4.7.3. Teste da diferença entre médias, amostras pareadas

4.7.3.1. Introdução

Vimos na seção anterior como comparar os resultado de dois tratamentos usando amostras *independentes*; veremos nesta seção outra forma de fazer estas comparações, usando amostras *pareadas*.

Suponha por exemplo que queiramos avaliar a eficácia de um medicamento para redução de peso, e planejamos um experimento para comparar as respostas de um grupo *tratamento* (pacientes que tomam o medicamento) com as de um grupo *controle* (pacientes que não tomam nada, tomam um placebo, ou tomam um outro medicamento já conhecido). Deve haver um grande número de fatores que tornam mais fácil ou mais difícil a um paciente perder peso; provavelmente fatores como sexo, idade, raça, hábito de fumar e nível de atividade física estarão atuando para ajudar ou dificultar o efeito do medicamento que está sendo avaliado. Para evitar que estes fatores influenciem na resposta, podemos fazer com que os dois grupos (tanto o *controle* quanto o *tratamento*) sejam compostos de pacientes nos quais estes fatores apareçam com a mesma frequência, pelo menos aproximadamente; por exemplo, se 20% dos pacientes do tratamento fumam, 20% dos pacientes do controle também fumam.

Uma maneira de fazer isto é usar amostras aleatórias *independentes*: simplesmente retiramos duas amostras aleatórias simples da população, sem nos preocuparmos em identificar os fatores que possam influenciar na resposta. O acaso se encarregará de fazer com que todos os fatores estejam representados nas duas amostras na mesma proporção em que estão nas populações. Este tipo de planejamento é bastante usado na prática, especialmente quando as amostras são grandes. Ele tem, contudo, suas desvantagens: uma vez que todos os fatores que influenciam a resposta estarão presentes nas amostras. de forma não-controlada, será mais difícil isolar que parcela da resposta se deve realmente ao tratamento, e que parcela se deve ao efeito de outros fatores.

Outro tipo de planejamento é aquele que usa amostras *pareadas*. Neste, cada elemento da Amostra 1 é associado a um elemento correspondente na Amostra 2, no qual todos fatores estejam presentes nos mesmos níveis. Se por exemplo acreditamos que os fatores *sexo*, *idade*, *raça*, *fumo* e *nível de atividade física* podem influenciar na resposta, organizamos as amostras de forma que, se um elemento da Amostra 1 é uma mulher, na faixa dos 40-50 anos de idade, branca, fumante e de atividade física moderada, ela será associada a um elemento da Amostra 2 que também terá estas mesmas características; estas duas mulheres formarão um *par* (por isso a amostra é chamada de *pareada*).

Os resultados do experimento são organizados como na Tab. 1. As variáveis X e Y medem as respostas ao tratamento, isto é, a perda de peso obtida em cada amostra. O que vai nos interessar, agora, não são mais as respostas x_i e y_i medidas em cada elemento das duas amostras, mas sim as diferenças $d_i = x_i - y_i$ observadas entre os elementos de cada par. A idéia é que, uma vez que os pacientes x_i e y_i são similares em todos os outros fatores, a única coisa que pode ter causado esta diferença nas respostas foi o tratamento, que é o que nos interessa verificar.

Tabela 1. Resultados de experimento com amostras pareadas

Par	Resposta ao Tratamento 1	Resposta ao Tratamento 2	Diferença
1	x_1	y_1	$d_1 = x_1 - y_1$
2	x_2	y_2	$d_2 = x_2 - y_2$
...	...	...	...
n	x_n	y_n	$d_n = x_n - y_n$
médias	$\overline{X}$	$\overline{Y}$	$\overline{d}$

4.7.3.2. Realização do teste

Quando as amostras são pareadas, a diferença entre as médias das populações X e Y,

$$\delta = \mu_X - \mu_Y$$

não será mais estimada pela diferença D entre as médias das amostras

$$D = \overline{X} - \overline{Y}$$

e sim pela *média das diferenças* calculadas em cada par:

$$\overline{d} = \frac{1}{n}\sum d_i$$

O desvio-padrão das diferenças d calculadas em cada par de observações será dado por

$$s_d = \sqrt{\frac{\sum (d_i - \overline{d})^2}{n-1}}$$

A hipótese nula é a de que a diferença entre as médias das duas populações é igual a um valor especificado δ (delta):

$$H_0: \quad \mu_1 - \mu_2 = \delta$$

Se as diferenças d_i entre cada par têm distribuição normal (este é um pressuposto importante!), a variável de teste t poderá ser calculada por:

$$t = \frac{\overline{d} - \delta}{s_d / \sqrt{n}} \tag{1}$$

A variável t neste caso terá distribuição de Student, com número de graus de liberdade dado por

$g.l. = n - 1$ onde n é o número de pares

O que o teste com amostras pareadas faz, em suma, é tratar a média $\overline{d}$ do mesmo jeito em que trata uma média $\overline{X}$ qualquer. Note que a fórmula para calcular a estatística t acima é a mesma usada no teste de média que vimos anteriormente (na seção 4.7.1.1):

$$t = \frac{\overline{X} - \mu}{s_{corr} / \sqrt{n}} \tag{2}$$

Se compararmos as eq. (1) e (2), veremos que para o teste da diferença com amostras pareadas, apenas substituímos $\overline{X}$ por $\overline{d}$, e trocamos a média da população especificada na hipótese nula de δ para μ.

4.7.3.3. Exemplo

Pesquisadores desejam verificar se um tratamento contra enxaqueca tem o efeito colateral não desejado de reduzir a pressão sanguínea das pacientes. Para isto, retiram uma amostra de 15 mulheres da mesma faixa etária e medem a pressão de cada uma delas; em seguida, aplicam a elas o tratamento durante um mês e medem novamente as pressões ao final deste período. Verifique se os resultados obtidos confirmam a hipótese de redução de pressão. A Tab. **2** mostra as pressões *antes* e *depois* do tratamento, e as diferenças *d* medidas em cada paciente.

Tabela 2. Pressões medidas em 15 pacientes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
antes	75	80	81	67	81	77	74	72	72	82	70	77	68	70	78
depois	58	65	81	72	72	70	71	68	80	75	57	74	65	56	67
d	−17	−15	0	5	−9	−7	−3	−4	8	−7	−13	−3	−3	−14	−11

A hipótese que nos interessa testar é a de que a pressão *depois* do tratamento é menor do que a pressão *antes* do tratamento. Esta hipótese pode ser escrita em termos da diferença δ nas populações da forma:

$$\delta = \mu_{depois} - \mu_{antes}$$
$$H_0: \quad \delta \geq 0$$
$$H_1: \quad \delta < 0$$

Usaremos para o teste as diferenças *d=depois–antes* medidas em cada paciente da amostra. As estatísticas calculadas para esta variável são:

número de pares	n = 15
média das diferenças	$\overline{d} = -6{,}2$
desvio-padrão das diferenças	$s_d = 7{,}24$

(i) Teste dos pressupostos

Antes de fazer o teste, devemos verificar o pressuposto da normalidade da distribuição das diferenças *d*. A Fig. **1** mostra o gráfico de separatrizes de *d*; os pontos estão razoavelmente alinhados ao longo de uma reta. O teste de normalidade de Shapiro dá como resultado um valor-p igual a 0.7524, o que também indica que a hipótese nula (de normalidade) não deve ser rejeitada. (O teste de Kolmogorov-Smirnov não pode ser usado aqui, pois há valores repetidos na variável *d*).

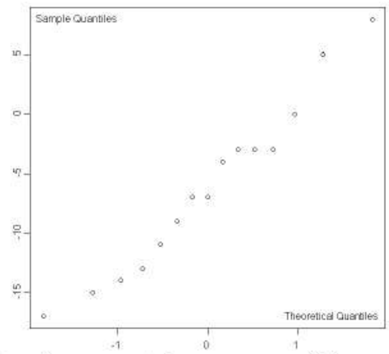

Figura 1. Gráfico de separatrizes para as diferenças em cada par

(ii) Cálculo da variável de teste t

Podemos portanto prosseguir com o teste. A estatística de teste *t* será calculada por:

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}} = \frac{-6{,}2 - 0}{7{,}24 / \sqrt{15}} = -3{,}32$$

O número de graus de liberdade *g.l.* desta estatística será dado por:

$g.l. = n - 1 = 15 - 1 = 14$

Usaremos um nível de significância α=0,05. Como o teste é unilateral, a área de rejeição, de probabilidade 0,05, deverá ser situada apenas do lado esquerdo da curva; na tabela, portanto, devemos procurar o valor crítico equivalente a um teste bilateral com α=0,10 (ou seja, um teste que coloque 0,05 de cada lado da curva). Na Tab. **3** (reprodução parcial da tabela de Student), encontramos para 14 graus de liberdade o valor crítico

$t_c = -1{,}76$.

Tabela 3. Probabilidades na distribuição de Student (parte)

	α													
g.l.	0,995	0,99	0,975	0,95	0,9	0,75	0,5	0,25	0,1	0,05	0,025	0,01	0,005	0,001
13	0,01	0,01	0,03	0,06	0,13	0,33	0,69	1,20	1,77	2,16	2,53	3,01	3,37	4,22
14	0,01	0,01	0,03	0,06	0,13	0,32	0,69	1,20	**1,76**	2,14	2,51	2,98	3,33	4,14
15	0,01	0,01	0,03	0,06	0,13	0,32	0,69	1,20	1,75	2,13	2,49	2,95	3,29	4,07

Conclusão: O valor da estatística *t* calculado na amostra é bem menor do que o valor crítico:

$t = -3{,}32 \ < \ t_c = -1{,}76$

Rejeitamos portanto a hipótese nula e aceitamos a hipótese alternativa: há evidências nos dados de que o tratamento realmente reduz a pressão média das pacientes. O cálculo feito no R, aliás, dá para este resultado um valor-p muito significativo, *p-value* = 0,005.

(iii) Comentários

Testes feitos com amostras pareadas têm em geral maior poder do que os feitos com amostras independentes. No exemplo acima, se analisarmos os dados sem levar em conta o pareamento (isto é, se fazemos as contas como se as duas amostras fossem independentes), obteremos um valor-p = 0.015, maior do que o do teste pareado; seria significativo se considerarmos $\alpha=0,05$, mas não se considerarmos $\alpha=0,01$. A razão disto é fácil de entender intuitivamente: no exemplo, a pressão sanguínea de uma mulher pode depender de sua idade, dieta, raça, etc.; se fazemos o teste pareado, o efeito de todos estes fatores será eliminado (dizemos que os fatores foram *controlados*), e o efeito que restar só pode ter sido causado pelo tratamento.

Testes pareados, porém, dão mais trabalho para organizar, pois para cada elemento de um par será preciso encontrar um correspondente no qual os mesmos fatores estejam presentes, o que geralmente não é fácil. Os testes *antes / depois* são uma maneira simples de resolver este problema – já que os elementos são os mesmos nas duas amostras, todos os fatores já estarão controlados. (Será preciso porém garantir que eles não sejam modificados durante o experimento; os pacientes não podem subitamente mudar seus hábitos, dietas, etc., porque isto afetaria o resultado). Este tipo de planejamento é por isto muito usado em pesquisas na área de Saúde, pois freqüentemente os experimentos são feitos com amostras pequenas, e o número de fatores que podem afetar o resultado é muito grande.

Resumo

1. Amostras pareadas são aquelas nas quais os elementos estão organizados em *pares*: para cada elemento de uma amostra há um correspondente na outra amostra que tem características similares. Se os elementos são pacientes em um experimento, por exemplo, os dois elementos de um par serão pacientes de mesmo sexo e idade, mesmo tipo de doença, no mesmo estágio, etc.
2. Testes feitos com amostras pareadas têm em geral maior poder do que os feitos com amostras independentes. Uma vez que os elementos de cada par têm basicamente as mesmas características, se houver alguma diferença entre eles na variável resposta, esta diferença só pode ter sido causada pelos tratamentos diferentes a que estes elementos foram submetidos.
3. Conceitos novos apresentados: *grupo controle, grupo tratamento*.