

Teste de Hipótese em Regressão Normal Linear
Múltipla
Análise do Teste $H_0 : \beta_1 = \mathbf{0}_p$

Curso de Inferência Estatística - PPGEST/UFPE

17 de novembro de 2025

1 Introdução e Especificação do Modelo

A regressão linear múltipla modela a relação entre uma variável resposta Y e múltiplas variáveis explicativas X_1, X_2, \dots, X_p . Os testes de hipótese permitem avaliar a significância estatística dos parâmetros e a relevância das variáveis explicativas.

1.1 Especificação do Modelo

O modelo de regressão linear múltipla é:

$$Y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

onde $\mu_i \triangleq \mu_i(\beta) = E(Y_i | X_i = x_i)$ e

$$\mu_i(\beta) = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (2)$$

com $x_i^T = [1, x_{i1}, \dots, x_{ip}]$, Y_i é variável aleatória, x_{i1}, \dots, x_{ip} são variáveis explicativas fixas, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ são parâmetros desconhecidos e ε_i é o erro aleatório.

Em notação matricial:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (3)$$

onde $\mathbf{y}^T = (y_1, \dots, y_n)$ é o vetor de variáveis resposta, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$, $\varepsilon^T = (\varepsilon_1, \dots, \varepsilon_n)$ é o vetor de erros e \mathbf{X} é a matriz modelo $n \times (p+1)$ de planejamento com primeira coluna de uns.

1.2 Pressupostos

Definição 1.1 (Pressupostos Clássicos). 1. **Normalidade**: $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$, independentes.

2. **Homocedasticidade**: $\text{Var}(Y_i | x_i) = \sigma^2$ constante.

3. **Independência**: $\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 I_0(i - j)$, $\forall i, j = 1, \dots, n$, onde $I_0(i - j) = 1$ se $i = j$ e 0 caso contrário.

4. **Não-colinearidade**: $\mathbf{X}^T \mathbf{X}$ é inversível.

5. **Linearidade**: $\mu_i(\beta) = x_i^T \beta$.

Sob esses pressupostos, $Y_i | X_i \sim N(\mu_i, \sigma^2)$ e $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$.

2 Fundamentação Teórica e Teste de Hipótese

2.1 Estimadores de Mínimos Quadrados

O estimador de mínimos quadrados (E.M.Q.) para β minimiza:

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \quad (4)$$

Diferenciando $S(\beta)$ em relação a β e igualando a zero, obtemos:

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\beta = 0 \quad (5)$$

Assumindo que $\mathbf{X}^T \mathbf{X}$ é inversível, o estimador de mínimos quadrados é:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

Teorema 2.1 (Propriedades). Sob os pressupostos clássicos: (i) $E[\hat{\beta}] = \beta$ (não-viesado); (ii) Eficiente (Gauss-Markov); (iii) $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$.

Particionando $\beta = (\beta_0, \beta_1^T)^T$ onde $\beta_1 = (\beta_1, \dots, \beta_p)^T$, temos:

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2(\mathbf{X}_1^T \mathbf{M}_1 \mathbf{X}_1)^{-1}) \quad (7)$$

O estimador não-viesado de σ^2 é:

$$\hat{\sigma}^2 = \frac{SSE}{n-p-1} = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})}{n-p-1} \quad (8)$$

onde SSE é a Soma dos Quadrados dos Erros. Sob os pressupostos, $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p-1}^2$ e é independente de $\hat{\beta}$.

2.2 Teste de Hipótese $H_0 : \beta_1 = \mathbf{0}_p$

Testamos se as variáveis explicativas têm efeito significativo sobre Y :

$$H_0 : \beta_1 = \mathbf{0}_p \quad \text{versus} \quad H_1 : \beta_1 \neq \mathbf{0}_p \quad (9)$$

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

A estatística F é:

$$F = \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n-p-1)} \quad (10)$$

onde $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ é a Soma dos Quadrados da Regressão, $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})$, $MSR = SSR/p$ e $MSE = SSE/(n-p-1)$.

Teorema 2.2 (Distribuição da Estatística F). Sob $H_0 : \beta_1 = \mathbf{0}_p$ e os pressupostos do modelo:

$$F \sim F_{p,n-p-1} \quad (11)$$

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

Para testes individuais $H_0 : \beta_j = 0$, utilizamos:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}} \sim t_{n-p-1} \quad (12)$$

Rejeitamos $H_0 : \beta_j = 0$ se $|t_j| > t_{n-p-1;\alpha/2}$.

2.3 Derivação da Estatística F

A estatística F pode ser derivada através da razão de verossimilhanças ou via decomposição da soma de quadrados. Sob $H_0 : \beta_1 = \mathbf{0}_p$, o modelo reduzido é $Y_i = \beta_0 + \varepsilon_i$, cujo estimador de mínimos quadrados é $\hat{\beta}_0 = \bar{y}$.

A soma de quadrados do modelo completo é:

$$SSE = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y} \quad (13)$$

A soma de quadrados do modelo reduzido (sob H_0) é:

$$SSE_0 = \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}^T\mathbf{y} - n\bar{y}^2 \quad (14)$$

A diferença $SSE_0 - SSE = SSR$ representa a redução na soma de quadrados devido à inclusão das variáveis explicativas. Sob H_0 , temos:

$$\frac{SSR}{\sigma^2} \sim \chi_p^2, \quad \frac{SSE}{\sigma^2} \sim \chi_{n-p-1}^2 \quad (15)$$

e essas quantidades são independentes. Portanto, a razão:

$$F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{(SSR/\sigma^2)/p}{(SSE/\sigma^2)/(n-p-1)} \sim F_{p,n-p-1} \quad (16)$$

segue uma distribuição F com p e $n - p - 1$ graus de liberdade.

2.4 Propriedades e Interpretação

A estatística F mede a razão entre a variância explicada pelo modelo e a variância residual. Valores grandes de F indicam que o modelo explica uma proporção significativa da variabilidade dos dados.

O quadrado médio da regressão $MSR = SSR/p$ representa a variância média explicada por cada variável explicativa, enquanto $MSE = SSE/(n - p - 1)$ é a variância residual média. A razão $F = MSR/MSE$ compara essas duas fontes de variabilidade.

Quando F é grande (e estatisticamente significativo), concluímos que pelo menos uma das variáveis explicativas contribui significativamente para explicar a variabilidade de Y . O teste F global é complementado pelos testes t individuais, que identificam quais variáveis específicas são significativas.

3 Aplicações e Considerações Finais

3.1 Exemplo Numérico

Considere um modelo com $n = 30$ observações e $p = 3$ variáveis explicativas. Após ajustar o modelo:

$$SSR = 450.2, \quad SSE = 89.5, \quad (17)$$

$$MSR = 150.07, \quad MSE = 3.44, \quad (18)$$

$$F_{\text{obs}} = \frac{150.07}{3.44} = 43.61 \quad (19)$$

Como $F_{\text{obs}} = 43.61 > F_{3,26;0.05} \approx 2.98$, rejeitamos $H_0 : \beta_1 = \mathbf{0}_3$ ao nível de 5%, concluindo que pelo menos uma variável explicativa tem efeito significativo.

3.2 Relação com Análise de Variância (ANOVA)

A decomposição fundamental 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:

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	$n - 1$		

Tabela 1: Tabela ANOVA para Regressão Linear Múltipla

O coeficiente de determinação $R^2 = SSR/SST$ mede a proporção da variabilidade explicada pelo modelo.

3.3 Considerações Finais

O teste $H_0 : \beta_1 = \mathbf{0}_p$ permite: (i) seleção de variáveis significativas; (ii) validação do modelo; (iii) comparação de modelos. Limitações importantes: (i) o teste F global não identifica variáveis específicas (requer testes t individuais); (ii) violação dos pressupostos compromete a validade; (iii) significância estatística não implica significância prática; (iv) multicolinearidade afeta precisão e interpretação.

Em resumo, o teste de hipótese em regressão linear múltipla fornece ferramentas poderosas para inferência estatística, mas requer verificação cuidadosa de pressupostos e interpretação contextual dos resultados.

Referências Bibliográficas

- Casella, G. & Berger, R. L. (2002). *Statistical Inference*. 2nd ed. Duxbury Press.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. 5th ed. Wiley.
- Seber, G. A. F. & Lee, A. J. (2012). *Linear Regression Analysis*. 2nd ed. Wiley.