

Exercício 2: Tabelas de contingência, distribuições teóricas

Gustavo Almeida Silva

1

Dada a seguinte tabela de contingência, que indica um estudo retrospectivo de coortes feito a partir dos prontuários eletrônicos dos primeiros 450 pacientes internados em um hospital, com COVID-19 confirmada pelo exame PCR.

Table 1: Resultado Método das Comparações Dois a Dois

	Cura	Obito	Permanece.internado	Total
>80	57	90	3	150
<80	180	83	37	300
Total	237	173	40	450

A - Quais as escalas utilizadas para as variáveis categóricas?

A variável idade é classificada como categórica ordinal, assim como a variável status

B - Qual a estrutura de dados apresentada na tabela acima, de acordo com Nishisato (2007)?

De acordo com Nishisato os dados utilizados são classificados como dados de incidência, por conta de seus elementos representem a ausência (0) ou presença (1) de um atributo, que nos dão as frequências de tais atributos.

C - Qual o modelo probabilístico para a população provavelmente utilizado? Justificar sua resposta.

O título especifica que a tabela foi construída com os 450 primeiros prontuários de Covid-19, ou seja, apenas o valor do tamanho amostral era fixado antes do estudo, assim sabendo que cada casela pode ser considerada como uma variável aleatória, o modelo a ser utilizado é o *Multinomial*

D - Qual a hipótese de interesse a ser testada?

A hipótese de interesse para a relação entre as variáveis é independência:

$$H_0 : \pi_{ij} = \pi_{i+} \times \pi_{+j}$$

E - Quais os estimadores para os parâmetros populacionais?

No caso de um modelo *Multinomial* o estimador de Máxima Verossimilhança é dado pela frequência da casela observada, ou seja:

$$\pi_{ij} = \frac{n_{ij}}{n}$$

2

O objetivo primário desse estudo prospectivo de coortes é: identificar os preditores para hospitalização e resultados clínicos desfavoráveis entre pacientes que se apresentarem para realização do exame do vírus SARS-Cov-2 no centro de testagem para COVID-19 na Universidade de Zurique (UZH), Suíça, entre Março e Dezembro de 2020.

Desfecho primário: Internação hospitalar por qualquer motivo entre pacientes que testaram positivo para SARS-Cov-2 até 30 dias após consulta inicial para realização de exame para SARS-Cov-2. As informações dos participantes serão coletadas através de chamadas telefônicas de seguimento nos dias 7, 14 e 60 após a consulta inicial, assim como através do registro de pacientes dos médicos de clínica geral e/ou médicos dos hospitais responsáveis pelo tratamento após 60 dias.

A - Se for estudada a relação entre etnia e internação de pacientes com Covid-19, quais os tipos de variáveis categóricas (escalas utilizadas)?

As variáveis categóricas que podem ser utilizadas para estudar a relação entre etnia e internação de pacientes com COVID-19 incluem:

- Etnia: essa variável categórica pode ser medida através de perguntas sobre a identidade étnica ou racial dos pacientes, como branco, negro, asiático, latino, etc.
- Gênero: essa variável categórica pode ser medida através de perguntas sobre o sexo dos pacientes, como masculino ou feminino.
- Idade: essa variável pode ser categorizada em faixas etárias, como 18-30, 31-40, 41-50, 51-60, 61-70, 71-80 e acima de 80 anos.
- Comorbidades: essa variável pode ser categorizada como a presença ou ausência de condições médicas preexistentes, como doenças cardiovasculares, diabetes, obesidade, etc.

B - Qual o modelo probabilístico para a população provavelmente utilizado para estudar tal relação?

Nesse exemplo se nota que o número de pacientes entrevistados não é fixado, onde serão entrevistados aqueles que tiveram internação hospitalar por qualquer motivo e testaram positivo para SARS-Cov-2 até 30 dias após consulta inicial para realização de exame para SARS-Cov-2. Assim, o modelo utilizado é o *Poisson*, onde cada casela possui uma distribuição *Poisson* independente

C - Qual a hipótese de interesse a ser testada?

A hipótese de interesse para a relação entre as variáveis é a de multiplicabilidade de médias

$$H_0 : \mu_{ij} = \frac{\mu_{i+} \times \mu_{+j}}{\mu_{++}}$$

D - Quais os estimadores para os parâmetros populacionais?

As frequências nas células (n_{ij}) são V.A com valores esperados $E[n_{ij}] = \mu_{ij}$. Os estimadores para os parâmetros das distribuições independentes Poisson são $\hat{\mu}_{ij} = \frac{n_{i+} \times n_{+j}}{n}$