_4^2
7	49	9	81	8	64	8	64
7	49	9	81	8	64	7	49
8	64	10	100	7	49	10	100
9	81	11	121	7	49	13	169
10	100	13	169	8	64	12	144
7	49	12	144	8	64	11	121
7	49	7	49	7	49	10	100
8	64	7	49	9	81	12	144
9	81	10	100	10	100	7	49
7	49	13	169	8	64	8	64
$\Sigma=79$		$\Sigma=101$		$\Sigma=80$		$\Sigma=98$	
$\Sigma=635$		$\Sigma=1.063$		$\Sigma=648$		$\Sigma=1.004$	

Para saber o valor da jornada média de cada categoria, utiliza-se o somatório (Σ) de X_1, X_2, X_3 e X_4 e divide-se o valor por n :

Cobrador estatal: $79/10 = 7,9$ horas
 Cobrador particular: $101/10 = 10,1$ horas
 Motorista estatal: $80/10 = 8,0$ horas
 Motorista particular: $98/10 = 9,8$ horas

Sejam A = Soma total dos valores:

$$A = 79 + 101 + 80 + 98 = 358$$

B = Soma total dos quadrados:

$$B = 635 + 1.063 + 648 + 1.004 = 3.350$$

$$C = \frac{(\sum x_i)^2}{n} = \frac{(79)^2}{10} + \frac{(101)^2}{10} + \frac{(80)^2}{10} + \frac{(98)^2}{10} = 3.244,60$$

Soma de Quadrados Total: SQT:

$$SQT = B - \frac{(A)^2}{n \times K} = 3.350 - \frac{(358)^2}{10 \times 4} = 145,90$$

Soma de Quadrados entre Amostras: SQE

$$SQE = C - \frac{(A)^2}{n \times K} = 3.244,60 - \frac{(358)^2}{10 \times 4} = 40,50$$

A Soma de Quadrados do Resíduo (SQR) é obtida pela diferença entre SQT e SQE:

$$SQR = 145,90 - 40,50 = 105,40$$

Assim, o quadro da ANOVA será:

F.V.	S.Q.	G.L.	Q.M.	F
Entre amostras	40,50	3	$40,5 \div 3 = 13,5$	$F_{obs} = 13,5 \div 2,93$
Resíduo	105,40	36	$105,40 \div 36 = 2,93$	$= 4,61$
Total	145,90	39		

O valor de F crítico obtido na tabela da distribuição F'Snedecor (Anexo IV), com 3 e 36 graus de liberdade e ao nível de 5% é $F_{crit} = 2,28$.

Conclusão: Sendo $F_{obs} > F_{crit}$, rejeita-se a hipótese nula, ou seja, pelo menos uma jornada média é diferente.

3.4.2 Amostras de tamanhos diferentes

Na tese de doutorado de FERREIRA que estudou o trabalho em turno, um dos objetivos era comparar a influência de diferentes sistemas de turno sobre a duração do sono e a avaliação subjetiva da qualidade do sono.

Para obter o valor da duração do sono, os 1.617 trabalhadores pesquisados assinalavam a hora em que iam dormir e a hora em que acordavam, depois de um dia de trabalho, obtendo-se uma boa estimativa da duração do sono. Analisaram-se 8 diferentes grupos de turno:

Tabela 12. Distribuição dos trabalhadores segundo o tipo de turno.

Grupo	Sistema	Horário	Nº de Trabalhadores
G1	3x8	6-14	777
		14-22	
		22-6	
G2	3x8	7-15	104
		15-23	
		23-7	
G3	2x8	8-16	79
		16-24	
G4	2x8	6-14	248
		14-22	
G5	2x8	5-13:30	90
		13:30-22	
G6	fixo	6-14	117
G7	fixo	14-22	151
G8	fixo	22-6	51

Fonte: FERREIRA

Vamos supor, por exemplo, que haja interesse de verificar se a duração média do sono no horário noturno dos turnos G1 (22-6), G2 (23-7) e o turno fixo G8 (22-6) é igual ou diferente. Sendo 3 médias a serem analisadas, a técnica estatística a ser utilizada é a Análise de Variância, com as seguintes hipóteses:

$$H_0: \mu_1 = \mu_2 = \mu_8$$

H_a : pelo menos uma média difere.

Utilizando-se dados hipotéticos e amostras menores, no grupo G1 foram analisados 10 trabalhadores, no grupo G2, 9 trabalhadores e 8 trabalhadores no grupo G8. Nota-se que as amostras são de tamanhos diferentes e para utilizar a Análise de Variância, há uma pequena modificação nos cálculos.

Distribuição da duração do sono dos trabalhadores segundo o tipo de turno (em horas).

G1	G2	G8
6:00	6:40	6:50
6:30	6:10	6:00
6:50	5:80	7:10
6:10	6:00	7:20
7:00	6:50	7:00
6:00	7:00	7:15
4:00	4:10	6:30
5:70	5:00	6:50
5:10	5:80	
5:50		

A seguir, será apresentado um quadro para facilitar os cálculos necessários para a ANOVA.

G1		G2		G8	
X1	X ² 1	X2	X ² 2	X3	X ² 3
6:00	36,00	6:40	40,96	6:50	42,25
6:30	39,60	5:80	33,64	6:00	36,00
6:50	42,25	6:50	42,25	7:10	50,41
6:10	37,21	4:10	16,81	7:20	51,84
7:00	49,00	5:80	33,64	7:00	49,00
6:00	36,00	6:10	37,21	7:15	51,12
4:00	16,00	6:00	36,00	6:30	39,69
5:70	32,49	7:00	49,00	6:50	42,25
5:10	26,01	5:00	25,00		
5:50	30,25				
$\Sigma=58,20$		$\Sigma=52,70$		$\Sigma=53,75$	
	$\Sigma=344,90$		$\Sigma=314,51$		$\Sigma=362,56$

Para o primeiro grupo (G1) tem-se $n = 10$, para G2 $n = 9$ e para G8, $n = 8$

A duração média de sono de cada turno será:

$$G1: X = 58,20 / 10 = 5,82 \text{ horas}$$

$$G2: X = 52,70 / 9 = 5,86 \text{ horas}$$

$$G8: X = 53,75 / 8 = 6,72 \text{ horas}$$

Seja: (SPIEGEL)

$$A = \text{soma total dos valores de } X\text{'s} \\ 58,20 + 52,70 + 53,75 = 164,65$$

$$B = \text{soma total dos quadrados:} \\ 344,90 + 314,51 + 362,56 = 1.021,97$$

$$C = (58,20)^2/10 + (52,70)^2/9 + (53,75)^2/8 = 1.008,4446$$

Soma de Quadrados Total: SQT

$$SQT = B - \frac{(A)^2}{\sum n_i} = 1.021,97 - \frac{(164,65)^2}{27} = 17,9099$$

Soma de Quadrados entre Amostras: SQE

$$SQE = C - \frac{(A)^2}{\sum n_i} = 1.008,44 - 1.004,06 = 4,3845$$

A Soma dos Quadrados do Resíduo (SQR) é obtida pela diferença entre SQT - SQE:

$$SQR = 17,9099 - 4,3845 = 13,5254$$

Os graus de liberdade (GL) serão:
entre amostras: $K - 1 = 3 - 1 = 2$
resíduo: $n_i - K = 27 - 3 = 24$
total: $n_i - 1 = 27 - 1 = 26$

Assim, o quadro da ANOVA é:

F.V.	S.Q	GL	QM	F _{obs}
entre amostras	4,3845	2	2,1923	3,89
resíduo	13,5254	24	0,5636	
Total	17,9099	26		

O valor de $F_{\text{crítico}}$ obtido na tabela da distribuição F'Snedecor (Anexo IV), para 2 e 24 graus de liberdade e ao nível de 5% é igual a 3,40. Sendo o valor de F_{obs} maior que o $F_{\text{crítico}}$, rejeita-se a hipótese nula, ou seja, a duração média do sono é diferente em pelo menos um dos turnos.

Referências bibliográficas

BUSCHINELLI, J.T.P. Monitoramento biológico por ar exalado. Revista Brasileira de Saúde Ocupacional, v.18, n.70, p.9-11, abr./jun., 1990.

COSTA NETO, P.L.O. Estatística. São Paulo : Edgard Blücher, 1977.

FERREIRA, L. L. Les travailleurs postés et leur sommeil: une étude dans l'industrie brésilienne. These (Doctorat d'Ergonomie) - Université de Paris XIII, 1988. 96p.

FERREIRA, L.L., BUSSACOS, M.A., SCHILITER, C.R.S. et al. Voando com os pilotos: condições de trabalho dos pilotos de uma empresa de aviação comercial. São Paulo: Associação dos Pilotos da VARIG, FUNDACENTRO, 1991.

FREIRE, N.B., SOUZA, M.F.M., UEHARA, M.K. et al. Estudo das condições de saúde e trabalho dos motoristas e cobradores urbanos de São Paulo. São Paulo : FUNDACENTRO, 1991.

MENDONÇA, E.M.C., SILVA, R.C.C. da, ALGRANTI, E. Função pulmonar e reatividade brônquica inespecífica em trabalhadores do Terminal Rodoviário do Tietê - São Paulo, expostos a produtos de exaustão de motores diesel. Revista Brasileira de Saúde Ocupacional, v.19, n. 74, p.20-4, jul./dez. 1991.

RODRIGUES, J.B. Monitoramento biológico da exposição ao benzeno em indústria petroquímica do Estado de São Paulo, pela dosagem de fenol urinário. Revista Brasileira de Saúde Ocupacional, v.19,n. 72, p.41-4, jan./mar. 1991.

SPIEGEL, M.R. Estatística. 3.ed. São Paulo: Makrow Books, 1993.

4.1 Relação entre duas variáveis quantitativas

Uma investigação será realizada numa fábrica de baterias, na qual os trabalhadores estão expostos ao chumbo inorgânico. Entre as estratégias a serem aplicadas, o pesquisador utilizará os indicadores biológicos ALA-U e PBs (obs.: A execução de um programa de controle biológico deve obedecer a alguns critérios, a fim de que esses programas tragam benefícios reais aos trabalhadores expostos e não sejam usados como "cobaia" ou amostrador do ambiente de trabalho (BUSCHINELLI):

ALA-U é um indicador de efeito, sua excreção aumenta na urina de trabalhadores expostos a chumbo devido à inibição da enzima ALA-dehidrose-(ALAD) e **PBs** é um indicador de exposição ou de dose mais utilizado, mede a quantidade de chumbo presente no sangue total e reflete a situação de exposição ou não ao agente em concentrações determinadas (SANTOS).

Assim, investigando-se 20 trabalhadores do Setor X, obteve-se a seguinte tabela (obs.: ALA-U corrigido por creatinina):

Tabela 13. Distribuição dos trabalhadores segundo os valores de ALA-U e PBs.

Trabalhador	ALA-U	PBs
1	9,0	66,0
2	5,1	53,0
3	2,0	36,0
4	10,6	68,0
5	4,3	60,0
6	5,6	71,0
7	9,5	79,0
8	20,0	71,5
9	6,10	70,0
10	8,40	75,0
11	9,0	66,0
12	6,6	51,0
13	5,6	65,0
14	6,6	68,0
15	5,7	62,0
16	7,4	43,0
17	14,2	73,0
18	9,6	82,0
19	7,5	76,0
20	9,0	66,0

Será que existe uma relação estatística entre estes dois indicadores?

4.1.1 Diagrama de dispersão:

Suponhamos que se queira verificar se existe relação entre duas variáveis quantitativas X e Y . Retira-se uma amostra e para cada elemento da amostra um par de observação: um valor de X e um valor de Y , que podem ser representados no seguinte diagrama:

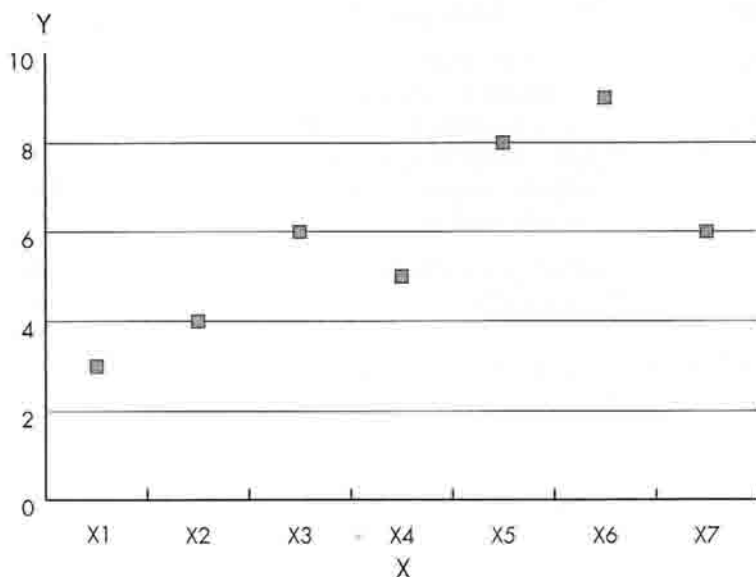
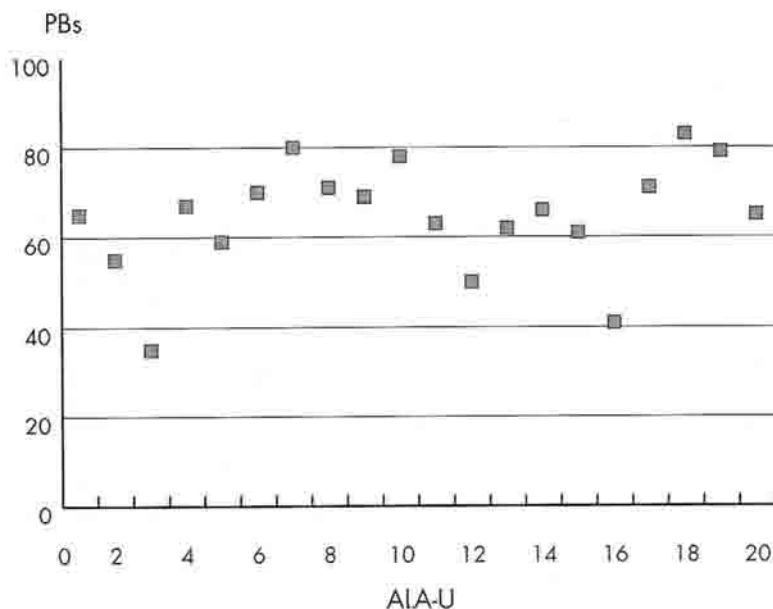


Figura 5. Diagrama de dispersão.

A esse gráfico, chama-se Diagrama de dispersão, que já fornece uma idéia intuitiva da eventual relação entre as duas variáveis (HOFFMANN). Nesse exemplo, verifica-se uma possível linearidade entre as duas variáveis.

Para o exemplo dos trabalhadores expostos ao chumbo, obtém-se o seguinte gráfico:

Distribuição dos Resultados do ALA-U e PBs Segundo Trabalhador



Fonte: Dados fornecidos por BUSCHINELLI

Figura 6. Distribuição dos resultados do ALA-U e PBs, segundo trabalhador.

Para poder “quantificar” essa relação, tem-se:

4.1.2 Coeficiente de Correlação Linear de Pearson: r

É uma medida do grau de dependência linear entre duas variáveis que, ao mesmo tempo, serve para verificar se existe ou não relação desse tipo entre elas.

O valor do coeficiente de correlação de Pearson varia de -1 até 1 ($-1 < r < 1$), se existir uma relação linear direta entre X e Y, o valor de r será positivo, se a relação linear for inversa, r será negativo, e se as variáveis não apresentarem relação linear, $r = 0$ (um erro freqüente é dizer que quando $r = 0$ não existe relação entre as duas variáveis, mas, na verdade, o que significa é que **não existe relação linear**.)

Para encontrar o valor desse coeficiente, aplica-se a seguinte fórmula (SPIEGEL):

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}}$$

Assim, no exemplo fornecido, tem-se:

ALA-U(x)	PBs(Y)	X•Y	X ²	Y ²
9,0	66,0	594,0	81,0	4.356,0
5,1	53,0	270,3	26,01	2.809,0
2,0	36,0	72,0	4,0	1.296,0
10,6	68,0	720,8	112,36	4.624,0
4,3	60,0	258,0	18,49	3.600,0
5,6	71,0	397,5	31,36	5.041,0
9,5	79,0	750,5	90,25	6.241,0
20,0	71,5	1.430,0	400,0	5.122,25
6,1	70,0	427,0	37,21	4.900,0
8,4	75,0	630,0	70,56	5.625,0
9,0	66,0	594,0	81,0	4.356,0
6,6	51,0	336,6	43,56	2.601,0
5,6	65,0	364,0	31,36	4.225,0
6,6	68,0	448,8	43,56	4.624,0
5,7	62,0	353,4	32,49	3.844,0
7,4	43,0	318,2	54,76	1.849,0
14,2	73,0	1.036,6	201,64	5.329,0
9,6	82,0	787,2	92,16	6.724,0
7,5	76,0	570,0	56,25	5.776,0
9,0	66,0	594,0	81,0	4.356,0
$\Sigma=161,8$	$\Sigma=1.301,5$	$\Sigma=10.953,0$	$\Sigma=1.589,02$	$\Sigma=87.288,25$

$$\text{logo, } r = \frac{10.953,0 - \frac{161,8 \times 1.301,5}{20}}{\sqrt{(1.589,02 - \frac{(161,8)^2}{20}) \times (87.288,25 - \frac{(1.301,5)^2}{20})}} = 0,50$$

Quando estamos trabalhando com o coeficiente de correlação de Pearson, quase sempre é através de uma amostragem, ou seja, r é uma estimativa por ponto de ρ , **coeficiente de correlação de Pearson populacional**. Devido às flutuações casuais, pode-se encontrar um $r \neq 0$, quando na verdade, $\rho = 0$. Então, deve-se testar o valor encontrado, com as seguintes hipóteses estatísticas:

$$H_0: \rho = 0$$

$$H_a: \rho \neq 0$$

Teste estatístico: t'Student, com n-2 graus de liberdade e

$$t_{\text{obs}} = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

No exemplo, ao nível de 5%, com 20-2=18 graus de liberdade, tem-se, na tabela de t'Student (Anexo II), o valor de 2,101 e

$$t_{\text{obs}} = \frac{0,5 \times \sqrt{18}}{\sqrt{1 - (0,5)^2}} = 2,45$$

O valor observado de t é maior do que o valor crítico, portanto, se rejeita a hipótese nula, ou seja, há relação linear entre ALA-U e PBs.

Referências bibliográficas

BUSCHINELLI, J.T.P. Conceitos, indicações e utilização de indicadores biológicos. In: COSTA, D.F.C., CARMO, J.C., SETTIMI, M.M. et al. Programa de saúde dos trabalhadores: a experiência da Zona Norte, uma alternativa em Saúde Pública. São Paulo: HUCITEC, 1989. Cap.10 p.299-348.

HOFFAMANN, R., VIEIRA, S. Análise de regressão. São Paulo : HUCITEC, EDUSP, 1977.

SANTOS, U.P. Manual prático de monitoragem e avaliação biológica em trabalhadores expostos a agentes químicos. In: COSTA, D.F., CARMO, J.C., SETTIMI, M.M. et al. Programa de saúde dos trabalhadores: a experiência da Zona Norte, uma alternativa em Saúde Pública. São Paulo: HUCITEC, 1989. Cap.10 p.320-48.

SPIEGEL, M.R. Estatística. 3.ed. São Paulo: Makrow Books, 1993.

5 ANÁLISE DE TABELA 2x2

5.1 Introdução

Quando há o objetivo de analisar uma tabela cujas as variáveis são dicotômicas, isto é, apresentam duas categorias [por exemplo, Agravado(sim/não), Exposição a um fator (sim/não), Sexo (fem./masc.), etc.], o pesquisador tem várias técnicas estatísticas para serem utilizadas e dentre estas, destacamos as seguintes, aplicadas no exemplo a seguir:

BUSCHINELLI* analisando 554 trabalhadores de uma usina de cana-de-açúcar, obteve a seguinte tabela para duas, entre as variáveis estudadas:

Tabela 14. Distribuição dos trabalhadores segundo a exposição a vibração de máquina agrícola e queixa gástrica.

	Com queixa	Sem queixa	Total
Exposto à vibração	106	188	294
Não exposto	29	231	260
Total	135	419	554

Fonte: BUSCHINELLI

* Este é um exemplo de um estudo epidemiológico de coorte, onde há dois grupos independentes, um estando exposto ao fator de risco e o outro não exposto a este fator, e verifica-se a distribuição do agravo nos dois grupos.

O objetivo é verificar se existe uma relação entre as duas variáveis (associação) ou se elas são independentes. Mas como realizar essa análise?

5.2 Qui-Quadrado: χ^2 (Prova de Independência)

De uma maneira geral, diz-se que duas variáveis estão relacionadas, ou são dependentes, quando as intensidades de uma delas determinam, de certo modo, as intensidades da outra.

Se não existe relação entre as duas variáveis, diz-se que elas são independentes. E por meio do Qui-Quadrado, um teste estatístico, que se poderá verificar se existe uma relação entre as variáveis ou não.

Hipóteses: **H₀:** Há independência entre as variáveis.

H_a: Há associação entre as variáveis.

Para decidir qual das duas hipóteses a ser aceita, para a tabela 2x2, com nível de significância de 5% ($p < 0,05$), o valor do Qui-Quadrado obtido na tabela do Anexo III, com 1 grau de liberdade, é igual a 3,84 e iremos comparar este valor com o resultado obtido no Qui-Quadrado calculado (utilizando-se uma fórmula).

Se o valor do Qui-Quadrado calculado for superior ao valor 3,84, **rejeita-se** a hipótese nula.

5.2.1 Qui-Quadrado χ^2 (sem correção):

Para a seguinte tabela hipotética:

	B+	B -	Total
A+	a	b	a + b
A-	c	d	c + d
Total	a + c	b + d	N

Aplica-se a seguinte fórmula para se obter o valor do Qui-Quadrado:

$$\chi^2 = \frac{N \times (a \times d - c \times b)^2}{(a + b) \times (c + d) \times (a + c) \times (b + d)}$$

No exemplo de BUSCHINELLI, tem-se o seguinte valor para o Qui-Quadrado:

$$\chi^2 = \frac{554 \times (106 \times 231 - 29 \times 188)^2}{135 \times 419 \times 294 \times 260} = 46,4$$

Conclusão: Como o valor do Qui-Quadrado calculado (46,4) é maior que o valor obtido na tabela (3,84), rejeita-se a hipótese nula, ou seja, há evidências estatísticas de uma associação entre a exposição à vibração e a queixa gástrica.

5.2.2 Qui-Quadrado χ^2 (com correção)

É utilizado para amostras pequenas ($n < 30$). Há bastante controvérsia em sua aplicação e, muitos preferem utilizar o teste exato de Fisher (FLEISS).

$$\text{Fórmula: } \chi^2 = \frac{N \times [\text{mod} (a \times d - b \times c) - 0,5]^2}{(a + b) \times (b + d) \times (a + c) \times (c + d)}$$

5.2.3. Qui-Quadrado de Mantel Haenszel:

Este Qui-Quadrado testa a hipótese alternativa de que há uma associação linear entre as duas variáveis (entre a linha e coluna), com 1 grau de liberdade (FLEISS):

$$\chi^2 = \frac{(N - 1) \times (axd - bxc)^2}{(a + b) \times (b + d) \times (a + c) \times (c + d)}$$

5.3 Outras medidas de associação

Os valores do Qui-Quadrado variam de zero a mais infinito, assim o seu resultado não permite inferir nem o grau nem o sinal da associação. Ou seja, o resultado obtido no estudo de BUSCHINELLI, $\chi^2 = 46,42$, será que representa uma "forte" associação entre o fator de exposição e o agravo? Para poder responder a esta pergunta, "desenvolveram-se várias tentativas de criar um coeficiente que fosse igual a zero, no caso de independência, e valores bem definidos, positivos ou negativos, no caso de uma associação perfeita positiva ou negativa, respectivamente" (BERQUÓ).

5.3.1 Coeficiente Phi - φ :

Este coeficiente deriva da estatística Qui-Quadrado e tem a qualidade de apresentar as duas condições: grau e sentido da associação.

O seu valor, para tabela 2x2, varia de -1 a +1 ($-1 < \varphi < +1$), assim, quando $\varphi =$ zero, há independência entre as variáveis; $\varphi = -1$, ocorre associação perfeita negativa e quando $\varphi = +1$, associação perfeita positiva (FLEISS).

$$\text{Fórmula: } \varphi = \frac{axd - bxc}{\sqrt{(a + b) \times (a + c) \times (c + d) \times (b + d)}}$$

Para as seguintes tabelas, com dados hipotéticos, tem-se exemplos de associação perfeita negativa, perfeita positiva e independência:

Associação Perfeita Negativa:

	B+	B-	Total
A+	0	50	50
A-	30	0	30
Total	30	50	80

Qui-Quadrado = 80,00

$$\varphi = \frac{0 - 30 \times 50}{\sqrt{30 \times 50 \times 50 \times 30}} = -1,0$$

Associação Perfeita Positiva:

	B+	B -	Total
A+	60	0	60
A-	0	40	40
Total	60	40	100

Qui-Quadrado = 100,00

$$\varphi = \frac{60 \times 40 - 0}{\sqrt{60 \times 40 \times 40 \times 60}} = 1,0$$

Independência:

	B+	B -	Total
A+	20	20	40
A-	20	20	40
Total	40	40	40

Qui-Quadrado = 0,00

$$\varphi = \frac{20 \times 20 - 20 \times 20}{\sqrt{40 \times 40 \times 40 \times 40}} = 0,0$$

No estudo de BUSCHINELLI, encontramos o seguinte valor para o coeficiente phi:

$$\varphi = \frac{106 \times 231 - 29 \times 188}{\sqrt{135 \times 419 \times 294 \times 260}} = 0,29$$

significando uma associação positiva entre o fator de exposição à vibração e queixa gástrica.

Para os estudos epidemiológicos, outras medidas são utilizadas, que complementam a associação entre as variáveis analisadas:

5.3.2 Risco Relativo: RR

Este risco, utilizado em estudo tipo coorte, indica quantas vezes maior é a probabilidade de os expostos ao fator apresentarem o agravo em relação aos não expostos. Basicamente, verifica-se a incidência do agravo nos expostos com a incidência do agravo

nos não expostos ao fator de risco. Esquemáticamente, tem-se a seguinte tabela:

	Com agravo	Sem agravo	Total
Exposto ao fator	a	b	a + b
Não exposto	c	d	c + d

Fórmula do Risco Relativo: $RR = \frac{a / a + b}{c / c + d}$

Para o trabalho apresentado por BUSCHINELLI, obteve-se o seguinte valor para o risco relativo:

$$RR = \frac{106 / 294}{29 / 260} = 3,27, \text{ ou seja, os expostos à vibração têm 3}$$

vezes mais de possibilidade de apresentarem queixas gástricas do que os não expostos.

Construindo-se um intervalo de confiança para o risco relativo é possível verificar a significância estatística desse risco, se o intervalo não conter o valor 1, diz-se que o risco relativo é significativo. Segundo KLEINBAUM, tem-se a seguinte fórmula para a construção deste intervalo de confiança:

Supondo-se a distribuição Normal (z), ao nível de 95% (p < 0,05):

(RR x exp (-z √ v), RR x exp (+ z √ v), onde z = 1,96 e v é definido como:

$$v = \frac{(1 - a / a + b)}{a} + \frac{(1 - c / c + b)}{c}$$

O valor de v nesse estudo será:

$$v = \frac{(1 - 106 / 294)}{106} + \frac{(1 - 29 / 260)}{29} = 0,0367$$

assim, o intervalo de confiança do risco relativo para verificar a significância estatística entre os expostos à vibração e a queixa gástrica será:

RR = 3,27, z = 1,96, v = 0,0367 e exp é o número neperiano (e), tem-se:

$(3,27 \times \exp(-1,96 \pm 0,0367), 3,27 \times \exp(+1,96 \pm 0,0367)) = (2,25, 4,76)$. Como este intervalo não contém a unidade, o risco relativo encontrado é estatisticamente significativo.

5.3.3 Odds Ratio: OR

É empregado quando o estudo epidemiológico é do tipo caso-controle, isto é, as pessoas com o agravo (caso) e as pessoas sem agravo (controle) são selecionadas, e as proporções de casos e controles que foram expostas a possíveis fatores de risco são então determinadas e comparadas. Esquemáticamente, tem-se o seguinte quadro:

	Caso	Controle
Exposto ao fator	a	b
Não exposto	c	d
Total	a + c	b + d

Para calcular o risco relativo, utiliza-se a sua estimativa, que é o Odds Ratio (OR), definido como:

$$OR = \frac{a \times d}{b \times c}$$

Como este parâmetro é uma estimativa pontual do risco relativo, deve estar acompanhado pelo seu respectivo intervalo de confiança, pois além de verificar a significância estatística, fornece com uma margem de confiança pré-fixada, a estimativa por intervalo. Conforme KLEINBAUM, esse intervalo é calculado da seguinte maneira:

$$(OR \times \exp(-z \times \sqrt{v}), OR \times \exp(+z \times \sqrt{v}))$$

onde: z = parâmetro da distribuição Normal, para 95%, z = 1,96

OR= odds ratio

$$v = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

Para ilustrar a aplicação desses cálculos, utilizaremos o resultado apresentado na tese de doutoramento de WÜNSCH, que realizou um estudo caso-controle com objetivo de verificar a relação entre os riscos ocupacionais e câncer de pulmão na Região Metropolitana de São Paulo. Para o fator de risco-Tabagismo, definido como a pessoa ter fumado 50 ou mais cigarros durante toda a vida, encontrou-se a seguinte tabela:

Tabela 15. Distribuição de casos e controles segundo tabagismo, região metropolitana de São Paulo, 1990/91

	Caso	Controle
Tabagismo	278	342
Não tabagismo	38	194
Total	316	536

Qui-Quadrado = 58,60

$$OR = \frac{278 \times 194}{38 \times 342} = 4,15$$

e o respectivo intervalo de confiança (sendo estatisticamente significativo quando exclui o valor 1) ao nível de 95% ($z = 1,96$) e

$$v = \frac{1}{278} + \frac{1}{342} + \frac{1}{38} + \frac{1}{194} = 0,038$$

Assim, o intervalo de confiança será:

$(4,15 \times \exp(-1,96 \times \sqrt{0,038}), 4,15 \times \exp(+1,96 \times \sqrt{0,038})) = (2,83, 6,08)^*$. Como o intervalo não contém a unidade, o Odds ratio encontrado é estatisticamente significativo.

*O método utilizado por Wunsch, que aplicou o software EPINFO, para a elaboração do intervalo de confiança, é diferente do método desenvolvido por Kleinbaum. O intervalo obtido na tese foi (2,79, 6,20).

Referências bibliográficas

BERQUÓ, E.S., SOUZA, J.M.P., GOTLIEB, S.L.D. Bioestatística. São Paulo : EPU, 1980.

BUSCHINELLI, J.T.P., MORO, G. Avaliação epidemiológica de sintomatologia gástrica e lombalgias em expostos a vibrações de baixa frequência (corpo inteiro). Revista Brasileira de Saúde Ocupacional, v.13, n.51, p.37-9, jul./set. 1985.

FLEISS, J.L. Statistical methods for rates and proportions. 2.ed. New York: John Wiley & Sons, 1981.

KLEINBAUM, D.G., KUPPER, L.L., MORGENTHAU, H. Epidemiologic research principles and quantitative methods. Belmont : Wadsworth, 1982.

SIEGEL, S. Estatística não paramétrica. São Paulo: McGraw Hill, 1975.

WÜNSCH, V.F. Riscos ocupacionais e câncer de pulmão. Dissertação (Doutoramento em Saúde Ocupacional) - Faculdade de Saúde Pública, Universidade de São Paulo, 1992.

A apresentação, a análise e a interpretação de dados são tarefas relativamente complexas. Os pesquisadores que não conhecem estatística podem pensar que é fácil escolher um tipo qualquer de análise, dentro da grande variedade de técnicas descritas em livros de estatísticas e disponíveis em vários "softwares" de microcomputadores.

Nesse sentido, nunca é demais lembrar que é necessário conhecer as bases da estatística para escolher com critério a melhor técnica estatística para a análise de seus dados (resultados) e poder tirar conclusões sobre eles.

Os erros estatísticos mais comuns vão desde o delineamento incorreto de experimento até o uso inadequado de métodos de análise. A explicação para esse fato é a de que os pesquisadores estão mais interessados nas conclusões do que na técnica estatística utilizada para chegar a estas conclusões. No entanto, é evidente que não se pode confiar em conclusões obtidas através de técnicas estatísticas erradas. Afinal, muitos pesquisadores ainda se deslumbram com o jargão estatístico e o uso da informática, esquecendo aspectos importantes da pesquisa, como a estrutura correta do experimento e a escolha do método adequado de análise (VIEIRA).

É importante ressaltar que existem poucas técnicas alternativas, quando se está diante de um problema específico, e, ao realizar um estudo, o pesquisador deve ter em mente as seguintes indagações: o que pretende fazer? Por quê? Como? Quando? E onde?

Muitas vezes, há uma inversão do planejamento da pesquisa, principalmente quando esta envolve levantamento de dados, ou seja, na hora de definir o objetivo com bastante clareza, que é a espinha dorsal da pesquisa, destina-se pouco tempo para a discussão dessa etapa e já se parte para a elaboração do questionário, que é o instrumento de coleta de dados. É nessa fase que há grande probabilidade de se iniciar a "**Teoria Já Que...**". Para melhor compreensão dessa "teoria", a seguir, um exemplo hipotético:

Um inquérito domiciliar será realizado numa cidade, tendo como objetivo investigar os acidentes do trabalho (típico, trajeto e doença ocupacional) da população empregada.

Os pesquisadores definiram as seguintes variáveis que irão compor essa investigação: quantos já sofreram acidente do trabalho, idade, sexo, as causas do acidente, o horário em que ocorreu, o ramo de atividade da empresa e tempo de afastamento.

Um dos fatores que desencadeiam a "Teoria Já Que" é a carência de informação que há, principalmente no Brasil, e os pesquisadores, na ânsia de cobrir essa ausência, tendem a introduzir variáveis que não necessariamente estão relacionadas com o objetivo do estudo. Por exemplo, "**Já que** estamos estudando os trabalhadores, poderíamos acrescentar o *tempo de trajeto casa-trabalho*, pois esta variável está associada com o acidente do trabalho. Também poderíamos saber a *renda*, o *tempo na função*, a *função* (Obs.: sem dúvida, a variável de maior complexidade para análise, muitas vezes, no relatório final, é deixada de lado).

Já que estamos realizando essa investigação domiciliar, poderíamos conhecer as *condições de trabalho* desses trabalhadores, tais como, , a *iluminação*, *presença de ruído*, *ventilação*, *piso*, *jornada de trabalho*, se o *local do trabalho é confortável* (?), *pressão da chefia*, *produção*, *existência de proteção coletiva*, *EPI*, se *utilizam os EPI*, se *eles os incomodam*, como é o *banheiro*?, o *de que gostam e não gostam no trabalho*...

Já que estamos analisando esses trabalhadores, poderíamos conhecer a condição de vida deles, por exemplo, o *tipo de moradia*, se há *esgoto*, *água encanada*, *coleta de lixo*, *quantas pessoas moram na casa*, *quantas trabalham*, *renda familiar*, *naturalidade*, *tempo de residência na cidade*, o *que fazem nas folgas*...

Já que estamos realizando esse levantamento, poderíamos relacionar algumas queixas (morbidade referida) com o trabalho, tais como: *dor nas costas*, *queimação no estômago*, *dor de cabeça*, *dor nas pernas*, *palpitação*, *zumbido*, *problemas de visão*, *audição*, se *procurou o serviço médico nos últimos meses*, *quais as doenças que teve nos últimos meses*, se *ficou afastado*, *fez tratamento*, *onde*?...

Já que estamos perguntando ao trabalhador, poderíamos saber se *ele fuma*, *quantos maços/dia*, se *bebe*, *quantas doses*, se *já experimentou alguma droga* (?), *qual*?...

E assim por diante. Nesse exemplo, o número originário de

variáveis era 7 é, no final, com a Teoria Já Que, chegou-se a 54 variáveis! É importante salientar que, embora a pesquisa seja hipotética, as variáveis mencionadas já foram utilizadas em outras pesquisas, recentemente, na área de Saúde Ocupacional.

O grande problema que o "Já Que" cria para o pesquisador é depará-lo com uma infinidade de informações, às vezes desconexas com o objetivo do estudo, as quais isoladamente não levam a nada ou não trazem qualquer significado, resultando num alto custo financeiro e de tempo.

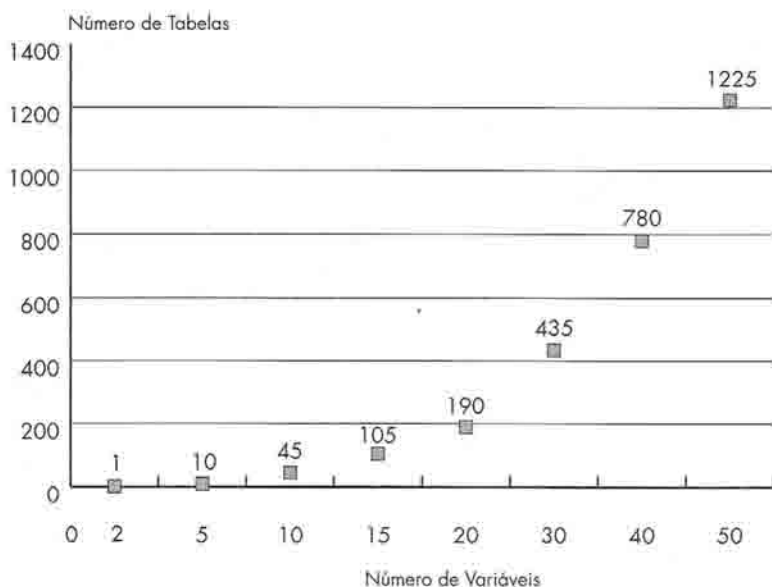
Por outro lado, um fator bastante crítico e menosprezado é o entrevistado (no exemplo, é o trabalhador) que além de "perder tempo" em responder às perguntas desconexas, tendenciosas ou que em si já apresentam um viés nas respostas, tendem a ser menos precisos quanto maior for o número de perguntas, ou seja, a precisão das respostas está inversamente relacionada ao número de questões a serem formuladas.

Quanto ao número de variáveis que deve ter um questionário, estatisticamente não existe uma regra, porém, usando-se o bom senso e com a restrição de alguns softwares estatísticos, cuja capacidade máxima pode variar de 30 a 40 variáveis, deve-se trabalhar tão somente com as variáveis que estão realmente relacionadas com o objetivo do estudo.

Cruzamentos: Esta é uma palavra chave nas pesquisas que envolvem levantamento de dados e que na teoria Já Que torna-se bastante evidente, ou seja, além do aumento do número de variáveis que o Já Que proporciona, os cruzamentos entre as variáveis tendem aumentar ainda mais a proliferação de papéis. A princípio, os pesquisadores solicitam aos estatísticos os possíveis cruzamentos entre as variáveis (normalmente, tabelas com duas variáveis) e verificar quais os cruzamentos são estatisticamente significativos.

Por exemplo, se num estudo houver 40 variáveis, o número de cruzamentos possíveis (duas a duas variáveis, sem repetição) resulta em **780** tabelas para serem analisadas! O gráfico seguinte apresenta a curva exponencial do número de tabelas geradas segundo o número de variáveis existentes no estudo:

Número de tabelas* geradas segundo o número de variáveis



*Combinação entre duas variáveis, sem repetição (cruzamento)

Figura 7. Número de tabelas geradas segundo o número de variáveis.

Essa situação, ainda que possa parecer inconcebível na área de pesquisa, infelizmente, é comum e a inversão dos princípios de metodologia científica (ou distanciamento da pergunta condutora*) se faz presente nas pesquisas de campo.

No capítulo 6 do livro de BERQUÓ, há uma síntese das etapas de um levantamento por amostragem, estando estas etapas intimamente ligadas aos princípios de metodologia científica, que iremos transcrevê-las, como um guia de orientação para minimizar os efeitos negativos da Teoria Já Que:

Etapas de um levantamento por Amostragem:

a) Explicitação dos objetivos com bastante clareza, a fim de evitar dúvidas posteriores ou mesmo esquecimentos, devendo ficar bem definida qual a unidade elementar (elemento) ou unidade de análise a ser trabalhada.

b) Definição da população a ser amostrada.

c) Escolha das variáveis a serem observadas em cada unidade de análise.

*Na dissertação de mestrado de NOVAES há uma explanação sobre a formulação da pergunta condutora.

Deve ser verificado se todos os dados que vão ser coligidos são relevantes para a pesquisa e se nenhum dado foi omitido. Existe uma tendência, particularmente ao se trabalhar com populações humanas, usando questionário, de se fazer muitas perguntas, um grande número das quais nunca são analisadas; questionários longos, em geral, levam a diminuir a qualidade da resposta.

d) Especificação do grau de precisão desejado.

e) Escolha dos instrumentos de medida e da forma de abordagem.

f) Escolha da unidade amostral.

g) Execução da prova piloto (ou pré-teste).

Nesta etapa, é feito um verdadeiro ensaio do trabalho a ser desenvolvido, sendo testados os instrumentos de medida, questionários, pessoal de campo, a sistemática proposta, a reação da população a ser amostrada. Orienta os reajustes necessários e pode dar informações valiosas sobre possível duração e custo da pesquisa e indicações da variabilidade do fenômeno pesquisado, o que permite calcular melhor o tamanho da amostra.

h) Seleção da amostra, após decidido qual deve ser o respectivo tamanho."

Uma outra orientação para não inserir a **Teoria Já Que** na pesquisa está exemplificada na seguinte situação:

Um pesquisador é procurado por um Sindicato de uma certa categoria profissional para que "estude" um problema que a aflige. Mas, com raríssimas exceções, o problema não está bem definido ou delimitado. A princípio, tanto o pesquisador como o Sindicato, têm dúvidas sobre o que realmente deva ser estudado (pois, às vezes, são vários os problemas ou situações a serem analisados) e é, nesse momento, que o diálogo entre o pesquisador e os componentes do sindicato deve-se ser o mais exaustivo possível, desde a compreensão da linguagem (por exemplo, o termo "amostra" para um estatístico tem significado diferente do que para um químico) até a "visualização" do problema em si.

Assim, com esta negociação preliminar, fica mais exequível a elaboração da pergunta condutora do estudo, reduzindo-se o risco de introduzir a Teoria Já Que.

Por fim, é importante considerar as atribuições da equipe, do questionário e da procura de informações.

A equipe: A partir do final da década de 60, algumas pesquisas de campo, principalmente na área de Saúde Pública, eram realizadas com o cognome de trabalho multiprofissional, que

consistia em reunir vários profissionais (médicos, engenheiros, químicos, sociólogos, etc.) trabalhando com o mesmo problema (os estágios de campo multiprofissional da Faculdade de Saúde Pública-USP são um exemplo da formação dessas equipes, que tinham como objetivo: "proporcionar experiências de aprendizagem de casos concretos, a partir de um trabalho de equipe multiprofissional, que favoreçam a elaboração de um ensaio de diagnóstico de saúde de uma determinada área"-PROGRAMA DO ESTÁGIO).

Quando da elaboração do questionário, o que ocorria era que cada profissional envolvido (ou cada área) elaborava o "seu questionário" e, posteriormente, juntava (grampeava) com os demais questionários, criando-se uma ilusão de integração multiprofissional e a presença da Teoria Já Que era inevitável. Posteriormente, o termo multiprofissional foi substituído por multidisciplinariedade mas, o viés, o questionário grampeado, com poucas exceções, persistiu.

Assim, quando houver uma equipe multidisciplinar para estudar um problema, é de extrema importância que haja um sinergismo entre os componentes da equipe, todos compreendendo o problema com a mesma visão, é de extrema importância que todos os componentes da equipe falem a mesma linguagem, ou pelo menos, se compreendam entre si quanto ao objetivo de sua participação. Isto não significa uma padronização e sim uma "unificação" do saber de diferentes formações (socialização do conhecimento).

Questionário não é tudo: Quanto maior a complexidade do problema, pior pode ser o emprego do questionário, pois é praticamente impossível abranger toda essa complexidade num único método. É preferível utilizar diversos métodos a restringir-se em uma única metodologia. O trabalho de FERREIRA, *Voando com os Pilotos*, é um exemplo de utilização de duas metodologias.

A procura de informações: Se é fundamental, por exemplo, conhecer a renda dos trabalhadores de uma empresa, antes de inseri-la no questionário, é necessário verificar se esta informação é possível de se obter na empresa estudada, ou seja, há informações que estão disponíveis ou são de fácil recuperação, não necessitando de serem incluídas no questionário.

Referências bibliográficas

BERQUÔ, E.S., SOUZA, J.M.P., GOTILIEB, S.L.D. Amostragem. In: Bioestatística. São Paulo: EPU, 1980. Cap.6, p.133 - 45.

FERREIRA, L.L., BUSSACOS, M.A., SCHILTER, C.R.S. et al. Voando com os pilotos: condições de trabalho dos pilotos de uma empresa de aviação comercial. São Paulo: Associação dos Pilotos da VARIG, FUNDACENTRO, 1992. 89p.

NOVAES, T.C.P. Bases metodológicas para abordagem ocupacional ao benzeno. Dissertação (Mestrado em Química Analítica) - Instituto de Química, Universidade de São Paulo, 1992.

UNIVERSIDADE DE SÃO PAULO. Faculdade de Saúde Pública. Programa de estágio de campo multiprofissional. São Paulo, 1978. (mimeografado).

VIEIRA, S. Consultoria estatística. Ciência e Cultura, v.38, n.2, p.318-23, 1986.

7 EXEMPLO DE APLICAÇÃO ESTATÍSTICA NA ÁREA DE PNEUMOLOGIA

7.1 Análise da curva de Peak-Flow

O ambulatório de doenças ocupacionais pulmonares da FUNDACENTRO atende a diversos trabalhadores, e um de seus objetivos é diagnosticar a asma ocupacional. E a utilização do método estatístico se faz presente.

"Após o diagnóstico de asma e uma história clínica e ocupacional compatíveis com Asma Ocupacional, é necessário o estabelecimento do nexo causal, pois a Asma Ocupacional não é diretamente enquadrada como doença profissional na nossa legislação.

E o método de estabelecimento do nexo causal é a realização de medidas seriadas de pico de fluxo expiratório (curva de Peak-Flow). A condução da curva depende de dados da história clínica e ocupacional. Em situações bem identificadas, que causam sintomas imediatos, os registros podem ser de curta duração, alguns dias a uma semana, com medidas seriadas a cada hora ou a cada duas horas, registradas em um formulário, no qual o paciente além de anotar os 'picos', informa se neste dia estava trabalhando ou afastado" (ALGRANTI).

O método estatístico consiste em uma análise descritiva de um histograma e do teste de comparação entre duas médias (comparação entre o período trabalhado com o período em que esteve afastado).

Vamos supor a seguinte situação: "Um paciente masculino, com 35 anos, branco, apresentou-se no ambulatório. Há nove anos trabalhando como ajudante geral em uma granja de aves, tem como função o cuidado dos galinheiros e o preparo semanal da ração. Quatro anos após o início desse trabalho apresentou 'pneumonia' e a partir de então quadros de broncoespasmo, que surgiam horas após o preparo das rações, notadamente no verão, quando lidava com milho estocado" (ALGRANTI). O paciente registrou os picos de Peak-Flow durante 41 dias, sendo 15 dias trabalhando e o restante afastado do serviço.

A seguir serão mostrados os procedimentos estatísticos utilizados no ambulatório de pneumologia da FUNDACENTRO:

7.1.1 Análise gráfica:

Os pacientes anotam os picos no formulário e a partir desse formulário, é calculada a média diária dos picos de Peak-Flow. Após esse cálculo, um histograma é construído, identificando-se os dias de trabalho, afastamento e os sábados e domingos.

Para o exemplo citado, tem-se o seguinte quadro:

Período	Média	Período	Média	Período	Média	Período	Média
S	650	T	490	D	565	A	605
D	660	T	560	A	510	A	610
T	650	T	580	A	590	S	620
T	490	S	615	A	610	D	605
T	470	D	610	A	600	A	600
T	520	A	590	A	610	A	605
T	450	A	510	S	615	A	604
S	530	A	500	D	610	A	610
D	480	A	550	A	615		
T	460	A	605	A	620		
T	450	S	560	A	615		

e o respectivo gráfico:

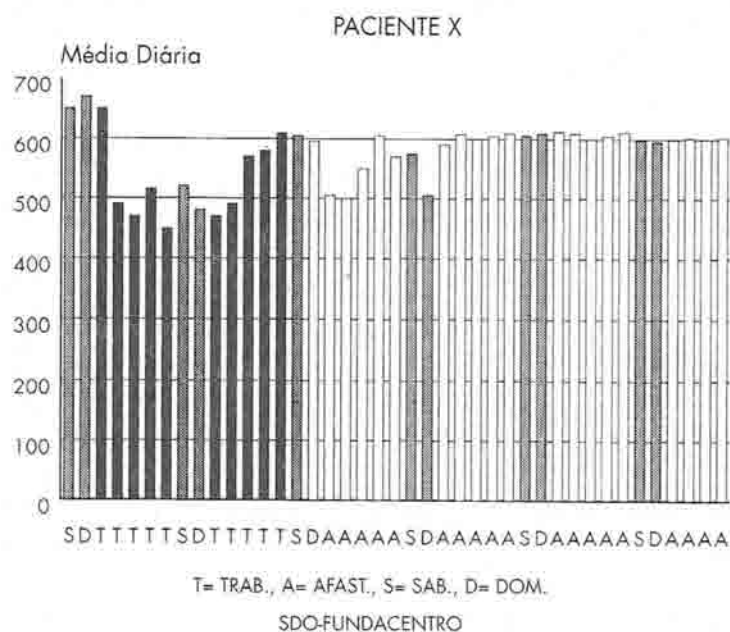


Figura 8. Curva de Peak-Flow.

A inspeção gráfica já orienta o pesquisador na evolução da curva de Peak-Flow, que, no exemplo, demonstra uma nítida queda do Peak-Flow, relacionada ao preparo da ração.

7.1.2 Comparação das duas médias: período trabalho x afastado

O objetivo da utilização desta comparação é verificar se as médias dos picos de Peak-Flow durante o período de trabalho são diferentes (menor) que as do período de afastamento. Assim, calculam-se a média e o desvio padrão para cada um dos períodos, incluindo-se os sábados e domingos.

Utilizando-se a fórmula fornecida no *Capítulo 3 de Teste de hipóteses* com os seguintes parâmetros:

Trabalho: Média (\bar{x}_t) = 537,00
Desvio Padrão (s_t) = 76,85
 $n = 15$

Afastado: Média (\bar{x}_a) = 590,15
Desvio Padrão (s_a) = 35,44
 $n = 26$

O desvio padrão ponderado será igual a $S_p = 54,0851$. Assim, tem-se o seguinte valor para o teste t'Student:

$t_o = -3,0310$ com 39 graus de liberdade, que resulta em $p = 2,157E-03$

Conclusão: As médias são estatisticamente diferentes, ou seja, as médias do período de afastamento são maiores que as do período de trabalho.

Essas duas análises estatísticas fornecem ao pesquisador subsídios importantes para o diagnóstico de Asma Ocupacional. Lembrando que não apenas essas duas análises devem ser feitas, sugerimos a leitura do artigo de ALGRANTI no qual este texto foi baseado.

Referências bibliográficas

ALGRANTI, E., MENDONÇA, E.M.C. Asma ocupacional. *Revista Brasileira de Clínica Terapêutica*, v.21, n.7, p.270-77, 1992.

**8 EXEMPLO DE PROGRAMA
INTERLABORATORIAL DE
CONTROLE DE QUALIDADE EM
ANÁLISES TOXICOLÓGICAS**

8.1 Introdução

“Dados analíticos são utilizados para tomar grande variedade de decisões. Resultados de análises toxicológicas servem para indicar exposição de indivíduos a agentes tóxicos, se eles estão ou não intoxicados, se foram tomadas medidas de higiene e engenharia adequadas para evitar ou diminuir as emissões destes agentes, e para o estabelecimento do nexo causal de uma doença ocupacional.

Dados incorretos podem surgir por inadequação na escolha da amostra, na maneira e no momento de colhê-la, nos métodos e técnicas utilizados, no registro e tratamento dos resultados, ou até na requisição de análise. Como consequência, as decisões que dependem destes resultados poderão ser também incorretas. Por exemplo, se a ausência de um agente tóxico é admitida erroneamente, o trabalhador permanecerá exposto ao xenobiótico, com risco de intoxicação.

Embora toda medida de grandeza física traga sempre consigo uma incerteza e qualquer tipo de determinação analítica tenha limitações inerentes, o conhecimento dos erros prováveis e o controle dos mesmos proporcionarão uma margem de confiança na qual os resultados numéricos obtidos serão **estimativas** estatísticas próximas de um valor verdadeiro.

Nenhum resultado analítico é absoluto. Diversos tipos de erro (determinados ou não) podem ser cometidos, produzindo variações nos resultados obtidos não só inter como intralaboratorialmente.

Um programa de controle de qualidade interlaboratorial tem como objetivo principal assegurar que os dados produzidos sejam válidos, segundo critérios estabelecidos que definem os intervalos de precisão e exatidão necessários para se atingirem determinados propósitos. Através da reunião e tratamento adequado dos resultados, pode ser feita a avaliação da performance dos laboratórios que participam do programa.

A demanda de análises toxicológicas para a área de Saúde Ocupacional tem crescido muito em nosso meio ultimamente, criando a necessidade de ações bem definidas no sentido de assegurar a qualidade dos resultados analíticos e possibilitar a comparação dos dados obtidos por diferentes métodos e laboratórios.

Quando se trabalha com amostras biológicas visando determinar índices de exposição sujeitos, via de regra, a diversas

interferências, não é tarefa muitas vezes simples estabelecer a exatidão e precisão dos dados obtidos. Mas com um programa de controle de qualidade, é possível estudar a influência sobre os resultados causada pela adoção de diferentes processos de colheita e tratamento de amostra, assim como de correção de dados, eliminando-se fontes de erro.

Com isso, torna-se viável não apenas o acompanhamento adequado de quadros individuais de saúde, mas também o estudo de grupos de indivíduos, a longo prazo, nos programas de vigilância epidemiológica" (KATO).

O programa interlaboratorial de Controle de Qualidade de Análise de Fenol Urinário (C.Q.A.), implantado em 1987 pela FUNDACENTRO, utiliza métodos estatísticos como um dos instrumentos de avaliação dos laboratórios participantes. A seguir, mostraremos os procedimentos para a análise estatística:

Por exemplo, uma amostra contendo fenol urinário foi enviada para 20 laboratórios participantes de um programa de controle de qualidade, tendo-se obtido a seguinte distribuição dos resultados:

Tabela 16. Distribuição dos resultados (duplicatas) da amostra X segundo os laboratórios participantes do programa.

Laboratório	Resultado	(mg/ml)
1	34	34
2	36	35
3	35	35
4	30	31
5	32	32
6	33	34
7	34	33
8	33	34
9	33	33
10	36	35
11	33	32
12	35	36
13	32	33
14	34	34
15	31	30
16	34	34
17	32	33
18	34	34
19	35	34
20	60	61

8.2 Identificação de outliers:

O termo "**outliers**" se traduz como um valor aberrante, discrepante, contaminado em relação a outros valores. A geração de um "outliers" em uma análise química pode dar-se por diversos fatores, tais como problemas na amostra, método analítico, equipamento sem calibração, entre outros.

A princípio, os resultados do laboratório 20 chamam a atenção. Será que estes resultados são "outliers"?

Há várias técnicas estatísticas para se detectar quando um ou mais valores de uma variável podem ser considerados "outliers" ou não e as utilizadas pelo C.Q.A. são:

a) $\bar{X} \pm 2 \times s$: Se a distribuição dos resultados pode ser considerada uma distribuição de probabilidade Normal, uma maneira simples de se eliminar um "outliers" é utilizar esse intervalo.

Primeiramente, calculam-se os dois parâmetros que identificam a distribuição, ou seja, média (\bar{x}) e o desvio padrão (s). Neste exemplo, tem-se:

$$\bar{x} = 34,8 \mu\text{g/ml} \quad \text{e} \quad s = 6,1 \mu\text{g/ml}$$

Construindo-se o intervalo $\bar{X} \pm 2 \times S$, tem-se: 22,6 a 47,0. Se os valores suspeitos de serem "outliers" não estiverem dentro deste intervalo (neste exemplo, os valores 60 e 61 estão fora do intervalo), então podem ser considerados "outliers" e devem ser eliminados para os cálculos posteriores da média e do desvio padrão.

b) **Estatística Dixon** : Essa estatística é baseada na distância do possível "outliers" a seu vizinho, ou vizinhos mais próximos. Para se utilizar essa estatística, é necessário ordenar os resultados (soma das duplicatas, quando houver), em ordem crescente, com o resultado de menor valor assumindo o posto 1 e assim sucessivamente. Para o exemplo, tem-se:

Laboratório	Σ das duplicatas	Posto
1	68	12,5
2	71	18
3	70	16
4	61	1,5
5	64	3
6	67	9
7	67	9
8	67	9
9	66	7
10	71	18
11	65	5
12	71	18
13	65	5
14	68	12,5
15	61	1,5
16	68	12,5
17	65	5
18	68	12,5
19	69	15
20	121	20

A soma 61, que é a menor, aparece duas vezes, assim o seu posto será:

$$p = (1 + 2) / 2 = 1,5 ,$$

o valor imediatamente superior é 64, que assume o posto 3 e assim sucessivamente.

Em função do número de resultados, que neste exemplo é igual a 20, para usar a estatística de Dixon, há um quociente que, aqui, será (YOUDEN):

$$D_n = \frac{X_n - X_{n-2}}{X_n - X_3} = \frac{X_{20} - X_{18}}{X_{20} - X_3} = \frac{121 - 71}{121 - 64} = 0,88$$

Por meio da tabela fornecida em YOUDEN, se o valor de D_n for superior ao valor crítico da tabela, o resultado X_n será um "outliers". Nesse exemplo, $D_n = 0,88$ e $D_c = 0,45$, logo a soma dos resultados do laboratório 20 é um "outliers". Portanto, para os cálculos posteriores, os resultados do laboratório 20 não serão considerados.

A etapa seguinte do procedimento adotado no C.Q.A., após terem sido eliminados os resultados aberrantes, é a estimação da concentração média.

8.3 Estimativa da concentração média

Quando um laboratório emite o resultado de uma análise, normalmente este valor é uma estimativa de um parâmetro, ou seja, é uma estimativa por ponto. A estimação por ponto consiste em fornecer a melhor estimativa possível para o parâmetro, sendo um valor único que corresponde a um ponto sobre o eixo de variação da variável.

Para proceder à estimativa por ponto, deve-se escolher o melhor estimador e, para a concentração média, parâmetro este usualmente utilizado pelos laboratórios, tem-se a estimativa por ponto da média populacional (μ), que é a concentração verdadeira. E o melhor estimador de que se dispõe é a média amostral (\bar{x}).

As estimativas por ponto são, em geral, utilizadas quando se necessita conhecer o valor do parâmetro para utilizá-lo em um teste estatístico ou na construção de uma carta de controle.

Entretanto, se a determinação de um dado parâmetro é a meta final do estudo estatístico, por exemplo, emitir um resultado de uma amostra, a estimação por ponto é insuficiente, pois a probabilidade de que a estimativa adotada venha a coincidir com o verdadeiro valor do parâmetro populacional é nula ou praticamente zero. Isso decorre de os estimadores serem variáveis aleatórias, muitas vezes contínuas, logo as estimativas obtidas quase certamente serão distintas do valor do parâmetro. Ou seja, é quase certo que se esteja cometendo um erro de estimação (COSTA NETO).

Devido a esse fato, surge a idéia de se construir um intervalo em torno da estimativa por ponto, de modo que esse intervalo tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro.

Se o número de resultados (n) for inferior ou igual a 30, a distribuição de probabilidade utilizada é a t -Student, com $n-1$ graus de liberdade e o intervalo de confiança para a média será:

$$\bar{X} - t_{n-1} \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} \times \frac{S}{\sqrt{n}}$$

onde \bar{X} = média, S = desvio padrão, t_{n-1} = valor de t' Student no nível de confiança pré-fixado e n = número de elementos.

Se o número de resultados for superior a 30, admita-se a distribuição de probabilidade Normal e o intervalo de confiança será (BUSSAB):

$$\bar{X} - z \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z \frac{S}{\sqrt{n}}$$

onde \bar{X} = média, S = desvio padrão, z = valor da distribuição Normal e n = número de elementos.

Os níveis de confiança usualmente utilizados são os de 95 e 99%.

No exemplo com os 20 laboratórios, após eliminados os resultados do laboratório 20, calculando-se novamente a média e o desvio padrão, tem-se:

Média (\bar{x}) = 33,5 mg/ml Desvio Padrão (s) = 1,5 mg/ml e
 n = 38 resultados

A estimativa da verdadeira concentração média desta amostra, ao nível de 95%, por exemplo, será:

sendo o número de elementos superior a 30 ($n = 38$), logo se utiliza a distribuição Normal, e o valor de z será igual a $z = 1,96$. Assim :

$$33,5 - 1,96 \cdot \frac{1,5}{\sqrt{38}} < \mu < 33,5 + 1,96 \cdot \frac{1,5}{\sqrt{38}}$$

$$33,0 \leq \mu \leq 34,0$$

Ou seja, tem-se 95% de confiança de que o intervalo contém o verdadeiro valor da concentração média de fenol urinário.

*Ver DUNCAN

8.4 Construção da carta de controle tipo Shewhart*

Para a montagem da carta de controle, utiliza-se a média e o desvio padrão obtidos no item anterior ($\bar{X} = 33,5$ e $s = 1,5$),

construindo-se dois intervalos, um ao nível de 95% e outro ao nível de 99%. Os resultados que estiverem fora do limite de 99% serão considerados discrepantes em relação aos demais resultados, ou seja, há algum problema no laboratório que o emitiu.

Esses intervalos conterão cerca de 95% a 99% dos resultados, respectivamente, e a distribuição de probabilidade Normal para estes níveis vale 1,96 e 2,58. Assim, os dois intervalos a serem construídos na carta de controle, neste exemplo, serão:

Intervalo de 95% : $\bar{X} \pm 1,96 \times s$

limite inferior: 30,6

limite superior: 36,4

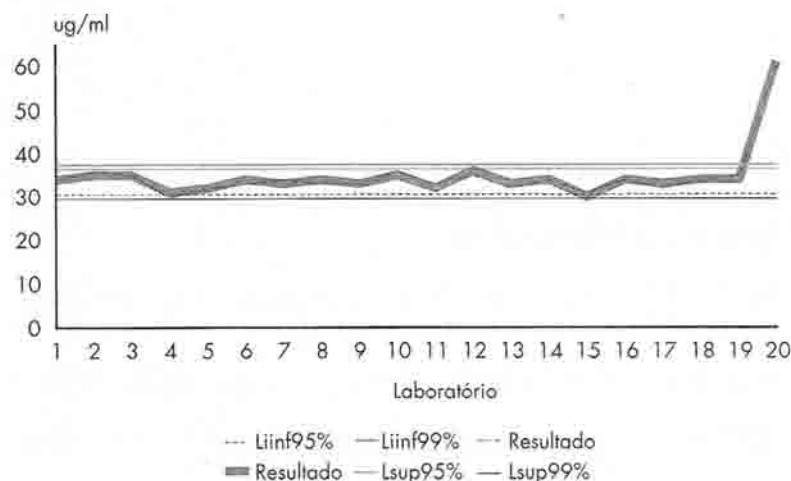
Intervalo de 99% : $\bar{X} \pm 2,58 \times s$

limite inferior: 29,6

limite superior: 37,4

Finalmente, constrói-se a carta de controle onde na ordenada tem-se os valores dos resultados obtidos pelos laboratórios, os limites dos dois intervalos (inferior e superior) e na abscissa, os laboratórios.

Carta de Controle



Pela análise do gráfico, verifica-se que o laboratório 20 é discrepante em relação aos demais laboratórios, ou seja, o resultado emitido por este laboratório é diferente.

Referências bibliográficas

BUSSAB, W.O., MORETTIN, P.A. Estatística básica. 3.ed. São Paulo: Atual, 1986.

COSTA NETO, P.O.L. Estatística. São Paulo: Edgard Blücher, 1977.

DUNCAN, A.J. Quality control and industrial statistic. [s.l.], [s.l.p.], 1974.

KATO, M. Programa interlaborial de controle de qualidade em análise toxicológica: documento interno. São Paulo: FUNDACENTRO, 1990.

YOUTEN, W.J., STELNER, E.H. Statistical manual of the Association of Official Analytical Chemists. Arlington : A.O.A.C., 1975.

**9 USO DA
DISTRIBUIÇÃO DE
POISSON NO CONTROLE
DE ANÁLISE DE FIBRAS
INORGÂNICAS POR
MICROSCOPIA ÓTICA**

9.1 Introdução

As concentrações de fibras inorgânicas no ar, em especial as fibras de asbesto existentes nos ambientes de trabalho, são geralmente determinadas pelo Método do Filtro-Membrana por Microscopia Ótica, mas a experiência tem mostrado que esse método nem sempre proporciona repetibilidade e reprodutibilidade de resultados (ASBESTOS INTERNATIONAL ASSOCIATION), mesmo porque os procedimentos de amostragem e análise por contagem são subjetivos, pois o resultado depende dos técnicos contadores.

Contudo esse é o método mais adequado para um monitoramento ambiental rotineiro de fibras dispersas no ar e, como tal, deve ser usado em conjunto com um programa de controle de qualidade analítico específico.

9.2 Procedimentos para avaliação de Fibras Inorgânicas

A amostra é coletada fazendo-se passar o ar por um filtro-membrana colocado em um dispositivo de coleta, e ligado a uma bomba de amostragem individual. A preparação do filtro-membrana para visualização no microscópio se faz por transparência do filtro-membrana e adição de um reagente para fornecer o contraste necessário durante a observação, usando um microscópio com contraste de fase.

As fibras são medidas, contadas, e a concentração ambiental é expressa em fibras por centímetro cúbico de ar, sendo calculada considerando-se o número de fibras sobre o filtro e o volume de ar amostrado.

A contagem das fibras sobre o filtro é realizada através de seleção dos campos aleatórios, procurando-se percorrer toda a área útil do filtro.

Durante a amostragem, preparação de amostras e contagem das fibras podem surgir problemas que o amostrador, preparador ou contador não percebiam, tais como a distribuição não uniforme de fibras sobre o filtro durante a amostragem, o deslocamento de fibras do filtro durante a preparação da amostra e uma contagem não fidedigna do filtro. Estes problemas podem ser identificados por meio da aplicação da distribuição de Poisson.

9.3 Distribuição de Poisson

Esta distribuição de probabilidade é de grande importância tanto teórica como prática. A sua aplicação surgiu, primeiramente, no campo da Biologia nas investigações associadas à contagem de células. Atualmente, é utilizada em vários campos, principalmente na Pesquisa Operacional em estudos referentes a chamadas telefônicas, emissões de partículas, etc.

Com relação à análise de fibras, sabe-se que a contagem segue a distribuição de Poisson e que o resultado de cada amostra deve-se apresentar de acordo com esta distribuição. Para tanto, devem ser calculadas as probabilidades de Poisson, definindo-se os parâmetros dessa distribuição, que são: a **média** (λ), que representa, nesta situação, a densidade de fibras por campo e o **número** (k), que representa o número de fibras encontradas por campo, onde k , pode variar de **zero** até um número n , de maneira que a probabilidade é definida como:

$$P(X = k) = \frac{(e^{-\lambda} \times \lambda^k)}{k!}, \quad K = 0, 1, 2, 3, \dots, n.$$

Assim, para verificar se os resultados obtidos de uma leitura estão ajustados à distribuição de Poisson, utiliza-se a **Prova de Aderência** [$N \cdot P(X = k)$, sendo N o número total de campos observados], a concordância entre a frequência observada e a frequência esperada é averiguada aplicando-se o teste de **Qui-Quadrado** (FELLER, HOEL).

Se o resultado de uma contagem não se enquadrar na distribuição de Poisson, provavelmente o técnico contador e/ou a amostra não estão bem adequados.

9.3.1 Exemplos de aplicação da distribuição de Poisson

Exemplo 1: Suponha que um técnico fez uma leitura de uma amostra e obteve o seguinte resultado:

Nº de fibras	Nº de campos* observados
0	16
1	27
2	19
3	9
4	4

* Entende-se por campo a área definida pelo retículo circular 3:1 do Walton & Beckett sobre o filtro-membrana.

Para calcular a densidade fibras/campo, faz-se o cálculo da média:

Nº fibras (x)	Nº campos (f)	x • f
0	16	0
1	27	27
2	19	38
3	9	27
4	4	16
Total	75	108

Assim, a média será: $\bar{x} = \frac{\sum x \cdot f}{n} = \frac{108}{75} = 1,44$

Este valor é a média λ da distribuição de Poisson, e o número total de campos observados foi igual a 75 :

$$N \cdot P(X = K) = 75 \cdot \frac{(e^{-1,44} \times (1,44)^K)}{K!}$$

Se o técnico contador (e/ou amostra) estiver adequado, qual será o número esperado de campos sem nenhuma fibra ?

Ou seja,
 $k = 0, \quad 75 \times (P X = 0) = 75 \times \frac{(e^{-1,44} \times (1,44)^0)}{0!} = 19 \text{ campos}$

Daí resulta que o número esperado de campos sem nenhuma fibra seja 19 campos.

Assim, para 1 fibra será: $K = 1$

$$75 \times \frac{(e^{-1,44} \times (1,44)^1)}{1!} = 26 \text{ campos}$$

Duas fibras: $K = 2$

$$75 \times \frac{(e^{-1,44} \times (1,44)^2)}{2!} = 18 \text{ campos}$$

Três fibras: $K = 3$

$$75 \times \frac{(e^{-1,44} \times (1,44)^3)}{3!} = 9 \text{ campos}$$

Quatro fibras: $K = 4$

$$75 \times \frac{(e^{-1,44} \times (1,44)^4)}{4!} = 18 \text{ campos}$$

Num quadro resumo, tem-se:

Nº de fibras	Nº de campos observados	Nº de campos esperados
0	16	19
1	27	26
2	19	18
3	9	9
4	4	3

Para aplicar a prova de aderência, utilizando-se o teste de Qui-Quadrado ($\chi^2 = \sum (O - E)^2 / E$, e: lembrando-se de que os valores encontrados no campo observado corresponde ao valor O (observado) e os valores encontrados no campo esperado correspondem ao valor E (esperado) do teste de Qui-Quadrado, tem-se:

Nº de fibras	Nº de campos observados (O)	Nº de campos esperados (E)	(O-E)	(O-E) ²	(O-E) ² /E
0	16	19	-4	16	0,4737
1	27	26	1	1	0,0385
2	19	18	1	1	0,0556
3	9	9	0	0	0,0
4	4	3	1	1	0,3333
Total	75	75			0,9011

O valor do Qui-Quadrado, com 4 graus de liberdade, ao nível de 5%, será igual ao valor 9,488 (consultando-se uma tabela com a distribuição do Qui-Quadrado do Anexo III). Assim, o Qui-Quadrado observado (valor calculado na tabela acima) é igual a 1,3525 que é menor que o valor crítico, concluindo-se daí que este resultado de contagem se ajusta à distribuição de Poisson.

A seguir, um exemplo em que o resultado de contagem não se ajuste à uma distribuição de Poisson:

Nº de fibras	Nº campos observados
0	18
1	50
2	26
3	5
4	1
Total	100

A densidade fibra/campo é igual a 1,21 e o número de campos observados é igual a 100, assim podem-se calcular os valores de campos esperados segundo o número de fibras, supondo-se uma distribuição de Poisson:

Para nenhuma fibra, $k = 0$, o número de campos esperado será:

$$100 \times \frac{(e^{-1,21} \times (-1,21)^0)}{0!} = 32 \text{ campos}$$

Para uma fibra, $k = 1$:

$$100 \times \frac{(e^{-1,21} \times (-1,21)^1)}{1!} = 36 \text{ campos}$$

Para duas fibras, $k=2$:

$$100 \times \frac{(e^{-1,21} \times (-1,21)^2)}{2!} = 21 \text{ campos}$$

Para três fibras, $k=3$:

$$100 \times \frac{(e^{-1,21} \times (-1,21)^3)}{3!} = 8 \text{ campos}$$

Para quatro fibras, $k=4$:

$$100 \times \frac{(e^{-1,21} \times (-1,21)^4)}{4!} = 3 \text{ campos}$$

Tem-se o seguinte quadro para o cálculo do Qui-Quadrado:

Nº de fibras	Nº campo observado	Nº campo esperado	(O-E)	(O - E)	(O-E)²/E
0	18	32	-14	196	6,125
1	50	36	14	196	5,444
2	26	21	5	25	0,429
3	5	8	-3	9	1,125
4	1	3	-2	4	1,333
Total	100	100			14,456

Conclusão:

O valor do Qui-Quadrado observado (14,456) é maior que o valor do Qui-Quadrado crítico (com 4 graus de liberdade e nível de 5%, $X = 9,4888$), logo a leitura não se ajusta a Poisson. Nesse exemplo podem ter ocorrido problemas na amostra ou na contragem.

9.4 Comentário final

Dentro das atividades de um laboratório de microscopia, normalmente, um lote de amostras é contado por um mesmo técnico, sendo que um número mínimo de amostras é recontado (DICKER). Portanto, em cada resultado de contagem deve-se utilizar a distribuição de Poisson como um controle na identificação dos problemas apontados. Somente devem ser aceitos os resultados de contagem que estiverem adequados à distribuição de Poisson, as amostras que não tiverem seus resultados de acordo devem ser recontadas, de preferência por um outro técnico, para que se possa identificar a provável causa da não aderência à distribuição de Poisson.

Se dois técnicos contadores não conseguirem obter resultados compatíveis, provavelmente a amostra não estava adequada para análise, tendo havido problema na amostragem ou na preparação da amostra; se um técnico contador não fornecer resultado satisfatório, para uma mesma amostra, e o outro forneceu, então é provável que o primeiro tenha tido problemas na contagem.

Referências bibliográficas

ASBESTOS INTERNATIONAL ASSOCIATION. Reference method for the determination of airborne asbestos fibre concentrations at workplaces by the light microscopy (membrane filter method). London, 1982 (AIA. Health and Safety Publication. Recommended Technical Method. No.1 (RTM 1).

DICKER, W.L. Asbestos identification and sampling. Occupational Health in Ontario, v.3, n.3, p.154-72, July 1982.

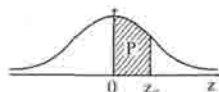
FELLER, W. An introduction to probability theory and its application. 2.ed. New York : John Wiley & Sons, 1965.

HOEL, P.G. Introduction to probability theory. Los Angeles : Houghton Mifflin Co., 1971.

ANEXOS

- 1) Tabela da distribuição Reduzida Normal**
- 2) Tabela da distribuição t'Student**
- 3) Tabela da distribuição Qui-Quadrado**
- 4) Tabela de distribuição F'Snedecor**

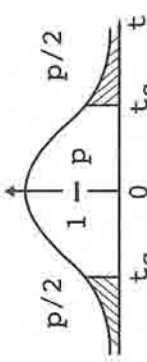
Distribuição normal reduzida: $N(0;1)$
 Probabilidades p tais que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda Decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
0,0	$p = 0$										0,0
0,1	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,1
0,2	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,2
0,3	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,3
0,4	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,4
0,5	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,5
0,6	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,6
0,7	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,7
0,8	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,8
0,9	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,9
1,0	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	1,0
1,1	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,1
1,2	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,2
1,3	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,3
1,4	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,4
1,5	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,5
1,6	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,6
1,7	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,7
1,8	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,8
1,9	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,9
2,0	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	2,0
2,1	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,1
2,2	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,2
2,3	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,3
2,4	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,4
2,5	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,5
2,6	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,6
2,7	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,7
2,8	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,8
2,9	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,9
3,0	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	3,0
3,1	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,1
3,2	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,2
3,3	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,3
3,4	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,4
3,5	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,5
3,6	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,6
3,7	49984	49985	49985	49986	49987	49987	49987	49988	49988	49989	3,7
3,8	49989	49990	49990	49991	49991	49992	49992	49992	49992	49992	3,8
3,9	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,9
4,0	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	4,0
4,1	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,1
4,2	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,2
4,3	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,3
4,4	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,4
4,5	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5
parte inteira e primeira decimal de Z_c	Segunda e terceira Decimais de Z_c										parte inteira e primeira decimal de Z_c
	05	15	25	35	45	55	65	75	85	95	
0,0	$p = 0$										0,0
0,1	00199	00598	00997	01396	01795	02193	02591	02989	03387	03784	0,1
0,2	04181	04578	04974	05369	05764	06159	06553	06946	07339	07730	0,2
0,3	08121	08512	08901	09290	09677	10064	10450	10834	11218	11600	0,3
0,4	11982	12362	12741	13119	13495	13871	14244	14617	14988	15358	0,4
0,5	15726	16093	16458	16822	17184	17545	17903	18261	18500	18970	0,5
0,6	19322	19672	20021	20368	20712	21055	21396	21735	22073	22408	0,6
0,7	22741	23072	23401	23729	24054	24377	24697	25016	25333	25647	0,7
0,8	25959	26270	26577	26883	27186	27488	27786	28083	28377	28669	0,8
0,9	28959	29246	29531	29814	30094	30372	30648	30921	31192	31461	0,9
1,0	31727	31990	32252	32511	32767	33021	33273	33522	33769	34013	1,0

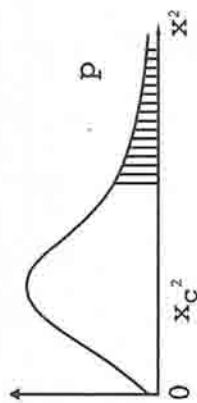
Distribuição de t' Student: St(n)

Valores críticos de t' tais que $P(-t_c < t < t_c) = 1 - p$

Graus de liberdade		Distribuição de t'Student: St(n) Valores críticos de t'tais que P(tc < t < tc) = 1 - p															Graus de liberdade
p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%			
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,657	318,309	636,619	1	
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,598	2	
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	3	
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,998	3,747	4,604	7,173	8,610	4	
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,756	3,365	4,032	5,893	6,869	5	
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6	
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7	
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,401	8	
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9	
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10	
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	3,925	4,437	11	
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12	
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13	
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14	
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,248	2,602	2,947	3,733	4,073	15	
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	16	
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17	
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18	
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19	
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20	
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21	
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22	
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23	
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24	
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,166	2,485	2,787	3,450	3,725	25	
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26	
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	27	
28	0,127	0,256	0,389	0,530	0,684	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28	
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	29	
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30	
35	0,126	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35	
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	40	
50	0,126	0,254	0,387	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,409	2,678	3,261	3,496	50	
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,466	60	
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	120	
x	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	x	
p=90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%			

FONTE: Bussab, W e Morelhi, P.A. - Estatística básica

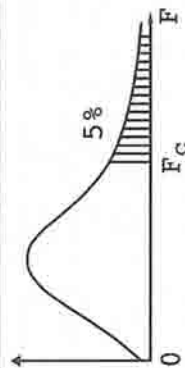
Distribuição de Qui-Quadrado: $\chi^2(n)$
Valores críticos de Qui-Quadrado, tais que $P(\chi^2 < \chi^2_c) = p$



Graus de liberdade n	p=99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%	Graus de liberdade n
1	0,016	0,063	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	4,218	5,024	5,412	6,635	9,550	10,827	1
2	0,020	0,040	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	6,438	7,378	7,824	9,210	12,429	13,815	2
3	0,115	0,185	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	8,311	9,348	9,837	11,345	14,796	16,266	3
4	0,297	0,429	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	10,026	11,143	11,668	13,277	16,924	18,467	4
5	0,554	0,752	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	11,644	12,832	13,388	15,086	18,907	20,515	5
6	0,872	1,134	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	13,198	14,449	15,033	16,812	20,791	22,457	6
7	1,239	1,564	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	14,703	16,013	16,662	18,475	22,601	24,322	7
8	1,646	2,032	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	16,171	17,534	18,168	20,090	24,352	26,125	8
9	2,088	2,532	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	17,608	19,023	19,679	21,666	26,056	27,877	9
10	2,558	3,059	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	19,021	20,483	21,161	23,209	27,722	29,588	10
11	3,053	3,609	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	20,412	21,920	22,618	24,725	29,354	31,264	11
12	3,571	4,178	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	21,785	23,337	24,054	26,217	30,957	32,909	12
13	4,107	4,765	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	23,142	24,736	25,472	27,688	32,535	34,528	13
14	4,660	5,368	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	24,485	26,119	26,873	29,141	34,091	36,123	14
15	5,229	5,985	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	25,816	27,488	28,259	30,578	35,628	37,697	15
16	5,812	6,614	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	27,136	28,845	29,633	32,000	37,146	39,252	16
17	6,408	7,255	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	28,445	30,191	30,995	33,409	38,648	40,790	17
18	7,015	7,906	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	29,745	31,526	32,346	34,805	40,136	42,312	18
19	7,633	8,567	8,906	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	31,037	32,852	33,687	36,191	41,610	43,820	19
20	8,260	9,237	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	32,321	34,170	35,020	37,566	43,072	45,315	20
21	8,897	9,915	10,283	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	33,597	35,479	36,343	38,932	44,522	46,797	21
22	9,542	10,600	10,982	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	34,867	36,781	37,659	40,268	45,962	48,288	22
23	10,196	11,293	11,688	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	36,131	38,076	38,968	41,638	47,391	49,728	23
24	10,856	11,992	12,401	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	37,389	39,360	40,270	42,940	48,812	51,179	24
25	11,524	12,697	13,120	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	38,642	40,646	41,566	44,314	50,223	52,620	25
26	12,198	13,409	13,844	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	39,889	41,923	42,856	45,642	51,627	54,052	26
27	12,879	14,125	14,573	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	41,132	43,194	44,140	46,963	53,022	55,476	27
28	13,565	14,847	15,308	16,928	18,939	21,588	23,647	27,336	31,319	34,027	37,916	41,337	42,370	44,461	45,419	48,278	54,411	56,893	28
29	14,256	15,574	16,047	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	43,604	45,722	46,693	49,588	55,792	58,302	29
30	14,953	16,306	16,791	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	44,834	46,979	47,962	50,892	57,167	59,703	30
p=99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%		

FONTE: Bussab, W e Morelli, P.A. - Estatística básica

Distribuição de Fisher. Snedecor - F
Valores críticos de F_r tais que $P(F > F_c) = 0,05$



Graus de liberdade do denominador de F_r : n_2		Graus de liberdade do numerador de F_r : n_1																				Graus de liberdade do denominador de F_r : n_2	
1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	24	30	40	60	120	∞	1	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,4	246,5	247,3	248,0	249,1	250,1	251,1	252,2	253,3	254,3	1	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,44	19,45	19,45	19,46	19,47	19,48	19,49	19,50	2	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,72	8,70	8,69	8,68	8,64	8,62	8,59	8,57	8,55	8,53	3	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,86	5,84	5,82	5,77	5,75	5,72	5,69	5,66	5,63	4	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,62	4,60	4,58	4,56	4,53	4,50	4,46	4,43	4,40	5	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,94	3,92	3,90	3,87	3,84	3,81	3,77	3,74	3,70	6	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,51	3,49	3,47	3,44	3,41	3,38	3,34	3,30	3,27	7	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,22	3,20	3,17	3,15	3,12	3,08	3,04	3,01	2,97	8	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	3,01	2,99	2,96	2,94	2,90	2,86	2,83	2,79	2,75	9	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,87	2,85	2,83	2,80	2,77	2,74	2,70	2,66	2,62	2,58	10	2,58
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,72	2,70	2,67	2,65	2,61	2,57	2,53	2,49	2,45	11	2,45
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,62	2,60	2,57	2,54	2,51	2,47	2,43	2,38	2,34	12	2,34
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,53	2,52	2,48	2,46	2,42	2,38	2,34	2,30	2,25	13	2,25
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,46	2,44	2,41	2,39	2,35	2,31	2,27	2,22	2,18	14	2,18
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,40	2,39	2,35	2,33	2,29	2,25	2,20	2,16	2,11	15	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,35	2,33	2,30	2,28	2,24	2,19	2,15	2,11	2,06	16	2,06
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,34	2,31	2,29	2,26	2,23	2,19	2,15	2,10	2,06	2,01	17	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,27	2,25	2,22	2,19	2,15	2,11	2,06	2,02	1,97	18	1,97
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,23	2,22	2,18	2,16	2,11	2,07	2,03	1,98	1,93	19	1,93
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,20	2,18	2,15	2,12	2,08	2,04	2,00	1,95	1,90	20	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,18	2,16	2,12	2,10	2,05	2,01	1,96	1,92	1,87	21	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,15	2,13	2,10	2,07	2,03	1,98	1,94	1,89	1,84	22	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,13	2,11	2,08	2,05	2,01	1,96	1,91	1,86	1,81	23	1,81
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,11	2,09	2,05	2,03	1,98	1,94	1,89	1,84	1,79	24	1,79
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,09	2,07	2,04	2,01	1,96	1,92	1,87	1,82	1,77	25	1,77
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,07	2,05	2,02	1,99	1,95	1,90	1,85	1,80	1,75	26	1,75
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,06	2,04	2,00	1,97	1,93	1,88	1,84	1,79	1,73	27	1,73
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,04	2,02	1,99	1,96	1,91	1,87	1,82	1,77	1,71	28	1,71
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,03	2,01	1,97	1,94	1,90	1,85	1,81	1,75	1,70	29	1,70
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	2,01	1,99	1,96	1,93	1,89	1,84	1,79	1,74	1,68	30	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,92	1,90	1,87	1,84	1,79	1,74	1,69	1,64	1,58	40	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,84	1,81	1,78	1,75	1,70	1,65	1,59	1,53	1,47	60	1,47
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,77	1,75	1,72	1,69	1,66	1,61	1,55	1,50	1,43	1,35	120	1,35
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,69	1,67	1,63	1,60	1,57	1,52	1,46	1,39	1,32	1,22	∞	1,22
1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	24	30	40	60	120	∞	1	∞

FONTE: Bussab, W e Morethi, P.A. - Estatística básica

SOBRE O LIVRO

Formato: 16x23 cm. *Tipologia:* Futura. *Papel:* Offset 90g/m2 (miolo). Papel Couché 210g/m2 (capa). *Impressão:* JCA. *Tiragem:* 2.000. *1ª Edição:* 1997.

EQUIPE DE REALIZAÇÃO

Supervisão editorial: Elisabeth Rossi. *Revisão gramatical:* WW Produções. *Editoração Eletrônica:* Fábio Cavalotti. *Criação da Capa:* Alfredo Maeoca.

FUNDACENTRO

FUNDAÇÃO JORGE DUPRAT FIGUEIREDO
DE SEGURANÇA E MEDICINA DO TRABALHO

Rua Capote Valente, 710 - São Paulo - SP

CEP - 05409-002

