

MI 602 A - Métodos Computacionais em Estatística
Primeiro semestre de 2012
Lista de Exercícios IV
Data da entrega: 18/06/2012 (todas as questões)

OBS: É PERMITIDO O USO DOS PACOTES “BOOT” E “BOOTSTRAP”.

1. Considere o seguinte conjunto de dados abaixo com $n = 30$ observações (referentes à produção de milho em kilos, por exemplo, 32.06 significa 32,06 kilos, sob diferentes combinações de nitrogênio e fosfato):

$\mathbf{x} = (32.06, 45.01, 42.80, 59.85, 36.40, 33.61, 62.06, 85.92, 54.77, 66.00, 94.28, 38.20, 61.82, 87.56, 77.97, 81.16, 80.92, 115.43, 39.35, 89.94, 90.26, 100.10, 67.97, 59.77, 115.02, 112.48, 54.19, 118.14, 134.45, 125.27)$

Responda os itens (sugestão: use a função `quantile` do R):

- a) Considerando que os dados podem ser apropriadamente modelados por uma distribuição exponencial (na parametrização adotada em nosso curso, ou seja, $\mathcal{E}(X) = \lambda$), realize um bootstrap paramétrico (estime λ via máxima verossimilhança) com $m = n$ e $R = 5000$ em que o interesse é estimar a mediana populacional. Utilizando a amostra bootstrap forneça: uma estimativa pontual, a medida de precisão associada à estimativa e um intervalo de confiança. Faça comentários com respeito à precisão de sua estimativa (inclusive comparando com a estimativa via fda, lembrando da definição de mediana) e à distribuição do estimador.
 - b) Repita o item acima realizando bootstrap não-paramétrico.
 - c) Compare os resultados obtidos nos itens acima. Qual dos procedimentos bootstrap, em sua opinião, você usaria para estimar a mediana populacional. Justifique, adequadamente, sua resposta. Sugestão: utilize algum procedimento para verificar a suposição de distribuição exponencial.
2. Suspeita-se que o conjunto de observações contidas no arquivo `snack.xls` pode ser modelado apropriadamente por uma distribuição $Weibull(\alpha, \beta)$ com valores apropriados para (α, β) , relativos à força de cisalhamento de snacks (veja o livro do Prof. Gilberto Paula, Modelo de Regressão e aplicações, para mais detalhes). Responda os itens:
 - a) Obtenha as estimativas de máxima verossimilhança de (α, β) e os respectivos erros-padrão assintóticos associados. Sugestão: utilize a função `optim` do R.
 - b) Compare o histograma dos dados com a curva teórica do modelo. Qual sua conclusão à respeito da conjectura proposta? Justifique, adequadamente, sua resposta.
 - c) Calcule a estatística de Kolmogorov-Smirnov, fornecendo os p-valores exato e aproximado associados à estatística do teste. Qual sua conclusão à respeito da conjectura proposta? Justifique, adequadamente, sua resposta. Sugestão: utilize a função `ks.test` do R.
 - d) Repita o item c), utilizando agora $R = 5000$ amostras de bootstrap com $m = n$, conforme visto em sala de aula. Qual sua conclusão à respeito da conjectura proposta? Justifique, adequadamente, sua resposta.

3. A razão de chances (“odds ratio”) é definida como a razão entre chances (razão entre probabilidades) de interesse. Por exemplo, considere a tabela abaixo que representa a quantidade de indivíduos que sofrem ou não da doença de Hodgkin (uma forma de câncer linfático) e que foram submetidos ou não à tonsilectomia (extração das amídalas). Ou seja, foram fixadas as quantidades de indivíduos que apresentam ou não a doença de Hodgkin e, dentro de cada grupo, foram verificados quantos indivíduos tiveram ou não de extrair suas amídalas. Ou seja, a tabela abaixo foi gerada por um mecanismo aleatório oriundo do produto de duas Binomiais independentes:

	tonsilectomia		Total
D. de Hodgkin	Sim	Não	
Sim	67(x_{11})	34(x_{12})	101
Não	43(x_{21})	64(x_{22})	107
Total	110	98	208

em que x_{ij} representa o número de indivíduos pertencentes à categoria i, j . Assim $X_{i1} \sim \text{Binomial}(n_i, \theta_{i1})$, $i = 1, 2$. Posto isso, temos que a razão de chances (do indivíduo ser submetido à tonsilectomia em relação à não ser) tendo ele a doença de Hodgkin em relação à um indivíduo que não tem é dado por:

$$RC = \frac{\theta_{11}(1 - \theta_{11})}{\theta_{21}(1 - \theta_{21})}.$$

Utilize as proporções amostrais (estimadores de máxima verossimilhança de θ_{i1}) para estimar as probabilidades de interesse. Sugestão: pesquise sobre o RC no livro de regressão do Prof. Gilberto Paula, Modelo de Regressão e Aplicações. Responda os itens:

- Obtenha a estimativa de máxima verossimilhança da razão de chances e seu respectivo erro-padrão assintótico. Obtenha a distribuição assintótica do estimador em questão. Sugestão: lembre-se do método Delta.
 - Através do bootstrap paramétrico (sob o modelo em questão e de posse das estimativas de máxima verossimilhança) gere $R = 5000$ amostras da tabela acima. Calcule a estimativa, o erro-padrão e um intervalo de confiança para o RC, usando as respectivas estimativas bootstrap. Obtenha também a distribuição empírica do estimador.
 - Compare os resultados acima. Eles concordam entre si ou não? Qual dos dois resultados você escolheria para fazer inferências com respeito à RC? Justifique, adequadamente sua resposta.
4. Os dados do arquivo pesca.dat correspondem a dados sobre técnicas de pesca nos municípios de Santos e Ubatuba. Maiores detalhes podem ser encontrados no livro “Modelos de regressão com apoio computacional”, do Prof. Gilberto Paula. Considere, além da variável resposta (CPUE), as covariáveis frota e ano (notem que elas são categorizadas). Proponha um modelo (que considere interação entre os fatores) para explicar a CPUE em função dessas duas variáveis categorizadas, com distribuição normal inversa e função de ligação log. Utilize o método de máxima verossimilhança com o algoritmo de mínimos quadrados ponderados via função glm no R.
- Utilize o bootstrap paramétrico com $R = 5000$ amostras de tamanho $m = n$, conforme visto em sala de aula. Obtenha as distribuições empíricas, estimativas pontuais, erros-padrão e intervalos de confiança, para cada um dos parâmetros do modelo. O que você pode dizer sobre a normalidade assintótica dos estimadores dos parâmetros em questão?

- b) De acordo com os resultados do item a), o que você pode dizer sobre a interação e os efeitos principais?
- c) Compara os resultados via máxima verossimilhança e com os resultados bootstrap. Eles são concordantes?
- d) Utilize o gráfico de envelope para verificar a suposição de normalidade inversa do modelo.