Lista1 - MAE0560

Guilherme NºUSP: 8943160 e Leonardo NºUSP: 9793436

Capítulo 1

Exercício 1

Em cada um dos itens a seguir, classifique uma das variáveis como resposta e as demais como variáveis explicativas.

(a) Infecção urinária (curada, não curada), sexo (feminino, masculino) e tratamento (A, B, C).

Resolução

Variáveis Resposta: Infecção urinária;

Variáveis Explicativas: Sexo e tratamento.

(b) Consumo de bebida alcoólica (sim, não), câncer de esôfago (sim, não) e histórico familiar (presente, ausente).

Resolução

Variáveis Resposta: Câncer de esôfago;

Variáveis Explicativas: Consumo de bebida alcoólica e histórico familiar.

(c) Alívio da dor de cabeça (0, 1, 2, 3, 4 horas), dosagem do medicamento (10, 20, 30 mg) e idade (< 30, ≥ 30 anos).

Resolução

Variáveis Resposta: Horas de alívio de dor de cabeça;

Variáveis Explicativas: Dosagem do medicamento e idade.

(d) Método de aprendizado preferido (individual, em grupo, em sala de aula) e período escolar frequentado (padrão, integral).

Resolução

Variáveis Resposta: Método de aprendizado preferido;

Variáveis Explicativas: Período escolar.

Exercício 2

Identifique a escala de medida mais apropriada (nominal ou ordinal) associada a cada uma das variáveis citadas no exercício anterior.

Resolução

(a)

Nominal: Infecção urinária, Sexo e Tratamento;

Ordinal: Horas de alívio de dor de cabeça.

(b)

Nominal: Câncer de esôfago e Histórico familiar;

Ordinal: Dosagem do medicamento.

(c)

Nominal: Consumo de bebida alcoólica;

Ordinal: Idade.

(d)

Nominal: Método de aprendizado e Período escolar;

Ordinal: Nenhuma.

Exercício 3

Em um estudo realizado com 39 pacientes com linfoma de Hodgkin, cada paciente foi classificado simultaneamente por sexo e anormalidades na função pulmonar. Os dados estão na Tabela 1.9.

(a) Identifique o tipo de estudo realizado.

Resolução

O estudo é transversal.

(b) Obtenha a prevalência de anormalidade pulmonar: i) entre os pacientes do sexo masculino; e ii) entre os pacientes do sexo feminino.

Resolução

i) Entre pacientes do sexo masculino $= \frac{14}{26} = 0,5385$; e ii) entre pacientes do sexo feminino $= \frac{12}{13} = 0,9231$.

Exercício 4

Com o objetivo de investigar a associação entre tabaco e câncer de pulmão, 2.000 pessoas (800 fumantes e 1.200 não fumantes) foram acompanhadas por 20 anos obtendo-se os dados na Tabela 1.10.

(a) Identifique o tipo de estudo realizado.

Resolução

O estudo é longitudinal.

(b) Obtenha a incidência de câncer de pulmão: i) entre os fumantes; e ii) entre os não fumantes.

Resolução

i) Entre fumantes $= \frac{90}{800} = 0,1125$; e ii) entre não fumantes $= \frac{10}{1200} = 0,0083$.

Caítulo 2

Exercício 1

Identifique o modelo probabilístico associado ao estudo de linforna de Hodgkin apresentado no exercício 3 do Capítulo 1.

Resolução

O modelo probabilístico é o modelo multinomial.

Exercício 5

Em um estudo descrito em Hueb et al. (2010), 611 pacientes com diagnóstico de doença coronária multiarterial foram aleatoriamente alocados a uma de três terapias: (I) medicamentosa ($n_1 = 203$), (II) cirurgia ($n_2 = 203$) e (III) angioplastia ($n_3 = 205$). Ao final de um período de acompanhamento de 10 anos, foram registrados 42, 22 e 29 óbitos associados, respectivamente, às terapias I, II e III.

(a) Apresente os dados desse estudo em uma tabela de contingência.

Resolução

Table 1: Contingência

	Obitos	Não obitos	Total
Medicamentosa	42	161	203
Cirurgia	22	181	203
Angioplastia	29	176	205
Total	93	518	611

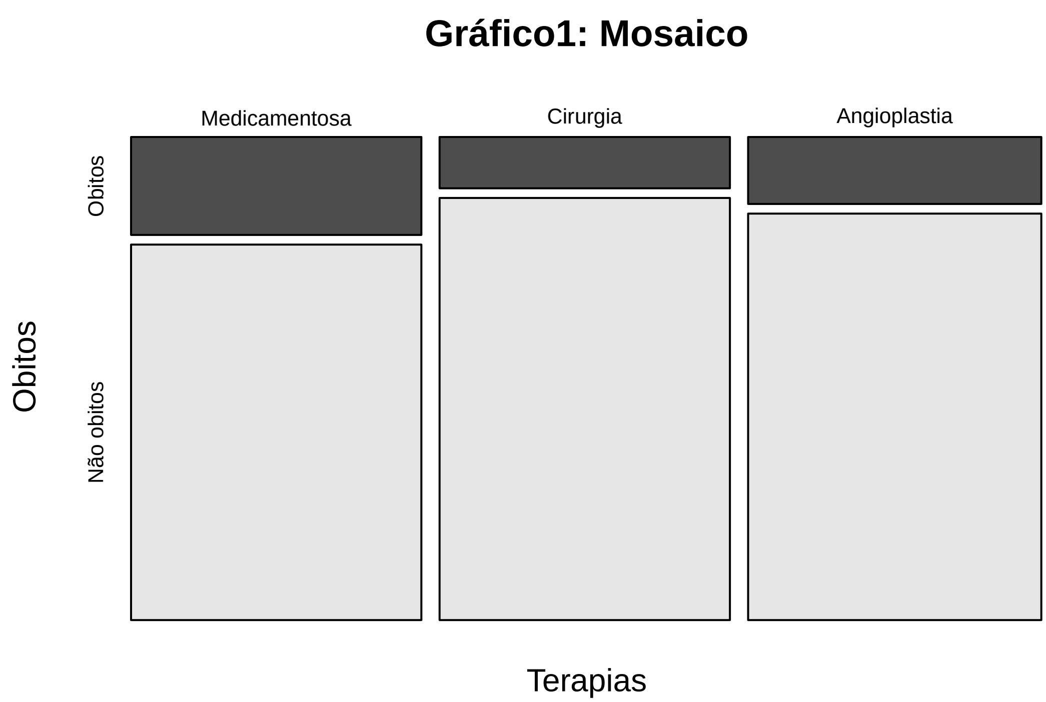
(b) Apresente o modelo probabilístico associado ao estudo.

Resolução

O modelo probabilístico é o modelo produto de binomiais.

(c) Construa um gráfico em que seja possível visualisar as proporções amostrais de óbitos e não óbitos ocorridos por terapia.

Resolução



Com o gráfico é possível ver que os grupos com mais obitos proporcionalmente são: Medicamentosa, Angioplastia e Cirurgia, respectivamente.

Exercício 6

Os dados mostrados na Tabela 2.4 são de um estudo realizado com 868 pacientes diagnosticados com artrite reumatoide e 1.194 que não possuem artrite reumatoide. O objetivo do estudo foi avaliar a associação entre o sexo e a doença.

(a) Apresente o modelo probabilístico associado ao estudo.

Resolução

O modelo probabilístico é o modelo produto de binomiais.

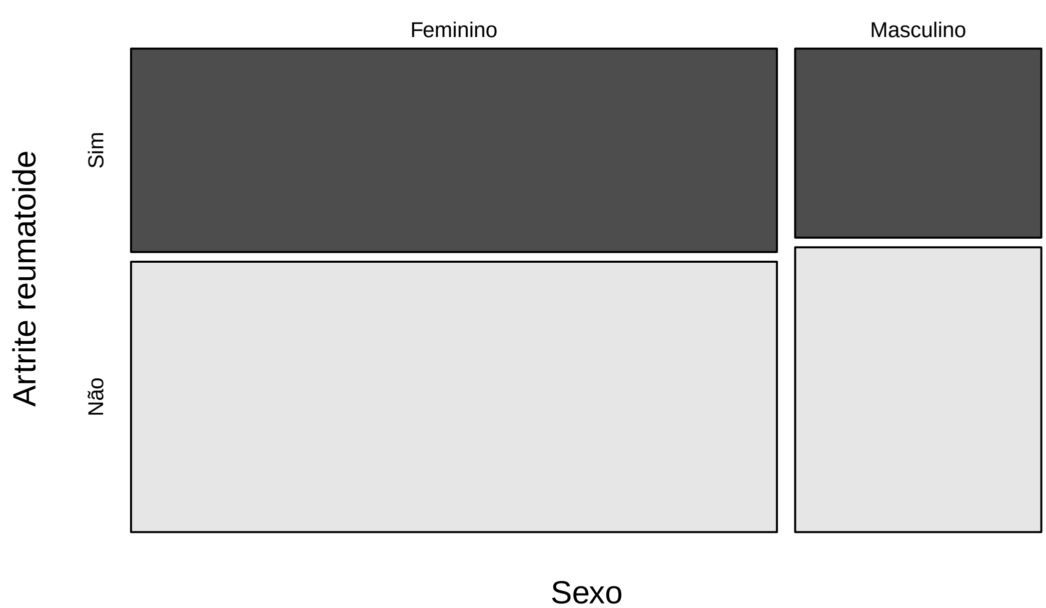
(b) Construa um gráfico em que seja possível visualizar as proporções amostrais dos casos e dos controles por sexo.

Resolução

Table 2: Contingência

	Sim	Não	Totais
Feminino	641	852	1493
Masculino	227	342	569
Totais	868	1194	2062

Gráfico 2: Mosaico



Apesar de a maioria da amostra ser do sexo feminino, a maioria das pessoas com artrite reumatoide também são do sexo feminino, de modo que a proporção do sexo feminino e masculino com artrite reumatoide são bem próximas.

Exercício 9

Mostre que a distribuição de probabilidades de $(N_{11}, N_{12}, N_{21}, N_{22})'$ condicional à soma $N = \sum_{i,j=1}^{2} Nij$, com N_{ij} Poisson(μ_{ij}) independentes, é a multinomial de parâmetros N e p = $(p_{11}, ..., p_{22})$, em que $p_{ij} = \frac{\mu_{ij}}{\sum_{i,j=1}^{2}\mu_{ij}}$, para i,j = 1,2.

Resolução

Sejam N_{11}, N_{12}, N_{21} e N_{22} variáveis aleatórias Poisson independentes com parâmetros μ_{11},μ_{12},μ_{21} e μ_{22}, respectivamente. Então,

$$P(N_{11} = n_{11}, N_{12} = n_{12}, N_{21} = n_{21}, N_{22} = n_{22})|N_{11} + N_{12} + N_{21} + N_{22} = n) =$$

$$\frac{P(N_{11} = n_{11}, N_{12} = n_{12}, N_{21} = n_{21}, N_{22} = n(n_{11} + n_{12} + n_{21}))}{P(N_{11} + N_{12} + N_{21} + N_{22} = n)} =$$

$$\frac{\frac{e^{\mu_{11}}(\mu_{11})^{n_{11}}}{n_{11}!}\frac{e^{\mu_{12}}(\mu_{12})^{n_{12}}}{n_{12}!}\frac{e^{\mu_{21}}(\mu_{21})^{n_{21}}}{n_{21}!}\frac{e^{\mu_{22}}(\mu_{22})^{n-(n_{11}+n_{12}+n_{21})}}{[n-(n_{11}+n_{12}+n_{21})]!}}{\frac{e^{(\mu_{11}+\mu_{12}+\mu_{21}+\mu_{22})}(\mu_{11}+\mu_{12}+\mu_{21}+\mu_{22})^{n}}{n!}} =$$

$$\frac{n!}{n_{11}!n_{12}!n_{21}![n(n_{11} + n_{12} + n_{21})]!}\frac{(\mu_{11})^{n_{11}}(\mu_{12})^{n_{12}}(\mu_{21})^{n_{21}}(\mu_{22})^{n(n_{11}+n_{12}+n_{21})}}{(\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22})^{n}} =$$

$$\frac{n!}{n_{11}!n_{12}!n_{21}![n(n_{11} + n_{12} + n_{21})]!}(\frac{\mu_{11}}{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}})^{n_{11}}(\frac{\mu_{12}}{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}})^{n_{12}}(\frac{\mu_{21}}{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}})^{n_{21}}$$

$$(\frac{\mu_{22}}{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}})^{n-(n_{11}+n_{12}+n_{21})} =$$

$$\frac{n!}{n_{11}!n_{12}!n_{21}![n(n_{11} + n_{12} + n_{21})]!}(p_{11})^{n_{11}}(p_{12})^{n_{12}}(p_{21})^{n_{21}}(p_{22})^{n-(n_{11}-n_{12}-n_{21})}$$

em que $p_{ij} = \frac{\mu_{ij}}{\sum_{i,j=1}^{2}\mu_{ij}}$ para i,j=1,2.

Exercício 10

Considerando o modelo multinomial expresso em (2.4), mostre que a distribuição marginal associada a cada variável N_{ij} que compõe o vetor aleatório N é a binomial com parâmetros n e p_{ij}.

Resolução

Seja $\mathbf{N} = (N_{11}, N_{12}, N_{21}, N_{22})^T$ com $\mathbf{N}|(n, \mathbf{p}) \sim Multi(n, \mathbf{p})$ com função de probabilidade expressa por:

$$\mathbb{P}(N = n) = \mathbb{P}(N_{11} = n_{11}, N_{12} = n_{12}, N_{21} = n_{21}, N_{22} = n_{22}) = n!\prod_{i=1}^{2}\prod_{j=1}^{2}\frac{(p_{ij})^{(n_{ij})}}{(n_{ij})!}$$

Como $\sum_{i,j=1}^{2} n_{i,j} = 1$ e $\sum_{i,j=1}^{2} p_{i,j} = 1$, temos que:

$$\mathbb{P}(N_{11} = n_{11}) = \sum_{N_{12},\ldots,N_{22}} \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!} p_{11}^{n_{11}} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}} = \frac{n!}{n_{11}!(n-n_{11})!} \sum_{N_{12},\ldots,N_{22}} \frac{(n-n_{11})!}{n_{12}!n_{21}!n_{22}!} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}}$$

Pelo teorema multinomial temos:

$$\frac{n!}{n_{11}!(n-n_{11})!} \sum_{N_{12},\ldots,N_{22}} \frac{(n-n_{11})!}{n_{12}!n_{21}!n_{22}!} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}} = (p_{12}+p_{21}+p_{22})^{n-n_{11}} = (1-p_{11})^{n-n_{11}}$$

Logo $\mathbb{P}(N_{11} = n_{11}) = \frac{n!}{n_{11}!(n-n_{11})!} p_{11}^{n_{11}} (1-p_{11})^{n-n_{11}} \Rightarrow N_{11} \sim Bin(n, p_{11})$ Para as variáveis N_{12}, N_{21} e N_{22} é análogo ao caso acima.

Capítulo 3

Exercício 2

Em um Programa de reabilitação de drogas, indivíduos do sexo masculino e com idade entre 25 e 34 anos, ao entrarem no programa, foram classificados segundo duas categorias étnicas (A ou B). Um ano após a entrada no programa, foi observado quantos haviam retornado ao uso das drogas. Os dados estão na Tabela 3.6.

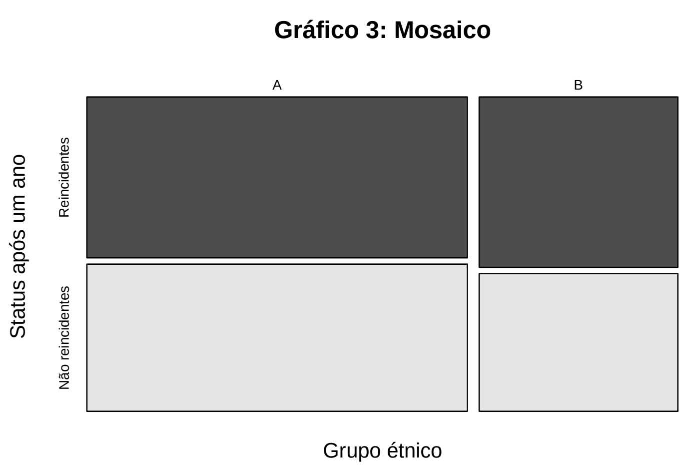
(a) Identifique o tipo de estudo realizado.

Resolução

O estudo é logitudinal

(b) Represente os dados graficamente.

Resolução



(c) Obtenha o risco relativo e seu respectivo $IC_{95\%}$. Interprete.

Resolução

Table 3: Risco Relativo

RR	li	ls
0.9440171	0.6826663	1.305423

O risco relativo é a razão entre a proporção de Reincidentes com o Total de A e de B. Essa medida varia de 0 a $+\infty$, logo quanto mais próximo de 1 há mais indícios que o Status após um ano não dependem do Grupo étnico.

Exercício 3

Para avaliar se um novo programa de acompanhamento de aleitamento materno seria mais eficiente do que o tradicional, foi realizado um estudo em duas maternidades. Na maternidade H, adotou-se o novo programa e, na maternidade A, manteve-se o tradicional.

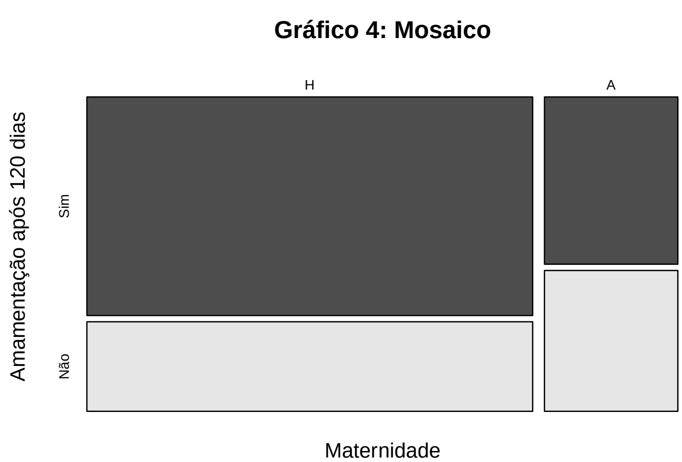
Por eficiência do programa, foi considerado se a mães, ao final dos 120 dias de acompanhamento, continuavam amamentando as crianças com leite materno. Os dados estão na Tabela 3.1.

Table 3.1: Estudo referente ao aleitamento materno

Materninade	Amamentação após 120 dias		Totais
	Sim	Não	
H	83	34	117
A	19	16	35
Totais	102	50	152

(a) Represente os dados graficamente.

Resolução



(b) Analise os dados e apresente conclusões (considere α= 5% e 10%).

Resolução

```
##
##  Pearson's Chi-squared test
##
## data:  dados
## X-squared = 3.3852, df = 1, p-value = 0.06578
```

Table 4: Risco Relativo 10%

RR	li1	ls1
1.306793	0.9438295	1.809338

Table 5: Risco Relativo 5%

RR	li2	ls2
1.306793	0.9945187	1.717119

Pelo Qui-quadrado de Pearson (teste de adenrência) onde as hipósteses são descritas abaixo:

$$\begin{cases} H_0 : \frac{\mu_{1j}}{\mu_{1+}} = \frac{\mu_{2j}}{\mu_{2+}},\ j = 1,2 \\ H_1 : \frac{\mu_{1j}}{\mu_{1+}} \neq \frac{\mu_{2j}}{\mu_{2+}},\ j = 1,2 \end{cases}$$

É possível concluir que ao nível de 5% observando um *p-value* de 0.0658 não rejeitamos a hipótese nula, já ao nível de significância de 10% rejeita-se a mesma hipótese.

Exercício 5

Mostre que a variância assintótica de $\widehat{f} = ln(\widehat{OR})$ pode ser estimada por $\widehat{V}(\widehat{f}) = (\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}})$.

Resolução

$$\widehat{Var}(\widehat{f}) = \widehat{Var}(ln(\widehat{OR})) = \widehat{Var}(ln(\frac{n_{11}n_{22}}{n_{12}n_{21}})) = \widehat{Var}(ln(n_{11}) + ln(n_{22}) - ln(n_{12}) - ln(n_{21})) =$$

Por independência temos:

$$\widehat{Var}(ln(n_{11})) + \widehat{Var}(ln(n_{22})) + \widehat{Var}(ln(n_{12})) + \widehat{Var}(ln(n_{21}))$$

Como sabemos pelo método delta que:

$$Var(f(X)) = (f'(X)) * Var(X)$$

Então assintóticamente temos:

$$\widehat{Var}(ln(n_{11})) = (\frac{1}{n_{11}})^2 * Var(n_{11}) = (\frac{1}{n_{11}})^2 * n_{11} = \frac{1}{n_{11}}$$

Logo,

$$\widehat{Var}(ln(\widehat{OR})) = (\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$$