Análise de Dados Categóricos
Resposta multicategórica
Seções 3.4, 3.4.1, 3.4.2 e 3.3.3

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Outubro, 2023

- **3.4 Modelos de regressão de resposta ordinal**
- **3.4.1 Odds ratio**
- **3.4.2 Tabelas de contingência**
- **3.4.3 Modelo de probabilidades não proporcionais**

Muitas variáveis de resposta categórica têm uma ordenação natural para seus níveis.

Por exemplo, uma variável de resposta pode ser medida usando uma escala Likert com categorias “discordo totalmente”, “discordo”, “neutro”, “concordo” ou “concordo totalmente”.

Se os níveis de resposta puderem ser organizados de modo que a categoria 1 < categoria 2 < $\cdots$ < categoria J em alguma escala conceitual de medida (por exemplo, quantidade de concordância), então os modelos de regressão podem incorporar essa ordenação por meio de uma variedade de transformações logit da resposta probabilidades. Nesta seção, focamos na modelagem de probabilidades acumuladas com base na ordenação das categorias.

A probabilidade acumulada para a categoria j de Y é $P(Y \leq j) = \pi_1 + \cdots + \pi_j$ para $j = 1, \cdots, J$. Observe que $P(Y \leq J) = 1$. Modelos de regressão para respostas multinomiais ordinais podem examinar os efeitos das variáveis explicativas $x_1, \cdots, x_p$ nas probabilidades ou log-odds acumulados, também chamadas de logits cumulativos,

$$logit\big(P(Y \leq j)\big) = \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right).$$

Em particular, o modelo de probabilidades proporcionais é um modelo especial que assume que o logit dessas probabilidades acumuladas muda linearmente à medida que as variáveis explicativas mudam e também que a inclinação dessa relação é a mesma independentemente da categoria j.

Formalmente, o modelo é definido como

$$logit\big(P(Y \leq j)\big) = \beta_{j0} + \beta_1 x_1 + \cdots + \beta_p x_p,$$

para $j = 1, \cdots, J$.

Observe que não há subscritos j nos parâmetros $\beta_1, \cdots, \beta_p$. O modelo assume que os efeitos das variáveis explicativas são os mesmos, independentemente de quais probabilidades acumuladas são usadas para formar os logaritmos.

Assim, o nome “proporcional odds” deriva de cada odd ser um múltiplo de $\exp(\beta_{j0})$.

Para um j fixo, aumentar x_r por c unidades muda cada log-odds na equação acima por $c\,\beta_r$ ao manter as outras variáveis explicativas constantes.

Por outro lado, a diferença no log-odds entre as categorias de resposta j e j' é constante, $\beta_{j0} - \beta_{j'0}$ e não depende dos valores de $x_1, \cdots, x_p$ quando eles são mantidos fixos.

Esses resultados estão diretamente relacionados às razões de chances, conforme detalhado na Seção 3.4.1. Além disso, observe que as chances devem aumentar à medida que j aumenta, porque colocamos progressivamente mais probabilidade no numerador, $P(Y \leq j)$. Isso implica que $\beta_{10} < \cdots < \beta_{J-1,0}$.

As probabilidades de observar uma determinada categoria de resposta j são encontradas observando que

$$\pi_j = P(Y = j) = P(Y \leq j) - P(Y \leq j - 1)$$

onde $P(Y \leq 0) = 0, P(Y \leq J) = 1$ e

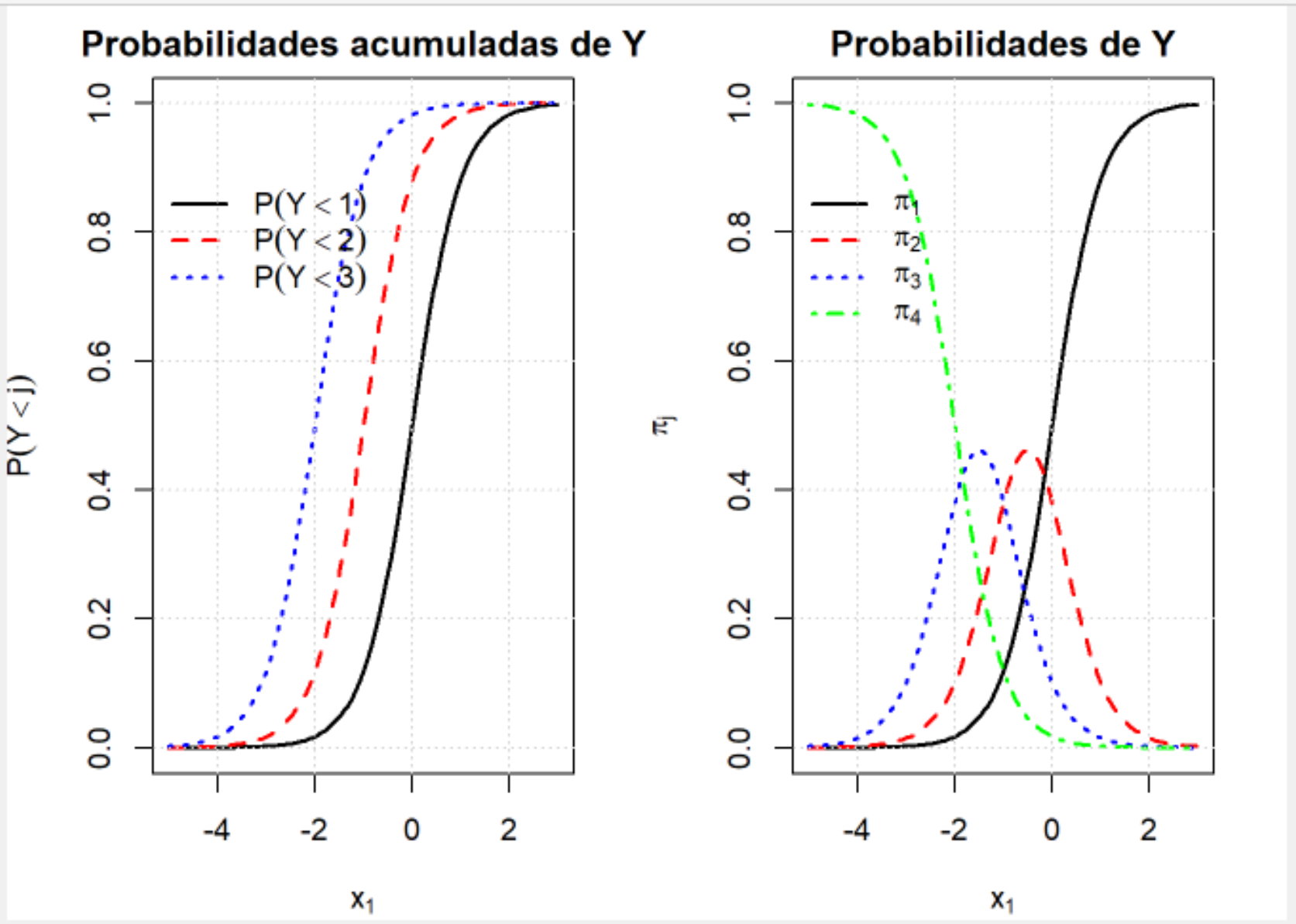
$$P(Y \leq j) = \frac{\exp\left(\beta_{j0} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}{1 + \exp\left(\beta_{j0} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}.$$

Por exemplo, a probabilidade para a categoria 1 é

$$\pi_1 = P(Y \leq 1) - P(Y \leq 0) = \frac{\exp\left(\beta_{10} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}{1 + \exp\left(\beta_{10} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}$$

e a probabilidade para a categoria J é

$$\begin{aligned} \pi_J &= P(Y \leq J) - P(Y \leq J-1) \\ &= \frac{\exp\left(\beta_{J-1,0} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}{1 + \exp\left(\beta_{J-1,0} + \beta_1 x_1 + \cdots + \beta_p x_p\right)}. \end{aligned}$$


Probabilidades acumuladas de Y
P(Y < j)
P(Y < 1)
P(Y < 2)
P(Y < 3)
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
x1
Probabilidades de Y
πj
π1
π2
π3
π4
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
x1

Os parâmetros do modelo de probabilidades proporcionais são estimados usando a máxima verossimilhança. Semelhante à Seção 3.3, a função de verossimilhança para uma amostra de tamanho *m* é simplesmente o produto de *m* distribuições multinomiais com parâmetros de probabilidade expressos como funções das variáveis explicativas. Procedimentos numéricos iterativos são então usados para ajustar o modelo.

A função **polr()** do pacote **MASS** executa os cálculos necessários. É importante garantir que os níveis da resposta categórica sejam ordenados da maneira desejada ao usar **polr()**; caso contrário, a ordenação dos níveis de *Y* não será considerada corretamente. Demonstraremos como verificar e, se necessário, alterar a ordenação no próximo exemplo. A matriz de covariância para as estimativas de parâmetros de regressão é encontrada usando procedimentos de verossimilhança padrão.

Os procedimentos de inferência baseados em Wald e LR também são realizados da maneira usual. As hipóteses associadas aos testes de parâmetros de regressão individuais são

$$H_0 : \beta_r = 0 \qquad \text{vs.} \qquad H_1 : \beta_r \neq 0.$$

Se a hipótese nula for verdadeira, isso diz que $J - 1$ log-odds comparando $P(Y \leq j)$ a $P(Y > j)$ não dependem de x_r, mantendo todas as outras variáveis explicativas constantes.

Se a hipótese alternativa for verdadeira, então as log-odds para cada probabilidade cumulativa aumentam ou diminuem com x_r dependendo do sinal de β_r. Por sua vez, isso impõe uma ordenação nas probabilidades de categorias individuais, como a vista no gráfico à direita da figura no exemplo anterior.

Compare o teste anterior envolvendo β_r com um teste correspondente envolvendo o modelo de regressão multinomial com uma resposta nominal da Seção 3.3,

$$H_0 : \beta_{2r} = \cdots = \beta_{Jr} = 0 \text{ vs. } H_1 : \text{Pelo menos um } \beta_{jr} \neq 0.$$

O modelo da hipótese alternativa coloca menos restrições sobre como x_r se relaciona com probabilidades de categorias individuais por meio do uso de mais parâmetros do que o modelo de probabilidades proporcionais. Assim, quando as hipóteses de probabilidades proporcionais são aplicáveis, o modelo de hipóteses alternativas para regressão multinomial descreve a relação entre a resposta e as variáveis explicativas de forma menos eficiente.

Exemplo 3.9: Grãos de trigo.

As razões de probabilidades ou odds ratios baseadas nas probabilidades acumuladas são facilmente formadas porque o modelo de probabilidades proporcionais iguala os logaritmos das probabilidades ao preditor linear.

Por exemplo, a razão de chances envolvendo uma variável explicativa, digamos x_1, é

$$\frac{Odds_{x_1+c,\cdots,x_p}(Y \leq j)}{Odds_{x_1,\cdots,x_p}(Y \leq j)} = \frac{e^{\beta_{j0}+\beta_1(x_1+c)+\cdots+\beta_p x_p}}{e^{\beta_{j0}+\beta_1 x_1+\cdots+\beta_p x_p}} = e^{c\,\beta_1},$$

onde deixamos $Odds_{x_1,\cdots,x_p}(Y \leq j)$ denotar as chances de observar a categoria j ou menor para Y .

Observe que esta formulação da razão de chances assume que $x_2, \cdots, x_p$ permanecem inalterados. A interpretação formal da razão de chances é:

As chances de $Y \leq j$ vs. $Y > j$ mudam em $e^{c\,\beta_1}$ vezes para um aumento de c unidades em x_1 enquanto as outras variáveis explicativas no modelo são mantidas constantes.

Curiosamente, a razão de chances permanece a mesma, não importa qual categoria de resposta seja usada para j. Esta característica chave do modelo de probabilidades proporcionais ocorre devido à ausência de um subscrito j nos parâmetros de inclinação $\beta_1, \cdots, \beta_p$.

Exemplo 3.10: Grãos de trigo.

A Seção 3.2 discute como testar a independência em uma tabela de contingência usando um modelo de regressão multinomial para uma resposta nominal.

Um teste semelhante para independência também pode ser realizado usando o modelo de chances proporcionais, mas a hipótese alternativa especifica o tipo de dependência que pode estar presente.

Isso ocorre porque o modelo de chances proporcionais assume uma estrutura específica para a associação entre uma variável explicativa categórica X e uma variável de resposta Y e, portanto, usa menos parâmetros do que o modelo nominal para resumir essa associação.

Em particular, para uma tabela de contingência $I \times J$ com categorias ordinais de Y, o modelo de chances proporcionais é

$$logit\big(P(Y \leq j)\big) = \beta_{j0} + \beta_2 x_2 + \cdots + \beta_I x_I,$$

$j = 1, \cdots, J-1$ onde $x_2, \cdots, x_I$ são variáveis indicadoras para as linhas $2, \cdots, I$, respectivamente.

Qualquer $\beta_i \neq 0$ para $i = 2, \cdots, I$ significa que as probabilidades envolvendo probabilidades acumuladas para Y não são as mesmas nas linhas 1 e i.

Categorias inferiores de Y são mais prováveis de serem observadas na linha i do que na linha 1 se $\beta_i > 0$ e menos prováveis se $\beta_i < 0$.

Assim, a independência é testada como

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_I = 0$$

vs.

$$H_1 : \text{qualquer } \beta_i \neq 0$$

$i = 2, \cdots, I$.

Se a hipótese nula de independência for rejeitada, a associação é resumida com o auxílio dos sinais e valores de $\widehat{\beta}_2, \cdots, \widehat{\beta}_I$, juntamente com intervalos de confiança para os parâmetros correspondentes.

A equação acima contém apenas $(I-1)+(J-1)$ parâmetros, em comparação com $I(J-1)$ no modelo de regressão multinomial correspondente. A redução nos parâmetros vem da suposição de chances proporcionais, que especifica que a diferença nos logits de probabilidades acumuladas para quaisquer duas linhas é controlada por apenas um parâmetro β_i, independentemente de j.

Assim, o teste de independência usando o modelo de chances proporcionais é mais poderoso do que um teste de independência usando o modelo de regressão multinomial quando a associação realmente segue essa estrutura. Caso contrário, o teste usando o modelo de chances proporcionais pode falhar em detectar qualquer associação, mesmo quando for muito forte.

Exemplo 3.11: Biscoitos enriquecidos com fibras.

Um modelo de probabilidades proporcionais é uma das formas preferidas de explicar uma resposta multinomial ordenada, isso é porque os parâmetros de regressão de inclinação são constantes nas categorias de resposta.

Embora isso possa simplificar bastante o modelo, impõe a suposição de que a associação afeta os logaritmos das probabilidades acumuladas da mesma forma para todo $j = 1, \cdots, J-1$. Isso pode não ser verdade em todas as situações. Um modelo alternativo que relaxa essa suposição é

$$logit\big(P(Y \leq j)\big) = \beta_{j0} + \beta_{j1}x_1 + \cdots + \beta_{jp}x_p,$$

onde $j = 1, \cdots, J-1$. Observe que todos os parâmetros de regressão agora podem variar entre os níveis de Y. A equação acima é chamada de modelo de probabilidade não proporcional.

Como o modelo de chances proporcionais é um caso especial da equação acima, podemos testar a suposição de chances proporcionais por meio das hipóteses

$$H_0 : \beta_{1r} = \cdots = \beta_{J-1,r}$$
$$H_1 : \text{Nem todos são iguais,}$$

para $r = 1, \cdots, p$.

O teste é conduzido como um LRT onde os graus de liberdade para a distribuição χ^2 de referência são encontrados a partir da diferença no número de parâmetros nos dois modelos,

$$(p+1)(J-1) - (p+J-1) = p(J-2).$$

Rejeitar a suposição de probabilidades proporcionais (H_0) sugere que o modelo de probabilidades não proporcionais pode ser preferido.

As probabilidades estimadas e as razões de chances têm uma forma diferente devido aos parâmetros extras. No entanto, o modelo de probabilidades proporcionais ainda pode ser preferido devido ao seu menor número de parâmetros.

Devemos observar que cada parâmetro estimado adiciona variabilidade à estimativa subsequente de probabilidades e razões de chances, de modo que se pode obter estimativas de razões de chances mais próximas da verdade usando um modelo menor com um defeito menor do que usando um modelo muito maior sem o defeito.

Além disso, um tamanho de amostra muito grande pode resultar na rejeição da hipótese nula, mesmo que os dados se desviem apenas ligeiramente da suposição de chances proporcionais.

Deixar de rejeitar a hipótese das probabilidades proporcionais não é prova de que ela seja verdadeira.

No entanto, oferece alguma garantia de que um modelo de probabilidades proporcionais fornece uma aproximação razoável para relações verdadeiras entre Y e as variáveis explanatórias.

Examinaremos maneiras adicionais de avaliar o ajuste de um modelo no Capítulo 5.

Exemplo 3.12: Biscoitos enriquecidos com fibras.

Um problema com o uso de modelos de chances não proporcionais para uso geral é que o modelo não restringe adequadamente os parâmetros para evitar $P(Y \leq j) < P(Y \leq j')$ para $j > j'$. Assim, as probabilidades cumulativas podem diminuir em algum ponto fazendo com que a probabilidade de uma categoria individual seja menor que 0.

Essa violação das regras de probabilidade ocorre porque o efeito que uma variável explicativa tem sobre $logit\big(P(Y \leq j)\big)$ pode mudar para cada j; isto é, os parâmetros $\beta_{j1}, \cdots, \beta_{jp}$ podem variar livremente sobre os níveis de Y. Por esse motivo, é necessário ter cuidado com esses modelos para garantir que probabilidades sem sentido não ocorram. O próximo exemplo ilustra um caso em que um modelo de probabilidades não proporcional é inadequado.

Exemplo 3.13: Gráficos de modelo de probabilidades não proporcionais.