

Trabalho de Analise Multivariada de Dados

Geovani Finoti Leitão

Exercício 1

- Foi analisado a matriz de variância total explicada, onde percebe-se que possui apenas um componente explicativo, no qual possui 68% de poder de explicação.
- Foi analisado a matriz de componentes, onde percebe-se que todas as variáveis possuem valores altos no componente 1, sendo que a variável “Ganho Bruto” possui o maior valor, se destacando um pouco das demais. Já no componente 2 as variáveis “Ganho Liquido” e “Patrimônio” possuem os maiores valores, porém significativamente abaixo do componente 1. E por fim no componente 3 os valores são abaixo dos demais componentes e muito próximos um do outro.

Exercício 2- Saída AF_Exercício

- Foi analisado a matriz de correlação, onde se percebeu que as variáveis apresentam bons níveis de correlação, apresentando pouquíssimos fatores neutros entre si.
- Foi analisado o Teste de KMO, no qual apresentou um resultado de 0,83 que representa uma BOA análise de componentes principais.
- Foi analisado a matriz de correlação anti-imagem no qual é possível verificar que toda a diagonal possui valores superior a 0,5, aceitando como um bom relacionamento entre as variáveis.
- Foi analisado a matriz de comunalidades onde foi verificado que as variáveis apresentam bons percentuais de explicação, com exceção de 3 variáveis (Mortalid.Infantil, Mort. CausasExternas e População Favelada) que apresentaram percentuais abaixo de 0,5, porém não foi necessário remover as da análise pois o percentual na matriz de anti-imagem das mesmas estava aceitável.
- Foi analisado a matriz de variância total explicada, onde percebe-se que foi dividido em dois componentes explicativos, onde o primeiro é capaz de explicar 49% e o segundo 18%, gerando um total 68% de poder de explicação.
- Analisando a matriz de componentes, iremos utilizar para o primeiro componente, nomeado como Perfil Demografico, as variáveis RendaFamiliar, QuotaRes, AnosdeEstudo, Perfil Etária,

Mortalid.Infantil e Mort.CausaExternas. Já para o segundo componente, nomeado como Densidade Populacional, foi utilizada apenas a variável Densid.PoP. Foram selecionadas apenas as variáveis com índices maiores que 0,6, pois a base analisada possui cerca de 96 observações.

Exercício 3 - Saída AD_PMRV_ML_V_End

- Foi analisado o Lambda de Wilks onde foi identificado que a variável Vendas possui a maior poder discriminatório sobre os dados disponibilizados.
- Foi analisado o a matriz Resultado do teste onde foi verificado que há variação significativa entre os grupos, pois a significância calculada foi de 0,646, consideravelmente maior que 0,05.
- Analisando a matriz de autovalores foi percebido que a primeira função tem um poder discriminatório de 99,2%.
- Analisando o Lambda de Wilks para as funções, percebe-se que a primeira função possui o maior poder discriminatório sobre as demais.

Coeficientes de função discriminante canônica			
	Função		
	1	2	3
PMRV	-,008	,021	-,023
Endividamento	-,005	,060	,007
Vendas	,002	,000	,000
Margem_Líquida	,018	,131	,392
(Constante)	-7,944	-2,782	-3,291

Coeficientes não padronizados

Classificar uma nova empresa que tenha:

- PMRV: 28 dias
- Endividamento: 32%
- Vendas: 4348,00 (R\$ x mil)
- Margem Líquida: 9,8%

$$Z1 = -7,944 + (28 * -0,008) + (32 * -0,005) + (4348 * 0,002) + (9,8 * 0,018) = 0,5444$$

$$Z2 = -2,782 + (28 * 0,021) + (32 * 0,060) + (4348 * 0) + (9,8 * 0,131) = 1,0098$$

$$Z3 = -3,291 + (28 * -0,23) + (32 * 0,007) + (4348 * 0) + (9,8 * 0,392) = -5,6654$$

R = Empresa pertence ao grupo 2, conforme resultado apresentado no Z2

E outra empresa com os seguintes dados:

- PMRV: 22 dias
- Endividamento: 22 %
- Vendas: 6348,00 (R\$ x mil)
- Margem Líquida: 12,3%

$$Z1 = -7,944 + (22 * -0,008) + (22 * -0,005) + (6348 * 0,002) + (12,3 * 0,018) = 4,6874$$

$$Z2 = -2,782 + (22 * 0,021) + (22 * 0,060) + (6348 * 0) + (12,3 * 0,131) = 0,6113$$

$$Z3 = -3,291 + (22 * -0,23) + (22 * 0,007) + (6348 * 0) + (12,3 * 0,392) = -3,3754$$

R = Empresa pertence ao grupo 1, conforme resultado apresentado no Z1

Exercício 4 - Saída RL_1

- Foi analisado o Teste de Omnibus do Modelo de coeficiente e visto que há pelo menos uma variável que discrimine o modelo.
- Foi analisado o valor R quadrado de Nagelkerke e visto que possui 43% de poder explicativo em cima da base analisada.
- Foi analisado a Tabela de classificação onde pode ser visto que o modelo acertou de modo geral 75,6% dos casos avaliados. Sendo que classificou de maneira correta 36 casos dos 50 possíveis para o pagamento igual a 0, e 100 casos dos 130 possíveis para o pagamento igual a 1.
- Foi analisado nas variáveis na equação que todas as variáveis possuem um grau de significância alta para o modelo.
- Foi analisado a Matriz de correlação que a variável idade e a variável constante possuem um bom grau de correlação.
- Foi analisado o gráfico de probabilidades preditas e visto que não há uma boa distribuição das informações.
- Foi analisado a lista entre casos e visto que há 7 casos onde os mesmos deveriam pertencer a outro grupo, dada suas características.

Formula da equação: $-3,616 - 1,301 * \text{Sexo} + 0,116 * \text{Idade} + 2,951 * \text{EstadoCivil}$