APÊNDICE D

Resumo de álgebra matricial

Este apêndice resume os conceitos de álgebra matricial, incluindo a álgebra da probabilidade, necessários para o estudo de modelos de regressão linear múltipla com matrizes do Apêndice E. Nenhum destes materiais é usado no texto principal.

D.1 Definições básicas

Definição D.1 (Matriz). Uma **matriz** é uma formação retangular de números. Mais precisamente, uma matriz $m \times n$ tem m linhas e n colunas. O inteiro positivo m é chamado de *dimensão da linha* e n é chamado de *dimensão da coluna*.

Utilizamos letras maiúsculas em negrito para indicar matrizes. Podemos escrever uma matriz $m \times n$ de forma genérica como

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & & & & \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix},$$

onde a_{ij} representa o elemento na i^{a} linha e j^{a} *coluna*. Por exemplo, a_{25} representa o número na segunda linha e quinta coluna de **A**. Um exemplo específico de matriz 2×3 é

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \tag{D.1}$$

onde $a_{13} = 7$. A abreviação $\mathbf{A} = [a_{ij}]$ é usada com frequência para definir operações matriciais.

Definição D.2 (Matriz quadrada). Uma **matriz quadrada** tem o mesmo número de linhas e de colunas. A dimensão de uma matriz quadrada é seu número de linhas e colunas.

Definição D.3 (Vetores).

(i) Uma matriz $1 \times m$ é chamada de **vetor linha** (da dimensão m) e pode ser escrita como $\mathbf{x} \equiv (x_1, x_2, ..., x_m)$.

(ii) Uma matriz $n \times 1$ é chamada de **vetor coluna** e pode ser escrita como

$$\mathbf{y} \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Definição D.4 (Matriz diagonal). Uma matriz quadrada **A** é uma **matriz diagonal** quando todos os seus elementos fora da diagonal forem zero, isto é, $a_{ij} = 0$ para todo $i \neq j$. Podemos sempre escrever uma matriz diagonal como

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

Definição D.5 (Matriz identidade e matriz nula).

(i) A **matriz identidade** $n \times n$, sinalizada por **I**, ou às vezes $\mathbf{I}_n$ para enfatizar sua dimensão, é a matriz diagonal com unidade (um) em cada posição diagonal e zero no restante:

$$\mathbf{I} \equiv \mathbf{I}_n \equiv \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

(ii) A **matriz nula** $m \times n$, indicada por **0**, é a matriz $m \times n$ com zero em todas as entradas. Ela não precisa ser uma matriz quadrada.

D.2 Operações matriciais

D.2a Soma de matrizes

Duas matrizes **A** e **B**, cada uma com dimensão $m \times n$, podem ser somadas elemento a elemento: $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$. Mais precisamente

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & & & \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}.$$

Por exemplo,

$$\begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -4 \\ 4 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 3 \\ 0 & 7 & 3 \end{bmatrix}.$$

Matrizes com dimensões diferentes não podem ser somadas.

D.2b Multiplicação escalar

Dado qualquer número real γ (frequentemente chamado de escalar), a **multiplicação escalar** é definida como $\gamma\mathbf{A} \equiv [\gamma a_{ij}]$, ou

$$\gamma\mathbf{A} = \begin{bmatrix} \gamma a_{11} & \gamma a_{12} & \dots & \gamma a_{1n} \\ \gamma a_{21} & \gamma a_{22} & \dots & \gamma a_{2n} \\ \vdots & & & \\ \gamma a_{m1} & \gamma a_{m2} & \dots & \gamma a_{mn} \end{bmatrix}.$$

Por exemplo, se $\gamma = 2$ e $\mathbf{A}$ é a matriz da equação (D.1), então

$$\gamma\mathbf{A} = \begin{bmatrix} 4 & -2 & 14 \\ -8 & 10 & 0 \end{bmatrix}.$$

D.2c Multiplicação de matrizes

Para multiplicar a matriz **A** pela matriz **B** e formar o produto **AB**, a dimensão da *coluna* de **A** deve ser igual à dimensão da *linha* de **B**. Portanto, seja **A** uma matriz $m \times n$ e **B** uma matriz $n \times p$. Assim, a **multiplicação de matrizes** é definida como

$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right].$$

Em outras palavras, o $(i, j)^{\text{o}}$ elemento da nova matriz **AB** é obtido pela multiplicação de cada elemento da i^a linha de **A** pelo elemento correspondente da j^a coluna de **B** e pela soma desses n produtos. Um esquema pode ajudar a tornar esse processo mais claro:

$$\begin{array}{cccc} & \mathbf{A} & \mathbf{B} & \mathbf{AB} \\ i^a \text{ linha} \rightarrow & \begin{bmatrix} \\ a_{i1}a_{i2}a_{i3}\dots a_{in} \\ \\ \end{bmatrix} & \begin{bmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{nj} \end{bmatrix} = & \begin{bmatrix} \\ \displaystyle\sum_{k=1}^{n} a_{ik}b_{kj} \\ \\ \end{bmatrix}, \\ & & \uparrow & \uparrow \\ & & j^a \text{ coluna} & (i, j)^{\text{o}} \text{ elemento} \end{array}$$

onde, pela definição do operador de soma do Apêndice A,

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}.$$

Por exemplo,

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}.$$

Também podemos multiplicar uma matriz e um vetor. Se **A** é uma matriz $n \times m$ e **y** é um vetor $m \times 1$, **Ay** será um vetor $n \times 1$. Se **x** for um vetor $1 \times n$, **xA** será um vetor $1 \times m$.

Soma de matrizes, multiplicação escalar e multiplicação de matrizes podem ser combinadas de diversas formas, e essas operações satisfazem várias regras que são familiares nas operações numéricas básicas. Na lista de propriedades a seguir, **A**, **B** e **C** são matrizes com dimensões apropriadas para cada operação, e α e β são números reais. A maioria dessas propriedades é fácil de ilustrar a partir das definições.

Propriedades de operações matriciais. (1) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$; (2) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$; (3) $(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$; (4) $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B}$; (5) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$; (6) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$; (7) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$; (8) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$; (9) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$; (10) $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$; (11) $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$; (12) $\mathbf{A} - \mathbf{A} = \mathbf{0}$; (13) $\mathbf{A0} = \mathbf{0A} = \mathbf{0}$; e (14) $\mathbf{AB} \neq \mathbf{BA}$, mesmo quando ambos os produtos forem definidos.

A última propriedade merece comentários adicionais. Se **A** for $n \times m$ e **B** for $m \times p$, então **AB** será definida, mas **BA** só seria definida se $n = p$ (a dimensão da linha de **A** fosse igual à dimensão da coluna de **B**). Se **A** fosse $m \times n$ e **B** fosse $n \times m$, então **AB** e **BA** seriam ambas definidas, mas normalmente elas não são as mesmas; na verdade, elas terão dimensões diferentes, a não ser que **A** e **B** sejam duas matrizes quadradas. Mesmo quando **A** e **B** forem quadradas, $\mathbf{AB} \neq \mathbf{BA}$, exceto sob circunstâncias especiais.

D.2d Transposição

Definição D.6 (Transposição). Seja $\mathbf{A} = [a_{ij}]$ uma matriz $m \times n$. A **transposição** de **A**, indicada por $\mathbf{A}'$ (chamada de **A** *primária*), é a matriz $n \times m$ obtida pela troca de linhas e colunas de **A**. Podemos escrevê-la como $\mathbf{A}' \equiv [a_{ji}]$.

Por exemplo,

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \quad \mathbf{A}' = \begin{bmatrix} 2 & -4 \\ -1 & 5 \\ 7 & 0 \end{bmatrix}.$$

Propriedades da transposição. (1) $(\mathbf{A}')' = \mathbf{A}$; (2) $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$ para qualquer escalar α; (3) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$; (4) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, onde **A** é $m \times n$ e **B** é $n \times k$; (5) $\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$, onde **x** é um vetor $n \times 1$; e (6) se **A** for uma matriz $n \times k$ com *linhas* dadas pelos vetores $1 \times k$ $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$, de modo que possamos escrever

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix},$$

então, $\mathbf{A}' = (\mathbf{a}'_1\mathbf{a}'_2 ... \mathbf{a}'_n)$.

Definição D.7 (Matriz simétrica). Uma matriz quadrada **A** será uma **matriz simétrica** se, e somente se, $\mathbf{A}' = \mathbf{A}$.

Se **X** for qualquer matriz $n \times k$, então $\mathbf{X}'\mathbf{X}$ será sempre definida e uma matriz simétrica, como podemos ver ao aplicar a primeira e a quarta propriedades de transposição (ver Problema 3).

D.2e Multiplicação de matriz particionada

Seja $\mathbf{A}$ uma matriz $n \times k$ com linhas dadas pelos vetores $1 \times k$ $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ e $\mathbf{B}$ uma matriz $n \times m$ com linhas dadas pelos vetores $1 \times m$ $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}.$$

Então,

$$\mathbf{A}'\mathbf{B} = \sum_{i=1}^{n} \mathbf{a}_i' \mathbf{b}_i,$$

onde, para cada i, $\mathbf{a}'_i\mathbf{b}'_i$ é uma matriz $k \times m$. Dessa forma, $\mathbf{A}'\mathbf{B}$ pode ser escrita como a soma de n matrizes, cada uma delas $k \times m$. Como caso especial, temos

$$\mathbf{A}'\mathbf{A} = \sum_{i=1}^{n} \mathbf{a}_i' \mathbf{a}_i,$$

onde $\mathbf{a}'_i\mathbf{a}_i$ é uma matriz $k \times k$ para todo i.

Uma forma mais geral de multiplicação de matrizes particionadas se mantém quando tivermos matrizes $\mathbf{A}$ $(m \times n)$ e $\mathbf{B}$ $(n \times p)$ escritas como

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix},$$

onde $\mathbf{A}_{11}$ é $m_1 \times n_1$, $\mathbf{A}_{12}$ é $m_1 \times n_2$, $\mathbf{A}_{21}$ é $m_2 \times n_1$, $\mathbf{A}_{22}$ é $m_2 \times n_2$, $\mathbf{B}_{11}$ é $n_1 \times p_1$, $\mathbf{B}_{12}$ é $n_1 \times p_2$, $\mathbf{B}_{21}$ é $n_2 \times p_1$ e $\mathbf{B}_{22}$ é $n_2 \times p_2$. Naturalmente, $m_1 + m_2 = m$, $n_1 + n_2 = n$ e $p_1 + p_2 = p$.

Quando formamos o produto $\mathbf{AB}$, a expressão fica igual a quando as entradas são escalares:

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}.$$

Note que cada uma das multiplicações da matriz que formam a partição à direita é bem definida porque as dimensões linha e coluna são compatíveis para multiplicação.

D.2f Traço

O traço de uma matriz é uma operação muito simples, definida apenas para matrizes *quadradas*.

Definição D.8 (Traço). Para qualquer matriz $n \times n$, o **traço de uma matriz A**, indicado por $\text{tr}(\mathbf{A})$, será a soma de seus elementos diagonais. Matematicamente,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

Propriedades do Traço. (1) $\text{tr}(\mathbf{I}_n) = n$; (2) $\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$; (3) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$; (4) $\text{tr}(\alpha\mathbf{A}) = \alpha\text{tr}(\mathbf{A})$ para qualquer escalar α; e (5) $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, onde $\mathbf{A}$ é $m \times n$ e $\mathbf{B}$ é $n \times m$.

D.2g Inverso

A noção de inverso de uma matriz é muito importante para matrizes quadradas.

Definição D.9 (Inverso). Uma matriz $n \times n$ $\mathbf{A}$ terá um **inverso**, indicado por $\mathbf{A}^{-1}$, desde que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ e $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$. Neste caso, $\mathbf{A}$ é dita *invertível* ou *não singular*. No caso contrário, é chamada de *não invertível* ou *singular*.

Propriedades do inverso. (1) Se existir um inverso, ele será único; (2) $(\alpha\mathbf{A})^{-1} = (1/\alpha)\mathbf{A}^{-1}$, se $\alpha \neq 0$ e $\mathbf{A}$ for invertível; (3) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, se $\mathbf{A}$ e $\mathbf{B}$ forem ambos $n \times n$ e invertíveis; e (4) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.

Não nos preocuparemos com os mecanismos de cálculo de uma matriz inversa. Qualquer texto sobre álgebra matricial contém exemplos detalhados destes cálculos.

D.3 Independência linear e posto matricial

Para um conjunto de vetores que têm a mesma dimensão, é importante saber se um vetor pode ser expresso como uma combinação linear dos vetores restantes.

Definição D.10 (Independência linear). Defina $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r\}$ como um conjunto de vetores $n \times 1$. Eles serão **vetores linearmente independentes** se, e somente se,

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_r\mathbf{x}_r = \mathbf{0} \tag{D.2}$$

indicar que $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$. Se (D.2) se mantiver para um conjunto de escalares que não sejam todos zero, então, $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r\}$ será *linearmente dependente*.

A afirmação de que $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r\}$ é linearmente dependente é equivalente a dizer que pelo menos um vetor nesse conjunto pode ser escrito como uma combinação linear dos outros.

Definição D.11 (Posto)

(i) Seja $\mathbf{A}$ uma matriz $n \times m$. O **posto de uma matriz A**, sinalizado por posto ($\mathbf{A}$), é o número máximo de colunas linearmente independentes de $\mathbf{A}$.

(ii) Se $\mathbf{A}$ é $n \times m$ e posto ($\mathbf{A}$) $= m$, então $\mathbf{A}$ tem um *posto de coluna completo*.

Se $\mathbf{A}$ for $n \times m$, seu posto pode ser, no máximo, m. Uma matriz tem posto de coluna completo se suas colunas formarem um conjunto linearmente independente. Por exemplo, a matriz 3×2

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 0 & 0 \end{bmatrix}$$

pode ter, no máximo, posto dois. Na verdade, seu posto é somente um porque a segunda coluna é três vezes a primeira coluna.

Propriedades de Posto. (1) posto($\mathbf{A}'$) = posto($\mathbf{A}$); (2) Se $\mathbf{A}$ for $n \times k$, então posto ($\mathbf{A}$) $\leq \min(n, k)$; e (3) Se $\mathbf{A}$ for $k \times k$ e posto($\mathbf{A}$) $= k$, então $\mathbf{A}$ será invertível.

D.4 Formas quadráticas e matrizes positivas definidas

Definição D.12 (Forma quadrática). Seja $\mathbf{A}$ uma matriz $n \times n$ *simétrica*. A **forma quadrática** associada com a matriz $\mathbf{A}$ é a função de valor real definida por todos os vetores $\mathbf{x}$ $n \times 1$:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} a_{ij}x_i x_j.$$

Definição D.13 (Positiva definida e positiva semidefinida)

(i) Uma matriz simétrica $\mathbf{A}$ é dita **positiva definida (p.d)** se

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \text{ para todo vetor } \mathbf{x}\ n \times 1, \text{ exceto } \mathbf{x} = \mathbf{0}.$$

(ii) Uma matriz simétrica $\mathbf{A}$ é **positiva semidefinida (p.s.d)** se

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \text{ para todo vetor } n \times 1.$$

Se uma matriz é positiva definida ou positiva semidefinida, ela é automaticamente suposta como simétrica.

Propriedades de matrizes positivas definidas e positivas semidefinidas. (1) Uma matriz p.d. tem elementos diagonais estritamente positivos, enquanto uma matriz p.s.d. tem elementos diagonais não negativos; (2) Se $\mathbf{A}$ é p.d, então $\mathbf{A}^{-1}$ existe e é p.d; (3) Se $\mathbf{X}$ é $n \times k$, então $\mathbf{X}'\mathbf{X}$ e $\mathbf{X}\mathbf{X}'$ são p.s.d.; e (4) Se $\mathbf{X}$ é $n \times k$ e posto($\mathbf{X}$) $= k$, então $\mathbf{X}'\mathbf{X}$ é p.d. (e, portanto, não singular).

D.5 Matrizes idempotentes

Definição D.14 (Matriz idempotente). Seja $\mathbf{A}$ uma matriz simétrica $n \times n$. Dessa forma, $\mathbf{A}$ é dita como uma **matriz idempotente** se, e somente se, $\mathbf{AA} = \mathbf{A}$.

Por exemplo,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

é uma matriz idempotente, como verificado pela multiplicação direta.

Propriedades de matrizes idempotentes. Defina $\mathbf{A}$ como uma matriz idempotente $n \times n$. (1) posto($\mathbf{A}$) $=$ tr($\mathbf{A}$) e (2) $\mathbf{A}$ é positiva semidefinida.

Podemos construir matrizes idempotentes muito geralmente. Seja $\mathbf{X}$ uma matriz $n \times n$ com posto ($\mathbf{X}$) $= k$. Defina

$$\mathbf{P} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$
$$\mathbf{M} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}_n - \mathbf{P}.$$

Assim, $\mathbf{P}$ e $\mathbf{M}$ são matrizes simétricas idempotentes com posto ($\mathbf{P}$) $= k$ e posto($\mathbf{M}$) $= n - k$. Os postos são obtidos mais facilmente usando a Propriedade 1: tr($\mathbf{P}$) $=$ tr[$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$] (da Propriedade 5 do traço) $=$ tr($\mathbf{I}_k$) $= k$ (a partir da Propriedade 1 do traço). Isso resulta facilmente em tr($\mathbf{M}$) $=$ tr($\mathbf{I}_n$) – tr($\mathbf{P}$) $= n - k$.

D.6 Diferenciação das formas linear e quadrática

Para determinado vetor **a** $n \times 1$, considere a função linear definida por

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x},$$

para todos os vetores **x** $n \times 1$. A derivada de f em relação a **x** é o vetor $1 \times n$ das derivadas parciais, que é simplesmente

$$\partial f(\mathbf{x})/\partial \mathbf{x} = \mathbf{a}'.$$

Para qualquer matriz simétrica **A** $n \times n$, defina a forma quadrática

$$g(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}.$$

Assim,

$$\partial g(\mathbf{x})/\partial \mathbf{x} = 2\mathbf{x}'\mathbf{A},$$

que é um vetor $1 \times n$.

D.7 Momentos e distribuições de vetores aleatórios

Com o objetivo de derivar o valor esperado e a variância dos estimadores de MQO usando matrizes, precisamos definir o valor esperado e a variância de um **vetor aleatório**. Como o próprio nome sugere, um vetor aleatório é simplesmente um vetor de variáveis aleatórias. Também precisamos definir a distribuição normal multivariada. Esses conceitos são somente extensões daqueles que foram tratados no Apêndice B.

D.7a Valor esperado

Definição D.15 (Valor esperado)

(i) Se **y** é um vetor aleatório $n \times 1$, o **valor esperado** de **y**, sinalizado por E(**y**), é o vetor dos valores esperados: $\mathrm{E}(\mathbf{y}) = [\mathrm{E}(y_1), \mathrm{E}(y_2), ..., \mathrm{E}(y_n)]'$.

(ii) Se **Z** é uma matriz aleatória $n \times m$, E(**Z**) é a matriz $n \times m$ dos valores esperados: $\mathrm{E}(\mathbf{Z}) = [\mathrm{E}(z_{ij})]$.

Propriedades do valor esperado. (1) Se **A** é uma matriz $m \times n$ e **b** é um vetor $n \times 1$, ambos não aleatórios, então $\mathrm{E}(\mathbf{Ay} + \mathbf{b}) = \mathbf{A}\mathrm{E}(\mathbf{y}) + \mathbf{b}$; e (2) Se **A** é $p \times n$ e **B** é $m \times k$, ambas não aleatórias, $\mathrm{E}(\mathbf{AZB}) = \mathbf{A}\mathrm{E}(\mathbf{Z})\mathbf{B}$.

D.7b Matriz de variância-covariância

Definição D.16 (Matriz de variância-covariância). Se **y** for um vetor aleatório $n \times 1$, sua **matriz de variância-covariância**, indicada por Var(**y**), será definida como

$$\mathrm{Var}(\mathbf{y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix},$$

onde $\sigma_j^2 = \text{Var}(y_j)$ e $\sigma_{ij} = \text{Cov}(y_i, y_j)$. Em outras palavras, a matriz de variância-covariância tem as variâncias de cada elemento de $\mathbf{y}$ em sua diagonal, com os termos de covariância fora das diagonais. Como $\text{Cov}(y_i, y_j) = \text{Cov}(y_j, y_i)$, verifica-se imediatamente que uma matriz de variância-covariância é simétrica.

Propriedades da variância. (1) Se $\mathbf{a}$ é um vetor não aleatório $n \times 1$, $\text{Var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'[\text{Var}(\mathbf{y})\mathbf{a} \geq 0$; (2) Se $\text{Var}(\mathbf{a}'\mathbf{y}) > 0$ para todo $\mathbf{a} \neq \mathbf{0}$, $\text{Var}(\mathbf{y})$ é positiva definida; (3) $\text{Var}(\mathbf{y}) = \text{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})']$, onde $\boldsymbol{\mu} = \text{E}(\mathbf{y})$; (4) Se os elementos de $\mathbf{y}$ forem não correlacionados, $\text{Var}(\mathbf{y})$ será uma matriz diagonal. Se, além disso, $\text{Var}(y_j) = \boldsymbol{\sigma}^2$ para $j = 1, 2, ..., n$, então $\text{Var}(\mathbf{y}) = \boldsymbol{\sigma}^2\text{I}_n$; e (5) Se $\mathbf{A}$ é uma matriz não aleatória $m \times n$ e $\mathbf{b}$ é um vetor não aleatório $n \times 1$, $\text{Var}(\mathbf{Ay} + \mathbf{b}) = \mathbf{A}[\text{Var}(\mathbf{y})]\mathbf{A}'$.

D.7c Distribuição normal multivariada

A distribuição normal de uma variável aleatória foi discutida amplamente no Apêndice B. Precisamos estender a distribuição normal a vetores aleatórios. Não forneceremos uma expressão para a função de distribuição de probabilidade, já que não precisamos de uma. É importante saber que um vetor aleatório normal multivariado é completamente caracterizado por sua média e sua matriz de variância-covariância. Assim, se $\mathbf{y}$ for um vetor aleatório normal multivariado $n \times 1$ com média $\boldsymbol{\mu}$ e matriz de variância-covariância $\boldsymbol{\Sigma}$, escreveremos $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Expressaremos agora várias propriedades úteis da **distribuição normal multivariada.**

Propriedades da distribuição normal multivariada. (1) Se $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, cada elemento de $\mathbf{y}$ será normalmente distribuído; (2) Se $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, y_ie y_j, quaisquer dois elementos de $\mathbf{y}$ serão independentes se, e somente se, forem não correlacionados, isto é $\sigma_{ij} = 0$; (3) Se $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{Ay} + \mathbf{b} \sim \text{Normal}(\mathbf{A}\boldsymbol{\mu}, \mathbf{b},\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$, onde $\mathbf{A}$ e $\mathbf{B}$ são não aleatórias; (4) Se $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então, para matrizes não aleatórias $\mathbf{A}$ e $\mathbf{B}$, $\mathbf{Ay}$ e $\mathbf{By}$ serão independentes se, e somente se, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$. Em particular, se $\boldsymbol{\Sigma} = \boldsymbol{\sigma}^2\mathbf{I}_n$, então $\mathbf{AB}' = \mathbf{0}$ é necessário e suficiente para a independência de $\mathbf{Ay}$ e $\mathbf{By}$; (5) Se $\mathbf{y} \sim \text{Normal}(\mathbf{0},\boldsymbol{\sigma}^2\text{I}_n)$,$\mathbf{A}$ é uma matriz não aleatória $k \times n$ e $\mathbf{B}$ é uma matriz simétrica idempotente $n \times n$, $\mathbf{Ay}$ e $\mathbf{y}'\mathbf{By}$ serão independentes se, e somente se, $\mathbf{AB} = \mathbf{0}$; e (6) Se $\mathbf{y} \sim \text{Normal}(\mathbf{0},\boldsymbol{\Sigma}^2\text{I}_n)$ e $\mathbf{A}$ e $\mathbf{B}$ são matrizes simétricas não aleatórias e idempotentes, $\mathbf{y}'\mathbf{Ay}$ e $\mathbf{y}'\mathbf{By}$ serão independentes se, e somente se, $\mathbf{AB} = \mathbf{0}$.

D.7d Distribuição qui-quadrada

No Apêndice B, definimos uma **variável aleatória qui-quadrada** como a soma dos *quadrados* de variáveis aleatórias normais padrão independentes. Em notação vetorial, se $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_n)$, então $\mathbf{u}'\mathbf{u} \sim \chi_n^2$.

Propriedades da distribuição qui-quadrada. (1) Se $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_n)$ e $\mathbf{A}$ é uma matriz idempotente e simétrica $n \times n$ com posto$(\mathbf{A}) = q$, então $\mathbf{u}'\mathbf{Au} \sim \chi_q^2$; (2) Se $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_n)$ e $\mathbf{A}$ e $\mathbf{B}$ são matrizes idempotentes simétricas $n \times n$ de modo que $\mathbf{AB} = \mathbf{0}$, $\mathbf{u}'\mathbf{Au}$ e $\mathbf{u}'\mathbf{Bu}$ serão variáveis aleatórias independentes e qui-quadradas; e (3) Se $\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{C})$, em que $\mathbf{C}$ é uma matriz não singular $m \times m$, $\mathbf{z}'\mathbf{C}^{-1} \sim \chi_m^2$.

D.7e Distribuição *t*

Também definimos a **distribuição *t*** no Apêndice B. Agora, acrescentamos uma importante propriedade.

Propriedade da distribuição t. Se $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{c}$ for um vetor não aleatório $n \times 1$, $\mathbf{A}$ for uma matriz não aleatória, simétrica $n \times n$ e idempotente com posto q, e $\mathbf{Ac} = \mathbf{0}$, então $\{\mathbf{c}'\mathbf{u}/(\mathbf{c}'\mathbf{c})^{1/2}\}/(\mathbf{u}'\mathbf{Au}/\mathbf{q})^{1/2} \sim t_q$.

D.7f Distribuição *F*

Lembre-se de que uma **variável aleatória *F*** é obtida ao selecionar duas variáveis aleatórias qui-quadradas *independentes* e ao encontrar a razão de cada uma delas, padronizada por graus de liberdade.

Propriedade da distribuição F. Se $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_n)$ e $\mathbf{A}$ e $\mathbf{B}$ forem matrizes $n \times n$ não aleatórias, simétricas e idempotentes com posto$(\mathbf{A}) = k_1$, posto$(\mathbf{B}) = k_2$ e $\mathbf{AB} = \mathbf{0}$, então $(\mathbf{u}'\mathbf{Au}/\mathrm{k}_1)/(\mathbf{u}'\mathbf{Bu}/\mathrm{k}_2) \sim F_{k_1, k_2}$.

Resumo

Este apêndice contém uma forma condensada das informações básicas necessárias para estudar o modelo linear clássico utilizando matrizes. Embora seja independente, este material foi elaborado principalmente como uma revisão para leitores que já estão familiarizados com álgebra matricial e estatística multivariada, e será usado amplamente no Apêndice E.

Termos-chave

Distribuição normal multivariada
Distribuição *t*
Forma quadrática
Inverso
Matriz
Matriz de variância-covariância
Matriz diagonal
Matriz idempotente
Matriz identidade
Matriz nula
Matriz quadrada
Matriz simétrica
Multiplicação de matrizes
Multiplicação escalar
Positiva definida (p.d)
Positiva semidefinida (p.s.d)
Posto de uma matriz A
Traço de uma matriz
Transposição
Valor esperado
Variável aleatória *F*
Variável aleatória qui-quadrada
Vetor aleatório
Vetor coluna
Vetor linha
Vetores linearmente independentes

Problemas

1 (i) Encontre o produto **AB** usando

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 1 & 6 \\ 1 & 8 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

(ii) O produto **BA** existe?

2 Se **A** e **B** são matrizes diagonais $n \times n$, mostre que $\mathbf{AB} = \mathbf{BA}$.

3 Seja **X** qualquer matriz $n \times k$. Mostre que $\mathbf{X}'\mathbf{X}$ é uma matriz simétrica.

4 (i) Use as propriedades de traço para defender que $\text{tr}(\mathbf{A}'\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{A}')$ em qualquer matriz **A** $n \times m$.

(ii) Para $\mathbf{A} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 3 & 0 \end{bmatrix}$, verifique se $\text{tr}(\mathbf{A}'\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{A}')$.

5 (i) Use a definição de inverso para provar o seguinte: se **A** e **B** são matrizes $n \times n$ não singulares, então$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

(ii) Se **A, B** e **C** são matrizes $n \times n$ não singulares, descubra $(\mathbf{ABC})^{-1}$ em termos de $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ e $\mathbf{C}^{-1}$.

6 (i) Mostre que, se **A** for uma matriz positiva definida $n \times n$ simétrica, então **A** deverá ter elementos diagonais estritamente positivos.

(ii) Desenhe uma matriz simétrica 2×2 com elementos diagonais estritamente positivos que *não* seja positiva definida.

7 Seja **A** uma matriz positiva definida $n \times n$ simétrica. Mostre que, se **P** for uma matriz não singular $n \times n$, $\mathbf{P}'\mathbf{AP}$ será positiva definida.

8 Prove a Propriedade 5 das variâncias para vetores usando a Propriedade 3.

9 Defina **a** como um vetor não aleatório $n \times 1$ e **u** como um vetor aleatório $n \times 1$ com $\text{E}(\mathbf{uu}') = \mathbf{I}_n$. Mostre que $\text{E}[\text{tr}(\mathbf{auu}'\mathbf{a}')] = \sum_{i=1}^{n} a_i^2$.

10 Partindo das propriedades da distribuição qui-quadrada listadas no texto, mostre que essas propriedades, ao lado da definição de uma variável aleatória *F*, implicam a propriedade fixa da distribuição *F* (em relação às razões das formas quadráticas).

11 Seja **X** uma matriz $n \times k$ particionada como

$$\mathbf{X} = (\mathbf{X}_1 \;\; \mathbf{X}_2),$$

em que $\mathbf{X}_1$ é $n \times k_1$ e $\mathbf{X}_2$ é $n \times k_2$.

(i) Mostre que

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}.$$

Quais são as dimensões de cada uma das matrizes?

(ii) Seja **b** um vetor $k \times 1$, particionado como

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix},$$

em que $\mathbf{b}_1$ é $k_1 \times 1$ e $\mathbf{b}_2$ é $k_2 \times 1$. Mostre que

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)\mathbf{b}_1 + (\mathbf{X}_1'\mathbf{X}_2)\mathbf{b}_2 \\ (\mathbf{X}_2'\mathbf{X}_1)\mathbf{b}_1 + (\mathbf{X}_2'\mathbf{X}_2)\mathbf{b}_2 \end{pmatrix}.$$

APÊNDICE E

O modelo de regressão linear em forma de matriz

Este apêndice deriva vários resultados para a estimação de mínimos quadrados ordinários do modelo de regressão linear múltiplo, utilizando notação de matriz e álgebra matricial (veja um resumo sobre o assunto no Apêndice D).

E-1 Modelo e estimação de mínimos quadrados ordinários

Por meio deste apêndice, usamos o t subscrito às observações do índice e um n para denotar o tamanho do amostra. É útil escrever o modelo de regressão linear múltipla com parâmetros k, como se segue:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t,\ t = 1, 2, \ldots, n, \qquad \text{(E.1)}$$

em que y_t é a variável dependente para observação t e x_{tj}, $j = 1, 2, \ldots, k$, são as variáveis independentes. Como de hábito, β_0 é o intercepto e $\beta_1, \ldots, \beta_k$ denota os parâmetros de inclinação.

Para cada t, defina um vetor $1 \times (k + 1)$, $\mathbf{x}_t = (1, x_{t1}, \ldots, x_{tk})$, e considere que $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ seja o vetor $(k + 1) \times 1$ de todos os parâmetros. Depois, podemos escrever (E.1) como

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t,\ t = 1, 2, \ldots, n. \qquad \text{(E.2)}$$

[Alguns autores preferem definir $\mathbf{x}_t$ como um vetor de coluna, caso $\mathbf{x}_t$ seja substituído por $\mathbf{x}_t'$ em (E.2). Matematicamente, faz mais sentido defini-lo como um vetor coluna]. Podemos escrever (E.2) em notação matricial integral ao definir de forma adequada os vetores de dados e matrizes. Considere que $\mathbf{y}$ denote o vetor de observações $n \times 1$ em y: o elemento t^{th} de $\mathbf{y}$ é y_t. Considere que $\mathbf{X}$ seja o vetor de observações $n \times (k + 1)$ sobre as variáveis explicativas. Em outras palavras, a coluna t^{th} de $\mathbf{X}$ consiste no vetor $\mathbf{x}_t$. Descrito em detalhes,

$$\underset{n \times (k+1)}{\mathbf{X}} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}.$$

Por fim, consideremos $\mathbf{u}$ o vetor de erros ou interferências não observáveis $n \times 1$. Sendo assim, podemos escrever (E.2) para todas as observações n em **notação de matriz**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \tag{E.3}$$

Lembre-se, já que $\mathbf{X}$ é $n \times (k + 1)$ e $\boldsymbol{\beta}$ é $(k + 1) \times 1$, $\mathbf{X}\boldsymbol{\beta}$ é $n \times 1$.

A estimação de $\boldsymbol{\beta}$ funciona por meio da minimização da soma de resíduos quadrados, conforme a Seção 3.2. Defina a soma da função de resíduos quadrados para qualquer vetor $\mathbf{b}$ de parâmetro possível $(k + 1) \times 1$, conforme

$$\text{SQR}(\mathbf{b}) \equiv \sum_{t=1}^{n} (y_t - \mathbf{x}_t\mathbf{b})^2.$$

O vetor de estimativas de mínimos quadrados ordinários $(k + 1) \times 1$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)'$, minimiza SQR($\mathbf{b}$) sobre todos os possíveis vetores $\mathbf{b}$ $(k + 1) \times 1$. Isso é um problema que ocorre em cálculo multivariável. Para que $\hat{\boldsymbol{\beta}}$ minimize a soma dos resíduos quadrados, é necessário resolver a **condição de primeira ordem**

$$\partial \text{SQR}(\hat{\boldsymbol{\beta}})/\partial \mathbf{b} \equiv 0. \tag{E.4}$$

Levando em conta o fato de que o derivativo de $(y_t - \mathbf{x}_t\mathbf{b})^2$ em relação a $\mathbf{b}$ é o $1 \times (k + 1)$ vetor $-2(y_t - \mathbf{x}_t\mathbf{b})\mathbf{x}_t$, (E.4) é equivalente a

$$\sum_{t=1}^{n} \mathbf{x}_t'(y_t - \mathbf{x}_t\hat{\boldsymbol{\beta}}) \equiv \mathbf{0}. \tag{E.5}$$

(Dividimos por -2 e fizemos a transposição). Podemos escrever essa condição de primeira ordem como

$$\begin{aligned}
&\sum_{t=1}^{n} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \dots - \hat{\beta}_k x_{tk}) = 0 \\
&\sum_{t=1}^{n} x_{t1}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \dots - \hat{\beta}_k x_{tk}) = 0 \\
&. \\
&. \\
&. \\
&\sum_{t=1}^{n} x_{tk}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \dots - \hat{\beta}_k x_{tk}) = 0,
\end{aligned}$$

que é idêntica às condições de primeira ordem na equação (3.13). Queremos escrevê-las na forma de matriz para tornar mais fácil de manipulá-las. Usando a fórmula para multiplicação particionada do Apêndice D, vemos que (E.5) é equivalente a

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \tag{E.6}$$

ou

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{E.7}$$

Pode ser mostrado que (E.7) sempre teve ao menos uma solução. Múltiplas soluções não nos ajudam caso estejamos procurando um conjunto único de estimativas MQO, visto nosso conjunto de dados. Presumindo que a matriz simétrica $(k + 1) \times (k + 1)$ $\mathbf{X}'\mathbf{X}$ não seja singular, podemos multiplicar previamente ambos os lados de (E.7) por $(\mathbf{X}'\mathbf{X})^{-1}$ para resolver o estimador MQO $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{E.8}$$

Essa é a fórmula essencial para a análise de matriz do modelo de regressão linear múltipla. A hipótese de que $\mathbf{X}'\mathbf{X}$ seja invertível equivale à hipótese de classificação rank$(\mathbf{X}) = (k + 1)$, o que significa que as colunas de $\mathbf{X}$ devem ser linearmente independentes. Essa é a versão matricial de RLM.3 do Capítulo 3.

Antes de continuarmos, é necessário deixar um alerta sobre (E.8). É tentador simplificar a fórmula para $\hat{\boldsymbol{\beta}}$, conforme segue:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{-1}\mathbf{y}.$$

A falha nesse raciocínio é que $\mathbf{X}$ geralmente não é uma matriz quadrada, ou seja, não pode ser invertida. Em outras palavras, não podemos escrever $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}$ a menos que $n = (k + 1)$, um caso que quase nunca surge na prática.

Os vetores de valores ajustados e os resíduos de MQO $n \times 1$ são dados por

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}, \ \hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}, \text{ respectivamente.}$$

Com (E.6) e a definição de $\hat{\mathbf{u}}$, podemos ver que a condição de primeira ordem para $\hat{\boldsymbol{\beta}}$ é a mesma que

$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}. \tag{E.9}$$

Já que a primeira coluna de $\mathbf{X}$ é composta inteiramente deles, (E.9) implica que os resíduos MQO sempre somatizam zero quando um intercepto é incluído na equação e que a covariância de amostra entre cada variável independente e os resíduos MQO é zero (discutimos ambas as propriedades no Capítulo 3).

A soma dos resíduos quadrados pode ser escrita como

$$\text{SQR} = \sum_{t=1}^{n}\hat{u}_t^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{E.10}$$

Todas as propriedades algébricas do Capítulo 3 podem ser derivadas usando álgebra matricial. Por exemplo, podemos mostrar que a soma total de quadrados é igual à soma explicada de quadrados, mais a soma dos resíduos quadrados [ver (3.27)]. O uso das matrizes não fornece uma prova mais simples do que a notação somatória, dessa forma, não fornecemos outra derivação.

A abordagem de matriz para regressão múltipla pode ser usada como base para a interpretação geométrica da regressão. Isso envolve conceitos matemáticos ainda mais avançados do que os abordados no Apêndice D [Ver Goldberger (1991) ou Greene (1997)].

E.1a Teorema Frisch-Waugh

Na Seção 3.2, descrevemos uma interpretação "parcialmente ausente" das estimativas de mínimos quadrados ordinários. Podemos estabelecer a interpretação parcialmente ausente de forma muito geral, usando notação matricial. Dividindo a matriz $\mathbf{X}$ $n \times (k + 1)$ como

$$\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2),$$

em que $\mathbf{X}_1$ é $n \times (k_1 + 1)$ e inclui o intercepto – mesmo que não seja exigido para que o resultado se mantenha – e $\mathbf{X}_2$ é $n \times k_2$. Ainda supomos que $\mathbf{X}$ tenha classificado $k + 1$, o que significa que $\mathbf{X}_1$ classificou $k_1 + 1$ e que $\mathbf{X}_2$ classificou k_2.

Considere as estimativas MQO $\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_2$ da (longa) regressão

$$\mathbf{y} \text{ em } \mathbf{X}_1, \mathbf{X}_2.$$

Como sabemos, os coeficientes de regressão múltipla em $\mathbf{X}_2$, $\hat{\boldsymbol{\beta}}_2$, geralmente diferem de $\tilde{\boldsymbol{\beta}}_2$ da regressão $\mathbf{y}$ em $\mathbf{X}_2$. Uma forma de descrever a diferença é compreender que podemos obter $\hat{\boldsymbol{\beta}}_2$ de uma regressão menor, mas que primeiro devemos "parcelar de forma ausente" $\mathbf{X}_1$ de $\mathbf{X}_2$. Considere o seguinte método, de duas etapas:

(i) Regrida (cada coluna de) $\mathbf{X}_2$ em $\mathbf{X}_1$ e obtenha a matriz dos resíduos, digamos $\ddot{\mathbf{X}}_2$. Podemos escrever $\ddot{\mathbf{X}}_2$ como

$$\ddot{\mathbf{X}}_2 = [\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1']\mathbf{X}_2 = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2 = \mathbf{M}_1\mathbf{X}_2,$$

em que $\mathbf{P}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ e $\mathbf{M}_1 = \mathbf{I}_{n\,-}\,\mathbf{P}_1$ são $n \times n$ matrizes idempotentes simétricas.

(ii) Regrida $\mathbf{y}$ em $\ddot{\mathbf{X}}_2$ e preveja o vetor $k_2 \times 1$ de coeficiente $\ddot{\boldsymbol{\beta}}_2$.

O **teorema de Frisch-Waugh (FW)** declara que

$$\ddot{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}_2.$$

Consideravelmente, o teorema FW em geral não diz nada sobre a igualdade das estimativas de regressão longa, $\hat{\boldsymbol{\beta}}_2$, e sobre as da regressão curta, $\tilde{\boldsymbol{\beta}}_2$. Geralmente, $\hat{\boldsymbol{\beta}}_2 \neq \tilde{\boldsymbol{\beta}}_2$. Contudo, se $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, então $\ddot{\mathbf{X}}_2 = \mathbf{M}_1\mathbf{X}_2$, em cujo caso $\ddot{\boldsymbol{\beta}}_2 = \tilde{\boldsymbol{\beta}}_2$; logo, $\hat{\boldsymbol{\beta}}_2 = \tilde{\boldsymbol{\beta}}_2$ resulta de FW. Também vale a pena ressaltar que vamos obter $\hat{\boldsymbol{\beta}}_2$ se também favorecermos $\mathbf{X}_1$ em detrimento de $\mathbf{y}$. Em outras palavras, consideremos $\ddot{\mathbf{y}}$ os resíduos de regredir $\mathbf{y}$ em $\mathbf{X}_1$, de forma que

$$\ddot{\mathbf{y}} = \mathbf{M}_1\mathbf{y}.$$

Logo, $\hat{\boldsymbol{\beta}}_2$ é obtido da regressão de $\ddot{\mathbf{y}}$ em $\ddot{\mathbf{X}}_2$. É importante entender que não é suficiente apenas favorecer $\mathbf{X}_1$ em detrimento de $\mathbf{y}$. Um passo importante é favorecer $\mathbf{X}_1$ em detrimento de $\mathbf{X}_2$. O Problema 6, no final deste capítulo, pede que você derive o teorema FW e investigue algumas questões relacionadas.

Outro resultado algébrico útil ocorre quando regredimos $\ddot{\mathbf{y}}$ em $\ddot{\mathbf{X}}_2$ e poupamos os resíduos, por exemplo, $\ddot{\mathbf{u}}$, estes são idênticos aos resíduos MQO da regressão (longa) original:

$$\ddot{\mathbf{y}} = \ddot{\mathbf{X}}_2\hat{\boldsymbol{\beta}}_2 = \ddot{\mathbf{u}} = \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2,$$

em que usamos o resultado FW $\ddot{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}_2$. Não obtemos os resíduos MQO originais se regredirmos $\mathbf{y}$ em $\ddot{\mathbf{X}}_2$ (mas obtemos $\ddot{\boldsymbol{\beta}}_2$).

Antes dos computadores eficazes, o resultado Frisch-Waugh era às vezes usado como dispositivo computacional. Atualmente, o resultado possui interesse mais teórico, e é muito útil para compreender os mecanismos de MQO. Por exemplo, lembre-se de que no Capítulo 10 usamos o teorema FW para estabelecer que adicionar uma tendência temporal a uma regressão múltipla é algebricamente equivalente a primeiro “detrend” linearmente todas as variáveis explicativas antes de estabelecer a regressão. O teorema FW também foi usado no Capítulo 14 para estabelecer que o estimador de efeitos fixos, que apresentamos como obtido de dados temporais reduzidos, também pode ser obtido da regressão de variável *dummy* (longa).

E.2 Propriedades finitas da amostra de MQO

Derivar o valor e a variação esperados do estimador MQO $\hat{\boldsymbol{\beta}}$ é facilitado por álgebra matricial, mas devemos demonstrar algum cuidado ao declarar as hipóteses.

Hipótese E.1 Linear em parâmetros

O modelo pode ser escrito como em (E.3), em que $\mathbf{y}$ é um vetor observado $n \times 1$, $\mathbf{X}$ é uma matriz observada $n \times (k + 1)$ e $\mathbf{u}$ é um vetor de erros ou distúrbios não observados $n \times 1$.

Hipótese E.2 Colinearidade imperfeita

A matriz $\mathbf{X}$ classifica $(k + 1)$.

Essa é uma declaração cuidadosa da hipótese, que descarta dependências lineares entre as variáveis explicativas. Sob a Hipótese E.2, $\mathbf{X}'\mathbf{X}$ não é singular, ou seja, $\hat{\boldsymbol{\beta}}$ é única e pode ser escrita como em (E.8).

Hipótese E.3 Medida condicional zero

Condicional à matriz $\mathbf{X}$ inteira, cada erro u_t tem medida zero: $E(u_t|\mathbf{X}) = 0$, $t = 1, 2, \ldots, n$.

Na forma de vetor, a Hipótese E.3 pode ser escrita como

$$E(\mathbf{u}|\mathbf{X}) = \mathbf{0}. \tag{E.11}$$

Essa hipótese é implicada por RLM.4 sob a hipótese de amostra aleatória RLM.2. Em aplicações de série temporal, a Hipótese E.3 impõe estrita exogeneidade nas variáveis explicativas, algo discutido longamente no Capítulo 10. Isso exclui variáveis explicativas cujos valores futuros sejam correlacionados com u_t; em especial, isso elimina variáveis dependentes defasadas. Sob a Hipótese E.3, podemos condicionar em x_{tj} quando calculamos o valor esperado de $\hat{\boldsymbol{\beta}}$.

TEOREMA E.1

INEXISTÊNCIA DE VIÉS DO MQO

Sob as Hipóteses E.1, E.2 e E.3, o estimador MQO $\hat{\boldsymbol{\beta}}$ é não viesado para $\boldsymbol{\beta}$.

Demonstração: Utilize as Hipóteses E.1, E.2 e álgebra simples para escrever

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u},\end{aligned} \tag{E.12}$$

em que usamos o fato de que $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}_{k+1}$. Colocar a expectativa condicional em $\mathbf{X}$ nos dá

$$\begin{aligned}\mathrm{E}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{u}|\mathbf{X}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \boldsymbol{\beta},\end{aligned}$$

logo, $\mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$, sob a hipótese E.3. Esse argumento claramente não depende do valor de $\boldsymbol{\beta}$, então mostramos que $\hat{\boldsymbol{\beta}}$ é não viesado.

Para obter a forma mais simples de matriz de variação-covariação de $\hat{\boldsymbol{\beta}}$, introduzimos as hipóteses de homoscedasticidade e nenhuma correlação serial.

Hipótese E.4. Homoscedasticidade sem correlação serial

(i) $\mathrm{Var}(u_t|\mathbf{X}) = \sigma^2$, $t = 1, 2, \ldots, n$. (ii) $\mathrm{Cov}(u_t,u_s|\mathbf{X}) = 0$, para todos $t \neq s$. Em forma de matriz, podemos escrever essas duas hipóteses como

$$\mathrm{Var}(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_{n'} \tag{E.13}$$

em que $\mathbf{I}_n$ é a matriz de identidade $n \times n$.

A parte (i) da Hipótese E.4 é a hipótese de homoscedasticidade: a variação de u_t não pode depender de qualquer elemento de **X**, e a variação deve ser constante nas variações, t. A parte (ii) é a hipótese de não correlação serial: os erros não podem ser correlacionados nas observações. Sob a amostragem aleatória, e em quaisquer outros esquemas de amostragem de corte transversal com observações independentes, a parte (ii) da Hipótese E.4 automaticamente se sustenta. Para aplicações de séries temporais, a parte (ii) exclui correlação nos erros com o tempo (ambos condicionais em **X** e incondicionalmente).

Por causa de (E.13), frequentemente dizemos que **u** tem uma **matriz de variação-covariação escalar** caso a Hipótese E.4 se mantenha. Podemos agora derivar **matriz de variação-covariação do estimador MQO.**

TEOREMA E.2

MATRIZ DE VARIAÇÃO-COVARIAÇÃO DO ESTIMADOR MQO

Sob a Hipótese E.1 em E.4,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{E.14}$$

Prova: Da última fórmula da equação (E.12), temos

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \text{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\text{Var}(\mathbf{u}|\mathbf{X})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Agora, usamos a Hipótese E.4 para obter

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

A fórmula (E.14) significa que a variação de $\hat{\beta}_j$ (condicional em $\mathbf{X}$) é obtida pela multiplicação de σ^2 pelo elemento diagonal j^{th} de $(\mathbf{X}'\mathbf{X})^{-1}$. Para os coeficientes de inclinação, damos uma fórmula interpretável à equação (3.51). A equação (E.14) também nos diz como obter a covariação entre qualquer uma das duas estimativas MQO: multiplicar σ^2 pelo elemento de $(\mathbf{X}'\mathbf{X})^{-1}$ fora da diagonal. No Capítulo 4, mostramos como evitar explicitamente encontrar covariações para obter intervalos de confiança e testes de hipótese ao reescrever o modelo de forma apropriada.

O Teorema de Gauss-Markov, em sua plena generalidade, pode ser provado.

TEOREMA E.3

TEOREMA DE GAUSS-MARKOV

Sob a Hipótese E.1 em E.4, $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear não viesado.

Prova: Qualquer outro estimador linear de $\boldsymbol{\beta}$ pode ser escrito como

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}'\mathbf{y}, \tag{E.15}$$

em que $\mathbf{A}$ é uma matriz $n \times (k + 1)$. Para que $\tilde{\boldsymbol{\beta}}$ seja um condicional não viesado em $\mathbf{X}$, $\mathbf{A}$ pode consistir em números e funções não aleatórios de $\mathbf{X}$ (por exemplo, $\mathbf{A}$ não pode ser uma função de $\mathbf{y}$). Para ver que restrições adicionais de $\mathbf{A}$ são necessárias, escreva

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = (\mathbf{A}'\mathbf{X})\boldsymbol{\beta} + \mathbf{A}'\mathbf{u}. \tag{E.16}$$

Logo,

$$\begin{aligned}\text{E}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \text{E}(\mathbf{A}'\mathbf{u}|\mathbf{X}) \\ &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\text{E}(\mathbf{u}|\mathbf{X}) \text{ porque } \mathbf{A} \text{ é uma função de } \mathbf{X} \\ &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} \text{ porque E } (\mathbf{u}|\mathbf{X}) = \mathbf{0}.\end{aligned}$$

Para que $\tilde{\boldsymbol{\beta}}$ seja um estimador não viesado de $\boldsymbol{\beta}$, é preciso que E $(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$ para todos os vetores $(k + 1) \times 1$ $\boldsymbol{\beta}$, ou seja,

$$\mathbf{A}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \text{ para todos } (k + 1) \times 1 \text{ vetores } \boldsymbol{\beta}. \tag{E.17}$$

Como $\mathbf{A}'\mathbf{X}$ é uma matriz $(k + 1) \times (k + 1)$, (E.17) se sustenta apenas se $\mathbf{A}'\mathbf{X} = \mathbf{I}_{k+1}$. As equações (E.15) e (E.17) caracterizam a classe de estimadores não viesados, lineares para $\boldsymbol{\beta}$.

Na sequência, de (E.16), temos

$$\text{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \mathbf{A}'[\text{Var}(\mathbf{u}|\mathbf{X})]\mathbf{A} = \sigma^2\mathbf{A}'\mathbf{A},$$

pela Hipótese E.4. Portanto,

$$\begin{aligned}\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) &= \sigma^2[\mathbf{A}'\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}]\\ &= \sigma^2[\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}] \text{ porque } \mathbf{A}'\mathbf{X} = \mathbf{I}_{k+1}\\ &= \sigma^2\mathbf{A}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{A}\\ &\equiv \sigma^2\mathbf{A}'\mathbf{M}\mathbf{A},\end{aligned}$$

em que $\mathbf{M} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Como $\mathbf{M}$ é simétrico e idempotente, $\mathbf{A}'\mathbf{M}\mathbf{A}$ é positivo e semidefinitivo para qualquer matriz $\mathbf{A}$ $n \times (k + 1)$. Isso estabelece que o estimador MQO $\hat{\boldsymbol{\beta}}$ é BLUE. Por que isso é importante? Leve em conta que $\mathbf{c}$ é um vetor $k + 1 \times 1$ qualquer e considere a combinação linear $\mathbf{c}'\boldsymbol{\beta} = c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k$ como escalar. Os estimadores não viesados de $\mathbf{c}'\boldsymbol{\beta}$ são $\mathbf{c}'\hat{\boldsymbol{\beta}}$ e $\mathbf{c}'\tilde{\boldsymbol{\beta}}$. Contudo

$$\mathrm{Var}(\mathbf{c}'\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}}|\mathbf{X}) = \mathbf{c}'[\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\hat{\beta}|\mathbf{X})]\mathbf{c} \geq 0,$$

porque $[\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\hat{\beta}|\mathbf{X})]$ é p.s.d. Logo, quando é usado para estimar qualquer combinação linear de $\boldsymbol{\beta}$, MQO rende a menor variação. Em particular, $\mathrm{Var}(\hat{\beta}_j|\mathbf{X}) \leq \mathrm{Var}(\tilde{\beta}_j|\mathbf{X})$ para qualquer outro estimador linear, não viesado de β_j.

O estimador não viesado da variação de erro σ^2 pode ser escrito como

$$\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n - k - 1),$$

significando o mesmo que a equação (3.56).

TEOREMA E.4

INEXISTÊNCIA DE VIÉS DE $\hat{\sigma}^2$

Segundo as Hipóteses de E.1 a E.4, $\hat{\sigma}^2$ é não viesada para $\hat{\sigma}^2$: $\mathrm{E}(\hat{\sigma}^2|\mathbf{X}) = \sigma^2$ para todo $\sigma^2 > 0$.

Prova: Escreva $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}$, em que $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ e a última equação ocorrem porque $\mathbf{M}\mathbf{X} = \mathbf{0}$. Como $\mathbf{M}$ é simétrico e idempotente,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u}.$$

Como $\mathbf{u}'\mathbf{M}$ é escalar, se iguala a seu traço. Portanto,

$$\begin{aligned}\mathrm{E}(\mathbf{u}'\mathbf{M}\mathbf{u}|\mathbf{X}) &= \mathrm{E}[\mathrm{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})|\mathbf{X}] = \mathrm{E}[\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')|\mathbf{X}]\\ &= \mathrm{tr}[\mathrm{E}(\mathbf{M}\mathbf{u}\mathbf{u}'|\mathbf{X})] = \mathrm{tr}[\mathbf{M}\mathrm{E}(\mathbf{u}\mathbf{u}'|\mathbf{X})]\\ &= \mathrm{tr}(\mathbf{M}\sigma^2\mathbf{I}_n) = \sigma^2\mathrm{tr}(\mathbf{M}) = \sigma^2(n - k - 1).\end{aligned}$$

A última igualdade ocorre a partir de $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = n - \mathrm{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = n - \mathrm{tr}(\mathbf{I}_{k+1}) = n - (k + 1) = n - k - 1$. Logo,

$$\mathrm{E}(\hat{\sigma}^2|\mathbf{X}) = \mathrm{E}(\mathbf{u}'\mathbf{M}\mathbf{u}|\mathbf{X})/(n - k - 1) = \sigma^2.$$

E.3. Inferência estatística

Quando adicionamos a hipótese do modelo linear clássico, $\hat{\boldsymbol{\beta}}$ apresenta distribuição normal multivariada, o que nos leva às distribuições t e F para as estatísticas de teste padrão abordadas no Capítulo 4.

Hipótese E.5 **Normalidade de erros**

Condicional em **X**, o u_t é independente e identicamente distribuído como Normal $(0,\sigma^2)$. De forma equivalente, visto **X**, **u** é distribuído como multivariado normal com significado zero e matriz de variação-covariação $\sigma^2\mathbf{I}_n$: $u \sim \text{Normal}\,(0,\sigma^2\mathbf{I}_n)$.

Sob a Hipótese E.5, cada u_t é independente das variáveis explicativas para todos os t. Ao se estabelecer séries temporais, essa é a hipótese de exogeneidade estrita.

TEOREMA E.5

NORMALIDADE DE $\hat{\boldsymbol{\beta}}$

Sob as hipóteses de modelo linear clássico de E.1 a E.5, o condicional $\hat{\boldsymbol{\beta}}$ em **X** é distribuído como normal multivariado com significado $\boldsymbol{\beta}$ e matriz de variação-covariação $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

O Teorema E.5 é a base para inferência estatística envolvendo $\boldsymbol{\beta}$. De fato, juntamente com as propriedades de qui-quadrado e as distribuições t e F resumidas no Apêndice D, podemos usar o Teorema E.5 para estabelecer que estatísticas t têm uma distribuição t segundo as Hipóteses de E.1 a E.5 (sob a hipótese nula) e igualmente para estatísticas F. Ilustramos essa situação com uma demonstração de estatísticas t.

TEOREMA E.6

DISTRIBUIÇÃO DE ESTATÍSTICA *T*

Sob as Hipóteses de E.1 a E.5,

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) \sim t_{n-k-1},\ j = 0, 1, \ldots, k.$$

***Prova*:** A prova exige várias etapas; as seguintes declarações são inicialmente condicionais em **X**. Primeiro, pelo Teorema E.5, $(\hat{\beta}_j - \beta_j)/\text{dp}(\hat{\beta}_j) \sim \text{Normal}\,(0, 1)$, em que $\text{dp}(\hat{\beta}_j) = \sigma\sqrt{C_{jj}}$, e c_{jj} são os j^{th} elementos diagonais de $(\mathbf{X}'\mathbf{X})^{-1}$. A seguir, de acordo com as Hipóteses de E.1 a E.5, condicionais em **X**,

$$(n - k - 1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-k-1}. \tag{E.18}$$

Isso ocorre porque $(n - k - 1)\ \hat{\sigma}^2/\hat{\sigma}^2 = (\mathbf{u}/\sigma)'\mathbf{M}(\mathbf{u}/\sigma)$, em que **M** é a matriz $n \times n$, simétrica, idempotente, definida no Teorema E.4. Contudo, $\mathbf{u}/\sigma \sim \text{Normal}\,(\mathbf{0},\mathbf{I}_n)$, de acordo com a Hipótese E.5. Ocorre na Propriedade 1 para distribuição qui-quadrada do Apêndice D que $(\mathbf{u}/\sigma)'\mathbf{M}(\mathbf{u}/\sigma) \sim \chi^2_{n-k-1}$ (já que **M** avaliou $n - k - 1$).

Também precisamos mostrar que $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ são independentes. Contudo, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ e $\hat{\sigma}^2 = \mathbf{u}'\mathbf{M}\mathbf{u}\ (n - k - 1)$. Agora, $[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{M} = \mathbf{0}$, já que $\mathbf{X}'\mathbf{M} = \mathbf{0}$. Ocorre, na Propriedade 5 da distribuição normal multivariada do Apêndice D, que $\hat{\boldsymbol{\beta}}$ e **Mu** são independentes. Já que $\hat{\sigma}^2$ é uma função de **Mu**, $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ são também independentes.

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) = [(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)]/(\hat{\sigma}^2/\sigma^2)^{1/2},$$

que é a proporção de uma variável aleatória normal padrão e a raiz quadrada de uma variável aleatória $\chi^2_{n-k-1}/(n - k - 1)$. Apenas mostramos que ela é independente, então, por definição de uma variável aleatória t, $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)$ possui a distribuição t_{n-k-1}. Como essa distribuição não depende de **X**, também é a distribuição incondicional de $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)$.

Podemos associar a partir desse teorema qualquer valor hipotético para β_j e usar a estatística t para testar hipóteses, como habitualmente.

Segundo as Hipóteses de E.1 a E.5, podemos calcular o que se conhece como limite inferior de *Cramer-Rao* para matriz de variação-covariação de estimadores não viesados de $\boldsymbol{\beta}$ (novamente, condicional em $\mathbf{X}$) [ver Greene (1997, Capítulo 4)]. Isso pode ser mostrado como $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, exatamente a matriz de variação-covariação do estimador MQO. Isso implica que $\hat{\boldsymbol{\beta}}$ é o **estimador não viesado de variância mínima** de $\boldsymbol{\beta}$ (condicional em $\mathbf{X}$): $\text{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})$ é positivo, semidefinitivo para qualquer outro estimador não viesado de $\tilde{\boldsymbol{\beta}}$; não precisamos mais restringir nossa atenção aos estimadores lineares em $\mathbf{y}$.

É fácil mostrar que os estimadores MQO são, de fato, os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, segundo a Hipótese E.5. Para cada t, a distribuição de y_t, dado $\mathbf{X}$, é Normal($\mathbf{x}_t\boldsymbol{\beta}$,$\sigma^2$). Como y_t é condicional independente em $\mathbf{X}$, a função de probabilidade para a amostra é obtida do produto das densidades:

$$\prod_{t=1}^{n}(2\pi\sigma^2)^{-1/2}\exp[-(y_t - \mathbf{x}_t\boldsymbol{\beta})^2/(2\sigma^2)],$$

em que Π denota produto. Maximizar essa função relacionando $\boldsymbol{\beta}$ e σ^2 é o mesmo que maximizar seu logaritmo natural:

$$\sum_{t=1}^{n}[-(1/2)\log(2\pi\sigma^2) - (y_t - \mathbf{x}_t\boldsymbol{\beta})^2/(2\sigma^2)].$$

Para obter $\hat{\boldsymbol{\beta}}$, basta fazer o mesmo que minimizar $\sum_{t=1}^{n}(y_t - \mathbf{x}_t\boldsymbol{\beta})^2$– a divisão por $2\sigma^2$ não afeta a otimização – justamente o problema que MQO resolve. O estimador de σ^2 que usamos, SQR/($n - k$), descobriu-se não ser o EMV de σ^2; o EMV é SQR/n, estimador viesado. Como o estimador não viesado de σ^2 resulta em estatísticas t e F, com exatas distribuições t e F sob a nula, ele é sempre usado em vez de MLE.

O fato de que o estimador MQO é o EMV, sob a Hipótese E.5, implica uma interessante propriedade de robustez do EMV com base em distribuição normal. O motivo é simples. Sabemos que o estimador MQO é não viesado sob as Hipóteses de E.1 a E.3; a normalidade dos erros não é usada em lugar algum na demonstração, assim como a Hipótese 4. Como a próxima seção mostra, o estimador MQO também é consistente sem normalidade, contanto que a lei de grandes quantidades se sustente (o que é amplamente verdade). Essas propriedades estatísticas do estimador MQO implicam que o EMV com base em função normal de verossimilhança logarítmica é robusto para especificação distribucional: a distribuição pode ser (quase) qualquer coisa e ainda assim obtemos um estimador consistente (e, de E.1 a E.3, não viesado). Como discutido na Seção 17.3, um estimador de máxima verossimilhança obtido ao se presumir que a distribuição esteja correta é frequentemente chamado de **estimador por quase máxima verossimilhança (EQMV)**.

Geralmente, a consistência do EMV depende de uma distribuição correta para concluir que é consistente para os parâmetros. Vimos que a distribuição normal é uma exceção notável. Existem outras distribuições que compartilham essa propriedade, incluindo a distribuição de Poisson – como discutido na Seção 17.3. Wooldridge (2010, Capítulo 18) discute outros exemplos úteis.

E.4 Algumas análises assimptóticas

A abordagem de matriz para o modelo de regressão múltipla também pode fazer derivações de propriedades assimptóticas mais concisas. De fato, fornecemos demonstrações gerais das alegações no Capítulo 11.

Começamos provando o resultado de consistência do Teorema 11.1. Lembre-se de que essas hipóteses contêm, como um caso especial, as hipóteses para análise de corte transversal sob amostragem aleatória.

Prova do Teorema 11.1. Como visto no Problema E.1 e usando a Hipótese TS.1′, escrevemos o estimador MQO como

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\mathbf{x}_t'y_t\right) = \left(\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\mathbf{x}_t'(\mathbf{x}_t\boldsymbol{\beta} + u_t)\right) \\ &= \boldsymbol{\beta} + \left(\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\mathbf{x}_t'u_t\right) \\ &= \boldsymbol{\beta} + \left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right).\end{aligned} \tag{E.19}$$

Agora, segundo a lei de grandes quantidades,

$$n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t \xrightarrow{p} \mathbf{A} \text{ e } n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'u_t \xrightarrow{p} \mathbf{0}, \tag{E.20}$$

em que $\mathbf{A} = \mathrm{E}(\mathbf{x}_t'\mathbf{x}_t)$ é uma $(k + 1) \times (k + 1)$ matriz não singular sob a Hipótese TS.2′, e usamos o fato de que $\mathrm{E}(\mathbf{x}_t'\mathbf{u}_t) = 0$ sob a Hipótese TS.3′. Agora, devemos usar a versão matricial de Propriedade PLIM.1 do Apêndice C. Especificamente, como A não é singular,

$$\left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}. \tag{E.21}$$

[Wooldridge (2010, Capítulo 3) apresenta uma discussão sobre esse tipo de resultado de convergência]. O resultado observado em (E.19), (E.20) e (E.21) é que

$$\text{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{A}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta}.$$

Isso completa a demonstração.

Na sequência, esboçamos uma demonstração do resultado da normalidade assimptótica do Teorema 11.2.

Prova do Teorema 11.2. Segundo a equação (E.19), podemos escrever

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right) \\ &= \mathbf{A}^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right) + o_p(1),\end{aligned} \tag{E.22}$$

em que o termo "$O_p(1)$" é um termo remanescente que, em probabilidade, converge a zero. Esse termo é igual a $[(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t)^{-1} - \mathbf{A}^{-1}](n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t)$. O termo entre parênteses converge em probabilidade a zero (pelo mesmo argumento usado na prova do Teorema 11.1), enquanto $(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t$ se limita em probabilidade porque converge a uma distribuição normal multivariada pelo teorema de limite central. Um resultado bem conhecido na teoria assimptótica é que o produto desses termos converge em probabilidade a zero. Além disso, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ herda sua distribuição assimptótica de $\mathbf{A}^{-1}(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t)$. Veja Wooldridge (2010, Capítulo 3) para mais detalhes sobre os resultados convergentes usados nessa demonstração.

Segundo o teorema de limite central, $n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t$ possui uma distribuição normal assimptótica com significado zero e, digamos, matriz $\mathbf{B}$ de variação-covariação $(k + 1) \times (k + 1)$. Logo, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ tem uma distribuição normal multivariada assimptótica com significado zero e matriz de variação-covariação $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$. Mostramos agora que, de acordo com as Hipóteses TS.4′ e TS.5′, $\mathbf{B} = \sigma^2\mathbf{A}$. (O termo geral é útil porque sustenta erros padrão robustos com relação à heteroscedasticidade e correlação serial para MQO, do tipo que discutimos no Capítulo 12). Em primeiro lugar, de acordo com a Hipótese TS.5′, $\mathbf{x}_t'u_t$ e $\mathbf{x}_s'u_s$ não são correlacionados para $t \neq s$. Por quê? Suponha $s < t$ para uma concretude. Então, pela lei das expectativas iteradas, $\mathrm{E}(\mathbf{x}_t'u_tu_s\mathbf{x}_s) = \mathrm{E}[\mathrm{E}(u_tu_s|\mathbf{x}_t'\mathbf{x}_s)\mathbf{x}_t'\mathbf{x}_s) = \mathrm{E}[\mathrm{E}(u_tu_s|\mathbf{x}_t'\mathbf{x}_s)\mathbf{x}_t'\mathbf{x}_s) = \mathrm{E}[0 \cdot \mathbf{x}_t'\mathbf{x}_s] = 0$. As covariações zero implicam que a variação da soma é a soma das variações. Contudo, $\mathrm{Var}(\mathbf{x}_t'\mathbf{x}_s) = \mathrm{E}(\mathbf{x}_t'u_tu_t\mathbf{x}_t) = \mathrm{E}(u_y^2\,\mathbf{x}_t'\mathbf{x}_t)$. Pela lei das expectativas iteradas, $\mathrm{E}(u_t^2\mathbf{x}_t'\mathbf{x}_t) = \mathrm{E}[\mathrm{E}(u_t^2\,\mathbf{x}_t'\mathbf{x}_t|\,\mathbf{x}_t')$ $= \mathrm{E}[\mathrm{E}(u_t^2\,|\,\mathbf{x}_t)\mathbf{x}_t'\mathbf{x}_t) = \mathrm{E}[\sigma^2\mathbf{x}_t'\mathbf{x}_t] = \sigma^2\mathrm{E}(\mathbf{x}_t'\mathbf{x}_t) = \sigma^2\mathbf{A}$, em que usamos $\mathrm{E}(u_t^2|\,\mathbf{x}_t) = \sigma^2$ de acordo com as Hipóteses TS.3′ e TS.4′. Isso mostra que $\mathbf{B} = \sigma^2\mathbf{A}$ e então, de acordo com as Hipóteses de TS.1′ a TS.5′, temos

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \mathrm{Normal}(\mathbf{0},\sigma^2\mathbf{A}^{-1}). \tag{E.23}$$

Isso completa a prova.

Segundo a equação (E.23), tratamos $\hat{\boldsymbol{\beta}}$ como se fosse aproximada e normalmente distribuído com significado $\boldsymbol{\beta}$ e matriz de variação-covariação $\sigma^2\mathrm{A}^{-1}/n$. Espera-se aqui a divisão pelo tamanho da amostra, n: a aproximação para a matriz de variação-covariação de $\hat{\boldsymbol{\beta}}$ diminui a zero à proporção $1/n$. Quando substituímos σ^2 por seu estimador consistente, $\hat{\sigma}^2 = \mathrm{SQR}/(n - k - 1)$, e substituímos A por seu estimador consistente, $n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t = \mathbf{X}'\mathbf{X}/n$, obtemos um estimador para a variação assimptótica de $\hat{\boldsymbol{\beta}}$:

$$\widehat{\mathbf{A}var(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{E.24}$$

Perceba como as duas divisões por n terminam e a parte direita de (E.24) é simplesmente a forma habitual com que estimamos a matriz de variação do estimador MQO sob as Hipóteses de Gauss-Markov. Para resumir, mostramos que, sob as Hipóteses de TS.1′ a TS.5′ – que contêm RLM.1 a RLM.5 como casos especiais –, os erros padrões habituais e estatísticas t são assimptoticamente válidos. É perfeitamente legítimo usar a distribuição t usual para obter valores críticos e p-valores para testar uma única hipótese. Curiosamente, na configuração geral do Capítulo 11, presumir a normalidade dos erros – digamos, u_t, visto que $\mathbf{x}_t$, u_{t-1}, $\mathbf{x}_{t-1}$, . . ., u_1, $\mathbf{x}_1$, é distribuído como Normal(0, σ^2) – não ajuda necessariamente, já que as estatísticas t não teriam geralmente estatísticas t exatas, de acordo com esse tipo de hipótese de normalidade. Quando não presumimos exogeneidade estrita das variáveis explicativas, os resultados distributivos exatos são difíceis, se não impossíveis, de obter.

Se modificarmos os argumentos acima, podemos derivar uma matriz de variação--covariação robusta com relação à heteroscedasticidade. A chave é estimar $\mathrm{E}(u_t^2\mathbf{x}_t'\mathbf{x}_t)$ separadamente, porque a matriz não é mais igual a $\sigma^2\mathrm{E}(\mathbf{x}_t'\mathbf{x}_t)$. Contudo, se $\hat{u}_t$ for os resíduos MQO, um estimador consistente seria

$$(n - k - 1)^{-1}\sum_{t=1}^{n}\hat{u}_t^2\mathbf{x}_t'\mathbf{x}_t, \tag{E.25}$$

em que a divisão por $n - k - 1$, em vez de n, representa graus de ajuste de liberdade que tipicamente ajudam as propriedades de amostras finitas do estimador. Quando usamos a expressão na equação (E.25), obtemos

$$\widehat{\mathbf{A}var(\hat{\boldsymbol{\beta}})} = [n/(n - k - 1)](\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{t=1}^{n}\hat{u}_t^2\mathbf{x}_t'\mathbf{x}_t\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{E.26}$$

As raízes quadradas dos elementos diagonais dessa matriz são os mesmos erros padrão robustos com relação à heteroscedasticidade obtidos na Seção 8.2, para o caso de puro corte transversal. Uma extensão de matriz dos erros padrão robustos com relação à correlação serial (e heteroscedasticidade) obtidos na Seção 12.5 também está disponível, mas a matriz que deve substituir (E.25) é complicada por causa da correlação serial. Veja, por exemplo, Hamilton (1994, Seção 10.5).

E.4 Estatística de Wald para o teste de múltiplas hipóteses

Argumentos similares podem ser usados para obter a distribuição assimptótica da **estatística Wald** para o teste de múltiplas hipóteses. Considere $\mathbf{R}$ uma matriz $q \times (k + 1)$, com $q \leq (k + 1)$. Presuma que as restrições q sobre o vetor de parâmetros $(k + 1) \times 1$, $\boldsymbol{\beta}$, possam ser expressas como H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, em que $\mathbf{r}$ é um vetor $q \times 1$ de constantes conhecidas. Levando em conta as Hipóteses de TS.1′ a TS.5′, pode ser mostrado que, sob H_0,

$$[\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})]'(\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}[\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})] \overset{a}{\sim} \chi_q^2, \tag{E.27}$$

em que $\mathbf{A} = (\mathbf{x}_t'\mathbf{x}_t)$, como nas demonstrações dos Teoremas 11.1 e 11.2. A intuição por trás da equação (E.25) é simples. Como $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ é aproximadamente distribuída como Normal($\mathbf{0}$, $\sigma^2\mathbf{A}^{-1}$), $\mathbf{R}[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] = \sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ é aproximadamente Normal($0,\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}'$) pela Propriedade 3 da distribuição normal multivariada do Apêndice D. Sob H_0, $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, logo, $\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \overset{a}{\sim}$ Normal ($\mathbf{0}$, $\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}'$) sob H_0. Segundo a Propriedade 3 da distribuição qui-quadrada, $z'(\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}z \sim \chi_q^2$ se $z \sim$ Normal ($\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}'$). Para obter formalmente o resultado final, precisamos usar uma versão assimptótica dessa propriedade, que pode ser encontrada em Wooldridge (2010, Capítulo 3).

Visto o resultado em (E.25), obtemos uma estatística computável ao substituir $\mathbf{A}$ e σ^2 por seus estimadores consistentes; fazer isso não muda a distribuição assimptótica. O resultado é a chamada estatística Wald, que, depois de anular os tamanhos de amostras e fazer uma pequena álgebra, pode ser descrita como

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})/\hat{\sigma}^2. \tag{E.28}$$

De acordo com H_0, $W \overset{a}{\sim} \chi^2_q$, em que lembramos que q é o número de restrições testadas. Se $\sigma^2 = \mathrm{SQR}/(n - k - 1)$, pode ser demonstrado que W/q é exatamente a estatística F que obtemos no Capítulo 4 para o teste de restrições lineares múltiplas. [Ver, por exemplo, Greene (1997, Capítulo 7).] Portanto, de acordo com as hipóteses de modelo linear clássico de TS.1 a TS.6, do Capítulo 10, W/q apresentou uma exata distribuição $F_{q,\,n-k-1}$. Sob as Hipóteses de TS.1′ a TS.5′, temos somente o resultado assimptótico em (E.26). Entretanto, é apropriado e comum tratar a estatística F usual como tendo uma distribuição $F_{q,\,n-k-1}$ aproximada.

Uma estatística Wald robusta com relação à heteroscedasticidade, de forma desconhecida, é obtida usando uma matriz em (E.26), no lugar de $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, e, de forma similar, para uma estatística de teste robusta com relação à heteroscedasticidade e também à correlação serial. As versões robustas de estatísticas de teste não podem ser calculadas por meio da soma de resíduos quadrados ou R-quadrados de regressões restritas e irrestritas.

Resumo

Este Apêndice forneceu uma breve abordagem do modelo de regressão linear usando notação de matriz. Este material foi incluso para classes mais avançadas, que utilizam álgebra matricial; contudo, não é necessário ler o texto. De fato, este apêndice prova alguns dos resultados que ambos estabelecemos sem demonstração, provando apenas em casos especiais ou através de um método ou demonstração mais difícil. Outros tópicos – como propriedades assimptóticas, estimação de variáveis instrumentais e modelos de dados em painel – podem receber tratamentos concisos usando matrizes. Para mais detalhes, consulte os textos avançados em econometria de Davidson e MacKinnon (1993), Greene (1997), Hayashi (2000) e Wooldridge (2010).

Termos-chave

Condição de primeira ordem
Estatística Wald
Estimador não viesado de variância mínima
Estimador por quase máxima verossimilhança (EQMV)
Matriz de variação-covariação do estimador MQO
Matriz de variação-covariação escalar
Notação de matriz
Teorema de Frisch-Waugh (FW)

Problemas

1 Considere $\mathbf{x}_t$ o vetor de variáveis explicativas $1 \times (k + 1)$ para a observação t. Mostre que o estimador MQO $\hat{\boldsymbol{\beta}}$ pode ser escrito como

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{x}_t' y_t\right).$$

Dividir cada somatória por n mostra que $\hat{\boldsymbol{\beta}}$ é uma função de médias amostrais.

2. Considere $\hat{\boldsymbol{\beta}}$ o vetor de estimativas MQO $(k + 1) \times 1$.

(i) Demonstre que, para cada vetor $\mathbf{b}$ $(k + 1) \times 1$, podemos escrever a soma de resíduos quadrados como

$$\text{SQR}(\mathbf{b}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b}).$$

{Dica: Escreva $(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = [\hat{\mathbf{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})]'[\hat{\mathbf{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})]$ e use o fato de que $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$.}

(ii) Explique como a expressão para SQR($\mathbf{b}$) da parte (i) prova que $\hat{\boldsymbol{\beta}}$ minimiza unicamente SQR($\mathbf{b}$) sobre todos os valores possíveis de $\mathbf{b}$, presumindo que $\mathbf{X}$ classificou $k + 1$.

3. Considere $\hat{\boldsymbol{\beta}}$ o estimador MQO da regressão de $\mathbf{y}$ em $\mathbf{X}$. Suponha que $\mathbf{A}$ seja uma matriz não singular $(k + 1) \times (k + 1)$ e defina $\mathbf{z}_t \equiv \mathbf{x}_t\mathbf{A}$, $t = 1, \ldots, n$. Portanto, $\mathbf{z}_t$ é $1 \times (k + 1)$ e é uma combinação não singular linear de $\mathbf{x}_t$. Considere $\mathbf{Z}$ a matriz $n \times (k + 1)$ com linhas $\mathbf{z}_t$. Considere que $\hat{\boldsymbol{\beta}}$ denote a estimativa MQO para regressão de $\mathbf{y}$ em $\mathbf{Z}$.

(i) Mostre que $\tilde{\boldsymbol{\beta}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

(ii) Considere $\hat{y}_t$ os valores ajustados da regressão original e $\hat{y}_t$ os valores ajustados da regressão de $\mathbf{y}$ em $\mathbf{Z}$. Mostre que $\tilde{y}_t = \hat{y}_t$, para todos $t = 1, 2, \ldots, n$. Como se comparam os resíduos das duas regressões?

(iii) Mostre que a matriz de variação estimada para $\hat{\boldsymbol{\beta}}$ é $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}^{-1\prime}$, em que $\hat{\sigma}^2$ é a estimativa de variação usual da regressão de $\mathbf{y}$ em $\mathbf{X}$.

(iv) Considere $\hat{\beta}_j$ as estimativas MQO da regressão de y_t em $x_{t1}, \ldots, x_{tk}$, e que $\hat{\beta}_j$ seja as estimativas MQO da regressão de y_t em $1, a_1x_{t1}, \ldots, a_kx_{tk}$, em que $a_i \neq 0, j = 1, \ldots, k$. Use os resultados do item (i) para encontrar a relação entre $\tilde{\beta}_j$ e $\hat{\beta}_j$.

(v) Presuma a organização do item (iv), e use o item (iii) para mostrar que $(\hat{\beta}_j) = \text{ep}(\hat{\beta}_j)/|a_j|$.

(vi) Presuma a organização do item (iv), e mostre que os valores absolutos das estatísticas t para $\tilde{\beta}_j$ e $\hat{\beta}_j$ são idênticos.

4. Presuma que o modelo $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ satisfaça as hipóteses de Gauss-Markov, considere $\mathbf{G}$ uma matriz não singular e não aleatória $(k + 1) \times (k + 1)$, e defina $\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\beta}$, de forma que $\boldsymbol{\delta}$ seja também um vetor $(k + 1) \times 1$. Considere $\hat{\boldsymbol{\beta}}$ o vetor de estimadores MQO $(k + 1) \times 1$ e defina $\hat{\boldsymbol{\delta}} = \mathbf{G}\hat{\boldsymbol{\beta}}$ como estimador MQO de $\boldsymbol{\delta}$.

(i) Mostre que $\text{E}(\hat{\boldsymbol{\delta}}|\mathbf{X}) = \boldsymbol{\delta}$.

(ii) Encontre $\text{Var}(\hat{\boldsymbol{\delta}}|\mathbf{X})$ em termos de σ^2, $\mathbf{X}$ e $\mathbf{G}$.

(iii) Use o problema E.3 para provar que $\hat{\boldsymbol{\delta}}$ e a estimativa adequada de $\text{Var}(\hat{\boldsymbol{\delta}}|\mathbf{X})$ são obtidos da regressão de $\mathbf{y}$ em $\mathbf{XG}^{-1}$.

(iv) Agora, considere $\mathbf{c}$ um vetor $(k + 1) \times 1$ com ao menos uma entrada com valor diferente de zero. Para concretude, presuma que $c_k \neq 0$. Defina $\theta = \mathbf{c}'\boldsymbol{\beta}$, de forma que θ seja escalar. Defina $\delta_j = \boldsymbol{\beta}_j, j = 0, 1, \ldots, k - 1$ e $\delta_k = \theta$. Mostre como definir uma matriz $\mathbf{G}$ não singular $(k + 1) \times (k + 1)$, de forma que $\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\beta}$ (*Dica*: Cada uma das primeiras fileiras k de $\mathbf{G}$ contêm zero k e um zero. Qual é a última fileira?)

(v) Mostre que, para a escolha de $\mathbf{G}$ na parte (iv),

$$\mathbf{G}^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & 0 & \cdot & \cdot & \cdot & 1 & 0 \\ -c_0/c_k & -c_1/c_k & \cdot & \cdot & \cdot & -c_{k-1}/c_k & 1/c_k \end{bmatrix}$$

Use essa expressão para $\mathbf{G}^{-1}$ e o item (iii) para concluir que x_{tk}/c_k e seu erro padrão são obtidos como coeficiente em x_{tk}/c_k, na regressão de

y_t on $[1 - (c_0/c_k)x_{tk}], [x_{t1} - (c_1/c_k)x_{tk}], \ldots, [x_{t,k-1} - (c_{k-1}/c_k)x_{tk}], x_{tk}/c_k, t = 1, \ldots, n.$

Essa regressão é a exatamente a obtida quando escrevemos β_k em termos de θ e β_0, $\beta_1, \ldots, \beta_{k-1}$, associando o resultado ao modelo original e rearranjando-o. Portanto, podemos justificar formalmente o truque que usamos no texto para obter o erro padrão de uma combinação linear de parâmetros.

5. Suponha que o modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ satisfaça as hipóteses de Gauss-Markov e considere $\hat{\boldsymbol{\beta}}$ o estimador MQO de $\boldsymbol{\beta}$. Suponha que $\mathbf{Z} = \mathbf{G}(\mathbf{X})$ seja uma função de matriz de $\mathbf{X}$ $n \times (k + 1)$ e presuma que a matriz $\mathbf{Z}'\mathbf{X}$[a$(k + 1) \times (k + 1)$] seja não singular. Defina um novo estimador de $\boldsymbol{\beta}$ por $\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$.

(i) Mostre que $\mathrm{E}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$, de forma que $\tilde{\boldsymbol{\beta}}$ também seja um condicional não viesado em $\mathbf{X}$.

(ii) Encontre $\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X})$. Certifique-se de que é uma matriz $(k + 1) \times (k + 1)$ simétrica que depende de $\mathbf{Z}$, $\mathbf{X}$ e σ^2.

(iii) Que estimador você prefere, $\hat{\boldsymbol{\beta}}$ ou $\tilde{\boldsymbol{\beta}}$? Explique.

6. Considere a organização do Teorema de Frisch-Waugh.

(i) Usando matrizes separadas, mostre que as condições de primeira ordem $(\mathbf{X}'\mathbf{X})\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ podem ser escritas como

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_1'\mathbf{y}$$
$$\mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{y}.$$

(ii) Multiplique o primeiro conjunto de equações por $\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_2'\mathbf{X}_1)^{-1}$ e subtraia o resultado do segundo conjunto de equações para mostrar que

$$(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{M}_1\mathbf{y},$$

em que $\mathbf{I}_{n\,-}\,\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$. Determine que

$$\hat{\boldsymbol{\beta}}_2 = (\ddot{\mathbf{X}}_2'\ddot{\mathbf{X}}_2)^{-1}\ddot{\mathbf{X}}_2'\mathbf{y}.$$

(iii) Use a parte (ii) para mostrar que

$$\hat{\boldsymbol{\beta}}_2 = (\ddot{\mathbf{X}}_2'\ddot{\mathbf{X}}_2)^{-1}\ddot{\mathbf{X}}_2'\,\ddot{y}.$$

(iv) Use o fato de que $\mathbf{M}_1\mathbf{X}_1 = 0$ para mostrar que os resíduos $\ddot{\mathbf{u}}$ da regressão de $\ddot{\mathbf{y}}$ em $\ddot{\mathbf{X}}_2$ são idênticos aos resíduos $\hat{u}$ da regressão $\mathbf{y}$ em $\mathbf{X}_1$, $\mathbf{X}_2$. [*Dica*: Por definição e pelo teorema FW,

$$\ddot{\mathbf{u}} = \ddot{\mathbf{y}} - \ddot{\mathbf{X}}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{M}_1(\mathbf{y} - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2) = \mathbf{M}_1(\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2).$$

Agora faça o resto.]

7. Considere que o modelo linear, escrito em notação de matriz,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

satisfaça as Hipóteses E.1, E.2 e E.3. Divida o modelo como

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u},$$

em que $\mathbf{X}_1$ é $n \times (k_1 + 1)$ e $\mathbf{X}_2$ é $n \times k_2$.

(i) Considere a seguinte proposta de estimação de $\boldsymbol{\beta}_2$. Primeiro regrida $\mathbf{y}$ em $\mathbf{X}_1$ e obtenha os resíduos, digamos, $\ddot{\mathbf{y}}$. Depois, regrida $\ddot{\mathbf{y}}$ em $\mathbf{X}_2$ para obter $\breve{\boldsymbol{\beta}}_2$. Mostre que $\breve{\boldsymbol{\beta}}_2$ é geralmente viesado e mostre qual é o viés. [Você deve encontrar $\mathrm{E}(\breve{\boldsymbol{\beta}}_2|\mathbf{X})$ em termos de $\boldsymbol{\beta}_2$, $\mathbf{X}_2$ e a matriz $\mathbf{M}_1$ de realização residual].

(ii) Como um caso especial, escreva

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \beta_k\mathbf{X}_k + \mathbf{u},$$

em que $\mathbf{X}_k$ é um vetor $n \times 1$ na variável x_{tk}. Mostre que

$$\mathrm{E}(\tilde{\beta}_k|\mathbf{X}) = \left(\frac{\mathrm{SQR}_k}{\sum_{t=1}^{n} x_{tk}^2}\right)\beta_k,$$

em que SQRk é a soma dos resíduos quadrados da regressão de x_{tk} em 1, x_{t1}, x_{t2}, . . ., $x_{t,\,k-1}$. Como o fator multiplicador de β_k nunca é maior que um?

(iii) Suponha que você conheça $\boldsymbol{\beta}_1$. Mostre que a regressão $\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1$ em $\mathbf{X}_1$ produz um estimador não viesado de $\boldsymbol{\beta}_2$ (condicional em $\mathbf{X}$).