4.7.1. Teste de média com amostras pequenas

Neste capítulo, veremos como fazer testes de hipóteses sobre médias e diferenças de médias usando amostras pequenas. O trabalho com amostras grandes é mais simples (em termos da matemática envolvida) e dá resultados mais precisos; contudo, amostras pequenas são muito usadas em várias áreas, como as engenharias e as ciências da saúde, porque freqüentemente é caro e trabalhoso conseguir amostras maiores.

4.7.1.1. A distribuição de *Student*

Na seção 4.5.3 vimos o Teorema 1, que afirma que quando amostras aleatórias simples são retiradas de uma população com distribuição normal, a média $\overline{X}$ destas amostras também será uma variável com distribuição normal. Reproduzimos abaixo este teorema:

Teorema 1. Distribuição das médias $\overline{X}$ de amostras de população normal
Se amostras aleatórias simples de tamanho n (qualquer) são retiradas de uma população normal de média μ e desvio-padrão σ, a média amostral $\overline{X}$ será uma variável com distribuição normal,

$$\overline{X} \sim N(\mu_{\overline{X}}, \sigma^2_{\overline{X}})$$

cujos parâmetros são $\mu_{\overline{X}} = \mu$ e $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Podemos portanto *padronizar* a variável $\overline{X}$, transformando-a numa variável de teste Z de distribuição normal padrão:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma} \qquad \text{onde } Z \sim \mathrm{N}(0,1) \qquad (1)$$

Note que este teorema faz duas exigências importantes: primeiro, que a distribuição da população seja *normal*; segundo, que o desvio-padrão σ desta distribuição seja conhecido. Nas aplicações práticas, é claro, estas exigências causam problemas: nem sempre a população será normal, e o desvio-padrão em geral será desconhecido. A solução então é usar amostras grandes, que resolvem os dois problemas. O primeiro, por que há outro teorema, o *Teorema do Limite Central*, que diz que a distribuição da média amostral $\overline{X}$ *tende* para a normal, quando cresce o tamanho da amostra, mesmo que a população não seja normal. O segundo, porque se a amostra for grande, o seu desvio-padrão s poderá ser usado como estimativa razoável do desvio-padrão σ da população. Note que a inferência exige então duas aproximações: que a distribuição de $\overline{X}$ seja aproximada pela normal; que o desvio-padrão da população seja aproximado pelo da população. Estas aproximações funcionam satisfatoriamente, se a amostra for grande (digamos, maior que n=50).

O que fazer, porém, quando as amostras são pequenas? Neste caso, é mais difícil resolver estes problemas. A exigência de que a distribuição da população seja normal não pode ser relaxada; se a população não for normal, a distribuição de $\overline{X}$ não poderá ser aproximada pelo modelo normal.

Além disso, pode ser demonstrado que o desvio-padrão s usual, dado por:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n}} \tag{2}$$

não é um bom estimador do desvio-padrão σ da população; em vez dele, temos que usar o *desvio-padrão corrigido*, definido por:

$$s_{corr} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}} \tag{3}$$

Mesmo assim, este desvio-padrão corrigido ainda não é uma aproximação muito boa do desvio-padrão σ, e faz com que a distribuição amostral de $\overline{X}$ siga um modelo diferente do normal. O problema agora é descobrir qual é este modelo.

A solução foi encontrada por William S. Gosset, que mostrou que, para amostras pequenas extraídas de população normal, a média amostral $\overline{X}$ pode ser transformada numa variável t cuja distribuição segue um modelo conhecido como *modelo de Student*. Sua distribuição é unimodal e simétrica, com forma parecida com a da normal, e se caracteriza por ter apenas um parâmetro, chamado de *número de graus de liberdade*. (Representamos este parâmetro por suas iniciais *g.l.*). Isto pode ser resumido no Teorema **2**.

Teorema 2. Estatística de teste *t*, com distribuição de Student
Se amostras aleatórias simples de tamanho n são retiradas de uma população de distribuição normal de média μ, a estatística de teste t definida como

$$t = \frac{\overline{X} - \mu}{s_{corr} / \sqrt{n}}$$ onde $\overline{X}$: média da amostra s_{corr}: desvio-padrão corrigido

terá distribuição de *Student*, com $g.l. = n - 1$ (4)

A Fig. **1** mostra distribuições de Student para três números de graus de liberdade (*g.l.* = 1, 2 e 5), comparadas com a distribuição normal padrão.

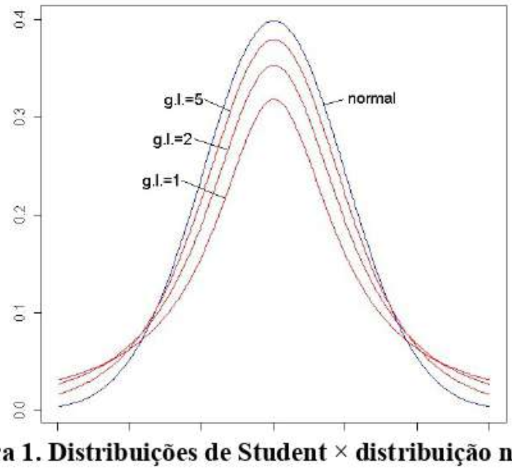

Figura 1. Distribuições de Student × distribuição normal

Note que a distribuição de Student é sempre mais achatada do que a normal, e tem caudas mais "grossas"; isto indica que os valores da distribuição de Student têm maior variabilidade, afastam-se mais da média do que os da distribuição normal. Não é difícil entender porque isto acontece. Na inferência feita sobre população de desvio-padrão σ desconhecido, será preciso estimar este σ antes de fazer o teste; no entanto, esta estimativa será imprecisa, porque feita com pouca informação (já que há poucos dados na amostra). Como conseqüência, a incerteza será maior, e os testes feitos com a distribuição de Student terão sempre menor *poder* do que os feitos com a distribuição normal. Os valores críticos para testes de hipóteses usando a variável t serão sempre maiores do que os dos testes feitos usando a variável Z de distribuição normal. Por exemplo, se fazemos testes bilaterais com nível de significância $\alpha = 0.05$, teremos como valores críticos:

na distribuição de Student	com g.l.= 4	$\rightarrow t_c = \pm\ 2{,}78$
	com g.l.= 8	$\rightarrow t_c = \pm\ 2{,}31$
	com g.l.= 15	$\rightarrow t_c = \pm\ 2{,}13$
	com g.l.= 30	$\rightarrow t_c = \pm\ 2{,}04$
	com g.l.= 100	$\rightarrow t_c = \pm\ 1{,}98$
na distribuição normal padrão	————————	$\rightarrow z_c = \pm\ 1{,}96$

Note, no gráfico acima e nos valores de t_c, que à medida que cresce o número de graus de liberdade *g.l.* (ou seja, a medida que cresce o tamanho n da amostra), a distribuição de Student tende para a normal. Na prática, portanto, para amostras razoavelmente grandes, a distribuição normal dá uma aproximação aceitável da distribuição de Student.

(i) Notas sobre notação e terminologia

(1) No Teorema **1**, podemos dizer que a variável Z calculada é uma *normalização* ou *padronização* da variável $\overline{X}$ original, porque a equação (**1**) transforma $\overline{X}$ numa variável Z de distribuição *normal padrão*. Não podemos porém dizer que a variável t no Teorema **2**, calculada pela eq. (**4**) seja uma *normalização* ou *padronização* da variável original $\overline{X}$, porque t não terá distribuição normal padrão, e sim distribuição de Student.

(2) Alguns programas têm funções que permitem que tanto a forma original (com n no denominador) quanto a forma corrigida (com $n - 1$) do desvio-padrão sejam calculadas a partir de uma amostra; outros (como o R) calculam apenas a forma corrigida. Como usaremos sempre o R, daqui em diante representaremos o desvio-padrão corrigido simplesmente por s, para simplificar a notação; fica subentendido que estamos nos referindo sempre aos valores corrigidos.

4.7.1.2. Teste usando a distribuição de *Student*

Os testes de hipóteses usando amostras pequenas, baseados na distribuição de Student, são organizados de forma bastante semelhante aos testes de médias que já vimos anteriormente (seção 4.5.3). É preciso dar os mesmos passos: decidir quais serão as hipóteses nula e alternativa, qual será o nível de significância, se o teste será unilateral ou bilateral, etc. A diferença principal é que os testes com amostras pequenas (não só o de uma média, mas todos os que veremos daqui em diante) têm sempre mais pressupostos do que os testes com amostras grandes; antes de fazer o teste, é preciso verificar se estes pressupostos são atendidos.

(i) Verificação dos pressupostos

Qualquer teste paramétrico feito com amostras pequenas se baseia em alguns pressupostos. O Teorema **2**, que define a distribuição da variável de teste t, tem dois pressupostos. O primeiro é o de que a amostra usada seja uma *amostra aleatória simples*, cujos elementos tenham sido escolhidos por meio de um sorteio entre os elementos da população. (Há vários outros tipos de amostra, veja seção 5.1). Este pressuposto não pode ser verificado a partir dos dados; não podemos descobrir a partir deles se os elementos da amostra foram escolhidos por sorteio, ou por outra técnica qualquer. Teremos que verificar como foi feito o planejamento do experimento (se os dados vieram de um experimento), como os dados foram coletados, etc.

O segundo pressuposto é o de que a distribuição da população seja *normal.* Quando estudamos a distribuição normal, vimos que ela podia ser usada como modelo para várias populações, simplesmente comparando a forma dos histogramas das amostras com a da curva normal (seção 3.4.4.3); o histograma deve ser unimodal, simétrico, com uma forma de sino, etc. Este tipo de comparação porém não pode ser aplicado quando temos uma amostra pequena. Se são poucos os dados, não faz sentido usar um histograma; além disso, qualquer gráfico que fizermos provavelmente irá ter uma forma assimétrica e irregular. A Fig. **2** mostra, como exemplo, diagramas de ramo-e-folhas de cinco amostras (n = 10) simuladas aleatoriamente a partir de uma distribuição normal padrão. Note que nenhuma delas é simétrica, há pontos discrepantes, aglomerados separados, etc.

```
-1 |             -1 | 1         -1 |             -1 | 80        -1 | 821
-0 | 532211      -0 | 3320      -0 |             -0 | 9881      -0 | 5541
 0 | 348          0 | 12777      0 | 3221179      0 | 0259       0 |
 1 |              1 |            1 | 27           1 |            1 | 345
 2 | 2            2 |            2 | 3            2 |            2 |
```

Figura 2. Amostras pequenas simuladas a partir de uma população normal padrão

É preciso sempre lembrar a distribuição normal é um modelo estatístico teórico, aplicado a uma população supostamente infinita. Pelas leis da Probabilidade, quando o tamanho da amostra cresce, a distribuição observada na amostra deverá tender para esta distribuição teórica; nas amostras pequenas, porém, não podemos prever qual será a forma da distribuição encontrada. (Isto é uma aplicação das leis básicas da Probabilidades, similar a que ocorre, por exemplo, quando lançamos uma moeda. Se lançamos duas vezes, não esperamos encontrar sempre uma cara e uma coroa; porém, se lançamos várias vezes, esperamos que o número de caras e de coroas tendam a se igualar, aproximando-se do previsto no modelo probabilístico teórico.)

Usar medidas numéricas como os coeficientes de assimetria e de curtose (seção 2.2.3) e comparar os valores medidos na amostra com os valores teóricos para uma distribuição normal, também não costuma funcionar bem para amostras pequenas. O que é feito com mais freqüência, na prática, é usar o *gráfico de separatrizes* (*quantile plot*) para fazer comparações gráficos, ou *testes de normalidade*, que testam a partir de amostras a hipótese de uma população ser normal. Veremos a seguir estas duas técnicas.

Gráfico de quantis (*quantile plot*)

O *gráfico de quantis* (já mencionado na seção 3.4.4) é um diagrama de dispersão no qual os quantis (percentis) de uma distribuição teórica são marcadas no eixo horizontal, e os quantis da distribuição encontrada na amostra são marcadas no eixo vertical. Cada valor observado na amostra é portanto representado por um ponto cujas coordenadas são a sua posição na amostra (em relação aos quantis) e a posição que ele deve ocupar na distribuição teórica. Se as duas distribuições (a observada e a teórica) coincidem aproximadamente, a maioria dos pontos devem estar próximos da reta identidade (a reta y=x). Se os pontos, ao invés de se alinharem ao longo de uma reta, se alinham ao longo de curvas, isto indica que a distribuição na amostra se afasta da normal de alguma forma (é assimétrica, tem excesso de curtose, etc.). Veremos mais detalhes sobre estes gráficos na seção 5.6.

Testes de normalidade

Existem vários tipos de testes estatísticos que verificam se um certo modelo probabilístico pode ou não ser usado para descrever uma população. Estes testes são em geral chamados de *testes de aderência*, pois testam quão bem um modelo *adere* ao dados observados. Os mais usados destes testes são os testes de *normalidade*, que verificam se a curva normal serve como modelo para uma determinada população.

Existem vários destes testes. Neste texto, usaremos os que são encontrados com mais freqüência, e que estão implementados no pacote *stats* do R: o teste de *Kolmogorov-Smirnov*, e o teste de *Shapiro*. Em ambos, as hipóteses nulas são as de que a distribuição da população é normal; se esta hipótese é rejeitada, isto indica que o modelo normal não se adequa bem à população, pelo que foi observado nos dados, e que portanto o teste *t* não pode ser usado. (Veremos mais detalhes sobre estes testes na seção 5.6).

(ii) Tabela da distribuição de Student

Se fizermos um teste de média com amostra pequena sem usar programas de Estatística, precisaremos de uma tabela para obter os valores críticos de *t*. A Tabela **1** mostra parte de uma tabela que fornece os valores críticos para testes bilaterais (a área de rejeição é dividida entre os dois extremos da curva; cada extremo tem área igual a $\alpha/2$).

Tabela 1. Valores críticos de *t*, para testes bilaterais

	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
1	0,01	0,02	0,04	0,08	0,16	0,41	1,00	2,41	6,31	12,71	25,45	63,66	127,32	636,58
2	0,01	0,01	0,04	0,07	0,14	0,37	0,82	1,60	2,92	4,30	6,21	9,92	14,09	31,60
3	0,01	0,01	0,03	0,07	0,14	0,35	0,76	1,42	2,35	3,18	4,18	5,84	7,45	12,92
4	0,01	0,01	0,03	0,07	0,13	0,34	0,74	1,34	2,13	**2,78**	3,50	4,60	5,60	8,61
5	0,01	0,01	0,03	0,07	0,13	0,34	0,73	1,30	2,02	2,57	3,16	4,03	4,77	6,87
6	0,01	0,01	0,03	0,07	0,13	0,33	0,72	1,27	1,94	2,45	2,97	3,71	4,32	5,96
7	0,01	0,01	0,03	0,06	0,13	0,33	0,71	1,25	1,89	2,36	2,84	3,50	4,03	5,41
8	0,01	0,01	0,03	0,06	0,13	0,33	0,71	1,24	1,86	**2,31**	2,75	3,36	3,83	5,04
9	0,01	0,01	0,03	0,06	0,13	0,33	0,70	1,23	1,83	2,26	2,69	3,25	3,69	4,78
10	0,01	0,01	0,03	0,06	0,13	0,33	0,70	1,22	1,81	2,23	2,63	3,17	3,58	4,59
11	0,01	0,01	0,03	0,06	0,13	0,33	0,70	1,21	**1,80**	2,20	2,59	3,11	3,50	4,44
12	0,01	0,01	0,03	0,06	0,13	0,33	0,70	1,21	1,78	2,18	2,56	3,05	3,43	4,32
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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

Para usar a tabela, procuramos na primeira linha o valor de α que queremos usar, e na primeira coluna o número de graus de liberdade, que depende do tamanho da amostra ($g.l. = n - 1$). Por exemplo: se usamos α=0,05 e temos 4 graus de liberdade, o valor crítico de t será $t_c = 2,78$ (destacado na tabela); se temos 8 graus de liberdade, o valor crítico será $t_c = 2,31$. Note que, a medida que aumenta o tamanho da amostra (e o número de graus de liberdade), o valor crítico tende para o da distribuição normal ($z_c = 1,96$), que está na última linha da tabela, correspondendo a infinitos graus de liberdade.

4.7.1.3. Exemplo

Faremos a seguir um exemplo de teste de média aritmética com amostra pequena. Mostraremos como o teste é feito, passo a passo; o mesmo exemplo também pode ser resolvido usando uma rotina em R.

Um fiscal deseja testar se as garrafas de um certo refrigerante têm em média um conteúdo igual ou superior a 250 ml. Para isto, retira uma amostra aleatória de 12 destas garrafas, mede seus conteúdos, e encontra estes valores (em ml):

244	246	249	254	245	239
254	244	261	243	236	247

O que ele pode concluir a partir destes dados?

A organização do teste é semelhante à que já vimos nos testes com amostras grandes. A hipótese a ser testada é a de que a média da população é maior ou igual a 250 ml. As hipóteses são:

$$H_0 : \quad \mu \geq 250 \text{ ml}$$
$$H_1 : \quad \mu < 250 \text{ ml}$$

Antes de fazer o teste, teremos primeiro que verificar o pressuposto mais importante, o da *normalidade* da população. A Fig. 2 mostra a gráfico de quantis com os dados da amostra; vemos que os pontos estão mais ou menos alinhados ao longo de uma reta, como esperado no caso de uma distribuição normal.

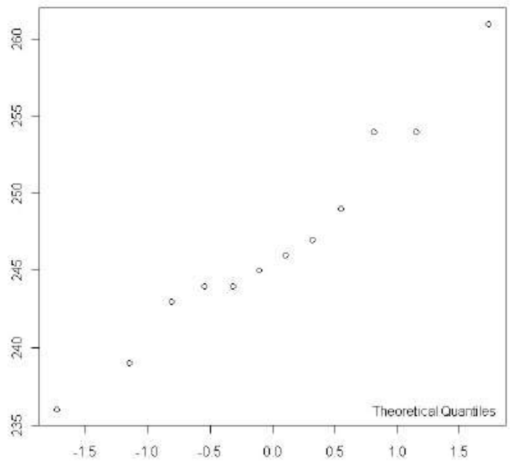

Figura 2. Gráfico de quantis

Dos dois testes de normalidade disponíveis no pacote *stats* do R, o de Kolmogorov-Smirnov não pode ser usado neste problema porque não funciona se houver valores repetidos na amostra. O teste de Shapiro dá como resultado:

W = 0.9574, p-value = 0.7564

Como o valor-p (*p-value*) é maior do que o nível de significância usual (α=0,05), a hipótese nula de que a população é normal não pode ser rejeitada. Podemos então prosseguir com o teste *t*. O número de graus de liberdade será

$$g.l. = n - 1 = 12 - 1 = 11$$

Usaremos um nível de significância $\alpha = 0{,}05$. Como o teste será unilateral (por causa de H_1), toda a área de rejeição deve estar localizada no extremo esquerdo da curva; temos que usar o valor crítico que corresponde a um teste bilateral com $\alpha = 0{,}10$. Na Tab. **1**, encontramos o valor $t_c = -1{,}80$ (destacado em azul; o sinal negativo é necessário porque o teste é feito no lado esquerdo da curva).

Na amostra, as estatísticas calculadas são:

$$\bar{X} = 246{,}83 \text{ ml}$$
$$s = 6{,}89 \text{ ml}$$

Calculando o valor de *t*:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{246{,}83 - 250{,}0}{6{,}89/\sqrt{12}} = -1{,}59$$

Como o valor de *t* encontrado é maior do que o valor crítico t_c,

$$t = -1{,}59 \; > \; t_c = -1{,}80$$

o valor de *t* não está na região de rejeição, e a hipótese nula não pode ser rejeitada. Não há como calcular o valor-*p* a partir da tabela; o cálculo no R fornece *p*=0.06974.

O resultado encontrado portanto é *não significativo*. Apesar de a média encontrada na amostra ser menor que 250 ml (e a maioria das garrafas conter menos de 250 ml), a conclusão é que não há evidência nestes dados para afirmarmos que a média das garrafas produzidas por esta fábrica seja *menor* do que 250 ml.

Resumo

1. O teste de uma hipótese sobre a média de uma população pode ser feito usando uma amostra pequena; a variável de teste será então *t* (com distribuição de *Student*) em vez de *Z* (com distribuição normal).
2. A distribuição de Student tem forma simétrica e unimodal, e apenas um parâmetro, denominado *número de graus de liberdade*.
3. Para calcular o valor da estatística de teste *t*, precisamos de uma estimativa do desvio-padrão da população; esta estimativa será dada pelo *desvio-padrão corrigido* calculado na amostra.
4. Todo teste *paramétrico* feito com amostra pequena é baseado em alguns pressupostos; é preciso verificar se eles foram atendidos, antes de fazer o teste. Se estes pressupostos não forem atendidos, podemos usar testes *não-paramétricos*, mas estes sempre têm menor poder.
5. O teste *t* para uma média populacional, além de pressupor que a distribuição na população seja *normal*, exige também que a amostra usada seja aleatória simples. A normalidade pode ser verificada a partir da amostra por meio de gráficos ou de *testes de normalidade*. Contudo, não é possível verificar a partir dos dados se a amostra foi aleatória simples.