ESTATÍSTICA DESCRITIVA E INFERENCIAL

BREVES NOTAS

Pedro Lopes Ferreira

2005

ÍNDICE

1 Pensar em termos estatísticos

1.1 Variável

Os dados são os ingredientes básicos da estatística enquanto disciplina definida como a colecção, organização, sumário, análise e interpretação dos dados.

Comecemos por definir alguns termos básicos. Assim, entidade é uma pessoa, local, data, hora ou coisa que fornece o atributo, contagem, ou a medição de interesse. Exemplos de entidades são, portanto:

- o número de empresas com a sede em Coimbra;
- o número de pessoas que sofreram de enfarto do miocárdio em Portugal, no ano de 2001;
- o número de automóveis que passam, por hora, num determinado cruzamento.

Algumas destas entidades estão associadas a contagens, outras a medições, outras ainda a atributos, representando todas elas um valor para um conceito. Para definir um conceito, usamos o termo variável como sendo uma qualidade ou quantidade com um valor alterável por entidade. A seguir apresentam-se alguns exemplos de variáveis, possíveis valores e entidades.

Variável	Possíveis valores	Entidade
Idade	30, 31, 32	pessoa
Sexo	feminino, masculino	pessoa
forma de pagamento	numerário, cheque, cartão de crédito	transacção
Metros quadrados da sala	25, 30, 50, …	casa
tipo de media	rádio, televisão, jornal, …	media

Quadro 1 - Exemplos de variáveis, valores e entidades

Como a entidade está relacionada com a unidade onde a expressão é realizada, este conceito está normalmente associado à palavra "por". Assim, a variável "número de horas de trabalho por dia" sugere a entidade dia. Do mesmo modo, à variável "número de horas de trabalho de um operário por mês" estão associadas duas entidades, pessoa e mês.

1.2 Quantitativo vs. qualitativo

É fácil ver diferenças entre os possíveis valores de duas variáveis como idade e sexo. Nesta última, os valores possíveis são palavras; na primeira, são números. Estes dois estilos de valores reflectem dois tipos diferentes de dados: quantitativos e qualitativos.

De uma maneira mais geral pode dizer-se que sempre que as entidades conduzem a etiquetas, categorias, atributos ou códigos, os dados resultantes são considerados qualitativos. Quando as entidades produzem uma contagem ou uma medição, os dados são considerados quantitativos.

Por vezes, a introdução de códigos constitui uma excepção a esta dicotomia números-palavras. Outra situação de excepção é o chamado dado ordenado (ranked data). Os respectivos valores estão ordenados, por exemplo de muito boa a muito má qualidade ou o número de estrelas da classificação de um hotel. Os dados ordenados e os códigos, apesar de numéricos, são normalmente considerados qualitativos.

1.3 Escalas de medição

A determinação de qual a análise estatística mais apropriada para os dados de uma determinada variável depende da escala de medição usada para essa variável. Existem quatro escalas de medição: nominal, ordinal, intervalar e de razão (ou proporcional). A escala de medição determina a quantidade de informação contida nos dados e indica quais as formas mais apropriadas para sumariar os dados e quais as análises estatísticas mais convenientes. Iremos, de seguida, descrever as quatro escalas de medição.

1.3.1 Escala nominal

A escala de medição para uma variável é nominal quando as observações para a variável são apenas etiquetas usadas para identificar (nomear, dar um nome a) um atributo de cada elemento. Exemplos de escalas nominais são as seguintes:

- sexo (masculino, feminino);
- estado civil (solteiro, casado, viúvo, divorciado);
- religião (várias possibilidades);
- código de peça (A13622, 12B63);
- situação profissional (empregado, desempregado);
- número de polícia (5654, 2712, 35, 624).

É importante notar que as operações aritméticas como a adição, a subtracção, a multiplicação e a divisão não fazem qualquer sentido em dados nominais. Assim, mesmo quando os dados nominais são numéricos, cálculos como soma ou média não são admissíveis.

1.3.2 Escala ordinal

A escala de medição para uma variável é ordinal quando (1) os dados têm propriedades nominais e (2) podem ser usados para ordenar as observações nessa variável. Um exemplo de uma escala ordinal é a utilizada em vários questionários de satisfação ou de avaliação da qualidade. Consideremos o seguinte caso de um questionário usado num restaurante:

	Excelente	Boa	Má	Comentário
Comida	❑	❑	❑	
Bebidas	❑	❑	❑	
Serviço	❑	❑	❑	
Empregado/a	❑	❑	❑	

Quadro 2 – Escalas ordinais

Neste caso pede-se aos clientes que avaliem a qualidade da comida, das bebidas, do serviço em geral e da forma como foram atendidos pelo/a empregado/a. As categorias de resposta são, para cada variável, excelente, boa e má.

As observações para cada variável possuem as características de dados nominais (cada resposta é uma forma de etiquetar como excelente, boa ou má a qualidade do serviço). Para além disso, as observações podem ser ordenadas em termos de qualidade. Por exemplo, considerando a variável 'qualidade da comida' e após recolher os dados, podemos ordenar as observações em termos de qualidade de comida começando com as observações de excelente comida, seguidas pelas observações de boa e, finalmente, pelas observações de má comida.

Tal como os dados obtidos de uma escala nominal, os obtidos de uma escala ordinal podem ser numéricos ou não numéricos. No caso anterior poderíamos usar os valores E, B e M ou, respectivamente, os códigos 1, 2 e 3. Também aqui não faz sentido qualquer manipulação obtida por operador aritmético.

1.3.3 Escala intervalar

A escala de medição para uma variável é intervalar quando (1) os dados têm propriedades ordinais e (2) o intervalo entre as observações pode ser expresso em termos de uma unidade fixa de medida.

A temperatura é um bom exemplo de uma variável que usa uma escala intervalar de medição. A unidade fixa de medida é o grau e uma observação registada num determinado ponto do tempo será um valor numérico que especifica a quantidade de graus. Estes dados possuem as propriedades dos dados ordinais, podendo os vários valores de temperatura ser ordenados do mais frio para o mais quente.

Para além disto, a escala intervalar possui a propriedade que o intervalo entre observações pode ser expresso em termos de uma unidade fixa de medida. Deste modo, o intervalo entre 35 e 40 graus é de 5 graus, o intervalo entre 35 e 85 graus é de 50 graus e o intervalo entre 85 e 90 graus é de 5

graus. Com dados nominais e ordinais, tais diferenças entre observações não tinham qualquer significado.

A unidade fixa de medida exigida por uma escala intervalar, significa que os dados têm necessariamente de ser numéricos. Então, já faz sentido somar, subtrair, multiplicar e dividir.

1.3.4 Escala de razão

A escala de medição para uma variável é de razão quando (1) os dados têm propriedades intervalares e (2) faz sentido dividir duas observações.

As variáveis distância, peso, comprimento e tempo medem-se através de escalas de razão, exigindo necessariamente a presença de um zero, representando a não existência de valor.

Consideremos uma variável indicando o preço de um automóvel. O ponto zero corresponde ao valor de um automóvel sem preço (gratuito). Deste modo, comparando o preço de 35.000€ com o de 17.500€ se pode deduzir que o primeiro custa duas vezes mais do que o segundo.

Os dados obtidos por uma escala de razão são também sempre numéricos.

1.4 Amostra vs. população

Outro par de conceitos que necessita ser mencionado é o de população e amostra. Se o conjunto de dados que possuímos é exaustivo dizemos simplesmente que forma uma população. Neste caso, toda e qualquer entidade tem de estar presente. Quando algumas, mas não todas, as entidades estão presentes, a colecção de valores obtidos denomina-se amostra.

Normalmente, para designarmos o número de entidades (tamanho) de uma população finita, usamos a letra maiúscula N e a letra minúscula n representa o número de entidades na amostra.

1.5 Estatística descritiva vs. inferencial

Há essencialmente dois tipos de procedimentos em estatísticas. A estatística descritiva tem como objectivo a descrição dos dados, sejam eles de uma amostra ou de uma população. Pode incluir:

- verificação da representatividade ou da falta de dados;
- ordenação dos dados;
- compilação dos dados em tabela;
- criação de gráficos com os dados;
- calcular valores de sumário, tais como médias;
- obter relações funcionais entre variáveis.

A estatística inferencial, o segundo tipo de procedimentos em estatística, preocupa-se com o raciocínio necessário para, a partir dos dados, se obter conclusões gerais. O seu objectivo é obter uma afirmação acerca de uma população com base numa amostra. Estas inferências ou generalizações podem também ser de dois tipos: estimações ou decisões (testes de hipóteses).

1.6 Amostragem

A amostragem é o processo pelo qual recolhemos dados. Isto dá-nos apenas uma imagem da população em estudo. No entanto, e independentemente da correcção dos processos usados, para recolher a amostra, há sempre a considerar o chamado erro de amostragem. Devemos sempre esperar algumas diferenças entre a amostra e a população.

Por outro lado, por exemplo, o erro pode residir não só na amostragem, mas também nos próprios dados. Erros não amostrais acontecem quando os valores recolhidos não pertencem aos valores possíveis da entidade (exemplo: registado o valor 21 para uma nota, quando deveria ter sido 12) ou quando apenas uma pequena proporção da população é recolhida.

1.7 Tipos de amostragem

De uma maneira geral, os tipos de amostragem podem ser de dois tipos: aleatórias e não aleatória. O método de amostragem não aleatória consiste em seleccionarmos entidades através de escolha pessoal. As amostras não aleatórias incluem:

- as de opinião quando as entidades são escolhidas porque compõem uma amostra representativa (os habitantes de duas freguesias podem ser usados como representativos dos eleitores de uma zona mais ampla do país, por exemplo);
- as de conveniência quando escolhemos as entidades apenas estas estarem mais próximas de nós (escolhemos os alunos de uma turma quando pretendemos obter a opinião de todos os alunos de uma escola;
- as de quota quando os elementos que compõem a amostra são de determinadas características (se soubermos que os consumidores de um determinado produto são 60% do sexo feminino, podemos dizer a um inquiridor que esteja à porta de um supermercado para entrevistar 60 pessoas do sexo feminino e 40 do sexo masculino, cabendo-lhe a ele a decisão de escolher quem entrevista).

Porque dependem de escolha pessoal, as amostras não aleatórias podem efectivamente não ser representativas de uma população, sendo difícil o cálculo do erro amostral. Para ultrapassarmos este problema, as amostras aleatórias deixam a escolha ao acaso, tendo em princípio cada elemento da população a mesma hipótese de ser escolhido.

Há quatro tipos de amostragens aleatórias. A primeira delas é a chamada amostra aleatória simples de tamanho n onde não só cada elemento da população tem as mesmas hipóteses de ser escolhido, como também qualquer conjunto de tamanho n pode ser escolhido.

Um dos instrumentos usados para se obter uma amostra aleatória simples de tamanho n é a chamada tabela de números aleatórios. A Tabela A1 (em anexo) é um exemplo de uma tabela de números aleatórios usada para identificar entidades se $N \leq 100$.

Para exemplificar o uso desta tabela, vejamos o caso de uma amostra de tamanho n = 6 de uma população de tamanho N = 76. De início, temos de especificar o ponto de partida e a orientação da leitura na tabela. Para ponto de partida suponhamos que escolhemos dois dígitos (1 e 8 por exemplo) que correspondem aos números de linha e de coluna na tabela. Na intersecção da 1ª linha com a 8ª coluna encontramos o valor 40, o primeiro elemento da nossa amostra. De seguida escolhemos o percurso na tabela (de cima para baixo, por exemplo). É então fácil identificarmos os seguintes cinco elementos: 61, 97, 12, 58 e 27. Como o tamanho da população é 76, é impossível recolhermos o elemento de ordem 97. Assim, ignoramos este número e escolhemos o próximo número possível, no nosso caso, exactamente o número 76.

Procedimento semelhante pode ser usado na Tabela A2 para obtermos amostras da população onde $N \leq 1000$.

O segundo tipo de amostragem é a amostragem estratificada. Neste tipo de amostragem, as entidades são agrupadas em estratos segundo características físicas ou materiais. Para assegurar que todos os estratos da população estudantil afectados por determinado diploma sejam considerados, escolhem-se, por exemplo, uma amostra aleatória de estudantes de cada um dos tipos de ensino: básico, secundário e superior. Uma única amostra aleatória simples não poderia garantir esta representação de estudantes dos três tipos de ensino.

Um terceiro tipo de amostragem é a chamada amostragem por cachos. Aqui, as entidades são classificadas em grupos ou cachos e é seleccionada uma amostra aleatória de cachos. Um censo (de toda a população) é então conduzido dentro dos cachos seleccionados.

Por fim, a amostra sistemática selecciona todas as entidades de ordem k numa população finita de tamanho N. Normalmente k é o valor arredondado de N/n.

1.8 Algumas ideias úteis

As estatísticas normalmente começam com um conjunto de medições em membros de uma população. O quadro 3 apresenta as condições atmosféricas no Outono em 23 cidades do Sul de Inglaterra.

Cidade	Sol (horas)	Chuva(poleg)	Temp (ºC)	Tempo (dia)
Folkestone	–	–	13,5	Nublado
Hastings	0,1	–	14	Nublado
Eastbourne	2,4	–	14,5	Claro
Brighton	0,1	–	14	Nublado
Worthing	1,2	–	14,5	Nublado
Littlehampton	1,4	–	14	Nublado
Bognor Regis	1,4	–	14,5	Claro
Southsea	1,9	–	15	Nublado
Sandown	1,8	–	15	Claro
Shanklin	0,8	–	15	Nublado
Ventnor	2,3	–	15	Claro
Bournemonth	3,8	–	15,5	Sol à tarde
Poole	4,4	–	16	Sol à tarde
Swanage	3,8	–	15	Sol à tarde
Weymonuth	3,9	0,01	15	Sol à tarde
Exmouth	3,1	0,04	15	Sol à tarde
Teignmouth	5,0	0,14	15,5	Sol à tarde
Torquay	3,7	0,15	15,5	Claro tarde
Falmouth	3,7	0,23	15,5	Chuva
Penzance	4,1	0,24	15,5	Chuva
Isles of Scilly	2,4	0,26	14,5	Sol à tarde
Jersey	5,0	0,04	16	Claro
Guernsey	5,8	–	16,5	Sol à tarde

Quadro 3 - Condições atmosféricas no Outono em 23 cidades do Sul de Inglaterra.

Nesta tabela vemos que os valores da temperatura variam entre 13.5 e 16.5. Uma boa maneira de ter uma boa imagem sobre estas temperaturas é através de um histograma de temperaturas.

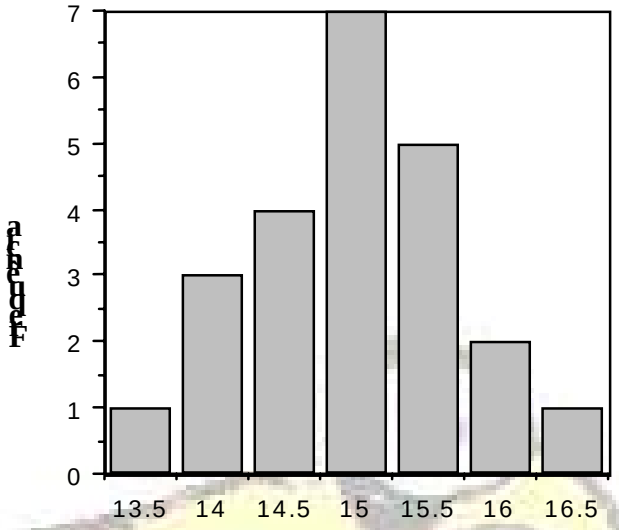


Figura 1 – Distribuição da temperatura

Vemos agora, de uma maneira mais fácil, que a temperatura mais frequente no Outono na costa Sul da Inglaterra é de 15º C.

Usando a fórmula $\overline{x} = \frac{\sum x_i}{n}$ poderíamos concluir que a temperatura média é de 14.98 ºC quase coincidindo com a temperatura mais frequente.

Para termos uma ideia da dispersão dos dados em relação à média, podemos determinar as diferenças de temperatura em cada cidade em relação à média. Há, no entanto, consequências matemáticas desagradáveis de se proceder deste modo e que resulta do facto de que, quando queremos ter uma ideia do desvio total, todos os desvios somam zero, sendo portanto, o desvio médio também igual a zero. É capaz de demonstrar?

Ultrapassa-se este problema calculando os quadrados de todos os desvios, calculando a média destes quadrados (a variância) e depois extraindo a raiz quadrada (desvio padrão). Para se calcular o desvio padrão (e a variância) é necessário calcular-se o seu desvio padrão em relação à média.

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

Cidade	Temp (ºC)	Desvios da média	Desvios quadrados
Folkestone	13,50	-1,48	2,19
Hastings	14,00	-0,98	0,96
Eastbourne	14,50	-0,48	0,23
Brighton	14,00	-0,98	0,96
Worthing	14,50	-0,48	0,23
Littlehampton	14,00	-0,98	0,96
Bognor Regis	14,50	-0,48	0,23
Southsea	15,00	0,02	0,00
Sandown	15,00	0,02	0,00
Shanklin	15,00	0,02	0,00
Ventnor	15,00	0,02	0,00
Bournemonth	15,50	0,52	0,27
Poole	16,00	1,02	1,04
Swanage	15,00	0,02	0,00
Weymonuth	15,00	0,02	0,00
Exmouth	15,00	0,02	0,00
Teignmouth	15,50	0,52	0,27
Torquay	15,50	0,52	0,27
Falmouth	15,50	0,52	0,27
Penzance	15,50	0,52	0,27
Isles of Scilly	14,50	-0,48	0,23
Jersey	16,00	1,02	1,04
Guernsey	16,50	1,52	2,31
TOTAL	344,50	-0,04	11,73
MÉDIA	14,98	0,00	0,51

Quadro 4 - Desvios em relação à temperatura média nas cidades costeiras do Sul

Um outro conceito que joga com os conceitos de média e de desvio padrão é o de distribuição normal. Trata-se de uma distribuição unimodal e simétrica com algumas propriedades matemáticas bem conhecidas. Entre elas destacam-se as apresentadas na figura abaixo onde aproximadamente 68% de todos os valores de uma variável normalmente distribuída estão entre ±1 desvio padrão da média; aproximadamente 95% entre ±2 d.p. e cerca de 99% está entre ±3 d.p. da média

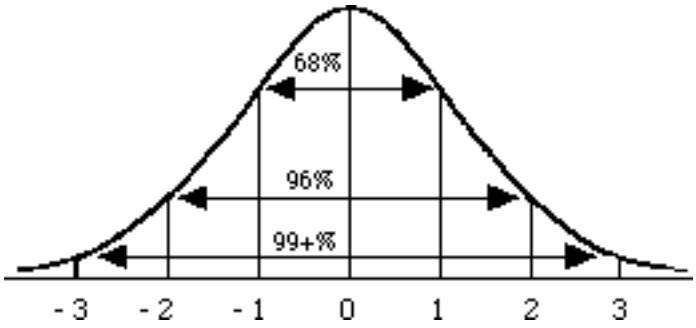


Figura 2 – Distribuição normal

No exemplo anterior das temperaturas $\overline{X}$ = 14.98, s = 0.71, ±1 d.p. corresponde ao intervalo [14.27,15.69], ±2 d.p. ao intervalo [13.56,16.40] e ±3 d.p. ao intervalo [12.85,17.11].

Importa também chamar a atenção de que toda a variável normalmente distribuída pode ser convertida naquilo que se pode chamar pontos (scores) padrão ou pontos z através da fórmula $z = \frac{x - \mu}{s}$

Os pontos padrão são aqueles aos quais a média da população é igual a zero e o desvio padrão é igual à unidade, variando, neste caso, entre -3 e +3.

Nem todas as variáveis são normalmente distribuídas. A variável precipitação (em polegadas) está longe de ser normalmente distribuída, estatísticamente, isto pode causar alguns problemas se pretendermos sumariar as variáveis em questão onde a média pode dar uma ideia errada da forma distribuição.

A menos que se saiba que a variável é normalmente distribuída, compensa normalmente também calcular a mediana e a moda.

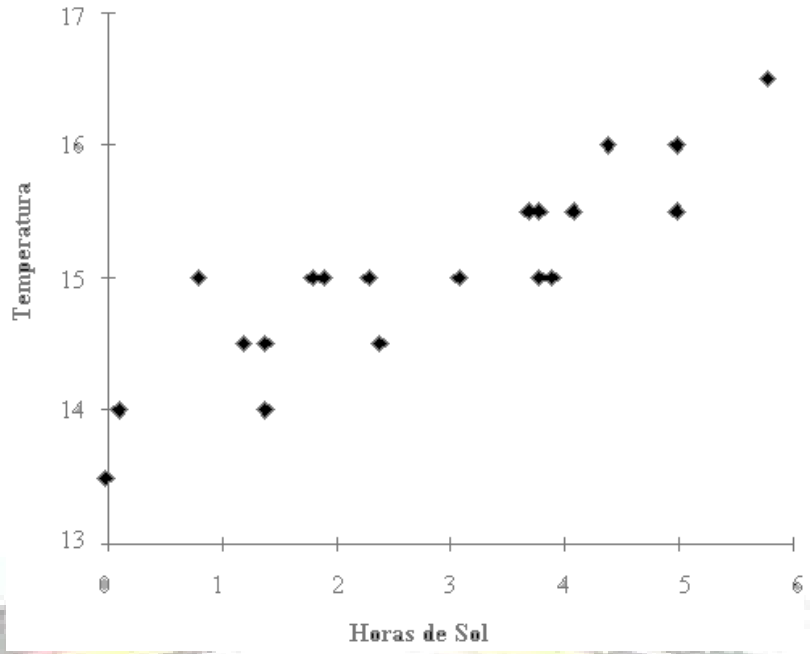


Figura 3 - Relação entre as temperaturas e o número de horas de sol

Um outro conceito extensamente utilizado em estatística é o de correlação, uma medida de aproximação entre duas variáveis. Utilizando os dados anteriores, podemos representar graficamente a relação entre as temperaturas e o número de horas de sol.

Como se pode ver pelo gráfico de pontos da Figura 1.3, há uma tendência no sentido de dizer que quanto mais quente for o dia, maior é o número de horas de sol. É possível desenhar-se uma linha recta ao longo da qual os dados poderão estar agrupados. Os procedimentos matemáticos de regressão encarregam-se de determinar a equação de tal recta.

No nosso caso, e pelas razões já apresentadas, a correlação diz-se positiva. Outras situações onde se assiste a que um aumento de valor de uma variável corresponde a um decréscimo da outra variável é chamado correlação negativa.

Os coeficientes de correlação são expressos numa escala de -1 a 0 até +1. A correlação de -1 significa uma correlação negativa perfeita; uma correlação de zero corresponde à não existência de correlação, e uma correlação de +1 indica uma correlação positiva perfeita.

Um outro processo de visualizar o significado do coeficiente de correlação é utilizarmos os pontos padrão com uma média nula e valores positivos e negativos.

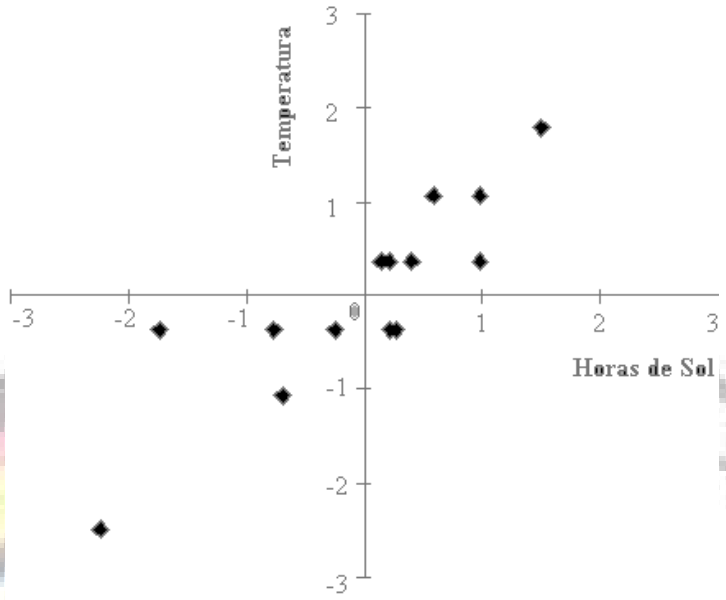


Figura 4 – Gráfico de valores normalizados

Calculando o coeficiente de correlação entre estas duas variáveis obtém-se um valor de r=0.9.O seu quadrado, r2, dá-nos uma medida da variabilidade entre as duas variáveis que pode ser obtida através do coeficiente de correlação.

No nosso caso $(0.9)^2 \approx 0.8$, ou 80%, da variabilidade entre temperatura e horas de sol foi tida em conta pelo coeficiente de correlação, isto é, 20% da variabilidade permanece sem explicação. Posto de uma outra maneira, conhecendo as horas de sol não nos dá uma maneira precisa para prever a temperatura. Para se ter uma melhor previsão da temperatura, temos de olhar para outras variáveis, além das horas de sol. Apesar disso reafirma-se que existe uma forte relação entre temperatura e horas de sol.

Uma outra maneira de visualizar a correlação entre duas variáveis é geometricamente através do ângulo entre duas rectas. As linhas rectas (vectores) representam as variáveis e o ângulo entre elas, a correlação. Estando as variáveis padronizadas, os vectores que as representam têm igual comprimento.

Assim, voltando ao nosso exemplo, e como o valor de correlação de 0.9 corresponde ao valor do cos de 25º a correlação pode ser representada pela figura abaixo.

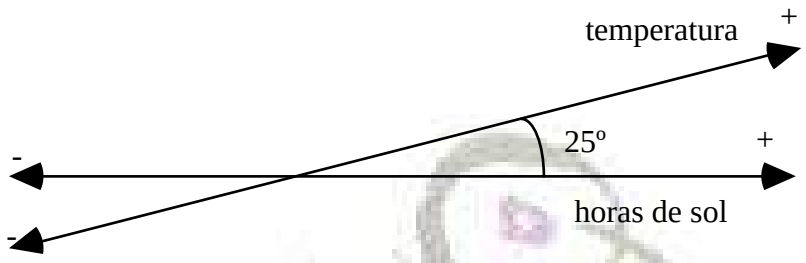


Figura 5 – Correlação entre variáveis

Nesta linguagem, ângulos agudos e obtusos correspondem, respectivamente, a correlações positivas e negativas. No caso de uma perfeita correlação (R=1), o ângulo entre os dois vectores será 0 e os vectores estarão sobrepostos. Duas variáveis completamente não correlacionadas são representadas por dois vectores fazendo um ângulo de 90º, isto é por dois vectores ortogonais.

Finalmente é bom chamar-se a atenção para aquilo que um coeficiente de correlação nos não diz. Só por si, um coeficiente de correlação não nos fornece qualquer informação acerca das razões por que as variáveis estão correlacionadas; não nos fornece qualquer relação causa-efeito.

Para ilustrar isto vejamos o quadro seguinte formado com números provenientes do *Bureau* de Estatística da UE que relaciona dias de greve per capita com o Produto Nacional Bruto.

País	**Dias de greve por 100 trabalhadores (1969-1974)**	**Aumento do PNB *per capita* em libras (1968-1973)**	**Taxa média anual de inflação (%)**
Alemanha Ocidental	240	1193	5.2
França	901	838	7.4
Reino Unido	9035	436	8.9
Itália	5083	389	8.0

Quadro 5 – Correlação entre variáveis

Há uma correlação negativa entre 'dias de greve per capita' e 'aumento do PNB *per capita*'; e uma correlação positiva entre 'dias de greve *per capita*' e 'taxa de inflação anual'.

Uma conclusão rápida deste quadro poder-nos-ia levar a pensar que as greves e, por associação, os sindicatos, foram a maior causa do declínio no Reino Unido. Outro tipo de conclusão bem diferente desta, poderia dizer que as greves constituem mais uma manifestação da insatisfação e descontentamento numa sociedade que não está a lidar convenientemente com as suas ineficiências e desigualdades.

2 Organização e apresentação dos dados

2.1 Classificação e características dos dados

A organização e a apresentação de dados não são independentes da classificação das variáveis em quantitativas ou qualitativas. Enquanto que é possível agrupar os dados qualitativos numa só categoria genérica, os quantitativos são divididos em dados discretos e dados contínuos.

Uma variável quantitativa que tenha valores separados em pontos específicos ao longo da linha dos números é chamada discreta; caso não seja possível encontrarem-se "buracos" entre os possíveis valores de uma variável, esta é dita contínua. Contar os empregados de uma empresa leva-nos a uma variável discreta, ao passo que a taxa de desemprego, obtida através da razão entre o número de desempregados e o tamanho da força de trabalho, constitui uma variável contínua.

Notar que nem sempre é evidente esta distinção entre variáveis contínuas e discretas. Devido a limitações de espaço e de precisão, os valores contínuos são normalmente truncados ou arredondados e portanto assemelham-se a valores de variáveis discretas.

2.2 Organização de dados quantitativos

Existem dois processos elementares de organizarmos os dados quantitativos: são eles a criação de tabelas de distribuição de frequências agrupadas e não agrupadas.

Para construir uma distribuição de frequências não agrupadas apenas temos de listar, numa coluna, todos os valores distintos da variável e, numa segunda coluna, as contagens das frequências de cada valor.

Exemplo: Para se estudar a mobilidade da nossa sociedade, uma empresa de consultoria inquiriu 47 indivíduos e perguntou-lhes quantas vezes, nos últimos três anos, é que tinham mudado de residência. As respostas e a distribuição de frequência não agrupada obtida foram as seguintes:

5	1	0	1	1	1	5	3
0	0	1	1	1	1	3	0
2	4	0	2	4	0	0	3
0	0	3	0	0	6	1	2
4	5	0	2	3	0	4	0
1	4	3	0	7	2	0	

X	f
0	16
1	10
2	5
3	6
4	5
5	3
6	1
7	1
	n = 47

A ideia subjacente à distribuição de frequências agrupadas é juntar valores próximos em classes ou intervalos de números. Para isso é necessário determinar-se o número de classes e o comprimento de cada uma delas.

Para determinarmos o número de classes, e apenas como orientação geral, podemos usar a tabela abaixo onde, por exemplo, se o tamanho da amostra estiver entre 16 e 32, se recomenda 4 a 5 classes.

Tamanho da amostra n	Número de classes c
11 - 16	3 - 4
16 - 32	4 - 5
32 - 64	5 - 6
64 - 128	6 - 7
128 - 256	7 - 8
256 - 512	8 - 9
Acima de 512	10 ou mais

Quadro 6 - Número aconselhável de classes por tamanho da amostra

Escolhendo um valor para c, necessitamos de encontrar o comprimento de cada classe. Este é dado por $w = \frac{\max - \min}{c}$ onde max e min correspondem, respectivamente, aos valores máximo e mínimo dos dados.

Por exemplo, se tivermos uma amostra de tamanho n = 60 com valores de 0,30 a 9,22, a tabela anterior aconselha-nos a agrupar os dados em 6 classes. Assim,

$$w = \frac{\max - \min}{c} = \frac{9{,}22 - 0{,}30}{6} = 1{,}487$$

Como a precisão dos dados é de duas casas decimais, w = 1,49. As classes obtidas poderão então ser as seguintes:

0,30	-	1,79
1,79	-	3,28
3,28	-	4,77
4,77	-	6,26
6,26	-	7,75
7,75	-	9,24

Para cada classe é normalmente aceite que o limite inferior lhe pertence e o limite superior lhe é externo. No nosso caso, a primeira classe seria então formada pelos valores X tal que $0{,}30 \leq X < 1{,}79$.

Entretanto, se a nossa amostra fosse

3,28	0,39	1,14	6,87	7,06	6,48	1,29	1,21	3,67	2,60
0,57	2,17	9,18	1,57	9,22	5,99	2,91	1,34	0,39	4,57
0,56	0,53	4,50	4,87	2,84	1,83	1,91	4,55	3,60	1,78
2,42	1,53	2,06	0,80	2,27	1,25	4,94	0,93	4,57	2,97
0,30	1,30	3,41	1,43	3,83	0,31	6,18	3,30	1,78	2,60
0,45	3,80	4,40	4,19	5,61	2,95	2,42	5,04	2,97	2,37

obter-se-ia a seguinte distribuição de frequências agrupadas:

Classes	f
0,30 - 1,79	21
1,79 - 3,28	15
3,28 - 4,77	13
4,77 - 6,26	6
6,26 - 7,75	3
7,75 - 9,24	2
	n = 60

Três outros conceitos importantes emergem de imediato. O primeiro é o da frequência relativa de uma classe, fr, definida como sendo a frequência das observações dessa classe dividida por n. O segundo é o da frequência acumulada, fa, de uma classe como sendo a soma das frequências nessa classe e em todas as que a precedem. Por fim, o terceiro conceito é o do ponto médio de uma classe, M, que é obtido pela soma dos pontos extremos dividida por dois.

No exemplo da nossa amostra de tamanho 60, podemos ter

Classes	f	f_r	f_a	M
0,30 - 1,79	21	0,350	21	1,045
1,79 - 3,28	15	0,250	36	2,535
3,28 - 4,77	13	0,217	49	4,025
4,77 - 6,26	6	0,100	55	5,515
6,26 - 7,75	3	0,050	58	7,005
7,75 - 9,24	2	0,033	60	8,495
	n = 60	1,000		

2.3 Organização de dados qualitativos

Os dados qualitativos são, por definição, caracterizados por palavras ou categorias, sendo relativamente simples a forma de os organizarmos. São exemplos de dados qualitativos os obtidos pelas seguintes questões, por exemplo:

Qual é o seu sexo? ❑1 FEMININO ❑2 MASCULINO

Qual é o seu estado civil?
❑1 CASADO
❑2 SOLTEIRO
❑3 SEPARADO
❑4 DIVORCIADO
❑5 VIÚVO

Como tem sido a sua qualidade de vida nas últimas 4 semanas? Como é que as coisas lhe têm corrido?

❑1 MUITO BOA: NÃO PODIA SER MELHOR
❑2 BOA
❑3 BOA E MÁ EM PARTES IGUAIS
❑4 MÁ
❑5 MUITO MÁ: NÃO PODIA SER PIOR

A organização dos dados qualitativos é normalmente feita calculando o número de respostas em cada uma das categorias, seguido da percentagem correspondente. Tabelas de uma via, como o quadro 7, são meios concisos e efectivos de organização.

Qualidade de vida	**Número**	**Perc.**
Muito boa	11	17,2
Boa	33	51,6
Em partes iguais	10	15,6
Má	8	12,5
Muito má	2	3,1

Quadro 2.2 - Distribuição das respostas referente à qualidade de vida

3 Medidas numéricas descritivas

Neste capítulo vão ser apresentados os alicerces fundamentais da estatística descritiva. Iremos ver vários processos numéricos destinados a sumariar os dados.

Também aqui é importante distinguir-se uma amostra de uma população. Assim chamaremos parâmetro a uma medida descritiva numérica de uma população; a correspondente medida descritiva da amostra é denominada estatística.

3.1 Medidas de localização central

As primeiras medidas descritivas de interesse são as medidas de localização central que têm como objectivo comum a determinação do centro do conjunto dos dados. Conforme a interpretação que damos à palavra 'centro' assim usamos diferentes medidas. Centro pode ser definido como o ponto médio da amplitude entre o dado mais pequeno e o maior. Podemos também definir centro como o local que divide ao meio toda a massa dos dados. Por fim, numa terceira interpretação, a palavra centro pode representar o centro de gravidade.

Nas linhas que se seguem analisaremos três medidas de localização central. São elas a moda, a mediana e a média. A moda, Mo, de um conjunto de n valores x_1, x_2, x_3, …, x_n é definida como o valor que ocorre mais frequentemente no conjunto de dados.

A figura 6 apresenta graficamente a moda, o valor x com maior frequência.

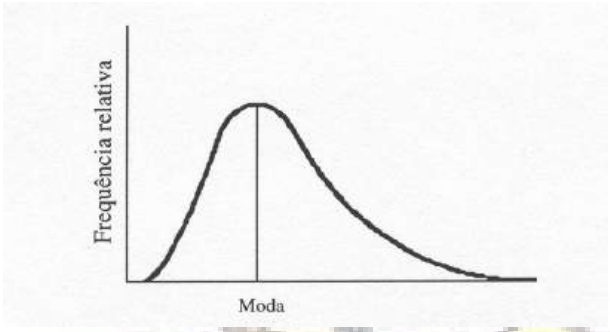


Figura 6 - Localização da moda de uma distribuição de frequência

Exemplificando, se tivermos o seguinte conjunto de dados:

1	4	1	0	2	1	1	3	2	9	1	2

e construirmos a distribuição de frequência, detectamos que o elemento que ocorre com maior frequência é o número 1, que aparece 5 vezes. A moda, neste caso, é dada pelo valor 1.

Em casos onde todos os valores ocorrem com a mesma frequência, a moda não existe. Se dois valores têm a mesma frequência e esta for a maior frequência de todo o conjunto de dados, dizemos que é bi-modal. Se estivermos a trabalhar com valores agrupados em classes, a classe modal é a que obtiver maior frequência, e a moda é o ponto médio dessa classe.

Como vimos, a moda apenas leva em conta o valor com maior frequência e não a posição desse valor no conjunto dos dados. A mediana, Md, por outro lado, separa o conjunto dos dados em duas metades. É definida como o número que divide ao meio um conjunto ordenado de dados.

Para encontrarmos a mediana, dispomos os dados segundo uma ordem crescente ou decrescente e dividimos a meio este conjunto de valores. Se tivermos um número ímpar de dados, a mediana é única e é obtida pelo elemento que se encontra na posição $\frac{n+1}{2}$. Por exemplo, se n = 9, a mediana é o dado que se encontra na quinta posição pois $\frac{n+1}{2} = \frac{9+1}{2} = 5$.

Se o tamanho da amostra for par, temos dois valores medianos e a mediana é então calculada como o ponto médio entre estes dois pontos medianos. Por exemplo, se tivermos a amostra

6,25	6,40	6,40	6,40	6,50	6,25	6,38	6,37	6,40

para determinar a mediana temos de ordenar os valores

6,25	6,25	6,37	6,38	6,40	6,40	6,40	6,40	6,50

Como n = 9, a mediana é dada pelo elemento que ocupa a 5ª posição, isto é, o elemento 6,40.

Se o número 6,50 não pertencesse à nossa amostra, a mediana seria dada pela meia distância entre os dois pontos medianos 6,38 e 6,40, ou seja, seria 6,39.

A mediana apresenta, no entanto, duas grandes desvantagens de manuseamento. A primeira é pecar por uma certa insensibilidade em relação aos valores dos dados. Para ilustrar esta afirmação consideremos uma variável representando, por exemplo, o número de telefones instalados em 12 residências dada pelos seguintes valores:

1	4	1	0	2	1	1	3	2	9	1	2

Se a décima residência, em vez de 9 telefones, tiver 90 telefones, isto não iria alterar o valor da mediana.

A segunda desvantagem reside no facto de termos de ordenar os dados antes de determinarmos a mediana. Isto é ainda mais saliente se soubermos que para amostras grandes (onde o tempo de ordenação se faz mais sentir) o valor da mediana se aproxima do valor de uma outra medida de localização central, a média.

A média é o número total de valores de um conjunto de dados dividido pelo número de valores. Para populações, a média é normalmente representada pela letra grega μ, enquanto que para amostras, usamos o símbolo $\overline{X}$.

O processo usado para determinarmos o valor da média de um conjunto de dados x_1, x_2, x_3, …, x_n depende da maneira como os dados estiverem agrupados. Se não houver qualquer agregação, a média é dada por

$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ onde X_i é um qualquer valor no conjunto dos dados e n é o tamanho da amostra.

Se os dados estiverem agrupados e se fi representar a frequência de ocorrência de x_i, a média é dada pelas fórmulas (para distribuições não agrupadas e agrupadas):

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} \quad \text{e} \quad \overline{X} = \frac{\sum_{i=1}^{n} f_i M_i}{n}$$

onde Mi representa o ponto médio da classe.

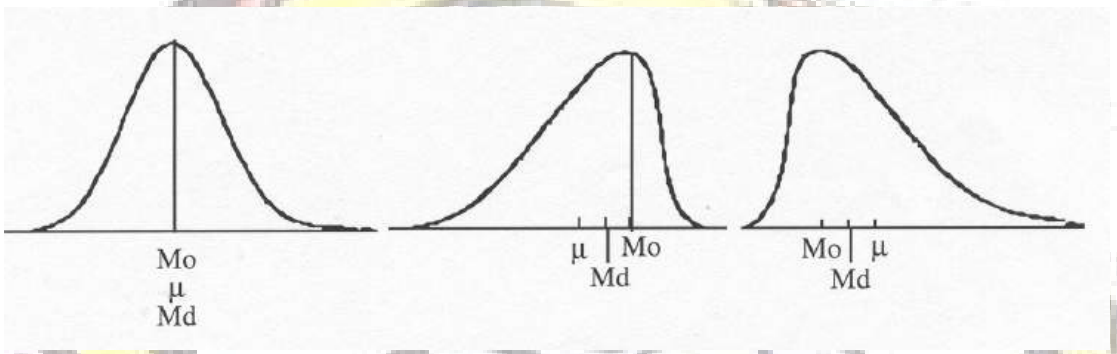


Figura 7 - Relações entre a média (μ), a mediana (Md) e a moda (Mo)

As relações entre a média (μ), a mediana (Md) e a moda (Mo) para uma distribuição com uma só moda pode ser examinada na Figura 3.2. Para uma distribuição simétrica os valores da média, mediana e moda coincidem. Se a distribuição é enviesada para a esquerda, μ < Md < Mo. Se a distribuição é enviesada para a direita, Mo < Md < μ. Em distribuições enviesadas, a mediana está sempre entre a média e a moda.

3.2 Medidas de dispersão

As medidas de localização central não são suficientes para descrever uma distribuição. Veja-se o caso das duas distribuições da Figura 3.3. Ambos os conjuntos de dados partilham dos mesmos valores para as medidas de localização central. Contudo, a distribuição B é mais dispersa do que a distribuição A.

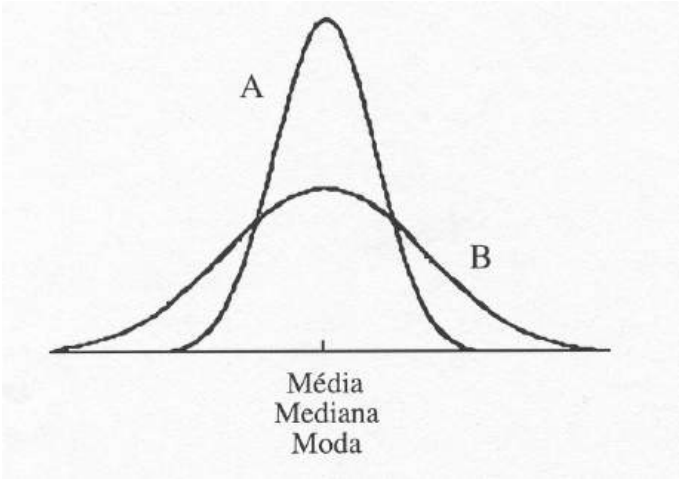


Figura 8 - Distribuições com iguais medidas de localização central e diferente dispersão

A variação é uma característica muito importante, não só em termos práticos, como também para construirmos mentalmente o gráfico da distribuição de frequências. Como existem várias medidas de variabilidade ou de dispersão, iremos ver as mais importantes.

A primeira delas é a amplitude, Amp, envolvendo a subtracção entre os valores máximo e mínimo do conjunto de dados:

$$Amp = Max - Min$$

Infelizmente, a amplitude ainda não é suficiente como medida de variação. As duas distribuições da Figura 3.4, têm ambas a mesma amplitude mas os dados da distribuição da direita apresentam uma variação maior do que os da esquerda.

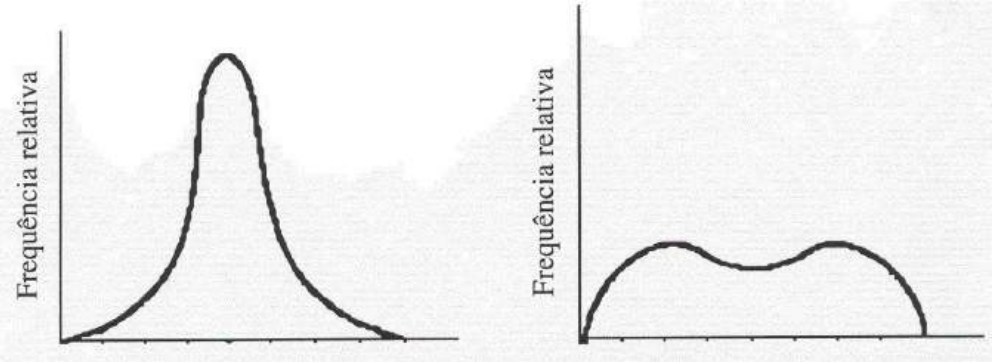


Figura 9 - Distribuições com igual amplitude e diferente variabilidade

Para ultrapassar esta situação, podemos introduzir as noções de quartil e percentil. Definimos os quartis como sendo os valores de x que dividem a área do histograma em quatro partes iguais (figura 10).

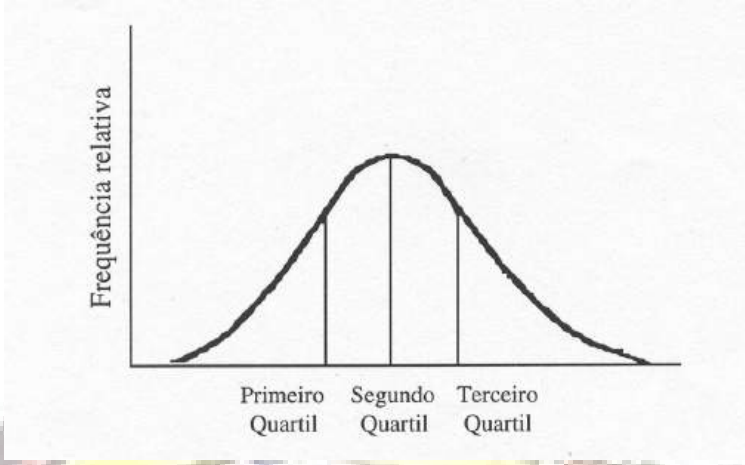


Figura 10 - Localização dos quartis

Principalmente quando manipulamos grandes quantidades de dados, é preferível usarmos os percentis. Se x_1, x_2, x_3, …, x_n representar um conjunto de n medições dispostas segundo a grandeza dos seus valores, o percentil p é o valor de x de tal modo que, quanto muito, 100p% das medições são menores do que o valor de x e, quanto muito, 100(1-p)% são maiores. Por exemplo, o percentil 90 de um conjunto de dados é um valor de x que excede 90% das medições e é menor do que 10%.

Podemos também ver a variabilidade em termos da distância entre as diversas medições e a média, isto é, pelos desvios ($x_1 - \bar{x}$). Se concordarmos com esta ideia, temos de utilizar uma medida de variação baseada nos desvios.

A primeira das alternativas é a utilização do desvio médio das observações, dado por $DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$ onde $|x_1 - \bar{x}|$ representa um desvio absoluto.

Outra das alternativas consiste em elevar ao quadrado os desvio absolutos. Obtemos assim a fórmula para a variância:

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

A variância é a média corrigida dos desvios quadrados. O termo corrigida é incluído pois estamos a trabalhar com amostras e usamos o denominador (n - 1). A variância da população é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{N} = \frac{\sum x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{N}$$

Calculemos a variância do seguinte conjunto de dados

4	9	5	4	4	10

Processo 1: Para determinar o valor de S2 começamos por obter o valor da média $\bar{x} = 6$ e calcular os desvios.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
4	-2	4
9	3	9
5	-1	1
4	-2	4
4	-2	4
10	4	16
		$\sum (x_i - \bar{x})^2 = 38$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = \frac{38}{6-1} = 7{,}6$$

Processo 2: Usando a fórmula de cálculo,

x_i	x_i^2
4	16
9	81
5	25
4	16
4	16
10	100

$\sum_{i=1}^{n} x_i = 36$ $\sum_{i=1}^{n} x_i^2 = 254$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1} = \frac{254 - \frac{36^2}{6}}{6-1} = 7{,}6$$

Para distribuições de frequências agrupadas, a técnica é semelhante.

Ao elevarmos ao quadrado os desvios resolvemos o problema da soma nula entre os desvios positivos e negativos. No entanto estamos a trabalhar com unidades diferentes das iniciais. Assim, se aplicarmos a função raiz quadrada à variância, regressamos às unidades originais. O desvio padrão é então a raiz quadrada positiva da variância.

O desvio padrão e a média são dois dos parâmetros ou estatísticas mais usadas para caracterizar (descrever) uma população ou uma amostra. Se tivermos dois conjuntos de dados com a mesma média e diferentes desvios padrão, o conjunto de dados com maior desvio padrão é mais disperso do que o outro com menor desvio padrão. Quando as médias são distintas, no entanto, uma mera comparação entre os desvios padrão não faz sentido. Nestes casos podemo-nos recorrer do chamado coeficiente de variação, CV, definido pela razão entre o desvio padrão e a média.

Por exemplo, consideremos a figura 11 com as distribuições de frequência de dois conjuntos de dados: A com média 7,6 e desvio padrão 3,2; B com média 6,8 e desvio padrão 2,5.

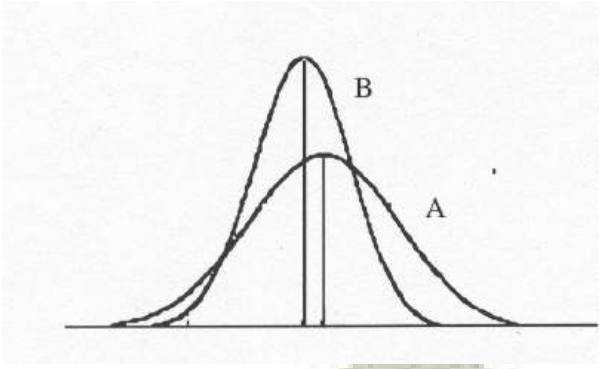


Figura 11 - Duas distribuições com médias e desvio padrão diferentes

Suponhamos também que estas distribuições correspondem, por exemplo, à quantidade do investimento que é recuperada se investirmos no projecto A ou no projecto B.

O projecto A parece melhor do que B pois a média de capital recuperado é maior. No entanto é também o que contém mais risco, sendo potencialmente mais volátil pois o desvio padrão é maior. Qual deverá ser a nossa decisão?

Calculando os coeficientes de dispersão, obteremos

$$CVA = \frac{s}{\mu} \times 100 = \frac{3,2}{7,6} \times 100 = 42,11\%$$

$$CVB = \frac{s}{\mu} \times 100 = \frac{2,5}{6,8} \times 100 = 36,76\%$$

Em termos de dispersão relativa, o projecto B é menos variável do que o projecto A. Se não quisermos correr riscos, o projecto B é a escolha apropriada; caso contrário, devemos escolher o projecto A.

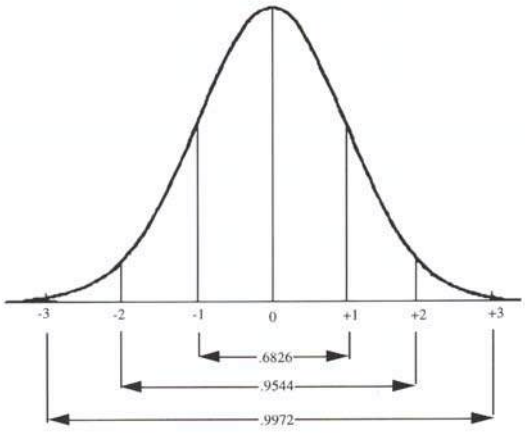


Figura 12 - Percentagem de áreas entre desvios padrão

Estas duas medidas descritivas – a média e o desvio padrão – podem também ser combinadas através de uma regra empírica (figura 12). Segundo esta regra empírica, cerca de dois terços do dados (0,6826) está à volta da média à distância de um desvio padrão. Dois desvio padrão acima e abaixo da média contêm cerca de 95% dos dados. Três desvio padrão conterão, segundo esta regra, quase a totalidade dos dados.

Notar que esta regra empírica apenas se aplica a conjuntos de dados com distribuição aproximadamente em forma de sino e simétrica.

4 Distribuição normal univariada

4.1 A curva normal

Como se pode ver na figura abaixo, a distribuição Normal é uma distribuição simétrica em relação à média μ, tendo esta valor idêntico à mediana e à moda. Os valores da variável aleatória X vão desde -∞ até +∞ e a área debaixo da curva é igual a 1.

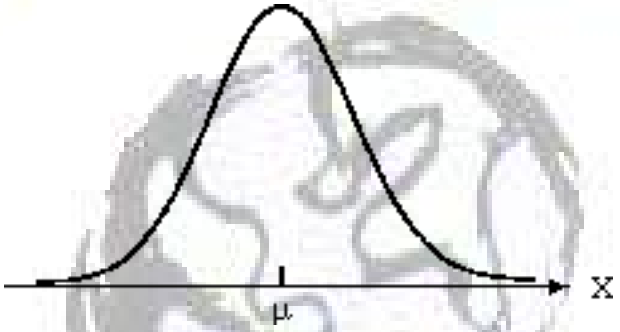


Figura 13 — Distribuição de probabilidade Normal

Pela fórmula seguinte da distribuição Normal, pode facilmente deduzir-se que os dois parâmetros desta distribuição são a média μ e o desvio padrão σ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < X < +\infty$$

Também aqui nos socorremos de tabelas para obter os valores para a variável normalmente distribuída. No entanto, e porque se torna impossível possuirmos uma tabela para todos os possíveis valores de μ e de σ, procedemos a uma transformação de variável e reduzimos a nossa distribuição em estudo a uma distribuição N(0,1), com média 0 e desvio padrão 1. Tal transformação é a seguinte:

$$Z = \frac{\overline{X} - \mu}{\sigma}$$

Por exemplo, o ponto X = 84 numa distribuição N(72,8) corresponde ao ponto Z = 1.5 numa distribuição N(0,1). De facto,

$$Z = \frac{\overline{X} - \mu}{\sigma} = \frac{84 - 72}{8} = 1{,}5$$

De notar que esta variável estandardizada Z corresponde à distância, medida em desvios padrão, desde a média até à localização do ponto.

Ainda para ilustrar a utilização da tabela da lei Normal centrada e reduzida, N(0,1), consideremos os discos rígidos de um determinado modelo e marca produzidos por uma empresa de material informático. O tempo de vida destes discos é uma variável aleatória normalmente distribuída com μ = 760 e σ = 140 horas.

Se quiséssemos determinar a probabilidade de que um disco deste tipo falhe antes das 1000 horas de funcionamento procederíamos da seguinte maneira:

$$P(X < 1000) = P\left(Z < \frac{1000 - 760}{140}\right) = P(Z < 1{,}71)$$

Na tabela da distribuição Normal, obteremos o resultado 0,9564. A figura seguinte pretende ilustrar a correspondência conseguida pela transformação linear.

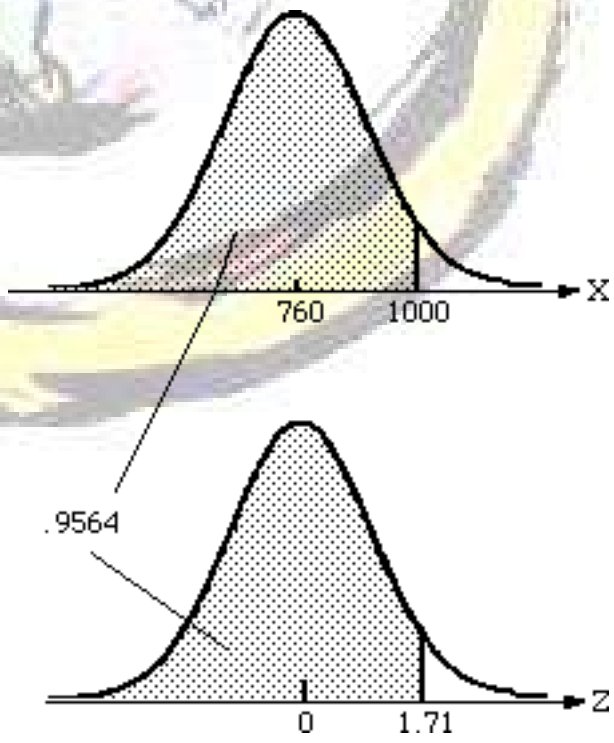


Figura 14 — Encontrar P(X < 1000), X ~ N(760,140)

Procederíamos do mesmo modo para determinar a probabilidade de um disco rígido durar mais do que 600 horas (ver figura 15):

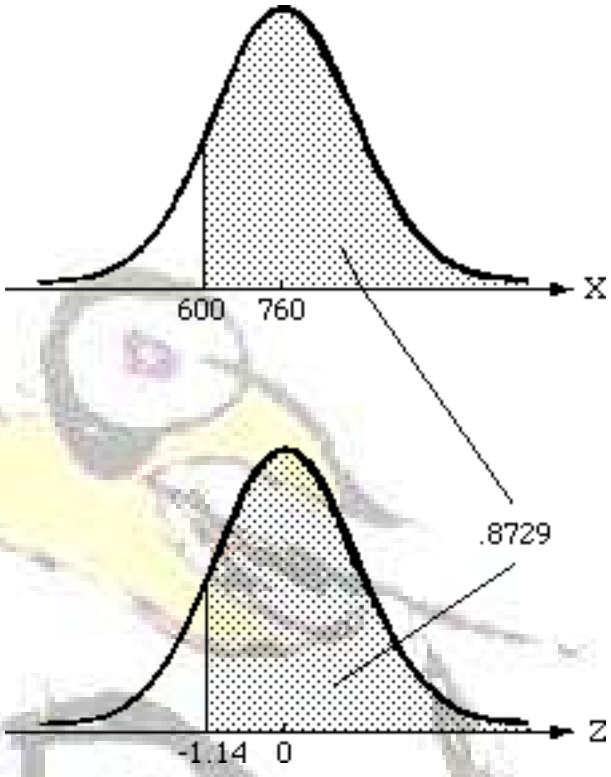


Figura 15 — Encontrar P(X > 600), X ~ N(760,140)

$$P(X > 600) = P\left(Z > \frac{600 - 760}{140}\right) = P(Z > -1{,}14) = P(Z < 1{,}14) = 0{,}8729)$$

Para determinarmos a probabilidade de que um determinado disco tenha uma vida entre 700 e 800 horas, equacionamos o nosso pedido da seguinte maneira:

$$P(700 < X < 800) = P\left(\frac{700 - 760}{140} < Z > \frac{800 - 760}{140}\right) = P(-0{,}43 < Z < 0{,}29) = 0{,}2805$$

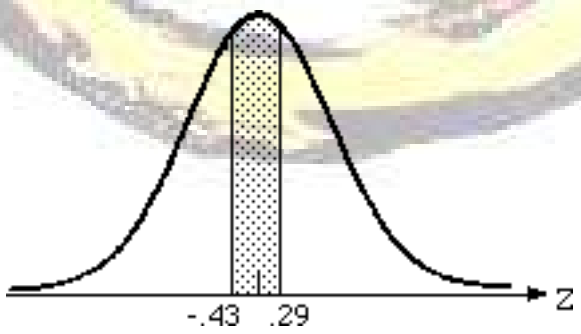


Figura 16 — Encontrar a probabilidade entre dois pontos

4.2 Distribuição da média amostral

Se tivermos uma população Normal N(μ,σ) de tamanho N e pretendermos extrair uma amostra de tamanho n, temos várias hipóteses possíveis, mais concretamente, $\frac{N!}{n!(N-n)!}$. Para cada uma destas possíveis amostras de tamanho n podemos calcular uma média. Temos assim vários valores para a variável que representa a média de uma amostra de tamanho n extraída de uma população de tamanho N.

Esta nova variável $\overline{X}$ é também uma distribuição Normal mas com média $\mu_{\overline{X}} = \mu$ e desvio padrão $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. Normalmente, mesmo que a população de origem não seja Normal, podemos utilizar a variável $\overline{X}$ como sendo normalmente distribuída.

Este resultado é-nos dado pelo Teorema do Limite Central:

Para quase todas as populações, sempre que o tamanho da amostra seja grande, a distribuição amostral de $\overline{X}$ é aproximadamente Normal com média μ e desvio padrão $\frac{\sigma}{\sqrt{n}}$, qualquer que seja a forma da distribuição.

Este teorema é importante pois, desde que a amostra seja grande, não necessitamos conhecer a distribuição de onde foi extraída. A noção de amostra grande é, no entanto, uma noção relativa. Se a distribuição da população é simétrica, uma amostra de n = 10 ou 15 pode ser de tamanho suficiente. Contudo, se a distribuição da população é moderadamente enviesada, 15 pode não ser um número suficientemente grande, sendo vulgarmente usado n ≥ 30 como condição para a utilização do teorema do limite central. Se a população é fortemente enviesada, é necessário uma amostra de tamanho muito maior.

Sempre que a população seja finita e conheçamos o valor de N, se n ≥ 0,05N podemos tornar mais compacta a distribuição de $\overline{X}$ usando uma correcção para populações finitas. Esta correcção consiste em multiplicar o valor do desvio padrão $\frac{\sigma}{\sqrt{n}}$ por $\sqrt{\frac{N-n}{N-1}}$ ficando a fórmula do desvio padrão com o seguinte aspecto:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

5 Organização e descrição de dados bivariados

Nos capítulos anteriores vimos como organizar os dados em tabelas e em gráficos, assim como obter medidas numéricas para melhor descrever o conjunto de dados em estudo. Neste capítulo iremos analisar o comportamento simultâneo de duas variáveis e a relação existente entre elas.

5.1 Dados quantitativos bivariados

Na análise bivariada dos dados há normalmente uma variável independente X e uma variável dependente Y cujos valores são previstos ou explicados com base nos valores da variável independente.

Vejamos por exemplo os dados bivariados da tabela abaixo referente à cilindrada de um motor e ao número de quilómetros percorridos com um litro de combustível.

Modelo de Automóvel	Tamanho do motor (em cm^3)	Km/litro estimados
A	4.311	9,8
B	3.966	10,2
C	2.426	13,2
D	2.131	12,8
E	2.950	11,5
F	5.212	7,7
G	2.622	11,5
H	3.311	10,6
I	4.999	7,7
J	3.540	11,1

Neste exemplo é fácil distinguir-se a variável X da variável Y. Isto porque parece óbvia uma relação causa-e-efeito entre as duas variáveis. No

entanto, nem sempre é assim. Por vezes não existe relação, embora continuem a existir as variáveis as variáveis X e Y.

Vejamos como exemplo os dados agrupados na tabela seguinte

País	Consumo individual diário de proteínas	Taxa de natalidade
Formosa	4,7	45,6
Malásia	7,5	39,7
Índia	8,7	33,0
Japão	9,7	27,0
Jugoslávia	11,2	25,9
Grécia	15,2	13,5
Itália	15,2	23,4
Bulgária	16,8	22,2
Alemanha	37,3	20,0
Irlanda	46,7	19,1
Dinamarca	56,1	18,3
Austrália	59,9	18,0
Estados Unidos	61,4	17,9
Suécia	62,6	15,0

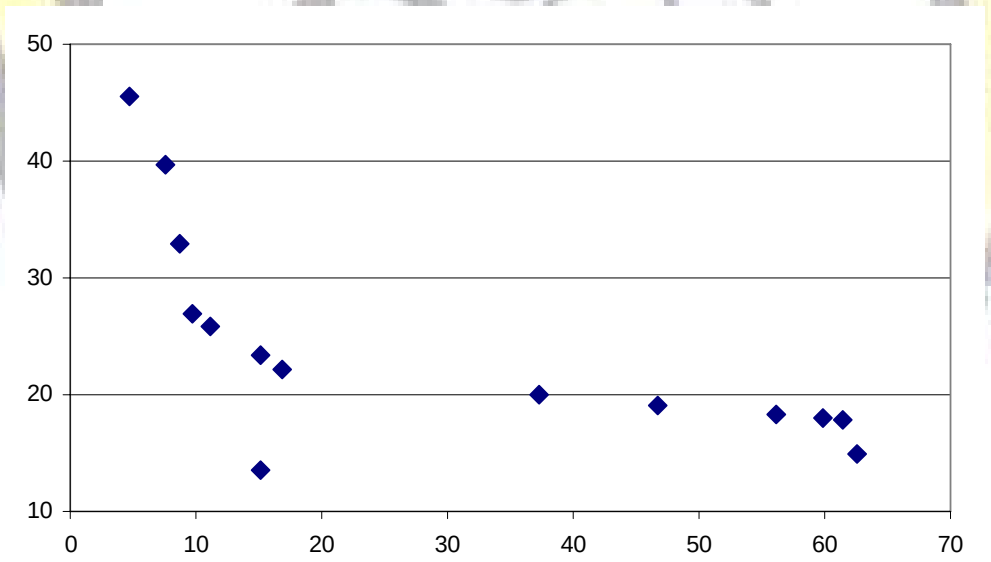


Figura 17 - Gráfico de pontos

Como se mostra na figura 17 existe uma correlação entre as variáveis "consumo diário de proteínas" e "taxa de natalidade". Este exemplo, além de ilustrar a correlação negativa, mostra bastante claramente que correlação não significa necessariamente uma relação causal, isto é, um fenómeno de causa-e-efeito. Destes dados não se pode tirar a conclusão de que um aumento de proteínas determine a redução da fertilidade.

Neste caso, a correlação negativa entre as duas variáveis poderia ser explicada pela qualidade de vida. É razoável admitir-se que uma melhoria da vida num país, determine tanto um aumento no consumo médio de proteínas, como uma diminuição do taxa de natalidade.

É importante frisar que se duas variáveis estão positivamente correlacionadas, isto apenas significa que variam no mesmo sentido. Não podemos afirmar que aumentos sucessivos numa das variáveis determinam aumentos sucessivos na outra variável. Podem sempre existir outras variáveis causando a variação das variáveis em estudo.

O processo mais vulgar de apresentar visualmente os dados quantitativos bivariados é através de gráficos de pontos, gráficos a duas dimensões com a variável independente no eixo horizontal e a variável dependente no eixo vertical. A Figura 4.2 representa um gráfico de pontos referente ao exemplo anterior.

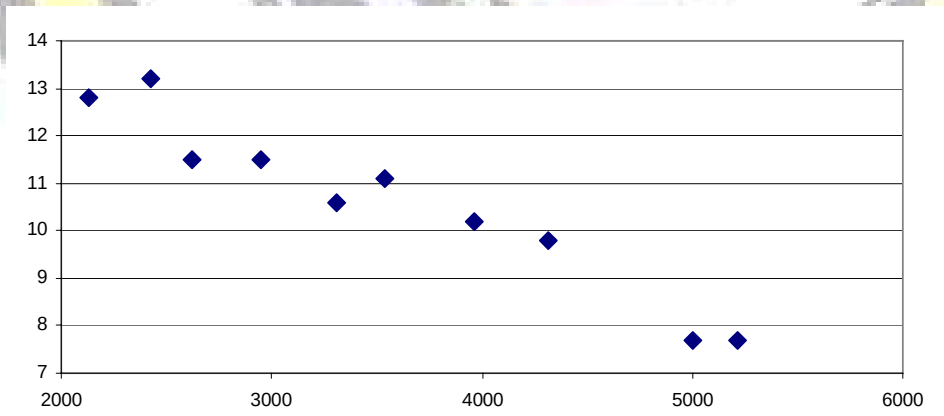


Figura 18 - Gráfico de pontos

Nesta figura é evidente que aos motores com maior cilindrada correspondem aos quilómetros percorridos por cada litro. Isto é representativo por uma relação negativa ou inversa entre estas duas variáveis.

Esta relação entre cilindrada de motor e quilómetros andados por litro pode também ser capturada pela linha recta da figura 19.

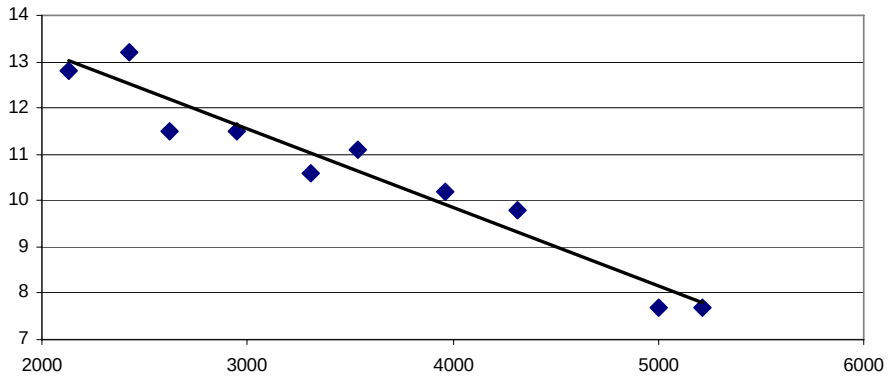


Figura 19 - Gráfico de pontos com recta ajustada

A figura 20 apresenta dois casos de relação linear directa e de não relação entre duas variáveis. Neste último caso, os dados não apresentam qualquer tendência ou padrão.

Figura 20 - Exemplos de relação

Finalmente, os dados podem estar relacionados mas não linearmente.

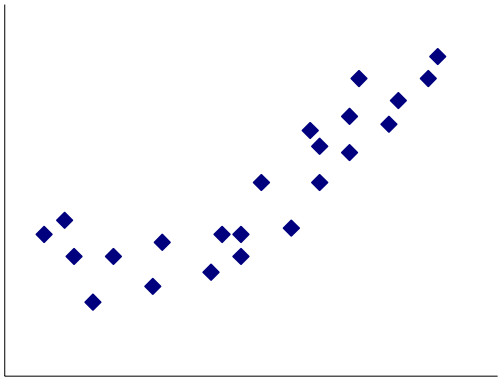


Figura 21 - Relação não linear

Quando se trata de uma relação linear, há por vezes interesse em medirmos a intensidade dessa relação. Nestas situações usamos uma medida denominada coeficiente de correlação, normalmente representado pela letra r. O coeficiente de correlação é um número sem unidades que descreve o grau de relação linear entre X e Y e cuja amplitude varia de -1 a +1. Uma relação negativa corresponde a uma relação indirecta e uma correlação positiva a uma relação directa.

O valor do coeficiente de correlação r é dado pela fórmula

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - \left(\sum X\right)^2}\sqrt{n\sum Y^2 - \left(\sum Y\right)^2}}$$

Para ilustrar os cálculos de determinação do coeficiente de correlação, consideremos o seguinte exemplo de pares (X,Y).

X	17	29	18	19	21	21	14	24	26	28
Y	46	60	42	43	50	47	39	58	53	58

O gráfico de pontos da figura seguinte dá-nos uma ideia de uma relação directa entre duas variáveis

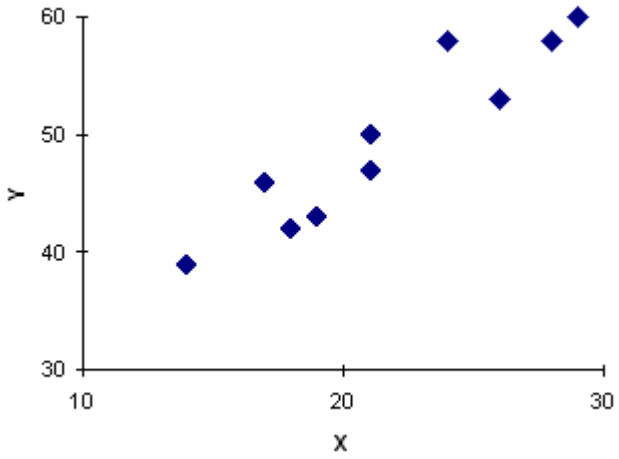


Figura 22 - Gráfico de pontos

O coeficiente de correlação pode ser calculado do seguinte modo

X	Y	X^2	XY	Y^2
17	46	289	782	2.116
29	60	841	1.740	3.600
18	42	324	756	1.764
19	43	361	817	1.849
21	50	441	1.050	2.500
21	47	441	987	2.209
14	39	196	546	1.521
24	58	576	1.392	3.364
26	53	676	1.378	2.809
28	58	784	1.624	3.364
$\Sigma X = 217$	$\Sigma Y = 496$	$\Sigma X^2 = 4.929$	$\Sigma XY = 11.072$	$\Sigma Y^2 = 25.096$

$$r = \frac{10(11.072) - 217(496)}{\sqrt{10(4.929) - 217^2}\sqrt{10(25.096) - 496^2}} = \frac{3.088}{\sqrt{2.201}\sqrt{4.944}} = 0{,}936$$

o que significa uma forte correlação positiva.

Para determinarmos a significância de um coeficiente de correlação, isto é, para decidirmos da existência de uma relação entre X e Y, podemos usar a Tabela A3 de pontos de corte para r. Segundo esta tabela, se tivermos uma amostra com n = 50 elementos e obtivermos um coeficiente de correlação r = 0,085, há forte informação para que não exista uma relação entre X e Y, pois |0,085| < 0,279.

Por vezes estamos interessados numa equação que melhor descreva a relação estatística entre X e Y, ou seja, em determinar a equação da recta de regressão.

Como se trata de uma recta a sua equação será da forma

$$\hat{Y} = b_0 + b_1 X$$

onde b_1 representa a inclinação, isto é, a alteração na variável Y associada à alteração de uma unidade de X;

b_0 representa a ordenada na origem, ou seja, o valor de Y correspondente a X = 0.

O "chapéu" colocado na letra Y significa que estamos a obter um valor estimado ou previsível para Y. À diferença entre os valores reais e os estimados, $(Y - \hat{Y})$, é chamado erro de aproximação, ou simplesmente, erro. A figura 23 representa exemplos de erros positivos e negativos de aproximação.

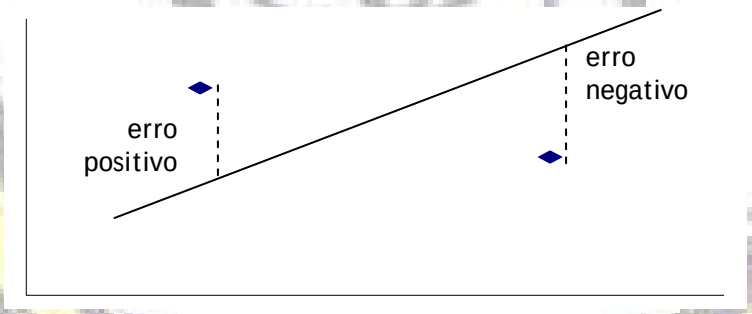


Figura 23 - Erros de aproximação

O processo de determinação dos valores para b_0 e b_1, ou seja, da equação da recta de regressão, é baseado na minimização dos erros da aproximação. É denominado critério dos mínimos quadrados.

Assim, uma relação estatística linear que satisfaça o critério dos mínimos quadrados é chamada a recta dos mínimos quadrado e é da forma

$$\hat{Y} = b_0 + b_1 X$$

onde $b_1 = \dfrac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2}$ e $b_0 = \overline{Y} - b_1 \overline{X}$

e onde $\overline{X} = \dfrac{\sum X}{n}$ é a média de X e $\overline{Y} = \dfrac{\sum Y}{n}$ é a média de Y.

Com base nos dados do exemplo anterior (da determinação do coeficiente de correlação) podemos obter os valores para determinar a equação da regressão linear.

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{10(11.072) - 217(496)}{10(4.929) - 217^2} = 1{,}4030$$

$$b_0 = \overline{Y} - b_1 \overline{X} = 49{,}6 - 1{,}4030(21{,}7) = 19{,}1549$$

Arredondando para duas casa decimais, a recta dos mínimos quadrados para os 10 pontos é

$$\hat{Y} = 19{,}15 + 1{,}40X$$

Esta última equação de regressão pode também ser usada com o objectivo de previsão. Suponhamos que tínhamos um valor X = 25. Neste caso, poderíamos prever ou estimar o valor de Y do seguinte modo:

$$\hat{Y} = 19{,}15 + 1{,}40(25) = 19{,}15 + 35{,}0 = 54{,}15$$

5.2 Dados qualitativos bivariados

Para apresentar os resultados referentes a dados qualitativos é comum usarem-se tabelas de frequência. Para dados bivariados, usamos as chamadas tabelas de contingência, ou tabelas de dupla entrada, onde uma variável é apresentada em linhas e outra em colunas.

Vejamos o seguinte caso. Num total de n = 309, defeitos em mobiliário foram registados e classificados em quatro tipo, de A a D. Ao mesmo tempo, cada peça de mobiliário foi identificada de acordo com o turno da produção em que foi manufacturada. A tabela de contingência obtida é

	Tipo de Defeito				
Turno	**A**	**B**	**C**	**D**	**Total**
1	15	21	45	13	94
2	26	31	34	5	96
3	33	17	49	20	119
Total	74	69	128	38	309

Esta tabela é chamada tabela de contingência 3 x 4 pois contem 3 linhas, 4 colunas e os números dentro da tabela representam frequências ou contagens. As frequências dos lados exteriores da tabela são chamadas frequências marginais e representam os totais univariados para cada uma das variáveis.

Também esta tabela de contingência pode ser apresentada em termos de frequências relativas. A tabela seguinte é um exemplo do que foi dito.

	Tipo de Defeito				
Turno	**A**	**B**	**C**	**D**	**Total**
1	0,05	0,07	0,15	0,04	0,30
2	0,08	0,10	0,11	0,02	0,31
3	0,11	0,06	0,16	0,06	0,39
Total	0,24	0,22	0,41	0,12	1,00

6 Inferência estatística

Neste capítulo iremos ver alguns tópicos da inferência estatística cujo objectivo é generalizar, para toda a população, os resultados obtidos da amostra. Assim, a média μ, a variância σ^2 e a proporção π são exemplos de parâmetros normalmente desconhecidos de populações e cujos valores pretendemos inferir através das correspondentes estatísticas $\overline{X}$, s^2 e π.

Existem dois processos para inferir estatisticamente. O primeiro é a técnica de estimação segundo a qual pretendemos encontrar um valor ou um intervalo para o parâmetro desconhecido. O outro é o teste de hipóteses no qual, com base em duas afirmações opostas, decidimos acerca dos possíveis valores do parâmetro.

No entanto, para que seja correcto inferir-se estatisticamente, é necessário que se defina com antecedência qual o objectivo da análise, se tenha acesso às população e que se proceda a uma amostragem aleatória. Só assim podemos utilizar a inferência estatística para apoiar as nossas tomadas de decisão.

6.1 Estimação

Conforme abordámos nos parágrafos anteriores, podemos considerar dois tipos de estimação:

(1) obter um valor - estimação pontual - que constitua a melhor aproximação para o parâmetro; ou

(2) obter um conjunto de valores - estimação intervalar - no qual seja provável que o parâmetro da população se encontre. O intervalo obtido é por vezes também chamado intervalo de confiança e é dependente da amostra que estamos a estudar mais directamente.

6.2 Teste de hipótese

O outro método para se inferir estatisticamente consiste em realizar um teste de hipóteses. Como o próprio nome indica, com este método testamos uma hipótese por nós formulada para explicar certas observações ou uma situação. Podemos, por exemplo, afirmar que a média de uma população é $\mu=7$, que a probabilidade de um determinado acontecimento ocorrer é $\pi=0,4$ ou que a variância da população em estudo é $\sigma^2=0,15$.

Para testarmos uma hipótese usamos o método da prova indirecta ou da redução ao absurdo, pressupondo como verdadeira a hipótese contrária àquela que queremos testar. Se chegarmos a uma contradição, podemos concluir que o nosso pressuposto estava errado e, por conseguinte, a hipótese inicial não é de excluir.

Existem, portanto, duas hipóteses de trabalho:

- H_0 hipótese nula, o ponto de partida da nossa investigação;
- H_1 hipótese alternativa, normalmente a negação da hipótese nula.

H_0 constitui uma afirmação acerca do valor de um parâmetro que aceitamos como verdadeira. No entanto, pelo facto de estarmos a trabalhar com uma amostra e não com toda a população, é evidente que qualquer que seja a conclusão do teste, este não é isento de erro. As nossas decisões são sempre baseadas na chamada evidência amostral, naquilo que podemos inferir a partir da amostra que recolhemos. Se existir suficiente evidência amostral para contradizer a amostra nula, então acreditamos que a hipótese alternativa é a afirmação mais razoável. Se a evidência amostral não for suficiente, continuamos a acreditar que a hipótese nula é a mais correcta para representar a situação em estudo.

Consideremos os seguintes exemplos de hipóteses nulas e alternativas:

(1)	$H_0: \mu = 10$	(2)	$H_0: \sigma^2 < 8$	(3)	$H_0: \pi = 0,60$
	$H_1: \mu \neq 10$		$H_1: \sigma^2 \geq 8$		$H_1: \pi > 0,60$

Repararemos que no exemplo (2) a hipótese nula é apresentada na forma de um intervalo e que no exemplo (3) a hipótese alternativa não constitui uma negação perfeita de H_0. Todas as formulações das hipóteses

alternativas, incluindo obviamente o exemplo (1), correspondem a uma interpretação do que se entende por afirmação oposta.

Tudo depende do enquadramento do nosso problema. Pode acontecer que o enquadramento do problema seja tal que o nosso único interesse seja testar se o valor do parâmetro é estritamente maior do que um valor especificado.

Como veremos adiante, qualquer teste de hipóteses inclui cinco elementos:

- 1. uma hipótese nula H_0;
- 2. uma hipótese alternativa H_1;
- 3. uma estatística do teste;
- 4. uma região de rejeição;
- 5. uma conclusão.

As hipóteses nula e alternativa representam as afirmações referentes ao parâmetro da população, a estatística do teste permite sumariar os dados amostrais e apresentar a evidência estatística, a região de rejeição é a zona que permite que a conclusão seja elaborada e, por fim, tomada a decisão.

Se a estatística do teste estiver na região de rejeição, concluímos que a amostra é inconsistente com a hipótese nula e rejeitaremos esta hipótese. Caso o teste estatístico não pertença à região de rejeição, há evidência de que a amostra é consistente com a hipótese nula e, portanto, não rejeitaremos esta hipótese.

A conclusão é baseada na evidência estatística obtida a partir da amostra. O que não significa que esteja isenta de erro. De facto, comparando as duas conclusões possíveis (rejeitar ou não H_0) com a situação real da hipótese nula (verdadeira ou falsa) temos quatro situações possíveis:

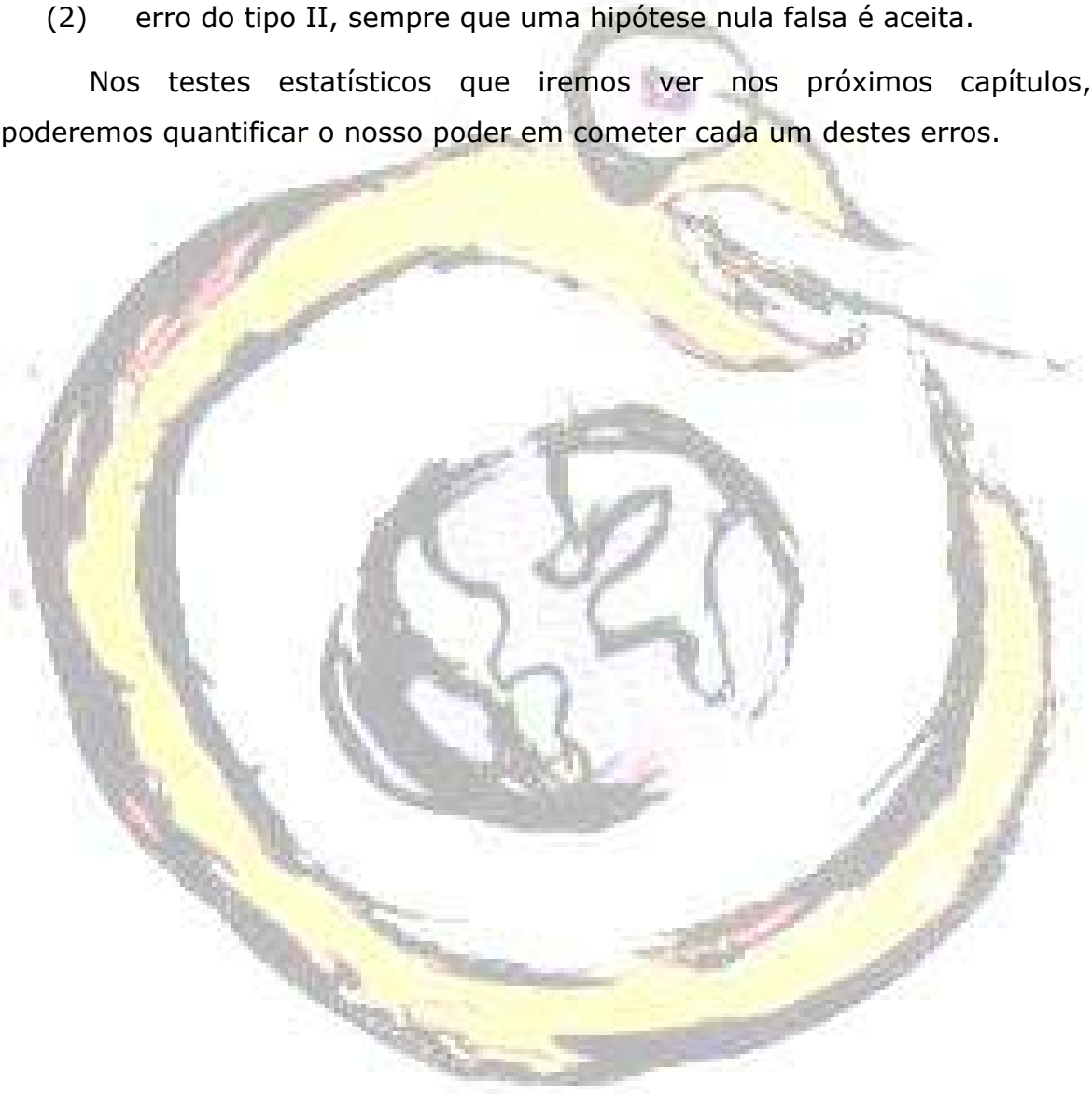
	Situação real	
Conclusão	H_0 verdadeira	H_0 falsa
Rejeitar H_0	ERRO DO TIPO I	CONCLUSÃO CORRECTA
Não rejeitar H_0	CONCLUSÃO CORRECTA	ERRO DO TIPO II

A conclusão é correcta se o teste estatístico nos levar a rejeitar H0 e este ser efectivamente de rejeitar ou se o teste estatístico nos levar a não rejeitar H_0 e este ser de aceitar. No entanto, nas outras duas situações temos uma conclusão errada, havendo dois tipos de erro:

(1) erro do tipo I, sempre que uma hipótese nula verdadeira é rejeitada;

(2) erro do tipo II, sempre que uma hipótese nula falsa é aceita.

Nos testes estatísticos que iremos ver nos próximos capítulos, poderemos quantificar o nosso poder em cometer cada um destes erros.



7 Inferência sobre μ

Este capítulo debruça-se sobre inferências acerca da média μ de uma população quando a amostra é grande (n≥30) e quando a amostra é pequena (n<30).

7.1 Amostras grandes

7.1.1 Estimações pontuais e intervalares

Exemplo: Suponhamos que um hospital adquire regularmente algodão em pacotes de 50 Kg. Há, no entanto, uma suspeita de que a mercadoria não esteja a ser entregue com o peso especificado. Assim, ao investigar-se a situação, numa amostra de 32 pacotes encontrou-se uma média de $\overline{X}$ = 49,5 Kg.

A média da amostra constitui uma estimação pontual para a média μ da população. No entanto, esta estimação pontual não nos fornece informação alguma acerca do eventual erro amostral. A solução reside num estimador intervalar, uma amplitude de valores centrada em $\overline{X}$ dentro do qual deverá estar o valor desconhecido da média μ.

As bases para a obtenção do estimador intervalar, ou intervalo de confiança, são o Teorema do Limite Central e a nossa capacidade em obter probabilidades através das tabelas de Z. Um intervalo de confiança a 98 % para Z centrado em 0 é apresentado na figura 24.

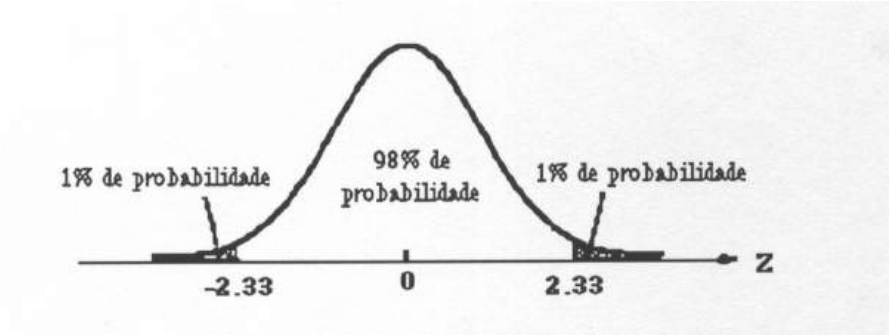


Figura 24 - Pontos Z para um intervalo de confiança a 98%

Como se pode ver nesta figura, o intervalo de confiança a 98% começa no primeiro percentil e acaba no percentil 99. Analisando a tabela de Z, os valores da variável correspondentes a estes dois percentis são Z =± 2,33. Por outras palavras,

$$P(-2{,}33 \leq Z \leq 2{,}33) = 0{,}98$$

Substituindo Z em função de $\overline{X}$ obtemos

$$P\left(-2{,}33 \leq \frac{X-\mu}{s/\sqrt{n}} \leq 2{,}33\right) = 0{,}98$$

$$P\left(\overline{X} - 2{,}33\,\sigma/\sqrt{n} \leq \mu \leq \overline{X} + 2{,}33\,\sigma/\sqrt{n}\right) = 0{,}98$$

Se recolhermos uma amostra aleatória, existe uma probabilidade de 98% para que o valor desconhecido da média μ pertença ao intervalo $\overline{X} \pm 2{,}33\,\sigma/\sqrt{n}$. Dito de outro modo, há um risco de 0,02 de que o intervalo $\overline{X} \pm 2{,}33\,\sigma/\sqrt{n}$ não contenha o valor de μ.

Em termos gerais, o intervalo de confiança para μ com σ conhecido e uma amostra grande é dado por

$$\overline{X} \pm z\frac{\sigma}{\sqrt{n}}$$

O valor 0,98 é normalmente denominado nível de confiança sendo os valores 0,99, 0,95 e 0,90 os mais prováveis.

Caso o valor de σ seja também desconhecido, a estimativa para μ é conseguida através do valor do desvio padrão amostral s. O intervalo de confiança para μ com σ desconhecido e uma amostra grande é dado por

$$\overline{X} \pm z \frac{s}{\sqrt{n}}$$

Vejamos um exemplo. Suponhamos que numa determinada prova de acesso a um lugar de técnico superior estagiário podem existir notas desde 200 a 800 pontos, levando as notas já obtidas neste exame a aceitar um desvio padrão σ igual a 100. Pretende-se encontrar, com base numa amostra de 75 elementos onde foi encontrada uma média $\overline{X}$ = 534, um intervalo de confiança a 90% para a média.

Como ao nível de confiança de 90% correspondem os valores de Z=±1,65, obtemos

$$534 \pm 1{,}96 \frac{100}{\sqrt{75}} \qquad \text{ou} \qquad 534 \pm 19$$

Outra maneira de apresentarmos este resultado é dizer que, com um risco de 10%, a média da população deverá pertencer ao intervalo de 515 a 553.

Outro exemplo. Um director de um hospital desenhou um estudo para determinar a média de gastos em papel feito pelos vários serviços. Ao analisar 147 requisições calculou-se uma média $\overline{X}$ = 318 e um desvio padrão s = 249. Criar um intervalo de confiança com um risco de apenas 5% de não incluir μ.

Apesar de se não conhecer o desvio padrão da população, a amostra é suficientemente grande para que se possa usar o seu desvio padrão na determinação do intervalo de confiança a 95%.

$$\overline{X} \pm z \frac{s}{\sqrt{n}} = 318 \pm 1{,}96 \frac{249}{\sqrt{147}} = 318 \pm 40{,}3$$

7.1.2 Determinação do tamanho da amostra

O método para se encontrar um intervalo de confiança pode também ser usado para se determinar o tamanho da amostra. Para isso é necessário conhecer-se:

(1) o erro amostral E máximo que estamos dispostos a tolerar;

(2) a dispersão na população de interesse, medida pelo desvio padrão σ;

(3) o nível de confiança para o intervalo, representado pelo valor da tabela Z.

A expressão que nos permite determinar à partida, o tamanho da amostra para atingir os nossos objectivos é a seguinte:

$$n = \left(\frac{Zs}{E}\right)^2$$

No exemplo anterior do teste de admissão, uma amostra de 75 elementos produziu o intervalo de confiança 534 ± 19. Suponhamos agora que se pretende um erro de amostragem não superior a 15. Qual deverá ser o tamanho da amostra para se encontrar o mesmo intervalo de confiança a 90%?

$$n = \left(\frac{Zs}{E}\right)^2 = \left(\frac{1{,}65(100)}{15}\right)^2 = 121$$

São então necessárias mais 46 observações para que o intervalo centrado em $\overline{X}$ tenha uma amplitude ±15.

7.1.3 Testes de hipóteses

O outro processo para se usar informação amostral na inferência para a população é testando uma hipótese. Num teste para a média, a hipótese é formulada da seguinte maneira:

$$H_0: \quad \mu = \mu_0$$

$$H_1: \quad \mu \neq \mu_0$$

onde μ_0 pode ser um qualquer valor. Se H_0 for verdadeira, então inferiremos que a evidência amostral provem de uma distribuição amostral normal para $\overline{X}$ centrada em μ_0 e que a média amostral está na vizinhança de μ_0.

Na figura 25 podemos ver que a região de aceitação para H_0 inclui grande parte da distribuição amostral e que constitui um intervalo onde é provável encontrar-se o valor de $\overline{X}$, caso H_0 seja verdadeira.

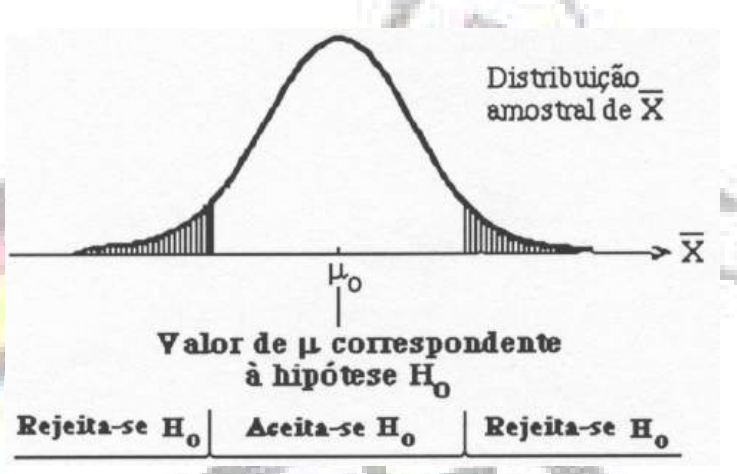


Figura 2.2 - Regiões de rejeição e de aceitação para teste bilateral

Para determinarmos se a hipótese H_0 é ou não de rejeitar, usamos a informação amostral (média, desvio padrão e tamanho da amostra) para calcular a chamada estatística do teste, que não é mais do que um valor Z* que combina o valor de μ da hipótese H_0 com a evidência amostral. Comparando este valor calculado Z* com o valor de corte Z, é possível tomar-se a decisão.

Quando testamos a hipótese referente à média da população usamos a seguinte estatística de teste:

$$z^* = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Se σ é desconhecido, podemos substituir σ pelo desvio padrão s da amostra, especialmente se n é superior a 30.

Para ilustrar este tipo de teste, vejamos o exemplo de um serviço cirúrgico que foi há cerca de ano e meio sujeito a uma auditoria. Uma das

conclusões desta auditoria é que o tempo médio de espera dos doentes era de 51.5 dias. O Director do serviço pretende agora, com base numa amostra de 75 facturas e com um nível de significância de 0,10, verificar se este número foi alterado.

O resultado da amostra foi $\overline{X}$ = 57,3 dias com s = 21,0 dias, sendo as hipóteses a testar

H_0: μ = 51,5 dias

H_1: μ ≠ 51,5 dias

Para um risco de 10% de cometer um erro do tipo I, o ponto de corte correspondente é ±1.65 e a estatística do teste é

$$z^* = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{57{,}3 - 51{,}5}{21/\sqrt{75}} = 2{,}39$$

Como Z*=2.39>1.65, rejeita-se a hipótese nula e, portanto poderemos dizer que o tempo médio de espera foi, de facto, alterado.

Neste exemplo procedemos a um teste bilateral com a região de rejeição a corresponder aos valores de $\overline{X}$ mais afastados de μ_0 para ambas as direcções. Por vezes na prática pretendemos apenas incluir na região de rejeição apenas uma das direcções. Isto constitui um chamado teste unilateral. Vejamos outro exemplo.

Há quatro meses, no intuito de diminuir o tempo de espera para uma intervenção, o Conselho de Administração de um hospital decidiu implementar um programa de incentivos. Pretendemos então testar H_0: μ=57 dias contra H_1: μ<57 dias a um nível de significância de 0.05.

Uma amostra de 40 casos de doentes à espera apresentou uma média $\overline{X}$ = 53,5 com s = 26 dias. Será que podemos concluir que o programa de incentivos fez diminuir o tempo de espera dos doentes?

Para α=0,05, o ponte de corte inferior para Z é -1.65 e a nossa estatística de teste é então

$$z^* = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{53{,}5 - 57}{26/\sqrt{40}} = 0{,}85$$

Como Z*=-0,85>-1,65, não rejeitamos H_0, ou seja a evidência amostral é consistente com a veracidade da hipótese nula.

7.2 Amostras pequenas

7.2.1 Estimações intervalares

Para amostras pequenas não é possível aplicar o Teorema do Limite Central e, portanto os procedimentos inferenciais vistos até agora neste capítulo não são teoricamente válidos. Por isso usaremos a distribuição t para representar o comportamento da variável $\overline{X}$.

A distribuição t é usada para realizar inferências sobre μ nas seguintes condições:

(1) O desvio padrão da população σ é desconhecido mas estimado à custa do desvio padrão amostral s;

(2) O tamanho da amostra é menor do que 30;

(3) É possível partir do pressuposto de que a população subjacente de X é normalmente distribuída.

A figura 26 mostra a forma genérica da distribuição t. Esta distribuição, apesar de ter uma forma semelhante à da normal, possui um desvio padrão superior a 1. É de facto uma família de curvas, cada uma delas associada a um diferente grau de liberdade.

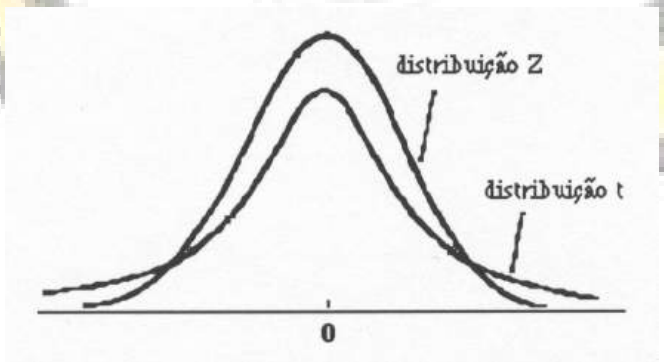


Figura 26 - Forma genérica da distribuição t relativamente à distribuição Z

O termo grau de liberdade diz respeito ao número de valores que são livres de variar, havendo algumas restrições nos dados. Por exemplo, ao calcularmos a variância através da fórmula

$$s^2 = \frac{n}{n-1}\sigma^2 = \frac{\sum\left(X - \overline{X}\right)^2}{n-1}$$

todos os valores, excepto um, podem variar. Esta excepção deverá satisfazer a restrição $\sum\left(X - \overline{X}\right) = 0$. Ao longo deste capítulo, os graus de liberdade associados a uma distribuição t serão sempre iguais a n-1, onde n representa o tamanho da amostra.

Se n é menor do que 30, devemos usar a distribuição t e a correspondente tabela. O intervalo de confiança é dado pela fórmula

$$\overline{X} \pm t\frac{s}{\sqrt{n}}$$

Um director de um centro de saúde pretende saber quantos utentes estão a utilizar as horas suplementares de abertura da dependência (das 9 às 11 horas da noite) todos os dias úteis. Ao longo de seis semanas contou os doentes nestes dias e horas e obteve os valores 74, 73, 82, 86, 81 e 78. Partindo do pressuposto de que X (número de doentes entre as 9 e as 11 horas da noite) é uma variável normalmente distribuída, o director pretende encontrar um intervalo de confiança a 90% para o número médio de doentes a usar as horas suplementares.

Desta amostra de n = 6, encontramos $\overline{X}$ = 79 e s = 4.98. Para n - 1 = 5 graus de liberdade e para α = .10, o valor de t é ± 2.015. Assim,

$$\overline{X} \pm t\frac{s}{\sqrt{n}} = 79 \pm 2{,}015\frac{4{,}98}{\sqrt{6}} = 79 \pm 4{,}10$$

O director do centro de saúde pode então estar 90% certo de que o número de doentes das 9 às 11 horas da noite está compreendido entre 75 e 83 clientes.

7.2.2 Testes de hipóteses

Do mesmo modo, a estatística do teste de média para amostras pequenas segue uma distribuição t e é dada pela fórmula

$$t^* = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Este valor é comparado com o valor de corte da tabela t com n-1 graus de liberdade e um nível especificado de confiança.

Vejamos o caso de uma companhia de produtos farmacêuticos que pretende construir mais uma empresa. De exemplos anteriores em muito mais pequenas escalas sabe-se que é possível aproveitar completamente 93.5% da matéria-prima. Após estar em funcionamento, foram analisados 12 lotes com os seguintes valores:

91,41	94,47	92,80	92,93	94,19	93,37
92,92	93,16	93,24	92,70	93,66	93,86

Usando um nível de significância de 0,05, testar a hipótese do aproveitamento médio da matéria-prima ser de 93.5%, isto é H_0: μ=93,5 contra H_1: μ≠93.5. O valor de corte para este nível de significância e n-1=11 graus de liberdade é t = ±2,201.

Para esta amostra de n=12 elementos, a média e o desvio padrão obtidos são respectivamente $\overline{X}$ = 93.23% e s = 80%. Então a estatística do teste será

$$t^* = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{93{,}23 - 93{,}5}{0{,}80/\sqrt{12}} = -1{,}17$$

Como |t*|≤2,201, a nossa conclusão é de que a evidência amostral seguiu H_0.

8 Inferência sobre π e σ

Neste capítulo iremos alargar a análise feita no capítulo anterior às amostragens binomiais onde analisamos a proporção π e o desvio padrão σ da população.

8.1 Inferências sobre a proporção π da população

Numa situação amostral de características binomiais, há que considerar a percentagem ou proporção p de êxito dada pela fórmula

$$p = \frac{X}{n} = \frac{Número\ de\ êxitos\ na\ amostra}{Número\ de\ items\ da\ amostra}$$

Tal como π, p é um número entre zero e um, inclusive, e segue uma distribuição aproximadamente normal com média igual a π e desvio padrão igual a $\sigma_p = \sqrt{\pi(1-\pi)/n}$. Deste modo, a equação para estandardizar resultados amostrais binomiais é

$$z = \frac{p - \pi}{s_p} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Suponhamos que pretendemos realizar uma sondagem à opinião pública e a um total de 1100 adultos é-lhes perguntado se são a favor ou contra uma certa legislação. Se soubermos que 40% dos adultos apoiam a legislação em questão, qual a probabilidade de que a amostra dê valores dentro de dois pontos percentuais à volta dos 40%?

Se π=0,40, o erro padrão é $\sqrt{0,4(0,6)/1100} = 0,01477$ e, portanto,

$$P(0,38 \leq p \leq 0,42) = P\left(\frac{0,38 - 0,40}{0,01477} \leq Z \leq \frac{0,42 - 0,40}{0,01477}\right)$$

$$= P(-1.35 \leq Z \leq 1.35) = 0,8230$$

Um resultado amostral baseado em 1.100 observações tem mais do que quatro em cinco hipóteses de estar a dois pontos percentuais de 0,40.

Analogamente ao já apresentado no capítulo anterior, um intervalo de confiança para a proporção de uma população é dado por

$$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

Exemplo: A pedido dos seus clientes, uma agência de publicidade levou a cabo um estudo para saber se as pessoas, quando vêm programas previamente gravados na televisão, avançam a fita quando surgem os anúncios. Numa amostra de 698 possuidores de gravadores de vídeo, foi detectado que 38% normalmente ou sempre passavam à frente as partes de publicidade.

Partindo do pressuposto de que esta amostra representa a população dos possuidores de gravadores de vídeo, pretende-se determinar um intervalo de confiança a 90%.

Este intervalo será $0,38 \pm 1,65\sqrt{\frac{0,38(0,62)}{698}} = 0,38 \pm 0,0303$ ou seja, de 35% a 41%.

Também para a estimação da proporção π há uma expressão que nos fornece o tamanho de amostra necessário

$$n = \frac{z^2\pi(1-\pi)}{E^2}$$

Como se pode ver esta expressão é idêntica à da média com a única diferença de que $\pi(1-\pi)$ substitui o valor de σ^2.

Exemplo: Um grupo de investigação de uma empresa produtora de material médico está interessado em estimar a proporção π dos serviços com material obsoleto. Pretende-se que esta estimação tenha um erro máximo ±0,03 com

90% de confiança, nada se sabendo acerca do valor de π. Qual deve ser o tamanho da amostra para se obter esta precisão desejada?

Como não temos um valor para π, estimamos π=0,50. Para 90% de confiança, z=1,65. Então

$$n = \frac{z^2\pi(1-\pi)}{E^2} = \frac{1{,}65^2(0{,}5)(0{,}5)}{0{,}03^2} = 757$$

O uso de π=0,5 constitui uma abordagem dita conservadora uma vez que qualquer outro par de valores (0,7 e 0,3 ou 0,8 e 0,2) levar-nos-ia a amostras mais reduzidas.

Também a lógica e os métodos da estatística dos testes referentes a amostragens binomiais são idênticos aos testes de Z e de t. A estatística do teste para a proporção binomial π é dada pela fórmula

$$z^* = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Ao estudar o processo de aprovação de crédito decidido há 18 meses, uma empresa com várias dependências em várias cidades detectou que o crédito era concedido a 65 por cento dos que o solicitavam. Pretende-se agora saber se esta percentagem ainda se mantém. Para isso recolheu uma amostra de 315 pedidos de crédito nos últimos 90 dias e verificou 101 recusas e 214 aprovações, ou seja, p = X/n = 214/315 = 0,6794.

Se π representar a proporção de todos os pedidos de crédito, temos H_0: π=0,65 contra H_1: $\pi \neq 0{,}65$. Para α=0,10, o ponto de corte para Z é ±1,65 e, portanto, a estatística do teste é

$$z^* = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}6794 - 0{,}65}{\sqrt{\frac{0{,}65(0{,}35)}{315}}} = 1{,}09$$

Como $-1{,}65 \leq Z^* \leq 1{,}65$, não podemos rejeitar H_0: podemos descrever a diferença entre p e π_0 (0,0294) como não significativa

8.2 Inferências sobre σ e σ^2

Até agora vimos métodos de inferência estatística aplicados à média μ e à proporção π. Nesta secção debruçar-nos-emos sobre intervalos de confiança e testes de hipóteses para a variância σ^2 ou para o desvio padrão σ.

Por razões de simplicidade matemática, apenas iremos inferir estatisticamente acerca da variância σ^2. A função raiz quadrada que liga a variância ao desvio padrão, permite-nos facilmente converter os resultados obtidos para o desvio padrão.

Todos estes procedimentos de inferência acerca de σ^2 e de σ partem do pressuposto de que a amostra em estudo provém de uma população com uma distribuição normal ou aproximadamente normal.

Começamos por reconhecer que existe um valor de população σ^2 que, apesar de constante, é desconhecido. Além disso, o valor s^2 da variância da amostra muito dificilmente será exactamente igual a σ^2, sendo portanto possível falar-se de uma distribuição para s^2. Esta distribuição depende não só de σ^2 como também do tamanho n da amostra. Assim, estandardizando, obtemos uma variável chamada qui-quadrado (χ^2) dada por

$$\chi^2 = \frac{(n-1)\sigma^2}{s^2}$$

Usando a tabela do qui-quadrado apresentada em Apêndice, podemos encontrar os pontos de corte correspondentes aos níveis aceitáveis de risco e a vários graus de liberdade. A expressão geral para o intervalo de confiança que envolva uma variável qui-quadrado é dada por

$$\chi_E^2 \leq \chi^2 \leq \chi_D^2$$

onde χ_E^2 e χ_D^2 correspondem, respectivamente, aos limites inferior (esquerdo) e superior (direito) do intervalo.

Assim, substituindo χ^2 pela expressão anterior temos

$$\chi_E^2 \leq \frac{(n-1)\sigma^2}{s^2} \leq \chi_D^2$$

ou seja,

$$\frac{(n-1)s^2}{\chi^2_D} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_E}$$

Exemplo: O grupo de controlo de qualidade de uma fábrica de motorizadas decidiu testar uma amostra de 20 motorizadas saídas da produção. Pretende-se estudar a precisão e a variabilidade dos odómetros para se poder avaliar a qualidade dos fornecimentos feitos por uma outra empresa.

As motorizadas percorreram uma distância de 200 quilómetros e forneceram os seguintes dados: $\overline{X}$ = 200,21 Km, s^2 = 0,23 e s = 0,48. Construir um intervalo de confiança a 95% para σ, o desvio padrão da população, partindo do pressuposto que as leituras dos odómetros seguem uma distribuição aproximadamente normal.

Para n = 20, os nossos graus de liberdade são gdl=n-1=19. Então, com um erro à direita e à esquerda de 0.025, obtêm-se os seguintes valores: χ^2_D=32,852 e χ^2_E=8,907. Assim, $\frac{19(0,23)}{32,852} \leq \sigma^2 \leq \frac{19(0,23)}{8,907}$ ou $0,13 \leq \sigma^2 \leq 0,49$.

O intervalo de confiança para σ é obtido calculando as raízes quadradas aos limites do intervalo para σ^2, produzindo 0,36 Km ≤ σ ≤ 0,70 Km.

Reparar que os limites inferior e superior não estão equidistantes da estimação pontual, devendo-se isto ao enviesamento na distribuição do qui-quadrado. A estimação pontual está sempre mais próxima do limite inferior do que do limite superior.

Um outro processo de inferir acerca de σ ou σ^2 é usando um teste de hipóteses para σ^2. A estatística do teste para a variância da população σ^2 é

$$\chi^{2*} = \frac{(n-1)s^2}{s_0^2}$$

onde o índice zero corresponde aos valores no caso da hipótese nula não ser de rejeitar.

Como exemplo, vejamos o caso de uma companhia farmacêutica que produz cápsulas de 40 mg, sendo esta produção, em princípio uniforme. No entanto há necessidade em se testar, sabendo-se que o desvio padrão da dosagem não deve ultrapassar 1,25 mg.

Uma amostra de 25 cápsulas apresentou $\overline{X}$ = 40,33 e s = 1,34. Será que a dosagem está a variar demasiado?

A nossa hipótese unilateral é:

H_0: $\sigma = 1{,}25$ $(\sigma^2 = 1{,}5625)$

H_1: $\sigma > 1{,}25$ $(\sigma^2 > 1{,}5625)$

Como n=25 e α=0,01, obtemos o valor σ^2=42,980, sendo então a estatística do teste

$$\chi^{2*} = \frac{(n-1)s^2}{s_0^2} = \frac{24(1{,}34)^2}{1{,}25^2} = 27{,}58$$

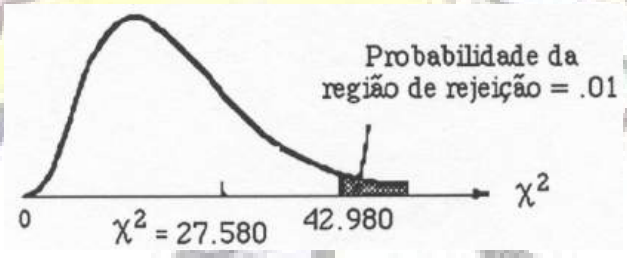


Figura 3.1 - Estatística do teste e região de rejeição

Como a estatística do teste é menor do que o ponto de corte dado pela tabela (ver Figura 3.1), podemos afirmar que a evidência amostral é consistente com a hipótese nula e, portanto, dizer que o processo está sob controlo.

9 Comparação entre duas populações

Neste capítulo iremos analisar o caso de duas populações diferentes com o intuito de as compararmos. Necessitamos de duas amostras para podermos inferir se as populações de onde elas provêm são a mesma ou diferem em relação a determinadas características.

A lógica que subjaz à criação dos intervalos de confiança e dos testes de hipóteses é idêntica às anteriormente utilizadas. Assim, o intervalo de confiança é dado por

$$\text{Estimação pontual} \pm (Z \text{ ou } t)\ (\text{erro padrão})$$

e o teste de hipótese é baseado na estatística

$$Z^* \text{ (ou } t^*) = \frac{\text{resultado da amostra} - \text{valor sob a hipótese nula}}{\text{erro padrão}}$$

9.1 Inferência sobre μ_1-μ_2: amostras independentes

Estamos a analisar a situação em que as amostras são independentes, ou seja, foram extraídas de duas populações. Os itens seleccionados de uma população não exercem qualquer influência sobre os itens seleccionados da segunda população.

No caso de amostras grandes, o estimador pontual para μ_1-μ_2 é dado pela diferença entre as médias amostrais $\overline{X}_1 - \overline{X}_2$. Aplicando o teorema do limite central, podemos dizer que a distribuição amostral de $\overline{X}_1 - \overline{X}_2$ tem uma forma aproximadamente normal com

$$\text{média} = \mu_1\text{-}\mu_2 \quad \text{e} \quad \text{desvio padrão} = \sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Caso as variáveis σ_1 e σ_2 das populações sejam desconhecidas, são utilizados os correspondentes valores s_1 e s_2 das amostras, obtendo um valor estimado para o desvio padrão

$$\hat{s}_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

O intervalo de confiança para a diferença entre as médias é obtido do seguinte modo:

$$\overline{X}_1 - \overline{X}_2 \pm z\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Exemplo: O responsável por uma pequena empresa pretende estimar a diferença entre médias das despesas feitas pelos clientes possuidores de um dos dois cartões de crédito *American Express* e *Visa*. Recolhida uma amostra aleatória dos talões dos últimos 3 meses obtiveram-se os seguintes resultados:

Cartão de Crédito	Amostra	$\overline{X}$	s
American Express	59	19 666	6 806
Visa	66	16 376	6 022

Pretende-se construir um intervalo de confiança a 90% para a diferença média das populações das quantias gastas. Para isso, e chamando amostra 1 à correspondente ao uso do cartão *American Express*, podemos calcular o seguinte intervalo:

$$\overline{X}_1 - \overline{X}_2 \pm z\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 19.666 - 16.376 \pm 1{,}65\sqrt{\frac{6.806^2}{59}+\frac{6.022^2}{66}} = 3.290 \pm 1.906{,}15$$

Podemos, portanto, estar 90% certos de que μ_1-μ_2 estará entre 1383,85 e 5196,15.

Um teste bilateral é normalmente usado com o objectivo de investigar a seguinte hipótese nula H_0: $\mu_1=\mu_2$ ou, o que é o mesmo, H_0: $\mu_1-\mu_2=0$.

Por outras palavras, esta hipótese nula implica que não exista diferença entre as médias das populações. Para se concluir que tal diferença existe, há necessidade de rejeitarmos a hipótese nula.

Sendo a hipótese nula verdadeira, as duas populações de onde foram extraídas as amostras têm a mesma média. Deste modo a estatística do teste é

$$z^* = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Exemplo: Num teste de mercado de uma nova marca de cereais para o pequeno almoço estamos a avaliar duas campanhas publicitárias. Uma das campanhas afirma que o produto é nutritivo para crianças, enquanto que, para a segunda campanha, o comer cereais pela manhã constitui um procedimento considerado natural em adultos.

A campanha orientada para as crianças é feita em 32 estabelecimentos de cinco cidades e a campanha orientada ara os adultos é realizada num conjunto diferente de 32 estabelecimentos em seis cidades do país.

Após um período de seis meses, obtiveram-se as seguintes vendas:

Campanha	Estabelecimentos	Médias de vendas	Desvio padrão
1. Adultos	32	13.9	2.3
2. Crianças	32	12.1	2.0

Pretende-se usar estes dados para decidir qual das campanhas parece mais promissora. Escolhe-se um erro de tipo I de 5%.

$$z^* = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(13{,}9 - 12{,}1) - 0}{\sqrt{\frac{2{,}3^2}{32} + \frac{2{,}0^2}{32}}} = \frac{1{,}8}{0{,}539} = 3{,}34$$

Como este valor é superior o ponto de corte para 5% de erro (±1.96), temos evidência suficiente para considerar que a hipótese nula não é verdadeira. A campanha orientada para adultos parece ser a mais apelativa.

Se estivermos a trabalhar com amostras pequenas temos de partir do pressuposto de que as duas populações partilham a mesma variância. Assim, com base nas variâncias amostras s_1^2 e s_2^2 podemos obter uma estimação para a variância comum das duas populações:

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Para ilustrarmos os cálculos para s_p^2, suponhamos que obtivemos amostras de cada uma das duas populações e delas os seguintes valores:

$$n_1 = 8 \qquad s_1 = 20{,}8 \qquad \overline{X}_1 = 88{,}3$$

$$n_2 = 12 \qquad s_2 = 24{,}2 \qquad \overline{X}_2 = 85{,}1$$

Partindo do pressuposto de que as populações têm a mesma média, podemos obter um valor de s_p^2 como estimador de σ^2.

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{20{,}8^2(8-1) + 24{,}2^2(12-1)}{8 + 12 - 2} = 526{,}14$$

O erro padrão é dado pela fórmula

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

e o intervalo de confiança para μ_1-μ_2 em amostras pequenas é

$$\overline{X}_1 - \overline{X}_2 \pm t\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Do mesmo modo a estatística do teste sobre μ_1-μ_2 em amostras pequenas é dada por

$$t^* = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

9.2 Inferência sobre μ_1-μ_2: amostras dependentes

As amostras dependentes, também chamadas emparelhadas, representam uma aplicação especial da inferência para duas amostras. Os estudos que envolvam este tipo de amostras são normalmente mais informativos e eficientes e as observações amostrais reduzem-se às diferenças di entre elementos correspondentes de ambas as amostras. Estas diferenças

constituem uma variável d com média $\bar{d}$ e desvio padrão s_d. Assim, a fórmulas para o intervalo de confiança para μ_d é

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

e a estatística do teste H_0: $\mu_d = 0$ é

$$t^* = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

Exemplo: Um novo medicamento para doentes com o nível de colesterol elevado foi ministrado a seis doentes num período de quatro meses. Se considerarmos este grupo de doentes como uma amostra aleatória de doentes com uma contagem de colesterol acima de 250 mg por 0.1 litros de sangue, pretende-se estimar com 95% de confiança a redução média de colesterol encontrada.

Doente	Antes Tratamento	Após Tratamento
1	252	211
2	311	251
3	280	241
4	293	248
5	312	258
6	327	268

Ao calcular as diferenças Antes - Após Tratamento encontramos os valores para a variável d

Doente	1	2	3	4	5	6
Antes - Após	41	60	39	45	54	59

e desta tabela, os valores $\bar{d} = 50$ e $s_d = 9{,}42$. Assim o nosso intervalo de confiança para a redução média no colesterol é

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}} = 50 \pm 2{,}571 \frac{9{,}42}{\sqrt{6}} = 50 \pm 10$$

Temos 95% de confiança de que a redução média da contagem de colesterol na população está compreendida entre 40 e 60.

9.3 Inferência sobre π_1 - π_2

Os estudos comparativos podem também envolver duas populações binomiais e, neste caso, estarmos interessados na inferência sobre $\pi_1 - \pi_2$. Aqui, partimos do pressuposto de que as expressões $\pi_1\,\pi_1$, $\pi_2\,\pi_2$, $\pi_1(1-\pi_1)$ e $\pi_2(1-\pi_2)$, são maiores ou iguais a 5 e de que as duas amostras binomiais são independentes.

Deste modo, o teorema do limite central garante-nos que a distribuição amostral de p_1-p_2 é aproximadamente normal com

Média = $\pi_1 - \pi_2$ e desvio padrão= $\sigma_{p_1-p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$

Como normalmente desconhecemos os valores de π_1 e de π_2, substituímos estas duas variáveis por p_1 e p_2 referentes à amostra. Assim o intervalo de confiança para $\pi_1 - \pi_2$, a diferença entre duas populações binomiais, é

$$p_1 - p_2 \pm z\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Exemplo: Um produtor tem dois campos de macieira tratadas com diferentes tipos de insecticidas no início da estação. Quando as maçãs amadureceram, é recolhida uma amostra de cada campo e obtidos os seguintes resultados

Insecticida	Tamanho da amostra	Número de infestadas
1	400	44
2	400	24

Pretende-se estimar a diferença nos índices de infestação nas duas populações de maçãs, com uma confiança de 90%.

Se X representar o número de maçãs infestadas, as estimações pontuais são

$$p_1 = \frac{X_1}{n_1} = \frac{44}{400} = 0{,}11 \qquad p_2 = \frac{X_2}{n_2} = \frac{24}{400} = 0{,}06$$

O intervalo de confiança correspondente será

$$p_1 - p_2 \pm z\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = 0{,}11 - 0{,}06 \pm 1{,}65\sqrt{\frac{0{,}11(0{,}89)}{400} + \frac{0{,}06(0{,}94)}{400}}$$

$$= 0{,}05 \pm 1{,}65(0{,}0196) = 0{,}05 \pm 0{,}032$$

Com 90% de confiança, acreditamos que a taxa de infestação foi maior quando o insecticida 1 foi usado, com uma percentagem a variar de 1,8% até 8,2%.

É também comum conduzir-se um teste para analisar se será razoável admitir-se que duas populações binomiais não difiram. A hipótese nula é então H_0: $\pi_1=\pi_2$ ou, o que é o mesmo, H_0: $\pi_1-\pi_2=0$.

Dependendo do objectivo do estudo, assim a hipótese alternativa pode ser unilateral ou bilateral.

Se $\bar{p}$ representar a média ponderada das proporções encontradas nas amostras, a sua expressão é dada por

$$\bar{p} = \frac{Número\ de\ êxitos\ nas\ amostras}{Tamanho\ total\ das\ amostras} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

e a correspondente estatística do teste é então

$$z^* = \frac{(p_1 - p_2) - 0}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}}$$

Exemplo: Um determinado banco pretende estudar se o estado civil dos indivíduos que recorrem a empréstimo para compra de automóvel tem algo a ver com o facto destes terem problemas com o pagamento das mensalidades do empréstimo dentro de um ano.

Para isso recolheu uma amostra de 950 pedidos de financiamento aprovados com os dados seguintes

Estado civil	Total de empréstimos	Número de problemas
Não casado	413	29
Casado	537	47

Iremos testar a hipótese H_0: $\pi_{NC}=\pi_C$ contra a hipótese H1: $\pi_{NC}\neq\pi_C$. Pelas amostras temos

$$\overline{p} = \frac{X_{NC} + X_C}{n_{NC} + n_C} = \frac{29 + 47}{413 + 537} = 0{,}08$$

A estatística do teste é então

$$z^* = \frac{(p_{NC} - p_C) - 0}{\sqrt{\frac{\overline{p}(1-\overline{p})}{n_{NC}} + \frac{\overline{p}(1-\overline{p})}{n_C}}} = \frac{(0{,}0702 - 0{,}075) - 0}{\sqrt{\frac{0{,}08(0{,}92)}{413} + \frac{0{,}08(0{,}92)}{537}}} = -0{,}98$$

Para α=0,10, não devemos rejeitar a hipótese H_0 a menos que o valor da estatística do teste exceda 1,65. No nosso caso, podemos declarar que a diferença entre as taxas encontradas de não pagamento não é estatisticamente significativa.

10 Ajustamento e independência

Neste capítulo iremos ver dois novos procedimentos. O primeiro permite analisar de que modo uma determinada variável qualitativa com três ou mais categorias se distribui (distribuição multinomial); o segundo permite-nos concluir se duas variáveis categóricas estão ou não relacionadas

10.1 Populações multinomiais

O gestor de uma loja pretende ver se é igualmente provável que cada um dos três métodos usuais de pagamento (numerário, cheque ou cartão de crédito) seja utilizado pelos jovens clientes. Uma amostra aleatória de 150 compras forneceu as seguintes frequências:

Tipo de pagamento	numerário	cheque	cartão de crédito
No de compras	30	52	68

Pretendemos testar

H_0: Todas as categorias (métodos de pagamento) são igualmente prováveis na população dos pagamentos

ou, simbolicamente, podemos escrever a hipótese nula como

H_0: $\pi_{\text{numerário}} = \pi_{\text{cheque}} = \pi_{\text{cartão de crédito}} = 1/3$

onde π, com um índice, representa a respectiva proporção dos pagamentos na população. Neste exemplo, utilizaremos o erro $\alpha = 0{,}05$.

Para podermos inferir, necessitamos de saber como é que os elementos da amostra se distribuem pelas várias formas de pagamento no caso da verdade da amostra nula H_0. A estas novas frequências chamamos frequências esperadas teoricamente. No nosso caso, com uma amostra de 150

pagamentos e supondo uma probabilidade igual para cada uma das formas de pagamento, as frequências esperadas serão um terço de 150, ou seja, 50. Colocando, lado a lado, as frequências observadas e esperadas temos:

Método de pagamento	Frequência observada O	Frequência esperada E
Numerário	30	50
Cheque	52	50
Cartão de crédito	68	50

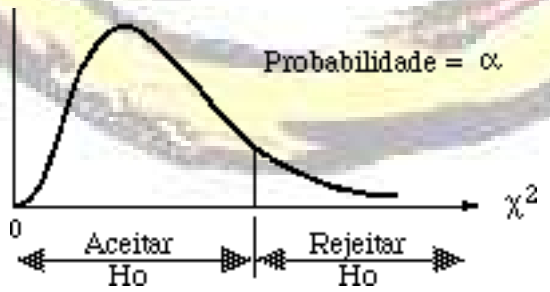
Se H_0 for verdadeira, as frequências observadas e esperadas de cada linha deverão ser semelhantes. Grandes discrepâncias entre os dois tipos de frequências sugerem que é pouco provável que H_0 seja verdade. A nossa estatística do teste será, portanto, uma quantidade que medirá a extensão da discordância entre as frequências observadas e as esperadas. Esta estatística do teste é denominada qui-quadrado e é obtida pela fórmula

$$\chi^{2*} = \sum_{\text{todas as células}} \frac{(\text{Frequência observada} - \text{Frequência esperada})^2}{\text{Frequência esperada}}$$

ou, de uma forma abreviada,

$$\chi^{2*} = \sum \frac{(O-E)^2}{E}$$

A forma geral da distribuição do χ^2 é dada pela figura abaixo.



A linha que separa a aceitação (não rejeição) da rejeição de H_0 depende do erro α de tipo I e dos graus de liberdade (gdl) associados à situação

amostral. Para testes que envolvam populações multinomiais, os graus de liberdade serão iguais ao número de celas ou categorias menos uma unidade.

Para o problema inicial, há três categorias de pagamento e, portanto, 2 graus de liberdade. Se pretendermos testar a nossa hipótese de iguais proporções com um nível de significância de 0,05, o valor correspondente na tabela do qui-quadrado é 5,991. Por outro lado, o valor da estatística do teste é dado por

$$\chi^{2*} = \sum \frac{(O-E)^2}{E} = \frac{(30-50)^2}{50} + \frac{(52-50)^2}{50} + \frac{(68-50)^2}{50} =$$

$$= 8{,}00 + 0{,}08 + 6{,}48 = 14{,}56$$

Como a estatística do teste excede o ponto de corte dado pela tabela, podemos afirmar que temos provas em número suficiente para rejeitar a hipótese nula. Concluímos então que H_0 é falsa. A cela "numerário" contribui 8,00 para a soma 14,56, enquanto que "cartão de crédito" contribuiu 6,48. Isto sugere que $\pi_{\text{numerário}}<1/3$ e que $\pi_{\text{cartão de crédito}}>1/3$. A cela "cheque" dá uma contribuição quase negligenciável para a soma 14,56.

Vejamos outro exemplo. Pretende-se estudar a queixas recebidas relativamente ao atendimento numa loja de uma cadeia de lojas. Depois de classificadas em grandes grupos, encontramos o seguinte:

A — qualidade do serviço: 31 cartas
B — cortesia do serviço: 25 cartas
C — preço de irem: 17 cartas
D — outras: 17 cartas.

Em relação a toda a cadeia a que pertence esta loja são conhecidas as percentagens de queixas para cada categoria: A = 40%, B = 25%, C = 20% e D = 15%. Testar a hipótese de que o padrão das queixas nesta loja não seja diferente do padrão geral. A hipótese nula é, então:

$$H_0: \quad \pi_A = 0{,}40, \ \pi_B = 0{,}25, \ \pi_C = 0{,}20, \ \pi_D = 0{,}15$$

Com base nesta hipótese, as 90 queixas da amostra deveriam distribuir-se do seguinte modo:

A: $n\pi_A = 90(0,40) = 36$ cartas

B: $n\pi_A = 90(0,25) = 22,5$ cartas

C: $n\pi_A = 90(0,20) = 18$ cartas

D: $n\pi_A = 90(0,15) = 13,5$ cartas

e a tabela de dupla entrada seria:

	Categoria			
Frequência	A	B	C	D
O (observada)	31	25	17	17
E (esperada)	36	22,5	18	13,5

A estatística do teste será

$$\chi^{2*} = \sum \frac{(O-E)^2}{E} = \frac{(31-36)^2}{36} + \frac{(25-22,5)^2}{22,5} + \frac{(17-18)^2}{18} + \frac{(17-13,5)^2}{13,5} =$$

$$= 0,694 + 0,278 + 0,056 + 0,907 = 1,935$$

Com quatro categorias, os graus de liberdade serão três. Se usarmos $\alpha = 0,10$ como erro correspondente à parte da direita da curva do qui-quadrado, encontramos o ponto de corte 6,251. Como $1,935 < 6,251$, aceitamos H_0 e consideramos a amostra de acordo com os dados de toda a cadeia.

Antes de prosseguir é importante salientar que os testes do qui-quadrado requerem que as frequências esperadas E sejam superiores ou iguais a 5.

10.2 Independência estatística

De seguida, iremos ver um teste do qui-quadrado que nos permite determinar se duas variáveis são independentes ou não. Estas variáveis estão dispostas numa tabela de dupla entrada ou, como a denominaremos, tabela de contingência.

Vejamos o exemplo da tabela de contingência obtida a partir de uma amostra de 90 residências, considerando o preço de venda e o tempo em que estas residências estiveram à venda no mercado:

	Dias no mercado		
Preço de venda	≤ 60 dias	> 60 dias	**Total**
Abaixo de 10 mil contos	18	12	30
De 10 a 50 mil contos	14	31	45
Acima de 50 mil contos	4	11	15
Total	36	54	90

Pretende-se saber se as duas variáveis são independentes. As hipóteses para o teste de independência são então:

H0: As duas variáveis são independentes

H1: As duas variáveis são dependentes

Testaremos a hipótese nula com um grau de significância de 0,05. Para calcular o teste do qui-quadrado necessitamos, no entanto, dos valores das frequências esperadas para cada uma destas seis celas. De uma forma intuitiva, e para os nossos dados, vemos que 40% das moradias (36 em 90) são vendidas em menos de 60 dias, depois de serem postas à venda. Assim, se o preço é independente do tempo no mercado, deveríamos ver os mesmos 40% em cada uma das categorias de preços. Por outras palavras, as frequências esperadas para a primeira linha deveriam ser 0,40(30)=12 e 0,60(30)=18; para a linha do meio 0,40(45)=18 e 0,60(45)=27; e, para a linha de baixo, 0,40(15)=6 e 0,60(15)=9. A tabela seguinte mostra, para cada cela, as frequências observadas e as frequências esperadas.

	Dias no mercado		
Preço de venda	≤ 60 dias	> 60 dias	Total
Abaixo de 10 mil contos	18 (12)	12 (18)	30
De 10 a 50 mil contos	14 (18)	31 (27)	45
Acima de 50 mil contos	4 (6)	11 (9)	15
Total	36	54	90

Outra maneira de obter as frequências esperadas é utilizar a fórmula

$$E_{ij} = \frac{\text{(Total da linha i) (Total da coluna j)}}{\text{Tamanho da amostra}}$$

onde i representa o número da linha e j o número da coluna. Também aqui todas as frequências esperadas devem ser, no mínimo, iguais a 5.

A estatística do teste é

$$\chi^{2^*} = \sum \frac{(O-E)^2}{E} =$$

$$= \frac{(18-12)^2}{12} + \frac{(12-18)^2}{18} + \frac{(14-18)^2}{18} + \frac{(31-27)^2}{27} + \frac{(4-4)^2}{4} + \frac{(11-9)^2}{9} =$$

$$= 3{,}0 + 2{,}0 + 0{,}889 + 0{,}593 + 0{,}667 + 0{,}444 = 7{,}593$$

Os graus de liberdade para uma tabela de confiança com nl linhas e nc colunas são dados por gdl=(nl-1)(nc-1). Neste exemplo, gdl=(3-1)(2-1)=2. Usando α=0,05 com gdl = 2, encontramos um ponto de corte de 5,991. Como a estatística do teste excede o valor do ponto de corte, rejeitamos H_0. Constatamos que as discrepâncias entre as frequências observadas e as esperadas são demasiado grandes para se aceitar a independência entre as variáveis.

Faz assim sentido analisar quais as celas que mais contribuíram para esta situação. No nosso exemplo, é visível que as celas da primeira linha têm o maior impacto, significando que as residências menos caras tendem a ser vendidas mais rapidamente do que as residências das outras categorias.

Vejamos outro exemplo. Um hotel decidiu, há alguns meses, adquirir 200 novos aparelhos de televisão, 80 de uma marca e 60 de cada uma de duas outras marcas. Foi então registado o número de queixas, durante esse tempo, feitas pelos hóspedes relativamente ao funcionamento dos televisores.

	Marca de televisores			
Nº de queixas	S	R	T	Total
Nenhuma	10	16	14	40
Uma	27	47	26	100
Duas ou mais	23	17	20	60
Total	60	80	60	200

Partindo do pressuposto que os aparelhos de televisão constituem amostras aleatórias das respectivas marcas, pretendemos testar, com 10% de risco, se existe alguma relação entre as duas variáveis, isto é, pretendemos testar:

H_0: Marca de televisor e número de queixas são independentes

H_1: Marca de televisor e número de queixas estão dependentes

Os graus de liberdade são gdl=(3-1)(3-1)=4. Para α=0,10, o valor de corte do qui-quadrado é 7,779. A tabela seguinte mostra-nos as frequências observadas e as correspondentes frequências esperadas, entre parêntesis.

	Marca de televisores			
Nº de queixas	S	R	T	Total
Nenhuma	10 (12)	16 (16)	14 (12)	40
Uma	27 (30)	47 (40)	26 (30)	100
Duas ou mais	23 (18)	17 (24)	20 (18)	60
Total	60	80	60	200

A estatística do teste é

$$\chi^{2*} = \sum \frac{(O-E)^2}{E} = \frac{(10-12)^2}{12} + \ldots + \frac{(11-9)^2}{9} = 6{,}378$$

Como 6,378 é menor do que o ponto de corte 7,779, não rejeitamos H_0. Não encontramos provas em número suficiente da relação entre as marcas de televisores e o número de queixas dos hóspedes.

11 Exercícios propostos

11.1 Estatística descritiva

1 Identifique as entidades e possíveis valores das seguintes variáveis:

a) nome do cartão de utente dos doentes em espera

b) número de consultas mensais realizadas num determinado centro de saúde

c) número de reclamações enviadas pelo correio por mês para um hospital

d) tipo de consulta (primeira/subsequente) de um doente na consulta externa de uma especialidade

e) número de telefone das pessoas que acorrem às urgências de um hospital

f) número de quartos por clínica em Coimbra

2 Em relação a cada variável de pergunta anterior, classifique os dados correspondentes como qualitativos ou quantitativos.

3 Partindo do pressuposto de que os dados da amostra estão isentos de erro, identifique os seguintes resultados como provenientes de estatística descritiva ou inferencial:

a) a quantidade de precipitação registada nos primeiros 10 dias do mês passado foi de 3 centímetros

b) após uma análise baseada em casas semelhantes da mesma área, pensa-se que esta casa poderá ser vendida por 250.000€.

c) uma sondagem feita uma semana antes das eleições revelou que o partido A seria o escolhido por 45% dos eleitores

4 Com base nas tabelas de números aleatórios fornecidas, identifique os números das entidades a serem incluídas na amostra, se for dada a seguinte informação:

a) N=100, n=12, ponto de partida: linha 7, coluna 5, para baixo

b) N=69, n=5, ponto de partida: linha 4, coluna 5, para a direita

c) N=811, n=10, ponto de partida: linha 40, coluna 9, para cima

5 Pretende-se obter uma amostra sistemática de 81 doentes seleccionados de uma população numerados entre 14.522 e 21.471. Suponha que um auditor escolheu o número 51 como o ponto aleatório de começo.

a) Será possível escolher 51 como ponto de partida?

b) Em caso afirmativo, quais são as 3 primeiras facturas seleccionadas? Em caso negativo, como é que então poderíamos proceder?

6 Classifique os seguintes dados como discretos ou contínuos e crie várias categorias para incluir os possíveis dados:

a) número de vezes por mês que um indivíduo recorre ao centro de saúde.

b) taxa de êxito por tipo de intervenção.

7 Pense numa variável que produza dados discretos e define-a em palavras. Repita este exercício para dados contínuos.

8 Considere a seguinte distribuição de frequência baseada num conjunto de dados:

X	0	1	2	3	4	5
f	10	17	14	9	3	1

a) Qual o tamanho da amostra?

b) Qual a frequência relativa de X=2?

c) Que tipo de distribuição de frequência é este?

d) Crie o conjunto das frequências acumuladas.

9 Quantas classes são necessárias para organizar um conjunto de dados com as seguintes observações?

a) n=60 **b)** n=200 **c)** n=15

10 Pretende-se organizar uma amostra de tamanho 63 numa distribuição de frequências agrupadas.

a) Quantas classes devem ser usadas?

b) Se os valores mínimo e máximo forem respectivamente 61,1 e 83,7 qual deve ser o valor do comprimento da classe.

c) Quais os limites extremos da primeira classe.

11 Considere a seguinte distribuição de frequências agrupadas:

No. classe	Classe	f
1	47 - 58	7
2	58 - 69	18
3	69 - 80	24
4	80 - 91	12
5	91 - 102	5

a) Determine o valor de w.

b) Qual o ponto médio da classe 2?

c) Qual a frequência relativa da classe 3?

d) Crie a coluna das frequências acumuladas.

12 Organize o seguinte conjunto de dados numa distribuição de frequências agrupadas:

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

13 O departamento de recursos humanos de um hospital registou o número de dias de doença de uma amostra de profissionais nos últimos seis meses. Os dados são os seguintes:

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

a) Defina a variável representada por estes números.

b) Será que produz dados discretos ou contínuos?

c) Qual a entidade neste problema?

d) Devem os dados ser organizados em distribuições de frequência agrupadas ou não agrupadas?

e) Apresente a distribuição de frequências apropriada.

14 Desenhe o histograma de frequências relativas correspondente à seguinte distribuição de frequências agrupadas:

Classes	f
0,45 - 8,05	3
8,05 - 15,65	5
15,65 - 23,25	8
23,25 - 30,85	4

a) Qual o valor de w?

b) Qual o ponto médio da classe 23,25 - 30,85?

c) Qual a frequência relativa acumulada no fim da classe 8,05 - 15,65?

15 A especificação técnica de uma sonda usada numa cirurgia indica que deve ter 5 milímetros de diâmetro. Uma amostra aleatória de 20 sondas possuía os diâmetros abaixo indicados:

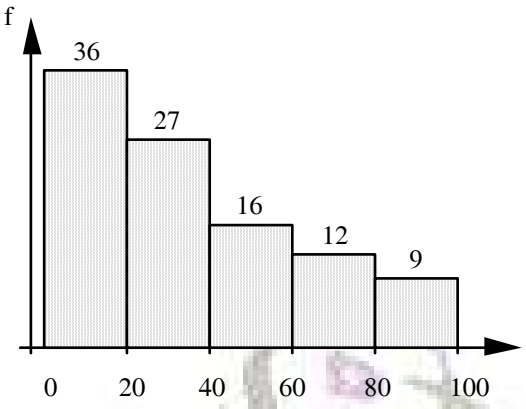
4,961	4,982	4,991	4,999	5,001
4,964	4,984	4,992	4,999	5,003
4,975	4,984	4,994	5,000	5,004
4,975	4,987	4,997	5,001	5,007

Organize estes dados numa distribuição de frequências agrupadas com frequência relativa, frequência relativa acumulada e ponto médio.

16 Considere o histograma de frequência da figura seguinte para um conjunto organizado de dados:

a) Reconstrua, a partir deste gráfico, a correspondente tabela de distribuição de frequências?

b) Qual o tamanho da amostra?



17 Construa um histograma de frequência para cada uma das seguintes distribuições de frequência:

a)

Classes	f
1,77 - 2,19	7
2,19 - 2,61	24
2,61 - 3,03	35
3,03 - 3,45	21
3,45 - 3,87	10

b)

X	f
10	1
11	5
12	7
13	0
14	12
15	3

c)

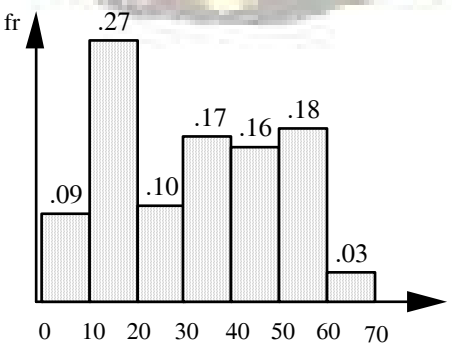
Classes	f
45 - 50	9
40 - 45	17
35 - 40	11
30 - 35	5
25 - 30	4
20 - 25	2

Identifique, se possível, a forma comum do histograma amostral.

18 Sendo dado as seguintes classe, em que classe incluiria o valor X = 14?

Classes
10 - 14
14 - 18
18 - 22

19 Considere o seguinte histograma de frequências relativas para um conjunto de 231 adultos, estando a ser estudado o número de dias de dores fortes de cabeça durante o último ano.



a) Qual é o comprimento de cada classe?

b) É possível determinar a frequência da última classe? Em caso afirmativo, qual é o seu valor?

20 No fim de um curso de formação, foi entregue aos vários participantes uma folha de avaliação onde se pedia que respondessem à seguinte questão: "Este curso terá um impacto imediato no meu desempenho como profisional de saúde". As respostas possíveis eram concordo muito, concordo, nem concordo nem discordo, discordo, discordo muito. Obteve-se a seguinte lista de respostas:

Participante	Resposta	Participante	Resposta
1	Concordo	15	Concordo
2	Nem concordo nem discordo	16	Concordo
3	Nem concordo nem discordo	17	Concordo muito
4	Discordo	18	Concordo
5	Concordo	19	Nem concordo nem discordo
6	Concordo	20	Nem concordo nem discordo
7	Discordo	21	Concordo
8	Nem concordo nem discordo	22	Concordo
9	Nem concordo nem discordo	23	Concordo
10	Concordo muito	24	Discordo
11	Concordo	25	Concordo
12	Concordo	26	Concordo
13	Concordo muito	27	Nem concordo nem discordo
14	Concordo	28	Nem concordo nem discordo

a) Compile os dados numa tabela de frequências.

b) Se fosse o responsável pelo referido seminário, considerá-lo-ia um êxito em termos do impacto imediato no desempenho dos participantes? Justifique.

21 Uma cadeia de restaurantes a nível nacional decidiu criar um índice de eficiência baseado em vários entre os quais factores orçamentais, de venda e de limpeza. Uma amostra de restaurantes forneceu os seguintes valores de índices (85,00 é o valor máximo possível):

a) Quantas classes devem ser usadas para agrupar estes dados?

b) Determine o valor do comprimento de cada classe.

c) Organize os dados numa distribuição de frequências agrupadas.

d) Desenvolva o conjunto das frequências acumuladas.

e) Desenhe o histograma para a referida distribuição.

73,65	74,01	78,02	63,15	69,73	63,68
58,14	71,25	74,28	73,90	76,26	69,27

73,13	74,20	78,06	71,32	71,88	72,45
70,12	72,35	79,00	56,85	74,53	64,43
70,75	72,11	68,48	70,30	77,53	74,22
78,64	79,01	72,62	78,26	74,89	69,56
70,87	78,63	73,26	79,74	80,20	52,36
66,81	73,53	71,90	81,14	53,88	65,90
79,66	59,58	73,61	78,16	74,08	70,11
60,35	66,92	74,24	59,15	67,92	60,09
51,92	65,70	66,32	54,90	71,54	63,58
60,89	56,56	60,26	53,41	60,03	74,92
71,64	69,43	68,04	69,84	80,14	69,30
67,06	59,56	75,73	74,18	70,20	64,32
73,81	68,80	71,60	71,54	78,81	78,54
57,86	65,35	67,47	73,92	68,22	76,37

22 Para cada uma das seguintes situações, indique se a medida numérica descritiva (em sublinhado) é uma estatística ou um parâmetro:

a) Um grupo de cinco notas é seleccionado de uma classe de 52 estudantes, e é calculado o total das cinco notas.

b) Recolhemos aleatoriamente o custo de um pacote de leite meio-gordo numa amostra de quatro supermercados da cidade de Coimbra. De seguida, calculamos a média destes preços.

c) Para estudar a ocupação dos automóveis que entram no perímetro de um grande hospital um observador colocou-se numa das principais entradas e contou o número de ocupantes em cada um de dez automóveis. Determinou-se a percentagem de automóveis com mais do que um ocupante.

23 Encontre a moda, a mediana, a média, a amplitude, o desvio médio, a variância e o desvio padrão das seguintes amostras de dados:

a) 12 9 11 12 10 14 10 11 10

b) 2 0 6 3 9 2 0 2

c) 8,9 -5,1 12,3 4,5 8,4 5,0

d) -4 -6 -1 0 -2

e) 0,08 0,03 0,05 0,03 7,2 0,07 0,01 0,05

24 Uma amostra aleatória de tamanho 37 produziu a seguinte distribuição de frequências não agrupadas:

X	0	1	2	3	4
f	12	1	6	11	7

Encontre os valores da moda, mediana, média e desvio padrão.

25 Considere a seguinte distribuição de frequências não agrupadas:

X	-4	-2	-1	0	1	3
f	24	56	33	17	8	2

Encontre os valores da moda, mediana, média e desvio padrão.

26 27 empresas candidatas ao fornecimento de um produto a um hospital foram sujeitas a uma bateria de testes de onde se obtiveram os seguintes valores. O maior valor possível é 25 e corresponde ao maior potencial de êxito.

Pontuação	f
4 - 8	3
8 - 12	8
12 - 16	7
16 - 20	7
20 - 24	2

a) Qual o valor do comprimento da classe?

b) Identifique a moda.

c) Determine a pontuação média.

d) Que número divide a meio o conjuntos das pontuações?

e) Determine os valores da variância e do desvio padrão.

27 Pretendeu-se estudar o número de vezes que os adolescentes tinham visto o seu filme favorito. As respostas a um inquérito foram distribuídas da seguinte maneira:

No. de vezes que o filme é visto	f
1	19
2	44
3	35
4	21
5	11
6	3
7	1

a) Será possível determinar quantos adolescentes responderam ao inquérito? Em caso afirmativo, determine este valor.

b) Qual o número médio de vezes que o filme favorito é visto?

c) Determine a mediana.

d) Será correcto afirmar-se que o mais provável é que os adolescentes vejam duas vezes o seu filme referido? Justifique.

28 Em meados da década de 80 começou-se a sentir uma escassez de pessoal de enfermagem nos hospitais norte-americanos. Os seguintes dados foram publicados pela American Hospital Association e dizem respeito ao número de enfermeiros por cama de hospital, por estado e em 1986:

Estado	Enferm	Estado	Enferm	Estado	Enferm
Alabama	0,72	Kentuky	0,75	North Dakota	0,58
Alaska	0,91	Louisiana	0,64	Ohio	0,83
Arizona	0,86	Maine	0,80	Oklahoma	0,70
Arkansas	0,76	Maryland	0,80	Oregon	0,85
California	0,88	Massachusetts	0,82	Pennsylvania	0,81
Colorado	0,77	Michigan	0,86	Rhode Island	0,86
Connecticut	0,79	Minnesota	0,61	S Carolina	0,78
Delaware	0,67	Mississippi	0,61	South Dakota	0,55
D,C,	0,88	Missouri	0,73	Tennessee	0,70
Florida	0,77	Montana	0,59	Texas	0,71
Georgia	0,76	Nebraska	0,61	Utah	0,95
Hawaii	0,80	Nevada	0,81	Vermont	0,76
Idaho	0,74	N Hampshire	0,87	Virginia	0,75
Illinois	0,76	New Jersey	0,74	Washington	0,88
Indiana	0,67	New Mexico	0,77	West Virginia	0,74
Iowa	0,62	New York	0,75	Wisconsin	0,62
Kansas	0,59	North Carolina	0,77	Wyoming	0,54

a) Partindo do pressuposto de que estes dados constituem uma amostra representativa, determine a média, a mediana e o número modal de pessoal de enfermagem por cama.

b) Organize os dados numa tabela de distribuição de frequências agrupadas. Determine a média, a mediana e a moda resultantes da distribuição e compare estes valores com os da alínea anterior. São idênticos? Porquê?

c) Encontre o valor do desvio padrão.

29 Recolheu-se o tempo de estadia num hospital de 161 doentes submetidos a cirurgia.

Dias de internamento	No. de doentes
9	18

10	21
11	29
12	25
13	10
14	22
15	8
16	12
17	5
18	11

a) Encontre a amplitude de valores para a variável X=duração (em dias) no hospital.

b) Calcule o desvio médio.

c) Qual o desvio em relação à média para X = 10?

d) Determine o valor do desvio padrão de X.

30 Suponha que lhe são dados os seguintes valores obtidos de uma amostra de tamanho n=34:

Mo = 38 Amp = 32

Md = 37 Min = 18

$\overline{X}$ = 35,4 S = 6,6

Esboce a forma provável da respectiva curva de frequências.

31 A tabela seguinte apresenta as coordenadas para n=5 pares de observações (x,y):

X	-2	-1	0	1	2
Y	0	0	1	1	3

a) Encontre a recta dos mínimos quadrados para os dados.

b) Como verificação dos cálculos da alínea a), represente os cinco pontos e a recta dos mínimos quadrados. Será que a recta é uma boa aproximação dos pontos de dados?

32 A tabela seguinte apresenta as coordenadas para n=5 pares de observações (x,y):

X	-3	-1	1	1	2
Y	6	4	3	1	1

a) Encontre a recta dos mínimos quadrados para os dados.

b) Como verificação dos cálculos da alínea a), represente os cinco pontos e a recta dos mínimos quadrados. Será que a recta é uma boa aproximação dos pontos de dados?

33 A tabela seguinte apresenta as coordenadas para n = 7 pares de observações (x,y):

X	7	8	2	3	5	3	7
Y	2	0	5	6	4	9	2

a) Encontre a recta dos mínimos quadrados para os dados.

b) Como verificação dos cálculos da alínea a), represente os sete pontos e a recta dos mínimos quadrados. Será que a recta é uma boa aproximação dos pontos de dados?

34 A tabela seguinte apresenta o lucro y pelas vendas (milhões de euros) de uma empresa de construção em nove projectos de centros de saúde e os número x de anos de experiência do responsável de cada projecto.

X	4	4	2	6	2	2	4	6	6
Y	2,0	3,5	8,5	4,5	7,0	7,0	2,0	6,5	8,0

a) Encontre a recta dos mínimos quadrados para os dados.

b) Como verificação dos cálculos da alínea a), represente os nove pontos e a recta dos mínimos quadrados. Será que a recta é uma boa aproximação dos pontos de dados?

35 Sendo apresentados os dados relativos a duas variáveis X e Y:

X	2	5	4	2	1
Y	2	2	7	7	2

a) Calcule o coeficiente de correlação para os dados apresentados na tabela.

b) Determine a regressão linear simples

c) Esboce graficamente o diagrama de dispersão e a regressão linear simples.

d) Comente o significado da recta de regressão encontrada.

37 O orçamento flexível é uma expressão das expectativas do gestor relativamente a receitas e custos para um certo período de tempo e serve para comunicar os objectivos da gestão aos vários gestores da organização. A gestão de uma empresa de manufactura está interessada em criar um orçamento flexível para estimar os custos extra da produção. Uma análise histórica forneceu os seguintes dados:

Produção (× $10000)	3	4	5	6	7	8	9
Custos extra (× $1000)	12,0	10,5	13,0	12,0	13,0	13,3	16,5

a) Encontre a recta dos mínimos quadrados que permita a estimação dos custos extra a partir da produção (isto é, a recta dos mínimos quadrados que relaciona custo com produção).

b) Represente graficamente os pontos e a recta.

38 Uma amostra aleatória referente a pedidos de manutenção de seis fotocopiadoras no último mês forneceu os seguintes dados:

Idade da máquina em meses	Nº de reparações pedidas
4	2
7	1
12	3
10	2
2	0
13	4

a) Apresente o diagrama de dispersão.

b) Determine o coeficiente de correlação.

c) Que pensa do significado da relação linear entre as duas variáveis?

39 Suponhamos que oito espécimes de um certo tipo de liga metálica para a construção de material cirúrgico são produzidos a diferentes temperaturas e que se registou a solidez de cada um. Os valores observados são dados na tabela a seguir, onde x_i representar a temperatura (em unidades codificadas) a que o espécime i foi produzido e yi representa a solidez (em unidades codificadas) desse espécime.

i	x_i	y_i
1	0,5	40
2	1,0	41
3	1,5	43
4	2,0	42
5	2,5	44
6	3,0	42
7	3,5	43
8	4,0	42

Encontre a recta da forma $y = b_0 + b_1 x$ para estes valores através dos mínimos quadrados.

11.2 Estatística inferencial

1 Seja X a variável aleatória que representa a pressão sanguínea. Suponhamos que a variável tem distribuição normal de média μ e desvio padrão σ=8. Suponha que se toma uma amostra de 100 indivíduos cuja média é $\overline{X}$=123. Determine um intervalo com 90% de confiança para a média.

2 Seja X a variável que representa a taxa normal de glicose no sangue. Suponhamos que essa distribuição é normal com desvio padrão σ=6 mg/100ml de sangue. Determine o intervalo de confiança com 95% de confiança para μ, sabendo que numa amostra aleatória de 25 indivíduos se encontrou uma média $\overline{X}$ = 5 mg/100ml.

3 Foi determinada a taxa de glicose em 8 ratos, tendo sido encontrados os seguintes valores: 100; 87,5; 110; 99,5; 92,5; 94; 100; 100. Calcule, com base nestes dados, o intervalo de 95% de confiança para a média da população.

4 Seja X a variável que representa a taxa de colesterol no plasma sanguíneo. A média desta variável é μ e o desvio padrão é σ=20 mg/100ml de plasma. Qual o intervalo de confiança para o parâmetro

desconhecido μ, quando foi recolhida uma amostra de 25 indivíduos para os quais a média encontrada foi de 198 mg/100ml.

a) Com um grau de confiança de 95%

b) Com um grau de confiança de 99%

5 Seja X uma variável aleatória que representa a taxa normal de colesterol, supondo que com base numa amostra de 25 indivíduos um investigador obteve a média $\overline{X}$ = 198 mg/100ml e o desvio padrão s = 30 mg/100ml.

a) Determine um intervalo de confiança a 90% para μ.

b) Suponha agora que eram 50 indivíduos. Quais os limites do intervalo de confiança?

6 Suponha a seguinte amostra proveniente de uma população com variância s^2 = 64: 1, 20, 18, 19, 19, 15, 31, 12. Determine os intervalos de confiança de 95% e de 99% para μ

7 Uma amostra de 100 votantes escolhidos aleatoriamente de um distrito indicaram que 55% eram a favor de um determinado candidato. Encontrar intervalos de confiança a 95% e 99% para a proporção de todos os votantes nesse candidato.

8 Com base no problema anterior, qual o tamanho que a amostra deveria ter para se ter 95% e 99% de confiança que o candidato seria eleito?

9 Em 40 lançamentos de uma moeda foram obtidas 24 caras. Encontrar um intervalo de confiança a 95% e a 99% para a proporção de caras que seriam encontradas num lançamento um número infinito de vezes da moeda.

10 Um empacotador de sacos de 50 Kg de sal seleccionou uma amostra de 30 sacos cheios nas últimas 24 horas. Cada saco é cuidadosamente cheio, tendo sido encontradas as seguintes estatísticas:

$$\overline{X} = 50{,}22 \text{ Kg} \qquad s^2 = 0{,}24 \text{ Kg}^2 \qquad s = 0{,}49 \text{ Kg}$$

a) Construa um intervalo de confiança a 99% para σ^2.

b) Encontre um intervalo de confiança a 99% para σ.

11 Teste H_0: $\sigma^2=10$ contra H_1 unilateral à direita, ao nível 0,05. Suponha que a sua amostra de 20 itens apresentou o valor $s^2=11,6$. Podemos rejeitar H_0?

12 Para testar a hipótese de equilíbrio de uma moeda, adoptou-se a seguinte regra de decisão: (1) aceitar a hipótese se o número de caras numa única amostra de 100 lançamentos esteja entre 40 e 60 inclusive, (2) rejeitar a hipótese no caso contrário.

a) Encontre a probabilidade de rejeitar a hipótese quando esta está correcta.

b) Interprete graficamente a regra de decisão e o resultado anterior.

c) Que conclusões tiraria se nos 100 lançamentos da moeda se obtiverem 53 caras? E 60 caras?

13 Numa experiência de percepção extra-sensorial pede-se a um indivíduo que adivinhe a cor (vermelho ou azul) de uma carta tirada de um baralho de 50 cartas. O indivíduo desconhece quantas cartas vermelhas e azuis estão no baralho. Se este identificar correctamente a cor de 32 cartas, determine se os resultados são significativos a um nível de significância de 0,05. E se este nível baixar para 0,01?

14 O fabricante de um medicamento afirma que ele é 90% eficaz no alívio de uma alergia no período de 8 horas. Numa amostra de 200 pessoas que tinham alergia, o medicamento mostrou ser eficaz em 160 pessoas. Determine a legitimidade da afirmação do fabricante.

15 Teste bilateralmente H_0: $\mu=48$ com base numa amostra de $n=60$ elementos onde se encontrou $s=6,1$ e $\overline{X}=46,1$. Use um nível de significância de 10%.

16 Uma determinada marca de detectores de fumo afirma que, em média, as peças que vende actuam sempre que sejam expostas a 375 partes por

milhão (ppm) de fumo no ar. Para testar esta frase, expuseram-se três dúzias de extintores a concentrações de fumo, num laboratório. Os valores, a partir dos quais os extintores actuaram, são os seguintes:

301	319	321	329	341	341
341	343	348	357	360	360
361	369	381	383	384	384
386	388	391	391	392	393
394	401	404	407	407	411
419	435	444	451	475	476

Usando um risco de erro do tipo I de 5%, veja se os dados contradizem o construtor.

17 Um gestor de uma dependência bancária afirma que a média de um levantamento de uma máquina é de 7500. Para se testar esta frase, a um nível de significância de 10%, recolheu-se uma amostra de 50 levantamentos e encontrou-se uma média de 6920 e um desvio padrão de 1850. Conclua se a diferença encontrada de 580 é estatisticamente significativa.

18 Teste H_0: $\mu=11.600$ contra uma hipótese unilateral à direita, usando o nível de significância de 0,10. Uma amostra de 127 elementos apresentou a média de 11.891 e um desvio padrão igual a 2.886.

19 No teste H_0: $\mu = 57$ dias contra H_1: $\mu<57$ dias, a estatística do teste for $z^*=-0,85$. Determine e interprete o valor de p.

20 No teste H_0: $\mu=36$ contra H_1: $\mu\neq36$, a estatística do teste for $z^*=0,65$. Determine e interprete o valor de p.

21 Uma amostra de 150 lâmpadas da marca A teve uma média de duração de 1400 horas com um desvio padrão de 120 horas. Uma amostra de 200 lâmpadas da marca B teve uma duração média de 1200 horas com um desvio padrão de 80 horas. Encontrar os limites de confiança a 95% e a 99% para a diferença entre os tempos médios de vida das populações das lâmpadas das marcas A e B.

22 Suponhamos dois grupos A e B, formados respectivamente por 50 e 100 doentes semelhantes. Ao grupo A foi administrado um novo tipo de medicamento para dormir e, ao grupo B, um medicamento convencional. Os doentes do grupo A estiveram uma média de 7,83 horas a dormir, com um desvio padrão de 0,24 horas; os doentes do grupo B estiveram 6,75±0,30 horas a dormir. Encontre os limites de confiança a 95% e 99% para a diferença do número médio de horas a dormir, induzidas por ambos os medicamentos.

23 Numa amostra de 400 adultos e 600 adolescentes que assistiam a um programa de televisão, 100 adultos e 300 adolescentes declararam que gostaram dele. Construa intervalos de confiança a 95% e a 99% para a diferença em proporções de todos os adultos e todos os adolescentes que viram o programa e gostaram dele.

24 Numa amostra de 200 peças produzidas por uma máquina foram encontradas 15 defeituosas, enquanto que numa amostra de 100 peças produzidas por outra máquina, apenas foram encontradas 12 defeituosas. Encontrar um intervalo de confiança a 95% e outro a 99% para a diferença entre proporções de peças defeituosas produzidas por ambas as máquinas.

25 Um mesmo exame foi dado a duas turmas com 40 e 50 alunos, respectivamente. Na primeira turma a média das notas foi 14,8 com um desvio padrão de 1,6, enquanto que na segunda turma houve uma média de 15,6 com um desvio padrão de 1,4. Existe alguma diferença significativa entre o desempenho das duas turmas a um nível de confiança de 95%?

26 Para testar a eficácia de um novo fertilizante na produção de trigo uma propriedade foi dividida em 60 áreas iguais. O novo fertilizante foi aplicado em 30 áreas e o velho fertilizante aplicado nas restantes. Nas áreas onde foi usado o novo fertilizante, o número médio de alqueires obtidos foi de 18,2, com um desvio padrão de 0,63 alqueires. Nas restantes áreas, o número médio de alqueires colhidos e o desvio padrão

foram, respectivamente, de 17,8 e de 0,54. Usando um nível de 0,05 teste a hipótese de que o novo fertilizante seja melhor o que o velho.

27 Uma amostra de 300 votantes de um concelho A e uma de 200 votantes do concelho B mostraram que 56% e 48% deles eram a favor de um candidato. Com o nível de significância de 0,05 teste a hipótese de que haja diferença entre os votantes de ambos os concelhos.

28 Dois grupos A e B de 100 indivíduos cada têm uma determinada doença. Aos doentes do grupo A é dado um medicamento em ensaio clínico e aos do grupo B não é ministrado esse medicamento. Em tudo o resto os doentes são tratados de uma forma semelhante. 75% e 65% dos doentes, respectivamente, dos grupos A e B recuperaram da doença. Teste a hipótese de que este medicamento cure os doentes usando um nível de significância de 0,01.

11.3 Ajustamento e independência

1 Use o teste do qui-quadrado de ajustamento para ver se as discrepâncias entre as frequências observadas e as esperadas dadas na tabela seguinte são demasiado grandes para poderem ser explicadas aleatoriamente.

	Categorias			
Frequências	A	B	C	D
E (esperadas)	26	30	25	15
O (observadas)	24	44	20	8

2 Num casino, um dado é lançado 90 vezes, com os seguintes resultados:

Face do dado	1	2	3	4	5	6
Frequência	14	22	18	14	9	13

Teste a hipótese de que o dado está equilibrado, com um risco de 10%

3 Uma empresa de venda por correio apresentou um novo tipo de gabardina para homem no seu catálogo de outono/inverso. Com base em experiência anterior com outros tipos de vestuário para homem consegue-se prever uma distribuição para as vendas dos vários tamanhos das gabardinas: S = 12%, M = 45%, L = 35% e XL = 8%. As primeiras 300 gabardinas encomendadas foram S = 22, M = 125, L = 119 e XL =

34. Considerando esta amostra como aleatória, será que é razoável manter as proporções originais? Use α = 0,10.

4 Uma empresa de camionagem pretende testar (α=0,10) se a distribuição de passageiros ao longo dos vários dias da semana é uniforme. Use os seguintes dados:

Dia	Seg	Ter	Qua	Qui	Sex
Frequência	57	63	64	69	54

5 O responsável pelo departamento comercial de uma empresa classifica os seus clientes de acordo com o tempo de pagamento das facturas: pronto pagamento, 1-30 dias, 31-60 dias e acima de 60 dias. Há seis meses a situação era a seguinte:

Pagamento	**%**
Pronto	68
1-30 dias	17
31-60 dias	10
Acima de 60 dias	5

Agora, que a empresa aumentou 5% a penalidade para o atraso no pagamento, recolheu-se uma amostra de 125 facturas e obtiveram-se os seguintes resultados: 91 pronto pagamentos, 19 pagamentos de 1 a 30 dias, 11 pagamentos de 31 a 60 dias e 4 pagamentos a acima de 60 dias. Teste a hipótese da não alteração, usando α=0,10.

6 Use a seguinte amostra para determinar se as variáveis A e B são, na população, dependentes ou independentes.

	Variável A		
Variável B	A_1	A_2	A_3
B_1	5	10	21
B_2	19	26	27

Obtenha as conclusões com um erro de 1%.

7 Uma amostra de 489 adultos respondeu a um inquérito acerca das suas atitudes em relação à publicidade feita por advogados. Afirmava-se que grande parte do preço cobrado pelos advogados era para pagar custos legais e pedia-se opinião aos respondentes. Os resultados foram:

	Resposta		
Idade	Concorda	Sem opinião	Descorda
+ Jovens (< 48 anos)	47	15	169
- Jovens (≥ 48 anos)	54	41	163

Partindo do pressuposto de que os dados provêm de uma amostra aleatória, teste a independência entre as respostas e a idade dos respondentes. Use 10% de significância.

8 A um grupo de jovens do sexo feminino foram dadas amostras de shampoos e pedido que dessem a sua opinião em relação a um novo produto. Os frascos tinham todos a mesma fórmula química, variando apenas na cor do líquido. Um terço das pessoas recebeu frascos de shampoo castanho, outro terço frascos de shampoo verde e o restante terço frascos de shampoo vermelho. As jovens usaram o shampoo durante três semanas e preencheram um questionário onde lhes era pedido, entre outras, que o avaliassem em relação a outros já experimentados. As respostas permitiram preencher a seguinte tabela:

	Cor		
Avaliação	Vermelha	Verde	Castanha
Acima da média	28	13	13
Na média	20	28	36
Abaixo da média	27	34	26

Tabelas

A1 Números com 2 dígitos aleatoriamente dispostos

Linha	Coluna									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
1	10	54	93	57	92	65	53	40	83	25
2	22	48	26	16	56	19	18	61	50	81
3	24	52	41	01	30	68	67	97	90	87
4	42	32	86	13	47	62	80	12	38	59
5	37	29	70	36	91	64	15	58	49	31
6	77	51	21	04	55	39	46	27	00	20
7	99	07	23	75	11	34	06	84	45	74
8	96	09	71	79	08	88	72	76	82	98
9	89	63	03	95	02	33	14	73	78	17
0	85	28	94	35	66	05	69	43	60	44

Nota: Esta tabela pressupõe uma amostragem sem substituição de uma população finita com, no máximo, 100 entidades. Cada dígito de 00 a 99 aparece apenas uma vez.

A2 Números com 3 dígitos aleatoriamente dispostos

	Coluna																							
Linha	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
1	815	258	826	409	935	707	253	276	128	395	901	224	704	566	038	809	788	405	525	969	575	242	49	770	435
2	296	963	290	069	204	876	299	027	466	895	430	078	784	728	527	338	277	081	970	589	972	497	741	292	540
3	007	333	143	394	041	493	073	141	119	882	438	994	709	151	588	050	289	651	614	249	937	853	752	138	342
4	053	780	683	887	055	144	672	761	922	402	807	833	756	584	810	547	499	377	318	544	768	260	858	838	029
5	919	862	064	356	474	982	464	791	920	975	638	212	346	821	502	618	746	535	668	369	940	750	347	506	868
6	005	454	012	188	231	533	827	366	339	508	261	718	733	094	252	449	820	434	751	758	354	158	421	933	372
7	690	814	631	909	237	278	781	173	516	008	186	852	686	116	946	280	004	076	373	902	477	677	932	251	691
8	259	017	468	279	859	742	496	900	305	662	297	925	884	938	560	725	894	327	720	945	999	311	193	847	112
9	097	385	987	735	174	101	491	437	157	799	625	051	154	957	717	904	942	398	363	757	621	640	241	285	198
10	915	865	329	298	236	954	301	149	531	213	648	453	881	043	427	696	238	022	565	991	813	436	414	956	234
11	179	467	848	958	570	142	129	908	086	971	428	786	835	062	159	732	762	310	708	445	462	011	898	314	514
12	465	424	649	879	207	417	886	312	095	719	864	134	079	532	330	873	541	203	611	412	538	576	678	981	365
13	921	590	812	641	870	879	552	341	610	655	178	397	617	797	917	667	585	216	841	842	619	955	545	910	181
14	145	099	703	256	265	337	561	699	605	352	953	361	111	375	201	978	136	548	808	918	063	629	724	871	118
15	984	239	303	390	020	512	773	283	376	793	490	426	452	353	092	519	928	636	300	036	977	624	734	404	507
16	349	926	702	680	831	199	943	122	045	869	739	806	486	343	378	817	567	787	659	358	609	581	220	082	018
17	700	795	948	048	993	286	664	170	025	470	485	688	671	850	801	878	250	098	067	014	927	612	088	494	460
18	539	083	961	722	469	415	964	153	644	645	068	000	393	591	766	660	171	536	607	367	293	169	461	646	551
19	760	180	052	205	319	546	248	121	221	528	152	860	191	803	571	060	233	601	413	263	100	986	966	382	84
20	907	856	381	573	663	383	846	656	523	628	934	165	196	941	804	598	684	443	247	302	564	037	874	070	459
21	643	796	211	042	729	583	444	077	457	647	420	056	384	572	620	529	730	163	223	823	160	488	877	380	114
22	089	032	849	264	391	676	267	557	331	006	345	765	743	418	779	798	370	811	480	792	015	335	723	776	705
23	950	976	669	047	511	534	863	071	710	243	313	359	482	939	127	857	117	861	687	013	021	175	713	996	120
24	156	439	578	698	066	255	189	135	130	168	738	187	172	885	774	235	890	245	284	652	592	906	521	854	093
25	164	603	407	657	992	176	679	432	219	227	753	947	693	689	777	195	503	087	472	836	883	131	580	586	368
26	182	218	080	579	924	031	308	721	340	125	355	271	892	952	001	501	558	951	107	307	745	140	190	965	635
27	731	824	124	834	074	061	040	840	244	888	542	974	455	775	065	759	230	650	626	362	044	210	019	217	747
28	574	637	009	425	240	364	200	379	388	830	692	802	495	476	562	505	225	492	471	778	024	711	030	139	226
29	304	309	332	563	893	162	023	844	257	595	197	324	484	889	785	755	106	072	749	998	988	392	085	034	027
30	166	374	194	185	054	325	396	058	829	764	410	281	416	348	322	851	387	429	184	137	673	577	109	593	771
31	967	115	763	896	035	670	016	294	473	597	923	555	697	633	727	401	772	515	675	983	423	845	685	046	513
32	389	371	642	627	475	979	344	209	828	504	816	654	931	110	867	192	790	246	155	553	604	980	326	968	602
33	316	275	202	075	613	622	232	899	510	229	543	715	661	448	419	463	682	517	442	403	962	498	905	272	091
34	789	653	350	639	606	616	587	634	599	113	837	989	960	596	033	712	269	262	959	489	985	559	929	714	594
35	039	500	839	282	270	930	658	479	183	875	800	866	126	288	518	825	386	855	406	891	783	321	456	736	550
36	744	912	167	549	400	632	148	441	487	568	408	767	133	483	524	608	805	520	716	903	973	944	681	274	320
37	090	913	351	295	059	615	105	665	273	206	254	147	754	317	287	132	726	666	537	782	458	694	623	108	334
38	422	446	819	522	266	872	911	323	880	740	997	222	914	411	916	146	228	096	600	123	177	431	451	530	291
39	161	556	818	674	630	949	794	150	440	399	357	995	336	706	843	748	102	028	306	447	268	010	481	701	769
40	214	215	104	003	554	526	990	450	208	328	822	360	737	569	002	936	509	103	478	832	582	695	433	057	315

A3 Valores de significância para r

Graus de Liberdade	Nível de Significância		Graus de Liberdade	Nível de Significância	
N-2	.05	.01	N-2	.05	.01
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.323
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

A4 Área debaixo da curva normal de 0 a X

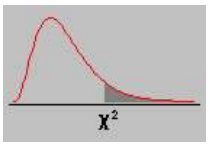
X	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998

A5 Valores críticos da distribuição t de Student para ν gdl

	Probabilidade de exceder os valores críticos				
ν	0.10	0.05	0.025	0.01	0.001
1	3.078	6.314	12.706	31.821	318.313
2	1.886	2.920	4.303	6.965	22.327
3	1.638	2.353	3.182	4.541	10.215
4	1.533	2.132	2.776	3.747	7.173
5	1.476	2.015	2.571	3.365	5.893
6	1.440	1.943	2.447	3.143	5.208
7	1.415	1.895	2.365	2.998	4.782
8	1.397	1.860	2.306	2.896	4.499
9	1.383	1.833	2.262	2.821	4.296
10	1.372	1.812	2.228	2.764	4.143
11	1.363	1.796	2.201	2.718	4.024
12	1.356	1.782	2.179	2.681	3.929
13	1.350	1.771	2.160	2.650	3.852
14	1.345	1.761	2.145	2.624	3.787
15	1.341	1.753	2.131	2.602	3.733
16	1.337	1.746	2.120	2.583	3.686
17	1.333	1.740	2.110	2.567	3.646
18	1.330	1.734	2.101	2.552	3.610
19	1.328	1.729	2.093	2.539	3.579
20	1.325	1.725	2.086	2.528	3.552
21	1.323	1.721	2.080	2.518	3.527
22	1.321	1.717	2.074	2.508	3.505
23	1.319	1.714	2.069	2.500	3.485
24	1.318	1.711	2.064	2.492	3.467
25	1.316	1.708	2.060	2.485	3.450
26	1.315	1.706	2.056	2.479	3.435
27	1.314	1.703	2.052	2.473	3.421
28	1.313	1.701	2.048	2.467	3.408
29	1.311	1.699	2.045	2.462	3.396
30	1.310	1.697	2.042	2.457	3.385
31	1.309	1.696	2.040	2.453	3.375
32	1.309	1.694	2.037	2.449	3.365
33	1.308	1.692	2.035	2.445	3.356
34	1.307	1.691	2.032	2.441	3.348
35	1.306	1.690	2.030	2.438	3.340
36	1.306	1.688	2.028	2.434	3.333
37	1.305	1.687	2.026	2.431	3.326
38	1.304	1.686	2.024	2.429	3.319
39	1.304	1.685	2.023	2.426	3.313
40	1.303	1.684	2.021	2.423	3.307
41	1.303	1.683	2.020	2.421	3.301
42	1.302	1.682	2.018	2.418	3.296
43	1.302	1.681	2.017	2.416	3.291
44	1.301	1.680	2.015	2.414	3.286
45	1.301	1.679	2.014	2.412	3.281
46	1.300	1.679	2.013	2.410	3.277
47	1.300	1.678	2.012	2.408	3.273
48	1.299	1.677	2.011	2.407	3.269
49	1.299	1.677	2.010	2.405	3.265

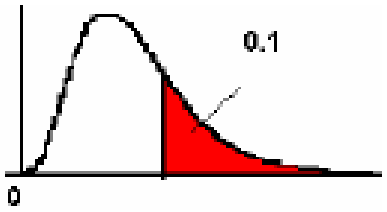
ν	0.10	0.05	0.025	0.01	0.001
50	1.299	1.676	2.009	2.403	3.261
51	1.298	1.675	2.008	2.402	3.258
52	1.298	1.675	2.007	2.400	3.255
53	1.298	1.674	2.006	2.399	3.251
54	1.297	1.674	2.005	2.397	3.248
55	1.297	1.673	2.004	2.396	3.245
56	1.297	1.673	2.003	2.395	3.242
57	1.297	1.672	2.002	2.394	3.239
58	1.296	1.672	2.002	2.392	3.237
59	1.296	1.671	2.001	2.391	3.234
60	1.296	1.671	2.000	2.390	3.232
61	1.296	1.670	2.000	2.389	3.229
62	1.295	1.670	1.999	2.388	3.227
63	1.295	1.669	1.998	2.387	3.225
64	1.295	1.669	1.998	2.386	3.223
65	1.295	1.669	1.997	2.385	3.220
66	1.295	1.668	1.997	2.384	3.218
67	1.294	1.668	1.996	2.383	3.216
68	1.294	1.668	1.995	2.382	3.214
69	1.294	1.667	1.995	2.382	3.213
70	1.294	1.667	1.994	2.381	3.211
71	1.294	1.667	1.994	2.380	3.209
72	1.293	1.666	1.993	2.379	3.207
73	1.293	1.666	1.993	2.379	3.206
74	1.293	1.666	1.993	2.378	3.204
75	1.293	1.665	1.992	2.377	3.202
76	1.293	1.665	1.992	2.376	3.201
77	1.293	1.665	1.991	2.376	3.199
78	1.292	1.665	1.991	2.375	3.198
79	1.292	1.664	1.990	2.374	3.197
80	1.292	1.664	1.990	2.374	3.195
81	1.292	1.664	1.990	2.373	3.194
82	1.292	1.664	1.989	2.373	3.193
83	1.292	1.663	1.989	2.372	3.191
84	1.292	1.663	1.989	2.372	3.190
85	1.292	1.663	1.988	2.371	3.189
86	1.291	1.663	1.988	2.370	3.188
87	1.291	1.663	1.988	2.370	3.187
88	1.291	1.662	1.987	2.369	3.185
89	1.291	1.662	1.987	2.369	3.184
90	1.291	1.662	1.987	2.368	3.183
91	1.291	1.662	1.986	2.368	3.182
92	1.291	1.662	1.986	2.368	3.181
93	1.291	1.661	1.986	2.367	3.180
94	1.291	1.661	1.986	2.367	3.179
95	1.291	1.661	1.985	2.366	3.178
96	1.290	1.661	1.985	2.366	3.177
97	1.290	1.661	1.985	2.365	3.176
98	1.290	1.661	1.984	2.365	3.175
99	1.290	1.660	1.984	2.365	3.175
100	1.290	1.660	1.984	2.364	3.174
∞	1.282	1.645	1.960	2.326	3.090

A6 Área à direita para a distribuição do qui-quadrado



gdl\area	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.14329
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.07050	12.83250
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.64464	12.59159	14.44938
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.01704	14.06714	16.01276
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	13.36157	15.50731	17.53455
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	14.68366	16.91898	19.02277
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	15.98718	18.30704	20.48318
11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.34100	13.70069	17.27501	19.67514	21.92005
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.34032	14.84540	18.54935	21.02607	23.33666
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.33976	15.98391	19.81193	22.36203	24.73560
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.16531	13.33927	17.11693	21.06414	23.68479	26.11895
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.03654	14.33886	18.24509	22.30713	24.99579	27.48839
16	5.14221	5.81221	6.90766	7.96165	9.31224	11.91222	15.33850	19.36886	23.54183	26.29623	28.84535
17	5.69722	6.40776	7.56419	8.67176	10.08519	12.79193	16.33818	20.48868	24.76904	27.58711	30.19101
18	6.26480	7.01491	8.23075	9.39046	10.86494	13.67529	17.33790	21.60489	25.98942	28.86930	31.52638
19	6.84397	7.63273	8.90652	10.11701	11.65091	14.56200	18.33765	22.71781	27.20357	30.14353	32.85233
20	7.43384	8.26040	9.59078	10.85081	12.44261	15.45177	19.33743	23.82769	28.41198	31.41043	34.16961
21	8.03365	8.89720	10.28290	11.59131	13.23960	16.34438	20.33723	24.93478	29.61509	32.67057	35.47888
22	8.64272	9.54249	10.98232	12.33801	14.04149	17.23962	21.33704	26.03927	30.81328	33.92444	36.78071
23	9.26042	10.19572	11.68855	13.09051	14.84796	18.13730	22.33688	27.14134	32.00690	35.17246	38.07563
24	9.88623	10.85636	12.40115	13.84843	15.65868	19.03725	23.33673	28.24115	33.19624	36.41503	39.36408
25	10.51965	11.52398	13.11972	14.61141	16.47341	19.93934	24.33659	29.33885	34.38159	37.65248	40.64647
26	11.16024	12.19815	13.84390	15.37916	17.29188	20.84343	25.33646	30.43457	35.56317	38.88514	41.92317
27	11.80759	12.87850	14.57338	16.15140	18.11390	21.74940	26.33634	31.52841	36.74122	40.11327	43.19451
28	12.46134	13.56471	15.30786	16.92788	18.93924	22.65716	27.33623	32.62049	37.91592	41.33714	44.46079
29	13.12115	14.25645	16.04707	17.70837	19.76774	23.56659	28.33613	33.71091	39.08747	42.55697	45.72229
30	13.78672	14.95346	16.79077	18.49266	20.59923	24.47761	29.33603	34.79974	40.25602	43.77297	46.97924

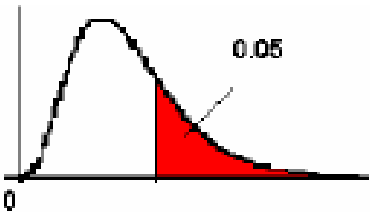
A7 Distribuição F de Fisher ($\alpha = 0,10$)



v2 \ v1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45	50	60	70	80	90	100	110	120	130
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71	61,07	61,35	61,57	61,74	62,05	62,26	62,42	62,53	62,62	62,69	62,79	62,87	62,93	62,97	63,01	63,04	63,06	63,08
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,43	9,44	9,44	9,45	9,46	9,46	9,47	9,47	9,47	9,47	9,48	9,48	9,48	9,48	9,48	9,48	9,48
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,20	5,19	5,18	5,17	5,17	5,16	5,16	5,16	5,15	5,15	5,15	5,15	5,15	5,14	5,14	5,14	5,14
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,88	3,86	3,85	3,84	3,83	3,82	3,81	3,80	3,80	3,80	3,79	3,79	3,78	3,78	3,78	3,78	3,78	3,77
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,25	3,23	3,22	3,21	3,19	3,17	3,16	3,16	3,15	3,15	3,14	3,14	3,13	3,13	3,13	3,12	3,12	3,12
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,88	2,86	2,85	2,84	2,81	2,80	2,79	2,78	2,77	2,77	2,76	2,76	2,75	2,75	2,75	2,74	2,74	2,74
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,64	2,62	2,61	2,59	2,57	2,56	2,54	2,54	2,53	2,52	2,51	2,51	2,50	2,50	2,50	2,49	2,49	2,49
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,48	2,45	2,44	2,42	2,40	2,38	2,37	2,36	2,35	2,35	2,34	2,33	2,33	2,32	2,32	2,32	2,32	2,31
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,35	2,33	2,31	2,30	2,27	2,25	2,24	2,23	2,22	2,22	2,21	2,20	2,20	2,19	2,19	2,19	2,18	2,18
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,26	2,23	2,22	2,20	2,17	2,16	2,14	2,13	2,12	2,12	2,11	2,10	2,09	2,09	2,09	2,08	2,08	2,08
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,18	2,16	2,14	2,12	2,10	2,08	2,06	2,05	2,04	2,04	2,03	2,02	2,01	2,01	2,01	2,00	2,00	2,00
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,12	2,09	2,08	2,06	2,03	2,01	2,00	1,99	1,98	1,97	1,96	1,95	1,95	1,94	1,94	1,93	1,93	1,93
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,07	2,04	2,02	2,01	1,98	1,96	1,94	1,93	1,92	1,92	1,90	1,90	1,89	1,89	1,88	1,88	1,88	1,87
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,02	2,00	1,98	1,96	1,93	1,91	1,90	1,89	1,88	1,87	1,86	1,85	1,84	1,84	1,83	1,83	1,83	1,83
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,99	1,96	1,94	1,92	1,89	1,87	1,86	1,85	1,84	1,83	1,82	1,81	1,80	1,80	1,79	1,79	1,79	1,78
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,95	1,93	1,91	1,89	1,86	1,84	1,82	1,81	1,80	1,79	1,78	1,77	1,77	1,76	1,76	1,75	1,75	1,75
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,93	1,90	1,88	1,86	1,83	1,81	1,79	1,78	1,77	1,76	1,75	1,74	1,74	1,73	1,73	1,72	1,72	1,72
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,90	1,87	1,85	1,84	1,80	1,78	1,77	1,75	1,74	1,74	1,72	1,71	1,71	1,70	1,70	1,69	1,69	1,69
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,91	1,88	1,85	1,83	1,81	1,78	1,76	1,74	1,73	1,72	1,71	1,70	1,69	1,68	1,68	1,67	1,67	1,67	1,66
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,89	1,86	1,83	1,81	1,79	1,76	1,74	1,72	1,71	1,70	1,69	1,68	1,67	1,66	1,65	1,65	1,65	1,64	1,64
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,84	1,81	1,79	1,78	1,74	1,72	1,70	1,69	1,68	1,67	1,66	1,65	1,64	1,63	1,63	1,63	1,62	1,62
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,83	1,80	1,78	1,76	1,73	1,70	1,68	1,67	1,66	1,65	1,64	1,63	1,62	1,62	1,61	1,61	1,60	1,60
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,81	1,78	1,76	1,74	1,71	1,69	1,67	1,66	1,64	1,64	1,62	1,61	1,61	1,60	1,59	1,59	1,59	1,58
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,80	1,77	1,75	1,73	1,70	1,67	1,65	1,64	1,63	1,62	1,61	1,60	1,59	1,58	1,58	1,57	1,57	1,57
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,79	1,76	1,74	1,72	1,68	1,66	1,64	1,63	1,62	1,61	1,59	1,58	1,58	1,57	1,56	1,56	1,56	1,55
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,77	1,75	1,72	1,71	1,67	1,65	1,63	1,61	1,60	1,59	1,58	1,57	1,56	1,56	1,55	1,55	1,54	1,54
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,76	1,74	1,71	1,70	1,66	1,64	1,62	1,60	1,59	1,58	1,57	1,56	1,55	1,54	1,54	1,53	1,53	1,53
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,75	1,73	1,70	1,69	1,65	1,63	1,61	1,59	1,58	1,57	1,56	1,55	1,54	1,53	1,53	1,52	1,52	1,52
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,75	1,72	1,69	1,68	1,64	1,62	1,60	1,58	1,57	1,56	1,55	1,54	1,53	1,52	1,52	1,51	1,51	1,51
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,74	1,71	1,69	1,67	1,63	1,61	1,59	1,57	1,56	1,55	1,54	1,53	1,52	1,51	1,51	1,50	1,50	1,50
35	2,85	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82	1,79	1,74	1,70	1,67	1,65	1,63	1,60	1,57	1,55	1,53	1,52	1,51	1,50	1,49	1,48	1,47	1,47	1,46	1,46	1,45
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,68	1,65	1,62	1,61	1,57	1,54	1,52	1,51	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,43	1,42	1,42
45	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	1,74	1,70	1,66	1,63	1,60	1,58	1,55	1,52	1,50	1,48	1,47	1,46	1,44	1,43	1,42	1,41	1,41	1,40	1,40	1,40
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,68	1,64	1,61	1,59	1,57	1,53	1,50	1,48	1,46	1,45	1,44	1,42	1,41	1,40	1,39	1,39	1,38	1,38	1,38
55	2,80	2,40	2,19	2,05	1,95	1,88	1,83	1,78	1,75	1,72	1,67	1,63	1,60	1,58	1,55	1,52	1,49	1,47	1,45	1,44	1,43	1,41	1,40	1,39	1,38	1,37	1,37	1,36	1,36
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,62	1,59	1,56	1,54	1,50	1,48	1,45	1,44	1,42	1,41	1,40	1,38	1,37	1,36	1,36	1,35	1,35	1,34
65	2,78	2,39	2,17	2,03	1,94	1,87	1,81	1,77	1,73	1,70	1,65	1,61	1,58	1,55	1,53	1,49	1,47	1,44	1,43	1,41	1,40	1,38	1,37	1,36	1,35	1,35	1,34	1,34	1,33
70	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72	1,69	1,64	1,60	1,57	1,55	1,53	1,49	1,46	1,43	1,42	1,40	1,39	1,37	1,36	1,35	1,34	1,34	1,33	1,32	1,32
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68	1,63	1,59	1,56	1,53	1,51	1,47	1,44	1,42	1,40	1,39	1,38	1,36	1,34	1,33	1,33	1,32	1,31	1,31	1,30
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67	1,62	1,58	1,55	1,52	1,50	1,46	1,43	1,41	1,39	1,38	1,36	1,35	1,33	1,32	1,31	1,30	1,30	1,29	1,29
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66	1,61	1,57	1,54	1,52	1,49	1,45	1,42	1,40	1,38	1,37	1,35	1,34	1,32	1,31	1,30	1,29	1,29	1,28	1,28
110	2,75	2,35	2,13	2,00	1,90	1,83	1,77	1,73	1,69	1,66	1,61	1,57	1,54	1,51	1,49	1,45	1,42	1,39	1,37	1,36	1,35	1,33	1,31	1,30	1,29	1,28	1,28	1,27	1,27
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,56	1,53	1,50	1,48	1,44	1,41	1,39	1,37	1,35	1,34	1,32	1,31	1,29	1,28	1,28	1,27	1,26	1,26

Tabela que fornece valores f tais que $P(F\nu_1, \nu_2 > f) = 0,1$

A8 Distribuição F de Fisher ($\alpha = 0,05$)



v2 \ v1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45	50	60	70	80	90	100	110	120	130
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,90	245,36	246,47	247,32	248,02	249,26	250,10	250,69	251,14	251,49	251,77	252,20	252,50	252,72	252,90	253,04	253,16	253,25	253,33
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	19,46	19,46	19,47	19,47	19,47	19,48	19,48	19,48	19,48	19,48	19,49	19,49	19,49	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,69	8,67	8,66	8,63	8,62	8,60	8,59	8,59	8,58	8,57	8,57	8,56	8,56	8,55	8,55	8,55	8,55
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	5,84	5,82	5,80	5,77	5,75	5,73	5,72	5,71	5,70	5,69	5,68	5,67	5,67	5,66	5,66	5,66	5,66
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	4,60	4,58	4,56	4,52	4,50	4,48	4,46	4,45	4,44	4,43	4,42	4,41	4,41	4,41	4,40	4,40	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	3,92	3,90	3,87	3,83	3,81	3,79	3,77	3,76	3,75	3,74	3,73	3,72	3,72	3,71	3,71	3,70	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	3,49	3,47	3,44	3,40	3,38	3,36	3,34	3,33	3,32	3,30	3,29	3,29	3,28	3,27	3,27	3,27	3,26
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	3,20	3,17	3,15	3,11	3,08	3,06	3,04	3,03	3,02	3,01	2,99	2,99	2,98	2,97	2,97	2,97	2,96
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	2,99	2,96	2,94	2,89	2,86	2,84	2,83	2,81	2,80	2,79	2,78	2,77	2,76	2,76	2,75	2,75	2,74
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,83	2,80	2,77	2,73	2,70	2,68	2,66	2,65	2,64	2,62	2,61	2,60	2,59	2,59	2,58	2,58	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	2,70	2,67	2,65	2,60	2,57	2,55	2,53	2,52	2,51	2,49	2,48	2,47	2,46	2,46	2,45	2,45	2,44
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	2,60	2,57	2,54	2,50	2,47	2,44	2,43	2,41	2,40	2,38	2,37	2,36	2,36	2,35	2,34	2,34	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	2,51	2,48	2,46	2,41	2,38	2,36	2,34	2,33	2,31	2,30	2,28	2,27	2,27	2,26	2,26	2,25	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	2,44	2,41	2,39	2,34	2,31	2,28	2,27	2,25	2,24	2,22	2,21	2,20	2,19	2,19	2,18	2,18	2,17
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	2,38	2,35	2,33	2,28	2,25	2,22	2,20	2,19	2,18	2,16	2,15	2,14	2,13	2,12	2,12	2,11	2,11
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	2,33	2,30	2,28	2,23	2,19	2,17	2,15	2,14	2,12	2,11	2,09	2,08	2,07	2,07	2,06	2,06	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	2,33	2,29	2,26	2,23	2,18	2,15	2,12	2,10	2,09	2,08	2,06	2,05	2,03	2,03	2,02	2,02	2,01	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	2,25	2,22	2,19	2,14	2,11	2,08	2,06	2,05	2,04	2,02	2,00	1,99	1,98	1,98	1,97	1,97	1,96
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	2,21	2,18	2,16	2,11	2,07	2,05	2,03	2,01	2,00	1,98	1,97	1,96	1,95	1,94	1,93	1,93	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	2,18	2,15	2,12	2,07	2,04	2,01	1,99	1,98	1,97	1,95	1,93	1,92	1,91	1,91	1,90	1,90	1,89
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	2,16	2,12	2,10	2,05	2,01	1,98	1,96	1,95	1,94	1,92	1,90	1,89	1,88	1,88	1,87	1,87	1,86
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	2,13	2,10	2,07	2,02	1,98	1,96	1,94	1,92	1,91	1,89	1,88	1,86	1,86	1,85	1,84	1,84	1,83
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	2,11	2,08	2,05	2,00	1,96	1,93	1,91	1,90	1,88	1,86	1,85	1,84	1,83	1,82	1,82	1,81	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	2,09	2,05	2,03	1,97	1,94	1,91	1,89	1,88	1,86	1,84	1,83	1,82	1,81	1,80	1,79	1,79	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	2,07	2,04	2,01	1,96	1,92	1,89	1,87	1,86	1,84	1,82	1,81	1,80	1,79	1,78	1,77	1,77	1,76
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	2,05	2,02	1,99	1,94	1,90	1,87	1,85	1,84	1,82	1,80	1,79	1,78	1,77	1,76	1,75	1,75	1,74
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	2,04	2,00	1,97	1,92	1,88	1,86	1,84	1,82	1,81	1,79	1,77	1,76	1,75	1,74	1,74	1,73	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	2,02	1,99	1,96	1,91	1,87	1,84	1,82	1,80	1,79	1,77	1,75	1,74	1,73	1,73	1,72	1,71	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	2,01	1,97	1,94	1,89	1,85	1,83	1,81	1,79	1,77	1,75	1,74	1,73	1,72	1,71	1,70	1,70	1,69
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	1,99	1,96	1,93	1,88	1,84	1,81	1,79	1,77	1,76	1,74	1,72	1,71	1,70	1,70	1,69	1,68	1,68
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,04	1,99	1,94	1,91	1,88	1,82	1,79	1,76	1,74	1,72	1,70	1,68	1,66	1,65	1,64	1,63	1,63	1,62	1,62
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	1,90	1,87	1,84	1,78	1,74	1,72	1,69	1,67	1,66	1,64	1,62	1,61	1,60	1,59	1,58	1,58	1,57
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	1,97	1,92	1,87	1,84	1,81	1,75	1,71	1,68	1,66	1,64	1,63	1,60	1,59	1,57	1,56	1,55	1,55	1,54	1,54
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95	1,89	1,85	1,81	1,78	1,73	1,69	1,66	1,63	1,61	1,60	1,58	1,56	1,54	1,53	1,52	1,52	1,51	1,51
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,93	1,88	1,83	1,79	1,76	1,71	1,67	1,64	1,61	1,59	1,58	1,55	1,54	1,52	1,51	1,50	1,49	1,49	1,48
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	1,82	1,78	1,75	1,69	1,65	1,62	1,59	1,57	1,56	1,53	1,52	1,50	1,49	1,48	1,47	1,47	1,46
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,90	1,85	1,80	1,76	1,73	1,68	1,63	1,60	1,58	1,56	1,54	1,52	1,50	1,49	1,47	1,46	1,46	1,45	1,44
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,89	1,84	1,79	1,75	1,72	1,66	1,62	1,59	1,57	1,55	1,53	1,50	1,49	1,47	1,46	1,45	1,44	1,44	1,43
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,82	1,77	1,73	1,70	1,64	1,60	1,57	1,54	1,52	1,51	1,48	1,46	1,45	1,44	1,43	1,42	1,41	1,40
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,86	1,80	1,76	1,72	1,69	1,63	1,59	1,56	1,53	1,51	1,49	1,46	1,44	1,43	1,42	1,41	1,40	1,39	1,39
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,75	1,71	1,68	1,62	1,57	1,54	1,52	1,49	1,48	1,45	1,43	1,41	1,40	1,39	1,38	1,38	1,37
110	3,93	3,08	2,69	2,45	2,30	2,18	2,09	2,02	1,97	1,92	1,84	1,78	1,74	1,70	1,67	1,61	1,56	1,53	1,50	1,48	1,47	1,44	1,42	1,40	1,39	1,38	1,37	1,36	1,36
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,78	1,73	1,69	1,66	1,60	1,55	1,52	1,50	1,47	1,46	1,43	1,41	1,39	1,38	1,37	1,36	1,35	1,35

Tabela que fornece valores f tais que $P(F\nu_1,\nu_2 > f) = 0,05$

A9 Distribuição F de Fisher ($\alpha = 0,01$)

v1 / v2	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45	50	60	70	80	90	100	110	120	130
1,0	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5928,3	5981,0	6022,4	6055,9	6106,7	6143,0	6170,0	6191,4	6208,7	6239,9	6260,4	6275,3	6286,4	6295,7	6302,3	6313,0	6320,9	6326,5	6330,7	6333,9	6336,7	6339,5	6341,4
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,42	99,43	99,44	99,44	99,45	99,46	99,47	99,47	99,48	99,48	99,48	99,48	99,48	99,48	99,49	99,49	99,49	99,49	99,49
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,92	26,83	26,75	26,69	26,58	26,50	26,45	26,41	26,38	26,35	26,32	26,29	26,27	26,25	26,24	26,23	26,22	26,21
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,25	14,15	14,08	14,02	13,91	13,84	13,79	13,75	13,71	13,69	13,65	13,63	13,61	13,59	13,58	13,57	13,56	13,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,77	9,68	9,61	9,55	9,45	9,38	9,33	9,29	9,26	9,24	9,20	9,18	9,16	9,14	9,13	9,12	9,11	9,10
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,60	7,52	7,45	7,40	7,30	7,23	7,18	7,14	7,11	7,09	7,06	7,03	7,01	7,00	6,99	6,98	6,97	6,96
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,36	6,28	6,21	6,16	6,06	5,99	5,94	5,91	5,88	5,86	5,82	5,80	5,78	5,77	5,75	5,75	5,74	5,73
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,56	5,48	5,41	5,36	5,26	5,20	5,15	5,12	5,09	5,07	5,03	5,01	4,99	4,97	4,96	4,95	4,95	4,94
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	5,01	4,92	4,86	4,81	4,71	4,65	4,60	4,57	4,54	4,52	4,48	4,46	4,44	4,43	4,41	4,41	4,40	4,39
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,60	4,52	4,46	4,41	4,31	4,25	4,20	4,17	4,14	4,12	4,08	4,06	4,04	4,03	4,01	4,00	4,00	3,99
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,29	4,21	4,15	4,10	4,01	3,94	3,89	3,86	3,83	3,81	3,78	3,75	3,73	3,72	3,71	3,70	3,69	3,68
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,05	3,97	3,91	3,86	3,76	3,70	3,65	3,62	3,59	3,57	3,54	3,51	3,49	3,48	3,47	3,46	3,45	3,44
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,86	3,78	3,72	3,66	3,57	3,51	3,46	3,43	3,40	3,38	3,34	3,32	3,30	3,28	3,27	3,26	3,25	3,25
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,70	3,62	3,56	3,51	3,41	3,35	3,30	3,27	3,24	3,22	3,18	3,16	3,14	3,12	3,11	3,10	3,09	3,09
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,56	3,49	3,42	3,37	3,28	3,21	3,17	3,13	3,10	3,08	3,05	3,02	3,00	2,99	2,98	2,97	2,96	2,95
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,45	3,37	3,31	3,26	3,16	3,10	3,05	3,02	2,99	2,97	2,93	2,91	2,89	2,87	2,86	2,85	2,84	2,84
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,35	3,27	3,21	3,16	3,07	3,00	2,96	2,92	2,89	2,87	2,83	2,81	2,79	2,78	2,76	2,75	2,75	2,74
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,27	3,19	3,13	3,08	2,98	2,92	2,87	2,84	2,81	2,78	2,75	2,72	2,70	2,69	2,68	2,67	2,66	2,65
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,19	3,12	3,05	3,00	2,91	2,84	2,80	2,76	2,73	2,71	2,67	2,65	2,63	2,61	2,60	2,59	2,58	2,58
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,13	3,05	2,99	2,94	2,84	2,78	2,73	2,69	2,67	2,64	2,61	2,58	2,56	2,55	2,54	2,53	2,52	2,51
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,07	2,99	2,93	2,88	2,79	2,72	2,67	2,64	2,61	2,58	2,55	2,52	2,50	2,49	2,48	2,47	2,46	2,45
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	3,02	2,94	2,88	2,83	2,73	2,67	2,62	2,58	2,55	2,53	2,50	2,47	2,45	2,43	2,42	2,41	2,40	2,40
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,97	2,89	2,83	2,78	2,69	2,62	2,57	2,54	2,51	2,48	2,45	2,42	2,40	2,39	2,37	2,36	2,35	2,35
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,93	2,85	2,79	2,74	2,64	2,58	2,53	2,49	2,46	2,44	2,40	2,38	2,36	2,34	2,33	2,32	2,31	2,30
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,89	2,81	2,75	2,70	2,60	2,54	2,49	2,45	2,42	2,40	2,36	2,34	2,32	2,30	2,29	2,28	2,27	2,26
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,86	2,78	2,72	2,66	2,57	2,50	2,45	2,42	2,39	2,36	2,33	2,30	2,28	2,26	2,25	2,24	2,23	2,23
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,82	2,75	2,68	2,63	2,54	2,47	2,42	2,38	2,35	2,33	2,29	2,27	2,25	2,23	2,22	2,21	2,20	2,19
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,79	2,72	2,65	2,60	2,51	2,44	2,39	2,35	2,32	2,30	2,26	2,24	2,22	2,20	2,19	2,18	2,17	2,16
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,77	2,69	2,63	2,57	2,48	2,41	2,36	2,33	2,30	2,27	2,23	2,21	2,19	2,17	2,16	2,15	2,14	2,13
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,74	2,66	2,60	2,55	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,18	2,16	2,14	2,13	2,12	2,11	2,10
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,74	2,64	2,56	2,50	2,44	2,35	2,28	2,23	2,19	2,16	2,14	2,10	2,07	2,05	2,03	2,02	2,01	2,00	1,99
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,56	2,48	2,42	2,37	2,27	2,20	2,15	2,11	2,08	2,06	2,02	1,99	1,97	1,95	1,94	1,93	1,92	1,91
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,61	2,51	2,43	2,36	2,31	2,21	2,14	2,09	2,05	2,02	2,00	1,96	1,93	1,91	1,89	1,88	1,86	1,85	1,84
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,56	2,46	2,38	2,32	2,27	2,17	2,10	2,05	2,01	1,97	1,95	1,91	1,88	1,86	1,84	1,82	1,81	1,80	1,79
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,53	2,42	2,34	2,28	2,23	2,13	2,06	2,01	1,97	1,94	1,91	1,87	1,84	1,82	1,80	1,78	1,77	1,76	1,75
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,39	2,31	2,25	2,20	2,10	2,03	1,98	1,94	1,90	1,88	1,84	1,81	1,78	1,76	1,75	1,74	1,73	1,72
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	2,47	2,37	2,29	2,23	2,17	2,07	2,00	1,95	1,91	1,88	1,85	1,81	1,78	1,75	1,74	1,72	1,71	1,70	1,69
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,45	2,35	2,27	2,20	2,15	2,05	1,98	1,93	1,89	1,85	1,83	1,78	1,75	1,73	1,71	1,70	1,68	1,67	1,66
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,42	2,31	2,23	2,17	2,12	2,01	1,94	1,89	1,85	1,82	1,79	1,75	1,71	1,69	1,67	1,65	1,64	1,63	1,62
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,39	2,29	2,21	2,14	2,09	1,99	1,92	1,86	1,82	1,79	1,76	1,72	1,68	1,66	1,64	1,62	1,61	1,60	1,59
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,37	2,27	2,19	2,12	2,07	1,97	1,89	1,84	1,80	1,76	1,74	1,69	1,66	1,63	1,61	1,60	1,58	1,57	1,56
110	6,87	4,80	3,96	3,49	3,19	2,97	2,81	2,68	2,57	2,49	2,35	2,25	2,17	2,10	2,05	1,95	1,88	1,82	1,78	1,74	1,72	1,67	1,64	1,61	1,59	1,58	1,56	1,55	1,54
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,23	2,15	2,09	2,03	1,93	1,86	1,81	1,76	1,73	1,70	1,66	1,62	1,60	1,58	1,56	1,55	1,53	1,52

Tabela que fornece valores f tais que $P(F\nu_1, \nu_2 > f) = 0,01$