

ME 714 - Análise de dados discretos  
Primeiro semestre de 2017  
Lista de Exercícios II

OBS1: A menos que o contrário seja mencionado, nos exercícios você deve considerar uma amostra aleatória (não necessariamente identicamente distribuída)  $X_1, \dots, X_n$  de  $X$  (a variável aleatória, ou vetor aleatório ou o modelo de regressão, especificado na questão).

OBS2: Obter um teste (exato ou assintótico) significa, a menos que o contrário seja mencionado, propor uma estatística do teste que seja apropriada para testar as hipóteses de interesse, sua distribuição sob  $H_0$ , as regiões crítica e de aceitação, bem como o valor  $p$  (p-valor).

OBS3: Para todas as questões (sempre que for pertinente), escreva o modelo probabilístico gerador da tabela de contingência e as respectivas hipóteses de interesse em termos dos parâmetros do modelo. Também, sempre que for pertinente, apresente as frequências esperadas sob a hipótese nula, calcule os resíduos ajustados (usando-os para identificar padrões específicos da tabela de contingência em termos das hipóteses de interesse) e apresente as estimativas dos parâmetros do modelo (irrestrito) junto com os respectivos erros-padrão. Naturalmente, após aplicar os testes, deve-se escrever as conclusões pertinentes.

1. Resolva os exercícios deixados em sala.
2. Seja  $Y|X = x \sim \text{Poisson}(x)$  e  $X \sim \text{gama}(\mu, \phi)$ , tal que  $\mathcal{E}(X) = \mu$  e  $\mathcal{V}(X) = \mu^2/\phi$ , encontre a distribuição marginal de  $Y$ .
3. Seja  $Y|X = x \sim \text{binomial}(n, x)$  e  $X \sim \text{beta}(a, b)$ . Encontre a distribuição marginal de  $Y$  (ela é chamada de distribuição beta binomial de parâmetros  $(p, a, b)$ ).
4. Dizemos que uma v.a.d (com suporte nos inteiros não negativos) é inflacionada no zero, se sua fdp é da forma  $P(Y = y) = g_Y(y) = [\pi + (1 - \pi)f_Z(0)] \mathbb{1}_{\{0\}}(y) + (1 - \pi)f_Z(y) \mathbb{1}_{\{1, 2, \dots\}}$ , em que  $f_Z(\cdot)$  representa a função de probabilidade de uma vad discreta de interesse (Poisson, geométrica, binomial negativa, etc). Responda os itens:
  - a) Encontre a média, variância e a função geradora de momentos de  $Y$  (usando a forma geral de sua fdp).
  - b) Particularize os resultados obtidos no item a) para as distribuições descritas no enunciado da questão.
5. Considere o vetor aleatório  $(X, Y)$  com a seguinte distribuição de probabilidade conjunta:

		X	
		-1	2
Y	-1	1/15	3/15
Y	0	0	4/15
Y	1	5/15	2/15

Responda os itens:

- a) Encontre as distribuições marginais e condicionais.
  - b) Encontre  $Cov(X, Y)$  e  $Corre(X, Y)$ .
6. Seja  $Y_1, \dots, Y_n \stackrel{ind.}{\sim} \text{Poisson}(\mu_i)$ . Encontre a distribuição de  $(Y_1, \dots, Y_{n-1}) | \sum_{i=1}^n Y_i = s$ .
  7. Encontre a fda de uma v.a. geométrica (p) definida como o número de fracassos e como o número de tentativas, até se obter o primeiro sucesso.
  8. Obtenha os estimadores de máxima verossimilhança para os parâmetros de uma distribuição multinomial e para o produto de multinomiais, conforme discutido em sala. Calcule também a Informação de Fisher. Repita o processo, para o produto de multinomiais, considerando agora que os parâmetros de tais distribuições são iguais, conforme visto em sala.
  9. Considere o exemplo visto em sala, sobre classificação do risco de cárie com a aplicação do teste de McNemar. Aplique o teste considerando as hipóteses equivalentes àquelas vistas em sala, para as tabelas  $2 \times 2$  formadas pelas classes de classificação de risco (baixa, média) e (média, alta).
  10. Considere a tabela relativa ao exemplo anterior e todas as tabelas ( $2 \times 2$ ) possíveis de serem formadas a partir dela, ou seja, (baixo, médio), (baixo, alto) e (médio, alto). Considere que os totais de indivíduos classificados pelo método convencional foram fixados para cada categoria (ou seja, temos um produto de duas binomiais independentes). O interesse é testar se  $H_0$  : não há concordância entre os métodos vs  $H_1$  : há concordância positiva entre os métodos. Use o teste exato de Fisher e compare os resultados com o teste de McNemar.
  11. (Extraído do livro: Paula (2013) Modelos de regressão com apoio computacional) A tabela abaixo refere-se à um estudo sobre o possível efeito cancerígeno do fungicida Avadex. No estudo, 403 camundongos são observados. Desses, 65 receberam o fungicida e foram acompanhados durante 85 semanas, verificando o desenvolvimento ou não de tumor cancerígeno. Os demais animais não receberam o fungicida (grupo controle) e também foram

acompanhados pelo mesmo período, verificando a ocorrência ou não de tumor. Dois fatores potenciais de confundimento, sexo e raça, foram considerados nas análises. Por enquanto, consideraremos apenas a presença ou não de tumor e o grupos tratado e controle. Suponha que o total para cada um desses grupos (controle e tratado) foi fixado. Teste se a proporção de ratos com câncer é a mesma entre os grupos (homogeneidade).

Tumor	Grupo	
	Tratado	Controle
Sim	11	21
Não	54	317
Total	66	338

12. Considere os dados da Questão 5, agrupados pelo sexo do camundongo, como segue abaixo. Repita o teste realizado na Questão 5, para cada tabela em separado. Suas conclusões foram as mesmas obtidas na Questão 5? Comente sobre isso.

Tumor	Macho		Tumor	Fêmea	
	Grupo			Grupo	
	Tratado	Controle		Tratado	Controle
Sim	6	8	Sim	5	13
Não	26	158	Não	28	159
Total	32	166	Total	33	172

13. Considere os dados da Questão 5, agrupados pelo sexo e raça do camundongo, como segue abaixo. Repita o teste realizado na Questão 5, para cada tabela em separado. Suas conclusões foram as mesmas obtidas nas Questão 5 e 6? Comente sobre isso.

Raça 1 e Macho			Raça 2 e Macho		
Tumor	Grupo		Tumor	Grupo	
	Tratado	Controle		Tratado	Controle
Sim	4	5	Sim	2	3
Não	12	74	Não	14	84
Total	16	79	Total	16	87

Raça 1 e Fêmea			Raça 2 e Fêmea		
Tumor	Grupo		Tumor	Grupo	
	Tratado	Controle		Tratado	Controle
Sim	4	10	Sim	1	3
Não	14	80	Não	14	79
Total	18	90	Total	15	82

14. (Extraído do livro: Paula (2013) Modelos de regressão com apoio computacional) No arquivo `recrutas.dat` são descritos os resultados de um estudo desenvolvido em 1990 com recrutas americanos referente a associação entre o número de infecções de ouvido e alguns fatores. Os dados são apresentados na seguinte ordem: hábito de nadar (ocasional ou frequente), local onde costuma nadar (piscina ou praia), faixa etária (15-19, 20-25 ou 25-29), sexo (masculino ou feminino) e número de infecções de ouvido diagnosticadas pelo próprio recruta. Suponha que o número total de recrutas foi fixado. A variável resposta é o número de infecções de ouvido enquanto que as demais são explicativas. Primeiramente, categorize a variável resposta considerando 0, se não houve infecção e 1 se houve pelo menos uma infecção. Considere todas as tabelas de contingência formadas a partir da variável resposta categorizada e cada uma das quatro variáveis explicativas (ou seja, teremos quatro tabelas). Para cada uma delas teste as hipóteses de independência versus dependência entre as variáveis.