Razão de Chances: Aplicações na Área da Saúde

Eduardo Bearzoti

17 de Fevereiro de 2025

Anotações

Anotações

Anotações

Introdução

Nesta apresentação, iremos abordar os importantes conceitos de *Chances* e *Razão de Chances*, a partir da análise do artigo científico:

SOLER, G.L.N.; SILVA, A.W.S.M.; SILVA, V.C.G.; TEIXEIRA, R.J. Doença hepática gordurosa não-alcoólica: associação com síndrome metabólica e fatores de risco cardiovascular. **Rev. SOCERJ**. n.21, v.2, p.94-100, 2008.

Anotações

Introdução

Adicionalmente, é feita uma discussão sobre o uso da Razão de Chances quando temos uma variável explicativa de regressão. Aqui foram considerados dois outros artigos:

- PERUCCI, L.O. *et al.* Neuroserpin: A potential biomarker for early-onset severe preeclampsia. **Immunobiology**. n.228, 2023.
- ROSÁRIO, N.S.A. *et al.* Exploring the effects of COVID-19-related traumatic events on the mental health of university students in Brazil: A cross-sectional investigation. **Acta Psychologica**. n.247, 2024.

Anotações

Anotações

Artigo 1

Artigo 1

A doença hepática gordurosa não-alcoólica (DHGNA) é uma doença com quadro patológico semelhante ao de lesão induzida por álcool, que ocorre em indivíduos sem ingestão etílica significativa.

Pacientes podem ser assintomáticos.

A pesquisa foi um estudo transversal de prevalência, que procurou relacionar a DHGNA com fatores de risco cardiovasculares.

Público: idosos participantes do projeto "Atividade Física na Vila".

Anotações

Anotações

Artigo 1

Neste artigo, foi apresentados valores de Razões de Chances para diferentes fatores de risco.

Por exemplo, os fatores obesidade e hipertensão:

Fatores de Risco	**DHGNA** $n = 22$	**Odds Ratio**	**Valor de** p
Obesidade ($n = 60$)	13	1,8	0,21
Hipertensão ($n = 60$)	19	5,7	0,007

Anotações

Artigo 1

Antes de mais nada, vamos entender como os dados foram coletados.

Por exemplo, em relação à obesidade, as informações do *slide* anterior permitem construir uma tabela de contingência incompleta:

	DHGNA		
Obesidade	Com DHGNA	Sem DHGNA	TOTAIS
Obeso	13		
Não obeso	9		
TOTAIS	22	38	60

Anotações

Artigo 1

Na publicação, verificamos que o número de obesos sem DHGNA foi igual a 17. Com isso somos capazes de preencher a Tabela completamente:

	DHGNA		
Obesidade	Com DHGNA	Sem DHGNA	TOTAIS
Obeso	13	17	30
Não obeso	9	21	30
TOTAIS	22	38	60

Anotações

Artigo 1

Procedendo de maneira análoga, podemos reconstruir a tabela de contingência referente à hipertensão:

	DHGNA		
Hipertensão	Com DHGNA	Sem DHGNA	TOTAIS
Hipertenso	19	20	39
Não hipertenso	3	18	21
TOTAIS	22	38	60

Anotações

Anotações

Chances e Razão de Chances

O "estranho" conceito (para nós, falantes de língua portuguesa) de chances:

Na língua inglesa, o termo "chances" é muito usado.

Na língua portuguesa, nós tendemos a confundir "chances" com o conceito de "probabilidade".

Exemplo: suponha que, antes das eleições americanas, houvesse um estadunidense (o Sr. John Doe) que acreditasse que Trump teria 75% de probabilidade de vencer o pleito.

Anotações

Chances e Razão de Chances

Se um órgão de pesquisa entrevistasse o Sr. John Doe, haveria duas maneiras de captar a sua opinião:

1) Pesquisa: "Na sua opinião, qual a *probabilidade* de Trump vencer as eleições?"
John Doe: "A probabilidade é de 75%."

2) Pesquisa: "Na sua opinião, quais as *chances* de Trump vencer as eleições?"
John Doe: "As chances são de 3 para 1."

Anotações

Chances e Razão de Chances

Ou seja, as *chances* são um parâmetro relativo, considerando a possibilidade de vitória (sucesso) em relação à possibilidade de derrota (fracasso).

Perceba: 3 para 1 (3:1) é a mesma coisa que 75% : 25%.

A partir das chances, podemos calcular a probabilidade de vitória, e vice-versa. Para calcular a probabilidade de vitória, a partir das chances 3:1, basta somar 3 com 1, e fazer:

$$\frac{3}{3+1} = \frac{3}{4} = 75\%$$

Anotações

Chances e Razão de Chances

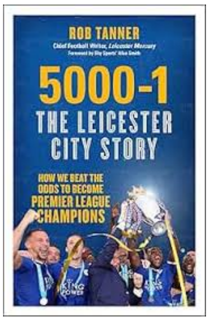
Exemplo: O Leicester City venceu a Premier League em 2016, contrariando odds de 5000:1, um dos maiores feitos da história do esporte.

Página da ESPN:

"Prior to the season, Leicester City was a 5,000-to-1 longshot to win the Premier League, according to English bookmaker William Hill."

Anotações

Chances e Razão de Chances



Anotações

Chances e Razão de Chances

Portal Terra:

Chances de perder o Campeonato Mineiro:

- Cruzeiro: 1,90
- Atlético-MG: 2,10

Anotações

Chances e Razão de Chances

Interpretação:

As chances de o Cruzeiro perder o campeonato (em relação a ganhar) são de 1,9 para 1.

As chances de o Atlético perder o campeonato (em relação a ganhar) são de 2,1 para 1.

Anotações

Chances e Razão de Chances

Perceba que as chances podem ser razões menores que 1:

> *As chances de o Cruzeiro ganhar o campeonato (em relação a perder) são de 1 para 1,9.*
>
> *As chances de o Atlético ganhar o campeonato (em relação a perder) são de 1 para 2,1.*

Anotações

Chances e Razão de Chances

No artigo, qual seria a *probabilidade* de um obeso apresentar DHGNA? E não apresentar?

	DHGNA		
Obesidade	Com DHGNA	Sem DHGNA	TOTAIS
Obeso	13	17	30
Não obeso	9	21	30
TOTAIS	22	38	60

$$\text{Apresentar: } \frac{13}{30} = 0,4333\ldots \approx 0,43 = 43\%$$

$$\text{Não apresentar: } 1 - 0,4333\ldots = 0,5666\ldots \approx 0,57 = 57\%$$

Anotações

Chances e Razão de Chances

E quais são as *chances* de um obeso apresentar DHGNA?

$$\text{Chances} = \text{Odds} = \frac{0,4333}{0,5666} = 0,76$$

Este número não é muito intuitivo, por ser menor que 1. Poderíamos expressá-lo, alternativamente:

$$0,76 = \frac{0,76}{1} = \frac{0,76/0,76}{1/0,76} = \frac{1}{1,3} = 1:1,3$$

Assim, coloquialmente, poderíamos falar: "as chances de um obeso desenvolver DHGNA (em relação a não desenvolver) é de 1 para 1,3".

Anotações

Chances e Razão de Chances

E no caso dos não-obesos?

	DHGNA		
Obesidade	Com DHGNA	Sem DHGNA	TOTAIS
Obeso	13	17	30
Não obeso	9	21	30
TOTAIS	22	38	60

$$\text{Apresentar DHGNA: } \frac{9}{30} = 0,3 = 30\%$$

$$\text{Não apresentar DHGNA: } 1 - 0,3 = 0,7 = 70\%$$

Anotações

Chances e Razão de Chances

Chances de um não-obeso apresentar DHGNA?

$$\text{Chances} = \text{Odds} = \frac{0,3}{0,7} = 0,42857\ldots \approx 0,4286$$

Como este número também é menor que 1, podemos expressá-lo, alternativamente:

$$0,4286 = \frac{0,4286}{1} = \frac{0,4286/0,4286}{1/0,4286} = \frac{1}{2,3} = 1:2,3$$

Assim, coloquialmente, poderíamos falar: "as chances de um não-obeso desenvolver DHGNA (em relação a não desenvolver) é de 1 para 2,3".

Anotações

Chances e Razão de Chances

Como as chances de desenvolver DHGNA para os obesos foram maiores do que as chances para os não obesos, daí é que surgiu a ideia de dividir as chances de uma categoria pelas chances de uma outra categoria, e isso levou à criação do conceito de Razão de Chances (em inglês, *Odds Ratio*, *OR*).

Se dividirmos as chances dos obesos pelas chances dos não obesos, encontramos:

$$OR = \frac{0,7647}{0,4286} = 1,78\ldots \approx 1,8$$

E este valor 1,8 é o que está constando na tabela publicada no artigo original.

Anotações

Chances e Razão de Chances

Interpretação:

> *As chances de um obeso apresentar DHGNA (em relação a não desenvolver) é 1,8 vezes maior que as dos não obesos.*

Anotações

Chances e Razão de Chances

A razão de chances, permite, por exemplo, comparar o Atlético com o Cruzeiro:

$$OR = \frac{\text{Atlético}}{\text{Cruzeiro}} = \frac{2,1}{1,9} = 1,1$$

> *As chances de o Atlético perder o campeonato (em relação a ganhar) é 1,1 vezes maior que as do Cruzeiro.*

ou...

> *As chances de o Cruzeiro ganhar o campeonato (em relação a perder) é 1,1 vezes maior que as do Atlético.*

Anotações

Chances e Razão de Chances

Para construirmos um intervalo de confiança para OR, ou fazer testes, podemos utilizar uma aproximação normal para o logaritmo da razão de chances.

Na escala logarítmica, o erro padrão de $\ln OR$ é dado por:

$$\hat{\sigma}(\ln OR) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

E assim um intervalo de confiança para o logaritmo da razão de chances pode ser construído por:

$$\ln OR \pm z_{\alpha/2}\hat{\sigma}(OR)$$

Anotações

Chances e Razão de Chances

Grande vantagem da OR:

Note:

$$\hat{OR} = \frac{\hat{p}_1(1-\hat{p}_2)}{\hat{p}_2(1-\hat{p}_1)} = \frac{\frac{n_{11}}{n_{1.}}\left(\frac{n_{2.}-n_{21}}{n_{2.}}\right)}{\frac{n_{21}}{n_{2.}}\left(\frac{n_{1.}-n_{11}}{n_{1.}}\right)} = \frac{\frac{n_{11}}{n_{1.}}\frac{n_{22}}{n_{2.}}}{\frac{n_{21}}{n_{2.}}\frac{n_{12}}{n_{1.}}}$$

Ou seja:

$$\hat{OR} = \frac{n_{11}n_{22}}{n_{21}n_{12}}$$

No numerador temos as células da *diagonal* da tabela; no denominador temos as células *fora da diagonal*

* a consequência é a de que, se trocarmos as linhas pelas colunas, a expressão de $\hat{OR}$ *permanece a mesma!*.

Anotações

Chances e Razão de Chances

Um estudo de caso-controle foi feito, considerando como casos pessoas de pressão elevada, e como controles pessoas de pressão normal. Em seguida observou-se se essas pessoas eram fumantes ou não:

	Hábito (Y)		
Pressão (X)	Fumante	Não-Fumante	TOTAIS
Elevada	218	305	523
Normal	99	424	523
TOTAIS	317	729	1046

Um teste de qui-quadrado revelou-se altamente significativo ($\chi^2_P = 64,10$).

Perceba que aqui a pressão é uma variável explicativa, e o hábito de fumo é uma *variável resposta*!

Anotações

Chances e Razão de Chances

Ou seja, podemos apenas concluir com o teste que há diferença quanto ao hábito de fumo, conforme a pressão.

O ideal, contudo, seria verificar a influência do hábito de fumo na pressão.

O parâmetro *Razão de Chances* é que fornece um critério. Podemos estimar:

$$\hat{OR} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{218 \times 424}{99 \times 305} = 3,06$$

Assim, embora tenhamos controlado X (número de casos e de controles), podemos afirmar que, no grupo dos fumantes, a chance de ter pressão alta, em relação a não ter, é 3,06 vezes maior do que a do grupo dos não fumantes.

Anotações

Anotações

Chances e GLM

As *chances* guardam uma estreita relação com a **Regressão Logística**, a qual é um GLM.

Componentes de Um GLM:

- Componente Aleatório
- Componente Sistemático
- Função de Ligação

Componente Aleatório

O componente aleatório se refere à distribuição de probabilidade escolhida para Y. Aqui utilizamos distribuições discretas, como a binomial.

Anotações

Chances e GLM

Componente Sistemático

O componente sistemático explicita as variáveis explicativas do modelo.

Nos GLM's o adjetivo "Linear" se refere ao fato de o componente sistemático ser uma combinação linear de parâmetros. Assim, se temos k variáveis explicativas, este componente seria:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Obs.: o componente sistemático também é chamado de *Preditor Linear*

Anotações

Chances e GLM

Função de Ligação

Vamos representar a esperança de Y como μ, ou seja, $\mu = E(Y)$.

A função de ligação é uma função g que relaciona μ com o preditor linear:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

A função de ligação mais simples que existe é a função de ligação *identidade*, na qual $g(\mu) = \mu$:

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Esta é a função em geral utilizada para modelos de regressão, por exemplo.

Anotações

Chances e GLM

Quando a variável resposta Y assume valores 0 ou 1, a função de ligação mais popular é:

$$\log \frac{\mu}{(1-\mu)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

ou seja, corresponde ao logaritmo das *chances* (*odds*).

Anotações

Chances e GLM

Esta função é chamada de *logit*, ou *logito*, e daí o nome de regressão logística.

Esta função corresponde à função natural (parâmetro natural) dos parâmetros do modelo, quando expressamos a binomial como uma família exponencial.

Quando isso ocorre, chamamos a função de ligação de ligação canônica.

Uma das vantagens de usarmos ligações canônicas é que as mesmas garantem a concavidade da verossimilhança e consequentemente muitos resultados assintóticos são obtidos mais facilmente.

Anotações

Chances e GLM

No caso da distribuição Bernoulli:

$$f(y_i|p_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

Expressando como uma família exponencial, temos:

$$f(y_i|p_i) = (1-p_i)\exp\left\{y_i \ln \frac{p_i}{1-p_i}\right\}$$

O parâmetro natural é, portanto $\ln \frac{p_i}{1-p_i}$, sendo esta portanto a ligação canônica:

$$\ln \frac{p_i}{1-p_i} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

representando uma regressão logística com k variáveis explicativas.

Anotações

Anotações

OR e Regressão Logística

Regressão Logística: Interpretação de Parâmetros

Vamos considerar uma regressão logística com uma única variável explicativa, ou seja:

$$\ln \frac{\mu}{(1-\mu)} = \beta_0 + \beta_1 X$$

lembrando que μ é a esperança de Y para um dado valor de X, ou seja, é a probabilidade de sucesso para um dado valor de X.

A função de ligação *logito* corresponde assim ao logaritmo da chance de ser um sucesso, em relação a ser um fracasso. Assim:

$$\frac{\mu}{(1-\mu)} = \exp(\beta_0 + \beta_1 X) = e^{\beta_0}\left(e^{\beta_1}\right)^X$$

Anotações

OR e Regressão Logística

Ou seja, a cada incremento de 1 unidade em X, a chance de ser um sucesso fica multiplicada por e^{β_1}, pois:

$$e^{\beta_0}\left(e^{\beta_1}\right)^{X+1} = e^{\beta_0}\left(e^{\beta_1}\right)^X e^{\beta_1}$$

Além disso, o termo e^{β_0} corresponde à chance quando $X = 0$.

Poderíamos ter interesse também em estimar *razões de chances*, considerando dois valores distintos de X.

Designando estes dois valores por $X = x_1$ e $X = x_2$, e as correspondentes probabilidades de sucesso por μ_1 e μ_2, temos:

Anotações

OR e Regressão Logística

$$\ln \frac{\mu_1}{(1-\mu_1)} = \beta_0 + \beta_1 x_1$$
$$\ln \frac{\mu_2}{(1-\mu_2)} = \beta_0 + \beta_1 x_2$$

Assim:

$$\ln \frac{\mu_1}{(1-\mu_1)} - \ln \frac{\mu_2}{(1-\mu_2)} = \ln \frac{\mu_1(1-\mu_2)}{\mu_2(1-\mu_1)} =$$
$$= \beta_0 + \beta_1 x_1 - (\beta_0 + \beta_1 x_2) = \beta_1(x_1 - x_2)$$

Portanto, o termo $\beta_1(x_1 - x_2)$ corresponde ao logaritmo da razão de chances. Ou seja, esta corresponde a:

$$\frac{\mu_1(1-\mu_2)}{\mu_2(1-\mu_1)} = e^{\beta_1(x_1 - x_2)}$$

Anotações

OR e Regressão Logística

Na obesidade, temos 1 grau de liberdade, porque temos duas categorias: obesos e não-obesos.

Assim, podemos utilizar um único parâmetro β_1, referente a uma variável X que assume valor 1 para os obesos, e 0 para os não-obesos.

Fazendo o ajustamento, encontramos $\hat{\beta}_1 = 0,579$

E assim, de fato:

$$e^{\beta_1(x_1 - x_2)} = e^{0{,}579(1-0)} = 1,78$$

conforme obtido anteriormente.

Anotações

OR e Regressão Logística

Outra ilustração: PERUCCI *et al.* 2023: idade *versus* pre-eclâmpsia

	Age 1 (years-old)				
Age 2 (years-old)	44	29	24	20	15
15	21,63	4,41	2,60	1,70	1
20	12,73	2,60	1,53	1	[1,1; 2,6]
24	8,33	1,70	1	[1,1; 2,2]	[1,2; 5,6]
29	4,90	1	[1,1; 2,6]	[1,2; 5,6]	[1,3; 14,8]
44	1	[1,3; 17,9]	[1,5; 46,7]	[1,6; 100,9]	[1,8; 263,7]

Outra ilustração: ROSÁRIO *et al.* 2024: traumas *versus* sintomas de depressão, ansiedade e estresse durante a pandemia.

Anotações

Introdução
Artigo 1: Soler et al., 2008
Chances e GLM
Artigos 2 e 3: Regressão Logística e OR

FIM

Obrigado!

Thales Tavares Correa Razão de Chances: Aplicações na Área da Saúde

Anotações

Anotações

Anotações