4.7.2. Teste da diferença entre médias, amostras pequenas independentes

---

Uma das tarefas mais comuns em Estatística é a de comparar as diferenças entre as médias de duas populações, a partir de duas amostras pequenas. Isto acontece, por exemplo, quando queremos comparar a produção de leite de duas raças de vacas, ou o desgaste de duas marcas de pneus, ou a capacidade aeróbica de pessoas que vivem em dois lugares de altitudes diferentes. Cada "população", nestes casos, conterá vacas de uma mesma raça, pneus de uma mesma marca, ou pessoas que vivem numa mesma altitude.

Técnicas estatísticas também são usadas para comparar os efeitos de dois *tratamentos* aplicados a uma mesma população. "Tratamento", aqui, é um termo genérico, aplicado a qualquer tipo de intervenção feitas sobre os elementos de uma amostra para produzir uma resposta desejada. Podemos estar interessados, por exemplo, em comparar os efeitos de dois medicamentos para controlar a pressão sanguínea de um paciente hipertensos, ou duas técnicas para aumentar a dureza de um certo tipo de aço, ou dois métodos de treinamento físico para aumentar a capacidade aeróbica de jogadores de futebol; os medicamentos, as técnicas de fabricação, e os métodos de treinamento, serão todos chamados genericamente de "tratamentos". O que é medido nas amostras é chamado genericamente de "resposta"; a pressão sanguínea, a dureza do aço ou a capacidade aeróbica. O objetivo do experimento é inferir quais seriam as respostas se os tratamentos fossem aplicados a populações teóricas compostas de todos os pacientes hipertensos, todo o aço daquele tipo, ou todos os jogadores de futebol.

Há duas maneiras de fazer este teste. A primeira, usada quando as duas amostras são *independentes* (como veremos nesta seção); a segunda, quando elas são *pareadas*, isto é, dependentes entre si (como veremos na seção 4.7.3).

Suponha, por exemplo, que queremos comparar o efeito de dois tipos de ração A e B sobre o crescimento de animais de certa raça, usando amostras *independentes*. Submetemos uma amostra destes animais ao tratamento A (isto é, os alimentamos com a ração A), a outra ao tratamento B (ração B), e medimos a variável de resposta que nos interessa nas duas amostras; por exemplo, o peso ao final do experimento. Em seguida, calculamos as médias destes pesos em cada amostra. A partir da diferença entre as médias calculadas nas duas amostras, podemos *testar* se a ração A faz as cobaias terem pesos significativamente maiores do que a ração B (como veremos a seguir), ou *estimar* a diferença que ocorreria nos pesos se estas rações fossem aplicadas na população destas cobaias (seção 4.8).

4.7.2.1. Distribuição amostral das diferenças de médias

Se retiramos amostras aleatórias simples de duas populações normais, a diferença entre as suas médias será uma variável também de distribuição normal, como afirma o Teorema 1.

Teorema 1: Distribuição amostral da diferença de médias, amostras pequenas de populações normais
Se duas populações normais têm médias μ_1 e μ_2 e desvios-padrões σ_1 e σ_2, e delas retiramos amostras aleatórias simples de tamanho n_1 e n_2, a diferença entre as médias destas amostras

$$D = \overline{X}_1 - \overline{X}_2$$

será uma variável com distribuição normal,

$$D \sim N(\mu_D, \sigma_D^2) \quad \text{cujos parâmetros são:} \quad \mu_D = \mu_1 - \mu_2 \quad \text{e} \quad \sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Portanto, podemos *padronizar* a variável D, transformando-a numa variável de teste Z de distribuição normal padrão, $Z \sim N(0,1)$, por meio de:

$$Z = \frac{D - \mu_D}{\sigma_D} = \frac{D - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{1}$$

Na maioria das aplicações, porém, não iremos conhecer os valores das variâncias σ_1 e σ_2. Se as amostras forem grandes, podemos simplesmente usar as variâncias das amostras s_1 e s_2 como estimativas de σ_1 e σ_2 (foi o que fizemos na seção 4.6.1). Quando as amostras forem pequenas, contudo, a substituição dos valores corretos σ por suas estimativas s irá aumentar a incerteza dos resultados; a variável de teste, ao invés de ter distribuição normal, passará a ter distribuição de Student, como afirma o Teorema **2** abaixo.

Teorema 2. Distribuição amostral da diferença de médias, amostras pequenas de populações normais de variâncias desconhecidas
Se duas populações normais têm médias μ_1 e μ_2 e variâncias desconhecidas, e delas retiramos amostras aleatórias simples pequenas de tamanho n_1 e n_2, a variável de teste t, com distribuição de *Student*, será calculada por meio de:

$$t = \frac{D - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{2}$$

onde s_1^2 e s_2^2 são as variâncias das duas amostras, e D é a diferença entre suas médias $D = \overline{X}_1 - \overline{X}_2$,

Note que a eq. (**2**) é similar à eq. (**1**), mas os desvios-padrões σ das populações foram substituídos pelos desvios-padrões s das amostras; por causa disto, a variável de teste resultante não é mais Z (normal) e sim t (Student).

Nas aplicações do teste, frequentemente é incluído mais um pressuposto, além da normalidade: o de que as variâncias das duas populações sejam iguais, isto é,

$$\sigma_1 = \sigma_2 = \sigma$$

Isto simplifica o problema, porque em vez de fazermos estimativas de dois desvios-padrões σ_1 e σ_2, faremos de apenas de um σ. Para fazer esta estimativa, podemos combinar as duas variâncias das amostras por meio de uma média ponderada, calculada pela eq. (**3**).

$$s_{comb} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{3}$$

Apesar de esta fórmula parecer complicada, todos os valores nela são conhecidos; o que ela faz é ponderar o desvio-padrão de cada amostra pelo tamanho da amostra (a amostra de maior n recebe maior peso). O valor s_{comb} é chamado de "desvio-padrão combinado".

Se aceitamos este pressuposto da igualdade das variâncias, calculamos a estatística de teste t por meio da eq. (**4**).

$$t = \frac{D - (\mu_1 - \mu_2)}{s_{comb}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{4}$$

A vantagem de fazer esta pressuposição é que a variável t neste caso passará a ter mais graus de liberdade. Na eq. (**4**), o número de graus de liberdade é dado por:

$$g.l. = n_1 + n_2 - 2 \tag{5}$$

Na eq. (**2**), o número é dado por:

$$g.l. = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{6}$$

O número de graus de liberdade calculado na eq. (**5**) é em geral maior do que o calculado na eq. (**6**). Isto significa que, quando aceitamos o pressuposto de que as variâncias são iguais, a distribuição de t estará mais aproximada da normal e seus valores críticos serão menores.

4.7.2.2. Verificação dos pressupostos

Antes de se fazer o teste, é importante verificar se os dois pressupostos foram atendidos. O primeiro é de que as duas populações tenham distribuições normais; como vimos na seção 4.7.1.2, esta verificação pode ser feita por meio de gráficos como o *diagrama de quantis*, ou por meio de testes como o de Kolmogorov-Smirnov ou o de Shapiro.

O segundo pressuposto é o de que as variâncias das duas populações sejam iguais, pelo menos aproximadamente. Como regra prática, alguns autores consideram que é permissível considerar que as duas variâncias populacionais (σ_1^2 e σ_2^2) sejam iguais, se a razão entre as duas variâncias amostrais (s_1^2 e s_2^2) for menor que 4 (isto é, uma delas for menos de 4 vezes maior que a outra). É possível, também, fazermos um *teste de igualdade de variâncias*, como o *teste de Levene*; nestes testes a hipótese nula é a de que as duas variâncias são iguais (estes testes serão vistos na seção 5.6).

Um opção simples é a de fazermos os dois testes de diferença de médias, um considerando as variâncias populacionais iguais, e outro considerando as variâncias diferentes. Como os programas atualmente permitem que os cálculos sejam feitos facilmente, calculamos das duas formas, e depois comparamos os resultados. Na maioria das vezes, eles irão concordar entre si (ou ambos os testes rejeitam H_0, ou ambos não rejeitam); apenas em raras ocasiões um deles (geralmente o que considera as variâncias iguais) irá rejeitar H_0, e o outro não.

4.7.2.3. Exemplo

O banco de dados *morley* (no pacote *datasets* do R) contém resultados dos experimentos feitos por Michelson e Morley para medir a velocidade da luz, em 1879. Foram feitos cinco experimentos, e em cada um deles 20 medições da velocidade. Iremos comparar neste exemplo as médias obtidas nos dois primeiros experimentos, para verificar se existe diferença significativa entre elas. A Fig. 1 mostra os diagramas de ramo-e-folhas dos resultados (para simplificar a notação, as velocidades foram registradas subtraindo-se 299.000 km/s)

```
 6 | 5                 7 | 699
 7 | 46                8 | 0001334
 8 | 1558              8 | 58888
 9 | 033566888         9 | 044
10 | 0007              9 | 66
```

Experimento 1   Experimento 2

Figura 1. Resultados de dois experimentos

As estatísticas obtidas nestas duas amostras foram as seguintes:

- tamanhos das amostras	$n_1 = 20$	$n_2 = 20$
- médias das amostras	$\overline{X}_1 = 909$	$\overline{X}_2 = 856$
- desvios-padrões das amostras	$s_1 = 104{,}93$	$s_2 = 61{,}16$

(Note que as velocidades médias de 909 e 856 significam na verdade 299,909 km/s e 299,856 km/s, respectivamente).

(i) Verificação dos pressupostos

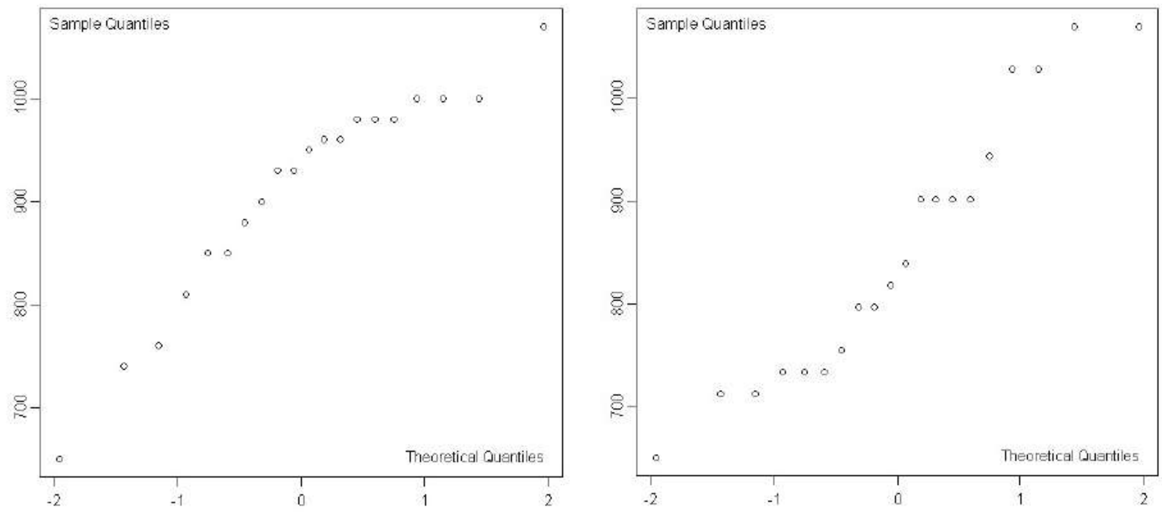
Antes de fazer o teste, temos que verificar se os pressupostos foram atendidos. O primeiro é o da normalidade das distribuições das populações. A Fig. 2 mostra os gráficos dos quantis, para as amostras resultantes dos dois experimentos. Em ambos, os pontos se localizam razoavelmente perto de uma reta, o que sugere que as populações sejam normais, pelo menos aproximadamente.

O teste de normalidade de Kolmogorov-Smirnov não pode ser usado neste exemplo, porque ambas as amostras têm valores repetidos. No teste de Shapiro, os valores-p encontrados para as duas amostras foram:

p-value = 0.099   p-value = 0.167

Estes valores indicam que a hipótese nula (de normalidade) não pode ser rejeitada, em nenhum dos experimentos.

O segundo pressuposto é o de que as variâncias das populações sejam iguais. Como a razão entre os desvios-padrões das amostras foi menor que 2, podemos aceitar este pressuposto.



Experimento 1 Experimento 2

Figura 2. Gráfico dos quantis, para os dois experimentos

(ii) Realização do teste de hipótese

As hipóteses nula e alternativa para este teste são:

$$H_0: \quad \mu_1 - \mu_2 = 0$$
$$H_1: \quad \mu_1 - \mu_2 \neq 0$$

O teste será portanto bilateral. Se supomos que as variâncias populacionais são iguais, o número de graus de liberdade da variável de teste t será calculado pela eq. (5):

$$g.l. = n_1 + n_2 - 2 = 38$$

Para um nível de significância α=0,05, procuraremos na tabela o valor crítico correspondente a 40 graus de liberdade (o mais próximo de 38); este valor é $|t_c|$=2,02 (em módulo, porque o valor pode ser positivo ou negativo). A Tab. 1 mostra parte da tabela de distribuição de Student.

Tabela 1. Valores críticos de *t* na distribuição de Student

	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
30	0,01	0,01	0,03	0,06	0,13	0,32	0,68	1,17	1,70	**2,04**	2,36	2,75	3,03	3,65
40	0,01	0,01	0,03	0,06	0,13	0,32	0,68	1,17	1,68	**2,02**	2,33	2,70	2,97	3,55
50	0,01	0,01	0,03	0,06	0,13	0,32	0,68	1,16	1,68	2,01	2,31	2,68	2,94	3,50
60	0,01	0,01	0,03	0,06	0,13	0,32	0,68	1,16	1,67	2,00	2,30	2,66	2,91	3,46
120	0,01	0,01	0,03	0,06	0,13	0,32	0,68	1,16	1,66	1,98	2,27	2,62	2,86	3,37
∞	0,01	0,01	0,03	0,06	0,13	0,32	0,67	1,15	1,64	1,96	2,24	2,58	2,81	3,29

Os desvios-padrões das amostras, s_1 e s_2, têm que ser combinados por meio da média ponderada na eq. (3), para fornecerem a estimativa do desvio-padrão da população:

$$s_{comb} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(20-1)\times 104{,}93^2 + (20_2-1)\times 61{,}16^2}{20+20-2}} = 85{,}88$$

O valor da estatística de teste t pode então ser calculado usando a eq. (**4**):

$$t = \frac{D-(\mu_1-\mu_2)}{s_{comb}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{53-0}{85{,}88\sqrt{\frac{1}{20}+\frac{1}{20}}} = 1{,}95$$

Como este valor de t é menor do que o valor crítico

$$t = 1{,}95 \ < \ t_c = 2{,}02$$

Concluímos que a hipótese nula não pode ser rejeitada; não há portanto evidências de que os resultados do primeiro experimento sejam significativamente diferentes, em média, dos resultados do segundo experimento.

Poderíamos também ter feito o teste sem aceitar o pressuposto de que as variâncias das populações sejam iguais. Neste caso, o valor da variável de teste t deve ser calculado pela eq. (**2**):

$$t = \frac{D-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{53-0}{\sqrt{\frac{104{,}93^2}{20}+\frac{61{,}16^2}{20}}} = 1{,}95$$

O número de graus de liberdade de t, agora, tem a expressão mais complicada na eq. (**6**):

$$g.l. = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{104{,}93^2}{20}+\frac{61{,}16^2}{20}\right)^2}{\frac{\left(\frac{104{,}93^2}{20}\right)^2}{20-1}+\frac{\left(\frac{61{,}16^2}{20}\right)^2}{20-1}} = 30{,}58$$

Usando o mesmo nível de significância α=0,05, procuramos na tabela o valor crítico correspondente a 30 graus de liberdade (o mais próximo de 30,58); este valor é $|t_c|$=2,04 (em módulo, porque o valor pode ser positivo ou negativo).

Como o valor calculado de t é menor do que este valor crítico,

$$t = 1{,}95 \ < \ t_c = 2{,}04$$

Concluímos novamente que a hipótese nula não pode ser rejeitada. Os dois testes, portanto, chegam à mesma conclusão.

(iii) Comentários

Vale a pena fazer alguns comentários sobre este experimento e seus resultados. Primeiro, qual sua finalidade? Para que serve comparar a velocidade da luz em diferentes direções?

Na época (final do século XIX), um dos problemas da Física era explicar como a luz se propaga no vácuo. De acordo com a teoria ondulatória, a luz é uma vibração (uma onda), da mesma forma que o som. O som que ouvimos nada mais é do que a sensação causada pela propagação de ondas de pressão em algum material - no ar, na água (se estivermos mergulhados), ou num sólido. O som não pode se propagar no vácuo; se um asteróide atinge a Lua e explode, por exemplo, podemos ver da Terra o brilho da explosão, mas não ouvimos seu som. Se a luz também é uma onda, como é que podemos ver a luz do Sol, se a maior parte da distância entre o Sol e a Terra não é preenchida por nenhum material, mas é simplesmente vácuo?

Para explicar isto, foi criada no século XIX a teoria de que todo o espaço é preenchido por uma substância invisível, chamada de "éter luminífero", na qual vibram as ondas luminosas. A idéia do experimento de Michelson e Morley foi: se existe mesmo este éter, a velocidade da luz deve depender da direção na qual estamos nos movendo. Se a Terra se move na direção do Sol, a luz deverá chegar à Terra com maior velocidade; se a Terra se afasta do Sol, a luz deverá chegar com menor velocidade. É o que acontece com o som: o som que ouvimos quando uma ambulância vem em nossa direção com a sirene ligada é mais agudo do que o que ouvimos quando a ambulância se afasta de nós, porque a velocidade do som nas duas direções é diferente (isto é chamado de "efeito Doppler"). Michelson e Morley fizeram várias medições da velocidade da luz, em direções ortogonais, e o resultado médio foi sempre o mesmo; isto destruiu a teoria do éter luminífero, e abriu caminho para a Teoria da Relatividade, proposta por Einstein, que considera que a velocidade da luz é sempre constante, mas a massa e o tempo são variáveis.

Segundo, a velocidade da luz pode ser constante, mas os resultados das medições sempre são variáveis, como podemos ver na Fig. **1**. Esta é a idéia básica da Teoria dos Erros, proposta por Gauss: existe um valor definido para uma grandeza, mas só podemos conhecê-lo por meio de medições – e toda medição é sujeita a erros. Sempre é preciso, portanto, usar Estatística para tirar alguma conclusão sobre o valor real desta grandeza.

Terceiro, apesar de sempre existir erro, as medições na Física são extremamente precisas. Nas duas amostras usadas neste exemplo, os coeficientes de variação das medições (feitas a 150 anos atrás) foram de apenas 0.035 % e 0.020 % (por isso, a Física é uma das ciências antes chamadas de "exatas").

Por último: neste teste, comparamos as medições da velocidade feitas em apenas duas direções. Existe uma técnica estatística, a *Análise de Variância* (ANOVA), que permite compararmos simultaneamente as medições feitas em diversas direções; veremos isto na seção 5.3.

Resumo

1. Para compararmos as médias de duas populações por meio de duas amostras pequenas independentes, a variável de teste t, com distribuição de Student, pode ser calculada por meio da eq. (2), desde que as duas populações tenham distribuição normal. É preciso verificar se este pressuposto da normalidade das populações é atendido, por meio de gráficos ou de testes de normalidade.
2. Se acrescentarmos o pressuposto de que as variâncias das duas populações são iguais, o valor de t pode ser calculado por meio da eq. (4). Este teste em geral tem mais poder do que o feito sem considerar este pressuposto, porque o número de graus de liberdade envolvidos é maior. Se aceitamos este pressuposto, precisamos verificar se ele foi atendido, antes de fazer o teste.
3. Conceitos novos apresentados: tratamento, variável resposta.