

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
MAT02035 - Modelos para dados correlacionados

Avaliação 02 (2×10 pontos)

Problema 1. No National Cooperative Gallstone Study (NCGS), um dos principais interesses era estudar a segurança da droga chenodiol para o tratamento de cálculos biliares de colesterol (Schoenfield et al., 1981; Wei e Lachin, 1984). Neste estudo, os pacientes foram aleatoriamente designados para dose alta (750 mg por dia), dose baixa (375 mg por dia) ou placebo. Nós nos concentramos em um subconjunto de dados em pacientes que tinham cálculos biliares flutuantes e que foram designados para os grupos de alta dose e placebo.

No NCGS, foi sugerido que o chenodiol dissolveria os cálculos biliares, mas, ao fazê-lo, poderia aumentar os níveis de colesterol sérico. Como resultado, o colesterol sérico (mg/dL) foi medido no início do estudo e aos 6, 12, 20 e 24 meses de acompanhamento. Muitas medições de colesterol estão faltando por causa de visitas perdidas, amostras de laboratório foram perdidas ou inadequadas, ou o acompanhamento do paciente foi encerrado.

- Os dados de colesterol sérico do NCGS estão armazenados em um arquivo externo: `colesterol.dta`.
- Cada linha do conjunto de dados contém as seguintes sete variáveis:
 - `Group ID Y1 Y2 Y3 Y4 Y5` Nota: A variável categórica `Group` é codificada 1 = Alta Dose, 2 = Placebo.

1. Carregando e descrevendo os dados.
 - a. Ler os dados do arquivo externo e mantê-los em formato “multivariado” ou “largo”.
 - b. Calcule as médias amostrais, desvios padrão e variâncias dos níveis de colesterol sérico em cada ocasião para cada grupo de tratamento.
 - c. Em um único gráfico, construa um gráfico de tempo que exiba a média do colesterol sérico versus o tempo (em meses) para os dois grupos de tratamento. Descreva as características gerais das tendências temporais para os dois grupos.
 - d. Em seguida leia os dados do arquivo externo e coloque os dados em formato “univariado” ou “longo”, com cinco “registros” por indivíduo (**Dica: resolver este item após o item [a]**).
2. Ajustando o modelo aos dados.
 - a. Assumindo uma matriz de covariância não estruturada, realize uma **análise dos perfis de resposta**. Determine se os padrões de mudança ao longo do tempo diferem nos dois grupos de tratamento.
 - b. Exiba as matrizes de covariância e correlação 5×5 estimadas (pelo modelo) para

as cinco medições repetidas de colesterol sérico.

- c. Tendo a linha de base (mês 0) e o grupo placebo (grupo 2) como grupo de referência, escreva o modelo de regressão para o colesterol sérico médio que corresponde à análise dos perfis de resposta no Questão [2.a] (**Dica: escreva a matriz de covariáveis para um indivíduo de cada grupo de tratamento**).
- d. Seja L uma matriz de pesos conhecidos e β o vetor de parâmetros de regressão linear do modelo assumido na Questão [2.c]. A hipótese nula de que os padrões de mudança ao longo do tempo não diferem nos dois grupos de tratamento pode ser expressa como $H_0 : L\beta = 0$. Descreva uma matriz de pesos L apropriada para esta hipótese nula.
- e. Mostre como os coeficientes de regressão estimados a partir de uma análise de perfis de resposta podem ser usados para construir as médias específicas do tempo nos dois grupos. Compare essas médias estimadas com as médias amostrais obtidas na Questão [1.b].

Problema 2. Em um estudo de ganho de peso (Box, 1950), os investigadores distribuíram aleatoriamente 30 ratos em três grupos de tratamento: o tratamento 1 era um controle (sem aditivo); os tratamentos 2 e 3 consistiram de dois aditivos diferentes (tiouracil e tiroxina, respectivamente) para a ingestão de água dos ratos. O peso, em gramas, foi medido no início (semana 0) e nas semanas 1, 2, 3 e 4. Devido a um acidente no início do estudo, os dados de 3 ratos do grupo tiroxina não estão disponíveis.

- Os dados brutos são armazenados em um arquivo externo: rat.dta
- Cada linha do conjunto de dados contém as seguintes sete variáveis:
 - ID Group Y1 Y2 Y3 Y4 Y5
- **Nota:** A variável Group é codificada 1 = controle, 2 = tiouracil e 3 = tiroxina.

1. Carregando e descrevendo os dados.
 - a. Ler os dados do arquivo externo e colocar os dados em formato “univariado” ou “longo”, com cinco “registros” por indivíduo.
 - b. Em um único gráfico, construa um gráfico de tempo que exiba o peso médio versus o tempo (em semanas) para os três grupos. Descreva as características gerais das tendências temporais para os três grupos.
2. Ajustando o modelo aos dados.
 - a. Suponha que a taxa de aumento em cada grupo seja **aproximadamente constante** ao longo da duração do estudo. Assumindo uma matriz de covariância não estruturada, construa um teste para verificar se a taxa de aumento difere nos grupos (**Dica: adote a abordagem de curvas paramétricas ou semiparamétricas; tendências lineares**).
 - b. Em um único gráfico, construa um gráfico de tempo que exiba o peso médio estimado versus tempo (em semanas) para os três grupos de tratamento a partir dos resultados gerados na Questão [2.a].
 - c. Com base nos resultados da Questão [2.a], qual é a taxa estimada de aumento do peso médio no grupo de controle (grupo 1)? Qual é a taxa estimada de aumento do peso médio no grupo tiouracil (grupo 2)? Qual é a taxa estimada de aumento do peso médio no grupo tiroxina (grupo 3)?
 - d. Os investigadores do estudo conjecturaram que haveria um aumento no peso, mas que a taxa de aumento se estabilizaria no final do estudo. Eles também conjecturaram que este padrão de mudança pode diferir nos três grupos de tratamento. Assumindo uma matriz de covariância não estruturada, construa um teste desta hipótese (**Dica: ajustar um modelo de tendências lineares por partes – splines lineares com um nó no tempo $t^* = 2$ semanas – com interação com efeito de grupo e testar a hipótese de ausência de interação**).
 - e. Compare e contraste os resultados das Questões [2.a] e [2.d]. Será que um modelo com apenas uma tendência linear no tempo explica adequadamente o padrão de mudança nos três grupos de tratamentos? Forneça resultados que apoiem sua conclusão (**Dica: comparação de modelos via teste de hipóteses ou algum critério estatístico**).
 - f. Dado os resultados de todas as análises anteriores, que conclusões podem ser tiradas sobre o efeito dos aditivos nos padrões de mudança de peso?