Pedro A. Morettin
Wilton de O. Bussab

ESTATÍSTICA BÁSICA

6ª edição
Revista e atualizada

Editora Saraiva

Capítulo 12

Testes de Hipóteses

12.1 Introdução

Vimos no Capítulo 10 que um dos problemas a serem resolvidos pela Inferência Estatística é o de testar uma hipótese. Isto é, feita determinada afirmação sobre uma população, usualmente sobre um parâmetro dessa, desejamos saber se os resultados experimentais provenientes de uma amostra contrariam ou não tal afirmação. Muitas vezes, essa afirmação sobre a população é derivada de teorias desenvolvidas no campo substantivo do conhecimento. A adequação ou não dessa teoria ao universo real pode ser verificada ou refutada pela amostra. O objetivo do teste estatístico de hipóteses é, então, fornecer uma metodologia que nos permita verificar se os dados amostrais trazem evidências que apóiem ou não uma hipótese (estatística) formulada.

Neste capítulo iremos introduzir o procedimento básico de teste de hipótese sobre um parâmetro de uma população. A idéia central desse procedimento é a de supor verdadeira a hipótese em questão e verificar se a amostra observada é "verossímil" nessas condições. No capítulo seguinte daremos alguns testes para comparação de parâmetros de duas populações.

12.2 Um Exemplo

Vamos introduzir a idéia de teste de uma hipótese por meio de um exemplo hipotético que, partindo de uma situação simples, será gradualmente ampliado para atender à situação geral do teste de hipóteses.

Exemplo 12.1. Uma indústria usa, como um dos componentes das máquinas que produz, um parafuso importado, que deve satisfazer a algumas exigências. Uma dessas é a resistência à tração. Esses parafusos são fabricados por alguns países, e as especificações técnicas variam de país para país. Por exemplo, o catálogo do país A afirma que a resistência média à tração de seus parafusos é de 145 kg, com desvio padrão de 12 kg. Já para o país B, a média é de 155 kg e desvio padrão 20 kg.

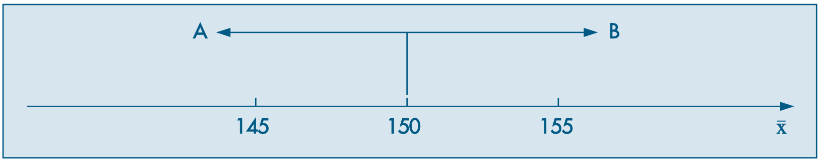
Um lote desses parafusos, de origem desconhecida, será leiloado a um preço muito convidativo. Para que a indústria saiba se faz ou não uma oferta, ela necessita saber qual país produziu tais parafusos. O edital do leiloeiro afirma que, pouco antes do leilão, será divulgada a resistência média $\bar{x}$ de uma amostra de 25 parafusos do lote. Qual regra de decisão deve ser usada pela indústria para dizer se os parafusos são do país A ou B?

Uma resposta que ocorre imediatamente é a que considera como país produtor aquele para o qual a média da amostra mais se aproximar da média da população. Assim, uma possível regra de decisão seria:

Se $\bar{x} \leqslant 150$ (o ponto médio entre 145 e 155), diremos que os parafusos são do país A; caso contrário, isto é, $\bar{x} > 150$, são do país B.

Na Figura 12.1 ilustramos essa regra de decisão.

Figura 12.1: Regra de decisão para o Exemplo 12.1.



Suponha que, no dia do leilão, fôssemos informados de que $\bar{x} = 148$; de acordo com nossa regra de decisão, diríamos que os parafusos são de origem A. Podemos estar enganados nessa conclusão? Ou, em outras palavras, é possível que uma amostra de 25 parafusos de origem B apresente média $\bar{x} = 148$? Sim, é possível. Então, para melhor entendermos a regra de decisão adotada, é interessante estudarmos os tipos de erros que podemos cometer e as respectivas probabilidades.

Podemos cometer dois tipos de erros, e vamos numerá-los para facilitar a linguagem:

Erro de tipo I: dizer que os parafusos são de A quando na realidade são de B. Isso ocorre quando uma amostra de 25 parafusos de B apresenta média $\bar{x}$ inferior ou igual a 150 kg.

Erro de tipo II: dizer que os parafusos são de B, quando na realidade eles são de A. Isso ocorre quando uma amostra de 25 parafusos de A apresenta média $\bar{x}$ superior a 150 kg.

Para facilitar ainda mais, vamos definir duas hipóteses também numeradas:

H_0: os parafusos são de origem B. Isso equivale a dizer que a resistência X de cada parafuso segue uma distribuição com média $\mu = 155$ e desvio padrão $\sigma = 20$.

H_1: os parafusos são de A, isto é, a média $\mu = 145$ e o desvio padrão $\sigma = 12$.

Finalmente, vamos indicar por RC a região correspondente aos valores menores que 150, ou seja,

$$RC = \{y \in I\!R \mid y \leqslant 150\}.$$

Com as notações indicadas acima, a probabilidade de se cometer cada um dos erros pode ser escrita:

$$P(\text{erro I}) = P(\bar{X} \in RC \mid H_0 \text{ é verdadeira}) = \alpha$$

e

$$P(\text{erro II}) = P(\bar{X} \notin RC \,|\, H_1 \text{ é verdadeira}) = \beta.$$

Quando H_0 for verdadeira, isto é, os parafusos forem de B, sabemos do TLC que $\bar{X}$ terá distribuição aproximadamente normal, com média 155 e desvio padrão igual a $20/\sqrt{25} = 4$, isto é,

$$\bar{X} \sim N(155{,}16).$$

Denotando por Z a v.a. com distribuição $N(0,1)$, temos

$$\begin{aligned} P(\text{erro I}) &= P(\bar{X} \in RC \,|\, H_0 \text{ é verdadeira}) \\ &= P(\bar{X} \leqslant 150 \,|\, \bar{X} \sim N(155{,}16)) \\ &= P\left(Z \leqslant \frac{150-155}{4}\right) \\ &= P(Z \leqslant -1{,}25) = 0{,}10565 = 10{,}56\% = \alpha. \end{aligned}$$

De modo análogo, quando H_1 for a alternativa verdadeira, teremos que a v.a. $\bar{X}$ é tal que, aproximadamente,

$$\bar{X} \sim N(145;\ 5{,}76).$$

Teremos, então,

$$\begin{aligned} P(\text{erro II}) &= P(\bar{X} \notin RC \,|\, H_1 \text{ é verdadeira}) \\ &= P(\bar{X} > 150 \,|\, \bar{X} \sim N(145;\ 5{,}76)) \end{aligned}$$

$$= P\left(Z > \frac{150-145}{2{,}4}\right) = P(Z > 2{,}08) = 0{,}01876 = 1{,}88\% = \beta.$$

Observando esses dois resultados, notamos que, com a regra de decisão adotada, estaremos cometendo o erro de tipo I com maior probabilidade do que o erro de tipo II. De certo modo, essa regra de decisão privilegia a afirmação de que os parafusos são de A. No Quadro 12.1 ilustramos as conseqüências que podem advir da regra de decisão adotada.

Quadro 12.1: Resumo do teste H_0: $\mu = 155$, H_1: $\mu = 145$, com $RC =]-\infty, 150]$.

<table>
<tr><td rowspan="3">Origem Real
dos
Parafusos</td><td colspan="2">Decisão</td></tr>
<tr><td colspan="2">RC 150 $\bar{X}$</td></tr>
<tr><td>A</td><td>B</td></tr>
<tr><td>A</td><td>Sem erro</td><td>Erro tipo II
$\beta = 1{,}88\%$</td></tr>
<tr><td>B</td><td>Erro tipo I
$\alpha = 10{,}56\%$</td><td>Sem erro</td></tr>
</table>

Desse quadro, podemos notar que, se os parafusos forem realmente de B (segunda linha) e a amostra tiver média superior a 150 (segunda coluna), diremos que são de B, e não cometeremos erro algum. Por outro lado, se a média $\bar{x}$ for inferior a 150 (primeira coluna), devemos dizer que são de A, e estaremos cometendo um erro cuja probabilidade nesse caso é de 10,56%. De modo análogo, teremos uma interpretação para o caso de os parafusos serem realmente de A (primeira linha).

Para cada regra de decisão adotada, isto é, se escolhermos um valor $\bar{x}_c$ em vez de 150 no Quadro 12.1, apenas as probabilidades α e β mudarão. Se $\bar{x}_c$ for escolhido menor que 150, notamos que α diminuirá e β aumentará. Logo, deve existir um ponto em que α seja igual a β, ou seja, uma regra de decisão em que a probabilidade de errar contra A seja a mesma que errar contra B. Mostre que esse ponto é $\bar{x}_c$ = 148,75, e nesse caso $\alpha = \beta$ = 5,94%.

Do exposto acima constatamos que, escolhido um valor de $\bar{x}_c$, podemos achar as probabilidades α e β de cometer cada tipo de erro. Mas também podemos proceder de modo inverso: fixar um dos erros, digamos α, e encontrar a regra de decisão que irá corresponder à probabilidade de erro de tipo I igual a α.

Por exemplo, fixemos α em 5%, e vejamos qual a regra de decisão correspondente. Temos

$$\begin{aligned} 5\% = P(\text{erro I}) &= P(\bar{X} \leqslant \bar{x}_c | \bar{X} \sim N(155{,}16)) \\ &= P(Z \leqslant -1{,}645), \end{aligned}$$

mas da transformação para a normal padrão sabemos que

$$-1{,}645 = \frac{\bar{x}_c - 155}{4},$$

ou seja, $\bar{x}_c$ = 148,42. Então, a regra de decisão será:

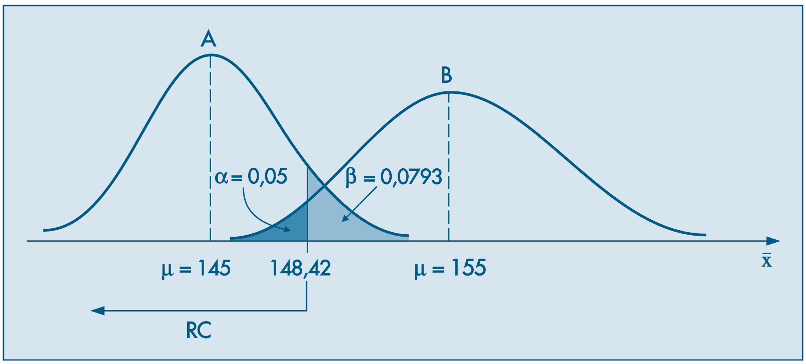
Se $\bar{x}$ for inferior a 148,42, dizemos que o lote é de A; caso contrário, dizemos que é de B.

Com essa regra, a probabilidade do erro de tipo II será

$$\begin{aligned} \beta = P(\text{erro II}) &= P(\bar{X} > 148{,}42 | \bar{X} \sim (145;\ 5{,}76)) \\ &= P(Z > 1{,}425) = 7{,}93\%. \end{aligned}$$

Veja a ilustração na Figura 12.2.

Figura 12.2: Ilustração dos erros de tipo I e II para o Exemplo 12.1.



Esse segundo tipo de procedimento é bastante utilizado, porque usualmente a decisão que devemos tomar não é apenas entre duas possíveis populações. Os parafusos poderiam ser produzidos por outros países além daqueles citados e, portanto, com outras características quanto à resistência média. Suponha, ainda, que interessa à indústria fazer uma proposta apenas no caso de o parafuso ser de origem B. Qual a regra de decisão que deve adotar?

A hipótese que nos interessa agora é:

H_0: os parafusos são de origem B (μ = 155 e σ = 20).

Caso essa não seja a hipótese verdadeira, a alternativa é muito mais ampla e pode ser expressa como:

H_1: os parafusos não são de origem B (μ e σ desconhecidos).

Aqui não podemos especificar os parâmetros sob a hipótese alternativa H_1, pois se não forem de origem B, os parafusos podem ser de vários outros países, cada um com suas próprias especificações. Alguns países podem ter técnicas mais sofisticadas de produção e, portanto, produzir com resistência média superior a 155. Outros, como no exemplo dado, com resistência menor. A especificação da hipótese alternativa depende muito do grau de informação que se tem do problema. Por exemplo, vamos admitir que a indústria do país B para esse caso seja a mais desenvolvida, e nenhum outro país possa produzir uma resistência média superior à dela. Então, nossa hipótese alternativa seria mais explícita:

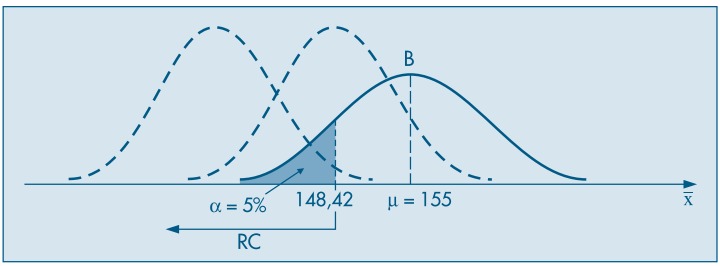
H_1: os parafusos não são de origem B ($\mu < 155$ e σ qualquer).

Isso significa que só iremos desconfiar de H_0 se $\bar{x}$ for muito menor do que 155. Ou seja, a nossa regra de decisão deverá ser semelhante à vista anteriormente. Como os parâmetros sob a hipótese alternativa são muitos, a melhor solução para construir a regra de decisão é fixar α, a probabilidade do erro de tipo I (rejeitar H_0 quando ela for verdadeira). Se fixarmos novamente α = 0,5, e nesse caso a regra de decisão depende apenas das informações de H_0, a regra de decisão será a mesma anterior:

Se $\bar{x}$ for superior a 148,42, diremos que o lote é de origem B; caso contrário, diremos que não é de origem B.

Com essa regra de decisão e com a hipótese alternativa mais ampla, não podemos encontrar β, pois não temos um único parâmetro μ como alternativa e nada sabemos sobre σ. Então, não podemos controlar o erro de tipo II. As implicações dessa regra de decisão estão resumidas na Figura 12.3 e no Quadro 12.2.

Figura 12.3: Teste $H_0 : \mu = 155$ vs $H_1 : \mu < 155$, com RC = $]-\infty; 148{,}42]$.



Quadro 12.2: Resumo do teste H_0: $\mu = 155$, H_1: $\mu < 155$, com RC = $]-\infty, 148{,}42]$.

Origem Real dos Parafusos	Decisão	
	RC — 148,42 — $\bar{x}$	
	não B	B
B	Erro tipo I, α = 5%	Sem erro
não B	Sem erro	Erro tipo II, β = ?

Podemos reescrever as hipóteses nessa situação da seguinte maneira:

$$H_0 : \mu = 155$$
$$H_1 : \mu < 155$$

O cálculo de β depende do valor de μ, que não é especificado. Mas podemos considerar a seguinte e importante função.

Definição. A função *característica de operação* (função CO) do teste acima é definida como

$$\beta(\mu) = P(\text{aceitar } H_0 | \mu) = P(\bar{X} > 148{,}42 | \mu).$$

Ou seja, $\beta(\mu)$ é a probabilidade de aceitar H_0, considerada como uma função de μ.

Usualmente, considera-se a função $\pi(\mu) = 1 - \beta(\mu)$, que é a probabilidade de se rejeitar H_0, como função de μ. Essa função é chamada *função poder do teste* e será estudada abaixo com certo detalhe. Nesses casos consideramos que σ é o mesmo para todos os valores de μ.

Admitamos, agora, que não exista razão alguma para acreditarmos que a resistência média dos parafusos de B seja maior ou menor do que a de outros países. Isso irá nos levar a duvidar que os parafusos não são de B, se a média observada for muito maior ou muito menor do que 155. Esta situação corresponde à seguinte hipótese alternativa:

H_1: os parafusos não são de origem B ($\mu \neq 155$).

Aqui, a regra de decisão deverá indicar dois pontos $\bar{x}_{c_1}$ e $\bar{x}_{c_2}$, tais que:

Se $\bar{x}$ estiver entre $\bar{x}_{c_1}$ e $\bar{x}_{c_2}$, diremos que os parafusos são de origem B; se $\bar{x}$ estiver fora do intervalo, diremos que não são de origem B.

Fixado α, a probabilidade do erro I, existirão muitos valores que satisfazem a essa condição. Daremos preferência àquelas soluções $\bar{x}_{c_1}$ e $\bar{x}_{c_2}$, simétricas em relação à média. Veja a Figura 12.4.

Voltando ao nosso problema, e fixado α em 5%, temos

$$0{,}05 = P(\text{erro I}) = P(\bar{X} < \bar{x}_{c_1} \text{ ou } \bar{X} > \bar{x}_{c_2} | \bar{X} \sim N(155{,}16))$$
$$= P(Z < -1{,}96 \text{ ou } Z > 1{,}96),$$

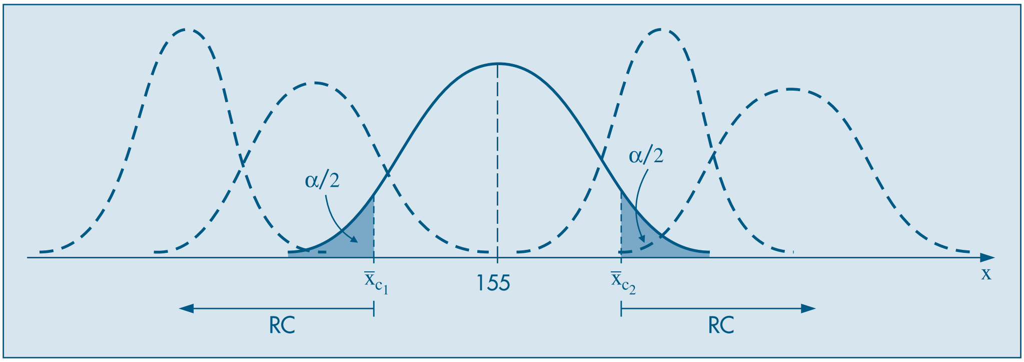
e daqui encontramos

$$-1{,}96 = (\bar{x}_{c_1} - 155)/4 \;\Rightarrow \bar{x}_{c_1} = 147{,}16$$

e

$$1{,}96 = (\bar{x}_{c_2} - 155)/4 \;\Rightarrow \bar{x}_{c_2} = 162{,}84.$$

Figura 12.4: Teste $H_0 : \mu = 155$ vs $H_1 : \mu \neq 155$.



Portanto, nesse caso, a região de rejeição da hipótese H_0 é (veja o Quadro 12.3)

$$RC = \{\bar{x} \in \mathbb{R} \,|\, \bar{x} < 147{,}16 \text{ ou } \bar{x} > 162{,}84\}.$$

Do apresentado nesta seção, vemos que, dependendo do grau de informação que se tem do problema, podemos ter regras de decisão unilaterais ou bilaterais. Na seção seguinte iremos dar os passos para a construção de um teste de hipótese.

Quadro 12.3: Resumo do teste H_0: $\mu = 155$, H_1: $\mu \neq 155$, com $RC =]-\infty, 147{,}16] \cup [162{,}84, +\infty[$.

Origem Real dos Parafusos	Decisão	
	RC 147,16	162,84 RC $\bar{x}$
	B	não B
B	Sem erro	Erro tipo II, β = ?
não B	Erro tipo I, α = 5%	Sem erro

Problemas

1. Para decidirmos se os habitantes de uma ilha são descendentes da civilização A ou B, iremos proceder do seguinte modo:

   (i) selecionamos uma amostra de 100 moradores adultos da ilha, e determinamos a altura média deles;

   (ii) se essa altura média for superior a 176, diremos que são descendentes de B; caso contrário, são descendentes de A.

   Os parâmetros das alturas das duas civilizações são:

   A : $\mu = 175$ e $\sigma = 10$;

   B : $\mu = 177$ e $\sigma = 10$.

   Definamos: Erro de tipo I — dizer que os habitantes da ilha são descendentes de B quando, na realidade, são de A.

   Erro de tipo II — dizer que são de A quando, na realidade, são de B.

   (a) Qual a probabilidade do erro de tipo I? E do erro de tipo II?

(b) Qual deve ser a regra de decisão se quisermos fixar a probabilidade do erro de tipo I em 5%? Qual a probabilidade do erro de tipo II, nesse caso?

(c) Se $\sigma_A = 5$, como ficariam as respostas de (b)?

(d) Quais as probabilidades do erro de tipo II, nas condições da questão (b), se a média $\mu_B = 178$? E $\mu_B = 180$? E $\mu_B = 181$? Coloque num gráfico os pares (μ_B, $P(\text{erro II} \mid \mu_B)$).

2. Fazendo o teste

$$H_0 : \mu = 1.150 \ (\sigma = 150) \text{ contra } H_1 : \mu = 1.200 \ (\sigma = 200),$$

e $n = 100$, estabeleceu-se a seguinte região crítica:

$$RC = [1.170, +\infty[.$$

(a) Qual a probabilidade α de rejeitar H_0 quando verdadeira?

(b) Qual a probabilidade β de aceitar H_0 quando H_1 é verdadeira?

(c) Qual deve ser a região crítica para que $\alpha = \beta$?

3. Nas situações abaixo, escolha como hipótese nula, H_0, aquela que para você leva a um erro de tipo I mais importante. Descreva quais os dois erros em cada caso.

(a) O trabalho de um operador de radar é detectar aeronaves inimigas. Quando surge alguma coisa estranha na tela, ele deve decidir entre as hipóteses:

1. está começando um ataque;
2. tudo bem, apenas uma leve interferência.

(b) Num júri, um indivíduo está sendo julgado por um crime. As hipóteses sujeitas ao júri são:

1. o acusado é inocente;
2. o acusado é culpado.

(c) Um pesquisador acredita que descobriu uma vacina contra resfriado. Ele irá conduzir uma pesquisa de laboratório para verificar a veracidade da afirmação. De acordo com o resultado, ele lançará ou não a vacina no mercado. As hipóteses que pode testar são:

1. a vacina é eficaz;
2. a vacina não é eficaz.

4. Se, ao lançarmos três vezes uma moeda, aparecerem 3 coroas, decidimos rejeitar a hipótese de que a moeda é "honesta". Quais as probabilidades de erro de tipo I e erro de tipo II, se $p = 2/3$?

5. A variável X, custo de manutenção de um tear, pode ser considerada como tendo distribuição normal de média μ e desvio padrão 20 unidades. Os valores possíveis de μ podem ser 200 ou 210. Para verificar qual dos dois valores é o mais provável, usar-se-á uma amostra de 25 teares. Defina:

(a) Uma hipótese a ser testada.

(b) Uma regra de decisão e encontre as probabilidades dos erros de tipo I e II.

12.3 Procedimento Geral do Teste de Hipóteses

A construção de um teste de hipóteses, para um parâmetro populacional, pode ser colocada do seguinte modo. Existe uma variável X associada a dada população e tem-se uma hipótese sobre determinado parâmetro θ dessa população. Por exemplo, afirmamos que o verdadeiro valor de θ é θ_0. Colhe-se uma amostra aleatória de elementos dessa população, e com ela deseja-se comprovar ou não tal hipótese.

Como já vimos anteriormente, iniciamos nossa análise explicitando claramente qual a hipótese que estamos colocando à prova e a chamamos de *hipótese nula*, e escrevemos

$$H_0 : \theta = \theta_0.$$

Em seguida, convém explicitar também a hipótese que será considerada aceitável, caso H_0 seja rejeitada. A essa hipótese chamamos de *hipótese alternativa*, e a sua caracterização estatística irá depender do grau de conhecimento que se tem do problema estudado. A alternativa mais geral seria

$$H_1 : \theta \neq \theta_0.$$

Poderíamos, ainda, ter alternativas da forma

$$H_1 : \theta < \theta_0 \quad \text{ou} \quad H_1 : \theta > \theta_0,$$

dependendo das informações que o problema traz.

Qualquer que seja a decisão tomada, vimos que estamos sujeitos a cometer erros. Para facilitar a linguagem, introduzimos as definições:

Erro de tipo I: rejeitar a hipótese nula quando essa é verdadeira. Chamamos de α a probabilidade de cometer esse erro, isto é,

$$\alpha = P(\text{erro do tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira}).$$

Erro de tipo II: não rejeitar H_0 quando H_0 é falsa. A probabilidade de cometer esse erro é denotada por β, logo

$$\beta = P(\text{erro do tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ é falsa}).$$

O objetivo do teste de hipóteses é dizer, usando uma estatística $\hat{\theta}$, se a hipótese H_0 é ou não aceitável. Operacionalmente, essa decisão é tomada através da consideração de uma região crítica RC. Caso o valor observado da estatística pertença a essa região, rejeitamos H_0; caso contrário, não rejeitamos H_0. Esta região é construída de modo que $P(\hat{\theta} \in RC | H_0$ é verdadeira) seja igual a α, fixado *a priori*. RC recebe o nome de *região crítica ou região de rejeição* do teste. Um fato importante a ressaltar é que a região crítica é sempre construída sob a hipótese de H_0 ser verdadeira. A determinação do valor de β já é mais difícil, pois usualmente não especificamos valores fixos para o parâmetro sob a hipótese alternativa. Mais adiante trataremos dessa situação, ao considerarmos o poder de um teste.

A probabilidade α de se cometer um erro de tipo I (ou de primeira espécie) é um valor arbitrário e recebe o nome de *nível de significância* do teste. O resultado da amostra é tanto mais significante para rejeitar H_0 quanto menor for esse nível α. Ou seja, quanto menor for α, menor é a probabilidade de se obter uma amostra com estatística pertencente à região crítica, sendo pouco verossímil a obtenção de uma amostra da população para a qual H_0 seja verdadeira. Usualmente, o valor de α é fixado em 5%, 1% ou 0,1%.

A fixação do valor de α envolve uma questionável arbitrariedade. Neste sentido há um modo alternativo de se proceder, que será considerado na seção 12.8.

12.4 Passos para a Construção de um Teste de Hipóteses

Vimos nas seções anteriores o procedimento que se deve usar para realizar um teste de hipóteses. Daremos abaixo uma seqüência que pode ser usada sistematicamente para qualquer teste de hipóteses.

Passo 1. Fixe qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1.

Passo 2. Use a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para testar a hipótese H_0. Obter as propriedades dessa estatística (distribuição, média, desvio padrão).

Passo 3. Fixe a probabilidade α de cometer o erro de tipo I e use este valor para construir a região crítica (regra de decisão). Lembre que essa região é construída para a estatística definida no passo 2, usando os valores do parâmetro hipotetizados por H_0.

Passo 4. Use as observações da amostra para calcular o valor da estatística do teste.

Passo 5. Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite H_0; caso contrário, rejeite H_0.

Procuraremos, sempre que fizermos teste de hipóteses, distinguir bem esses cinco passos. Finalmente um comentário sobre H_0 e o erro de tipo I. Devemos tomar como H_0 aquela hipótese, que, rejeitada, conduza a um erro de tipo I mais importante de evitar. Vejamos um exemplo devido a Neyman (1978). Suponha um experimento para se determinar se um produto A é ou não cancerígeno. Após realizado o teste, podemos concluir: (i) A é cancerígeno ou (ii) A não é cancerígeno. Cada uma dessas conclusões pode estar errada e temos os dois tipos de erro já mencionados, dependendo de qual hipótese seja H_0. Do ponto de vista do usuário do produto, a hipótese a ser testada deve ser

$$H_0 : \text{A é cancerígeno},$$

pois a probabilidade de erro na rejeição dessa hipótese, se ela for verdadeira, deve ser um valor muito pequeno. Outros exemplos estão contidos no Problema 3.

12.5 Testes sobre a Média de uma População com Variância Conhecida

Vejamos, agora, uma aplicação dos cinco passos definidos na seção anterior, para testar a hipótese de que a média de uma população μ seja igual a um número fixado μ_0, supondo-se a variância σ^2 dessa população conhecida.

Exemplo 12.2. Uma máquina automática para encher pacotes de café enche-os segundo uma distribuição normal, com média μ e variância sempre igual a 400 g^2. A máquina foi regulada para μ = 500 g. Desejamos, periodicamente, colher uma amostra de 16 pacotes e verificar se a produção está sob controle, isto é, se $\mu = 500$ g ou não. Se uma dessas amostras apresentasse uma média $\bar{x} = 492$ g, você pararia ou não a produção para regular a máquina?

Vejamos como testar essa hipótese.

Passo 1. Indiquemos por X o peso de cada pacote; então, $X \sim N(\mu, 400)$. E as hipóteses que nos interessam são:

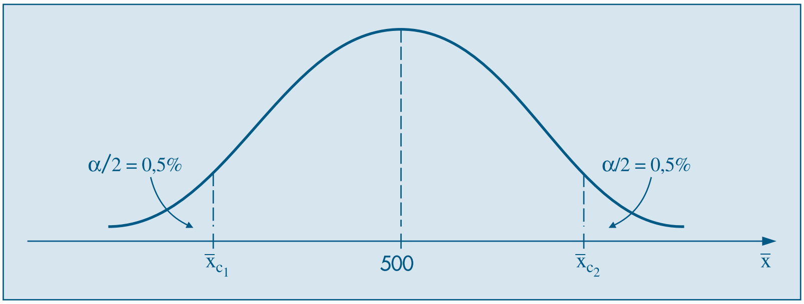
$$H_0 : \mu = 500 \text{ g},$$
$$H_1 : \mu \neq 500 \text{ g},$$

pois a máquina pode desregular para mais ou para menos.

Passo 2. Pela afirmação do problema, $\sigma^2 = 400$ será sempre a mesma; logo, para todo μ, a média $\bar{X}$ de 16 pacotes terá distribuição $N(\mu, 400/16)$, de modo que o desvio padrão (ou erro padrão) de $\bar{X}$ é $\sigma_{\bar{x}} = 5$. Em particular, se H_0 for verdadeira, $\bar{X} \sim N(500{,}25)$.

Passo 3. Vamos fixar $\alpha = 1\%$; pela hipótese alternativa, vemos que H_0 deve ser rejeitada quando $\bar{X}$ for muito pequena ou muito grande (dizemos que temos um teste bilateral). Portanto, nossa região crítica será como a da Figura 12.5.

Figura 12.5: Região crítica para o teste $H_0 : \mu = 500$ vs $H_1 : \mu \neq 500$ do Exemplo 12.2.



Da tabela da curva normal padronizada obtemos que

$$z_1 = -2{,}58 = (\bar{x}_{c_1} - 500)/5 \Rightarrow \bar{x}_{c_1} = 487{,}1,$$
$$z_2 = 2{,}58 = (\bar{x}_{c_2} - 500)/5 \Rightarrow \bar{x}_{c_2} = 512{,}9.$$

Segue-se que a região crítica é

$$RC = \{\bar{x} \in \mathbb{R} \mid \bar{x} \leqslant 487{,}1 \text{ ou } \bar{x} \geqslant 512{,}9\}.$$

Passo 4. A informação pertinente da amostra é sua média, que nesse caso particular é $\bar{x}_0 = 492$.

Passo 5. Como $\bar{x}_0$ não pertence à região crítica, nossa conclusão será não rejeitar H_0. Ou seja, o desvio da média da amostra para a média proposta por H_0 pode ser considerado como devido apenas ao sorteio aleatório dos pacotes.

A situação analisada não é muito realista: conhecer a variância da população. O caso mais geral, de média e variância desconhecidas, será tratado na seção 12.10.

Problemas

6. Sabe-se que o consumo mensal *per capita* de um determinado produto tem distribuição normal, com desvio padrão 2 kg. A diretoria de uma firma que fabrica esse produto resolveu que retiraria o produto da linha de produção se a média de consumo *per capita* fosse menor que 8 kg. Caso contrário, continuaria a fabricá-lo. Foi realizada uma pesquisa de mercado, tomando-se uma amostra de 25 indivíduos, e verificou-se que $\sum_{i=1}^{25} X_i = 180$ kg, onde X_i representa o consumo mensal do i-ésimo indivíduo da amostra.
   (a) Construa um teste de hipótese adequado, utilizando $\alpha = 0{,}05$, e com base na amostra colhida determine a decisão a ser tomada pela diretoria.
   (b) Qual a probabilidade β de se tomar uma decisão errada se, na realidade, a média populacional for $\mu = 7{,}8$ kg?
   (c) Se a diretoria tivesse fixado $\alpha = 0{,}01$, a decisão seria a mesma? (Justifique sua resposta.)
   (d) Se o desvio da população fosse 4 kg, qual seria a decisão, com $\alpha = 0{,}05$? (Justifique sua resposta.)

7. A associação dos proprietários de indústrias metalúrgicas está muito preocupada com o tempo perdido com acidentes de trabalho, cuja média, nos últimos tempos, tem sido da ordem de 60 horas/homem por ano e desvio padrão de 20 horas/homem. Tentou-se um programa de prevenção de acidentes, após o qual foi tomada uma amostra de nove indústrias e medido o número de horas/homens perdidas por acidente, que foi de 50 horas. Você diria, no nível de 5%, que há evidência de melhoria?

8. O salário médio dos empregados das indústrias siderúrgicas de um país é de 2,5 salários mínimos, com um desvio padrão de 0,5 salários mínimos. Uma indústria é escolhida ao acaso e desta é escolhida uma amostra de 49 empregados, resultando um salário médio de 2,3 salários mínimos. Podemos afirmar que esta indústria paga salários inferiores à média nacional, com o nível de 5%?

9. Uma companhia de cigarros anuncia que o índice médio de nicotina dos cigarros que fabrica apresenta-se abaixo de 23 mg por cigarro. Um laboratório realiza 6 análises desse índice, obtendo: 27, 24, 21, 25, 26, 22. Sabe-se que o índice de nicotina se distribui normalmente, com variância igual a 4,86 mg^2. Pode-se aceitar, no nível de 10%, a afirmação do fabricante?

12.6 Teste para Proporção

Vamos usar os passos descritos na seção 12.4 para mostrar a construção do teste para proporções.

Passo 1. Temos uma população e uma hipótese sobre a proporção p de indivíduos portadores de certa característica. Esta hipótese afirma que essa proporção é igual a certo valor p_0. Então,

$$H_0 : p = p_0.$$

O problema fornece informações sobre a alternativa, que pode ter uma das três formas abaixo:

(i) $H_1 : p \neq p_0$ (teste bilateral);
(ii) $H_1 : p > p_0$ (teste unilateral à direita); e
(iii) $H_1 : p < p_0$ (teste unilateral à esquerda).

Passo 2. Como vimos na seção 10.9, a estatística $\hat{p}$, a proporção amostral, tem uma distribuição aproximadamente normal, a saber,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

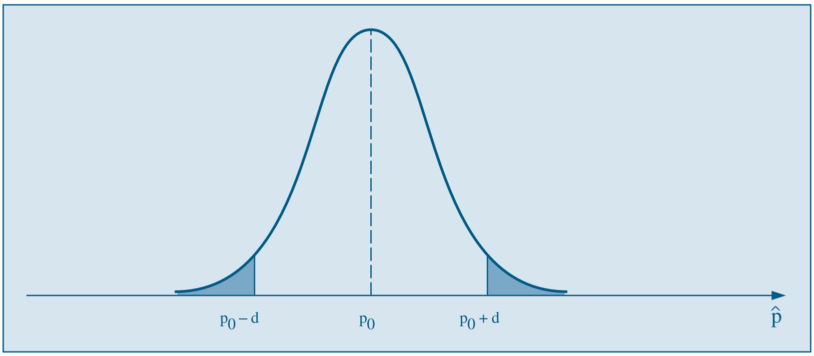
Passo 3. Fixado um valor de α, devemos construir a região crítica para p, sob a suposição de que o parâmetro definido por H_0 seja o verdadeiro. Ou seja, podemos escrever

$$\hat{p} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right),$$

e, conseqüentemente, teremos a região crítica da Figura 12.6, supondo a alternativa (i) acima; sendo que $d = Z(1-\alpha/2)\sqrt{p_0(1-p_0)/n}$ e $Z(p)$ é o p-quantil da normal padrão.

O quarto e quinto passos irão depender da amostra, e o procedimento está descrito no exemplo seguinte.

Figura 12.6: Região crítica para o teste $H_0 : p = p_0$ vs $H_1 : p \neq p_0$.



Exemplo 12.3. Uma estação de televisão afirma que 60% dos televisores estavam ligados no seu programa especial da última segunda-feira. Uma rede competidora deseja contestar essa afirmação e decide usar uma amostra de 200 famílias para um teste. Qual deve ser o procedimento adotado para avaliar a veracidade da afirmação da estação? No passo 4 a seguir daremos o resultado da amostra, pois é importante ficar claro que esse resultado não deve influenciar a escolha da alternativa.

Passo 1. Vamos colocar à prova a afirmação da estação, isto é,

$$H_0 : p = 0{,}60.$$

Sabemos que, se essa hipótese não for verdadeira, espera-se uma proporção menor, nunca maior. A estação divulgaria o máximo possível. Isso nos leva à hipótese alternativa

$$H_1 : p < 0{,}60.$$

Passo 2. A estatística a ser usada é $\hat{p}$, a proporção de 200 famílias que assistiram ao programa na última segunda-feira, e da teoria sabemos que

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{200}\right).$$

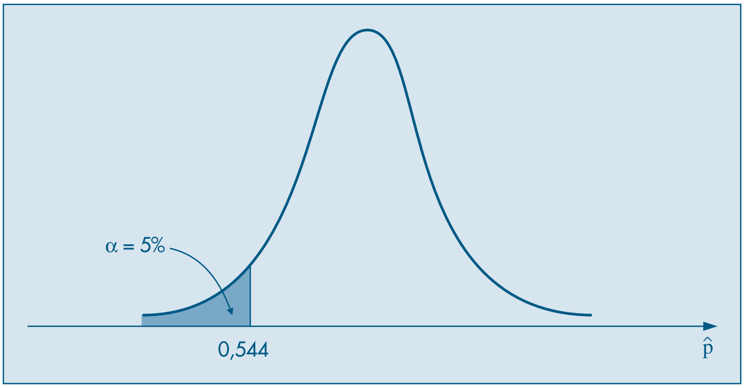
Passo 3. Fixaremos $\alpha = 0{,}05$ e sob a suposição que H_0 seja verdadeira,

$$\hat{p} \sim N(0{,}60,\ 0{,}24/200),$$

o que irá fornecer a região crítica (veja a Figura 12.7)

$$RC = \{\hat{p} \in \mathbb{R} \mid \hat{p} \leqslant 0{,}544\}.$$

Figura 12.7: Região crítica para o teste $H_0 : p = 0{,}60$ vs $H_1 : p < 0{,}60$ do Exemplo 12.3.



De fato, devemos achar o valor $\hat{p}_c$, tal que $P(\hat{p} \leqslant \hat{p}_c) = 0{,}05$, e usando a aproximação normal acima, teremos

$$P\left(Z \leqslant \frac{\hat{p}_c - 0{,}60}{\sqrt{0{,}24/200}}\right) = 0{,}05,$$

o que implica

$$\frac{\hat{p}_c - 0{,}60}{\sqrt{0{,}24/200}} = -1{,}645,$$

o valor $-1{,}645$ sendo obtido da normal padronizada. Segue-se que $\hat{p}_c = 0{,}544$, correspondendo à região crítica acima.

Passo 4. Admitamos que, da pesquisa feita com as 200 famílias, obtivemos 104 pessoas que estavam assistindo ao programa. A proporção da amostra será $\hat{p} = 104/200 = 0{,}52$.

Passo 5. Do resultado do passo anterior, vemos que $0{,}52 \in RC$; portanto, somos levados a rejeitar H_0. Isto é, há evidências que a audiência do programa de segunda-feira não foi de 60% e sim inferior a esse número.

Problemas

10. Uma pessoa gaba-se de adivinhar qual será o resultado do lance de uma moeda, mas é preciso que os presentes não o perturbem com pensamentos duvidosos. Para testar tal capacidade, lançou-se uma moeda perfeita 6 vezes, e o adivinhador acertou 5. Qual seria sua conclusão?

11. O consumidor de um certo produto acusou o fabricante, dizendo que mais de 20% das unidades fabricadas apresentam defeito. Para confirmar sua acusação, ele usou uma amostra de tamanho 50, onde 27% das peças eram defeituosas. Mostre como o fabricante poderia refutar a acusação. Utilize um nível de significância de 10%.

12. Um fabricante garante que 90% dos equipamentos que fornece a uma fábrica estão de acordo com as especificações exigidas. O exame de uma amostra de 200 peças desse equipamento revelou 25 defeituosas. Teste a afirmativa do fabricante, nos níveis de 5% e 1%.

13. Os produtores de um programa de televisão pretendem modificá-lo se for assistido regularmente por menos de um quarto dos possuidores de televisão. Uma pesquisa encomendada a uma empresa especializada mostrou que, de 400 famílias entrevistadas, 80 assistem ao programa regularmente. Com base nos dados, qual deve ser a decisão dos produtores?

12.7 Poder de um Teste

Vimos que, na construção de um teste de hipóteses, procuramos controlar o erro de tipo I, fixando sua probabilidade de ocorrência, α, e construindo a região crítica de modo que $P(\text{RC}|H_0 \text{ verdadeira}) = \alpha$. Ou seja, admitindo que H_0 seja verdadeira, estamos admitindo conhecido(s) o(s) parâmetro(s) que define(m) a distribuição da estatística usada no teste.

Por outro lado, a probabilidade do erro do tipo II, na maioria dos casos, não pode ser calculada, pois a hipótese alternativa usualmente especifica um conjunto de valores para o parâmetro. Voltemos ao exemplo da seção anterior.

Exemplo 12.2. (continuação) No exemplo da máquina de encher pacotes de café, a v.a. X, que descrevia o peso de cada pacote, tinha uma distribuição normal com média μ e variância 400, de modo que a média amostral $\bar{X} \sim N(500, 25)$, sob a hipótese H_0. Esse fato foi utilizado para determinar a região crítica $\text{RC} = \{\bar{x} \in \mathbb{R} \,|\, \bar{x} < 487{,}1 \text{ ou } \bar{x} > 512{,}9\}$ e nossa regra de decisão para verificar se a máquina estava ou não produzindo sob controle foi:

Se $\bar{x} \in RA$, a máquina está sob controle; se $\bar{x} \in RC$, não está,

onde RA é a região de aceitação do teste, isto é, o complementar de RC em relação a $\mathbb{R}$ e, portanto, dada no nosso caso por $\text{RA} = \{\bar{x} \in \mathbb{R} | 487{,}1 \leqslant \bar{x} \leqslant 512{,}9\}$.

A probabilidade β do erro de tipo II não pode ser calculada, a menos que se especifique um valor alternativo para μ. Segue-se que a função característica de operação do teste é dada por

$$\begin{aligned}\beta(\mu) &= P(\text{aceitar } H_0|\mu) = P(\bar{X} \in RA|\mu) \\ &= P(487{,}1 \leqslant \bar{X} \leqslant 512{,}9)|\mu).\end{aligned}$$

Por exemplo, se a máquina se desregular para μ = 505, teremos

$$\beta(505) = P(\overline{X} \in \text{RA}\,|\,\mu = 505) = P(-3{,}58 \leqslant Z \leqslant 1{,}58) = 94{,}28\%,$$

usando o fato que agora $\overline{X} \sim N(505, 25)$. Lembre-se de que supomos que $\sigma^2 = 400$, sempre!

Para qualquer outro valor do parâmetro μ podemos encontrar o respectivo valor de β, para a regra de decisão adotada. No Quadro 12.4 temos as decisões que podemos tomar e suas respectivas implicações.

Quadro 12.4: Decisões possíveis para o teste $H_0 : \mu = 500$ *versus* $H_1 : \mu \neq 500$

Decisão	Valor real do parâmetro	
	$H_0 : \mu = 500$	$H_1 : \mu \neq 500$
a máquina está sob controle: $\mu = 500$	$P(RA \mid H_0) = 0{,}99$	$P(RA \mid H_1) = \beta$ depende de valor alternativo de μ
a máquina não está sob controle: $\mu \neq 500$	$P(RC \mid H_0) = 0{,}01$	$P(RC \mid H_1) = 1 - \beta$ depende de valor alternativo de μ

Observe, por exemplo, que $1 - \beta(500) = P(\text{rejeitar } H_0|\mu = 500) = \alpha = 0{,}01$.

A quantidade $1 - \beta(\mu)$ é usualmente chamada de *poder* ou *potência do teste*, e é a probabilidade de rejeitar a hipótese H_0, dado um valor qualquer de μ, especificado ou não pela hipótese alternativa, e será denotado por $\pi(\mu)$. No nosso exemplo,

$$\pi(\mu) = P(\text{rejeitar } H_0|\mu) = P(\overline{X} < 487{,}1 \text{ ou } \overline{X} > 512{,}9\,|\,\mu).$$

Na Tabela 12.1 temos alguns valores de $\beta(\mu)$ e de $\pi(\mu)$, para diferentes valores de μ, e na Figura 12.8 a representação gráfica da determinação dessa probabilidade. Observe que quanto maior for a distância entre o valor fixado em $H_0(\mu = 500)$ e o valor atribuído para a hipótese alternativa, maior será a probabilidade de tomar a decisão correta. Na Figura 12.9 temos o gráfico de $\pi(\mu)$ para os valores de μ da Tabela 12.1.

Tabela 12.1: Valores de $\beta(\mu)$ e $\pi(\mu)$, usando a regra de decisão $RC = \{\bar{x} \in I\!R\,|\,\bar{x} \leqslant 487{,}1 \text{ ou } \bar{x} \geqslant 512{,}9\}$

Verdadeiro valor de μ		$\pi(\mu)$ (em %)	$\beta(\mu)$ (em %)
À esquerda de 500	À direita de 500		
500	500	1,0	99,0
498	502	1,7	98,3
495	505	5,7	94,3
492	508	16,4	83,6
490	510	28,1	71,9
487	513	49,0	51,0
485	515	66,3	34,7
480	520	92,1	7,9
475	525	99,2	0,8

Figura 12.8: Determinação do poder para o teste do Exemplo 12.2.

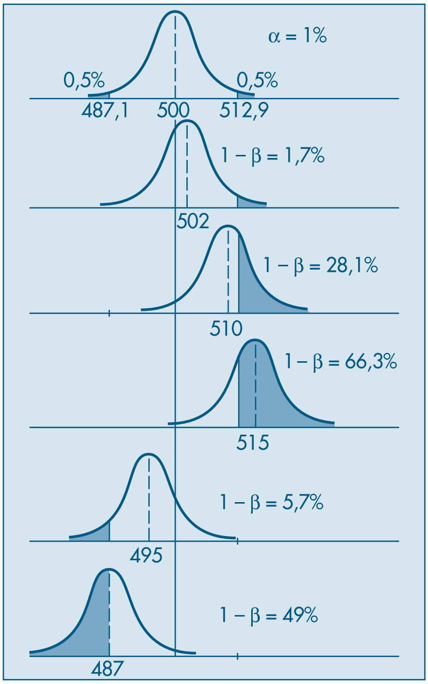
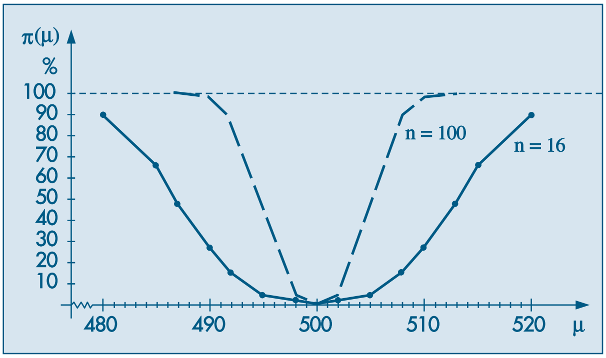


Figura 12.9: Curva de poder para o Exemplo 12.2.



As seguintes propriedades de $\pi(\mu)$ são facilmente verificadas:

(i) $\pi(-\infty) = \pi(+\infty) = 1$;

(ii) $\pi(500) = \alpha$;

(iii) π decresce para $\mu < 500$ (isto é, $d\pi/d\mu < 0$ para $\mu < 500$) e π cresce para $\mu > 500$ (isto é, $d\pi/d\mu > 0$, para $\mu > 500$).

Vemos que $\pi(\mu)$ indica a probabilidade de uma decisão correta, para as diversas alternativas do parâmetro e pode ser usada para decidir entre dois testes para uma mesma hipótese.

Exemplo 12.4. Se, no Exemplo 12.2, a amostra colhida fosse de 100 pacotes em vez de 16, e mantivéssemos o mesmo nível de significância $\alpha = 1\%$, a nova região crítica seria

$$RC = \{\bar{x} \in \mathbb{R} \mid \bar{x} \leqslant 494{,}8 \quad \text{ou} \quad \bar{x} \geqslant 505{,}2\}.$$

Construindo a função poder para esse teste, obtemos a curva tracejada na Figura 12.9. Verifique essas afirmações.

Observando as duas curvas na Figura 12.9, notamos que para todos os valores sob a hipótese alternativa, a probabilidade de uma decisão correta é maior para amostras de tamanho 100 do que de tamanho 16. Dizemos, nesse caso, que o teste baseado em amostras de tamanho 100 é *mais poderoso* do que o teste baseado em amostras de tamanho 16. Esse fato está de acordo com a intuição de que um teste com amostras maiores deve levar a melhores resultados.

De modo geral, se quisermos testar

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0,$$

e determinada a RC do teste, baseada na estatística $\hat{\theta}$, podemos dar a seguinte definição geral.

Definição. A função poder (ou potência) do teste de H_0 contra H_1 é definida por

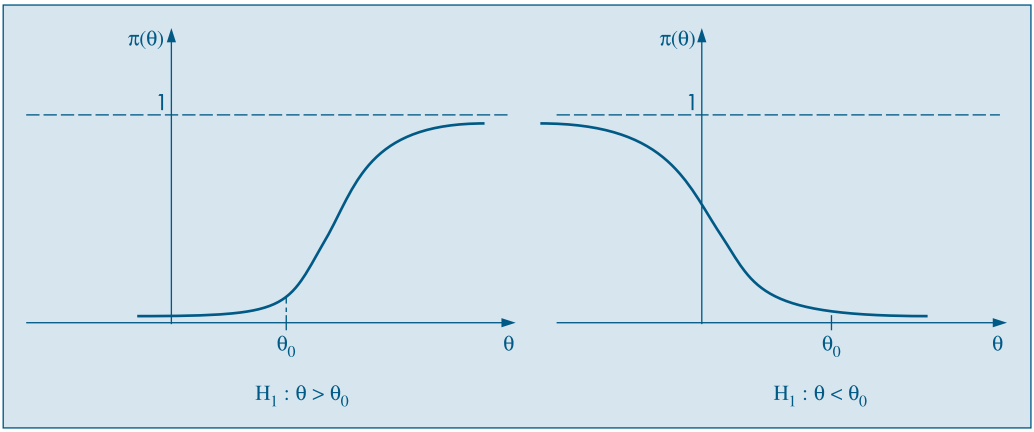
$$\pi(\theta) = P(\hat{\theta} \in RC|\theta),$$

ou seja, é a probabilidade de rejeitar a hipótese nula, como função de θ.

O gráfico dessa função é semelhante àqueles da Figura 12.9, e $\pi(\theta)$ tem as propriedades (i)–(iii) acima, substituindo 500 por θ_0.

Se tivermos hipóteses alternativas unilaterais, da forma $H_1 : \theta < \theta_0$ ou $H_1 : \theta > \theta_0$, obteremos os gráficos da Figura 12.10.

Figura 12.10: Curvas de poder para alternativas unilaterais.



Nos exemplos anteriores fixamos o tamanho da amostra, n, e o nível de significância, α. Suponha que queiramos determinar o tamanho da amostra e os limites da RC, para alcançarmos dado poder para determinado valor do parâmetro. No Exemplo 12.2 poderíamos, por exemplo, fixar $\pi(510) = 0{,}80$ e $\pi(500) = 0{,}05$ (o nível de significância). Dados esses valores, podemos determinar n e a RC. Veja o Problema 33.

Problemas

14. Suponha que estejamos testando $H_0 : p = 0{,}5$ contra $H_1 : p \neq 0{,}5$, e que, para uma amostra de tamanho $n = 10$, decidimos pela região crítica $RC = \{0, 1, 2, 8, 9, 10\}$.
    (a) Determine o nível de significância α.
    (b) Calcule o poder do teste para $p = 0{,}2, 0{,}4, 0{,}6, 0{,}8$. Faça um gráfico do poder como função de p.
    (c) Qual o poder do teste para $p = 0{,}5$?

15. Sendo X o custo de manutenção de um tear, sabe-se que $X \sim N(\mu, 400)$. Para testar a hipótese $H_0 : \mu = 200$, contra a alternativa $H_1 : \mu > 200$, será usada uma amostra de 25 teares.
    (a) Fixando-se $\alpha = 5\%$, encontre a correspondente RC.
    (b) Atribuindo-se valores arbitrários para μ, esboce a função poder do teste.
    (c) Para que valores de μ o poder será maior do que 50%?

12.8 Valor-*p*

O método de construção de um teste de hipóteses, descrito nas seções anteriores, parte da fixação do nível de significância α. Pode-se argumentar que esse procedimento pode levar à rejeição da hipótese nula para um valor de α e à não-rejeição para um valor menor. Outra maneira de proceder consiste em apresentar a *probabilidade de significância* ou *nível descritivo* ou ainda *valor-p* do teste. Os passos são muito parecidos aos já apresentados; a principal diferença está em não construir a região crítica. O que se faz é indicar a probabilidade de ocorrer valores da estatística mais extremos do que o observado, sob a hipótese de H_0 ser verdadeira.

Exemplo 12.5. Voltemos ao Exemplo 12.3, onde

$$H_0 : p = 0{,}60.$$

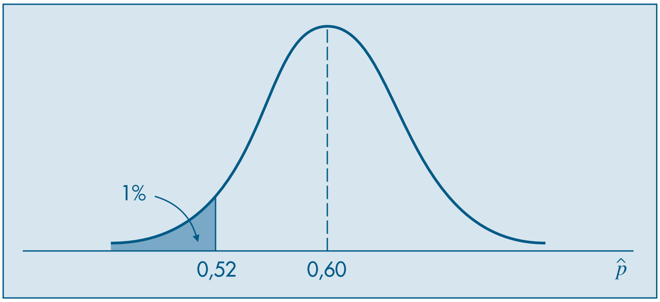
Como vimos, admitindo essa hipótese verdadeira, $\hat{p} \sim N(0{,}60; 0{,}24/200)$. Colhida a amostra obtivemos $\hat{p}_0 = 104/200 = 0{,}52$. Portanto, podemos calcular qual a probabilidade de ocorrerem valores de $\hat{p}$ mais desfavoráveis para H_0 do que esse. É evidente que quanto menor for $\hat{p}$, maior será a evidência contra $H_0 : p = 0{,}60$. Assim, calculemos

$$P(\hat{p} < 0{,}52 \mid p = 0{,}60) = P\left(Z < \frac{\sqrt{200}(0{,}52 - 0{,}60)}{\sqrt{0{,}24}}\right)$$
$$= P(Z < -2{,}30) = 0{,}01 = 1\%.$$

Esse resultado mostra que, se a audiência do programa fosse de 60% realmente, a probabilidade de encontrarmos uma amostra de 200 famílias com 52% ou menos de audiência é de 1%. Isso sugere que, ou estamos diante de uma amostra rara de ocorrer, 1 em 100, ou então a hipótese formulada não é aceitável. Nesse caso, somos levados a essa segunda opção, ou seja, os dados da amostra sugerem que a hipótese H_0 deve ser rejeitada.

O procedimento está ilustrado na Figura 12.11. O valor-p do teste será $\hat{\alpha} = 0{,}01$.

Figura 12.11: Determinação do valor-p para o Exemplo 12.5.



Exemplo 12.6. Um antibiótico A traz em sua bula a seguinte citação: "Nas broncopneumonias, a ação antiinflamatória de A é colocada em evidência pelo estudo dos parâmetros ventilatórios em duplo-cego contra placebo. Durante o tratamento com A pode-se observar uma melhora significativa em relação ao placebo, da capacidade vital ($p < 0{,}05$) e o VEMS($p < 0{,}001$) e do débito respiratório máximo ($p < 0{,}001$)".

Esse exemplo ilustra o uso cada vez mais difundido em muitas áreas aplicadas do conceito de valor-p. As afirmações do tipo "$p < 0{,}05$" acima referem-se a esse conceito. Vale a pena comentar um pouco sobre "estudos duplo-cego", mencionados acima. Nesse tipo de estudo, um número n de indivíduos é dividido em dois grupos de tamanhos aproximadamente iguais; a seleção dos indivíduos que vão pertencer a cada grupo é aleatória. Os indivíduos de um grupo recebem o tratamento (o antibiótico A, no caso), e os do outro grupo recebem placebo (uma substância inóqua). Os pesquisadores que acompanham o experimento não sabem quem recebeu tratamento e quem recebeu placebo, o mesmo acontecendo com os pacientes, daí o nome duplo-cego.

Podemos considerar probabilidades de significância bilaterais. Um procedimento é tomar o valor-p bilateral como sendo igual a duas vezes o valor-p unilateral. Esta prática é razoável quando a distribuição da estatística do teste, sob H_0, for simétrica.

Exemplo 12.7. Uma companhia de serviços de ônibus intermunicipais planejou uma nova rota para servir vários locais situados entre duas cidades importantes. Um estudo preliminar afirma que a duração das viagens pode ser considerada uma v.a. normal, com média igual a 300 minutos e desvio padrão 30 minutos. As dez primeiras viagens realizadas nessa nova rota apresentaram média igual a 314 minutos. Esse resultado comprova ou não o tempo médio determinado nos estudos preliminares?

Passo 1. Indicando por X a duração de cada viagem e por $\mu = E(X)$, queremos testar

$$H_0 : \mu = 300,$$
$$H_1 : \mu \neq 300.$$

Passo 2. Amostras de dez viagens terão média $\bar{X} \sim N(\mu, \sigma^2/10)$.

Passo 3. Sob a hipótese de que H_0 é verdadeira, e pelo fato de σ^2 ser conhecido (σ = 30), teremos

$$\bar{X} \sim N(300, 900/10).$$

Passo 4. Como o valor observado $\bar{x}_0 = 314$, podemos encontrar a probabilidade de ocorrerem amostras com valores de $\bar{X}$ mais extremos do que esse:

$$P(\bar{X} > 314) = P\left(Z > \frac{314 - 300}{9{,}49}\right) = P(Z > 1{,}48) = 0{,}07.$$

Como a distribuição de $\bar{X}$ é normal, portanto simétrica, tomamos $\hat{\alpha} = 0{,}14$. Nosso problema consiste em decidir se essa probabilidade corresponde ou não à chance de ocorrer um evento raro. Por ser uma probabilidade não muito pequena, podemos concluir que não existe muita evidência para rejeitar H_0. Assim, os estudos preliminares parecem estar corretos.

Um problema que pode ocorrer com o procedimento acima, de dobrar a probabilidade, é que o valor de $\hat{\alpha}$ pode ser maior do que um. Por isso, às vezes é preferível anunciar o valor do valor-p unilateral e a direção segundo a qual a observação afasta-se de H_0. No exemplo, o resultado indica que a chance de ocorrerem amostras com médias iguais ou superiores a 314 é 7%, que é um valor ainda não pequeno. Para outro método, ver o Problema 43.

Se indicarmos genericamente por $\hat{\alpha}$ o valor-p, rejeitaremos H_0 para aqueles níveis de significância α maiores do que $\hat{\alpha}$. No Exemplo 12.7, rejeitaremos H_0, por exemplo, se $\alpha = 0{,}10$, mas não a rejeitaremos se $\alpha = 0{,}05$ ou $\alpha = 0{,}01$. Ou seja, se o nível descritivo for muito pequeno, como o caso $\hat{\alpha} < 0{,}01$ do Exemplo 12.6, há evidências de que a hipótese não seja válida. Como vimos nesse exemplo, a probabilidade de significância é muitas vezes denotada por p na literatura (p-value).

Em nosso procedimento de testar uma hipótese estamos usando uma escala de evidências sugerida por Fisher (1954). Suponha que estejamos testando H_0 contra H_1 e, como vimos, rejeitamos H_0 se o valor-p $\hat{\alpha}$ for "bastante pequeno". A Tabela 12.2, extraída de Efron e Gous (1997), ilustra a escala de Fisher, contra H_0 (ou a favor de H_1).

Tabela 12.2: Escala de significância de Fisher.

valor-p	0,10	0,05	0,025	0,01	0,005	0,001
Natureza da evidência	marginal	moderada	substancial	forte	muito forte	fortíssima

Assim, um valor de $\hat{\alpha} = 0{,}01$ indica uma evidência forte contra a validade de H_0, $\hat{\alpha} = 0{,}05$ indica uma evidência moderada etc. É interessante notar que Fisher tomou como ponto de referência o valor 0,05: valores do valor-p menores do que 0,05 indicam que devemos rejeitar a hipótese nula. As considerações feitas por Fisher referiam-se a testes do qui-quadrado (veja o Capítulo 14).

Problemas

16. Suponha que queiramos testar $H_0 : \mu = 50$ contra $H_1 : \mu > 50$, onde μ é a média de uma normal $N(\mu, 900)$. Extraída uma amostra de $n = 36$ elementos da população, obtemos $\bar{x} = 52$. Calcule o valor-*p* $\hat{\alpha}$ do teste.

17. Os novos operários de uma empresa são treinados a operarem uma máquina, cujo tempo X (em horas) de aprendizado é anotado. Observou-se que X segue de perto a distribuição $N(25, 100)$. Uma nova técnica de ensino, que deve melhorar o tempo de aprendizado, foi testada em 16 novos empregados, o quais apresentaram 20,5 horas como tempo médio de aprendizado. Usando o valor-*p*, você diria que a nova técnica é melhor que a anterior?

12.9 Teste para a Variância de uma Normal

Um teste sobre a variância desconhecida de uma variável, com distribuição normal, irá usar a distribuição qui-quadrado, introduzida na seção 7.6.

Considere a média amostral $\bar{X}$ e a variância amostral S^2, ambas obtidas de uma amostra de tamanho n, $(X_1, \ldots, X_n)$ de $X \sim N(\mu, \sigma^2)$. A soma

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \ldots \left(\frac{X_n - \mu}{\sigma}\right)^2$$

terá distribuição $\chi^2(n)$, pois cada $(X_i - \mu)/\sigma$ terá distribuição $N(0,1)$. Logo, se definirmos

$$\hat{\sigma}_*^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2, \qquad (12.1)$$

vemos que

$$Y = \frac{n\hat{\sigma}_*^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \qquad (12.2)$$

tem distribuição $\chi^2(n)$. Observe que o estimador $\hat{\sigma}_*^2$ é muito parecido com o estimador $\hat{\sigma}^2$, definido em (11.6), com μ tomando o lugar de $\bar{X}$. É muito importante conhecer a distribuição de $\sum_{i=1}^{n} (X_i - \bar{X})^2$, para se ter a distribuição de S^2, que será usada no teste desta seção. Note inicialmente que

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\{(X_i - \bar{X}) + (\bar{X} - \mu)\}^2$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \bar{X}) + n(\bar{X} - \mu)^2,$$

e de $\sum_i (X_i - \bar{X}) = 0$, vem que

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2. \qquad (12.3)$$

Dividindo ambos os membros por σ^2, e reescrevendo (12.3) de forma conveniente, teremos

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2+\left(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right)^2. \tag{12.4}$$

O primeiro membro da expressão (12.4) tem distribuição $\chi^2(n)$, como vimos acima. O último termo de (12.4) tem distribuição $\chi^2(1)$. Seria, então, razoável supor que o primeiro termo do segundo membro tenha distribuição $\chi^2(n-1)$. A comprovação desse fato exige recursos fora do alcance deste livro, mas podemos resumir o resultado da seguinte maneira.

Teorema 12.1. Seja $(Z_1, ..., Z_n)$ uma amostra aleatória simples retirada de uma população $N(0,1)$. Então:

(i) $\overline{Z}$ tem distribuição $N(0,1/n)$;

(ii) as variáveis $\overline{Z}$ e $\sum_{i=1}^{n}(Z_i-\overline{Z})^2$ são independentes; e

(iii) $\sum_{i=1}^{n}(Z_i-\overline{Z})^2$ tem distribuição $\chi^2(n-1)$.

Corolário 12.1. A variável aleatória $(n-1)S^2/\sigma^2$ tem distribuição $\chi^2(n-1)$.

Prova. De fato,

$$\frac{(n-1)S^2}{\sigma^2}=\frac{n-1}{\sigma^2}\,\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2=\sum_{i=1}^{n}(Z_i-\overline{Z})^2,$$

bastando escrever $(X_i-\overline{X})/\sigma=(X_i-\mu)/\sigma-(\overline{X}-\mu)/\sigma$.

A expressão (12.4) e a própria definição de χ^2 garantem uma propriedade muito útil: a soma de duas v.a. independentes, cada uma com distribuição χ^2, é uma v.a. também com distribuição χ^2:

$$\chi^2(p)+\chi^2(q)=\chi^2(p+q).$$

Voltemos ao nosso problema original. Queremos testar

$$H_0: \sigma^2=\sigma_0^2,$$
$$H_1: \sigma^2\neq\sigma_0^2.$$

Nossas suposições são que $X_i \sim N(\mu, \sigma^2)$, $i = 1, ..., n$ e os X_i são independentes. A estatística do teste será, sob H_0,

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1). \tag{12.5}$$

Como temos um teste bilateral, a região crítica será da forma RC$=(0, \chi_1^2] \cup [\chi_2^2, +\infty)$, tal que

$$P(\chi^2\in \mathrm{RC}|H_0)=P(0<\chi^2<\chi_1^2 \text{ ou } \chi^2>\chi_2^2)=\alpha,$$

sendo α o nível de significância do teste, fixado *a priori*.

Observado o valor s_0^2 da estatística S^2, obteremos o valor $\chi_0^2 = \dfrac{(n-1)s_0^2}{\sigma_0^2}$. Se $\chi_0^2 \in$ RC, rejeitamos H_0; caso contrário, aceitamos H_0.

Exemplo 12.8. Uma das maneiras de manter sob controle a qualidade de um produto é controlar sua variabilidade. Uma máquina de encher pacotes de café está regulada para enchê-los com média de 500 g e desvio padrão de 10 g. O peso de cada pacote X segue uma distribuição $N(\mu, \sigma^2)$. Colheu-se uma amostra de 16 pacotes e observou-se uma variância de $S^2 = 169$ g². Com esse resultado, você diria que a máquina está desregulada com relação à variância?

Estamos interessados em testar, então,

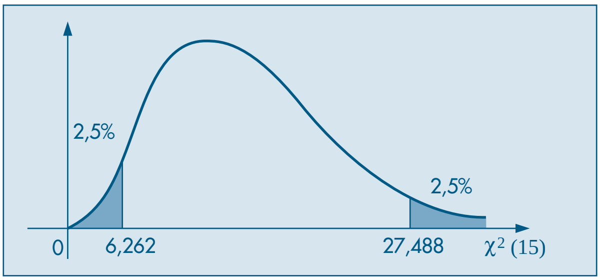
$$H_0 : \sigma^2 = 100,$$
$$H_1 : \sigma^2 \neq 100.$$

A estatística para realizar o teste é (12.5), com $n = 16$. Fixado o nível de significância α em 5%, teremos da Tabela IV que a região crítica é dada por RC = $\{\chi^2\colon 0 \leqslant \chi^2 \leqslant 6{,}262$ ou $\chi^2 \geqslant 27{,}488\}$. Veja a Figura 12.12. O valor observado da estatística é

$$\chi_0^2 = \frac{(n-1)s_0^2}{\sigma_0^2} = \frac{(15)(169)}{100} = 25{,}35.$$

Como $\chi_0^2 \notin$ RC, somos levados a aceitar H_0, isto é, a máquina está sob controle quanto à variância.

Figura 12.12: Região crítica para o teste do Exemplo 12.8.



A construção do IC(σ^2; γ) é feita a partir da expressão

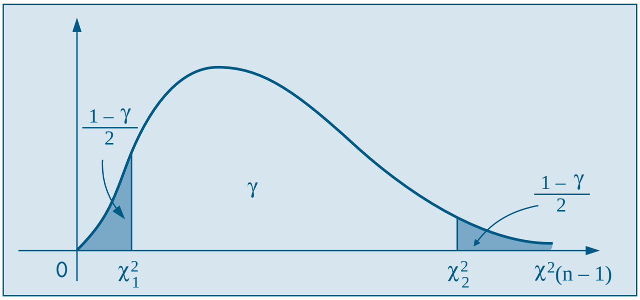
$$P\left(\chi_1^2 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant \chi_2^2\right) = \gamma, \qquad (12.6)$$

que permite obter a seguinte desigualdade:

$$\frac{(n-1)S^2}{\chi_2^2} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\chi_1^2}, \qquad (12.7)$$

que será o IC procurado. Veja a Figura 12.13.

Figura 12.13: Valores críticos para a construção de um intervalo de confiança para a variância.



Exemplo 12.9. Os dados abaixo referem-se às vendas diárias, em reais, durante uma semana, de carros de uma revendedora. Construir um IC(σ^2; 90%).

$$\text{Vendas} : 253,\ 187,\ 96,\ 450,\ 320,\ 105.$$

Inicialmente, calculamos a variância amostral, que é $s_0^2 = 18.460$; em seguida, os valores χ_1^2 e χ_2^2 que satisfaçam (12.6):

$$P(1{,}145 \leqslant \chi^2(5) \leqslant 11{,}070) = 0{,}90.$$

Substituindo em (12.7) obtemos

$$\text{IC}(\sigma^2;\ 0{,}90) = [8.338;\ 80.611].$$

Problemas

18. De uma população $X \sim N(50, 100)$ retira-se uma amostra de dez elementos e calculam-se os valores de $\hat{\sigma}_*^2$ e S^2. Encontre os valores pedidos abaixo, com a maior precisão possível.
    - *(a)* Se $P(\hat{\sigma}_*^2 > a) = 10\%$, encontre o valor de a.
    - *(b)* Sabendo-se que $P(S^2 < a) = 5\%$ e $P(S^2 > b) = 5\%$, encontre a e b.
    - *(c)* $P(S^2 < 163{,}16) = \alpha$, encontre α.
    - *(d)* $P(S^2 > 100) = \alpha$, encontre α.
    - *(e)* $P(S^2 < 18) = \alpha$, encontre α.
    - *(f)* Se o valor observado de S^2 foi 180, qual a probabilidade de encontrar uma amostra que produza um S^2 maior do que o observado?

19. Observou-se a produção mensal de uma indústria durante vários anos, verificando-se que ela obedecia a uma distribuição normal, com variância 300. Foi adotada uma nova técnica de produção e, durante 24 meses, observou-se a produção mensal. Após esse período, constatou-se que $\bar{x} = 10.000$ e $s^2 = 400$. Há razões para se acreditar que a variância mudou, ao nível de 20%?

20. Numa linha de produção, é muito importante que o tempo gasto numa determinada operação não varie muito de empregado para empregado.
    - *(a)* Que parâmetro estatístico poderia ser usado para avaliar esse fato? Por quê?

(*b*) Se 11 empregados apresentam os tempos abaixo para realizar essa operação, qual seria a estimativa para a parâmetro acima?

125	135	115	120	150	130
125	145	125	140	130	

12.10 Teste sobre a Média de uma Normal com Variância Desconhecida

Vimos, na seção 12.5, como testar a média de uma normal, supondo que a variância seja conhecida. Comentamos que essa não é uma suposição realista, logo iremos supor agora que temos uma v.a. X, com distribuição normal, com média μ e variância σ^2 desconhecidas.

No Capítulo 7 introduzimos a distribuição t de Student. Veremos, a seguir, como ela pode ser usada para testar hipóteses sobre μ nessa situação.

Consideremos a estatística

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}. \tag{12.8}$$

Inicialmente, dividamos o numerador e denominador pelo desvio padrão σ da população, e teremos

$$\frac{((\sqrt{n}(\bar{X}-\mu)/\sigma)}{(S/\sigma)}$$

O numerador $Z = (\sqrt{n}\,(\bar{X}-\mu))/\sigma$ tem distribuição $N(0, 1)$, como já foi visto. O quadrado do denominador pode ser escrito como

$$\frac{(n-1)S^2}{\sigma^2} / (n-1) = \frac{Y}{n-1},$$

onde $Y = (n-1)S^2/\sigma^2$. Mas, como foi visto na seção anterior, se os X_i forem normais, Y tem distribuição $\chi^2(n-1)$; logo, a estatística (12.8) é o quociente entre uma v.a $N(0, 1)$ e a raiz quadrada de uma v.a $\chi^2(n-1)$, dividida pelo número de graus de liberdade, e pelo Teorema 7.1 temos que

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1). \tag{12.9}$$

Observe que Z e Y são independentes, pois $\bar{X}$ e S^2 são independentes, pelo Teorema 12.1 (ii).

Estamos, agora, em condições de testar as hipóteses

$$H_0 : \mu = \mu_0,$$
$$H_1 : \mu \neq \mu_0.$$

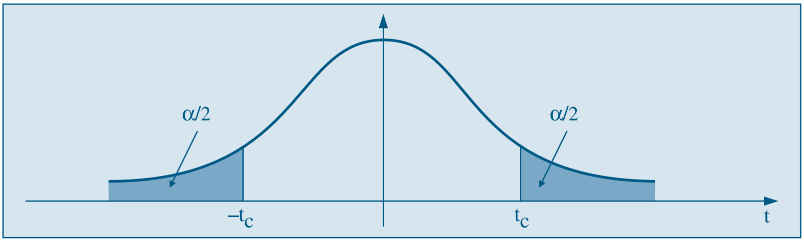
A hipótese alternativa poderia ser $\mu > \mu_0$ ou $\mu < \mu_0$, o que mudaria apenas a região de rejeição de bilateral para unilateral (à direita ou à esquerda, respectivamente).

A estatística a ser usada é

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S}, \qquad (12.10)$$

que sabemos agora ter uma distribuição t de Student com $(n - 1)$ graus de liberdade. Fixado o valor de α, podemos usar a Tabela V e encontrar o valor t_c, tal que $P(|T| < t_c) = 1 - \alpha$. Veja a Figura 12.14.

Figura 12.14: Valores críticos para o teste t.



Colhida a amostra de n indivíduos, calculamos os valores $\bar{x}_0$ e s_0^2 das estatísticas $\overline{X}$ e S^2, respectivamente, e depois o valor $t_0 = \sqrt{n}(\bar{x}_0 - \mu_0)/s_0$ de T. Se o valor dessa estatística for inferior a $-t_c$, ou superior a t_c, rejeita-se H_0. Caso contrário, aceita-se H_0.

Para a construção de intervalos de confiança, temos que

$$P\left(-t_\gamma < \frac{\sqrt{n}(\overline{X} - \mu)}{S} < t_\gamma\right) = \gamma,$$

da qual segue o intervalo de confiança

$$\text{IC}(\mu; \gamma) = \overline{X} \pm t_\gamma \frac{S}{\sqrt{n}}, \qquad (12.11)$$

muito parecido com aquele da variância conhecida.

Exemplo 12.10. Um fabricante afirma que seus cigarros contêm não mais que 30 mg de nicotina. Uma amostra de 25 cigarros fornece média de 31,5 mg e desvio padrão de 3 mg. No nível de 5%, os dados refutam ou não a afirmação do fabricante?

Passo 1. As hipóteses aqui são:

$$H_0 : \mu = 30,$$
$$H_1 : \mu > 30.$$

Passo 2. Supondo que X, a quantidade de nicotina por cigarro, tenha distribuição $N(\mu, \sigma^2)$, a estatística

$$T = \frac{\sqrt{25}(\overline{X} - 30)}{S}$$

terá distribuição $t(24)$.

Passo 3. Por ser um teste unilateral, devemos procurar o valor t_c tal que

$$P(T > t_c) = 0{,}05.$$

Da Tabela V, obtemos $t_c = 1{,}711$, ou seja, a região crítica para a estatística T é RC=[1,711; +∞[.

Passo 4. O valor observado da estatística é

$$t_0 = \frac{5(31{,}5 - 30)}{3} = 2{,}5.$$

Passo 5. Como t_0 pertence à região crítica, rejeitamos H_0, ou seja, há evidências de que os cigarros contenham mais de 30 g de nicotina.

Outra maneira de proceder é calcular o valor-p, ou seja,

$$\hat{\alpha} = P(T > t_0|H_0) = P(T > 2{,}5|H_0) = 0{,}01.$$

Esse valor pequeno de $\hat{\alpha}$ leva à rejeição de H_0.

Para construir um IC(μ; 0,95), verificamos na Tabela V que o valor $t_\gamma = 2{,}064$ e, portanto,

$$\text{IC}(\mu;\ 0{,}95) = 31{,}5 \pm (2{,}064)\ 3/\sqrt{25},$$

ou seja,

$$\text{IC}(\mu;\ 0{,}95) =]30{,}26;\ 32{,}74[.$$

Antes de encerrar este capítulo cabe uma observação. Quando aceitamos uma hipótese, estamos concluindo que temos algum conhecimento sobre a distribuição da variável de interesse. Já quando rejeitamos a hipótese, a distribuição da variável não fica especificada. A construção de intervalos de confiança desempenha um papel importante nessa situação. Ressaltamos, também, que temos usado a expressão “aceitamos” a hipótese, quando o mais correto talvez fosse “não rejeitamos” a hipótese.

Problemas

21. Da população $X \sim N(50,\ 100)$ retirou-se uma amostra casual simples de tamanho $n = 10$, calculando-se o valor de $\overline{X}$, S e o respectivo valor de t.
    (a) Se $P(|\overline{X} - 50| < tS/\sqrt{10}) = 90\%$, encontre o valor de t.
    (b) Se $\overline{X} = 48$ e $S^2 = 120$, qual a probabilidade de encontrar um valor de t menor que o produzido por essa amostra?
    (c) Se $S^2 = 120$, calcule a $P(|\overline{X} - 50| < 2)$.

22. O tempo médio, por operário, para executar uma tarefa, tem sido 100 minutos, com um desvio padrão de 15 minutos. Introduziu-se uma modificação para diminuir esse tempo, e, após certo período, sorteou-se uma amostra de 16 operários, medindo-se o tempo de execução de cada um. O tempo médio da amostra foi 85 minutos, e o desvio padrão foi 12 minutos. Estes resultados trazem evidências estatísticas da melhora desejada? Em caso

Capítulo 13

Inferência para Duas Populações

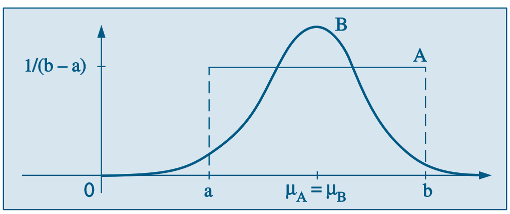
13.1 Introdução

Neste capítulo abordaremos o tópico importante de comparar duas populações P_1 e P_2, baseados em dados fornecidos por amostras dessas populações. Como vimos, uma grande parte das técnicas usadas em Estatística supõe que as variáveis aleatórias envolvidas tenham distribuição normal. Alguns testes que trataremos envolverão a normal. Contudo, se essa suposição de normalidade for violada, procedimentos mais “robustos” têm de ser utilizados, e veremos exemplos de tal situação.

Uma pergunta que aparece freqüentemente em Ciência é a seguinte: o método A é melhor do que o B? Em termos estatísticos, ela equivale a comparar dois conjuntos de informações, resultantes das medidas obtidas da aplicação dos dois métodos a dois conjuntos de objetos ou indivíduos.

Uma das dificuldades que enfrentamos é a de caracterizar adequadamente a “igualdade” ou “equivalência” de duas populações. Por exemplo, suponha que estamos interessados em saber se alunos de duas regiões, A e B, tiveram desempenhos iguais em um mesmo teste nacional. Mais ainda, suponha que tenhamos os resultados do teste para “todos os alunos” das duas regiões, isto é, conhecemos as duas populações. Suponha que cálculos posteriores revelem que as médias e desvios padrões das duas populações sejam iguais, isto é, $\mu_A = \mu_B$ e $\sigma_A = \sigma_B$. Será que isso equivale a dizer que os desempenhos nas duas regiões são equivalentes? Se uma análise mais cuidadosa não for feita, poderemos ser levados a responder afirmativamente a essa questão. Entretanto, observando a Figura 13.1, vemos que é possível ter duas distribuições com os mesmos parâmetros acima, mas formas bastante distintas.

Figura 13.1: Distribuições das populações A e B, com $\mu_A = \mu_B = 4$, $\sigma_A = \sigma_B = 1{,}16$.



Esse fato nos remete à necessidade de também mencionarmos a forma da distribuição. Especificada a forma, a igualdade dos parâmetros que identificam a curva implica a igualdade ou coincidência das duas populações. É bem pouco provável que um mesmo fenômeno obedeça a formas de distribuições distintas, como no exemplo da Figura 13.1. Seguir uma mesma distribuição, porém com parâmetros distintos, é mais verossímil. Como a normal é um modelo importante e seguido por muitas variáveis de interesse prático, estaremos admitindo essa forma, a não ser quando uma análise dos dados nos diga o contrário.

Neste capítulo trataremos de várias situações, que passamos a descrever.

1. *Inferências para duas médias: amostras independentes.*

Aqui temos dados na forma de duas amostras, extraídas independentemente de cada população. É muito comum em experimentos do tipo "controle" *versus* "tratamento", nos quais o interesse principal é verificar o efeito desse último. O caso típico é aquele de comparar uma nova droga com uma padrão, usadas para o tratamento de uma doença.

Exemplo 13.1.

(a) Um curso de Estatística é ministrado pela televisão para um grupo de alunos e ao vivo para outro grupo. Queremos testar a hipótese de que o curso ao vivo é mais eficaz que o curso por meio da televisão.

(b) Queremos comparar o efeito de duas rações, A e B, sobre o crescimento de porcos.

Dois grupos de porcos em crescimento foram alimentados com as duas rações e após cinco semanas verificam-se quais foram os ganhos de peso dos porcos dos dois grupos.

(c) 20 canteiros foram plantados com uma variedade de milho. Em dez deles um novo tipo de fertilizante é aplicado e nos outros um fertilizante padrão. Examinando-se as produções dos dois canteiros, queremos saber se há diferenças significativas entre as produções.

Na maioria das vezes fica claro o que chamamos de controle e tratamento. No exemplo (c) acima, os canteiros tratados com o novo fertilizante seriam o grupo de tratamento, enquanto os demais, tratados com o fertilizante usual, constituiriam o grupo de controle. Mas nos exemplos (a) e (b) essa distinção é apenas convencional.

Formalmente, o *modelo* para o problema das duas amostras é o seguinte: as v.a. $X_1, \ldots, X_m$ representam as respostas do grupo de controle e são consideradas v.a. independentes, com a mesma distribuição, P_1; $Y_1, \ldots, Y_n$ representam as respostas do grupo de tratamento e são v. a. independentes, com a mesma distribuição, P_2. Além disso, $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ são independentes entre si.

A hipótese a ser testada é

$$H_0 : P_1 = P_2, \tag{13.1}$$

ou seja, queremos testar a homogeneidade das populações de onde as amostras foram extraídas. H_0 é chamada *hipótese de homogeneidade*.

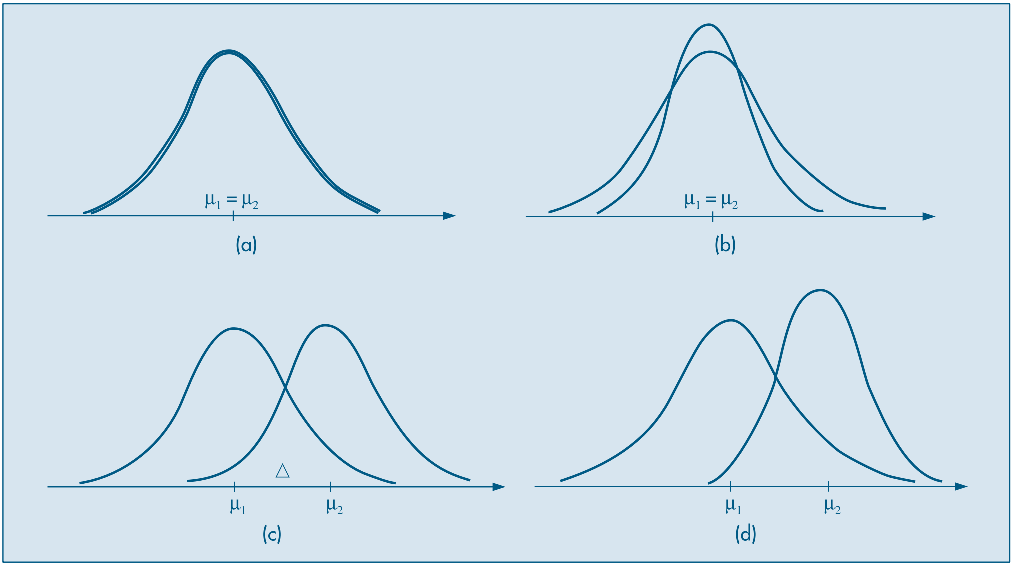
O significado de (13.1) dependerá muito do interesse do pesquisador em considerar qual "tipo" de igualdade implicará a coincidência das duas distribuições. Admitamos que tanto P_1 como P_2 sigam uma distribuição normal, ou seja, $P_1 \sim N(\mu_1, \sigma^2_1)$ e $P_2 \sim N(\mu_2, \sigma^2_2)$. Na Figura 13.2 temos as quatro situações possíveis. Observando os gráficos da Figura 13.2 não temos dúvidas em reconhecer que as duas populações são iguais no caso (a) e diferentes no caso (d). Já nos outros dois casos, podem existir situações em que elas possam ser consideradas iguais ou não. Por exemplo, uma pesquisa para verificar se o salário médio da região P_1 é o mesmo da região P_2 aceita como resposta verdadeira tanto a situação (a) como a (b). Outra pesquisa para verificar se dois processos produzem peças com a mesma qualidade em termos de dispersão aceita como verdadeiras as situações (a) ou (c).

Assim, a estratégia para comparar duas populações, por meio de seus parâmetros, envolve suposições sobre a forma das distribuições, para depois testar médias e variâncias. É comum estarmos interessados em testar apenas que P_1 e P_2 difiram em *localização* (ou posição), isto é, a alternativa a H_0 é que P_1 esteja à direita de P_2, ou o contrário, mas que ambas tenham a mesma dispersão (caso $\mu_1 \neq \mu_2$ e $\sigma_1 = \sigma_2$ da figura). Nesse caso, H_0 será equivalente a

$$H_0\colon \Delta = 0, \tag{13.2}$$

com $\Delta = \mu_2 - \mu_1$.

Figura 13.2: (a) $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$ (b) $\mu_1 = \mu_2$, $\sigma_1 \neq \sigma_2$ (c) $\mu_1 \neq \mu_2$, $\sigma_1 = \sigma_2$ (d) $\mu_1 \neq \mu_2$, $\sigma_1 \neq \sigma_2$.



Os testes t de Student e de Wilcoxon, descritos a seguir, são apropriados para esse tipo de situação. O teste t é aplicável quando P_1 e P_2 supostas são normais, com médias μ e $\mu + \Delta$, respectivamente, e com a mesma variância. O teste de Wilcoxon aplica-se para P_1 e P_2 quaisquer, mas suponha-se que a escala de medidas seja pelo menos ordinal. A análise fica mais fácil quando a P_1 e P_2 são atribuídas distribuições de variáveis contínuas. Discutiremos a razão desta suposição adicional.

Outro caso de interesse é aquele em que queremos testar se as duas médias são iguais, mas as variâncias são diferentes. Na Figura 13.1, as duas curvas teriam dispersões diferentes ao redor de suas médias. Então, um teste preliminar de igualdade de variâncias seria necessário. O teste t de Student para o caso de populações normais será apresentado neste capítulo.

A hipótese (13.1) ou (13.2) nos diz que não há efeito do tratamento. A alternativa usual para H_0 é que o efeito do tratamento é o de aumentar as respostas. Isto é, P_2 gera valores maiores que P_1, com maior freqüência. Mas pode ocorrer o contrário: diminuir as respostas. Por exemplo, o "tratamento" visa a diminuir o tempo para executar determinada tarefa.

2. *Inferências para duas médias: amostras dependentes*

Quando se comparam as médias de duas populações, pode ocorrer uma diferença significativa por causa de fatores externos não-controlados. Por exemplo, no caso do Exemplo 13.4 abaixo, poderia ocorrer que um dos grupos tivesse vendedores mais experientes e habilidosos do que o outro. Logo, a diferença seria devido a esses fatos, e não ao mérito real da técnica de vendas. Um modo de contornar esse problema é coletar as observações em pares, de modo que os dois elementos de cada par sejam homogêneos em todos os sentidos, exceto no que diz respeito ao fator que queremos comparar.

Por exemplo, no caso do Exemplo 13.1 (a), para testar os dois métodos de ensino, poderíamos usar n pares de gêmeos, sendo que um elemento de cada par recebe aulas pela TV e outro ao vivo. Esse procedimento pretende controlar o maior número possível de fatores externos que possam afetar o aprendizado. Se houver diferença no aprendizado, essa dever-se-á realmente ao método.

Esse procedimento também é usado quando observações das duas amostras são feitas no mesmo indivíduo, por exemplo, medindo uma característica do indivíduo antes e depois de ele ser submetido a um tratamento.

O teste t de Student para observações pareadas (ou emparelhadas), supondo normalidade, é apropriado para essas situações.

3. *Inferências para duas variâncias: amostras independentes*

Como vimos no item 1, podemos testar se duas amostras independentes provêm de duas populações com variâncias iguais, desconhecidas. Se essas variâncias forem diferentes, o teste tem de ser modificado. Esse teste, sob a suposição de normalidade das duas populações, usa uma estatística que tem uma distribuição especial, chamada F de Snedecor.

Finalizando esta seção, ressaltamos que poderemos ter mais do que duas amostras, e técnicas semelhantes podem ser desenvolvidas. Veja o Capítulo 15.

13.2 Comparação das Variâncias de Duas Populações Normais

A situação que vamos considerar nesta seção envolve a utilização da distribuição F, estudada na seção 7.7. A descrição a seguir é importante.

Uma das distribuições amostrais mais usadas, e que corresponde a uma distribuição F, resulta do seguinte problema. Suponha que temos duas amostras independentes, de tamanhos n_1 e n_2, retiradas de duas populações normais com a mesma variância σ^2. Indiquemos os estimadores de σ^2 obtidos das amostras por S_1^2 e S_2^2, respectivamente. Já vimos que

$$U = \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1),$$

$$V = \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1),$$

e portanto a v.a.

$$\frac{S_1^2}{S_2^2} = \frac{\dfrac{U}{n_1 - 1}}{\dfrac{V}{n_2 - 1}} \sim F(n_1 - 1, n_2 - 1). \tag{13.3}$$

Essa variável será usada no teste desta seção.

Consideremos, agora, uma amostra $X_1, \dots, X_n$ de uma população com distribuição $N(\mu_1, \sigma_1^2)$ e uma amostra $Y_1, \dots, Y_m$ de uma população com distribuição $N(\mu_2, \sigma_2^2)$. Suponhamos que as duas amostras sejam independentes.

Queremos testar

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Chamemos de S_1^2 e S_2^2 as variâncias amostrais respectivas. De (13.3) e sob a suposição de H_0 ser verdadeira, isto é $\sigma_1^2 = \sigma_2^2$, temos que

$$W = S_1^2/S_2^2 \sim F(n - 1, m - 1). \tag{13.4}$$

Fixado α, encontramos dois números f_1 e f_2, da Tabela VI, tais que

$$P(W \in \text{RC}) = P(W < f_1 \text{ ou } W > f_2) = \alpha.$$

Os valores f_1 e f_2 são determinados de modo que $P(W < f_1) = \alpha/2 = P(W > f_2)$. Na prática, consideramos o quociente (13.4) de tal sorte que $S_1^2/S_2^2 > 1$.

Colhidas as amostras de n e m indivíduos, respectivamente, das duas populações, calculamos os valores observados s^2_{10} e s^2_{20} e o valor observado de W, ou seja, $w_0 = s^2_{10}/s^2_{20}$.

Se w_0 pertencer à região crítica, rejeitamos H_0; caso contrário, a aceitamos.

Exemplo 13.2. Queremos verificar se duas máquinas produzem peças com a mesma homogeneidade quanto à resistência à tensão. Para isso, sorteamos duas amostras de seis peças de cada máquina, e obtivemos as seguintes resistências:

Máquina A:	145	127	136	142	141	137
Máquina B:	143	128	132	138	142	132

As hipóteses a serem testadas são:

$$H_0: \sigma_A^2 = \sigma_B^2 = \sigma^2$$
$$H_1: \sigma_A^2 \neq \sigma_B^2.$$

Sob a suposição de normalidade das medidas de resistência à tensão, para as duas máquinas, temos que a v.a. W, definida por (13.4), tem uma distribuição $F(5,5)$. Fixando $\alpha = 0{,}10$ e consultando a Tabela VI, teremos

$$RC =]0, (5{,}05)^{-1}[\cup]5{,}05, +\infty[.$$

Das amostras encontramos $s_A^2 = 40$ e $s_B^2 = 37$, portanto $w_0 = 1{,}08$. Como esse valor não pertence à região crítica, aceitamos H_0, ou seja, as máquinas produzem com a mesma homogeneidade quanto à variabilidade.

Caso tivéssemos rejeitado a hipótese de igualdade das variâncias, seria conveniente obter um intervalo de confiança para o quociente das duas variâncias. De (13.3) podemos escrever, quando $\sigma_1^2 \neq \sigma_2^2$,

$$W = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{U/(n-1)}{V/(m-1)} \sim F(n-1, m-1),$$

e para um dado γ, $0 < \gamma < 1$, podemos encontrar dois valores f_1 e f_2, tais que

$$P(f_1 < F(n-1, m-1) < f_2) = \gamma.$$

Dessa igualdade, segue-se que, com probabilidade γ,

$$f_1 < \frac{S_1^2}{S_2^2}\,\frac{\sigma_2^2}{\sigma_1^2} < f_2,$$

ou seja, o $IC(\sigma_2^2/\sigma_1^2; \gamma)$ será dado por

$$f_1\,\frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < f_2\,\frac{S_2^2}{S_1^2} \qquad (13.5)$$

Exemplo 13.3. Suponha que para outras seis medidas para as máquinas A e B do Exemplo 13.2 tivéssemos $S_A^2 = 85$ e $S_B^2 = 8$. Como $w_0 = 85/8 = 10{,}62$, rejeitaríamos H_0. Então, o IC dado por (13.5) ficaria, com $\gamma = 0{,}90$,

$$\frac{1}{5{,}05}\,\frac{8}{85} < \frac{\sigma_B^2}{\sigma_A^2} < 5{,}05\,\frac{8}{85},$$

ou seja,

$$0{,}019 < \frac{\sigma_B^2}{\sigma_A^2} < 0{,}475.$$

Invertendo-se, obtemos, também,

$$2{,}10 < \frac{\sigma_A^2}{\sigma_B^2} < 52{,}6,$$

que indica a variação possível, no nível fixado, da razão entre as duas variâncias. Note que, sob H_0, temos $\sigma_A^2/\sigma_B^2 = 1$, que não pertence a esse intervalo.

Problemas

1. Da população $X \sim N(50, 100)$ retirou-se uma amostra casual simples de $n = 10$ elementos. Da população $Y \sim N(60, 100)$ retirou-se uma amostra casual simples de $m = 6$ indivíduos, independente da primeira. Obtemos as variâncias amostrais S_1^2 e S_2^2, respectivamente.
   (a) Encontre o valor de a, tal que $P(S_1^2/S_2^2 < a) = 95\%$.
   (b) Encontre o valor de b, tal que $P(S_1^2/S_2^2 > b) = 95\%$.
2. Por que em (13.3) as v.a. U e V são independentes?
3. Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é por meio do desvio padrão de seus salários. A fábrica A diz ser mais coerente na política salarial do que a fábrica B. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de A, e 15 de B, obtendo-se os desvios padrões $s_A = 1.000$ reais e $s_B = 1.600$ reais. Qual seria a sua conclusão?
4. Deseja-se comparar a qualidade de um produto produzido por duas fábricas. Essa qualidade será definida pela uniformidade com que o produto é produzido em cada fábrica. Tomaram-se duas amostras, uma de cada fábrica, medindo-se o comprimento dos produtos (o resumo dos resultados está no quadro abaixo). A qualidade das duas fábricas é a mesma? Caso a sua resposta seja negativa, dê um intervalo de confiança para indicar a intensidade dessa desigualdade.

Estatísticas	Fábrica A	Fábrica B
Amostra	21	17
Média	21,15	21,12
Variância	0,0412	0,1734

13.3 Comparação de Duas Populações: Amostras Independentes

Nesta seção estudaremos o caso onde temos duas amostras independentes, $X_1, ..., X_n$ e $Y_1, ..., Y_m$, de duas populações P_1 e P_2, respectivamente.

Estaremos interessados em comparar as médias dessas populações, verificando se elas podem ser consideradas iguais ou não. No caso de populações normais, teremos, preliminarmente, de usar o que aprendemos na seção anterior, para testar se as variâncias de P_1 e P_2 são iguais.

Consideraremos duas situações: na primeira, iremos supor que as populações sejam normais (reveja os Problemas 32, 33 e 34 do Capítulo 10, os Problemas 31 e 32 do Capítulo 11 e o Problema 29 do Capítulo 12); na segunda, essa suposição não é necessária.

13.3.1 Populações Normais

Aqui, $P_1 \sim N(\mu_1, \sigma_1^2)$ e $P_2 \sim N(\mu_2, \sigma_2^2)$.

Queremos testar a hipótese (13.1), que aqui fica escrita na forma

$$H_0 : \mu_1 = \mu_2.$$

Na situação da Figura 13.2 (c), a alternativa adequada é

$$H_1 : \mu_2 > \mu_1,$$

mas supondo as variâncias iguais. Se estivermos apenas interessados em verificar se existe diferença entre as médias das duas populações, não importando a direção, então a alternativa adequada será

$$H_1 : \mu_1 \neq \mu_2.$$

Para cada amostra calculamos os estimadores da média e da variância:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2;$$

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i, \qquad S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2.$$

Sob a hipótese H_0, isto é, $\mu_1 = \mu_2$,

$$E(\bar{X} - \bar{Y}) = 0, \tag{13.6}$$

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}. \tag{13.7}$$

Como $\bar{X} - \bar{Y}$ tem distribuição normal, se as variâncias fossem conhecidas, a estatística

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \tag{13.8}$$

teria distribuição normal padrão, *sob a hipótese nula* H_0, e poderia ser usada para testar H_0 contra H_1. Contudo, nas situações de interesse prático, as variâncias não são conhecidas, devendo ser substituídas por estimativas convenientes. Aqui, a distribuição *t* de Student desempenha papel importante. Notemos que, da definição da v.a. t de Student, $t = \sqrt{n}(\bar{X} - \mu)/S$, podemos obter

$$t^2 = \frac{\left(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}{[(n-1)S^2/\sigma^2]/(n-1)} \sim F(1, n-1), \tag{13.9}$$

o que mostra uma relação entre as distribuições $t(n-1)$ e $F(1, n-1)$. Observe que o numerador de (13.9) é o quadrado de uma $N(0, 1)$ e, portanto, tem uma distribuição $\chi^2(1)$, e o denominador é o quociente de uma v.a. $\chi^2(n-1)$ por $(n-1)$.

Vamos considerar dois casos.

(a) Mesma Variância, Desconhecida

Suponha que, ao testar a hipótese de igualdade de variâncias, esta não seja rejeitada, isto é, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, porém essa variância comum é desconhecida. Como S_1^2 e S_2^2 são dois estimadores não-viesados de σ^2, podemos combiná-los para obter um estimador comum

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{n+m-2}, \tag{13.10}$$

que também é um estimador não-viesado de σ^2. Mais ainda, cada parcela do numerador de (13.10), quando dividida por σ^2, terá distribuição qui-quadrado, com $(n-1)$ e $(m-1)$ graus de liberdade, respectivamente. Logo, teremos que

$$\frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi^2(n+m-2). \tag{13.11}$$

Pelo Teorema 7.1, a estatística

$$T = \frac{\dfrac{\bar{X} - \bar{Y}}{\sigma\sqrt{1/n + 1/m}}}{S_p/\sigma} = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n + 1/m}} \tag{13.12}$$

terá uma distribuição t de Student, com $(n+m-2)$ graus de liberdade, *sob a hipótese* H_0, isto é, se $\mu_1 = \mu_2$.

Tabela 13.1: Dados para duas técnicas de vendas.

Dados	Vendas	
	Técnica A	Técnica B
Média	68	76
Variância	50	75
Vendedores	12	15

Exemplo 13.4. Duas técnicas de venda são aplicadas por dois grupos de vendedores: a técnica A, por 12 vendedores, e a técnica B, por 15 vendedores. Espera-se que a técnica B produza melhores resultados. No final de um mês, obtiveram-se os resultados da Tabela 13.1.

Vamos testar, para o nível de significância de 5%, se há diferenças significativas entre as vendas resultantes das duas técnicas. Informações adicionais permitem supor que as vendas sejam normalmente distribuídas, com uma variância comum σ^2, desconhecida.

As hipóteses a serem testadas ficam

$$H_0\colon \mu_{\mathrm{A}} = \mu_{\mathrm{B}}$$
$$H_1\colon \mu_{\mathrm{A}} < \mu_{\mathrm{B}}.$$

Pelas suposições acima, podemos usar a estatística (13.12), com $n = 12$, $m = 15$ e $S_p^2 = (11S_A^2 + 14S_B^2)/25$. Da Tabela V obtemos RC =]1,708, + ∞[.

Da Tabela 13.1 calculamos

$$s_p^2 = \frac{11(50) + 14(75)}{25} = 64,$$

$$t_0 = \frac{76 - 68}{8\sqrt{1/12 + 1/15}} = 2{,}56.$$

Como $t_0 \in$ RC, rejeitamos H_0, ou seja, existe evidência de que a técnica B produz melhores resultados do que a técnica A.

Encontrada diferença entre os métodos, a continuação natural é construir um intervalo de confiança para a diferença $\Delta = \mu_B - \mu_A$. Do resultado (13.12) é fácil verificar que

$$\mathrm{IC}(\Delta;\, \gamma) = (\bar{x}_0 - \bar{y}_0) \pm t_\gamma s_p \sqrt{1/n + 1/m}.$$

Para o nosso exemplo, com $\gamma = 0{,}95$, esse intervalo reduz-se a

$$\mathrm{IC}(\Delta;\, 0{,}95) = 8 \pm (2{,}06)(8)\sqrt{1/12 + 1/15}$$
$$= 8 \pm 6{,}38 =]1{,}62;\ 14{,}38[.$$

(b) Variâncias Desiguais, Desconhecidas

Quando a hipótese de igualdade de variâncias for rejeitada, devemos usar a estatística

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n + S_2^2/m}}. \tag{13.13}$$

Pode-se provar que, sob a veracidade de H_0, a v.a. T aproxima-se de uma distribuição t de Student, com o número de graus de liberdade dado aproximadamente por

$$\nu = \frac{(A+B)^2}{A^2/(n-1) + B^2/(m-1)}, \tag{13.14}$$

na qual

$$A = s_1^2/n, \quad B = s_2^2/m.$$

Como esse valor é geralmente fracionário, arredonde para o inteiro mais próximo para obter o número de graus de liberdade.

Exemplo 13.5. Queremos testar as resistências de dois tipos de vigas de aço, A e B. Tomando-se $n = 15$ vigas do tipo A e $m = 20$ vigas do tipo B, obtemos os valores na Tabela 13.2. Usando um teste F com nível $\alpha = 10\%$ rejeitamos a hipótese de variâncias iguais.

Tabela 13.2: Médias e variâncias para dois tipos de vigas de aço.

Tipo	Média	Variância
A	70,5	81,6
B	84,3	161,5

Consideremos as hipóteses

$$H_0: \mu_A = \mu_B$$
$$H_1: \mu_A \neq \mu_B.$$

A estatística a ser usada é (13.13), com $\nu = (182,66)/(2,11 + 3,43) = 32,9$, logo tomamos $\nu = 33$. Com $\alpha = 0,05$, obtemos da Tabela V que RA =]–2,0345; 2,0345[. Com os dados da Tabela 13.2, temos $t_0 = (-13,8)/3,68 = -3,75$.

Como $t_0 \in$ RC, rejeitamos H_0, ou seja, há evidências de que os dois tipos de vigas têm resistências médias diferentes.

Problemas

5. Num estudo comparativo do tempo médio de adaptação, uma amostra aleatória, de 50 homens e 50 mulheres de um grande complexo industrial, produziu os seguintes resultados:

Estatísticas	Homens	Mulheres
Médias	3,2 anos	3,7 anos
Desvios padrões	0,8 anos	0,9 anos

Que conclusões você poderia tirar para a população de homens e mulheres dessa indústria? (Indique as suposições feitas para resolver o problema.)

6. Diversas políticas em relação às filiais de uma rede de supermercados estão associadas ao gasto médio dos clientes em cada compra. Deseja-se comparar esse parâmetro para duas novas filiais, por meio de duas amostras de 50 clientes cada. As médias obtidas

foram 62 e 71, respectivamente. Sabe-se que o desvio padrão, em ambos os casos, deve ser da ordem de 20 unidades. É possível afirmar que o gasto médio nas duas filiais seja o mesmo? Caso contrário, dê um intervalo de confiança para a diferença.

7. Uma fábrica de embalagens para produtos químicos está estudando dois processos para combater a corrosão de suas latas especiais. Para verificar o efeito dos tratamentos, foram usadas amostras cujos resultados estão no quadro abaixo (em porcentagem de corrosão eliminada). Qual seria a conclusão sobre os dois tratamentos?

Método	Amostra	Média	Desvio Padrão
A	15	48	10
B	12	52	15

8. No Problema 4, teste a hipótese de que as médias dos comprimentos do produto produzido pelas duas fábricas são iguais.

9. Para investigar a influência da opção profissional sobre o salário inicial de recém-formados, investigaram-se dois grupos de profissionais: um de liberais em geral e outro de formados em Administração de Empresas. Com os resultados abaixo, expressos em salários mínimos, quais seriam suas conclusões?

Liberais	6,6	10,3	10,8	12,9	9,2	12,3	7,0	
Administradores	8,1	9,8	8,7	10,0	10,2	8,2	8,7	10,1

13.3.2 Populações Não-Normais

Passamos, agora, a descrever um teste que não faz suposições a respeito da forma das distribuições P_1 e P_2, a não ser que as variáveis envolvidas tenham uma escala de medida pelo menos ordinal. Ou seja, podemos abordar o caso de variáveis qualitativas ordinais e variáveis quantitativas. Esse teste (chamado de Wilcoxon ou de Mann-Whitney) pertence a uma categoria de procedimentos chamados *não-paramétricos* ou *livres de distribuição*.

Teremos para análise amostras *independentes* das duas populações e queremos testar a hipótese (13.1) contra a alternativa de que as distribuições diferem em localização: estaremos interessados em saber se uma população tende a ter valores maiores do que a outra, ou se elas têm a mesma mediana ou média.

O teste de Wilcoxon é baseado nos *postos* dos valores obtidos combinando-se as duas amostras. Isso é feito ordenando-se esses valores, do menor para o maior, independentemente do fato de qual população cada valor provém. A estatística do teste é a *soma* dos postos associados aos valores amostrados de uma população, P_1, por exemplo. Se essa soma for grande, isso é uma indicação de que os valores dessa população tendem a ser maiores do que os valores de P_2, e, então, rejeitamos (13.1).

No caso de termos uma v.a. qualitativa ordinal, comumente associamos números às diversas categorias (ou classes, ou atributos), segundo as quais a variável é classificada. Por exemplo, podemos ter 1 para *bom*, 2 para *muito bom* e 3 para *ótimo*. Vemos, então, que esses valores são os postos, nesse caso, e em outras situações é preferível trabalhar com postos do que com valores arbitrários associados à v.a. qualitativa.

Quando trabalhamos com v.a. quantitativas poderemos ter valores repetidos nas amostras. Veremos como associar postos nesse caso. Para evitar esses *empates*, uma possibilidade é supor que a v.a. seja contínua, de modo que se X for uma tal variável, $P(X = x_0) = 0$. Essa suposição é eventualmente necessária para o desenvolvimento teórico do teste, mas na prática, quer X seja contínua ou discreta, valores repetidos poderão aparecer.

(a) Observações Distintas

Suponha que tenhamos N observações $Z_1, Z_2, ..., Z_N$. Ordenando-as da menor para a maior obtemos as estatísticas de ordem, $Z_{(1)} \leqslant Z_{(2)} \leqslant ... \leqslant Z_{(N)}$. Inicialmente, suponha que não haja observações coincidentes, de modo que os sinais de $\leqslant$ são substituídos por $<$. Então, associamos números (normalmente 1, 2, ..., N), chamados *postos*, que correspondem às posições das observações na ordenação. O posto de Z_i é igual a 1 + (número de $Z_j < Z_i$). Assim, dadas as observações

$$Z_1 = 0{,}3,\ Z_2 = 1{,}5,\ Z_3 = -0{,}5,\ Z_4 = 2{,}0,$$

os postos de Z_1, Z_2, Z_3 e Z_4 serão, respectivamente,

$$R_1 = 2,\ R_2 = 3,\ R_3 = 1,\ R_4 = 4,$$

já que a ordenação resulta em

$$-0{,}5 < 0{,}3 < 1{,}5 < 2{,}0, \quad \text{ou} \quad Z_3 < Z_1 < Z_2 < Z_4.$$

Exemplo 13.6. Num estudo sobre um novo método para ensinar Matemática elementar, foram selecionadas cinco crianças. Destas, três são escolhidas ao acaso e ensinadas segundo o novo método, enquanto as outras duas funcionaram como controle e receberam instrução por um método tradicional. Após um período de cinco semanas é feito um teste, e as crianças são ordenadas segundo seu desempenho: a criança que tiver menor nota recebe posto 1, etc., até a criança que tiver maior nota recebe posto 5.

O método de ensino será considerado eficaz se as três crianças que recebem o novo método tiverem postos altos nessa ordenação combinada das cinco crianças. Seja H_0 a hipótese nula que especifica que o tratamento (novo método) não tem efeito, isto é, a nota da criança não é afetada se ela for ou não ensinada pelo novo método. Se H_0 for verdadeira, o posto atribuído a cada criança é determinado somente pela sua inteligência, ou seja, a ordenação das crianças não depende de qual recebe tratamento e qual funciona como controle. A Tabela 13.3 mostra todos os casos possíveis para a ordenação, onde C indica controle e T, tratamento.

Tabela 13.3: Valores de W_S para o Exemplo 13.6.

Postos					W_S
1	2	3	4	5	
C	C	T	T	T	12
C	T	C	T	T	11
T	C	C	T	T	10
C	T	T	C	T	10
T	C	T	C	T	9
C	T	T	T	C	9
T	C	T	T	C	8
T	T	C	T	C	7
T	T	T	C	C	6
T	T	C	C	T	8

Vemos que as crianças e seus postos podem ser divididos em dois grupos (tratados e controles) de $\binom{5}{3} = 10$ maneiras diferentes. A suposição de que as três crianças recebendo o tratamento são selecionadas ao acaso e de que os tratamentos são equivalentes, implica que todas as dez possibilidades têm a mesma probabilidade 1/10.

Consideremos a estatística

$$W_S = S_1 + S_2 + S_3, \tag{13.15}$$

onde S_1, S_2 e S_3 são os postos das crianças que receberam o tratamento na amostra combinada.

Poderíamos considerar como regra de decisão para rejeitar H_0 a ocorrência de $W_S = 12$, correspondendo à ocorrência de CCTTT, clara superioridade do tratamento. Qual seria a probabilidade de esse evento ocorrer por mero acaso, ou seja, quando os dois métodos são equivalentes? Nesse caso teremos

$$P(W_S = 12 | H_0 \text{ verdadeiro}) = 0{,}10,$$

que é a probabilidade do erro de tipo I, ou seja, o nível de significância do teste. Mas, como vimos antes, usualmente procedemos de maneira oposta, ou seja, fixamos α e não a regra de decisão.

Como vimos acima, rejeitamos H_0 para valores grandes de W_S, ou seja, $W_S \geqslant c$, onde c é uma constante determinada a partir do nível de significância do teste, α. Obtemos o *teste de Wilcoxon*:

"Rejeite H_0 se $W_S \geqslant c$, onde c é determinada por $P(W_S \geqslant c | H_0$ é verdadeira$) = \alpha$".

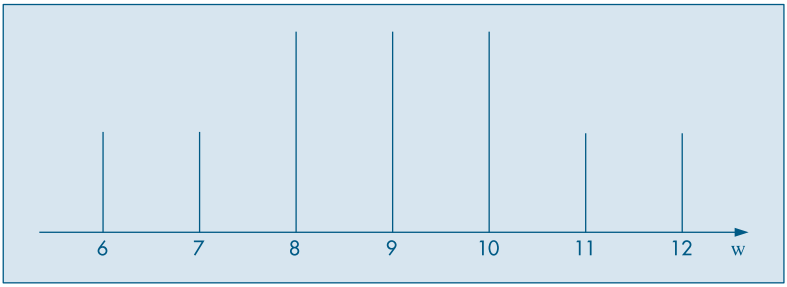
A distribuição nula (isto é, sob H_0) de W_S é obtida da Tabela 13.3 e está na Tabela 13.4.

Tabela 13.4: Distribuição de W_S, observações distintas.

w	6	7	8	9	10	11	12
$P(W_S = w)$	1/10	1/10	2/10	2/10	2/10	1/10	1/10

A distribuição de W_S é simétrica ao redor do valor 9 que, como veremos, representa a média de W_S, dada por $n(N + 1)/2$, com $N = n + m$ (Ver Figura 13.3).

Figura 13.3: Distribuição de W_S para o Exemplo 13.6.



Se, por exemplo, $\alpha = 0{,}05$, não existe valor satisfazendo $P(W_S \geqslant c) = 0{,}05$. Podemos encontrar c somente para valores de α iguais a 0,1; 0,2; 0,4 etc. Por exemplo, se $\alpha = 0{,}1$, então

$$P(W_S \geqslant 12) = 0{,}1 \quad \text{e} \quad c = 12.$$

Consideremos, agora, a situação geral. Queremos testar (13.1). Temos duas amostras independentes, $X_1, ..., X_n$, de P_1, e $Y_1, ..., Y_m$, de P_2. Seja $N = n + m$ e combinamos as duas amostras numa só, ordenamos os N valores no menor para o maior e chamemos $S_1 < S_2 < ... < S_m$ os postos dos Y_i (tratamentos) e $R_1 < R_2 < ... < R_n$ os postos dos X_i (controles). Estamos supondo que não haja empates. Seja

$$W_S = S_1 + S_2 + ... + S_m \tag{13.16}$$

a soma dos postos dos tratamentos. Rejeitamos H_0 se $W_S \geqslant c$.

No caso bilateral, rejeite H_0 se $W_S < c_1$ ou $W_S > c_2$, para dado α.

Não é difícil verificar que, se a distribuição de P_1 for contínua, então

$$P(S_1 = s_1, ..., S_m = s_m) = \frac{1}{\binom{N}{m}}, \tag{13.17}$$

onde $s_1 < s_2 < ... < s_m$ e $s_i \in \{1, 2, ..., N\}$, $N = n + m$.

Observação. Por (13.17) vemos que a distribuição dos postos e portanto de W_S não depende de P_1. Isso não ocorrerá se P_1 não for contínua. Se as distribuições P_1 e P_2 forem contínuas, há ausência de empates (isto é, coincidência entre valores de X e de Y). Isso significa que poderíamos considerar nossas medidas de X e Y de tal sorte que coincidências seriam evitadas. Na prática, contudo, as medidas são feitas em geral com o mesmo número de casas decimais, de modo que empates podem ocorrer. Essa situação é analisada abaixo.

A distribuição sob H_0 de W_S pode ser encontrada como no Exemplo 13.6. Para dado valor de w, verificamos quantas amostras de tamanho m, retiradas de $P = \{1, 2, ..., N\}$ fornecem o valor de w. Se # $(w; n, m)$ indicar esse número, então, por (13.17),

$$P(W_S = w \mid H_0 \text{ é verdadeira}) = \frac{\#(w;\, n,\, m)}{\binom{N}{m}}. \qquad (13.18)$$

Pode-se provar o seguinte resultado (veja, por exemplo, Lehmann, 1975):

Teorema 13.1. Para a estatística W_S temos:

$$E(W_S) = \frac{m(N+1)}{2}, \qquad (13.19)$$

$$Var(W_S) = \frac{nm(N+1)}{12}. \qquad (13.20)$$

Além disso, a distribuição de W_S pode ser aproximada pela distribuição normal; quando $n, m \to \infty$, a v.a.

$$Z = \frac{W_S - E(W_S)}{\sqrt{Var(W_S)}} \qquad (13.21)$$

tem uma distribuição aproximada $N(0, 1)$.

Uma estatística equivalente a W_S é

$$U_S = W_S - \frac{1}{2} m(m+1), \qquad (13.22)$$

chamada *estatística de Mann-Whitney*. Há duas vantagens em se usar U_S:

(a) a distribuição de U_S para $n = n_1$ e $m = m_1$ é a mesma que a distribuição de U_S quando os tamanhos são invertidos, isto é, para $n = m_1$ e $m = n_1$. Isso não acontece com W_S;

(b) o valor mínimo de W_S é obtido quando os postos dos m tratamentos são 1, 2, ..., m e $1 + 2 + ... + m = m(m+1)/2$; logo, o valor mínimo de U_S é zero, para quaisquer valores de n e m, simplificando a construção de tabelas. A Tabela VIII do Apêndice dá os valores de $P(U_S \leqslant u)$.

Para essa estatística temos o resultado seguinte.

Teorema 13.2. A média e variância de U_S são dadas por

$$E(U_S) = \frac{nm}{2} \qquad (13.23)$$

e

$$Var(U_S) = \frac{nm(N+1)}{12}, \qquad (13.24)$$

respectivamente. Além disso, a distribuição de U_S pode também ser aproximada por uma normal.

Exemplo 13.7. Suponha que $m = n = 10$ e queremos calcular $P(W_S \leqslant 87)$. O valor tabelado é 0,0952, que é encontrado na Tabela VIII com $n = m = 10$, e levando-se em conta que $U_S = 87 - 10 \times 11/2 = 32$ e, portanto, $P(U_S \leqslant 32) = 0{,}0952$.

Por outro lado, usando a aproximação normal, $E(W_S) = 105$, $\text{Var}(W_S) = 175$, temos

$$P(W_S \leqslant 87) = P\left(\frac{W_S - 105}{\sqrt{175}} \leqslant \frac{87 - 105}{\sqrt{175}}\right) = P(Z \leqslant -1{,}36) \approx 0{,}087,$$

que está bem próxima do valor encontrado usando-se a tabela.

A aproximação pode ser melhorada usando-se a correção de continuidade discutida na seção 7.5, pois aqui também estamos aproximando a distribuição de uma v.a. discreta (W_S) por uma distribuição de variável contínua (normal). Verifique que, usando essa correção, obtemos $P(W_S \leqslant 87) \approx 0{,}0934$.

(b) Observações Não Todas Distintas

Consideremos, agora, a situação em que haja observações coincidentes, ou empates.

Suponha, por exemplo, que $n = 3$, $m = 2$ e as observações são

1,3; 1,5; 1,5; 2,1; 2,5.

Nesse caso, usamos *postos médios*. Associamos o posto 1 à observação 1,3; às duas observações empatadas 1,5 associamos a média dos postos 2 e 3, que seriam atribuídas se as observações fossem distintas, ou seja, atribuímos o posto $(2 + 3)/2 = 2{,}5$; à observação 2,1 atribuímos o posto 4 e à observação 2,5 atribuímos o posto 5.

Embora a atribuição de postos seja diferente nesse caso, continuaremos a usar a mesma notação anterior para os postos das observações X_i e Y_i. A distribuição da estatística W_S não é mais dada por (13.17), pois os valores de $S_1, \ldots, S_m$ não são mais os anteriores. Retomemos o exemplo dado. Temos que a distribuição conjunta dos postos S_1 e S_2 será:

$$P(S_1 = 1, S_2 = 2{,}5) = 2/10, \qquad P(S_1 = 1, S_2 = 4) = 1/10,$$
$$P(S_1 = 1, S_2 = 5) = 1/10, \qquad P(S_1 = S_2 = 2{,}5) = 1/10,$$
$$P(S_1 = 2{,}5, S_2 = 4) = 2/10, \qquad P(S_1 = 2{,}5, S_2 = 5) = 2/10,$$
$$P(S_1 = 4, S_2 = 5) = 1/10,$$

pois ainda cada uma das $\binom{5}{2} = 10$ escolhas de dois dos postos médios como S_1 e S_2 são igualmente prováveis. Portanto a distribuição de $W_S = S_1 + S_2$ é dada pela Tabela 13.5.

Tabela 13.5: Distribuição de W_S, observações não-distintas.

w	3,5	5,0	6,0	6,5	7,5	9,0
$P(W_S = w)$	2/10	2/10	1/10	2/10	2/10	1/10

Observe que a distribuição da v.a. W_S nesse caso não é simétrica; será simétrica ao redor de $m(N + 1)/2$ se $n = m$.

Genericamente, o teste de Wilcoxon, no caso de observações empatadas, rejeita H_0 usando a mesma regra de decisão que no caso de observações não empatadas, exceto que a distribuição de W_S vai depender de n, m e dos números de observações empatadas em cada valor, ao contrário da situação de não empates, para a qual a distribuição de W_S depende somente de n e m.

Exemplo 13.8. Supondo $n = 3$, $m = 2$, as observações dos controles são 1,3, 1,5 e 2,1, e as observações dos tratamentos são 1,5 e 2,5. Então,

$$S_1 = 2{,}5, \quad S_2 = 5, \quad R_1 = 1, \quad R_2 = 2{,}5, \quad R_3 = 4 \quad \text{e} \quad W_S = S_1 + S_2 = 7{,}5.$$

Pelo que vimos acima, o valor-p será

$$\hat{\alpha} = P(W_S \geqslant 7{,}5) = 2/10 + 1/10 = 0{,}3,$$

logo não rejeitaremos H_0 nos níveis usuais.

Suponha que temos d_1 observações empatadas no menor valor, d_2 observações empatadas no segundo menor valor etc. até d_e observações empatadas no maior valor, onde e é o número de valores distintos. Denominamos $(e;\, d_1, \ldots, d_e)$ de *configuração de empates*, e a distribuição de W_S dependerá dessa configuração. Assim sendo, tabelas teriam de ser construídas para cada configuração de empates, o que não é prático. O que se faz é o seguinte: se o número de empates for pequeno, continue a usar a Tabela VIII. Caso contrário, use a aproximação normal. Nesse caso, a média de W_S é a mesma anterior, mas a variância é igual à anterior menos uma correção devida aos empates:

$$\operatorname{Var}(W_S) = \frac{mn(N+1)}{12} - \frac{mn}{12N(N-1)} \sum_{i=1}^{e} (d_i^3 - d_i). \qquad (13.25)$$

A aproximação normal será adequada se m e n forem relativamente grandes, e as proporções d_i/N não forem próximas de 1.

Exemplo 13.9. Em aparelhos dentários são usados grampos de dois tipos: um modelo em T e outro circunferencial, C. O objetivo é verificar se a resistência à remoção de grampos em T é a mesma do modelo C. Foram usados 40 corpos de provas (dente-grampo), sendo 20 para o modelo T e 20 para o modelo C, com cinco leituras para cada corpo de prova, num total de 100 observações para cada modelo. As Figuras 13.4 e 13.5 mostram os histogramas para os dois modelos, a resistência sendo medida em kg.

Figura 13.4: Resistência à remoção, em kg, para o modelo C.

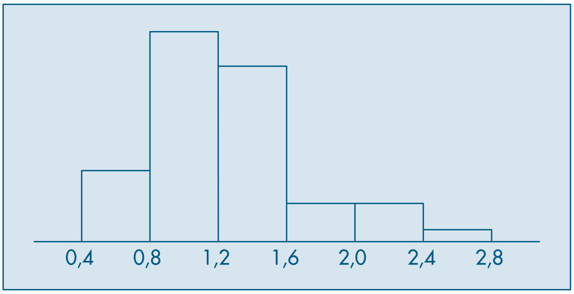
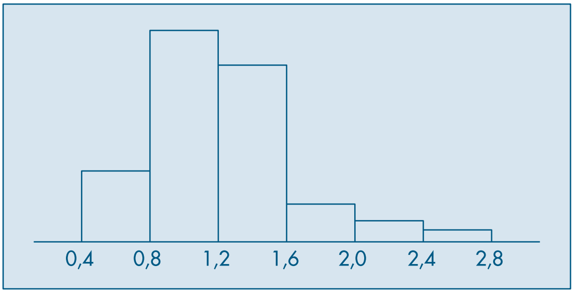


Figura 13.5: Resistência à remoção, em kg, para o modelo T.



Vemos que há assimetrias nos histogramas, sugerindo que a aplicação do teste t de Student não é adequada nessa situação. A Tabela 13.6 mostra as médias das 5 leituras para cada corpo de prova, para o modelo T e para o modelo C (em ordem crescente).

Admitamos que o grupo de controle seja aquele em que os grampos sejam do tipo T, e grampos do tipo C constituam o tratamento. Ordenando as médias da Tabela 13.6 e atribuindo postos obtemos a Tabela 13.7.

Tabela 13.6: Valores de resistência à remoção para os dois modelos.

T	C	T	C
0,60	0,52	1,19	1,19
0,63	0,77	1,20	1,20
0,83	0,79	1,26	1,34
0,85	0,79	1,28	1,36
0,91	0,81	1,30	1,38
0,95	0,81	1,37	1,43
1,01	0,89	1,45	1,64
1,03	0,98	1,54	1,71
1,03	1,01	1,68	2,16
1,16	1,18	2,20	2,25

Tabela 13.7: Postos para o Exemplo 13.9.

Média	0,52	0,60	0,63	0,77	0,79	0,79	0,81	0,81	0,83	0,85
Tipo	C	T	T	C	C	C	C	C	T	T
Posto	1	2	3	4	5,5	5,5	7,5	7,5	9	10
Média	0,89	0,91	0,95	0,98	1,01	1,01	1,03	1,03	1,16	1,18
Tipo	C	T	T	C	C	T	T	T	T	C
Posto	11	12	13	14	15,5	15,5	17,5	17,5	19	20
Média	1,19	1,19	1,20	1,20	1,26	1,28	1,30	1,34	1,36	1,37
Tipo	C	T	T	C	T	T	T	C	C	T
Posto	21,5	21,5	23,5	23,5	25	26	27	28	29	30
Média	1,38	1,43	1,45	1,54	1,64	1,68	1,71	2,16	2,20	2,25
Tipo	C	C	T	T	C	T	C	C	T	C
Posto	31	32	33	34	35	36	37	38	39	40

Aqui $n = m = 20$ e queremos testar

H_0 : a resistência à remoção é a mesma para os dois tipos de grampos;

H_1 : o tipo C apresenta menor resistência à remoção do que o do tipo T.

A soma dos postos dos tratamentos é

$$W_S = S_1 + S_2 + \ldots + S_{20} = 406{,}5.$$

Usando a aproximação normal, a v.a.

$$Z = \frac{W_S - E(W_S)}{\sqrt{\mathrm{Var}(W_S)}}, \tag{13.26}$$

onde $\mathrm{Var}(W_S)$ é dada por (13.25), e terá distribuição aproximadamente $N(0, 1)$. Consultando a Tabela 13.7, temos

$d_1 = d_2 = d_3 = d_4 = 1, \quad d_5 = 2, \quad d_6 = 2, \quad d_7 = \ldots = d_{12} = 1,$

$d_{13} = 2, \quad d_{14} = 2, \quad d_{15} = d_{16} = 1, \quad d_{17} = 2, \quad d_{18} = 2,$

$d_{19} = \ldots = d_{34} = 1.$

Aqui, temos $e = 34$ valores distintos e

$$E(W_S) = (20 \times 41)/2 = 410,$$

$$\mathrm{Var}(W_S) = (20 \times 20 \times 41)/12 - (20 \times 20)/(12 \times 40 \times 39)\,[(8 - 2) \times 6]$$

$$= 1.366{,}667 - 2{,}857 = 1.363{,}810.$$

O valor de (13.26) é

$$Z = (406{,}5 - 410)/36{,}93 = -0{,}095.$$

Como rejeitaremos H_0 se $W_S \leqslant c$, no nível $\alpha = 0{,}05$, devemos comparar esse valor com o valor $-1{,}64$ da normal padrão, portanto não rejeitamos H_0.

Vemos que o valor-p do teste é

$$\hat{\alpha} = P(W_S \leqslant 406{,}5) \approx P(Z \leqslant -0{,}095) = 0{,}46,$$

que é uma indicação de que a hipótese H_0 deve ser aceita.

Observação. Comparação entre o Teste t e o Teste de Wilcoxon.

O teste t baseia-se na suposição de que as populações P_1 e P_2 sejam normais. Uma violação dessa suposição altera a distribuição da estatística usada no teste e muda as probabilidades dos erros de tipo I e II. Dizemos que um teste é *robusto* contra a violação de uma suposição se suas probabilidades de erro de tipo I e II não são afetadas de forma apreciável pela violação.

Pode-se mostrar que o teste t é pouco sensível à heterogeneidade de variâncias se $m = n$, mas ele será mais afetado se as variâncias forem diferentes e $m \neq n$.

Os testes t e de Wilcoxon são comparados através de seus poderes em termos de uma quantidade chamada *eficiência relativa assintótica*, mas não entraremos em detalhes aqui sobre esse assunto. Mas podemos resumir a situação da seguinte maneira:

(a) O teste t é mais poderoso quando temos populações normais, mas a perda de eficiência do teste de Wilcoxon é pequena (menos de 5%) nesse caso;

(b) haverá pouca diferença entre os dois testes para distribuições próximas da normal;

(c) o teste de Wilcoxon é mais eficiente para distribuições que têm caudas "mais pesadas" do que a normal.

Para se ter uma idéia do que significa mais pesada, observamos que as distribuições t e Cauchy têm distribuições com caudas mais pesadas que a normal. Se P_1 e P_2 forem ambas uniformes, pode-se provar que os dois testes são igualmente eficientes e se P_1 e P_2 forem ambas exponenciais, o teste de Wilcoxon é três vezes mais eficiente.

Problemas

10. Vinte canteiros foram plantados com milho. Em dez deles um novo tipo de fertilizante foi aplicado, obtendo-se as produções abaixo. Há diferenças significativas entre as produções? A alternativa é que o novo fertilizante tende a produzir valores maiores. Tome $\alpha =$ 0,05. Calcule $\hat{\alpha}$.

Controle	7,1	6,0	8,0	7,0	6,6	7,4	7,0	7,0	6,9	6,8
Tratamento	6,9	6,8	7,5	6,8	6,9	6,8	6,8	6,8	6,7	6,6

11. Obtenha a distribuição nula de W_S para os casos:

    (a) $m = 2, n = 2$; (b) $m = 2, n = 4$; (c) $m = n = 3$.

12. Calcule as seguintes probabilidades, usando a Tabela VIII e a aproximação normal.

    (a) $m = 6, n = 7, P(W_S \leqslant 48)$
    (b) $m = 8, n = 10, P(W_S \leqslant 65)$
    (c) $m = 10, n = 10, P(W_S \geqslant 63)$

13. Encontre a distribuição nula de W_S no caso de empates, para os casos:

(a) $m = n = 3, d_1 = d_2 = 1, d_3 = 2, d_4 = d_5 = 1$
(b) $m = n = 3, d_1 = d_2 = d_3 = 2$
(c) $m = 2, n = 3, d_1 = d_2 = 1, d_3 = 3$

14. Faça os histogramas para W_S nos Problemas 11 e 13.

15. Suponha que as observações dos tratamentos sejam 3, 3, 5 e 7, e as observações dos controles sejam 1, 4 e 8, e que o teste de Wilcoxon rejeite para valores grandes de W_S. Calcule $\hat{\alpha} = P(W_S \geqslant w)$, onde w é o valor observado de W_S.

13.4 Comparação de Duas Populações: Amostras Dependentes

Na seção 13.1 já discutimos essa situação. Aqui, temos duas amostras $X_1, ..., X_n$ e $Y_1, ..., Y_n$, só que agora as observações são pareadas, isto é, podemos considerar que temos na realidade uma amostra de pares $(X_1, Y_1), ..., (X_n, Y_n)$. Se definirmos a v.a. $D = X - Y$, teremos a amostra $D_1, D_2, ..., D_n$, resultante das diferenças entre os valores de cada par. Observe que reduzimos a um problema com uma única população, conforme estudado nos capítulos anteriores.

Consideraremos dois casos: no primeiro, supomos que a população das diferenças é normal; no segundo, supomos que essa população é simétrica.

13.4.1 População Normal

Nessa situação, faremos a seguinte suposição: a v.a. D tem distribuição normal $N(\mu_D, \sigma^2_D)$. Podemos deduzir daqui que

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i = \frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i) = \bar{X} - \bar{Y} \tag{13.27}$$

terá distribuição $N(\mu_D, \sigma_D^2/n)$.

Considere

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2. \tag{13.28}$$

Pelo Teorema 7.1, a estatística

$$T = \frac{\sqrt{n}(\bar{D} - \mu_D)}{S_D} \tag{13.29}$$

terá distribuição t de Student, com $(n - 1)$ graus de liberdade.

Como

$$\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2,$$

qualquer afirmação sobre o $\mu_1 - \mu_2$ corresponde a uma afirmação sobre μ_D.

Exemplo 13.10. Cinco operadores de certo tipo de máquina são treinados em máquinas de duas marcas diferentes, A e B. Mediu-se o tempo que cada um deles gasta na realização de uma mesma tarefa, e os resultados estão na Tabela 13.8.

Tabela 13.8: Tempos para realização de tarefa para cinco operadores.

Operador	Marca A	Marca B
1	80	75
2	72	70
3	65	60
4	78	72
5	85	78

Com o nível de significância de 10%, poderíamos afirmar que a tarefa realizada na máquina A demora mais do que na máquina B?

Estamos interessados em testar

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A > \mu_B.$$

Essas hipóteses são equivalentes a

$$H_0 : \mu_D = 0$$
$$H_1 : \mu_D > 0.$$

Como é o mesmo operador que realiza a tarefa nas duas máquinas, estamos diante do caso em que se pode usar variáveis emparelhadas. Vamos admitir que, sob H_0, a diferença de tempo segue uma distribuição normal $N(0, \sigma^2_D)$. Logo, usamos a estatística (13.29).

Para determinar a região crítica, note que, devido à forma de H_1, devemos encontrar t_c tal que $P(T > t_c) = 0{,}10$, sendo que T tem distribuição $t(4)$. Usando a Tabela V, obtemos

$$\text{RC} =]1{,}54;\ +\infty[.$$

Da Tabela 13.8 obtemos os valores de D:

$$d_i : 5,\ 2,\ 5,\ 6,\ 7$$

e, portanto,

$$\bar{d} = 5 \quad \text{e} \quad s_D^2 = 3{,}5.$$

O valor observado da estatística T é $t_0 = (5/1{,}87)(\sqrt{5}) = 5{,}98$. Segue-se que rejeitamos H_0, ou seja, demora-se mais para realizar a tarefa com a máquina A.

Podemos construir um intervalo de confiança para μ_D; para $\gamma = 0{,}90$,

$$\text{IC}(\mu_A - \mu_B;\ 0{,}90) = \text{IC}(\mu_D;\ 0{,}90) = 5 \pm (2{,}13)(1{,}87)/\sqrt{5},$$

ou seja,

$$\text{IC}(\mu_D;\ 0{,}90) =]3{,}22;\ 6{,}78[.$$

13.4.2 População Não-Normal

Vamos considerar, agora, um teste baseado nos postos das diferenças D_i: o chamado *teste dos postos sinalizados de Wilcoxon*. Para esse teste, supomos que a escala das diferenças seja pelo menos intervalar e que os pares (X_i, Y_i) constituam uma AAS.

Isso implica, em particular, que os D_i são independentes, com a mesma mediana. Suponha, ainda, que cada D_i tenha uma distribuição simétrica. Ou seja, as médias e medianas coincidem.

Exemplo 13.11. Suponha que se possa simular um modelo por meio de duas linguagens computacionais, que chamaremos A e B. Supostamente, o tempo usando B é menor que o tempo usando A. Cinco pares de alunos são selecionados para o teste, de modo que cada membro de um par tenha a mesma habilidade computacional nas duas linguagens do que o outro. Um membro de cada par é escolhido ao acaso e este vai usar a linguagem B; o outro usará A. O tempo de simulação (em segundos) de cada linguagem é anotado, obtendo-se a Tabela 13.9.

Tabela 13.9: Tempos de simulação (em segundos) para as linguagens A e B.

Par	1	2	3	4	5
tempo de B(X)	300	410	420	410	400
tempo de A(Y)	350	390	490	435	440
$D = X - Y$	–50	20	–70	–25	–40
Posto de $\|D\|$	4	1	5	2	3
Posto sinalizado	–4	+1	–5	–2	–3

Queremos testar a hipótese de que os tempos são semelhantes contra a hipótese de que os tempos de B são menores. Ou, ainda,

$$H_0 : \mu_B - \mu_A = \mu_D = 0,$$
$$H_1 : \mu_B - \mu_A = \mu_D < 0.$$

Na quarta linha da Tabela 13.9 estão apresentadas as diferenças D_i, e os postos são calculados a partir das variáveis $|D_i|$, ou seja, os módulos (ou valores absolutos) dos D_i (quinta linha). A sexta linha, "posto sinalizado", é obtida atribuindo-se ao posto de $|D_i|$ o sinal correspondente de D_i. Por exemplo, para a primeira observação, $D_1 = 300 - 350 = -50$, com $|D_1| = 50$, que tem posto 4 e, portanto, posto sinalizado –4.

Notamos que só há um posto positivo, +1. Se indicarmos por T^+ a soma dos postos positivos, rejeitaremos H_0 se T^+ for "pequeno". É claro que podemos trabalhar com os postos negativos também, e considerar $T^- = -$ (soma dos postos negativos). No exemplo, $T^+ = 1$ e $T^- = 14$. Usando T^-, rejeitaremos H_0 se esta for "grande". Note que $T^+ + T^- = 15$, que é a soma de todos os postos dos $|D_i|$, que, por sua vez, é $n(n+1)/2$, sendo $n = 5$ o número de pares. Em geral, devemos usar a menor soma.

Trabalhemos com T^+. Para conduzir o teste, devemos obter a distribuição dessa estatística, sob a hipótese nula H_0. Para isso, note que, se H_0 for verdadeira, cada posto tem a mesma probabilidade de ser associado com um sinal + ou com um sinal –. Logo, a seqüência de postos sinalizados é uma de todas as possíveis combinações de ±1, ±2, ..., ±5. Há $2^5 = 32$ tais combinações, todas eqüiprováveis sob H_0, ou seja, com probabilidade 1/32 .

Na Tabela 13.10 temos todas as possibilidades juntamente com o valor de T^+. Na Tabela 13.11 temos a distribuição de T^+. Note que a distribuição de T^+ é simétrica, com média e mediana iguais a 7,5.

Tabela 13.10: Sinais possíveis para os postos, Exemplo 13.10.

1	2	3	4	5	T^+	1	2	3	4	5	T^+
+	+	+	+	+	15	+	+	−	+	−	7
−	+	+	+	+	14	−	+	−	−	+	7
+	−	+	+	+	13	−	−	+	+	−	7
+	+	−	+	+	12	+	−	−	−	+	6
−	−	+	+	+	12	+	+	+	−	−	6
+	+	+	−	+	11	−	+	−	+	−	6
−	+	−	+	+	11	+	−	−	+	−	5
+	+	+	+	−	10	−	+	+	−	−	5
−	+	+	−	+	10	−	−	−	−	+	5
+	−	−	+	+	10	+	−	+	−	−	4
−	+	+	+	−	9	−	−	−	+	−	4
−	−	−	+	+	9	+	+	−	−	−	3
+	−	+	−	+	9	−	−	+	−	−	3
+	+	−	−	+	8	−	+	−	−	−	2
+	−	+	+	−	8	+	−	−	−	−	1
−	−	+	−	+	8	−	−	−	−	−	0

Tabela 13.11: Distribuição de T^+ sob H_0.

T^+	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Freqüência	1	1	1	2	2	3	3	3	3	3	3	2	2	1	1	1

O valor-*p* do teste é $P(T^+ \leqslant 1|H_0) = 2/32 = 0{,}06$, usando a Tabela 13.11. Ou seja, há indicação de que o tempo de simulação usando a linguagem B é menor do que o tempo de A. Observe que temos poucos pares, e o valor $\hat{\alpha} = 0{,}06$ não é tão pequeno (reveja a Tabela 12.2). Mas como temos somente um posto positivo dentre cinco, somos levados a duvidar da validade de H_0.

Vejamos, agora, o caso geral. Tomemos os valores absolutos das diferenças, ou seja,

$$|D_i| = |X_i - Y_i|,\ i = 1, \ldots m.$$

Quando $X_i = Y_i$ omitir a diferença correspondente e seja n o número de diferenças estritamente diferentes de zero. Associemos a cada par (X_i, Y_i) o posto do módulo de D_i correspondente. Use postos médios, se houver D_i coincidentes.

A hipótese a ser testada é que a média (ou a mediana) das diferenças seja igual a zero contra a alternativa que não seja. Testes unilaterais podem, também, ser considerados. Ou seja, dada a simetria da distribuição dos D_i, iremos testar

$$H_0 : \mu_D = 0,$$
$$H_1 : \mu_D \neq 0,$$

onde μ_D representa, como antes, a média das diferenças.

Considere

$$R_i = \begin{cases} R(X_i, Y_i), & \text{se } D_i > 0, \\ -R(X_i, Y_i), & \text{se } D_i < 0, \end{cases} \qquad (13.30)$$

onde $R(X_i, Y_i)$ é o posto associado a (X_i, Y_i).

Temos dois casos a tratar:

(A) Se não houver empates, use a estatística

$$T^+ = \sum (R_i \text{ com } D_i > 0), \qquad (13.31)$$

ou seja, a soma dos postos positivos. Use a Tabela IX, pág. 506, para obter os quantis w_p da estatística, ou seja, o valor, tal que $P(T^+ < w_p) \leqslant p$ e $P(T^+ > w_p) \leqslant 1 - p$, se H_0 for verdadeira. Para $n > 50$ use a aproximação normal, com média e variância dados no teorema abaixo. Para $p > 0{,}5$ o quantil é dado por

$$w_p = \frac{n(n+1)}{2} - w_{1-p}.$$

(B) Se houver empates, use a estatística

$$V = \frac{\sum_{i=1}^{N} R_i}{\sqrt{\sum_{i=1}^{N} R_i^2}}, \qquad (13.32)$$

que tem uma distribuição aproximadamente $N(0,1)$, sob a hipótese nula.

Teorema 13.3. A média e variância de T^+ são dadas por

$$E(T^+) = \frac{n(n+1)}{4} \qquad (13.33)$$

e

$$Var(T^+) = \frac{n(n+1)(2n+1)}{24}, \qquad (13.34)$$

respectivamente.

Exemplo 13.11. (continuação) Obtivemos aqui $T^+ = 1$. A região crítica é unilateral à esquerda, logo rejeitamos H_0 se $T^+ < w_\alpha$, onde w_α é o quantil dado pela Tabela IX. Se fixarmos $\alpha = 0{,}025$ ou $\alpha = 0{,}01$, obteremos $w_\alpha = 0$, com $n = 5$, e, portanto, aceitaremos

H_0. Se α = 0,05, então w_α = 1, e o valor observado estará na fronteira da região crítica e teremos dúvidas em aceitar ou rejeitar H_0. Como salientamos antes, a decisão, nesse caso, dependerá de uma análise cuidadosa dos resultados, dado o pequeno valor de *n*.

13.5 Comparação de Proporções em Duas Populações

Nosso objetivo agora é a comparação das proporções de duas populações P_1 e P_2. Sendo mais explícitos, queremos comparar as proporções populacionais p_1 e p_2, por meio dos estimadores $\hat{p}_1$ e $\hat{p}_2$ obtidos de amostras independentes de tamanhos n_1 e n_2 respectivamente. Das seções 10.9 e 12.6 temos

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right), \ \hat{p}_2 \sim N\left(p_2, \frac{p_1(1-p_2)}{n_2}\right).$$

Comparando com o resultado da seção 13.3.1, e também do Problema 10.32, obtemos

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right),$$

e portanto, a estatística de decisão, tanto para a construção de intervalos de confiança como para testes de hipóteses, será

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \sim N(0,1).$$

Mas como os valores dos parâmetros são desconhecidos, substituem-se as variâncias pelas seus estimadores, obtendo-se, como visto em 13.3.1(b), uma distribuição aproximadamente t de Student. Entretanto, estudos envolvendo proporções utilizam amostras grandes e os valores da distribuição t aproximam-se de valores da normal padronizada. Desse modo, para comparação de duas proporções recomenda-se sempre o uso da estatística:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1). \qquad (13.35)$$

Exemplo 13.12: Para lançamento da nova embalagem do sabonete SEBO a divisão de criação estuda duas propostas:

A: amarela com letras vermelhas, ou

B: preta com letras douradas.

Eles acreditam que a proposta A chama a atenção em pelo menos 5% a mais do que a proposta B. Para verificar a validade de tal informação conduziu-se o seguinte experimento: em cada um de dois supermercados “semelhantes” foram colocados sabonetes com cada tipo de embalagem, e a clientes selecionados aleatoriamente foi perguntado se tinham notado o sabonete e que descrevessem qual a embalagem. Abaixo estão os resultados.

Proposta	Notaram?		Total
	Sim	Não	
A	168	232	400
B	180	420	600
Total	348	652	1000

Os resultados da pesquisa justificam ou não as suposições da divisão de criação? Aqui, consideramos

$$H_0\colon p_A - p_B = 0{,}05, \qquad H_1\colon p_A - p_B > 0{,}05.$$

Da tabela obtemos: $\hat{p}_1 = 0{,}42$ e $\hat{p}_2 = 0{,}30$, e aplicando a fórmula (13.35) obtemos:

$$Z = \frac{(0{,}42 - 0{,}30) - 0{,}05}{\sqrt{\dfrac{(0{,}42)(0{,}58)}{400} + \dfrac{(0{,}30)(0{,}70)}{600}}} = 2{,}26.$$

Consultando a Tabela III, encontramos o valor-*p* $\hat{\alpha} = 1{,}19\%$, o que leva a rejeição de H_0. O passo seguinte seria a construção de um Intervalo de Confiança, e novamente aplicado a expressão (13.35), obtém-se:

$$IC(p_A - p_B\colon 95\%) = (0{,}42 - 0{,}30) \pm 1{,}96\sqrt{\frac{(0{,}42)(0{,}58)}{400} + \frac{(0{,}30)(0{,}70)}{600}},$$

$$IC(p_A - p_B : 95\%) = 0{,}12 \pm 0{,}036 = [0{,}084; 0{,}156].$$

Para testar a hipótese de igualdade de proporções, $p_1 = p_2$, e usando as mesmas argumentações apresentadas na seção 13.3.1(a), deve-se usar uma estimativa comum das variâncias dada por $\hat{p}_c(1 - \hat{p}_c)$, onde $\hat{p}_c = (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$, resultando no teste:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_c(1 - \hat{p}_c)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0, 1) \qquad (13.36)$$

Exemplo 13.12 (continuação). Voltando ao problema do sabonete SEBO, suponha que eles não sabem se uma embalagem é ou não mais atraente do que outra, e a pesquisa foi feita para responder a essa questão. Portanto o teste agora será:

$$H_0\colon p_A = p_B, \quad H_1\colon p_A \neq p_B.$$

Da tabela obtemos $\hat{p}_c = (348/1000) = 0{,}348$, substituindo em (13.36), obtemos:

$$Z = \frac{0{,}42 - 0{,}30}{\sqrt{0{,}348(0{,}652)\left(\frac{1}{400} + \frac{1}{600}\right)}} = 3{,}90.$$

Consultando a Tabela III, encontramos valor-*p* próximo de zero, o que leva a rejeição de H_0. Como esse resultado mostra que as variâncias também são diferentes, a construção do Intervalo de Confiança é obtida do mesmo modo acima.

Problemas

16. Para investigar a lealdade de consumidores a um determinado produto, sorteou-se uma amostra de 200 homens e 200 mulheres. Foram classificados como tendo alto grau de fidelidade 100 homens e 120 mulheres. Os dados trazem evidências de diferença de grau de fidelidade entre os sexos? Em caso afirmativo construa um intervalo de confiança para a diferença.

17. Em uma amostra de 500 famílias da cidade A, constatou-se que 298 haviam comprado, durante os últimos 30 dias, o refrigerante Meca-Mela em sua nova versão incolor. Na cidade B esse número foi de 147 em 300 famílias entrevistadas. Na cidade A foi feita uma campanha publicitária através da rádio local, e não na cidade B. Os resultados trazem evidências de que as campanhas locais aumentam as vendas?

18. Um partido afirma que a porcentagem de votos masculinos a seu favor será 10% a mais que a de votos femininos. Em uma pesquisa feita entre 400 homens, 170 votariam no partido, enquanto que entre 625 mulheres, 194 lhe seriam favoráveis. A afirmação do partido é verdadeira ou não? Caso rejeite a igualdade, dê um IC para a diferença.

19. Para investigar os resultados do segundo turno de uma eleição estadual tomaram-se duas amostras de 600 eleitores cada: uma da capital e outra do interior. Da primeira, 276 disseram que votariam no candidato A, enquanto que 312 eleitores do interior também o fariam.
    (a) Estime a proporção de eleitores da capital que votariam em A. Dê um IC.
    (b) Existe diferença nas proporções entre capital e interior?
    (c) Que tamanho igual deveriam ter ambas as amostras para que a diferença entre as proporções fosse estimada com erro inferior a 2%?
    (d) Qual a proporção esperada de votos que irá receber o candidato A no estado?
    (e) De uma amostra de 120 indivíduos da classe A e B, 69 são favoráveis a eleição em dois turnos, enquanto que em uma amostra de 100 indíviduos da classe C, 48 é que são favoráveis. Existe evidência e diferenças de opiniões em relação à classe social?

20. Para verificar a importância de um cartaz nas compras de certo produto, procedeu-se do seguinte modo:
    (a) formaram-se sete pares de lojas;
    (b) os pares foram formados de modo que tivessem as mesmas características quanto à localização, ao tamanho e ao volume de vendas;
    (c) num dos elementos do par, colocou-se o cartaz; no outro, não;
    (d) as vendas semanais foram registradas, e os resultados estão a seguir.

    Qual seria a sua conclusão sobre a eficiência do cartaz? Use o teste *t*, fazendo as suposições necessárias.