

Análise de Dados Longitudinais

Aula 10.10.2018

José Luiz Padilha da Silva - UFPR
www.docs.ufpr.br/~jlpadilha

1 Análise de Resíduos e Diagnóstico

- Introdução
- Resíduos Usuais
- Resíduos Transformados

2 Estudo de Caso

Análise de Resíduos e Diagnóstico

Pontos Principais:

- A análise de dados longitudinais não fica completa sem a examinação dos resíduos. Ou seja, a verificação das suposições impostas ao modelo e ao processo de inferência.
- As ferramentas usuais de análise de resíduos para a regressão convencional (com observações independentes) podem ser estendidas para a estrutura longitudinal.

Resíduos

- Defina o vetor de resíduos para cada indivíduo

$$r_i = Y_i - X_i \hat{\beta}, \quad i = 1, \dots, N,$$

que é um estimador para o vetor de erros

$$\varepsilon_i = Y_i - X_i \beta, \quad i = 1, \dots, N.$$

- Tratando-se de dados longitudinais, sabemos que os componentes do vetor de resíduos r_i são correlacionados e não necessariamente têm variância constante.

Utilidade dos Resíduos r_j

Gráficos:

- Gráfico de r_{ij} vs \hat{Y}_{ij} : é útil para identificar alguma tendência sistemática (por exemplo, presença de curvatura) e presença de pontos extremos (“outliers”). O modelo corretamente especificado não deve apresentar nenhuma tendência neste gráfico.

Limitação: este gráfico não tem necessariamente uma largura constante. Ou seja, cuidado ao interpretar este gráfico com relação à homocedasticidade.

- Gráfico de r_{ij} vs t_{ij} : é também útil para identificar alguma tendência sistemática da média no tempo.

Resíduos

Para fins práticos, podemos aproximar a covariância dos resíduos por

$$\text{Cov}(r_i) \approx \text{Cov}(\varepsilon_i) = \Sigma_i.$$

- Diagnósticos usuais dos resíduos para examinar homogeneidade da variância dos resíduos ou autocorrelação entre os resíduos devem ser evitados.
- Embora os resíduos de uma regressão linear sejam não correlacionados com as covariáveis, os resíduos de uma análise longitudinal podem ser correlacionados com as covariáveis.
- Como resultado, pode haver uma aparente tendência sistemática no gráfico de dispersão dos resíduos contra a covariável.

Solução: Examinar resíduos transformados

- Há muitas possibilidades para transformar os resíduos.
- A transformação deve ser realizada de forma que os resíduos “imitem” aqueles da regressão linear padrão.
- Os resíduos r_i^* definidos a seguir são não-correlacionados e têm variância unitária:

$$r_i^* = L_i^{-1} r_i,$$

em que L_i é a matriz triangular superior resultante da decomposição de Cholesky da matriz de covariâncias estimada $\widehat{Var}(Y_i)$, ou seja, $\widehat{Var}(Y_i) = L_i L_i'$.

Resíduos transformados

- Podemos aplicar a mesma transformação ao vetor de valores preditos \hat{Y}_i , ao vetor da variável resposta Y_i e à matriz de covariáveis \mathbf{X}_i :

$$\hat{Y}_i^* = L_i^{-1} \hat{Y}_i$$

$$Y_i^* = L_i^{-1} Y_i$$

$$\mathbf{X}_i^* = \hat{L}_i^{-1} \mathbf{X}_i$$

e então todos os diagnósticos de resíduos usuais para a regressão linear padrão podem ser aplicados para r_i^* .

Gráficos de Adequação

- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus os valores preditos transformados \hat{Y}_{ij}^* : não deve apresentar nenhum padrão sistemático para um modelo corretamente especificado. Ou seja, deve apresentar um padrão aleatório em torno de uma média zero. Útil para verificar homocedasticidade.
- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus covariáveis transformadas X_{ij}^* (em especial, idade ou tempo): verificar padrões de mudança na resposta média ao longo do tempo.
- QQ-plot de r_i^* : verificar normalidade e identificar outliers.

Estudo de Caso: Influência da menarca nas mudanças do percentual de gordura corporal

Faremos a análise de um estudo prospectivo do desenvolvimento de obesidade em uma coorte de garotas.

Vamos examinar os resíduos do modelo ajustado para avaliar a adequação geral do modelo.

Os resíduos podem também ser úteis para detectar indivíduos com padrões de resposta incomuns.