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

OBS1: A menos que o contrário seja mencionado, nos exercícios você deve considerar uma amostra aleatória (não necessariamente identidicamente distribuída) $X_1, ..., X_n$ 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, apresente as análises descritivas pertinentes 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.

OBS4: Assuma que as condições de regularidade são válidas nas situações descritas nos exercícios. Assim, a distribuição assintótica normal dos estimadores de MV e/ou sua convergência em probabilidade para os verdadeiros valores dos parâmetros são válidas.

OBS 5: Sempre que falarmos da distribuição multinomial (e, consequentemente, do produto de multinomiais independentes), considere que $\boldsymbol{X} = (X_1, ..., X_{p-1})' \sim \text{multinomial}(n, \boldsymbol{\theta})$, em que $\boldsymbol{\theta} = (\theta_1, ...\theta_{p-1})$.

1. Resolva os exercícios deixados em sala.
2. Obtenha as distribuições assintóticas dos estimadores de máxima verossimilhança para os parâmetros de uma distribuição multinomial e para o produto de multinomiais, obtidos na Questão 8 da Lista II.
3. Em relação ao exercício 2 desta lista considere o modelo multinomial e o interesse em testar as hipóteses $H_0 : \boldsymbol{C}_{(r\times(p-1))}\boldsymbol{\theta}_{((p-1)\times 1)} = \boldsymbol{M}_{(r\times 1)}$ vs $H_1 : \boldsymbol{C}_{(r\times(p-1))}\boldsymbol{\theta}_{((p-1)\times 1)} \neq \boldsymbol{M}_{(r\times 1)}$. Utilizando os estimadores de máxima verossimilhança, proponha um teste para as hipóteses em questão, com base no que foi visto em sala.
4. Repita a questão anterior, considerando agora o modelo produto de s multinomiais. Note que, agora, as hipóteses de interesse serão: $H_0 : \boldsymbol{C}_{(r\times s(p-1))}\boldsymbol{\theta}_{(s(p-1)\times 1)} = \boldsymbol{M}_{(r\times 1)}$ vs $H_1 : \boldsymbol{C}_{(r\times s(p-1))}\boldsymbol{\theta}_{(s(p-1)\times 1)} \neq \boldsymbol{M}_{(r\times 1)}$.

5. Considere o modelo linear normal homocedástico em sua forma matricial, visto em sala. Defina $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, $\widehat{Y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, $\boldsymbol{R} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ e $\widehat{\boldsymbol{\theta}}_i = \boldsymbol{C}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{M}_i$ em que $\boldsymbol{C}_{i(a_i \times p)}$ são matrizes não-aleatórias tais que $r(\boldsymbol{C}_i) = a_i < p$ e $\boldsymbol{M}_{i(a_i \times 1)}$ são vetores não aleatórios, $i = 1, 2...$. Obtenha as distribuições de i) $\widehat{\boldsymbol{\beta}}$ ii) $\widehat{\boldsymbol{\theta}}_i$ iii) $(\widehat{\boldsymbol{Y}}', \boldsymbol{R}')'$, iv) $(\widehat{\boldsymbol{\theta}}_1', \widehat{\boldsymbol{\theta}}_2')$ e v) $(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\theta}}_1')'$.

6. Considere o modelo $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \xi_i, i = 1, 2, ..., n$, em que $\xi \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Encontre a forma escalar dos estimadores de MQO e as respectivas distribuições marginais.

7. Para as questões 11, 12 e 14 da Lista II, escreva as hipóteses de interesse na forma $H_0 : \boldsymbol{B\pi} = \boldsymbol{M}$ vs $H_1 : \boldsymbol{B\pi} \neq \boldsymbol{M}$ (sempre que possível) e utilize as respectivas metodologias, apresentadas em sala de aula, para testá-las. Quando H_0 não for rejeitada, ajuste o modelo reduzido pertinente, apresentando as estimativas pontuais, intervalares e erros-padrão e os respectivos gráficos pertinentes (quando possível).

8. Repita a questão anterior através de modelos lineares e log-lineares.

9. Considere as tabelas apresentadas na Questão 11 da Lista II. O objetivo continua sendo verificar se existe dependência entre a utilização do fungicida com o surgimento dos tumores em função dos grupos formados pelas combinações dos níveis dos fatores: sexo e raça. Escreva as hipóteses de interesse, em termos das razões de chances de cada estrato(grupo) e utilize o teste de Teste de Mantel Hanszel para testá-las. Caso a hipótese nula seja rejeitada, identifique em quais estratos a dependência se verifica, utilizando metodologias apropriadas. Antes de aplicar os testes, faça uma análise das razões de chanches de cada grupo, conforme visto em sala de aula (calculando estimativas pontuais e intervalares). Faça o mesmo também para a razão de chances comum. Assim como na Questão 11 da Lista II, considere que os totais de cada linha foram fixados.

10. Repita a questão anterior utilizando as metodologias para testar as hipóteses $H_0 : \boldsymbol{B\pi} = \boldsymbol{M}$ vs $H_1 : \boldsymbol{B\pi} \neq \boldsymbol{M}$, bem como modelos lineares e log-lineares.

Sexo	Raça	Grupo	Tumor		Total
			Sim	Não	
Macho	1	Tratado	4	12	16
		Controle	5	74	79
	2	Tratado	2	14	16
		Controle	3	84	87
Fêmea	1	Tratado	4	14	18
		Controle	10	80	90
	2	Tratado	1	14	15
		Controle	3	79	82

11. Considera os dados das questões 1, 2, 8 e 11, (começando na página 268) do livro "Modelos de regressão com apoio computacional". Escreva as hipóteses de interesse na forma $H_0 : \boldsymbol{B\pi} = \boldsymbol{M}$ vs $H_1 : \boldsymbol{B\pi} \neq \boldsymbol{M}$ (sempre que possível) e utilize as respectivas metodologias, apresentadas em sala de aula, para testá-las. Quando H_0 não for rejeitada, ajuste o modelo reduzido pertinente, apresentando as estimativas pontuais, intervalares e erros-padrão e os respectivos gráficos pertinentes (quando possível).

12. Repita a questão anterior através de modelos lineares e log-lineares.

13. Os dados do artigo Tormin (2000) (fonte: http://www.ime.usp.br/ jmsinger/doku.php?id=start) são provenientes de um estudo na área de Ortodontia. Um dos objetivos era comparar as distribuições do dâmetro mesiodistal de incisivos. O pesquisador tinha interesse em saber se as distribuições dependiam do arco (direito ou esquerdo) e da posição (lateral ou central). Primeiramente, considere apenas o fator "arco". Especifique (matricial e escalarmente) um modelo normal linear homocedástico que permita avaliar a influência desse fator nas distribuições mencionadas. Utilize as quatro parametrizações vistas em sala. Interprete os parâmetros em cada caso (se possível). Escolha uma das parametrizações e ajuste o modelo em questão, testando as hipóteses de interesse, através da metodologia (re)vista em sala de aula.

14. Considere os dados da questão anterior, levando em conta, agora, os dois fatores (arco e posição). Especifique (matricial e escalarmente) um modelo linear que permita avaliar a influência desses dois fatores e de sua interação nas distribuições mencionadas. Utilize as quatro parametrizações vistas em sala e interprete os parâmetros em cada caso (se possível). Especifique (escalar e matricialmente) as hipóteses de interesse, interpretando-as em termos do problema. Escolha uma das parametrizações e ajuste o modelo em questão, testando as hipóteses de interesse, através da metodologia (re)vista em sala de aula.