Teste de Hipótese em Regressão Normal Linear Múltipla

Caio César Barros de Araújo

Universidade Federal de Pernambuco

18 de novembro de 2025

1 Modelo e Fundamentos Teóricos

A regressão linear múltipla modela a relação entre uma variável resposta Y e múltiplas variáveis explicativas $X_1, X_2, \ldots, X_p$. Os testes de hipótese permitem avaliar a significância estatística dos parâmetros e a relevância das variáveis explicativas. Este trabalho apresenta o teste de hipótese $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$ no modelo de regressão normal linear múltipla, analisando sua fundamentação teórica.

1.1 Especificação do Modelo

O modelo de regressão linear múltipla é especificado como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

onde $\mu_i(\boldsymbol{\beta}) = \mathbf{x}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$, $\mathbf{y}^T = (y_1, \ldots, y_n)$ é o vetor de variáveis resposta, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ são parâmetros desconhecidos, $\mathbf{X}$ é a matriz modelo $n \times (p+1)$ de planejamento com primeira coluna de uns, e $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \ldots, \varepsilon_n)$ é o vetor de erros aleatórios.

1.2 Pressupostos Clássicos

Definição 1.1 (Pressupostos do Modelo Normal Linear)**.**

1. **Linearidade**: $\mu_i(\boldsymbol{\beta}) = \mathbf{x}_i^T\boldsymbol{\beta}$.
2. **Normalidade**: $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$, independentes.
3. **Homocedasticidade**: $\mathrm{Var}(Y_i \mid \mathbf{x}_i) = \sigma^2$ constante.
4. **Independência**: $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $\forall i \neq j$.
5. **Não-colinearidade**: $\mathbf{X}^T\mathbf{X}$ é inversível.

Sob esses pressupostos, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$. A violação da não-colinearidade (multicolinearidade), mesmo mantendo $\mathbf{X}^T\mathbf{X}$ inversível, aumenta drasticamente a variância dos estimadores $\hat{\boldsymbol{\beta}}$, alargando intervalos de confiança e reduzindo o poder dos testes t individuais [2, Seção 3.10].

1.3 Estimadores de Mínimos Quadrados

O estimador de mínimos quadrados (E.M.Q.) para $\boldsymbol{\beta}$ é definido como o minimizador da soma de quadrados dos resíduos:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} S(\boldsymbol{\beta}) \tag{2}$$

onde

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \tag{3}$$

Calculando a derivada de primeira ordem e igualando a zero, obtemos as **equações normais**:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \tag{4}$$

Assumindo que $\mathbf{X}^T\mathbf{X}$ é inversível (condição de não-colinearidade), o estimador de mínimos quadrados é:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{5}$$

Como a segunda derivada $\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = 2\mathbf{X}^T\mathbf{X}$ é positiva definida, $\hat{\boldsymbol{\beta}}$ é o minimizador global de $S(\boldsymbol{\beta})$.

Teorema 1.1 (Propriedades do Estimador). *Sob os pressupostos clássicos:*

(i) $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ *(não-viesado);*

(ii) $\hat{\boldsymbol{\beta}}$ *é o melhor estimador linear não-viesado (Teorema de Gauss-Markov);*

(iii) $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$.

Demonstração. A propriedade (i) segue diretamente da linearidade da esperança. Para (ii), pelo Teorema de Gauss-Markov, qualquer estimador linear não-viesado $\tilde{\boldsymbol{\beta}}$ possui $\text{Var}(\tilde{\boldsymbol{\beta}}) = \text{Var}(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{C}\mathbf{C}^T$, onde $\mathbf{C}\mathbf{C}^T$ é semidefinida positiva, garantindo $\text{Var}(\tilde{\boldsymbol{\beta}}) \geq \text{Var}(\hat{\boldsymbol{\beta}})$ no sentido matricial [1, Teorema 11.2.1]. A propriedade (iii) segue da normalidade de $\mathbf{y}$ e da linearidade de $\hat{\boldsymbol{\beta}}$. □

O estimador não-viesado de σ^2 é dado por:

$$\hat{\sigma}^2 = \frac{SSE}{n-p-1}, \quad \text{onde} \quad SSE = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{6}$$

Proposição 1.1 (Distribuição de $\hat{\sigma}^2$ e Independência). *Sob os pressupostos do modelo,* $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p-1}$ *e é independente de* $\hat{\boldsymbol{\beta}}$.

Demonstração. Seja $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ a matriz de projeção ortogonal no espaço coluna de $\mathbf{X}$. Note que $\mathbf{P}$ é simétrica e idempotente, com $\text{tr}(\mathbf{P}) = p + 1$. Para aplicar o Teorema de Cochran, consideramos $\mathbf{Q}_1 = \mathbf{P}$ e $\mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}$. Como $\mathbf{P}$ é idempotente, $\mathbf{I}_n - \mathbf{P}$ também é idempotente e $\mathbf{P}(\mathbf{I}_n - \mathbf{P}) = \mathbf{0}$, garantindo ortogonalidade. Além disso, $\mathbf{Q}_1 + \mathbf{Q}_2 = \mathbf{I}_n$ e $\text{tr}(\mathbf{Q}_1) + \text{tr}(\mathbf{Q}_2) = n$.

Como $SSE = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P})\mathbf{y}$ e $\mathbf{I}_n - \mathbf{P}$ é uma matriz de projeção ortogonal com $\text{tr}(\mathbf{I}_n - \mathbf{P}) = n - p - 1$, segue do Teorema de Cochran que $\frac{SSE}{\sigma^2} \sim \chi^2_{n-p-1}$.

A independência entre $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ decorre da ortogonalidade: como $\hat{\boldsymbol{\beta}}$ depende linearmente de $\mathbf{P}\mathbf{y}$ e SSE depende de $(\mathbf{I}_n - \mathbf{P})\mathbf{y}$, e $\mathbf{P}(\mathbf{I}_n - \mathbf{P}) = \mathbf{0}$ implica que $\mathbf{P}\mathbf{y}$ e $(\mathbf{I}_n - \mathbf{P})\mathbf{y}$ são vetores normais independentes, segue que $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ são independentes [3, Teorema 3.5(iii)]. □

1.4 Teorema de Cochran e Distribuições Qui-Quadrado

O resultado fundamental sobre distribuições qui-quadrado de formas quadráticas é dado pelo Teorema de Cochran, que fundamenta teoricamente a decomposição de somas de quadrados.

Teorema 1.2 (Teorema de Cochran). *Seja* $\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ *e sejam* $\mathbf{Q}_1, \ldots, \mathbf{Q}_k$ *matrizes simétricas idempotentes tais que* $\sum_{i=1}^k \mathbf{Q}_i = \mathbf{I}_n$ *e* $\sum_{i=1}^k \text{tr}(\mathbf{Q}_i) = n$. *Se* $\mathbf{Q}_i\boldsymbol{\mu} = \mathbf{0}$ *para* $i = 1, \ldots, k$, *então as formas quadráticas* $\mathbf{y}^T\mathbf{Q}_i\mathbf{y}$, $i = 1, \ldots, k$, *são independentes e* $\frac{\mathbf{y}^T\mathbf{Q}_i\mathbf{y}}{\sigma^2} \sim \chi^2_{\nu_i}$, *onde* $\nu_i = \text{tr}(\mathbf{Q}_i)$.

O Teorema de Cochran estabelece que sob condições adequadas, somas de quadrados podem ser decompostas em componentes qui-quadrado independentes, sendo fundamental para a análise de variância e inferência em modelos lineares. A fundamentação teórica deste resultado (Teorema 2.7 e Exemplo 2.13) e sua aplicação prática para testes de hipótese e análise de variância são detalhadas em Seber e Lee [3, Capítulos 2.4, 4 e 8].

2 Teste de Hipótese $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$

2.1 Formulação do Teste

Particionando $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T)^T$ onde $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_p)^T$, o teste avalia se as variáveis explicativas têm efeito significativo sobre Y:

$$H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}_p \tag{7}$$

onde $\mathbf{0}_p$ é o vetor nulo de dimensão p. Sob H_0, o modelo reduz-se a $Y_i = \beta_0 + \varepsilon_i$.

2.2 Estatística F e Distribuição

A estatística de teste F é definida como:

$$F = \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n-p-1)} \tag{8}$$

onde $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ é a Soma dos Quadrados da Regressão e $MSR = SSR/p$, $MSE = SSE/(n-p-1)$ são os quadrados médios.

Teorema 2.1 (Distribuição da Estatística F)**.** *Sob $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$ e os pressupostos do modelo:*

$$F \sim F_{p,n-p-1} \tag{9}$$

O teste F pode ser interpretado como uma transformação monotônica da estatística associada ao Teste da Razão de Verossimilhança (Λ). Especificamente, para o modelo normal linear, a estatística LRT $\Lambda = \left(\frac{SSE}{SSE_0}\right)^{n/2}$ está relacionada ao teste F através de uma transformação monotônica. Esta equivalência reforça que o teste F não é apenas uma conveniência baseada na decomposição de soma de quadrados, mas sim o **UMP** sob a suposição de normalidade, em virtude da optimalidade do LRT nesse cenário. [1, Teorema 10.1.1].

2.3 Derivação via Decomposição de Soma de Quadrados e Teorema de Cochran

A escolha da estatística F para o teste $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$ decorre da estrutura distribucional do modelo normal linear. Sob H_0, o modelo reduz-se a $Y_i = \beta_0 + \varepsilon_i$ com E.M.Q. $\hat{\beta}_0 = \bar{y}$. A soma de quadrados do modelo reduzido é $SSE_0 = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_0)\mathbf{y}$, onde $\mathbf{P}_0 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ projeta no espaço gerado por $\mathbf{1}_n$, contido no espaço coluna de $\mathbf{X}$.

A diferença $SSR = SSE_0 - SSE$ representa a redução na soma de quadrados devido à inclusão das p variáveis explicativas. Em termos de matrizes de projeção: $SSR = \mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}$ e $SSE = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P})\mathbf{y}$, onde $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Para aplicar o Teorema de Cochran, consideramos $\mathbf{Q}_1 = \mathbf{P} - \mathbf{P}_0$, $\mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}$ e $\mathbf{P}_0$. Verificamos explicitamente as condições: (i) $\mathbf{Q}_1$ é idempotente, pois como $\mathbf{P}_0$ projeta em subespaço de $\mathbf{P}$, temos $\mathbf{P}\mathbf{P}_0 = \mathbf{P}_0 = \mathbf{P}_0\mathbf{P}$, logo $(\mathbf{P} - \mathbf{P}_0)^2 = \mathbf{P} - \mathbf{P}_0$; (ii) $\mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}$ é idempotente; (iii) $\mathbf{Q}_1\mathbf{Q}_2 = (\mathbf{P} - \mathbf{P}_0)(\mathbf{I}_n - \mathbf{P}) = \mathbf{0}$, confirmando ortogonalidade; (iv) $\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{P}_0 = \mathbf{I}_n$ e $\mathrm{tr}(\mathbf{Q}_1) + \mathrm{tr}(\mathbf{Q}_2) + \mathrm{tr}(\mathbf{P}_0) = p + (n-p-1) + 1 = n$; (v) Sob H_0, quando $\boldsymbol{\beta}_1 = \mathbf{0}_p$, temos $\mathbf{X}\boldsymbol{\beta} = \beta_0\mathbf{1}_n$ e $\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} = (\mathbf{P} - \mathbf{P}_0)\mathbf{1}_n = \mathbf{0}$.

Portanto, pelo Teorema de Cochran, sob H_0:

$$\frac{SSR}{\sigma^2} = \frac{\mathbf{y}^T\mathbf{Q}_1\mathbf{y}}{\sigma^2} \sim \chi^2_p, \quad \frac{SSE}{\sigma^2} = \frac{\mathbf{y}^T\mathbf{Q}_2\mathbf{y}}{\sigma^2} \sim \chi^2_{n-p-1} \tag{10}$$

e essas quantidades são independentes, pois $\mathbf{Q}_1\mathbf{Q}_2 = \mathbf{0}$ garante independência via ortogonalidade das projeções.

Como a estatística F é a razão entre duas variáveis χ^2 independentes divididas por seus graus de liberdade, ela elimina o parâmetro de escala desconhecido σ^2, tornando-se uma **estatística pivotal**:

$$F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{(SSR/\sigma^2)/p}{(SSE/\sigma^2)/(n-p-1)} \sim F_{p,n-p-1} \tag{11}$$

Portanto, F possui distribuição completamente especificada sob H_0 e não depende de parâmetros desconhecidos. Valores grandes de F_{obs} indicam rejeição de H_0, pois a variância explicada (MSR) é significativamente maior que a variância residual (MSE).

2.4 Região de Rejeição e Testes Complementares

Para nível de significância α, rejeitamos H_0 se $F > F_{p,n-p-1;\alpha}$ ou se p-valor $= P(F_{p,n-p-1} > F_{\text{obs}}) < \alpha$.

Para testes individuais $H_0 : \beta_j = 0$, utiliza-se $t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim t_{n-p-1}$, que complementa o teste F global para identificar quais variáveis específicas são significativas. Note que o teste F global avalia $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$ (todas as variáveis simultaneamente), enquanto o teste t para $H_0 : \beta_j = 0$ avalia a contribuição de X_j **dado que as outras variáveis já estão no modelo**. Os coeficientes $\hat{\beta}_j$ são, portanto, **coeficientes parciais** que medem o efeito de X_j controlando pelas demais variáveis explicativas. A multicolinearidade pode resultar em rejeição de H_0 pelo teste F global enquanto os testes t individuais falham em identificar variáveis significativas.

3 Análise de Variância e Conclusão

3.1 Decomposição ANOVA

A decomposição fundamental da variabilidade total é $SST = SSR + SSE$, onde $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ é a Soma Total dos Quadrados. A tabela ANOVA resume essa decomposição:

Tabela 1: Tabela ANOVA para Regressão Linear Múltipla

Fonte	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F
Regressão	SSR	p	$MSR = SSR/p$	$F = MSR/MSE$
Erro	SSE	$n-p-1$	$MSE = SSE/(n-p-1)$	
Total	SST	$n-1$		

3.2 Interpretação e Considerações Finais

O teste $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}_p$ permite avaliar a significância global das variáveis explicativas no modelo. Rejeitar H_0 significa que pelo menos uma das variáveis explicativas contribui de forma estatisticamente significativa para explicar a variabilidade de Y. O teste F global deve ser complementado por testes t individuais para identificar quais variáveis específicas são responsáveis pela significância.

É relevante salientar que a rejeição da hipótese nula H_0, indicando significância estatística, não implica que o modelo seja adequado para previsão ou descrição. A avaliação definitiva da adequação do modelo exige a verificação dos pressupostos estatísticos — notadamente a normalidade dos resíduos e a homocedasticidade — por meio de análise residual. Tal procedimento é fundamental para determinar se o modelo apresenta bom ajuste aos dados observados e se as condições subjacentes à aplicação do teste F foram devidamente satisfeitas [2, Capítulos 4 e 5].

Referências

[1] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, CA, 2nd edition, 2002.

[2] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. Wiley, Hoboken, NJ, 5th edition, 2012.

[3] George A. F. Seber and Alan J. Lee. *Linear Regression Analysis*. Wiley, Hoboken, NJ, 2nd edition, 2012.