

OBS: Os exercícios referentes à análise de correspondência seguem a notação definidas nas notas de aulas disponíveis no site

1. Resolva todos os exercícios deixados em sala, em particular:

Prove que a estatística Q (de qui-quadrado) para testar a independência entre duas variáveis aleatórias numa tabela de contingência, pode ser escrita como:

$$Q = n (\mathbf{P} - \mathbf{P}_E)' \mathbf{D}_{\mathbf{P}_E}^{-1} (\mathbf{P} - \mathbf{P}_E)$$

e que as matrizes \mathbf{Z} e \mathbf{W} são tais que $\mathbf{Z} = \mathbf{W}'$.

2. Considere os dados disponíveis na homepage do curso (“Saúde mental”), relacionados ao estudo sobre a relação entre status socio-econômico dos pais com a condição de saúde mental de 1760 indivíduos. Ou seja, cada um desses indivíduos foi classificado de acordo com o status sócio-econômico dos pais bem como sua saúde mental. Os dados encontram-se também abaixo:

Saúde mental	Status sócio-econômico dos pais				
	A (alto)	B	C	D	E (baixo)
Boa	121	57	72	36	21
Presença fraca de sintomas	188	105	141	97	71
Presença moderada de sintomas	112	65	77	54	54
Debilitado	186	60	94	78	71

Responda os itens:

- a) Utilize o teste de qui-quadrado usual para testar a hipótese de independência entre as duas variáveis aleatórias. Escreva as hipóteses em termos dos parâmetros envolvidos, apresente a estatística do teste, o p-valor correspondente e escreva, de modo adequado, sua conclusão. Utilize $\alpha = 5\%$.
- b) Calcule os perfis de linhas e colunas e os interprete da forma mais abrangente possível. Coloque os resultados em tabelas adequadas.

- c) Faça uma análise de correspondência do modo mais amplo possível. Apresente a tabela com a inércia e a proporção da variabilidade explicada pelas duas primeiras componentes. Apresente o “bi-plot”. Interprete os resultados obtidos da forma mais ampla possível, tendo em mente os objetivos da pesquisa.
3. Este exemplo é sobre uma pesquisa feita na internet acerca de formas de pagamento. As pessoas davam sua opinião sobre a segurança de algumas formas de pagamento (não se sabe sobre quantas formas de pagamento cada pessoa opinava). Os dados estão apresentados na tabela a seguir:

	Tipo de Pagamento Total					
	Cartão	Boleto	Din. Elet.	Vale	Dep. Banc.	Cheque
Inseguro	882	113	621	316	339	428
Pouco seguro	704	225	733	570	461	561
Razoavel seguro	904	1103	873	961	894	871
Muito Seguro	184	1185	287	623	862	68
Total	2674	2626	2514	2470	2556	2554

Responda os itens:

- a) Utilize o teste de qui-quadrado usual para testar a hipótese de independência entre as duas variáveis aleatórias. Escreva as hipóteses em termos dos parâmetros envolvidos, apresente a estatística do teste, o p-valor correspondente e escreva, de modo adequado, sua conclusão. Utilize $\alpha = 5\%$.
- b) Calcule os perfis de linhas e colunas e os interprete da forma mais abrangente possível. Coloque os resultados em tabelas adequadas.
- c) Faça uma análise de correspondência do modo mais amplo possível. Apresente a tabela com a inércia e a proporção da variabilidade explicada pelas duas primeiras componentes. Apresente o “bi-plot”. Interprete os resultados obtidos da forma mais ampla possível, tendo em mente os objetivos da pesquisa.
4. (Os dados se encontram no arquivo Salmon.txt) A fronteira do Canadá com o Alasca é uma área de pesca de salmão. Sabidamente, os salmões nascem em água doce mas migram para o mar retornando, posteriormente, para o local onde nasceram, para fins de reprodução. Existem basicamente duas espécies de salmão nessa região: uma que nasce no Alasca (doravante grupo 1) e outra que nasce no Canadá (doravante grupo 2). Existe um acordo entre Estados Unidos e Canadá, de modo a proibir a pesca de um tipo de salmão por pescadores do outro país. Ou seja, pescadores canadenses devem pescar somente salmões provenientes do Canadá e o mesmo vale para pescadores do

Alasca. Para se ter algum controle sobre isso, pretende-se criar uma regra de classificação utilizando duas variáveis medidas em 50 salmões provenientes do Alasca e em 50 salmões provenientes do Canadá. Tal regra visa poder identificar mais facilmente a origem de salmões pescados. As variáveis medidas são : X_1 = diâmetro das guelras durante a fase em água doce (em mm) e X_2 = diâmetro das guelras durante a fase no mar (em mm). Considere, para construir tal regra, distribuição normal bivariada para os dois grupos, homocedasticidade, custos iguais de classificação errada e probabilidades a priori de classificação iguais. Responda os itens: