Bioestatística Entrega Trabalho Prático

Manuel Curral

November 2024

Caso de Estudo 1

Um estudo caso-controlo foi conduzido para investigar a associação entre o consumo de uma substância E e o cancro da bexiga (D). Os dados são apresentados em duas tabelas separadas para fumadores e não fumadores. Assume-se que o efeito potencial de confusão da idade foi controlado por restrição. O objetivo é calcular:

- (a) O **odds ratio (OR) não ajustado** com base nos dados combinados e fornecer o intervalo de confiança de 95%;
- (b) Comparar os dois **odds ratios específicos por tabagismo (fumadores e não fumadores)** com o OR não ajustado, interpretando os resultados.

Resolução

(a) Odds Ratio Não Ajustado

Os dados combinados podem ser obtidos somando os valores das tabelas de fumadores e não fumadores:

Grupo	**D (Casos)**	**Não D (Controles)**
$E(Consumidores)$	$35 + 10 = 45$	$20 + 20 = 40$
Não E (Não Consumidores)	$5 + 15 = 20$	$10 + 30 = 40$

O odds ratio (OR) é calculado usando a fórmula:

$$OR = \frac{a \cdot d}{b \cdot c},$$

onde:

- $a = 45$: Casos que consumiram E,
- $b = 40$: Controles que consumiram E,
- $c = 20$: Casos que não consumiram E,
- $d = 40$: Controles que não consumiram E.

Substituindo os valores:

$$OR = \frac{45 \cdot 40}{40 \cdot 20} = \frac{1800}{800} = 2.25.$$

Portanto, o **odds ratio não ajustado** é $OR = 2.25$.

Intervalo de Confiança de 95%

O intervalo de confiança (IC) para o odds ratio pode ser calculado usando a fórmula:

$$\ln(OR) \pm Z \cdot \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}},$$

onde $Z = 1.96$ para um nível de confiança de 95%.

Primeiro, calculamos o erro padrão (SE):

$$SE = \sqrt{\frac{1}{45} + \frac{1}{40} + \frac{1}{20} + \frac{1}{40}}.$$

Resolvendo:

$$SE = \sqrt{0.0222 + 0.025 + 0.05 + 0.025} = \sqrt{0.1222} \approx 0.3496.$$

Agora, calculamos o intervalo de confiança em logaritmo natural:

$$\ln(OR) = \ln(2.25) \approx 0.8109,$$

$$IC_{\ln} = 0.8109 \pm 1.96 \cdot 0.3496.$$

Resolvendo:

$$IC_{\ln} = 0.8109 \pm 0.6852,$$

$$IC_{\ln} = [0.1257, 1.4961].$$

Convertendo de volta para a escala original ($e^{\ln}$):

$$IC = [e^{0.1257}, e^{1.4961}] \approx [1.134, 4.464](3cd).$$

Conclusão

O **odds ratio não ajustado** é 2.25, com um intervalo de confiança de 95% de $[1.13, 4.47]$. Isso indica que consumidores da substância E têm aproximadamente 2.25 vezes mais probabilidade de desenvolver cancro da bexiga em comparação com não consumidores, com o intervalo de confiança excluindo 1, sugerindo uma associação estatisticamente significativa.

(b) Comparação dos Odds Ratios Específicos

Agora, calculamos os odds ratios específicos para fumadores e não fumadores:

1. Odds Ratio para Fumadores

Da tabela para fumadores:

Grupo	D (Casos)	Não D (Controles)
$E(Consumidores)$	35	20
Não E (Não Consumidores)	5	10

O odds ratio (OR) para fumadores é:

$$OR_{\text{fumadores}} = \frac{35 \cdot 10}{20 \cdot 5} = \frac{350}{100} = 3.5.$$

Erro Padrão para $OR_{\textbf{fumadores}}$

O erro padrão (SE) para o logaritmo do odds ratio é calculado como:

$$SE\left(\ln(OR)\right) = \sqrt{\frac{1}{35} + \frac{1}{20} + \frac{1}{5} + \frac{1}{10}}.$$

Calculando:

$$SE\left(\ln(OR)\right) = \sqrt{0.02857 + 0.05 + 0.2 + 0.1} = \sqrt{0.37857} \approx 0.6151.$$

Intervalo de Confiança para $OR_{\textbf{fumadores}}$

O intervalo de confiança de 95% para $\ln(OR)$ é calculado como:

$$\ln(OR) \pm Z \cdot SE\left(\ln(OR)\right),$$

onde $Z = 1.96$ para um nível de confiança de 95%.

Substituindo os valores:

$$\ln(3.5) \pm 1.96 \cdot 0.615 \implies 1.253 \pm 1.2056.$$

Convertendo para a escala original ($e^{\ln}$):

$$\text{Limite Inferior} = e^{1.253-1.96*0.615} = e^{0.0472} \approx 1.048,$$

$$\text{Limite Superior} = e^{1.253+1.96*0.615} = e^{2.4584} \approx 11.705.$$

Portanto, o **OR específico para fumadores** é 3.5 com um intervalo de confiança de 95% de $[1.048, 11.705](3cd)$.

Conclusão

O odds ratio específico para fumadores (3.5) com intervalo de confiança amplo indica uma associação significativa entre o consumo de E e o cancro da bexiga para fumadores. Esse resultado é mais forte do que o odds ratio não ajustado (2.25), o que reflete o impacto do tabagismo como modificador do efeito.

2. Odds Ratio para Não Fumadores

Da tabela para não fumadores:

Grupo	**D (Casos)**	**Não D (Controles)**
E (Expostos)	10	20
Não E (Não expostos)	15	30

O **odds ratio** é calculado utilizando a fórmula:

$$OR = \frac{a \cdot d}{b \cdot c},$$

onde:

- $a = 10$: Casos expostos (E),

- $b = 20$: Controles expostos (E),
- $c = 15$: Casos não expostos (não E),
- $d = 30$: Controles não expostos (não E).

Substituindo os valores:

$$OR_{\text{não fumadores}} = \frac{10 \cdot 30}{20 \cdot 15} = \frac{300}{300} = 1.0.$$

Cálculo do Intervalo de Confiança de 95% para Não Fumadores

Primeiro, calculamos o erro padrão (SE) da escala logarítmica:

$$SE(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}.$$

Substituindo os valores:

$$SE(\ln(OR)) = \sqrt{\frac{1}{10} + \frac{1}{20} + \frac{1}{15} + \frac{1}{30}} = \sqrt{0.1 + 0.05 + 0.0667 + 0.0333}.$$

Somando os termos:

$$SE(\ln(OR)) = \sqrt{0.25} = 0.5.$$

Agora, calculamos o intervalo de confiança em escala logarítmica:

$$\ln(OR) = \ln(1) = 0,$$

$$IC_{\ln} = 0 \pm 1.96 \cdot 0.5 \implies IC_{\ln} = 0 \pm 0.98.$$

Isso resulta em:

$$IC_{\ln} = [-0.98, 0.98].$$

Convertendo de volta para a escala original:

$$IC = [e^{-0.98}, e^{0.98}] \approx [0.375, 2.664](3cd).$$

Conclusão para Não Fumadores

O **odds ratio para não fumadores** é $OR = 1.0$, com um intervalo de confiança de 95% de $[0.375, 2.664](3cd)$. Este resultado sugere que não há uma associação estatisticamente significativa entre o consumo da substância E e o cancro da bexiga entre os não fumadores, uma vez que o intervalo de confiança inclui 1.

Comparação dos Odds Ratios

- **Odds Ratio para Fumadores:** $OR = 3.5$ (95% IC: [1.048, 11.705](3cd)).
- **Odds Ratio para Não Fumadores:** $OR = 1.0$ (95% IC: [0.375, 2.664](3cd)).
- **Odds Ratio Não Ajustado:** $OR = 2.25$ (95% IC: [1.134, 4.464]).

A comparação dos **odds ratios específicos** mostra que o OR é consideravelmente mais elevado para fumadores (3.5) em comparação com não fumadores (1.0). O OR não ajustado (2.25) representa uma média entre os dois subgrupos, indicando que o tabagismo atua como um **modificador de efeito** na associação entre o consumo da substância E e o cancro da bexiga.

Embora os intervalos de confiança se sobreponham ligeiramente, o OR mais elevado entre fumadores indica que a associação é mais forte nesse grupo. Entre os não fumadores, não se observa uma associação significativa, reforçando o papel do tabagismo como fator modificador.

Interpretação Final

O tabagismo parece modificar a relação entre o consumo da substância E e o cancro da bexiga. Enquanto a análise não ajustada indica uma associação geral ($OR = 2.25$), a análise estratificada revela que esta associação está presente principalmente entre fumadores, com um OR elevado (3.5) em comparação com os não fumadores (1.0). O tabagismo, portanto, amplifica o risco associado ao consumo de E.

O OR é:

$$OR_{\text{não fumadores}} = \frac{10 \cdot 30}{20 \cdot 15} = \frac{300}{300} = 1.0.$$

3. Interpretação

Os resultados indicam que o consumo da substância E está associado a um maior risco de cancro da bexiga apenas em fumadores ($OR = 3.5$), enquanto que, em não fumadores, não há associação ($OR = 1.0$).

Conclusão A discrepância entre o OR não ajustado (2.25) e os ORs específicos (3.5 para fumadores e 1.0 para não fumadores) destaca que o **tabagismo é um modificador de efeito** na relação entre o consumo de E e o cancro da bexiga. O OR não ajustado é uma média dos dois subgrupos, enquanto os ORs específicos revelam a influência distinta do tabagismo.

Caso de Estudo 2

A **incidência** refere-se ao número de novos casos de uma condição num dado período. Neste estudo, representa os participantes que desenvolveram hipertensão arterial durante o acompanhamento de 2 anos.

A **prevalência**, por outro lado, refere-se ao número total de casos existentes num momento específico. Aqui, é o número de participantes com hipertensão arterial ao final do estudo.

A discrepância observada nos valores do risco relativo deve-se a estas definições:

- Quando o desfecho é medido pela **incidência**, o risco relativo é 2, indicando que os participantes com alto nível de stresse têm o dobro do risco de desenvolver hipertensão arterial em comparação com os de baixo stresse.
- Quando medido pela **prevalência**, o risco relativo é 4, sugerindo que os participantes com alto nível de stresse têm 4 vezes mais probabilidade de apresentar hipertensão ao final do estudo.

Esta diferença ocorre porque a prevalência é influenciada não apenas pelos novos casos, mas também pela duração da condição e pela eficácia do tratamento. No estudo, participantes com hipertensão receberam tratamento, o que pode ter normalizado os níveis de alguns, afetando a prevalência final.

Questão Bónus

Modelagem do efeito da SBP no risco de CHD, de acordo com os modelos de regressão relativos

(a) Para modelar o efeito da SBP no risco de CHD utiliza-se uma variável numérica única, ou seja, um modelo de regressão logística. A variável X representaria a SBP codificada em uma escala ordinal, definida como:

$$X = \begin{cases} 0, & \text{se SBP} < 117 \\ 1, & \text{se } 117 \leq \text{SBP} \leq 126 \\ 2, & \text{se } 127 \leq \text{SBP} \leq 136 \\ \vdots & \\ 6, & \text{se } 167 \leq \text{SBP} \leq 176 \\ 7, & \text{se SBP} \geq 177 \end{cases}$$

Essa codificação permite capturar o efeito gradual da SBP no risco de CHD. O modelo de regressão logística estimará a probabilidade de ocorrência de CHD como função da variável X. Esta abordagem é adequada, dado que a SBP possui uma relação monotônica (não necessariamente linear) com o risco de CHD, refletida na estrutura ordinal.

(b) Ao ajustar o modelo de regressão logística, podemos analisar as relações estruturais entre as oito categorias de SBP e o log das chances (*log odds*) de CHD. Isso permitirá avaliar se o efeito da SBP no risco de CHD segue uma tendência constante ou apresenta descontinuidades em certas faixas.

(c) Para estudar a variação da incidência de CHD com a SBP, podemos utilizar a função `glm()` no `R` para ajustar o modelo de regressão logística especificado em (a). Este procedimento fornecerá estimativas dos coeficientes, interpretados como o efeito da SBP na probabilidade de CHD.

Exemplo de Código em R

Por exemplo, programaticamente:

```r
# Load dos dados
data <- read.csv("dados_chd_sbp.csv")

# Ajuste o modelo de regress o log stica
modelo <- glm(CHD ~ X, data = data, family = "binomial")

# Exiba um resumo do modelo
summary(modelo)
```

(d) A razão de chances (*odds ratio*) para comparar um indivíduo com SBP entre 137 e 146 mmHg com outro cuja SBP está entre 117 e 126 mmHg pode ser calculada como:

$$\text{Odds Ratio} = \exp(\beta_3 - \beta_1),$$

onde β_3 e β_1 são os coeficientes estimados para as respectivas categorias.

(e) Para avaliar a adequação do modelo ajustado, podemos construir um gráfico comparando as *odds* observadas de CHD para cada categoria de SBP com as *odds* estimadas pelo modelo. Este gráfico permite uma análise visual da qualidade do ajuste do modelo aos dados.

Exemplo de Código em R

Por exemplo, programaticamente na sequência do código anterior:

```r

# C lculo das log odds observadas para cada categoria de SBP
observed <- aggregate(CHD ~ X, data = data, function(y) log(
   mean(y) / (1 - mean(y))))

# Extra  o das log odds estimadas pelo modelo
data$log_odds_est <- predict(modelo, type = "link")
estimated <- aggregate(log_odds_est ~ X, data = data, mean)

# Constru  o do gr fico
plot(observed$X, observed$CHD, type = "b", col = "blue", pch
   = 16, lwd = 2,
     xlab = "Categoria de SBP (X)", ylab = "Log Odds de CHD",
     main = "Log Odds Observadas e Estimadas de CHD")
lines(estimated$X, estimated$log_odds_est, type = "b", col =
   "red", pch = 17, lwd = 2)

# Legenda
legend("topleft", legend = c("Log Odds Observadas", "Log Odds
    Estimadas"),
       col = c("blue", "red"), pch = c(16, 17), lwd = 2)
```

O objetivo principal é utilizar técnicas estatísticas apropriadas, como a regressão logística, para explorar de forma fundamentada a relação entre pressão arterial sistólica e o risco de doença cardíaca coronariana (CHD).