NOTAS

[1] Desenvolvida por PAYKEL, E. et al. *Archives of General Psychiatry*, v. 75, p. 340-7, 1971.
[2] É igual a $gl_2(gl_2 - 2)$, que geralmente está próximo de 1 a não ser que n seja muito pequeno.
[3] MULLER, A. *British Medical Journal*, v. 324, p. 23-5, 2002.
[4] http://www.heritage.org/Research/Family/cda02-05.cfm.
[5] http://www.economist.com/media/pdf/QUALITYOFLIFE.pdf.
[6] SHACKELFORD, T., BESSER, A. *Individual Differences Research*, v. 5, p. 106-14, 2007.

12 COMPARANDO GRUPOS: MÉTODOS DE ANÁLISE DE VARIÂNCIA (ANOVA)

O Capítulo 7 apresentou métodos para comparar as médias de dois grupos. Nesse capítulo veremos como esses métodos podem ser estendidos para comparar as médias de *vários* grupos.

O Capítulo 8 apresentou métodos para analisar associações entre duas variáveis *categóricas*. Os Capítulos 9 e 11 apresentaram métodos de regressão para analisar a associação entre variáveis *quantitativas*. Os métodos para comparar médias para vários grupos relacionam a associação entre uma variável resposta *quantitativa* e uma variável explicativa *categórica*. A média da variável resposta quantitativa é comparada entre os grupos que são categorias da variável explicativa. Por exemplo, para uma comparação da renda média anual entre negros, brancos e hispânicos, a variável resposta quantitativa é a renda anual e a variável explicativa categórica é o *status* étnico-racial.

O método inferencial para comparar várias médias é denominado de **análise de variância** e é abreviado por **ANOVA**. A Seção 12.1 mostra que o nome se refere à forma como um teste de significância se concentra em dois tipos de variabilidade dos dados. A Seção 12.2 apresenta os intervalos de confiança comparando as médias dos grupos. A Seção 12.3 mostra que as inferências são casos especiais de uma análise de regressão múltipla. As Seções 12.4 e 12.5 estendem esses métodos para incorporar variáveis explicativas adicionais – por exemplo, para comparar a renda média por meio das categorias do *status* étnico-racial e o gênero.

As Seções 12.1 a 12.5 apresentam análises para *amostras independentes*. Como a Seção 7.1 explicou, quando cada amostra tem os mesmos sujeitos, em vez de amostras não emparelhadas, as amostras são *dependentes* e diferentes métodos são aplicados. As Seções 12.6 e 12.7 apresentam esses métodos.

12.1 COMPARANDO VÁRIAS MÉDIAS: O TESTE *F* DA ANÁLISE DE VARIÂNCIA

O notável estatístico britânico Ronald A. Fisher desenvolveu o método de análise de variância em torno de 1920. O ponto principal dessa análise é o teste de significância, usando a distribuição F para detectar as diferenças entre um conjunto de médias populacionais.

Suposições do teste *F* para comparar médias

Considere g a representação do número de grupos a serem comparados, como $g = 3$, como acima, na comparação entre negros, brancos e hispânicos. As médias da variável resposta para as populações correspondentes são $\mu_1, \mu_2, \dots, \mu_g$, como μ_1 para a renda média anual de negros, μ_2 para dos brancos e μ_3 para a dos hispânicos. As médias amostrais são $\bar{y}_1, \bar{y}_2, ..., \bar{y}_g$.

A análise de variância (ANOVA) é um teste F para testar se:

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

H_a: pelo menos duas das médias populacionais são diferentes.

Se H_0 é falsa, talvez todas as médias populacionais sejam diferentes, talvez algumas sejam diferentes ou talvez somente uma média difira das demais. O teste analisa se as diferenças observadas entre as médias amostrais poderiam ter, plausivelmente, ocorrido por acaso, se H_0 fosse verdadeira.

As suposições para o teste são as seguintes:

- Para cada grupo, a distribuição populacional da variável resposta y é normal.
- O desvio padrão da distribuição populacional é a mesma em cada grupo. Representamos o valor comum por σ.
- As amostras retiradas da população são aleatórias e *independentes*.

A Figura 12.1 exibe uma representação das duas primeiras suposições. Sob estas suposições, a hipótese nula declara que a distribuição da população não depende do grupo ao qual o sujeito pertence. O teste da ANOVA é um *teste de independência* entre uma variável resposta quantitativa e uma variável explicativa categórica.

As suposições sobre as distribuições populacionais são severas e, de fato, nunca são inteiramente satisfeitas na prática. Como de costume, a suposição da amostragem aleatória é a mais importante. A última seção deste capítulo discute os efeitos da violação das suposições.

EXEMPLO 12.1 Ideologia política por partido político

A Tabela 12.1 resume as observações sobre a ideologia política para três grupos, baseadas em dados de sujeitos com idades entre os 18 e os 30 anos da PSG de 2004. Os três grupos são as categorias (Democratas, Independentes, Republicanos) da variável explicativa, identificação do partido político (ID). A ideologia política, a variável resposta, é mensurada em uma escala de sete pontos, variando de extremamente liberal (1) a extremamente conservador (7). Para cada ID do partido, a Tabela 12.1 mostra o número de sujeitos que deu cada resposta. Por exemplo, de 91 Democratas, 9 responderam extremamente liberal, 20 responderam liberal, ..., 0 responderam extremamente conservador.

Visto que a Tabela 12.1 exibe os dados como frequências em uma tabela de contingência, poderíamos usar métodos para dados categóricos (Capítulo 8). O teste qui-quadrado trata ambas as variáveis como nominais, embora a ideologia política seja ordinal. Esse teste não foi projetado para detectar se as respostas têm uma média mais alta ou mais baixa em alguns grupos do que em outros. Da mesma forma, os métodos ordinais do Capítulo 8 (como a medida gama de associação) são inapropriados porque eles requerem que ambas as variáveis sejam ordinais. Aqui, os grupos, que são categorias de identificação partidária, são nominais.

Quando uma resposta ordinal tem várias categorias, na prática é comum atribuir escores para seus níveis e tratá-la como uma variável quantitativa. Essa é uma estratégia razoável quando queremos trabalhar com uma medida de centro como a média em vez de proporções em categorias. Para a Tabela 12.1, por exemplo, o foco de interesse pode estar em quão liberais ou conservadoras as respostas tendem a ser para cada grupo, em um sentido médio, em vez das proporções de cada categoria. Analisamos esses dados atribuindo os escores (1, 2, 3, 4, 5, 6, 7) aos níveis de ideologia política e, então, iremos comparar as médias. Quanto mais alto o escore médio, mais conservadoras as respostas do grupo tendem a ser.

Para esses escores, a Tabela 12.1, também, mostra a média e o desvio padrão para cada grupo. A média geral da amostra é $\bar{y} = 3{,}89$, bem próxima do escore de 4,0 correspondendo à ideologia moderada. Devemos testar se as três populações têm médias iguais. A hipótese nula é H_0: $\mu_1 = \mu_2 = \mu_3$, onde μ_1 é a ideologia média da população para os Democratas, μ_2 para os Independentes e μ_3 para os Republicanos. ■

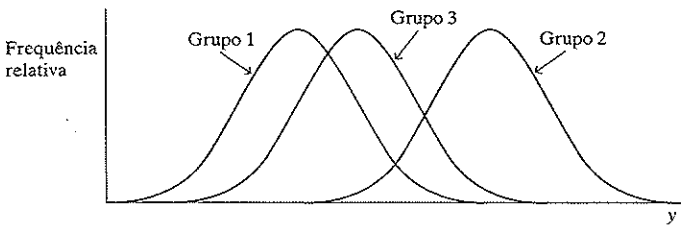


Figura 12.1 Suposições sobre as distribuições populacionais: normais com desvios padrão iguais (σ).

Variabilidade entre e dentro dos grupos

Por que um método para comparar as *médias* das populações é chamado de uma análise de *variância*? A razão é que a estatística-teste compara as médias usando duas estimativas da variância, σ^2, para cada grupo. Uma estimativa usa a variabilidade *entre* cada média amostral $\bar{y}_i$ e a média geral $\bar{y}$. A outra estimativa usa a variabilidade *dentro* de cada grupo de observações amostrais sobre suas médias individuais – as observações do primeiro grupo sobre $\bar{y}_1$, as observações do segundo grupo sobre $\bar{y}_2$ e, assim por diante.

Para ilustrar, suponha que observações aleatoriamente amostradas desses grupos estão exibidas na Figura 12.2a. Parece claro que as médias das populações que essas médias amostrais estimam não são iguais. A base para esta conclusão é que a variabilidade *entre* as médias amostrais é grande e a variabilidade das observações *dentro* de cada amostra é pequena.

Por outro lado, olhe para a Figura 12.2b. Ela tem as médias amostrais iguais às da Figura 12.2a, assim a variabilidade *entre* as médias amostrais é a mesma. Mas, na Figura 12.2b, a variabilidade *dentro* dos grupos é muito maior do que na Figura 12.2a. O desvio padrão amostral para cada grupo é muito maior na Figura 12.2b. Agora, não está claro se as médias populacionais diferem. Geralmente, quanto maior a variabilidade entre as médias amostrais e menor

Tabela 12.1 Ideologia política *versus* identificação partidária, para sujeitos com idades entre 18 e 30 anos

Grupo (ID Partidária)	Ideologia política							Tamanho da amostra	Média	Desvio padrão
	1	2	3	4	5	6	7			
Democrata	9	20	17	36	4	5	0	91	3,23	1,28
Independente	7	11	17	48	12	11	5	111	3,90	1,43
Republicano	0	2	7	23	23	17	2	74	4,70	1,10

Nota: Para a ideologia política, 1 = extremamente liberal, 2 = liberal, 3 = moderadamente liberal, 4 = moderado, 5 = moderadamente conservador, 6 = conservador, 7 = extremamente conservador.

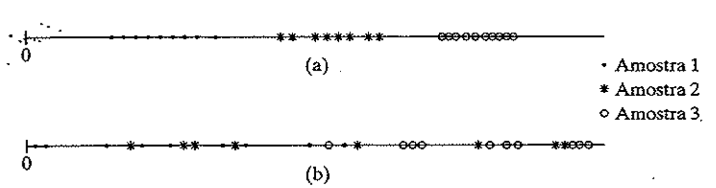


Figura 12.2 Duas amostras. As médias são as mesmas em cada caso, assim a variabilidade entre grupos é a mesma, mas a variabilidade dentro dos grupos é maior no segundo conjunto.

a variabilidade dentro de cada grupo, mais forte é a evidência contra a hipótese nula de médias populacionais iguais.

A estatística-teste *F* é a razão entre duas estimativas da variância

Para testar H_0: $\mu_1 = \mu_2 = \ldots = \mu_g$, a estatística-teste é a razão entre as duas estimativas da variância da população. A estimativa que usa a variabilidade *entre* cada média amostral $\bar{y}_i$ e a média global $\bar{y}$ é chamada de **estimativa entre grupos**. A estimativa que usa a variabilidade *dentro* de cada amostra é chamada de **estimativa dentro dos grupos**.

A estatística-teste *F* tem a forma:

$$F = \frac{\text{Estimativa da variância entre os grupos}}{\text{Estimativa da variância dentro dos grupos}}.$$

Esse valor é denominado de **estatística *F* da análise de variância** ou **estatística *F* da ANOVA** para abreviar. Deixaremos os detalhes computacionais das estimativas da variância para mais tarde nesta seção.

A estimativa dentro dos grupos é uma estimativa não tendenciosa de σ^2 não importando se H_0 é verdadeira. Por causa disto, iremos usá-la também na fórmula para os intervalos de confiança comparando médias. Em contraposição, a estimativa entre grupos não é tendenciosa apenas quando H_0 é verdadeira. Ela, então, assume o mesmo valor da estimativa dentro dos grupos. Esperamos, então, valores de *F* próximos a 1,0, excetuando-se o erro amostral. Se H_0 é falsa, contudo, a estimativa entre grupos tende a superestimar σ^2. Ela tende a ser maior do que a estimativa dentro dos grupos e a estatística *F* tende a ser maior do que 1,0, crescendo com o tamanho da amostra.

Quando H_0 é verdadeira a estatística *F* tem uma distribuição amostral *F*. Como nos testes *F* para os parâmetros da regressão múltipla (p. ex., Seção 11.4 na página 375), o valor-*p* é a probabilidade da cauda direita de que a estatística-teste *F* exceda o valor observado *F*. Quanto maior for a estatística-teste *F*, menor será o valor-*p*.

EXEMPLO 12.2 Teste *F* comparando a média da ideologia política pela identidade partidária

O *software* exibe os resultados dos testes *F* da ANOVA em uma tabela similar àquela usada para apresentar as somas dos quadrados na análise de regressão. Esta tabela é chamada de **tabela da ANOVA**. A Tabela 12.2 mostra tal tabela para o teste *F* de H_0: $\mu_1 = \mu_2 = \mu_3$ comparando a média da ideologia política populacional para três identidades partidárias com a Tabela 12.1.

Na tabela da ANOVA,

- As duas "médias dos quadrados" são as estimativas da variância populacional (σ^2) entre e dentro dos grupos.
- A estatística-teste *F* é a razão entre as duas médias dos quadrados.

Da coluna das Médias dos Quadrados da Tabela 12.2, a estimativa entre grupos da variância é 44,21 e a dentro dos grupos é 1,68. A estatística-teste $F = 44{,}21/1{,}68 = 26{,}3$. Em outras palavras, a estimativa entre grupos é superior a mais do que 25 vezes a estimativa dentro dos grupos. Lembre que, se H_0 é verdadeira, esperamos valores de *F* próximos a 1,0, desconsiderando o erro amostral. Assim, a estatística-teste fornece uma forte evidência contra H_0: $\mu_1 = \mu_2 = \mu_3$. A Tabela 12.2 informa que o valor-*p* é 0,000, arredondado para três casas decimais. Concluímos que existe uma diferença entre os valores da ideologia política média na população para as três identificações partidárias. ■

Tabela 12.2 Tabela da ANOVA para o teste *F* com os dados da Tabela 12.1. A estatística-teste *F* é a razão entre as médias dos quadrados

Fonte	Soma dos quadrados	*gl*	Média dos quadrados	*F*	Sig
Entre grupos (ID partidária)	88,43	2	44,21	26,3	0,000
Dentro dos grupos (erro)	459,52	273	1,68		
Total	547,95	275			

Estimativa da variância dentro dos grupos*

Agora veremos como construir as estimativas da variância que formam a estatística *F*. Cada estimativa é uma medida da variação dividida por um grau de liberdade.

A estimativa dentro dos grupos da variância da população σ^2 combina a soma dos quadrados das observações de cada grupo em relação às suas médias. Para as n_1 observações do primeiro grupo, $\sum(y - \bar{y}_1)^2$ é a soma dos quadrados das observações em relação à sua média. Esta soma dos quadrados tem $n_1 - 1$ graus de liberdade, o denominador para a variância amostral s_1^2 para o grupo 1. Da mesma forma, para as n_2 observações do segundo grupo, $\sum(y - \bar{y}_2)^2$ é a soma dos quadrados das observações em relação à sua média amostral, com $n_2 - 1$ graus de liberdade. A soma destes termos de soma de quadrados para toda a amostra (grupos) é denominada **soma dos quadrados dentro dos grupos**, visto que as somas dos quadrados são calculadas *dentro* de cada grupo que forma a amostra.

A soma dos quadrados dentro dos grupos tem graus de liberdade iguais à soma dos valores dos *gl* dos grupos componentes:

$$\begin{aligned} gl &= (n_1 - 1) + (n_2 - 1) + \ldots + (n_g - 1) \\ &= (n_1 + n_2 + \ldots + n_g) - g = N - g \\ &= \text{Tamanho total da amostra} - \text{número de grupos}, \end{aligned}$$

onde *N* representa o tamanho total da amostra. A razão

$$s^2 = \frac{\text{Soma dos quadrados dentro dos grupos}}{gl} = \frac{\text{SQ dentro dos grupos}}{N - g}$$

é a estimativa dentro dos grupos da variância da população σ^2 para os *g* grupos que compõem a amostra.

Essa estimativa resume a informação sobre a variabilidade dos vários grupos. A estimativa de σ^2 usando somente o primeiro grupo é:

$$s_1^2 = \frac{\sum(y - \bar{y}_1)^2}{n_1 - 1}.$$

Na Tabela 12.1, por exemplo, esse resulado é o quadrado do desvio padrão apresentado, $s_1 = 1{,}28$. Da mesma forma, a variância amostral para o segundo grupo é $s_2^2 = \sum(y - \bar{y}_2)^2/(n_2 - 1)$, e assim por diante para os grupos restantes. Sob a suposição de que as variâncias da população são idênticas, todos estes termos estimam o mesmo parâmetro, σ^2. O numerador e o denominador de s^2 combina as informações dessas estimativas somando seus numeradores e denominadores. A estimativa resultante se relaciona às estimativas separadas das variâncias amostrais por:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_g - 1)s_g^2}{N - g}.$$

Essa estimativa é uma média ponderada das variâncias amostrais separadas, com uma ponderação maior dada às amostras maiores. Com tamanhos amostrais iguais, s^2 é a média simples das g variâncias amostrais.

Para os dados da ideologia política da Tabela 12.1, poderíamos calcular s^2 calculando a soma dos quadrados dentro dos grupos a partir dos dados brutos. Visto que a Tabela 12.1 informa os desvios padrão, entretanto, é mais simples usar a fórmula recém dada ou, melhor ainda, usar um *software*. Os tamanhos dos grupos são $n_1 = 91$, $n_2 = 111$, $n_3 = 74$, para um tamanho total da amostra de $N = 276$. Da Tabela 12.1, segue:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{N - 3}$$
$$= \frac{(91 - 1)(1{,}28)^2 + (111 - 1)(1{,}43)^2 + (74 - 1)(1{,}10)^2}{276 - 3} = \frac{459{,}5}{273} = 1{,}68.$$

Em resumo, a soma dos quadrados dentro dos grupos é igual a 459,5, com $gl = 273$, fornecendo uma estimativa da variância dentro dos grupos de 1,68. A estimativa do desvio padrão $s = \sqrt{1{,}68} = 1{,}30$ resume os desvios padrão das três amostras da Tabela 12.1.

Estimativa da variância entre os grupos*

A estimativa de σ^2 baseada na variabilidade entre cada média amostral e a média amostral total é igual a:

$$\frac{\sum_i n_i(\bar{y}_i - \bar{y})^2}{g - 1}$$
$$= \frac{n_1(\bar{y}_1 - \bar{y})^2 + \cdots + n_g(\bar{y}_g - \bar{y})^2}{g - 1}.$$

O Exercício 12.57 induz essa fórmula. Visto que essa estimativa descreve a variabilidade entre as médias g, seu

$$gl = g - 1 = \text{Número de grupos} - 1,$$

que é o denominador da estimativa.

O numerador dessa estimativa é chamado de **soma dos quadrados entre os grupos**. A diferença ao quadrado entre cada média amostral e a média geral é ponderada pelo tamanho da amostra sobre o qual ela é baseada. Quando as médias da população são desiguais, os valores $\bar{y}_i$ tendem a variar mais do que se as médias da população fossem iguais. Quanto mais longe a média da população estiver da hipótese H_0: $\mu_1 = ... = \mu_g$, maior será a soma dos quadrados entre os grupos (SQ), a estimativa entre grupos e a estatística-teste F.

Para a Tabela 12.1, a média geral é $\bar{y} = 3{,}89$. A soma dos quadrados entre os grupos é igual a:

$$\text{SQ entre grupos} =$$
$$n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2$$
$$+ n_3(\bar{y}_3 - \bar{y})^2 = 91(3{,}23 - 3{,}89)^2$$
$$+ 111(3{,}90 - 3{,}89)^2 + 74(4{,}70 - 3{,}89)^2$$
$$= 88{,}43.$$

Visto que g = número de grupos = 3, esta soma dos quadrados tem $gl = g - 1 = 3 - 1 = 2$. A estimativa entre grupos da variância é, então:

$$\frac{\text{SQ entre grupos}}{g - 1} = \frac{88{,}43}{2} = 44{,}2.$$

Soma dos quadrados nas tabelas da ANOVA*

Vamos, agora, olhar a tabela da ANOVA (Tabela 12.2) para a ideologia política e identificação partidária. Para o teste F,

$$gl_1 = g - 1 = \text{Número de grupos} - 1 \text{ e}$$
$$gl_2 = N - g = \text{Tamanho da amostra total} - \text{Número de grupos}.$$

Eles são informados na coluna gl (*df*) da tabela. Para estes dados, $gl_1 = g - 1 = 3 - 1 = 2$ e $gl_2 = N - g = 276 - 3 = 273$.

Na linha do "Entre grupos" da tabela da ANOVA, a SQ entre grupos é dividida pelo gl_1 originando o quadrado médio, 88,43/2 = 44,21. Na linha "Dentro dos grupos", a SQ dentro dos grupos é dividida pelo gl_2 fornecendo o outro quadrado médio, 459,52/273 = 1,68. A estatística-teste F para H_0: $\mu_1 = \mu_2 = \mu_3$ é a razão das estimativas da variância, que é a razão entre os dois quadrados médios, $F = 44{,}21/1{,}68 = 26{,}3$. Os dois termos dos gl para o teste são os denominadores das duas estimativas da variância.

A soma da soma dos quadrados dentro dos grupos com a entre grupos é denominada de **soma dos quadrados total**. Na verdade, isto é igual a

$$\text{SQT} = \sum(y - \bar{y})^2 = \text{SQ entre grupos} + \text{SQ dentro dos grupos},$$

a soma dos quadrados da amostra combinada de N observações sobre a média total, $\bar{y}$. A Tabela 12.2 mostra que SQT = 547,95 = 88,43 + 459,52.

A ANOVA particiona a variabilidade de total em relação à média geral, SQT, em duas partes independentes. Uma parte, a SQ entre grupos, é a porção do total explicado pelas diferenças entre a média de cada grupo e a média geral. Isto também é chamado de *soma dos quadrados dos grupos*, e a maioria dos *softwares* substitui o rótulo "Entre grupos" da Tabela 12.2 pelo nome da variável do grupo (por exemplo, "PARTYID"). A outra parte, a SQ dentro dos grupos é a porção da variabilidade total que não pode ser explicada pelas diferenças entre os grupos. Ela representa a variabilidade que permanece após classificar as observações em grupos separados. A soma dos quadrados dentro dos grupos é também chamada de *soma dos quadrados do erro* e a maioria dos *softwares* substitui o rótulo "Dentro dos grupos" da Tabela 12.2 pelo termo "erro". A Seção 12.3 explica a analogia entre estas somas dos quadrados e a soma dos quadrados da análise de regressão.

O teste *F* *versus* vários testes *t*

Com dois grupos, a Seção 7.5 (página 225) mostrou como um teste t pode comparar as médias sob a suposição de desvios padrão populacionais iguais. Na verdade, se aplicarmos o teste F da ANOVA para dados de $g = 2$ grupos, a estatística-teste F é igual ao quadrado da estatística-teste t. O valor-p para o teste F é exatamente o mesmo que o valor-p bilateral para o teste t. Podemos usar qualquer um destes testes para executar a análise.

Com vários grupos, por que não usar o teste t para comparar cada par de médias, em vez de usar o teste F? Uma das razões é que usando um único teste F em vez de vários testes t controlamos a probabilidade geral de erro do Tipo I. Com um nível α de 0,05 no teste F, a probabilidade de rejeitar impropriamente uma H_0 verdadeira é fixa em 0,05. Ao contrário, quando executamos um teste t separado para cada par de médias, a probabilidade do erro do Tipo I se aplica a *cada* comparação. Nesse caso, então, não controlamos a taxa geral do erro do Tipo I para todas as comparações. A próxima seção mostra como estas mesmas considerações são relevantes para os intervalos de confiança.

12.2 COMPARAÇÕES MÚLTIPLAS DE MÉDIAS

O teste F da análise de variância de H_0: $\mu_1 = \mu_2 = ... = \mu_g$ é um teste global de independência das variáveis resposta e as explicativas, como o teste F de H_0: $\beta_1 = ... = \beta_k = 0$ para modelos de regressão múltipla ou o teste qui-quadrado para tabelas de contingência. Quando o valor-p é pequeno, isto não indica quais médias são diferentes ou quão diferentes elas são. Os intervalos de confiança são os que nos auxiliam a determinar isso. Mesmo se o valor-p não for pequeno, ele ainda é informativo para estimar quão grandes poderiam ser as diferenças das médias.

Intervalos de confiança comparando médias

Na prática, é mais informativo estimar as médias da população do que simplesmente testar se elas são iguais. Podemos construir um intervalo de confiança para cada média ou para cada diferença entre pares de médias.

- Um intervalo de confiança para μ_i é:

$$\bar{y}_i \pm t\frac{s}{\sqrt{n_i}}.$$

Nesta fórmula, s é a raiz quadrada da estimativa dentro dos grupos de σ^2 que é o denominador da estatística-teste F da ANOVA. O valor-t para o nível de confiança escolhido é baseado no gl para aquela estimativa, $gl = N - g$.

- Um intervalo de confiança para $\mu_i - \mu_j$ é:

$$\left(\bar{y}_i - \bar{y}_j\right) \pm ts\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Nesta fórmula, o $gl = N - g$ para o valor-t tabelado. Existe evidência de uma diferença entre μ_i e μ_j quando o intervalo não contém 0. (Para $g = 2$ grupos, $gl = N - g = n_1 + n_2 - 2$; esse intervalo de confiança, então, simplifica o introduzido na Seção 7.5 para $\mu_2 - \mu_1$ assumindo um desvio padrão comum.)

EXEMPLO 12.3 Comparando a ideologia média de Democratas e Republicanos

Para a Tabela 12.1, vamos comparar a ideologia média da população de Democratas (grupo 1) com a dos Republicanos (grupo 3). Da Tabela 12.1, $\bar{y}_1 = 3{,}23$ para $n_1 = 91$ Democratas e $\bar{y}_3 = 4{,}70$ para $n_3 = 74$ Republicanos. Da Tabela 12.2, a estimativa do desvio padrão da população é $s = \sqrt{1{,}68} = 1{,}30$, com $gl = 273$. Para um intervalo de 95% de confiança com $gl = 273$, o escore-t é $t_{0,025} = 1{,}97$ (essencialmente o escore $z_{0,025}$). O intervalo de confiança para $\mu_3 - \mu_1$ é, então:

$$(\bar{y}_3 - \bar{y}_1) \pm t_{0,025}s\sqrt{\frac{1}{n_1} + \frac{1}{n_3}}$$

$$= (4{,}70 - 3{,}23) \pm 1{,}97(1{,}30)\sqrt{\frac{1}{91} + \frac{1}{74}}$$

$$= 1{,}47 \pm 0{,}40 \quad \text{que é} \quad (1{,}07; 1{,}87).$$

Inferimos que a ideologia média na população está entre 1,07 e 1,87 unidades acima para os Republicanos do que para os Democratas. Visto que o intervalo contém somente números positivos, concluímos que $\mu_3 - \mu_1 > 0$; isto é, μ_3 excede μ_1. Na média, os Republicanos são mais conservadores do que os Democratas, com a diferença aproximadamente entre uma a duas categorias em uma escala de sete categorias. ■

Taxa de erro com um grande número de intervalos de confiança

Com g grupos, existem $g(g - 1)/2$ pares de grupos para comparar. Quando g é relativamente grande, o número de comparações pode ser muito grande. Os intervalos de confiança para alguns pares de médias podem sugerir que eles são diferentes *mesmo se todas as médias populacionais são iguais*.

Quando $g = 10$, por exemplo, existem $g(g - 1)/2 = 45$ pares de médias. Suponha que formamos um intervalo de 95% de confiança para a diferença entre cada par. A probabilidade de erro de 0,05 se aplica para cada comparação. Para as 45 comparações, esperaríamos que $45(0{,}05) = 2{,}25$ dos intervalos não conteriam a verdadeira diferença entre as médias.

Para intervalos de 95% de confiança, a probabilidade de erro de 0,05 é a probabilidade de que qualquer intervalo de confiança em particular não conterá a verdadeira diferença das médias populacionais. Quando formamos um grande número de intervalos de confiança, a probabilidade de que *pelo menos* um intervalo de confiança esteja errado é muito maior do que a probabilidade de erro para qualquer intervalo em particular. Quanto maior o número dos grupos comparados, maior é a chance de que pelo menos uma inferência esteja incorreta.

Comparações múltiplas de médias de Bonferroni

Quando planejamos muitas comparações, existem métodos disponíveis que controlam a probabilidade de que *todos* os intervalos contenham as diferenças verdadeiras. Tais métodos são chamados de métodos de **comparações múltiplas**. Eles fixam a probabilidade de que *todos* os intervalos contenham as verdadeiras diferenças das médias populacionais *simultaneamente* em vez de individualmente.

Por exemplo, com um método de comparações múltiplas aplicado a $g = 10$ médias com 95% de confiança, a probabilidade é igual a 0,95 de que *todos* os intervalos contenham as diferenças pareadas $\mu_i - \mu_j$. De forma equivalente, a probabilidade de que *pelo menos um* intervalo esteja errado é igual a 0,05. Esta probabilidade é chamada de **taxa de erro das comparações múltiplas**.

Primeiro, apresentaremos o método de **comparações múltiplas de Bonferroni**, visto que ele é simples e se aplica a uma grande variedade de situações. O método Bonferroni usa as mesmas fórmulas dos intervalos de confiança introduzidos no início desta seção. Entretanto, ele usa um nível de confiança mais estrito para cada intervalo, para assegurar que o nível de confiança geral é suficientemente alto.

Para ilustrar, suponha que gostaríamos de comparações múltiplas com uma taxa de erro de 0,10, isto é, uma probabilidade de 0,90 de que todos os intervalos de confiança estejam simultaneamente corretos. Se planejarmos quatro comparações de médias, então o método Bonferroni usa a probabilidade de erro de 0,10/4 = 0,025 para cada comparação. Isto é, ele usa um nível de 97,5% de confiança para cada intervalo. Essa abordagem é um tanto conservadora, pois assegura que a taxa de erro geral real é *no máximo* de 0,10 e que o nível de confiança geral é *no máximo* de 0,90. O método é baseado em uma desigualdade mostrada pelo matemático italiano Carlo Bonferroni em 1935. Ele declara que a probabilidade de que pelo menos um dos conjuntos de eventos ocorra não pode ser maior do que a soma das probabilidades separadas dos eventos. Por exemplo, se a probabilidade de erro é igual a 0,025 para cada um de quatro intervalos de confiança, então a probabilidade de que pelo menos um dos quatro intervalos esteja errado não é maior do que (0,025 + 0,025 + 0,025 + 0,025) = 0,10.

EXEMPLO 12.4 Intervalos de Bonferroni para as comparações da ideologia política

Para o $g = 3$ identificações partidárias da Tabela 12.1, vamos comparar as médias das ideologias políticas μ_1 com μ_2, μ_1 com μ_3 e μ_2 com μ_3. Construímos intervalos de confiança (IC) tendo um nível de confiança geral de pelo menos 0,95. Para as três comparações múltiplas com taxa de erro de 0,05, o método Bonferroni usa a probabilidade de erro de 0,05/3 = 0,0167 para cada intervalo. Elas utilizam o escore-*t* com a probabilidade bicaudal igual a 0,0167 ou a probabilidade unicaudal de 0,0083. Para um valor *gl* grande aqui (*gl* = 273), o escote-*t* é igual a 2,41, próximo ao escore-*z* de 2,39. Lembre, também, que $s = 1{,}30$.

O intervalo para $\mu_3 - \mu_1$, a diferença entre a ideologia média populacional de Republicanos e Democratas, é:

$$(\bar{y}_2 - \bar{y}_1) \pm ts\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (4{,}70 - 3{,}23) \pm 2{,}41(1{,}30)\sqrt{\frac{1}{91} + \frac{1}{74}}$$

$$= 1{,}47 \pm 0{,}49 \quad \text{que é} \quad (0{,}98; 1{,}96)$$

Construímos os intervalos para os outros dois pares de médias de uma forma semelhante. A Tabela 12.3 os exibe. Nenhum dos intervalos contém 0. Eles mostram evidência significativa de uma diferença entre cada par das médias populacionais.

Os intervalos de 95% de confiança das comparações múltiplas de Bonferroni são mais amplos do que os intervalos de 95% de confiança individualmente determinados. Por exemplo, o intervalo de 95% de confiança comum comparando Republicanos e Democratas é (1,07; 1,87), enquanto o intervalo de Bonferroni é (0,98; 1,96). Isso ocorre porque o método das comparações múltiplas usa um nível de confiança mais alto para cada intervalo separado para assegurar a obtenção do nível de confiança geral que se aplica a todo o conjunto de comparações. ■

Comparações múltiplas de médias de Tukey

Dentre outros métodos disponíveis para comparações múltiplas recomendamos o **método de Tukey**. Proposto pelo importante estatístico John Tukey, que também desenvolveu métodos de análise exploratória de dados como os diagramas de caixa e bigodes e de caule e folhas e também o termo *software*. Esse método tem intervalos que são levemente menores do que os intervalos de Bonferroni. Isso porque eles são projetados para *aproximar* o nível de confiança nominal em vez de ter *pelo menos* aquele nível. O método Tukey usa a distribuição de probabilidade (o intervalo *estudentizado*) da diferença entre as maiores e menores médias amostrais. Não apresentaremos essa distribuição neste livro, assim contamos com um *software* em vez de uma fórmula para determinar os intervalos de Tukey.

A Tabela 12.3 mostra os intervalos de Tukey para os dados da ideologia política. Para propósitos práticos, eles fornecem as mesmas conclusões do que os intervalos de Bonferroni.

☑ **Tabela 12.3** Comparações múltiplas de 95% de Bonferroni e Tukey da ideologia partidária média para três grupos de identidade partidária. A confiança de 95% se aplica ao conjunto dos três intervalos, em vez de para cada intervalo individuamente

Grupos	Diferença das médias $\mu_i - \mu_j$	Diferença estimada $\bar{y}_i - \bar{y}_j$	IC de 95% de Bonferroni	IC de 95% de Tukey
(Independente, Democrata)	$\mu_2 - \mu_1$	0,67	(0,23; 1,11)*	(0,24; 1,10)*
(Republicano, Democrata)	$\mu_3 - \mu_1$	1,47	(0,98; 1,96)*	(0,99; 1,95)*
(Republicano, Independente)	$\mu_3 - \mu_2$	0,80	(0,33; 1,27)*	(0,34; 1,26)*

Nota: Um asterisco (*) indica uma diferença significativa.

12.3 EXECUTANDO A ANOVA PELOS MODELOS DE REGRESSÃO

O Capítulo 11 usou a regressão múltipla para modelar o relacionamento entre a média de uma variável resposta quantitativa e um conjunto de variáveis explicativas *quantitativas*. A ANOVA modela o relacionamento entre a média de uma variável resposta quantitativa e uma variável explicativa *categórica*; as categorias são os grupos comparados. Na verdade, a ANOVA é um caso especial da regressão múltipla. Variáveis explicativas artificiais em um modelo de regressão podem representar os grupos.

A regressão com variáveis auxiliares

Estabelecemos uma variável artificial para ser igual a 1 se uma observação vem de um grupo em particular e 0 de outra forma. Com três grupos, como no exemplo da ideologia política e identificação partidária, usamos duas variáveis artificiais. A primeira, representada por z_1, é igual a 1 para observações do primeiro grupo e é igual a 0 de outra forma. A segunda, representada por z_2, é igual a 1 para observações do segundo grupo e igual a 0 de outra forma. Isto é:

$z_1 = 1$ e $z_2 = 0$: observações do grupo 1

$z_1 = 0$ e $z_2 = 1$: observações do grupo 2

$z_1 = 0$ e $z_2 = 0$: observações do grupo 3.

É desnecessário e redundante criar uma variável para o último (terceiro) grupo porque os valores de 0 para z_1 e z_2 identificam as observações desse grupo.

As variáveis artificiais z_1 e z_2 são chamadas de **variáveis auxiliares**. Elas indicam o grupo de uma observação. Isto é, elas dão uma classificação, não uma magnitude, para um previsor categórico. A Tabela 12.4 resume as variáveis auxiliares para os três grupos.

Para as variáveis auxiliares recém definidas, considere a equação da regressão múltipla

$$E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2.$$

Para as observações do grupo 3, $z_1 = z_2 = 0$. A equação, então é simplificada a

$$E(y) = \alpha + \beta_1(0) + \beta_2(0) = \alpha.$$

Assim, α representa a média da população μ_3 de y para o último grupo. Para as observações do grupo 1, $z_1 = 1$ e $z_2 = 0$, assim

$$E(y) = \alpha + \beta_1(1) + \beta_2(0) = \alpha + \beta_1$$

é igual à média da população μ_1 para aquele grupo. De forma similar, $\alpha + \beta_2$ é

☑ **Tabela 12.4** As duas variáveis auxiliares para os três grupos

Grupo	z_1	z_2
1	1	0
2	0	1
3	0	0

Tabela 12.5 Interpretação dos coeficientes das variáveis auxiliares no modelo $E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2$

Grupo	z_1	z_2	Média do y	Interpretação do β
1	1	0	$\mu_1 = \alpha + \beta_1$	$\beta_1 = \mu_1 - \mu_3$
2	0	1	$\mu_2 = \alpha + \beta_2$	$\beta_2 = \mu_2 - \mu_3$
3	0	0	$\mu_3 = \alpha$	

igual à média da população μ_2 para o grupo 2 (considere $z_1 = 0$ e $z_2 = 1$).

Visto que $\alpha + \beta_1 = \mu_1$ e $\alpha = \mu_3$, β_1 representa a diferença $\mu_1 - \mu_3$. Da mesma forma, $\beta_2 = \mu_2 - \mu_3$. A Tabela 12.5 resume os parâmetros do modelo de regressão e sua correspondência com as médias populacionais. O coeficiente β de uma variável auxiliar representa a diferença entre a média para o grupo que aquela variável auxiliar representa e a média do grupo que não tem a sua própria variável auxiliar.

A codificação das variáveis auxiliares funciona porque ela permite que as médias da população assumam valores arbitrários, sem assumir uma distância entre os grupos. Usar uma única variável artificial com código tal como $z = 1$ para o grupo 1, $z = 2$ para o grupo 2 e $z = 3$ para o grupo 3, não funcionaria. O modelo $E(y) = \alpha + \beta_z$ iria, então, assumir uma ordem assim como distâncias iguais entre os grupos. Ele trata a variável categórica como se ela fosse quantitativa, o que é impróprio. Visto que ele assume somente um termo em um modelo de regressão para representar o efeito linear de uma variável explicativa quantitativa, ele requer $g - 1$ termos para representar as categorias g de uma variável categórica.

EXEMPLO 12.5 Modelo de regressão para a ideologia política e identidade partidária

Para a Tabela 12.1, a variável do grupo (Identidade Partidária) tem três categorias. O modelo de regressão para o procedimento da ANOVA com y = ideologia política é:

$$E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2.$$

As variáveis auxiliares satisfazem $z_1 = 1$ somente para Democratas, $z_2 = 1$ somente para os Independentes e $z_1 = z_2 = 0$ para Republicanos. A Tabela 12.6 mostra uma parte de uma saída para o ajuste deste modelo. Nenhuma estimativa da variável auxiliar aparece na tabela para o Partido 3 (Republicanos) porque é redundante incluir uma variável auxiliar para o último grupo.

A equação de previsão é $\hat{y} = 4{,}70 - 1{,}47z_1 - 0{,}80z_2$. Os coeficientes na equação de previsão se relacionam às médias amostrais da mesma forma que os parâmetros da regressão se relacionam às médias populacionais. Assim como $\alpha = \mu_3$, da mesma forma será a sua estimativa 4,70 $= \bar{y}_3$, a média amostral para os Republicanos. Do mesmo modo, o coeficiente de z_1 é $-1{,}47 = \bar{y}_1 - \bar{y}_3$ e o coeficiente de z_2 é $-0{,}80 = \bar{y}_2 - \bar{y}_3$. ■

Tabela 12.6 Saída para o ajuste do modelo de regressão $E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2$ para os dados de y = ideologia política com variáveis auxiliares z_1 e z_2 para a identificação partidária

Variável dependente: IDEOLOGIA

Parâmetro		Estimativa	Erro padrão	t	Sig
(Constante)		4,703	0,0759	59,73	0,0001
PARTIDO	1	−1,472	0,1033	−6,94	0,0001
	2	−0,802	0,1054	−5,13	0,0001
	3	0,000	0,0	0,0	0,0

Regressão para o teste ANOVA comparando médias

Para três grupos, a hipótese nula no teste F da ANOVA é H_0: $\mu_1 = \mu_2 = \mu_3$. Se H_0 é verdadeira, então $\mu_1 - \mu_3 = 0$ e $\mu_2 - \mu_3 = 0$. Lembre que $\mu_1 - \mu_3 = \beta_1$ e $\mu_2 - \mu_3 = \beta_2$ no modelo de regressão $E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2$ com variáveis auxiliares. Assim, a hipótese da ANOVA é equivalente a H_0: $\beta_1 = \beta_2 = 0$ naquele modelo. Se todos os valores-β no modelo são iguais a 0, então a média da variável resposta é igual a α para cada grupo. Estabelecendo as variáveis auxiliares, podemos executar o teste da ANOVA usando o teste F de $H_0 = \beta_1 = \beta_2 = 0$ para este modelo.

A suposição da análise de regressão de que as distribuições condicionais de y sobre a equação de regressão são normais com desvio padrão constante implica aqui que as distribuições da população para os grupos sejam normais, com o mesmo desvio padrão para cada grupo. Estas são precisamente as suposições para o teste F da ANOVA.

EXEMPLO 12.6 Regressão para comparar as médias da ideologia política por identificação partidária

A Tabela 12.7 mostra a soma dos quadrados para ajustar o modelo de regressão, com variáveis auxiliares, aos dados da ideologia política e identificação partidária. Observe a similaridade entre esse procedimento e a tabela da ANOVA na Tabela 12.2. A "soma dos quadrados entre grupos" na ANOVA é a "soma dos quadrados da regressão" (também chamada de "soma dos quadrados do modelo" por alguns *softwares*) na análise de regressão. A "soma dos quadrados dentro dos grupos" na ANOVA é a "soma dos quadrados dos resíduos" (também chamada de "soma dos quadrados dos erros") e representada por SQE. Esta é a variabilidade dentro dos grupos não explicada pela inclusão dos parâmetros no modelo para explicar as diferenças entre as médias. A soma dos quadrados dos erros dividida pelos seus graus de liberdade é o erro quadrático médio (EQM), que é a estimativa dentro dos grupos $s^2 = 1{,}68$ da variância das observações para cada grupo. A média dos quadrados da regressão é a estimativa entre grupos.

A razão entre a média dos quadrados da regressão e o erro quadrático médio é a estatística $F = 26{,}3$, com $gl_1 = 2$ e $gl_2 = 273$, para testar H_0: $\beta_1 = \beta_2 = 0$. Esta hipótese é equivalente a H_0: $\mu_1 = \mu_2 = \mu_3$ para as três identificações partidárias. A análise de regressão fornece a mesma estatística F que a ANOVA forneceu na Seção 12.1. ■

Tabela 12.7 Saída mostrando a soma dos quadrados para o modelo de regressão $E(y) = \alpha + \beta_1 z_1 + \beta_2 z_2$ para a modelagem de ideologia política em termos da identificação partidária

	Soma dos quadrados	gl	Média dos quadrados	Valor F	Sig
Regressão	88,43	2	44,21	26,3	0,000
Resíduo	459,52	273	1,68		
Total	547,95	275			

Por que usar a regressão para executar uma ANOVA?

Visto que o teste F da ANOVA é fácil e visto que a ANOVA está amplamente disponível nos *softwares*, por que deveríamos nos importar que ela possa ser feita usando a regressão? Uma razão é que é útil ter uma abordagem unificada para a qual a maioria dos métodos estatísticos conduzidos na prática são casos especiais. (Na verdade, veremos no Capítulo 14 que a regressão, ela própria, é um caso especial de um tipo ainda mais geral de modelo que pode tratar de variáveis respostas que não tenham distribuições normais, tais como as variáveis categóricas.) Outra razão é que ser capaz de tratar dos previsores categóricos utilizando um modelo de regressão nos fornece um mecanismo para modelar *vários* previsores que podem ser categóricos ou uma mistura de categórico e quantitativo. Aprenderemos sobre tais métodos na Seção 12.5 e no próximo capítulo.

12.4 ANÁLISE DE VARIÂNCIA DE DOIS FATORES

Vimos como comparar as médias para grupos que são categorias de uma variável explicativa categórica. Algumas vezes, os grupos se referem a duas (ou mais) variáveis categóricas. Por exemplo, os grupos (homens brancos, mulheres brancas, homens negros, mulheres negras) resultam de uma classificação cruzada de raça e gênero. O método para comparar a média de uma variável resposta quantitativa por intermédio das categorias de cada uma de duas variáveis categóricas é chamado de **ANOVA de dois fatores**.

A ANOVA discutida até agora, tendo uma única variável explicativa, é chamada de **ANOVA de um fator**. Ela ignora as outras variáveis. Os Capítulos 10 e 11 mostraram que tais análises geralmente não são informativas como as análises multivariadas que controlam outras variáveis. O restante deste capítulo trata da ANOVA de dois fatores e de métodos mais complexos para variáveis explicativas categóricas e de controle.

Hipóteses do efeito principal na ANOVA de dois fatores

A ANOVA de dois fatores compara as médias da população por meio das categorias de duas variáveis explicativas. Cada hipótese nula declara que as médias da população são idênticas ao longo das categorias de uma variável categórica, controlada pelo outra.

Para ilustrar a ANOVA de dois fatores, analisamos a ideologia política média usando as variáveis explicativas identificação partidária e gênero. Seis médias resultam das $2 \times 3 = 6$ combinações das suas categorias, como a Tabela 12.8 mostra. Os rótulos subscritos identificam os grupos, tal como a média da população μ_{MD} para mulheres Democratas.

Uma análise possível compara a ideologia política média para as três identificações partidárias, controladas pelo gênero. Para mulheres, comparamos as médias μ_{MD}, μ_{MI} e μ_{MR} para as três identificações partidárias; para homens, comparamos μ_{HD}, μ_{HI} e μ_{HR}. Outra análise possível compara a ideologia política média para homens e mulheres, controlada pela identificação partidária, comparando as médias dentro de cada coluna da tabela.

A Tabela 12.9a exibe um conjunto de médias da população satisfazendo a hipótese nula de que a ideologia política média é a mesma para as três identificações partidárias, controladas pelo gênero.

A Tabela 12.9b exibe um conjunto de médias populacionais satisfazendo a hipótese nula de que a média da ideologia partidária é idêntica para os dois gêneros, controlada a identificação partidária. Os efeitos dos previsores individuais testados nestas duas hipóteses nulas são chamados de **efeitos principais**.

Tabela 12.8 Uma classificação de dois fatores da média populacional de ideologia partidária por identificação partidária e gênero

	Identificação partidária		
Gênero	Democrata	Independente	Republicano
Mulher	μ_{MD}	μ_{MI}	μ_{MR}
Homem	μ_{HD}	μ_{HI}	μ_{HR}

Tabela 12.9 Ideologia política média da população satisfazendo o efeito principal da hipótese nula: (a) sem efeito da identificação partidária, (b) com efeito do gênero

		Identificação partidária		
Tabela	Gênero	Democrata	Independente	Republicano
(a)	Mulher	3,0	3,0	3,0
	Homem	5,0	5,0	5,0
(b)	Mulher	3,0	4,0	5,0
	Homem	3,0	4,0	5,0

Testes F sobre os efeitos principais

Os testes F para a ANOVA de dois fatores têm as mesmas suposições que as do teste F para a ANOVA de um fator: aleatorização, distribuição normal da população e o mesmo desvio padrão em cada grupo. Agora os grupos são formados pelas células da classificação cruzada das duas variáveis explicativas.

As estatísticas-teste para a ANOVA de dois fatores têm fórmulas complexas exceto quando os tamanhos da amostra em todas as células são iguais. Por isso utilizaremos um *software*. Como ocorreu com a ANOVA de um fator, o teste para o efeito do previsor usa duas estimativas da variância para cada grupo. Essas estimativas aparecem na coluna da média dos quadrados (MQ) da tabela da ANOVA. Para testar o efeito principal para um previsor, a estatística-teste é a razão das médias dos quadrados,

$$F = \frac{\text{MQ para o previsor}}{\text{MQ do erro}}.$$

A MQ para o previsor é a estimativa da variância baseada na variação entre grupos para aquele previsor. Essa estimativa é

Tabela 12.10 Dados da PSG sobre ideologia política por identificação partidária e gênero

		Ideologia política									
Identificação partidária	Gênero	1	2	3	4	5	6	7	Tamanho da amostra	Média	Desvio padrão
Democrata	Mulher	5	30	35	98	20	24	3	215	3,85	1,26
	Homem	6	20	25	41	15	15	3	125	3,77	1,43
Independente	Mulher	4	17	27	83	16	17	5	169	3,95	1,24
	Homem	4	16	20	59	21	23	1	144	4,04	1,30
Republicano	Mulher	2	10	17	63	32	33	5	162	4,43	1,26
	Homem	0	9	13	36	33	28	9	128	4,66	1,31

inflacionada quando H_0 não é verdadeira. O erro da MQ (representado por EQM) é a estimativa da variância dentro dos grupos que sempre é não tendenciosa e é, também, usada em intervalos de confiança.

As estimativas da variância, listadas na tabela da ANOVA na coluna da média dos quadrados, divide a soma dos quadrados pelo valor dos seus graus de liberdade. Os graus de liberdade para as estatísticas F são $gl_1 = gl$ para a estimativa do numerador e $gl_2 = gl$ para o EQM. O valor do gl_1 é sempre um a menos do que o número dos grupos sendo comparados; por exemplo, $gl_1 = 2$ para comparar as três identificações partidárias. Como sempre, valores maiores F fornecem uma evidência mais forte contra H_0, assim cada valor-p é uma probabilidade unicaudal à direita.

EXEMPLO 12.7 ANOVA de dois fatores para a ideologia política por identificação partidária e gênero

A Tabela 12.10 mostra os dados da PSG para a ideologia política por identificação partidária (sem restrição de idade). A tabela também mostra as médias amostrais e desvios padrão da ideologia política, baseado nos escores (1, 2, 3, 4, 5, 6, 7).

O *software* apresenta um resumo da análise na Tabela 12.11. O erro quadrático médio (EQM) estima a variância da população σ^2 dentro de cada célula. Para testar o efeito principal ele é igual a:

$$s^2 = \text{EQM} = \frac{\text{SQE}}{gl} = \frac{1569{,}53}{939} = 1{,}67.$$

Para a hipótese nula de nenhuma diferença na ideologia política para as três identificações partidárias, controladas pelo gênero, a Tabela 12.11 mostra que a estatística-teste F é:

$$F = \frac{\text{Média dos quadrados da identificação partidária}}{\text{Erro quadrático médio}} = \frac{42{,}13}{1{,}67} = 25{,}2,$$

com $gl_1 = 2$ e $gl_2 = 939$, o valor-p é 0,0001. Existe uma evidência muito forte de uma diferença na média da ideologia política entre as três identificações partidárias, controladas pelo gênero. Para a hipótese nula de nenhuma diferença na média da ideologia política entre mulheres e homens, controlados pela identificação partidária, a estatística-teste F é:

$$F = \frac{\text{Média dos quadrados do gênero}}{\text{Erro quadrático médio}} = \frac{1{,}31}{1{,}67} = 0{,}78,$$

com $gl_1 = 1$ e $gl_2 = 939$, o valor-p é 0,38. Não existe evidência de que a média da ideologia política varie por gênero, dentro de cada identificação partidária. ■

Tabela 12.11 Tabela da ANOVA para uma análise de dois fatores dos efeitos principais da identificação partidária e gênero sobre a ideologia política média

Variável dependente: IDEOLOGIA

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Modelo	86,693	3	28,898	17,29	0,0001
Erro	1569,525	939	1,671		
Total	1656,218	942			
Fonte	SQ do tipo III	gl	Média dos quadrados	F	Sig
PARTIDO	84,2516	2	42,1258	25,20	0,0001
GÊNERO	1,3110	1	1,3110	0,78	0,3760

Tabela 12.12 Médias populacionais para uma classificação de dois fatores não mostrando interação

Gênero	Democrata	Independente	Republicano
Mulher	3,0	3,5	5,0
Homem	4,0	4,5	6,0

Interação na ANOVA de dois fatores

Na prática, antes de realizarmos os testes dos efeitos principais recém descritos, primeiro testamos outra hipótese nula. As Seções 10.3 (página 344) e 11.5 (página 380) mostraram que o estudo de *interação* é importante sempre que analisamos relacionamentos multivariados. Uma falta de interação entre duas variáveis explicativas significa que o efeito de ambas as variáveis na variável resposta (na população) não muda para diferentes níveis da outra.

Suponha que não exista interação entre gênero e identificação partidária nos seus efeitos sobre a ideologia política. Então, a diferença entre mulheres e homens na média populacional de ideologia política é a mesma para cada identificação partidária. A Tabela 12.12 mostra as médias populacionais satisfazendo isso. A diferença entre mulheres e homens na média ideologia política é −1,0 para cada partido. De forma similar, a diferença entre cada par de partidos na média da ideologia política é a mesma para cada gênero. A diferença entre Republicanos e Democratas, por exemplo, é igual a 2,0 tanto para as mulheres quanto para os homens. A Figura 12.3 traça as médias para as categorias da identificação partidária, dentro de cada gênero. A ordem das categorias no eixo horizontal não tem importância, visto que a identificação partidária é nominal. A ausência de interação é indicada pelas sequências paralelas dos pontos.

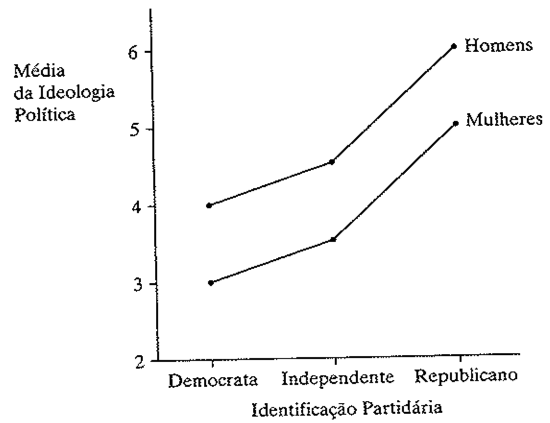


Figura 12.3 Média da ideologia política, por identificação partidária e gênero não mostrando interação.

☑ **Tabela 12.13** Médias populacionais para uma classificação de dois fatores mostrando interação

Gênero	Democrata	Independente	Republicano
Mulheres	3,0	4,0	5,0
Homens	5,0	4,0	3,0

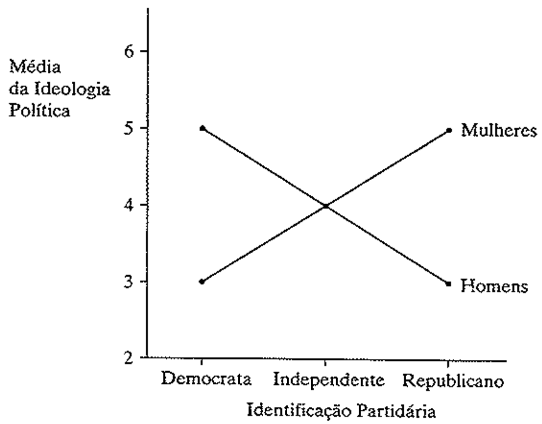
Ao contrário, a Tabela 12.13 e a Figura 12.4 mostram as médias populacionais com interação. A diferença entre mulheres e homens na média da ideologia política é −2 para Democratas, 0 para Independentes e +2 para Republicanos. Assim, a diferença depende da identificação partidária. Da mesma forma, o efeito da identificação partidária na ideologia difere entre mulheres e homens. Para as mulheres, as Republicanas são as mais conservadoras, enquanto que, para os homens, os Democratas são os mais conservadores.

Na Tabela 12.13, suponha que o número de homens e mulheres seja igual para cada identificação partidária. Então, a média geral da ideologia política, ignorando o gênero, é 4,0 para cada partido. A diferença geral nas médias entre quaisquer duas identificações partidárias é igual a 0. Em uma comparação de um fator, a interação implica diferenciarem-se os efeitos da identificação partidária para homens e mulheres. Como em outras análises multivariadas, o efeito de um previsor pode mudar dramaticamente quando controlamos outra variável.

Teste F de H_0: Nenhuma interação

Antes de testar a hipótese de efeitos principais em uma ANOVA de dois fatores, primeiro testamos H_0: nenhuma interação. A estatística-teste F é a razão da média dos quadrados baseada no grau de interação amostral dividido pelo EQM. Um valor-p pequeno sugere que cada previsor categórico tem um efeito na resposta, mas o tamanho do efeito varia de acordo com a categoria do outro previsor.

Quando existe uma interação, não é significativo testar a hipótese dos efeitos



☑ **Figura 12.4** Média da ideologia política, por identificação partidária e gênero mostrando interação.

principais. Quando rejeitamos H_0: nenhuma interação, concluímos que cada variável tem um efeito, mas a natureza daquele efeito muda de acordo com a categoria da outra variável. Então, é melhor comparar as médias para um previsor separadamente dentro das categorias da outra. Por outro lado, se a evidência de interação não for forte (isto é, se o valor-p não for pequeno), então testamos as duas hipóteses dos efeitos principais.

A Tabela 12.10 mostrou a média amostral da ideologia política para as seis combinações da identificação partidária e gênero. Essas médias não mostram evidência de interação. O único efeito amostral que parece ser substancial é a diferença entre Republicamos e Democratas e o efeito parece similar para cada gênero – a saber, uma média mais alta para Republicanos. Na próxima seção veremos que o teste H_0: nenhuma interação tem $F = 1{,}06$ e um valor-p de 0,34. Portanto, uma ausência de interação é plausível e os testes dos efeitos principais são válidos.

A seguir, é natural usar intervalos de confiança para descobrir mais sobre o efeito da filiação partidária, controlada por gênero. Para fazer isto, é útil estudar a ANOVA de dois fatores no contexto do modelo de regressão. Como no teste da ANOVA de um fator, os testes F na ANOVA de dois fatores são testes equivalentes sobre os parâmetros em um modelo de regressão. A próxima seção mostra como realizar a análise usando modelos de regressão múltipla com variáveis auxiliares para os previsores categóricos.

12.5 ANOVA DE DOIS FATORES E REGRESSÃO

Para realizar os testes F como casos especiais dos testes sobre os parâmetros de um modelo de regressão múltipla, organizamos as variáveis auxiliares para cada previsor. Ilustramos com a Tabela 12.10, usando a identificação partidária e gênero (sexo) como previsores da ideologia política.

Usamos o símbolo p para as variáveis auxiliares da identificação partidária e s como uma variável auxiliar para sexo (se o sujeito é uma mulher). Isto é,

$$p_1 = \begin{cases} 1 & \text{se o sujeito é Democrata} \\ 0 & \text{caso contrário} \end{cases}$$

$$p_2 = \begin{cases} 1 & \text{se o sujeito é Independente} \\ 0 & \text{caso contrário} \end{cases}$$

Ambos p_1 e p_2 são iguais a 0 quando o sujeito é Republicano. Também,

$$s = \begin{cases} 1 & \text{se o sujeito é mulher} \\ 0 & \text{se o sujeito é homem} \end{cases}$$

É redundante incluir variáveis auxiliares para as categorias finais.

☑ **Tabela 12.14** Médias populacionais para ideologia política para a classificação de dois fatores de identificação partidária e gênero (sexo) não assumindo interação

		Variáveis auxiliares			Média populacional de y
Gênero	Identificação partidária	p_1	p_2	s	$\alpha + \beta_1 p_1 + \beta_2 p_2 + \beta_3 s$
Mulher	Democrata	1	0	1	$\alpha + \beta_1 + \beta_3$
	Independente	0	1	1	$\alpha + \beta_2 + \beta_3$
	Republicano	0	0	1	$\alpha + \beta_3$
Homem	Democrata	1	0	0	$\alpha + \beta_1$
	Independente	0	1	0	$\alpha + \beta_2$
	Republicano	0	0	0	α

Modelo de regressão não assumindo interação

Para maior clareza, assumiremos primeiro que não existe interação. Na prática, verificaríamos esta suposição primeiro, como discutimos na seção anterior. O modelo de regressão, nesse caso, é:

$$E(y) = \alpha + \beta_1 p_1 + \beta_2 p_2 + \beta_3 s.$$

Para encontrar a correspondência entre as médias populacionais e os parâmetros de regressão, substituímos os valores possíveis para as variáveis auxiliares. Para ilustrar, para Republicanos ($p_1 = p_2 = 0$), a média da ideologia política é:

$$\text{Homens } (s = 0): \mu = \alpha + \beta_1(0) + \beta_2(0) + \beta_3(0) = \alpha.$$

$$\text{Mulheres } (s = 1): \mu = \alpha + \beta_1(0) + \beta_2(0) + \beta_3(1) = \alpha + \beta_3.$$

Para as seis combinações de identificação partidária e gênero, a Tabela 12.14 mostra as médias populacionais em termos dos parâmetros de regressão. Para cada identificação partidária, a diferença nas médias entre mulheres e homens é igual a β_3, isto é, o coeficiente de β_3 da variável auxiliar s para sexo é igual à diferença entre mulheres e homens na média da ideologia política, controlada pela identificação partidária. A hipótese nula de inexistência de diferença média entre mulheres e homens, controlada pela identificação partidária, é H_0: $\beta_3 = 0$.

O teste de H_0: $\beta_3 = 0$ de inexistência do efeito de gênero compara o modelo completo

$$E(y) = \alpha + \beta_1 p_1 + \beta_2 p_2 + \beta_3 s$$

ao modelo reduzido

$$E(y) = \alpha + \beta_1 p_1 + \beta_2 p_2.$$

O modelo reduzido não tem o efeito do gênero. A tabela da ANOVA em Tabela 12.11 mostrou que:

Fonte	SQ do tipo III	gl	Média dos quadrados	F	Sig
PARTIDO	84,2516	2	42,1258	25,20	0,0001
GÊNERO	1,3110	1	1,3110	0,78	0,3760

A soma dos quadrados do gênero de 1,31 é a quantidade de variação responsável pela introdução do termo $\beta_3 s$ no modelo, uma vez que os outros termos já estão lá. Ele representa a diferença entre as somas dos quadrados dos erros (SQE) quando esses termos são omitidos e quando são incluídos. Existe uma diferença de um parâmetro nos dois modelos, assim esta soma dos quadrados tem $gl = 1$. A tabela mostra que a estatística-teste F é igual a 0,78, com valor-p de 0,38.

No modelo completo, β_1 é a diferença entre as médias de Democratas e Republicanos, e β_2 é a diferença entre as médias de Independentes e Republicanos, controlados por gênero. As interpretações são similares aos valores-β para o modelo de regressão da ANOVA de dois fatores, exceto que aqui também controlamos o gênero. A hipótese nula de inexistência de diferença entre os partidos na ideologia política média, controlada por gênero, é H_0: $\beta_1 = \beta_2 = 0$. A Tabela 12.11 ou o seu excerto acima mostra que a estatística-teste F é igual a 25,2, com um valor-p de 0,0001.

A Tabela 12.15 mostra uma saída para o ajuste do modelo de regressão completo. A equação de previsão é:

$$\hat{y} = 4{,}58 - 0{,}71p_1 - 0{,}54p_2 - 0{,}08s.$$

O coeficiente de s é $-0{,}08$. Isto é a diferença estimada entre mulheres e homens na ideologia política média para cada identificação partidária. O teste do efeito principal do gênero indicou que esta diferença não é estatisticamente significativa. O coeficiente de p_1 é $-0{,}71$. Esta é a diferença estimada entre Democratas e Republicanos na ideologia política média para cada gênero. O coeficiente de p_2 é $-0{,}54$. Isto é a diferença estimada entre Independentes e Republicanos para cada gênero. A diferença estimada entre Democratas e Independentes é $(-0{,}71) - (-0{,}54) = -0{,}17$ para cada gênero.

Substituir os valores da variável auxiliar na equação de previsão gera médias estimadas que satisfazem o modelo de inexistência de interação. Por exemplo, para mulheres Republicanas, $p_1 = p_2 = 0$ e $s = 1$, assim $\hat{y} = 4{,}58 - 0{,}71(0) + 0{,}54(0) - 0{,}08(1) = 4{,}50$.

Tabela 12.15 Ajuste do modelo de regressão para a análise de dois fatores da ideologia política média por identificação partidária e gênero (sexo), assumindo a inexistência de interação. A estimativa é 0 no último nível de cada previsor porque uma variável auxiliar para aquele nível não é necessária e seria redundante

Variável Dependente: IDEOLOGIA

Parâmetro		B	Erro padrão	t	Sig
Intercepto		4,5768	0,0897	51,02	0,0001
PARTIDO	1	−0,7112	0,1035	−6,87	0,0001
	2	−0,5423	0,1054	−5,15	0,0001
	3	0	.	.	.
GÊNERO	1	−0,0758	0,0856	−0,89	0,3760
	2	0	.	.	.

Modelo de regressão com interação

O modelo considerado até agora é inadequado quando existe interação. A Seção 11.5 (página 380) mostrou que os termos do produto cruzado em um modelo de regressão múltipla podem representar interação. Aqui, pegamos os produtos cruzados das variáveis auxiliares para obter um modelo de regressão que permite efeitos de interação.

O modelo de interação para a classificação de dois fatores da identificação partidária e gênero é:

$$E(y) = \alpha + \beta_1 p_1 + \beta_2 p_2 + \beta_3 s + \beta_4(p_1 \times s) + \beta_5(p_2 \times s).$$

Os dois últimos termos usam produtos cruzados para a interação. Não é necessário pegar produtos cruzados das variáveis auxiliares das categorias do mesmo previsor categórico, tal como $p_1 \times p_2$. Isto é porque não mais do que uma variável auxiliar para um previsor dado pode ser não zero para qualquer observação, visto que uma observação não pode estar em mais do que uma categoria. Todos esses produtos cruzados seriam iguais a 0.

A Tabela 12.16 mostra a tabela da ANOVA para o modelo que permite interação. A soma dos quadrados para a interação, exibida na linha com o rótulo do produto PARTIDO*GÊNERO (PARTY*GENDER), é a quantidade da variação explicada pelos dois termos de interação. Ele é igual à diferença entre SQE sem e com estes termos no modelo. A média dos quadrados da interação é uma estimativa de σ^2 baseada em:

$$\frac{\text{Interação SQ}}{gl} = \frac{3{,}64}{2} = 1{,}82.$$

Testamos H_0: sem interação, isto é, H_0: $\beta_4 = \beta_5 = 0$, usando:

Tabela 12.16 Tabela da ANOVA para a análise de dois fatores da ideologia política média por identificação partidária e gênero (sexo), permitindo interação

Variável dependente: IDEOLOGIA

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Modelo	90,332	5	18,066	10,81	0,0001
Erro	1565,886	937	1,671		
Total	1656,218	942			
Fonte	SQ do tipo III	gl	Média dos quadrados	F	Sig
PARTIDO	87,795	2	43,898	26,27	0,0001
GÊNERO	1,488	1	1,488	0,89	0,3456
PARTIDO*GÊNERO	3,640	2	1,820	1,09	0,3370

$$F = \frac{\text{Média dos quadrados da interação}}{\text{Erro quadrático médio}} = \frac{1,82}{1,67} = 1,09.$$

Aqui, $gl_1 = 2$ para a interação SQ, porque o modelo permitindo a interação tem dois parâmetros extras. Como de costume, o valor do gl_2 para SQE é igual ao tamanho da amostra total menos o número de parâmetros no modelo. Para o modelo completo, existem 6 parâmetros, assim $gl_2 = 943 - 6 = 937$. Para a distribuição F com $gl_1 = 2$ e $gl_2 = 943$, o valor-p é 0,34.

Não existe muita evidência de interação. É sensato remover os termos dos produtos cruzados e usar modelos mais simples discutidos previamente. Visto que uma ausência de interação é plausível, os testes de efeitos principais apresentados na Tabela 12.11 para a identificação partidária e gênero são válidos. Quando não existe evidência significativa de interação, é melhor usar o modelo sem os termos de interação no teste dos efeitos principais dos previsores e na construção dos intervalos de confiança para os efeitos.

Somas parciais dos quadrados

A soma dos quadrados para partido, gênero e sua interações nas Tabelas 12.11 e 12.16 são chamadas de **somas parciais dos quadrados** (veja Exercício 11.60 na página 408). Alguns *softwares* as rotulam de **somas dos quadrados do Tipo III**. Elas representam a variabilidade em y explicada por estes termos, uma vez que os outros termos já estão no modelo. Isto é igual à diferença entre os valores do SQE para o modelo sem aqueles termos e o modelo com eles.

Suponha que cada célula na classificação cruzada dos dois previsores tenha o mesmo número de observações. Isto acontece com experimentos projetados que atribuem o mesmo número de sujeitos para cada célula. Então, os previsores são independentes, no sentido (para tabelas de contingência do Capítulo 8) de que as frequências nas células da classificação cruzada satisfazem a condição de independência. Nesse caso, essas somas dos quadrados explicam completamente porções separadas da variabilidade em y. Elas, então, somam a soma dos quadrados do modelo (chamada de SQ da *regressão* por alguns *softwares*). O modelo da soma dos quadrados é igual a STQ − SQE. Ele representa a variabilidade de explicada conjuntamente com todos os termos do modelo de regressão.

Para pesquisas com levantamentos de dados observacionais, os previsores são raramente independentes. O gênero, por exemplo, está de alguma forma associado à identidade partidária. Por causa disso, a soma parcial dos quadrados para a identificação partidária se sobrepõe um pouco à soma parcial dos quadrados para o gênero. Consequentemente, as somas dos quadrados do Tipo III listadas na Tabelas 12.11 e 12.16 não dão exatamente o mesmo resultado da soma dos quadrados do modelo.

Quando os previsores categóricos estão associados, a soma parcial dos quadrados explicada por um previsor depende da existência dos termos de interação no modelo. As somas parciais dos quadrados para o partido e para o gênero diferem levemente entre as Tabelas 12.11 e 12.16.

Comparações múltiplas seguindo a ANOVA de dois fatores

Na prática, usamos intervalos de confiança para estimar os tamanhos dos efeitos. Suponha que H_0: sem interação pareça plausível. Então, podemos tratar a diferença nas médias da população entre duas categorias para um previsor como a mesma em cada categoria da outra. Assim, construímos um único conjunto de comparações em vez de um conjunto separado em cada categoria da outra variável. No Exemplo 12.6, estimamos as diferenças na média da ideologia política entre cada par da identificação partidária, controlado por gênero, num total de três comparações. Podemos fazer isto usando intervalos de confiança comuns para os parâmetros da regressão. A forma é a usual da estimativa mais e menos o escore-t vezes o erro padrão, com gl para t sendo o valor do gl para o erro quadrático médio.

Por exemplo, para o modelo assumindo a inexistência de interação, $\hat{\beta}_1 = -0,71$ é a diferença estimada entre Democratas e Republicanos na ideologia política média, controlada por gênero. O erro padrão dessa estimativa, informado na Tabela 12.15, é 0,104. Um intervalo de 95% de confiança é $-0,71 \pm 1,96(0,104)$ ou $(-0,9; -0,5)$. Os Democratas são menos conservadores, em média, para cada gênero.

A abordagem Bonferroni (Seção 12.2) para a ANOVA de um fator se estende para a ANOVA com maior número de fatores. A comparação de todos os três pares da identificação partidária, com a taxa de erro das comparações múltiplas de 0,05, usa uma probabilidade de erro de 0,05/3 = 0,0167 na determinação do escore-t para cada intervalo. Para estes dados, obtemos intervalos similares àqueles exibidos na Tabela 12.3 seguindo a ANOVA de um fator.

Quando existe um grau de interação praticamente significativo não é apropriado fazer resumo das comparações das categorias de uma variável, controlada pela outra. Ao contrário, compare os pares das linhas separadamente dentro de cada coluna ou compare os pares das colunas separadamente dentro de cada linha.

ANOVA fatorial

Os métodos da ANOVA de dois fatores se estendem a vários previsores. As variáveis explicativas categóricas na ANOVA são geralmente chamadas de **fatores**. Uma ANOVA multifator com observações em todas as combinações dos fatores é chamada de **ANOVA fatorial**.

Por exemplo, com três fatores, a **ANOVA de três fatores** considera as possíveis interações bem como os efeitos principais para esses fatores. Para fatores representados por A, B e C, o modelo completo contém um fator principal para cada fator, A × B, A × C e B × C interações de dois fatores e A × B × C interação de três fatores. O modelo de regressão tem um conjunto de variáveis auxiliares para cada fator, produtos cruzados de pares de variáveis auxiliares para interações de dois fatores e produtos de três fatores de variáveis auxiliares de todos os três fatores para a interação de três fatores.

Na ANOVA de três fatores, primeiro testamos a interação dos três fatores. Se o valor-p é pequeno, comparamos os pares de categorias para uma variável em cada combinação de categorias das outras duas. Caso contrário, é melhor tirar o termo de três fatores do modelo e testar as interações de dois fatores. Suponha, por exemplo, que o valor-*p* é pequeno para a interação A × B, mas não para as outras. Após ajustar novamente o modelo com os efeitos principais e a interação A × B, podemos testar o efeito principal C e comparar pares de médias para vários pares de categorias de C. Por causa da interação A × B, comparamos médias das categorias A separadamente em cada categoria B e comparamos médias das categorias B separadamente em cada categoria A.

Quando você tem dois ou mais fatores, por que não executar ANOVAs de um fator separadas? Por exemplo, você poderia comparar a média da ideologia política para mulheres e homens usando a ANOVA de um fator, ignorando a informação sobre a identificação partidária. Da mesma forma, você poderia executar separado uma ANOVA de um fator para comparar as médias para as três identificações partidárias, ignorando a informação sobre o gênero. A razão principal é que com esta ANOVA fatorial descobrimos se existe interação. Quando existe, é mais informativo comparar níveis de um fator separadamente em cada nível do outro fator. Isso nos permite investigar como o efeito depende daquele outro fator.

Outro benefício da ANOVA fatorial é que a variabilidade residual, que afeta MQ do erro e os denominadores das estatísticas-teste *F*, tende a diminuir. Quando usamos dois ou mais fatores para prever a variável resposta, geralmente tendemos a conseguir melhores previsões (isto é, menos variabilidade residual) do que quando usamos um fator. Com menos variabilidade residual (dentro dos grupos), conseguimos estatísticas-teste maiores e, por conseguinte, maior poder para rejeitar a hipótese nula.

12.6 ANÁLISE DE VARIÂNCIA DE MEDIDAS REPETIDAS*

Os métodos apresentados até agora assumem que as amostras nos grupos são *independentes*, cada grupo tendo uma amostra separada de sujeitos. Em muitos estudos, entretanto, cada grupo tem os mesmos sujeitos. Isto acontece mais comumente quando existem *medidas repetidas* dos sujeitos ao longo do tempo ou em várias variáveis respostas relacionadas. As amostras são, então, *dependentes* e a análise deve levar isto em consideração.

EXEMPLO 12.8 Influências positivas e negativas nas crianças

Uma PSG recente pediu aos sujeitos para responder o seguinte: "As crianças são expostas a muitas influências no seu dia a dia. Que tipo de influência cada um dos seguintes itens têm nas crianças? 1. Filmes, 2. Programas na TV a cabo, 3. *Rock*." As respostas possíveis foram (muito negativa, negativa, neutra, positiva, muito positiva). A Tabela 12.17 mostra as respostas para 12 dos sujeitos amostrados, usando escores (−2, −1, 0, 1, 2) para as respostas possíveis. Isto é parte de um arquivo de dados muito maior de mais de 1000 respondentes. Analisamos somente essa pequena amostra aqui para ajudar na explicação dos conceitos e assim você pode facilmente usar o seu próprio *software* para tentar replicar os resultados. ■

ANOVA de um fator com medidas repetidas

Para a Tabela 12.17, H_0 é a mesma da ANOVA de um fator comum: médias populacionais iguais para vários grupos. Existe evidência de que as médias populacionais diferem para as três influências? A ANOVA comum é não é apropriada porque as três amostras para as categorias de influência não são independentes. Cada amostra tem os mesmos sujeitos.

Suponha que consideramos tanto as linhas da Tabela 12.17 quanto as colunas, como um fator. Então, o *layout* dos dados se parece uma ANOVA de dois fatores. Cada cruzamento de uma célula classifica um sujeito (uma linha) com uma influência (uma coluna). Um modelo de regressão poderia expressar a resposta esperada como uma função de duas variáveis auxiliares para as 3 influências e 11 variáveis auxiliares para os 12 sujeitos. O teste comparando as médias populacionais para as três influências é, então, o teste do efeito principal para a variável coluna na ANOVA de dois fatores. Na verdade, este é o teste apropriado para dados desse tipo.

A Tabela 12.18 mostra os resultados para a ANOVA de dois fatores. Considere:

H_0: Médias populacionais iguais para as três influências.

A estatística-teste *F* é a média dos quadrados para a influência dividida pelo erro quadrático médio, que é $F = 1{,}44/0{,}57 = 2{,}55$. Os valores dos *gl* são $gl_1 = 2$ e $gl_2 = 22$. O valor-*p* é igual a 0,10. A evidência contra H_0 não é forte. Mas, com somente

☑ **Tabela 12.17** Opiniões sobre três influências nas crianças. Os escores representam −2 = muito negativa, −1 = negativa, 0 = neutra, 1 = positiva, 2 = muito positiva

	Influência		
Sujeito	**Filmes**	**TV**	***Rock***
1	−1	0	−1
2	1	0	0
3	0	1	−2
4	2	0	1
5	0	−1	−1
6	−2	−2	−2
7	−1	−1	0
8	0	1	−1
9	−1	−1	−1
10	1	0	1
11	1	1	−1
12	−1	−1	−2
Média	−0,08	−0,25	−0,75

☑ **Tabela 12.18** Tabela da ANOVA para medidas repetidas da variável opinião por tipo de influência. Esta tabela mostra resultados do uso do *software* para a ANOVA de dois fatores, tratando o sujeito como um segundo fator

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Modelo	27,861	13	2,143	3,79	0,003
Erro	12,444	22	0,566		
Total	40,306	35			
Fonte	**SQ do tipo III**	**gl**	**Média dos quadrados**	**F**	**Sig**
INFLUÊNCIA	2,889	2	1,444	2,55	0,101
SUJEITO	24,972	11	2,270	4,01	0,003

12 sujeitos, se H_0 é falsa, o poder é provavelmente baixo.

A Tabela 12.19 mostra que obtemos resultados similares se usarmos um *software* especializado para medidas repetidas, tal como o SPSS, usando a opção *Repeated Measures* (Medidas Repetidas) após selecionar *General Linear Model* (Modelo Linear Geral) no menu *Analyse* (Analisar).

Quando existem somente dois grupos, com os mesmos sujeitos em cada um, a Seção 7.4 (página 222) mostrou que a inferência usa a distribuição t com diferenças dos escores. Para testar a igualdade de médias, a estatística F da ANOVA é então igual ao quadrado da estatística t do teste t pareado.

A suposição de esfericidade e simetria combinada

A ANOVA de medidas repetidas tradicional assume **esfericidade**. Falando em termos gerais, isto significa o seguinte. Para cada par de grupos, considere a diferença entre duas observações, uma para cada grupo. Esta diferença é uma variável e a condição de esfericidade declara que o desvio padrão da distribuição dessa diferença é idêntico para cada par de grupos. É mais fácil perceber um caso especial de esfericidade, chamado de **simetria combinada**. Esta condição se sustenta quando os grupos diferentes têm os mesmos desvios padrão e quando cada par de respostas tem a mesma correlação.

Se a suposição de esfericidade é fortemente violada, o valor-p tende a ser muito pequeno. A maioria dos *softwares* fornece um teste de significância formal (teste de Mauchly) para a hipótese de esfericidade. Quando os dados contradizem fortemente a suposição, um teste aproximado ajusta os graus de liberdade para baixo para a estatística-teste usual, usando o *ajuste de Greenhouse-Geisser*. Os detalhes técnicos para estes testes e ajustes estão além do conteúdo deste livro, mas os *softwares* informam esses resultados.

Usar um delineamento de medidas repetidas pode melhorar a precisão da estimativa. Ter os mesmos sujeitos em cada grupo ajuda a eliminar fontes de erro externas. Por exemplo, outras variáveis que afetam a resposta têm os mesmos valores para cada grupo, assim as diferenças entre as médias do grupo não podem refletir as diferenças entre grupos naquelas variáveis. Controlar os possíveis fatores de confusão mantendo-os fixos em cada linha do arquivo dos dados é denominado de **blocagem**. Para detalhes na ligação dos procedimentos da análise de variância com delineamentos experimentais, veja Howell (2006), Kirk (1995) e Winer e colaboradores (1991).

Intervalos de confiança comparando amostras dependentes

Como de costume, aprendemos mais estimando os parâmetros. A Tabela 12.17 mostrou as médias amostrais para os três fatores de influência. Visto que o tamanho da amostra é pequeno, enfraquecemos um pouco o nível de confiança das comparações múltiplas de forma que os intervalos não sejam excessivamente grandes. Os intervalos de confiança de 90% de Bonferroni usam a probabilidade de erro 0,10/3 = 0,0333 para cada intervalo. Os gl = 22 do erro e o escore-t com probabilidade 0,0333/2 = 0,0167 em cada cauda é t = 2,27. A raiz quadrada do erro quadrático médio é igual a $s = \sqrt{0{,}566} = 0{,}75$. Cada grupo tem 12 observações, assim a margem de erro para cada intervalo de confiança é:

$$ts\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} =$$

$$2{,}27(0{,}75)\sqrt{\frac{1}{12} + \frac{1}{12}} = 0{,}70.$$

Por exemplo, o intervalo de confiança para a diferença entre os valores médios dos filmes e do rock é $(-0{,}08) - (-0{,}75) \pm 0{,}70$ ou $(-0{,}03; 1{,}37)$. É plausível que as médias sejam iguais, mas também plausível que a média dos filmes esteja mais na direção positiva do que a do *rock*. A Tabela 12.20 mostra todas as três comparações de Bonferroni. É também plausível que a média para TV esteja mais na direção positiva do que a do *rock*. Os intervalos de confiança podem transmitir informação útil mesmo se a estatística-teste geral não for significativa.

Efeitos fixos e efeitos aleatórios

O modelo de regressão para a análise anterior é:

$$E(y) = \alpha + \beta_1 m + \beta_2 t + \gamma_1 s_1 + \gamma_2 s_2 + \ldots + \gamma_{11} s_{11}.$$

Aqui, y é a avaliação de uma influência, m é uma variável auxiliar para filmes (isto é, $m = 1$ para uma resposta sobre filmes, 0 caso contrário), t é uma variável auxiliar para TV ($t = 1$ para uma resposta sobre TV, 0 caso contrário) e $m = t = 0$ para uma resposta sobre *rock*. Da mesma forma, s_1 é uma variável auxiliar para o sujeito 1, igualando 1 para as três respostas daquele sujeito e 0 caso contrário e do mesmo modo para outras 10 variáveis auxiliares do sujeito. Usamos γ (gama) em vez de β para os coeficientes destes termos por conveniência, assim o índice do parâmetro concorda com o índice da variável auxiliar. Como de costume, cada fator tem uma variável auxiliar a menos do que seu número de categorias.

Uma forma abreviada de escrever esse modelo de regressão é:

$$E(y) = \alpha + \beta_j + \gamma_i.$$

Aqui, β_j representa o efeito para a influência j e γ_i é o efeito para o sujeito i, onde $\beta_3 = 0$ e $\gamma_{12} = 0$ para a categoria final de cada variável. Esta equação expressa a resposta esperada na célula da linha i e coluna j em termos de um efeito principal da linha e um da coluna. Testar a igualdade da média populacional y para as três influências corresponde a testar H_0: $\beta_1 = \beta_2 = 0$.

Nesse modelo, o foco principal é comparar os parâmetros da influência $\{\beta_j\}$ e não os parâmetros do sujeito $\{\gamma_i\}$. O $\{\gamma_i\}$ depende de quais sujeitos são escolhidos

Tabela 12.19 Saída parcial do SPSS para a ANOVA de medidas repetidas das variáveis opinião por tipo de influência

Fonte	Teste dos efeitos dentre sujeitos Soma dos quadrados do Tipo III	gl	Média dos quadrados	F	Sig
Esfericidade assumida					
Influência	2,889	2	1,444	2,55	0,101
Erro	12,444	22	0,566		

Tabela 12.20 Intervalos de 90% de confiança das comparações múltiplas de Bonferroni para comparar as respostas médias para as três influências

Influências	Diferença das médias	Intervalo de confiança
Filmes, TV	0,17	(−0,53; 0,87)
Filmes, *Rock*	0,67	(−0,03; 1,37)
TV, *Rock*	0,50	(−0,20; 1,20)

para a amostra. O efeito do sujeito é chamado de **efeito aleatório**, porque as categorias do fator do sujeito representam uma amostra aleatória de todos os sujeitos possíveis. Ao contrário, o fator que define os grupos, o tipo de influência, é chamado de **efeito fixo**. A análise usa *todas* as categorias de interesse de um efeito fixo em vez de uma amostra aleatória delas. Os modelos estudados nas seções anteriores desse capítulo continham somente efeitos fixos.

Quando as variáveis de classificação são uma mistura de efeitos aleatórios e fixos, tal como nesse exemplo, o modelo é denominado de **misto**. Para modelos mistos mais complexos do que esse, as estatísticas teste diferem para alguns testes da sua forma quando todas as variáveis de classificação são efeitos fixos. A próxima seção discute um caso importante.

12.7 ANOVA DE DOIS FATORES COM MEDIDAS REPETIDAS EM UM FATOR*

Conjuntos de dados de medidas repetidas geralmente têm mais do que um efeito fixo. As medidas repetidas normalmente ocorrem nas categorias de um fator, mas as categorias do(s) outro(s) fator(es) têm amostras independentes. O próximo exemplo ilustra a situação.

EXEMPLO 12.9 Comparação de três tratamentos para a anorexia

Para 72 jovens sofrendo de anorexia, a Tabela 12.21 mostra os seus pesos antes e após um período experimental. As meninas foram designadas aleatoriamente para receber uma de três terapias durante o período. Um grupo, o de controle, recebeu a terapia padrão. O estudo analisou se um tratamento é melhor do que os outros, com as meninas tendendo a ganhar mais peso durante o tratamento. As análises de partes desse conjunto de dados foram apresentadas nas páginas 144, 177 e 226.

A Figura 12.5 mostra diagramas de caixa e bigodes, descrevendo graficamente as distribuições da resposta antes e depois do período experimental de cada tratamento. A Tabela 12.22 mostra o resumo das médias amostrais. Os três tratamentos têm, originalmente, distribuições similares. Isto não é surpreendente porque os sujeitos foram aleatoriamente alocados aos três grupos naquele momento. Existe alguma evidência de um maior ganho médio de peso no grupo da terapia familiar, embora existam alguns valores atípicos baixos nos pesos desse grupo. ■

Medidas repetidas em um de dois efeitos fixos

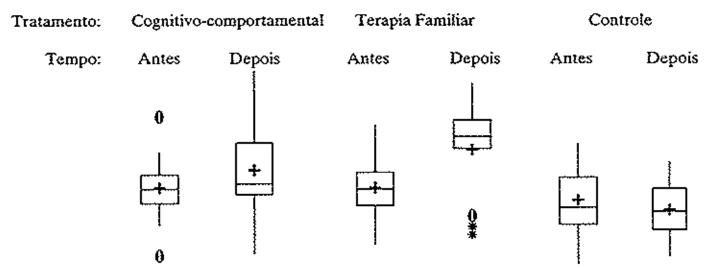
As Tabelas 12.21 e 12.22 têm dois efeitos fixos. Um deles, "Tratamento", tem categorias (CC = cognitivo-comportamental, TF = terapia familiar, C = controle). Ele define três grupos de meninas, representadas pelas três amostras independentes. O segundo efeito fixo, "Tempo", consiste nos dois períodos de observações, (antes, depois). Cada tempo tem os mesmos sujeitos, assim as amostras nos seus respectivos níveis são independentes. O tempo é denominado de **fator dentre sujeitos**, porque as comparações das suas categorias usam medidas repetidas nos sujeitos. O tratamento é denominado de **fator entre sujeitos**, porque as comparações das suas categorias usam sujeitos diferentes.

Embora os dois fatores (tratamento e tempo) sejam efeitos fixos, a análise difere da ANOVA de dois fatores normal. Isto é porque as medidas repetidas no fator dentre sujeitos (tempo) cria um terceiro efeito, um efeito aleatório para os sujeitos. Cada sujeito é mensurado em cada categoria do tempo. Os sujeitos estão sendo **cruzados** com o fator tempo. Cada sujeito ocorre em somente uma categoria dos sujeitos entre

☑ **Tabela 12.21** Pesos de meninas anoréxicas, em libras, antes e depois de receber um de três tratamentos

Cognitivo-comportamental		Terapia familiar		Controle	
		Pesos			
Antes	Depois	Antes	Depois	Antes	Depois
80,5	82,2	83,8	95,2	80,7	80,2
84,9	85,6	83,3	94,3	89,4	80,1
81,5	81,4	86,0	91,5	91,8	86,4
82,6	81,9	82,5	91,9	74,0	86,3
79,9	76,4	86,7	100,3	78,1	76,1
88,7	103,6	79,6	76,7	88,3	78,1
94,9	98,4	76,9	76,8	87,3	75,1
76,3	93,4	94,2	101,6	75,1	86,7
81,0	73,4	73,4	94,9	80,6	73,5
80,5	82,1	80,5	75,2	78,4	84,6
85,0	96,7	81,6	77,8	77,6	77,4
89,2	95,3	82,1	95,5	88,7	79,5
81,3	82,4	77,6	90,7	81,3	89,6
76,5	72,5	83,5	92,5	78,1	81,4
70,0	90,9	89,9	93,8	70,5	81,8
80,4	71,3	86,0	91,7	77,3	77,3
83,3	85,4	87,3	98,0	85,2	84,2
83,0	81,6			86,0	75,4
87,7	89,1			84,1	79,5
84,2	83,9			79,7	73,0
86,4	82,7			85,5	88,3
76,5	75,7			84,4	84,7
80,2	82,6			79,6	81,4
87,8	100,4			77,5	81,2
83,3	85,2			72,3	88,2
79,7	83,6			89,0	78,8
84,5	84,6				
80,8	96,2				
87,4	86,7				

Fonte: Agradecimentos ao Professor Brian Everitt do Institute of Psychiatry, Londres, por estes dados.



☑ **Figura 12.5** Diagramas de caixa e bigodes para meninas anoréxicas, por tratamento e tempo de avaliação.

Tabela 12.22 Peso médio da amostra, por tratamento e por tempo avaliado no estudo da anorexia

	Tempo	
Tratamento	**Antes**	**Depois**
Cognitivo-comportamental (CC)	82,7	85,7
Terapia familiar (TF)	83,2	90,5
Controle (C)	81,6	81,1

fatores (tratamento). Os sujeitos estão **aninhados** dentro do fator tratamento.

Como na ANOVA normal de dois fatores, podemos testar cada efeito principal como também suas interações. Contudo, os testes sobre o fator dentre sujeitos (tanto no efeito principal como na sua interação como o outro fator fixo) usam um termo erro diferente do que o teste sobre o efeito principal entre sujeitos. O termo normal da soma dos quadrados dos erros é particionado em duas partes. Uma parte usa a variabilidade entre os escores médios dos sujeitos. Ela forma um termo erro para testar o fator entre sujeitos. A outra parte é baseada em como o padrão dos escores dentre sujeitos varia entre os sujeitos. Ela forma um termo do erro para qualquer teste envolvendo o fator dentro dos grupos.

A Figura 12.6 mostra a partição da soma dos quadrados total para uma ANOVA de dois fatores com medidas repetidas em um fator. O *software* executa automaticamente esta divisão e cria estatísticas F na forma apropriada para testar os efeitos principais e a interação.

EXEMPLO 12.10 Testes *F* da ANOVA comparando os tratamentos da anorexia

A Tabela 12.23 mostra uma saída do SPSS para a análise dos dados do estudo da anorexia. Cada soma dos quadrados resume a variação que seus termos explicam em um modelo de regressão correspondente. Esta é a redução no SQE quando acrescentamos termos ao modelo para aquele efeito. Visto que o tratamento tem três categorias, ele tem duas variáveis auxiliares no modelo de regressão e sua soma dos quadrados tem $gl = 2$. Visto que tempo tem dois níveis, ele tem uma variável auxiliar e sua soma dos quadrados tem $gl = 1$. A interação entre estes efeitos tem dois termos no modelo, baseados no produto cruzado das duas variáveis auxiliares para o tratamento com a única variável auxiliar para o tempo, assim seu $gl = 2$.

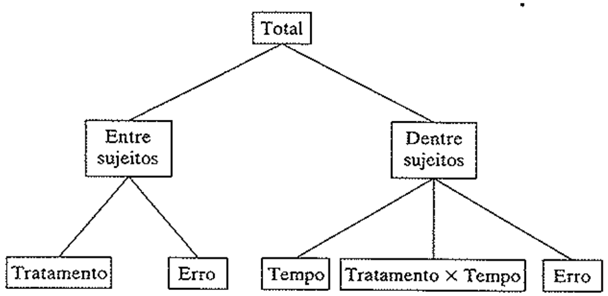


Figura 12.6 Partição da variabilidade na ANOVA de dois fatores com o tratamento e fatores tempo e com medidas repetidas no tempo. Os testes envolvendo o fator dentre sujeitos (tempo) usa um termo erro separado.

Tabela 12.23 Saída do computador para a análise de variância de dois fatores da Tabela 12.21 com efeitos fixos do tratamento e do tempo com medidas repetidas no tempo

Teste dos efeitos dentre sujeitos

Fonte	Soma dos quadrados do erro do Tipo III	gl	Média dos quadrados	F	Sig
TEMPO	366,04	1	366,04	12,92	0,001
TEMPO*TRATAMENTO	307,32	2	153,66	5,42	0,006
Erro(TEMPO)	1955,37	69	28,34		

Testes dos efeitos entre sujeitos

Fonte	Soma dos quadrados do erro do Tipo III	gl	Média dos quadrados	F	Sig
TRATAMENTO	644,23	2	322,12	6,20	0,003
Erro	3584,03	69	51,94		

O termo do erro para a parte da tabela entre sujeitos tem $gl = 69$, baseados nas 28 variáveis auxiliares para os 29 sujeitos recebendo a terapia CC, 16 variáveis auxiliares para os 17 sujeitos recebendo TF e 25 variáveis auxiliares para os 26 sujeitos no grupo C ($28 + 16 + 25 = 69$). A variabilidade restante, não responsável por este termo erro ou pelos efeitos principais e termos de interação, é a soma dos quadrados dos erros para testar os efeitos dentre sujeitos.

A linha da tabela da ANOVA TEMPO*TRATAMENTO (TIME*TREATMENT) indica que a interação é altamente significativa. O valor-p é igual a 0,006. A diferença entre as médias populacionais para os dois tempos difere de acordo com o tratamento e a diferença entre as médias populacionais para um par de tratamentos varia de acordo com o tempo. Por causa da interação significativa, não testamos os efeitos principais. Precisamos, ao contrário, usar intervalos de confiança para descrever a interação. ■

Acompanhando os intervalos de confiança

A Tabela 12.22 mostrou as médias amostrais para as seis combinações dos dois fatores. A evidência de interação é clara. As médias amostrais para os três tratamentos são similares no tempo inicial. No tempo final, ao contrário, a média para o grupo de controle é similar à sua média no tempo inicial, mas a média é maior para os outros dois tratamentos do que suas médias iniciais, especialmente para o grupo TF.

Para construir intervalos de confiança para comparar as médias nos dois tempos, para cada tratamento, a estimativa apropriada comum do desvio padrão é a raiz quadrada do erro quadrático médio da análise dentre sujeitos. Da Tabela 12.23, isto é igual a $\sqrt{28,34} = 5,3$, com $gl = 69$. Ilustramos construindo um intervalo de 95% de confiança comparando as duas médias para a terapia familiar (TF), que 17 meninas receberam em cada tempo. O escore-t para 95% de confiança quando $gl = 69$ é igual a 1,99. O intervalo de confiança tem margem de erro igual a este escore-*t* vezes a raiz do erro quadrático médio vezes a raiz quadrada do fator envolvendo o inverso de cada tamanho da amostra (17 para cada tempo). Este intervalo é igual a:

$$(90{,}5 - 83{,}2) \pm 1{,}99(5{,}3)\sqrt{\frac{1}{17} + \frac{1}{17}}, \text{ ou}$$
$$7{,}3 \pm 3{,}6 \text{ ou ainda } (3{,}6; 10{,}9).$$

Para a terapia TF, concluímos que o peso médio da população está entre 3,6 e 10,9 libras mais alto depois do período de tratamento. De forma similar, um intervalo comparando as duas médias é igual a (0,2; 5,8) para a terapia CC e (−3,4; 2,5) para o grupo de controle. Existe evidência de um aumento, embora pequeno, para a terapia CC, mas nenhuma evidência de mudança para o grupo de controle.

Para fazer comparações dos tratamentos entre sujeitos, para cada tempo, não podemos usar o erro quadrático médio da análise entre sujeitos. A razão é que estas comparações separadas envolvem tanto o efeito principal do tratamento quanto a interação e estas duas fontes de variação têm termos dos erros diferentes nas medidas repetidas da ANOVA. Em um tempo em particular, entretanto, os sujeitos nos três tratamentos são amostras independentes. Portanto, podemos comparar três médias em um tempo dado usando o teste *F* da ANOVA de um fator ou usando intervalos de confiança apenas para aqueles dados.

Por exemplo, para as 72 observações no tempo = final, a estatística-teste *F* para a ANOVA de um fator comparando as três médias é 8,6, com $gl_1 = 2$ e $gl_2 = 69$. O valor-*p* é 0,0004, evidência muito forte de uma diferença entre as médias dos tratamentos. Para esta ANOVA de um fator, a raiz quadrada do EQM é igual a $s = 7{,}3$. O intervalo de 95% de confiança para a diferença das médias entre os tratamentos TF e CC, baseados em 17 + 29 observações para os dois grupos, é igual a:

$$(90{,}5 - 85{,}7) \pm 1{,}99(7{,}3)\sqrt{\frac{1}{17} + \frac{1}{29}}, \quad \text{ou} \quad 4{,}8 \pm 4{,}4 \quad \text{ou ainda} \quad (0{,}4; 9{,}2).$$

Concluímos que, no tempo do acompanhamento, o peso médio está entre 0,4 e 9,2 libras mais alto com o tratamento TF do que com o tratamento CC. As médias verdadeiras podem ser essencialmente iguais, mas, se elas diferem, a vantagem poderia ser bem considerável para a terapia familiar. A Tabela 12.24 mostra os intervalos de confiança para cada par de tratamentos.

Em resumo, existe evidência de que o peso médio aumenta durante o período experimental para ambos os tratamentos não controlados. Existe uma evidência marginal de que a média é mais alta, após o experimento, para o tratamento TF do que para o tratamento CC.

Neste momento, um interesse adicional pode estar relacionado a se a mudança nas médias, entre o tempo depois e antes, são direrentes para os dois tratamentos não controlados. Isto é, se a diferença dos escores para o tratamento TF tem uma média significativamente maior do que a diferença dos escores para o tratamento CC? A diferença dos escores tem uma média de (90,5 − 83,2) = 7,3 para o tratamento TF e (85,7 − 82,7) = 3,0 para o tratamento CC e usamos isto como a base dos intervalos de confiança separados para a mudança média. Visto que os dois grupos são amostras independentes, a variância da diferença dessas médias é a soma das variâncias. Um intervalo de 95% de confiança para a diferença entre as mudanças médias nos pesos é, então:

☑ **Tabela 12.24** Intervalos de 95% de confiança comparando as médias dos tratamentos após o período de acompanhamento

Tratamentos comparados	Diferença das médias amostrais	Intervalos de confiança	Intervalos de Bonferroni
TF − CC	4,8	(0,4; 9,2)	(−0,7; 10,3)
TF − C	9,4	(4,9; 13,9)	(3,8; 15,0)
CC − C	4,6	(0,7; 8,5)	(−0,2; 9,4)

$$(7{,}3 - 3{,}0) \pm 1{,}99(5{,}3)\sqrt{\frac{1}{17} + \frac{1}{17} + \frac{1}{29} + \frac{1}{29}}, \quad \text{ou} \quad 4{,}3 \pm 4{,}6 \quad \text{ou ainda} \quad (-0{,}3; 8{,}8).$$

Embora a mudança média pudesse ser consideravelmente maior para o tratamento TF, é também possível que elas não sejam diferentes.

Comparações múltiplas de Bonferroni dos tratamentos

Para controlar a taxa de erro geral para as várias comparações, podemos usar o método das comparações múltiplas de Bonferroni. Suponha que usamos três intervalos de confiança para comparar tratamentos no tempo final e três intervalos para comparar os tempos dentro dos tratamentos. Para assegurar pelo menos 90% de confiança para todo o conjunto, visto que 0,10/6 = 0,0167, usamos um intervalo de 98,33% de confiança para cada comparação individual.

Tais intervalos são maiores do que aqueles recém informados, visto que usam um escore-*t* de 2,45 em vez de um de 1,99. A Tabela 12.24 mostra esses intervalos para as comparações pareadas de tratamentos no tempo final. Com esta abordagem mais conservadora, somente a diferença entre os tratamentos TF e C é significativa, com o respectivo intervalo não contendo 0.

Análise de medidas repetidas mais complexas

A ANOVA de dois fatores com medidas repetidas em um fator se estende para delineamentos mais complexos. Suponha, por exemplo, que o estudo tem três fatores, A, B e C, com medidas repetidas em C. Isto é, os sujeitos são cruzados com C, mas aninhados dentro de combinações de níveis A e B. Os efeitos entre sujeitos, a saber, os efeitos principais A e B e a interação A × B, são testados com um erro quadrático médio baseado na variabilidade entre sujeitos. Todos os efeitos envolvendo o fator C dentre sujeitos, a saber, o efeito principal C, a interação A × C, a interação B × C e a interação A × B × C, são testados com um erro quadrático médio separado.

Em alguns estudos com dois efeitos fixos, as medidas repetidas ocorrem em ambos os fatores. Por exemplo, podemos observar os mesmos sujeitos para cada tratamento em cada um de vários períodos de tempo. Então, os sujeitos (um efeito aleatório) são cruzados com ambos os fatores (efeitos fixos) e uma observação ocorre para cada sujeito em cada combinação dos níveis dos fatores. Como na ANOVA de dois fatores normal, os efeitos de interesse se referem aos efeitos fixos – seus efeitos principais e a interação. O fator de complicação é que cada teste requer um erro quadrático médio separado, mas um *software* pode facilmente executar essa análise.

Medidas repetidas em mais do que dois tempos

No Exemplo 12.8 as medidas repetidas ocorrem em dois períodos de tempo. Quando as observações são feitas em vários períodos de tempo, as medidas repetidas da ANOVA são mais complexas. Em particular, os resultados dependem das suposições sobre a estrutura da correlação das medidas repetidas. O teste padrão para o efeito dentre sujeitos assume *esfericidade*, como na ANOVA de um fator de medidas repetidas. Os testes dos efeitos entre sujeitos não são afetados pela violação da esfericidade, assim nenhum ajuste é necessário para aquele teste F.

As abordagens da ANOVA *multivariada* alternativa (MANOVA) para testar os efeitos dentre sujeitos (tais como teste lambda da razão da verossimilhança de Wilks) fazem poucas suposições, mas geralmente têm um poder fraco. Outros métodos recentemente desenvolvidos permitem uma variedade maior de tipos de efeitos aleatórios de modelagem da estrutura de correlações para as respostas repetidas, fornecendo outras opções além da esfericidade. Discutiremos tais abordagens na Seção 16.1.

12.8 EFEITOS DAS VIOLAÇÕES DAS SUPOSIÇÕES DA ANOVA

Cada método da ANOVA apresentado neste capítulo assume (além da aleatorização) que os grupos tenham distribuições normais com os mesmos desvios padrão. Essas são suposições limitadas. Como nas suposições da regressão comum, elas nunca são exatamente satisfeitas na prática.

Robustez dos testes *F*

Desvios moderados da normalidade nas distribuições da população podem ser tolerados. A distribuição F ainda fornece uma boa aproximação para a distribuição amostral real de uma estatística-teste F. Isto é particularmente verdadeiro para tamanhos grandes de amostras, visto que as distribuições amostrais, então, têm uma dependência fraca da forma da distribuição populacional. Desvios moderados da suposição de mesma variabilidade dentro dos grupos podem, também, ser tolerados. Quando os tamanhos da amostra são idênticos para todos os grupos, o teste F é bastante robusto à violações desta suposição.

Construir histogramas para cada distribuição dos dados amostrais ajuda a verificar desvios extremos destas suposições. Resultados enganosos podem ocorrer nos testes F se as distribuições da população são altamente assimétricas e o tamanho da amostra é pequeno ou se existem relativamente grandes diferenças entre os desvios padrão populacionais (digamos, o desvio padrão máximo é várias vezes maior do que o mínimo) e os tamanhos da amostras são desiguais. A Seção 14.5 discute abordagens de regressão alternativas para estas violações grosseiras. Quando as distribuições são altamente assimétricas, a média pode nem mesmo ser uma boa medida.

Intervalos de confiança, como os testes, não são altamente dependentes da suposição de normalidade. Quando os desvios padrão são bem diferentes, com a razão do maior para o menor excedendo aproximadamente 2, é preferível usar fórmulas para os intervalos baseadas em desvios padrão separados para os grupos em vez de um único valor agrupado. Por exemplo, o método do intervalo de confiança apresentado na Seção 7.3 (página 219) para dois grupos não assume a igualdade dos desvios padrão.

Como nas outras inferências, a qualidade da amostra é crucial. Na ANOVA de um fator, por exemplo, as conclusões podem ser inválidas se as observações nos grupos separados comparados não são amostras aleatórias independentes. Os procedimentos da ANOVA não são robustos às violações das suposições amostrais.

Vamos considerar a validade das suposições para a ANOVA de dois fatores para os dados da Tabela 12.10 sobre a ideologia política por identificação partidária e gênero. Os desvios padrão da amostra são similares para os seis grupos (entre 1,24 e 1,43). Além disso, os tamanhos da amostra são grandes (125 ou mais), assim a suposição de normalidade não é crucial. Toda a amostra da PSG foi obtida aleatoriamente, assim podemos considerar as seis amostras como amostras aleatórias independentes. A ANOVA é adequada para estes dados.

O teste de Kruskal-Wallis: uma abordagem não paramétrica

O **teste de Kruskal-Wallis** é uma alternativa para a ANOVA de um fator para comparar vários grupos. É um método não paramétrico, não requerendo a suposição de normalidade. A estatística-teste usa somente a informação ordinal dos dados. Ela classifica as observações e compara as classificações médias para vários grupos. A estatística-teste é maior quando as diferenças entre as classificações médias são maiores. Ela tem uma distribuição aproximadamente qui-quadrado com $gl = g - 1$.

Este teste é especialmente usado para amostras pequenas nas quais os efeitos de desvios severos da normalidade podem ser influentes. Não apresentaremos a estatística-teste aqui. Seu resultado é similar ao do teste do qui-quadrado para o efeito de um previsor categórico para uma resposta normal apresentada na Seção 15.4. Na prática, é mais informativo usar uma abordagem de modelagem porque as estimativas do parâmetro do modelo nos fornecem informação sobre os tamanhos do efeito. Além disso, a estratégia de modelagem se adapta melhor às análises multivariadas.

Existem, também, testes não paramétricos para análises mais complexas. Por exemplo, o **teste de Friedman** é uma alternativa ao teste F da Seção 12.6 para comparar grupos quando os mesmos sujeitos ocorrem em cada um. Uma vantagem dos métodos paramétricos é que eles mais facilmente generalizam a modelagem multivariada e a estimativa dos efeitos, que são mais importantes do que o teste de significância. Lehmann (1975) é uma boa fonte para detalhes sobre métodos não paramétricos.

12.9 RESUMO DO CAPÍTLO

Este capítulo apresentou os métodos da **análise de variância (ANOVA)** para comparar vários grupos de acordo com suas médias em uma variável resposta quantitativa. Os grupos são categorias de uma ou mais variáveis explicativas categóricas.

- Os métodos da **ANOVA de um fator** comparam as médias para categorias de uma única variável explicativa.
- Os métodos da **ANOVA de dois fatores** comparam as médias ao longo das categorias de cada duas variáveis explicativas. Não assumindo interação, os efeitos principais descrevem o efeito de cada previsor enquanto o outro é controlado.
- Os métodos de **comparações múltiplas** fornecem intervalos de confiança para a diferença entre cada par de médias, enquanto controlam a probabilidade de erro do Tipo I. O método **Bonferroni** faz isto usando uma probabilidade de erro para cada comparação que é igual à probabilidade do erro desejada dividida pelo número de comparações.
- Os métodos de análises de variâncias são casos especiais da análise de regressão múltipla. As **variáveis auxiliares** no modelo de regressão representam categorias que definem os grupos.

Cada variável auxiliar é igual a 1 para uma categoria em particular e 0 caso contrário.

Os métodos da ANOVA comuns comparam grupos com amostras aleatórias *independentes* dos grupos. Para alguns estudos, amostras diferentes, ao contrário, têm os mesmos sujeitos, como quando um experimento observa sujeitos repetidamente ao longo do tempo. Os métodos para a **ANOVA de medidas repetidas** resultam de modelos de regressão com **efeitos principais** que representam os efeitos da amostra aleatória dos sujeitos observados. Tais métodos tratam os efeitos **dentre sujeitos** (para medidas repetidas nos sujeitos) de forma diferente dos efeitos **entre sujeitos** (para amostras independentes de sujeitos).

Os Capítulos 9 e 11 apresentaram modelos de regressão para uma variável resposta quantitativa quando as variáveis explicativas são também *quantitativas*. Este capítulo modelou uma variável resposta quantitativa como uma função das variáveis explicativas *categóricas*. A Tabela 12.25 resume os testes estatísticos discutidos nas Seções 12.1, 12.4 e 12.6.

Os modelos para o próximo capítulo incluem variáveis explicativas categóricas e também quantitativas.

EXERCÍCIOS

Praticando o básico

12.1 Um estudo[1] das Forças Armadas Norte-Americanas que serviram no Iraque ou Afeganistão relatou que, daqueles que estiveram em combate, o valor-*p* era 0,001 para a comparação de combates relatados por soldados servindo no Afeganistão, no Iraque e fuzileiros navais servindo no Iraque. Explique como isto se relaciona à análise de variância, identificando a variável resposta, a explicativa e a hipótese que poderia ser testada para gerar este valor-*p*.

12.2 Uma PSG perguntou quantos bons amigos as pessoas tinham. Isto está associado ao signo astrológico do respondente (os 12 signos do Zodíaco)? A tabela da ANOVA para os dados da PSG informa $F = 0{,}61$ baseado em $gl_1 = 11$, $gl_2 = 813$.

(a) Especifique a hipótese nula e a alternativa para essa ANOVA.

☑ **Tabela 12.25** Testes da ANOVA para comparar vários grupos em uma variável resposta

Elemento do Teste	ANOVA de um fator	ANOVA de dois fatores	ANOVA de medidas repetidas
1. Amostras	Independentes	Independentes	Dependentes
2. Hipóteses	H_0: Médias idênticas H_a: pelo menos duas médias desiguais	H_0: Médias idênticas nas linhas H_0: Médias idênticas nas colunas H_0: Sem interação	H_0: Médias idênticas H_a: pelo menos duas médias desiguais
3. Estatística-teste	$F = \frac{\text{MQ entre grupos}}{\text{MQ dentro dos grupos}}$ Distribuição F $gl_1 = g - 1$ $gl_2 = N - g$	$F = \frac{\text{MQ efeitos}}{\text{MQ erros}}$ Distribuição F $gl_1 = gl$ para efeito $gl_2 = gl$ para erro	$F = \frac{\text{MQ efeitos}}{\text{MQ erros}}$ Distribuição F $gl_1 = gl$ para o efeito $gl_2 = gl$ para o erro
4. Valor-*p*	Probabilidade da cauda direita	Probabilidade da cauda direita	Probabilidade da cauda direita

(b) Baseado no que você sabe sobre a distribuição *F*, você acharia que $F = 0{,}61$ fornece uma evidência forte contra H_0? Explique.

(c) O *software* informa um valor-*p* de 0,82. Explique como interpretá-lo.

12.3 Para os dados da PSG comparando o número relatado de bons amigos para aqueles que são (casados, viúvos, divorciados, separados, nunca casados), uma tabela da ANOVA informa que $F = 0{,}80$.

(a) Especifique a hipótese nula e a alternativa para essa ANOVA.

(b) Baseado no que você sabe sobre a distribuição *F*, você acharia que $F = 0{,}80$ fornece uma evidência forte contra H_0? Explique.

(c) Um *software* informa um valor-*p* de 0,53. Explique como interpretá-lo.

12.4 Uma PSG recente perguntou: "Qual é o número ideal de filhos para uma família?". As respostas tendem a depender da afiliação religiosa? Os resultados de uma ANOVA estão na Tabela 12.26 para as categorias religiosas (Protestante, Católico, Judeu, Outra ou Nenhuma).

(a) Especifique as hipóteses testadas nesta tabela.

☑ **Tabela 12.26**

Fonte	SQ	gl	Média dos quadrados	F	Sig
Religião	11,72	3	3,91	5,48	0,001
Erro	922,82	1295	0,71		
Total	934,54	1298			

(b) Informe o valor da estatística-teste *F* e o valor-*p*. Interprete o valor-*p*.

(c) Baseado em (b), você pode concluir que *cada* par das afiliações religiosas tem médias populacionais diferentes para o tamanho ideal de uma família? Explique.

12.5 Uma PSG recente perguntou: "Com que frequência você vai a um bar ou taverna?". A Tabela 12.27 mostra as estatísticas descritivas e uma ANOVA para comparar o número médio de amigos informado nos três níveis dessa variável.

(a) Declare (i) a hipótese, (ii) o valor da estatística-teste, (iii) o valor-*p*, (iv) uma decisão para $\alpha = 0{,}05$ – nível do teste.

(b) Algum aspecto do resumo sugere que uma suposição para o teste da ANOVA pode ser seriamente violada? Explique.

☑ **Tabela 12.27**

	Com que frequência você vai a um bar ou taverna?		
	Muito frequentemente	Ocasionalmente	Nunca
Número médio bons amigos	12,1	6,4	6,2
Desvio padrão	21,3	10,2	14,0
Tamanho da amostra	41	166	215

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Grupo	1116,8	2	558,4	3,03	0,049
Erro	77171,8	419	184,2		
Total	78288,5	421			

12.6 A Tabela 12.28 mostra os escores do primeiro teste (máximo de 10 pontos) de um curso inicial de francês. Os estudantes do curso estão agrupados como segue:

Grupo A: Nunca estudou uma língua estrangeira antes, mas tem boas habilidades em inglês

Grupo B: Nunca estudou uma língua estrangeira antes; tem péssimas habilidades em inglês.

Grupo C: Estudou outra língua estrangeira.

☑ **Tabela 12.28**

Grupo A	Grupo B	Grupo C
4	1	9
6	5	10
8		5

(a) A Tabela 12.29 fornece resultados de uma ANOVA. Informe as suposições, hipóteses, estatística-teste e valor-*p*. Interprete o valor-*p*.

(b) Suponha que a primeira observação no segundo grupo era realmente 9,

☑ **Tabela 12.29**

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Entre Grupos	30,000	2	15,000	2,5000	0,1768
Dentro dos Grupos	30,000	5	6,00		
Total	60,000	7			

Grupo	n	Média	Desvio Padrão	Erro Padrão
Grupo 1	3	6,0000	2,0000	1,1547
Grupo 2	2	3,0000	2,8284	2,0000
Grupo 3	3	8,0000	2,6458	1,5275

não 1. Então, os desvios padrão são os mesmos que os relatados na tabela, mas as médias amostrais são 6, 7 e 8 em vez de 6, 3 e 8. Você acha que a estatística-teste F seria maior, a mesma ou menor? Explique o seu raciocínio sem fazer nenhum cálculo.

(c) Suponha que você tinha as mesmas médias que as mostradas na tabela, mas os desvios padrão eram 1,0; 1,8 e 1,6, em vez de 2,0; 2,8 e 2,6. Você acha que a estatística-teste F seria maior, a mesma ou menor? Explique o seu raciocínio.

(d) Suponha que você tinha as mesmas médias que as mostradas na tabela, mas os tamanhos das amostras eram 30, 20 e 30, em vez de 3, 2 e 3. Você acha que a estatística-teste F seria maior, a mesma ou menor? Explique o seu raciocínio.

(e) Em (b), (c) e (d), o valor-p seria maior, o mesmo ou menor? Por quê?

12.7 Use um *software* para reproduzir os resultados da ANOVA informados no exercício anterior.

12.8 Um grupo de proteção ao consumidor compara três tipos de parachoques dianteiros para uma marca de automóvel. Um teste é conduzido dirigindo um automóvel em direção a um muro de concreto a 15 milhas por hora. A resposta é a quantidade de danos no carro, mensurados como o custo do conserto, em centenas de dólares. Devido à potencialidade de grandes custos, o estudo conduz somente dois testes com cada tipo de parachoques. A Tabela 12.30 mostra os resultados.

☑ **Tabela 12.30**

Parachoque A	Parachoque B	Parachoque C
1	2	11
3	4	15

(a) Encontre a soma dos quadrados dentro dos grupos e a estimativa da variância associada.

(b) Encontre a soma dos quadrados entre grupos e sua estimativa da variância associada.

(c) Teste a hipótese de que os custos médios de conserto são os mesmos para os três tipos de parachoques. Informe a hipótese nula e a alternativa, a estatística-teste, os valores dos gl e o valor-p. Apresente os resultados em uma tabela da ANOVA e interprete.

12.9 Considere o exercício anterior. Construa intervalos de 94% de confiança de comparações múltiplas para as diferenças nos custos médios de reparos para cada par de parachoques. Interprete os resultados e indique quais os tipos de parachoques que são considerados diferentes quanto ao custo médio do conserto.

12.10 Em um estudo para comparar a satisfação do consumidor nos centros de serviços para o suporte técnico de PCs em San Jose (Califórnia), Toronto (Canadá) e Bangalore (Índia), cada centro amostrou aleatoriamente 100 pessoas que ligaram durante um período de duas semanas. Aqueles que ligaram avaliaram a sua satisfação em uma escala de 0 a 10, com escores mais altos representando maior satisfação. As médias amostrais foram 7,6 para San Jose, 7,8 para Toronto e 7,1 para Bangalore. A Tabela 12.31 mostra os resultados de uma ANOVA.

☑ **Tabela 12.31**

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Grupo	26,00	2	13,00	27,6	0,000
Erro	140,00	297	0,47		
Total	60,00	299			

(a) Explique como obter o valor da estatística-teste F informado na tabela dos valores da média dos quadrados mostrada. Informe os valores gl_1 e gl_2 para a distribuição F e informe e interprete o valor-p.

(b) Explique por que a margem de erro para intervalos de 95% de confiança separados é a mesma (0,19) para comparar as médias populacionais para cada par de cidades. Construa e interprete os três intervalos.

(c) A margem de erro para os intervalos de 95% de confiança das comparações múltiplas de Bonferroni ou de Tukey é 0,23. Por que ela é diferente da margem de erro em (b) e qual é a vantagem dessa abordagem?

(d) Com variáveis auxiliares para representar os centros de serviço, a equação de previsão é $\hat{y} = 7,1 + 0,5z_1 + 0,7z_2$. Mostre como os termos nesta equação se relacionam às médias amostrais de 7,6; 7,8 e 7,1.

12.11 A PSG de 2004 perguntou a 1829 sujeitos quantas horas por dia eles assistiam à televisão, em média. As médias amostrais foram 2,75 para brancos (n = 1435), 4,14 para negros (n = 305) e 2,61 para outros (n = 89). O intervalo de 95% de confiança comparando as médias populacionais de negros e brancos foi de (1,1; 1,7), de (−0,4; 0,7) para brancos e outra categoria e (0,9; 2,1) pra negros e outra categoria. Baseado nos intervalos de confiança, indique quais pares de médias são significativamente diferentes e interprete.

12.12 Uma PSG recente perguntou: "Você diria que está muito feliz, moderadamente feliz ou infeliz?" (variável "HAPPY"). A mesma PSG também perguntou: "Aproximadamente quantos bons amigos você tem?" (variável "NUMFRIEND"). A Tabela 12.32 resume os resultados.

(a) Proponha uma pergunta de pesquisa que você poderia responder com estes dados.

(b) Interprete os resultados do teste F.

(c) Um *software* informa que o intervalo de 95% de confiança de Tukey é (0,3; 5,7) para comparar muito feliz e moderadamente feliz, (−2,3; 6,5) para compar muito feliz e infeliz e (−5,1; 3,3) para comparar moderadamente feliz e infeliz. Interprete e indique quais grupos são significativamente diferentes.

12.13 Para g grupos com n = 100 cada, planejamos comparar todos os pares de médias populacionais. Queremos que a probabilidade seja igual a, pelo menos, 0,80 de que todo o conjunto de intervalos de confiança contenha as verdadeiras diferenças. Para o método Bonferroni, qual escore-t da tabela devemos usar para cada intervalo se **(a)** g = 10, **(b)** g = 5? Descreva como o escore-t depende do número de grupos e explique a consequência para à largura dos intervalos.

12.14 Uma psicóloga compara a quantidade média do tempo REM de sono para sujeitos sob três condições. Ela usa três grupos de sujeitos, com quatro sujeitos em cada grupo. A Tabela 12.33 mostra uma saída SAS para a análise. Interprete todas as informações dessa saída.

Tabela 12.32

	Muito feliz	Moderadamente feliz	Infeliz		
Média	10,4	7,4	8,3		
Desvio padrão	17,8	13,6	15,6		
Tamanho da amostra	276	468	87		
Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Grupo	1626,8	2	813,4	3,47	0,032
Erro	193900,9	828	234,2		
Total	195527,7	830			

Tabela 12.33

Fonte	GL	Soma dos quadrados	Média dos quadrados	Valor-F	Pr > F
Modelo	2	72,000	36,000	0,79	0,4813
Erro	9	408,000	45,333		
Total	11	480,000			

Testes T de Bonferroni para a variável: TEMPO

Alfa = 0,05 gl = 9 EQM = 45,33333

Valor crítico de T = 2,93

Diferença mínima significativa = 13,965

As médias com as mesmas letras não são significativamente diferentes.

Agrupamento Bon	Média	N	Grupo
A	18,000	4	3
A			
A	15,000	4	2
A			
A	12,000	4	1

12.15 Considere o exercício anterior.

(a) Estabeleça variáveis auxiliares para um modelo de regressão tal que um teste F para os parâmetros da regressão seja equivalente ao teste F da ANOVA. Expresse H_0 tanto em termos das médias populacionais quanto dos parâmetros da regressão.

(b) A equação de previsão obtida no ajuste da equação de regressão de (b) é $\hat{y} = 18 - 6z_1 - 3z_2$. Mostre como as estimativas do parâmetro se relacionam às médias amostrais da Tabela 12.33.

12.16 O Exercício 12.8 mostrou uma ANOVA para um experimento comparando três parachoques, com a média amostral dos danos de 2 para o parachoque A, 3 para o B e 13 para o C (em centenas de dólares).

(a) Estabeleça um modelo de regressão com variáveis auxiliares para a ANOVA.

(b) Mostre a correspondência entre a hipótese nula para as médias da ANOVA e a hipótese nula para os parâmetros da regressão.

(c) Informe a equação de previsão que você obteria para o modelo em (a). (Se não estiver seguro, use um *software* para encontrá-la.)

12.17 Quando usamos a PSG de 2004 para avaliar como o número médio de horas por dia assistindo à televisão depende do sexo e raça, obtemos os resultados exibidos na Tabela 12.34. As médias amostrais foram de 2,71 para mulheres brancas, 2,79 para homens brancos, 4,13 para mulheres negras e 4,16 para homens negros. Explique como estes resultados são compatíveis com os resultados dos testes exibidos.

Tabela 12.34

Fonte	SQ	gl	MQ	F	Sig
Sexo	2,22	1	2,22	0,35	0,555
Raça	489,65	1	489,65	76,62	0,000
Erro	11094,16	1737	6,39		
Total	11583,81	1739			

12.18 Uma PSG recente perguntou: "Qual é o número ideal de filhos para uma família?". A Tabela 12.35 mostrou os resultados de usar a PSG para avaliar os efeitos de gênero e raça.

(a) Explique como interpretar os resultados dos testes F.

(b) Considere $s = 1$ para mulheres e 0 para homens e considere $r = 1$ para negros e 0 para brancos. O modelo sem interação tem $\hat{y} = 2{,}42 + 0{,}04s + 0{,}37r$. Encontre a média estimada para cada combinação de gênero e raça. Explique como estas médias satisfazem "sem interação".

Tabela 12.35

Fonte	SQ	gl	MQ	F	Valor-p
Sexo	0,25	1	0,25	0,36	0,550
Raça	16,98	1	16,98	24,36	0,000
Erro	868,67	1246	0,70		
Total	886,12	1248			

12.19 A Tabela 12.15 da página 431 forneceu a equação de previsão $\hat{y} = 4{,}58 - 0{,}71p_1 - 0{,}54p_2 - 0{,}08s$ relacionando ideologia política à identificação partidária e ao sexo. Encontre as médias estimadas para as seis células e mostre que elas satisfazem a falta de interação.

12.20 Uma análise de regressão dos salários[2] do corpo docente de uma faculdade continha vários previsores, incluindo uma variável auxiliar para gênero (homem = 1) e uma variável auxiliar para raça (não branco = 1). Para a renda anual mensurada em milhares de dólares, os coeficientes estimados foram 0,76 para gênero e 0,62 para raça. Em uma configuração particular dos outros previsores, o salário médio estimado para mulheres brancas foi de 30,2 mil dólares. Estime as médias para os outros três grupos.

12.21 Quando usamos a PSG de 2004 e regredimos y = número de horas por dia assistindo à televisão em s = sexo (1 = masculino, 0 = feminino) e afiliação religiosa ($r_1 = 1$ para Protestante, $r_2 = 1$ para Católico, $r_3 = 1$ para Judeu, $r_1 = r_2 = r_3 =$ 0 para nenhuma ou outra), obtemos $\hat{y} = 2{,}4 + 0{,}2s + 0{,}5r_1 + 0{,}8r_2 - 0{,}1r_3$.

(a) Interprete o coeficiente de r_1.

(b) Determine um modelo correspondente para a população e indique quais parâmetros devem ser iguais a zero para y ser independente da afiliação religiosa, para cada sexo.

12.22 Nos Estados Unidos, o salário médio por hora da população atual de homens é de aproximadamente $22 para funcionários administrativos, $11 para prestadores de serviços e $14 para operários. Para as mulheres, as médias são $15 para funcionárias administrativas, $8 para prestadoras de serviços e $10 para operárias.

(a) Identifique a variável resposta e os dois fatores.

(b) Mostre estas médias em uma ANOVA de dois fatores.

(c) Compare as diferenças entre homens e mulheres para (i) funcionários administrativos (ii) para prestadores de serviços. Explique por que existe interação e descreva-a.

12.23 A U.S. Census Bureau (Agência do Censo dos Estados Unidos) recentemente declarou que a renda mediana da população era de $29661 para mulheres brancas, $25736 para mulheres negras, $40350 para homens brancos, $30886 para homens negros.

(a) Identifique a variável resposta e os dois fatores e mostre estas medianas em uma classificação de dois fatores.

(b) Explique por que existe interação em termos da mediana.

(c) Mostre quatro rendas medianas da população que iriam satisfazer H_0: sem interação.

12.24 A Tabela 12.36 resume as respostas sobre a ideologia política da PSG de 2004 por religião e sexo. O valor-p é 0,03 para testar H_0: sem interação. Explique o que isto significa no contexto desse exemplo e indique um lugar na tabela que possa ser responsável pelo valor-p pequeno.

12.25 A Tabela 12.37 mostra os resultados de uma ANOVA em y = índice de depressão e os previsores gênero e estado civil

Tabela 12.36

Religião		Ideologia política Média	Desvio padrão		Ideologia política Média	Desvio padrão
Protestante	Mulher	4,23	1,34	Homem	4,44	1,41
Católica	Mulher	4,10	1,33	Homem	4,22	1,42
Judia	Mulher	2,59	1,22	Homem	3,82	1,68
Nenhuma	Mulher	3,34	1,34	Homem	3,58	1,42

(casado, nunca casado, divorciado). Determine o tamanho da amostra e preencha os espaços em branco na tabela da ANOVA. Interprete os resultados.

Tabela 12.37

Fonte	Soma dos quadrados	*gl*	Média dos quadrados	*F*	Sig
Gênero	100	—	—	—	—
Estado civil	200	—	—	—	—
Interação	100	—	—	—	—
Erro	—	—	—		
Total	4000	205			

12.26 Use um *software* com a Tabela 12.10 da página 425.

(a) Ajuste o modelo sem interação e verifique os resultados obtidos anteriormente.

(b) Ajuste o modelo de interação. Mostre como a diferença entre SQE para este modelo e SQE para o modelo sem interação se relaciona à soma dos quadrados do numerador para testar H_0: sem interação.

(c) Usando a equação de previsão para o modelo com interação, encontre seis médias estimadas das células e compare-as às médias amostrais. (*Nota*: o modelo usa seis parâmetros para resumir seis médias, assim ele tem um ajuste perfeito.)

12.27 A equação de previsão $\hat{y} = 16 + 2s + 3r + 8(s \times r)$ relaciona s = sexo ($s = 1$ para homens, $s = 0$ para mulheres) e r = raça ($r = 1$ para brancos, $r = 0$ para negros). Encontrando as quatro médias previstas para esta equação, mostre que o coeficiente 8 do termo de interação é a quantia pela qual a média para um dos quatro grupos deve aumentar ou diminuir para que a interação desapareça.

12.28 Para a PSG de 2000, a Tabela 12.38 mostra médias amostrais da ideologia política classificada por gênero e por raça. Para H_0: sem interação, um *software* informa $F = 21{,}7$, $gl_1 = 1$, $gl_2 = 2508$ e o valor-$p = 0{,}001$.

Tabela 12.38

Gênero	Raça: Negra	Raça: Branca
Homem	4,06 (n = 256)	4,04 (n = 1144)
Mulher	3,74 (n = 139)	4,25 (n = 973)

(a) Suponha que, em vez da ANOVA de dois fatores, você executou uma ANOVA de um fator com o gênero como previsor e uma ANOVA de um fator separada com a raça como previsor. Suponha que a ANOVA para gênero não mostra um efeito significativo. Explique como isto poderia acontecer, embora a ANOVA de dois fatores tenha mostrado um efeito do gênero para cada raça. (*Dica:* as médias amostrais gerais para homens e mulheres serão mais similares do que elas são para cada raça?)

(b) Resuma o que você iria aprender sobre o efeito do gênero de uma ANOVA que você não iria aprender de uma ANOVA de um fator?

12.29 Considere a Tabela 12.17 da página 435 sobre as três influências em crianças.

(a) Usando um *software*, execute a análise de medidas repetidas da Seção 12.6.

(b) Suponha que você fez um escore das categorias de influência (−3, −2, 0, 2, 3). O que isto iria presumir sobre as categorias da resposta? Repita a análise usando estes escores. As conclusões são suscetíveis à escolha dos escores?

12.30 Recentemente, uma PSG perguntou aos respondentes: "Comparado com dez anos atrás, você diria que as crianças norte-americanas hoje estão (1) em uma situação bem melhor, (2) em uma situação melhor, (3) no mesmo, (4) pior ou (5) em uma situação bem pior?". A Tabela 12.39 mostra as respostas para dez dos sujeitos, em três itens: qualidade da sua educação, segurança dos bairros que residem e ter assistência médica quando necessitam.

(a) Para cada um dos seguintes indique se é um efeito fixo, um efeito aleatório ou uma variável resposta: (i) opinião, (ii) item, (iii) sujeito.

(b) Teste a hipótese de que as médias populacionais são iguais. Informe e valor-p e interprete.

Tabela 12.39

Sujeito	*Status* das crianças: Educação	Bairro	Assistência médica
1	4	4	3
2	2	4	2
3	3	3	4
4	1	2	1
5	3	4	3
6	2	5	4
7	1	4	2
8	3	3	3
9	4	5	3
10	2	4	2

12.31 Considere o exercício anterior. Os cinco primeiros respondentes eram mulheres e os cinco últimos homens. Analise estes dados usando tanto o gênero quanto o item como fatores.

12.32 A PSG pediu aos respondentes para avaliar vários grupos usando uma "escala de sensibilidade" onde 0 (mais desfavorável) a 100 (mais favorável). Planejamos estudar como a média se comporta para avaliar liberais e conservadores, nas avaliações de 2006 e de 1986. Explique por que uma ANOVA de dois fatores usando tempo (1986, 2006) e grupo avaliado (Liberal, Conservador) como fatores iria requerer métodos para medidas repetidas. Identifique os fatores dentre sujeitos e os entre sujeitos.

12.33 A Tabela 12.40 mostra resultados usando do SAS para analisar os dados sobre anorexia da Tabela 12.21.

(a) Explique como usar a informação desta tabela para testar H_0: sem interação entre tratamento e tempo.

(b) Explique por que não faz sentido conduzir testes para os efeitos principais.

12.34 Usando *software*, execute uma ANOVA de medidas repetidas com os dados da anorexia da Tabela 12.21 (página 439), disponíveis no *site* do livro.

Conceitos e aplicações

12.35 Considere o arquivo de dados *Student Survey* (Exercício 1.11 da página 25), com a variável resposta: número de horas semanais empregadas em esportes e outros exercícios físicos. Usando um *software* execute uma análise de variância e uma estimação de acompanhamento e prepare um relatório interpretando e resumindo suas análises usando o:

(a) Gênero como previsor.

(b) Gênero e se é vegetariano como previsores.

12.36 Para y = número de vezes que usou o transporte público na última semana e x = número de carros na família (que assume valores 0, 1 ou 2 para a amostra dada), explique a diferença entre reali-

☑ **Tabela 12.32**

	Muito feliz	Moderadamente feliz	Infeliz		
Média	10,4	7,4	8,3		
Desvio padrão	17,8	13,6	15,6		
Tamanho da amostra	276	468	87		
Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Grupo	1626,8	2	813,4	3,47	0,032
Erro	193900,9	828	234,2		
Total	195527,7	830			

☑ **Tabela 12.33**

Fonte	GL	Soma dos quadrados	Média dos quadrados	Valor-F	Pr > F
Modelo	2	72,000	36,000	0,79	0,4813
Erro	9	408,000	45,333		
Total	11	480,000			

Testes T de Bonferroni para a variável: TEMPO

Alfa = 0,05 gl = 9 EQM = 45,33333

Valor crítico de T = 2,93

Diferença mínima significativa = 13,965

As médias com as mesmas letras não são significativamente diferentes.

Agrupamento Bon	Média	N	Grupo
A	18,000	4	3
A			
A	15,000	4	2
A			
A	12,000	4	1

12.15 Considere o exercício anterior.

(a) Estabeleça variáveis auxiliares para um modelo de regressão tal que um teste F para os parâmetros da regressão seja equivalente ao teste F da ANOVA. Expresse H_0 tanto em termos das médias populacionais quanto dos parâmetros da regressão.

(b) A equação de previsão obtida no ajuste da equação de regressão de (b) é $\hat{y} = 18 - 6z_1 - 3z_2$. Mostre como as estimativas do parâmetro se relacionam às médias amostrais da Tabela 12.33.

12.16 O Exercício 12.8 mostrou uma ANOVA para um experimento comparando três parachoques, com a média amostral dos danos de 2 para o parachoque A, 3 para o B e 13 para o C (em centenas de dólares).

(a) Estabeleça um modelo de regressão com variáveis auxiliares para a ANOVA.

(b) Mostre a correspondência entre a hipótese nula para as médias da ANOVA e a hipótese nula para os parâmetros da regressão.

(c) Informe a equação de previsão que você obteria para o modelo em (a). (Se não estiver seguro, use um *software* para encontrá-la.)

12.17 Quando usamos a PSG de 2004 para avaliar como o número médio de horas por dia assistindo à televisão depende do sexo e raça, obtemos os resultados exibidos na Tabela 12.34. As médias amostrais foram de 2,71 para mulheres brancas, 2,79 para homens brancos, 4,13 para mulheres negras e 4,16 para homens negros. Explique como estes resultados são compatíveis com os resultados dos testes exibidos.

☑ **Tabela 12.34**

Fonte	SQ	gl	MQ	F	Sig
Sexo	2,22	1	2,22	0,35	0,555
Raça	489,65	1	489,65	76,62	0,000
Erro	11094,16	1737	6,39		
Total	11583,81	1739			

12.18 Uma PSG recente perguntou: "Qual é o número ideal de filhos para uma família?". A Tabela 12.35 mostrou os resultados de usar a PSG para avaliar os efeitos de gênero e raça.

(a) Explique como interpretar os resultados dos testes F.

(b) Considere $s = 1$ para mulheres e 0 para homens e considere $r = 1$ para negros e 0 para brancos. O modelo sem interação tem $\hat{y} = 2{,}42 + 0{,}04s + 0{,}37r$. Encontre a média estimada para cada combinação de gênero e raça. Explique como estas médias satisfazem "sem interação".

☑ **Tabela 12.35**

Fonte	SQ	gl	MQ	F	Valor-p
Sexo	0,25	1	0,25	0,36	0,550
Raça	16,98	1	16,98	24,36	0,000
Erro	868,67	1246	0,70		
Total	886,12	1248			

12.19 A Tabela 12.15 da página 431 forneceu a equação de previsão $\hat{y} = 4{,}58 - 0{,}71p_1 - 0{,}54p_2 - 0{,}08s$ relacionando ideologia política à identificação partidária e ao sexo. Encontre as médias estimadas para as seis células e mostre que elas satisfazem a falta de interação.

12.20 Uma análise de regressão dos salários[2] do corpo docente de uma faculdade continha vários previsores, incluindo uma variável auxiliar para gênero (homem = 1) e uma variável auxiliar para raça (não branco = 1). Para a renda anual mensurada em milhares de dólares, os coeficientes estimados foram 0,76 para gênero e 0,62 para raça. Em uma configuração particular dos outros previsores, o salário médio estimado para mulheres brancas foi de 30,2 mil dólares. Estime as médias para os outros três grupos.

12.21 Quando usamos a PSG de 2004 e regredimos y = número de horas por dia assistindo à televisão em s = sexo (1 = masculino, 0 = feminino) e afiliação religiosa ($r_1 = 1$ para Protestante, $r_2 = 1$ para Católico, $r_3 = 1$ para Judeu, $r_1 = r_2 = r_3 =$ 0 para nenhuma ou outra), obtemos $\hat{y} = 2{,}4 + 0{,}2s + 0{,}5r_1 + 0{,}8r_2 - 0{,}1r_3$.

(a) Interprete o coeficiente de r_1.

(b) Determine um modelo correspondente para a população e indique quais parâmetros devem ser iguais a zero para y ser independente da afiliação religiosa, para cada sexo.

12.22 Nos Estados Unidos, o salário médio por hora da população atual de homens é de aproximadamente $22 para funcionários administrativos, $11 para prestadores de serviços e $14 para operários. Para as mulheres, as médias são $15 para funcionárias administrativas, $8 para prestadoras de serviços e $10 para operárias.

(a) Identifique a variável resposta e os dois fatores.

(b) Mostre estas médias em uma ANOVA de dois fatores.

(c) Compare as diferenças entre homens e mulheres para (i) funcionários administrativos (ii) para prestadores de serviços. Explique por que existe interação e descreva-a.

12.23 A U.S. Census Bureau (Agência do Censo dos Estados Unidos) recentemente declarou que a renda mediana da população era de $29661 para mulheres brancas, $25736 para mulheres negras, $40350 para homens brancos, $30886 para homens negros.

(a) Identifique a variável resposta e os dois fatores e mostre estas medianas em uma classificação de dois fatores.

(b) Explique por que existe interação em termos da mediana.

(c) Mostre quatro rendas medianas da população que iriam satisfazer H_0: sem interação.

12.24 A Tabela 12.36 resume as respostas sobre a ideologia política da PSG de 2004 por religião e sexo. O valor-p é 0,03 para testar H_0: sem interação. Explique o que isto significa no contexto desse exemplo e indique um lugar na tabela que possa ser responsável pelo valor-p pequeno.

12.25 A Tabela 12.37 mostra os resultados de uma ANOVA em y = índice de depressão e os previsores gênero e estado civil

☑ **Tabela 12.36**

Religião		Ideologia política			Ideologia política	
		Média	Desvio padrão		Média	Desvio padrão
Protestante	Mulher	4,23	1,34	Homem	4,44	1,41
Católica	Mulher	4,10	1,33	Homem	4,22	1,42
Judia	Mulher	2,59	1,22	Homem	3,82	1,68
Nenhuma	Mulher	3,34	1,34	Homem	3,58	1,42

(casado, nunca casado, divorciado). Determine o tamanho da amostra e preencha os espaços em branco na tabela da ANOVA. Interprete os resultados.

☑ **Tabela 12.37**

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig
Gênero	100	—	—	—	—
Estado civil	200	—	—	—	—
Interação	100	—	—	—	—
Erro	—	—	—		
Total	4000	205			

12.26 Use um *software* com a Tabela 12.10 da página 425.

(a) Ajuste o modelo sem interação e verifique os resultados obtidos anteriormente.

(b) Ajuste o modelo de interação. Mostre como a diferença entre SQE para este modelo e SQE para o modelo sem interação se relaciona à soma dos quadrados do numerador para testar H_0: sem interação.

(c) Usando a equação de previsão para o modelo com interação, encontre seis médias estimadas das células e compare-as às médias amostrais. (*Nota*: o modelo usa seis parâmetros para resumir seis médias, assim ele tem um ajuste perfeito.)

12.27 A equação de previsão $\hat{y} = 16 + 2s + 3r + 8(s \times r)$ relaciona s = sexo ($s = 1$ para homens, $s = 0$ para mulheres) e r = raça ($r = 1$ para brancos, $r = 0$ para negros). Encontrando as quatro médias previstas para esta equação, mostre que o coeficiente 8 do termo de interação é a quantia pela qual a média para um dos quatro grupos deve aumentar ou diminuir para que a interação desapareça.

12.28 Para a PSG de 2000, a Tabela 12.38 mostra médias amostrais da ideologia política classificada por gênero e por raça. Para H_0: sem interação, um *software* informa $F = 21{,}7$, $gl_1 = 1$, $gl_2 = 2508$ e o valor-p = 0,001.

☑ **Tabela 12.38**

Gênero	Raça: Negra	Raça: Branca
Homem	4,06 (n = 256)	4,04 (n = 1144)
Mulher	3,74 (n = 139)	4,25 (n = 973)

(a) Suponha que, em vez da ANOVA de dois fatores, você executou uma ANOVA de um fator com o gênero como previsor e uma ANOVA de um fator separada com a raça como previsor. Suponha que a ANOVA para gênero não mostra um efeito significativo. Explique como isto poderia acontecer, embora a ANOVA de dois fatores tenha mostrado um efeito do gênero para cada raça. (*Dica:* as médias amostrais gerais para homens e mulheres serão mais similares do que elas são para cada raça?)

(b) Resuma o que você iria aprender sobre o efeito do gênero de uma ANOVA que você não iria aprender de uma ANOVA de um fator?

12.29 Considere a Tabela 12.17 da página 435 sobre as três influências em crianças.

(a) Usando um *software*, execute a análise de medidas repetidas da Seção 12.6.

(b) Suponha que você fez um escore das categorias de influência (−3, −2, 0, 2, 3). O que isto iria presumir sobre as categorias da resposta? Repita a análise usando estes escores. As conclusões são suscetíveis à escolha dos escores?

12.30 Recentemente, uma PSG perguntou aos respondentes: "Comparado com dez anos atrás, você diria que as crianças norte-americanas hoje estão (1) em uma situação bem melhor, (2) em uma situação melhor, (3) no mesmo, (4) pior ou (5) em uma situação bem pior?". A Tabela 12.39 mostra as respostas para dez dos sujeitos, em três itens: qualidade da sua educação, segurança dos bairros que residem e ter assistência médica quando necessitam.

(a) Para cada um dos seguintes indique se é um efeito fixo, um efeito aleatório ou uma variável resposta: (i) opinião, (ii) item, (iii) sujeito.

(b) Teste a hipótese de que as médias populacionais são iguais. Informe e valor-p e interprete.

☑ **Tabela 12.39**

	Status das crianças		
Sujeito	Educação	Bairro	Assistência médica
1	4	4	3
2	2	4	2
3	3	3	4
4	1	2	1
5	3	4	3
6	2	5	4
7	1	4	2
8	3	3	3
9	4	5	3
10	2	4	2

12.31 Considere o exercício anterior. Os cinco primeiros respondentes eram mulheres e os cinco últimos homens. Analise estes dados usando tanto o gênero quanto o item como fatores.

12.32 A PSG pediu aos respondentes para avaliar vários grupos usando uma "escala de sensibilidade" onde 0 (mais desfavorável) a 100 (mais favorável). Planejamos estudar como a média se comporta para avaliar liberais e conservadores, nas avaliações de 2006 e de 1986. Explique por que uma ANOVA de dois fatores usando tempo (1986, 2006) e grupo avaliado (Liberal, Conservador) como fatores iria requerer métodos para medidas repetidas. Identifique os fatores dentre sujeitos e os entre sujeitos.

12.33 A Tabela 12.40 mostra resultados usando SAS para analisar os dados sobre anorexia da Tabela 12.21.

(a) Explique como usar a informação desta tabela para testar H_0: sem interação entre tratamento e tempo.

(b) Explique por que não faz sentido conduzir testes para os efeitos principais.

12.34 Usando *software*, execute uma ANOVA de medidas repetidas com os dados da anorexia da Tabela 12.21 (página 439), disponíveis no *site* do livro.

Conceitos e aplicações

12.35 Considere o arquivo de dados *Student Survey* (Exercício 1.11 da página 25), com a variável resposta: número de horas semanais empregadas em esportes e outros exercícios físicos. Usando um *software* execute uma análise de variância e uma estimação de acompanhamento e prepare um relatório interpretando e resumindo suas análises usando o:

(a) Gênero como previsor.

(b) Gênero e se é vegetariano como previsores.

12.36 Para y = número de vezes que usou o transporte público na última semana e x = número de carros na família (que assume valores 0, 1 ou 2 para a amostra dada), explique a diferença entre reali-

Tabela 12.40

Análise de Variância de Medidas Repetidas

Testes de Hipóteses para efeitos entre Sujeitos

Fonte	GL	SQ ANOVA	Média dos quadrados	Valor-F	Pr > F
TRATAMENTO	2	644,23	322,12	6,20	0,0033
Erro	69	3584,03	51,94		

Testes de Hipóteses de uma variável para os efeitos dentre sujeitos

Fonte: TEMPO

GL	ANOVA SQ	Média dos quadrados	Valor-F	Pr > F
1	275,0069444	275,0069444	9,70	0,0027

Fonte: TEMPO*TRATAMENTO

GL	ANOVA SQ	Média dos quadrados	Valor-F	Pr > F
1	307,3218334	153,6609167	5,42	0,0065

Fonte: ERRO (TEMPO)

GL	ANOVA SQ	Média dos quadrados
69	1955,3712221	28,3387134

zar um teste de independência usando o teste F da ANOVA para comparar as três médias e usar o teste t da regressão para o coeficiente do número de carros em um modelo usual de regressão com um efeito linear sobre o número de carros. Dê um exemplo de três médias populacionais para as quais o teste de regressão seria menos apropriado do que o teste da ANOVA. (*Dica*: o que o teste do modelo de regressão linear assume que o teste t da ANOVA não assume?)

12.37 Vá ao *site* da PSG sda.berkeley.edu/GSS e clique em [*Comparison of Means*] (Comparação de Médias) no menu *Analysis* (Análise) e execute uma ANOVA para o ano mais recente disponível para comparar a ideologia política média ("POLVIEWS") em relação (a) às categorias da classe social (baixa, trabalhadora, média, alta) ("CLASS") e (b) às nove regiões do país ("REGION"). Informe os resultados e interprete.

12.38 Uma PSG recente perguntou: "Com que frequência você participa de cerimônias religiosas?". A Tabela 12.41 mostra os resultados da ANOVA para comparar os três níveis desta variável (pelo menos 2 – 3 vezes por mês para o grupo "alta", várias vezes ao ano a aproximadamente uma vez por mês para o grupo "média", e no máximo uma vez por ano para o grupo "baixa") em relação ao número de bons amigos. Use os resultados relatados para fazer inferências (um teste de significância e uma estimação). Resuma as suas análises e interpretações.

12.39 Um experimento usou quatro grupos selecionados aleatoriamente de cinco indivíduos cada. A média amostral geral foi de 60.

(a) Como ficariam os dados se uma ANOVA de um fator para comparar as médias teve uma estatística-teste $F = 0$?

(b) Como ficariam os dados se $F = \infty$?

Tabela 12.41

			Com que frequência participa de serviços religiosos?		
			Alta	Média	Baixa
Número médio de bons amigos			12,1	6,4	6,2
Tamanho da amostra			337	159	330
Fonte	Média dos quadrados	gl	Soma dos quadrados	F	Sig
Grupo	6748,2	2	3374,1	14,2	0,000
Erro	196136,9	823	238,3		
Total	202885,1	825			

12.40 Um estudo (Baltimore Longitudinal Study on Aging)[3] relacionando fumo e personalidade usou uma amostra de 1638 adultos. Os sujeitos formaram três grupos de acordo com o *status* de fumante (nunca, ex-fumante, ativo). Cada sujeito completou um questionário sobre a personalidade, o qual forneceu escores em várias escalas da personalidade delineadas para ter médias gerais de aproximadamente 50 e desvios padrão de aproximadamente 10. A Tabela 12.42 mostra alguns resultados para três características pessoais, dando as médias com os desvios padrão entre parênteses. O estudo utilizou 35 escalas de personalidade e informou um teste F comparando os três grupos de fumantes para cada escala. Os pesquisadores mencionaram terem feito uma correção de Bonferroni para os 35 testes F. Se a probabilidade do erro do Tipo I nominal era de 0,05 para os 35 testes, quão pequeno deve ser o valor-p para um dado teste ser significativo? Interprete os resultados na tabela.

12.41 Um estudo[4] comparou a memória verbal de homens e mulheres para palavras abstratas e para palavras concretas. O estudo encontrou um efeito principal de gênero em favor das mulheres. Ele também informou: "Não havia a interação sexo × tipo de palavra ($F = 0{,}408$, $p = 0{,}525$), indicando que as mulheres estavam igualmente favorecidas nos dois tipos de palavras". Como você iria explicar o que esta sentença significa para alguém que nunca estudou estatística?

12.42 **(a)** Explique cuidadosamente a diferença entre a probabilidade do erro do Tipo I de 0,05 para uma única comparação de duas médias e uma taxa de erro de comparação múltipla de 0,05 para comparar todos os pares de médias.

(b) Nas comparações múltiplas seguidas de uma ANOVA de um fator com tamanhos de amostras iguais, a margem de erro com um intervalo de 95% de confiança para comparar cada par de médias é igual a 10. Apresente três médias amostrais ilustrando que é possível que o grupo A não seja significativamente diferente de grupo B e que o grupo B não seja significativamente diferente do grupo C, porém o grupo A seja significativamente diferente do grupo C.

12.43 A Tabela 12.43 resume as respostas sobre a ideologia política da PSG de 2004 por raça e gênero (lembre que 1 = extremamente liberal, 4 = moderado, 7 = extremamente conservador). A Tabela 12.44 mostra os resultados das ANOVAs de um fator e dois fatores comparando os quatro grupos. Escreva um parágrafo explicando o que você aprendeu da

Tabela 12.42

	Nunca fumou	Ex-fumante	Fumante ativo	F
	(n = 828)	(n = 694)	(n = 116)	
Neurótico	46,7 (9,6)	48,5 (9,2)	51,9 (9,9)	17,77
Extrovertido	50,4 (10,3)	50,2 (10,0)	50,9 (9,4)	0,24
Consciente	51,8 (10,1)	48,9 (9,7)	45,6 (10,3)	29,42

ANOVA de dois fatores que você não pode aprender da ANOVA de um fator.

☑ **Tabela 12.43**

			Ideologia política	
Raça	**Gênero**	***n***	**Média**	**Desvio padrão**
Branca	Feminino	553	4,15	1,40
	Masculino	501	4,42	1,42
Negra	Feminino	100	3,95	1,47
	Masculino	54	3,93	1,30

☑ **Tabela 12.44**

ANOVA de um fator

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	P
Grupos	34,46	3	11,49	5,78	0,001
Erro	2393,83	1204	1,99		
Total	2428,29	1207			

ANOVA de dois fatores

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	P
Raça	12,74	1	12,74	6,41	0,011
Sexo	16,42	1	16,42	8,26	0,004
Interação	2,66	1	2,66	1,34	0,248
Erro	2393,83	1204	1,99		
Total	2428,29	1207			

12.44 A Tabela 7.24 da página 248 resume um estudo que relatou o número médio de encontros nos últimos três meses. Para homens, a média foi de 9,7 para os mais atraentes e 9,9 para os menos atraentes. Para mulheres, a média foi de 17,8 para as mais atraentes e 10,6 para as menos atraentes. Identifique a variável resposta e os fatores e indique se estes dados parecem mostrar interação. Explique.

12.45 Construa um exemplo numérico de médias para uma classificação de dois fatores sob as seguintes condições:

(a) Os efeitos principais estão presentes somente para a variável na linha.

(b) Os efeitos principais estão presentes para cada variável, sem interação.

(c) Os efeitos de interação estão presentes.

(d) Não existem efeitos presentes.

12.46 A hipótese nula da igualdade das médias para um fator é rejeitada em uma ANOVA de dois fatores. Isto implica que a hipótese será rejeitada em um teste *F* de uma ANOVA de um fator se os dados forem combinados com os níveis da segunda variável? Explique.

12.47 Para uma dupla classificação de médias pelos fatores A e B, em cada nível de B as médias são iguais aos níveis de A. Isto implica que as médias gerais são iguais nos vários níveis de A, ignorando B? Explique as implicações em termos de como os resultados podem diferir entre a ANOVA de dois fatores e a ANOVA de um fator.

12.48 As 25 professoras no departamento de humanas de uma faculdade têm um salário médio de $66000, enquanto as cinco do departamento de ciências têm um salário médio de $80000. Por outro lado, os 20 homens do departamento de humanas têm um salário médio de $65000 e os 30 homens do departamento de ciências têm um salário médio de $79000.

(a) Construa uma tabela para as médias amostrais da renda para uma classificação cruzada 2 × 2 de gênero por departamento. Encontre as médias gerais para homens e mulheres. Interprete.

(b) Discuta como os resultados de uma comparação de um fator das rendas médias por gênero iriam diferir dos resultados de uma comparação de dois fatores das rendas médias por gênero, controlando pelo departamento. (*Nota*: essa inversão no qual o gênero tem o salário médio mais alto, de acordo com quem controla o departamento, ilustra o *paradoxo de Simpson*. Veja o Exercício 10.14 na página 354.)

12.49 Uma amostra aleatória de 26 universitárias da Universidade da Flórida foi pesquisada sobre sua atitude em relação ao aborto. Cada uma recebeu um escore sobre a atitude em relação ao aborto de acordo com quantas possíveis razões, de uma lista de oito, ela iria aceitar como uma razão legítima para uma mulher praticar um aborto. A Tabela 12.45 exibe os escores, classificados por religião e frequência à igreja. Usando um *software*, analise os dados e informe as suas descobertas.

12.50 Verdadeiro ou falso? Suponha que, para sujeitos com menos de 50 anos, exista pouca diferença na média anual das despesas médicas para fumantes e não fumantes, mas para sujeitos acima dos 50 anos exista uma grande diferença. Portanto, não existe uma interação entre o *status* de fumante e a idade nos seus efeitos sobre as despesas médicas anuais.

Selecione a(s) resposta(s) correta(s) nos Exercícios 12.51 a 12.54. (Mais de uma resposta pode estar correta.)

12.51 A análise de variância e a regressão são similares no sentido de que:

(a) Ambas assumem uma variável resposta quantitativa.

(b) Ambas têm testes *F* para testar que a variável resposta é estatisticamente independente da(s) variável(eis) explicativa(s).

(c) Para propósitos inferenciais, ambas assumem que a variável resposta *y* é normalmente distribuída com o mesmo desvio padrão em todas as combinações dos níveis da(s) variável(eis) explicativa(s).

(d) Ambas fornecem formas de particionar a variação em *y* em componentes "explicados" e "não explicados".

12.52 Uma ANOVA de um fator fornece relativamente mais evidência de que H_0: $\mu_1 = ... = \mu_g$ é falsa:

(a) quanto menor a variação entre grupos e maior a variação dentro dos grupos.

(b) quanto menor a variação entre grupos e menor a variação dentro dos grupos.

(c) quanto maior a variação entre grupos e menor a variação dentro dos grupos.

(d) quanto maior a variação entre grupos e maior a variação dentro dos grupos.

12.53 Para quatro médias, um método de comparação múltipla fornece intervalos de 95% de confiança para as diferenças entre os seis pares. Então:

(a) para cada intervalo de confiança, existe uma chance de 0,95 de que ele contenha a diferença populacional.

(b) *P*(de que todos os seis intervalos de confiança estejam corretos) = 0,70.

(c) *P*(de que todos os seis intervalos de confiança estejam corretos) = 0,95.

(d) *P*(de que todos os seis intervalos de confiança estejam corretos) = $(0,95)^6$.

(e) *P*(de que pelo menos um intervalo não contenha a diferença verdadeira) = 0,05.

(f) os intervalos de confiança serão maiores do que os intervalos separados de 95% de confiança para cada diferença.

12.54 Os termos de interação são necessários em um modelo ANOVA de dois fatores quando:

(a) cada par de variáveis está associado.

(b) ambas as variáveis explicativas têm efeitos significativos no modelo sem os termos de interação.

(c) a diferença nas médias entre duas categorias de uma variável explicativa varia grandemente entre as categorias da outra variável explicativa.

(d) a média dos quadrados para interação é enorme comparada ao erro quadrático médio.

***12.55** Você conhece a média amostral, o desvio padrão e o tamanho da amostra para cada um dos três grupos. Você pode conduzir um teste *F* da ANOVA comparando as médias da população ou você necessitaria de mais informação?

***12.56** Você forma um intervalo de 95% de confiança em cinco situações diferentes com amostras independentes.

(a) Encontre a probabilidade que (i) *todos* os cinco intervalos contenham os parâmetros que foram projetados para estimar, (ii) pelo menos um intervalo esteja errado. (*Dica*: use a distribuição binomial.)

(b) Se você usa um nível de confiança de 0,9898 para cada intervalo, a probabilidade de que todos os cinco intervalos contenham os parâmetros é igual a 0,95. Explique por quê. (*Dica:* o que é $(0{,}9898)^5$?) Compare 0,9898 ao coeficiente de confiança para cada intervalo no método Bonferroni.

*12.57 Este exercício motiva a fórmula para a estimativa da variância entre grupos da ANOVA de um fator. Suponha que os tamanhos da amostra sejam todos iguais a n e as médias da população sejam iguais a μ. A distribuição amostral para cada $\bar{y}_i$, então, tem média μ e variância σ^2/n. A média amostral dos valores de $\bar{y}_i$ é $\bar{y}$.

(a) Tratando $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_g$ como observações g tendo média amostral $\bar{y}$, explique por que

$$\sum(\bar{y}_i - \bar{y})^2/(g - 1)$$

estima a variância σ^2/n da distribuição amostral dos valores $\bar{y}_i$.

(b) Usando (a), explique por que $\sum n(\bar{y}_i - \bar{y})^2/(g - 1)$ estima σ^2. Para o caso de tamanhos de amostras diferentes, substituindo n por n_i geramos a estimativa entre grupos.

☑ NOTAS

1 HOGE, C. et al. *New England J. Medic.*, v. 351, p. 13-21, 2004.
2 BELLAS, M. *American Sociological Review*, v. 59, p. 807, 1994.
3 TERRACCIANO, A., COSTA, P. *Addiction*, v. 99, p. 472-81, 2004.
4 KIMURA, D., CLARKE, P. *Psych. Reports*, v. 91, p. 1137-42, 2002.

☑ 13

COMBINANDO REGRESSÃO E ANOVA: PREVISORES QUANTITATIVOS E CATEGÓRICOS

O Capítulo 11 introduziu a regressão múltipla para analisar o relacionamento entre uma variável resposta quantitativa e variáveis explicativas *quantitativas*. O Capítulo 12 mostrou que a regressão múltipla pode também lidar com variáveis explicativas *categóricas*, como na análise de variância com variáveis auxiliares. Não surpreendentemente, a regressão múltipla pode também lidar simultaneamente com variáveis explicativas categóricas e quantitativas. O modelo combina elementos da análise de regressão comum, para a qual os previsores são categóricos.

Controlando uma covariável

A ANOVA de um fator compara a média da variável resposta para vários grupos. A ANOVA de dois fatores compara as médias enquanto controla outra variável categórica. Em muitas aplicações, é útil comparar as médias enquanto controlamos uma variável quantitativa. Por exemplo, na comparação da renda média de homens e mulheres, podemos controlar os diferentes níveis de experiência no trabalho entre homens e mulheres. A variável controle quantitativa é chamada de **covariável**. O uso da regressão para esse tipo de comparação é geralmente chamado de **análise de covariância**. Essa é uma das muitas contribuições estatísticas de Ronald A. Fisher, um brilhante estatístico britânico.

Pelo fato de que os efeitos podem mudar após o controle de uma variável, os resultados da análise de covariância podem diferir dos resultados da análise de variância. Por exemplo, a experiência no emprego é normalmente correlacionada com a renda. Se os homens tendem a ter níveis mais altos de experiência do que as mulheres em determinado emprego, os resultados de uma comparação da renda média entre homens e mulheres dependerá do controle da experiência.

Para simplificar, este capítulo ilustra conceitos usando um único previsor categórico e um único previsor quantitativo, mas as ideias básicas se estendem a múltiplos previsores. A primeira seção mostra representações gráficas dos efeitos potenciais. As Seções 13.2 e 13.3 mostram que os modelos de regressão com variáveis auxiliares fornecem a base para a análise. As Seções 13.4 e 13.5 apresentam inferências para os modelos. A Seção 13.6 mostra como ajustar as médias amostrais de y para refletir seus valores previstos após controlar a covariável.

13.1 COMPARANDO MÉDIAS E COMPARANDO LINHAS DE REGRESSÃO

Neste capítulo representamos a variável explicativa quantitativa por x e a variável