

MAE 0560 - Análise de Dados Categorizados

2º Semestre de 2023

Prof^a Márcia D'Elia Branco

Lista 1

- Uma pesquisa foi conduzida com interesse de saber a opinião das pessoas sobre aborto. O seguinte questionamento foi feito a 1824 pessoas “Você é a favor de uma mulher grávida obter um aborto legal se ela for casada?”. Um total de 842 pessoas disseram “sim” e 982 responderam “não”. Seja p a proporção populacional de respostas “sim”.
 - Encontre a estimativa pontual da proporção de respostas “sim”. Apresente um IC(0.95) para este parâmetro. Quais as suposições que estão sendo consideradas?
 - Teste a hipótese $H_0 : p = 0.5$ × $H_a : p \neq 0.5$. Apresente o p-valor e conclua.
- Quando questionados, na Pesquisa Social Geral de 2010 nos EUA, se aceitariam cortes em seus padrões de vida para proteção do meio ambiente, 486 de 1374 indivíduos responderam *sim*.
 - Estime a proporção da população que diria *sim*. Apresente e interprete um IC(0.99) para este parâmetro.
 - Formalize e conduza um teste de significância para determinar se a maioria ou a minoria da população diria *sim*. Apresente e interprete o valor-p.
- Em um experimento sobre a herança da clorofila no milho um total de 1103 mudas de plantas verdes heterozigotas autofecundadas foram analisadas. Desse total, 854 mudas eram verdes e 249 eram amarelos. A teoria prevê que a proporção de verde para amarelo é de 3:1. Formalize e conduza um teste a hipótese de que 3:1 é a proporção verdadeira. Relate e interprete o p-valor.
- Um estudo com 100 mulheres sofrendo de sangramento menstrual excessivo considera se um novo analgésico resulta em maior alívio do que o analgésico padrão. Dessas mulheres, 40 alegaram maior alívio com o analgésico padrão.
 - Formalize e teste a hipótese de que a probabilidade de alívio com o analgésico padrão é a mesma que a probabilidade de alívio com o novo analgésico. Apresente e interprete o valor-p.
 - Construa e interprete um IC(0.95) para a probabilidade de maior alívio com o novo analgésico.
- Sejam Y_1 e Y_2 variáveis aleatórias independentes Poisson com médias μ_1 e μ_2 .
 - Obtenha a distribuição condicional de (Y_1, Y_2) dado $Y_1 + Y_2 = n$.
 - Use o resultado anterior para realizar o teste de hipóteses $H_0 : \mu_1 = \mu_2$, usando o teste de hipótese sobre o parâmetro π da distribuição Binomial.

6. Suponha que em um ensaio clínico com 10 voluntários, sucesso foi observado em 8 destes. Considerando a distribuição binomial para o número de sucessos sem considerar a aproximação normal, determine:
- O valor-p para testar (i) $H_0 : \pi \leq 0.5$ vs. $H_a : \pi > 0.5$, (ii) $H_0 : \pi \geq 0.5$ vs. $H_a : \pi < 0.5$.
 - O valor mid-p para testar (i) e (ii) como no item anterior.
 - Porque a soma dos valores-p exatos unilaterais excede um e se isso ocorre também com a soma dos valores mid-p unilaterais.
 - Um IC(0.95) para π baseado no valor mid-p utilizando um software estatístico.
7. Uma pesquisa perguntou a 25 pessoas se eram vegetarianas. Nenhuma respondeu de forma afirmativa.
- Qual é a estimativa de máxima verossimilhança para a proporção populacional de vegetarianos? Construa, interprete e diferencie IC(0.95) do tipo Score e do tipo Wald para a proporção de vegetarianos na população.
 - Assumindo uma priori $Beta(1/2, 1/2)$ para a proporção de vegetarianos, encontre a distribuição a posteriori e sua média para esse parâmetro. Apresente e interprete o intervalo de credibilidade *a posteriori* simétrico de 95%. Qual a probabilidade *a posteriori* de que a proporção de vegetarianos não excede 0,5 nessa população?
 - Baseado no item anterior, discuta como a média da distribuição a posteriori desloca a média amostral na direção da média da distribuição a priori.
8. Assuma uma amostra aleatória $X_1, \dots, X_n \stackrel{iid}{\sim} Bernouli(\pi)$. O intervalo de confiança de $100(1 - \alpha)\%$ do tipo Score para o parâmetro π consiste em todos os valores de π_0 tais que

$$\sqrt{n} \frac{|\hat{\pi} - \pi_0|}{\sqrt{\pi_0(1 - \pi_0)}} < z_{1-\frac{\alpha}{2}} \quad .$$

Obtenha a fórmula para o cálculo desse intervalo de confiança.

9. 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 a seguir.

Sexo	Anormalidade		Totais
	Presente	Ausente	
Masculino	14	12	26
Feminino	12	01	13
Totais	26	13	39

- Identifique o tipo de estudo realizado.
- Obtenha a prevalência de anormalidade pulmonar entre os pacientes do sexo masculino e entre os pacientes do sexo feminino. Compare.
- Identifique o modelo probabilístico associado ao estudo.
- Construa o IC(0.90) para a probabilidade de anomalia pulmonar condicional ao sexo. Quais suposições estão sendo consideradas para construção desses intervalos?

10. Considere $X \sim binomial(n, \pi)$ e uma distribuição *a priori* uniforme no intervalo $(0, 1)$ para π .
- Obtenha a distribuição a posteriori para π , sua média e sua moda. [Prove os resultados]
 - Usando os dados do exercício anterior, obtenha os intervalos de credibilidade 0.95 para as probabilidades de anomalia pulmonar condicionais ao sexo. Interprete esses intervalos.
 - Construa um intervalo de credibilidade 0.95 para a diferença das probabilidades de anomalia pulmonar $\pi_{Fem} - \pi_{Masc}$. Compare os grupos. [Sugestão: contruir o intervalo usando simulação]
11. Com o objetivo de investigar a associação entre tabaco e câncer de pulmão, 2000 pessoas (800 fumantes e 1200 não fumantes) foram acompanhadas por 20 anos obtendo-se os dados na tabela a seguir.

Situação	Câncer		Totais
	Sim	Não	
Fumante	90	710	800
Não fumante	10	1190	1200
Totais	100	1900	2000

- Identifique o tipo de estudo realizado.
 - Obtenha a incidência de câncer no pulmão entre os pacientes fumantes e entre os não fumantes. Compare.
 - Identifique o modelo probabilístico associado ao estudo.
 - Construa o IC(0.90) para a probabilidade de câncer de pulmão condicional à situação do indivíduo. Quais hipóteses estão sendo consideradas para construção desses intervalos?
12. Em um estudo 611 pacientes com diagnóstico de doença coronária foram aleatoriamente alocados em uma das 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.
- Identifique o tipo de estudo e organize os dados em uma tabela de contingência.
 - Qual o modelo probabilístico associado? Escreva a função de verossimilhança associada a tabela.
 - Construa um gráfico que seja possível visualizar as proporções amostrais de óbitos e não óbitos ocorridos por terapia.
13. Considere $N_{ij} \sim Poisson(\mu_{ij})$ independentes, $i = 1, \dots, r$ e $j = 1, \dots, c$.
- Obtenha a distribuição conjunta de N_{ij} condicional a $N = \sum_{i=1}^r \sum_{j=1}^c N_{ij}$.
 - Para $c = r = 2$, considere $Y_1 = N_{11}$, $Y_2 = N_{11} + N_{12}$, $Y_3 = N_{11} + N_{21}$. Mostre que a distribuição condicional de Y_1 dado Y_2, Y_3 e N é uma hipergeométrica.
14. Considere $(N_{11}, N_{12}, N_{21}, N_{22}) \sim Multinomial(n, (p_{11}, p_{12}, p_{21}, p_{22}))$.
- Mostre que a distribuição marginal de cada componente é $Binomial(n, p_{ij})$.
 - Obtenha a correlação entre as componentes do vetor.

- (c) Obtenha o estimador de máxima verossimilhança (EMV) para $(p_{11}, p_{12}, p_{21}, p_{22})$.
- (d) Obtenha o EMV restrito à condição $p_{ij} = p_{i+}p_{+j}$, em que $p_{i+} = p_{i1} + p_{i2}$ e $p_{+j} = p_{1j} + p_{2j}$.

15. Uma amostra de 309 alunos foi classificada segundo o seu desempenho em determinada disciplina e o nível socio econômico. O resultados são apresentados na tabela de dupla entrada a seguir.

Nível socio-eco	Níveis de desempenho				Totais
	A	B	C	D	
I	12	15	24	22	73
II	37	31	35	06	109
III	35	34	53	05	127
Totais	84	80	112	33	309

- (a) Qual é o modelo probabilístico associado ao plano amostral?
- (b) Para cada nível socio econômico, encontre os IC(0.90) para as diferenças de probabilidades $\pi_A - \pi_B$ e $\pi_C - \pi_D$. Interprete.
- (c) Realize um teste de hipóteses adequado para testar a independência entre Nível Socio Econômico e Nível de Desempenho. Especifique as hipóteses em função dos parâmetros do modelo probabilístico associado. Quais as suposições que estão sendo considerada para realização deste teste?
16. Uma pesquisa conduzida examinou a ocorrência de anticorpos anti-Leptospira spp em um conjunto de $n = 329$ cães segundo as suas dietas. Os dados na tabela a seguir classificam os cães de acordo com sua dieta e se apresentam ou não anti-corpos

Table 1: Número de cães segundo a dieta.

Dieta	Resultado		Totais
	Positivo	Negativo	
Comida caseira	081	231	312
Ração comercial	009	008	017
Totais	090	239	329

- (a) Identifique o tipo de estudo realizado e o modelo probabilístico associado.
- (b) Obtenha a prevalência de resultados positivos entre cães de dieta com comida caseira e cães de dieta com ração comercial. Compare.
- (c) Construa um IC(0.9) para as probabilidades da multinomial.
17. Para um experimento Binomial com $n = 25$ e $y = 9$ desenhe um gráfico com os valores das estatísticas do teste de Wald ao quadrado, considerando $\pi = 0.1, 0.2, \dots, 0.9$. Em seguida, trace uma reta indicando o quantil 0.95 da distribuição qui-quadrado com 1 graus de liberdade. O que você pode concluir? Faça o mesmo para a estatística Escore.