

Avaliação do teste ANOCVA
para comparação de clusteres
através de simulações

Davi Augusto Caetano de Jesus

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Mestrado em Estatística
Orientador: Prof. Dr. Alexandre G. Patriota

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CNPq

São Paulo, fevereiro de 2017

Avaliação do teste ANOCVA
para comparação de clusteres
através de simulações

Esta é a versão original da dissertação elaborada pelo
candidato Davi Augusto Caetano de Jesus, tal como
submetida à Comissão Julgadora.

Resumo

Caetano de Jesus, D. A. **Avaliação do teste ANOCVA para comparação de clusters através de simulações**. 2017. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

ANOCVA (*Analysis Of Cluster Structure Variability*) é um teste estatístico proposto para comparação da estrutura de clusterização entre populações. Esse teste pode ser aplicado em ressonância magnética funcional para testar uma das hipóteses da neurociência que afirma que a presença ou não de uma desordem psiquiátrica pode ser explicada na forma com que os neurônios formam clusters no cérebro. O presente trabalho tem como objetivo expandir as simulações feitas com esse teste para estudar suas limitações e propriedades quando aplicados em situações reais. Para isso as simulações já publicadas foram reproduzidas e posteriormente expandidas através da alteração de parâmetros que originalmente foram considerados constantes, a saber: variância e número de *voxels*. Um novo teste marginal foi proposto através de uma alteração na hipótese nula com o objetivo que a estatística do teste cumpra algumas propriedades esperadas.

Palavras-chave: clusterização, fMRI, teste estatístico.

Abstract

Caetano de Jesus, D.A. **Evaluation of ANOCVA test for cluster comparison through simulation.** 2017. Master's thesis - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

ANOCVA (Analysis Of Cluster Structure Variability) is a statistical test proposed to compare the structure of clustering between populations. This test can be applied to Functional Magnetic Resonance Imaging (fMRI) to test one of the hypotheses of neuroscience that states that the presence or absence of a psychiatric disorder can be explained through the clusterization structures formed by neurons in the brain. The present work aims to present some simulations with the purpose to study its properties and properties when the method is applied in actual situations. Previously published simulations are reproduced and also expanded by varying some parameters that were previously considered constant, namely, variance and number of voxels. A new marginal test was proposed through a changing in the null hypothesis for the test statistic to hold some expected properties.

Keywords: clustering, fmri, statistical test.

Sumário

1	Introdução	1
2	Revisão de Conceitos	3
2.1	FMRI	3
2.2	Clusterização	4
2.3	Estatística de <i>Silhouette</i>	7
3	ANOCVA	10
3.1	Definição das estatísticas de teste	12
3.2	Teste estatístico	13
3.3	Estatística marginal proposta	15
4	Simulações	17
4.1	Simulações originais	17
4.2	Simulações propostas	18
4.3	Análise das estatísticas para variáveis degeneradas	19
4.3.1	Teste marginal original	21
4.3.2	Teste marginal proposto	22
4.4	Simulações	23
4.4.1	Simulações: $N_1 = 10$ e $N_2 = 1$ com distância Euclidiana	24
4.4.2	Simulações: $N_1 = 10$ e $N_2 = 2$ com distância Euclidiana	26
4.4.3	Simulações: $N_1 = 10$ e $N_2 = 5$ com distância Euclidiana	28
4.4.4	Simulações: $N_1 = 10$ e $N_2 = 1$ com distância Euclidiana padronizada por desvio absoluto médio	30
5	Conclusões e estudos futuros	32
5.1	Conclusões	32
5.2	Estudos futuros	33
A	Fórmulas	34
A.1	Desenvolvimento das fórmulas da estatística do teste	34
A.2	Generalização das fórmulas da estatística do teste	36
B	Gráficos	41
B.1	Tabela resumo dos gráficos	42
B.2	Gráficos	50

B.2.1	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste conjunto	50
B.2.2	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 11$	52
B.2.3	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 11$	54
B.2.4	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 1$	56
B.2.5	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 1$	58
B.2.6	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 12$	60
B.2.7	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 12$	62
B.2.8	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste conjunto	64
B.2.9	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 11$	66
B.2.10	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 11$	68
B.2.11	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 1$	70
B.2.12	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 1$	72
B.2.13	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 13$	74
B.2.14	Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 13$	76
B.2.15	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste conjunto	78
B.2.16	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 11$	80
B.2.17	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 11$	82
B.2.18	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 1$	84
B.2.19	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 1$	86
B.2.20	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 16$	88
B.2.21	Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 16$	90

C Gráficos 92

C.1	Tabela resumo dos gráficos para as simulações com distância Euclidiana padronizada pelo desvio absoluto médio	93
C.2	Gráficos	95
C.2.1	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste conjunto com distância Euclidiana padronizada pelo desvio absoluto médio	95
C.2.2	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 11$ com distância Euclidiana padronizada pelo desvio absoluto médio	97
C.2.3	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 1$ com distância Euclidiana padronizada pelo desvio absoluto médio	99
C.2.4	Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 12$ com distância Euclidiana padronizada pelo desvio absoluto médio	101

Referências Bibliográficas

Capítulo 1

Introdução

Avanços nas tecnologias de sensores digitais, comunicação, computação e armazenamento de dados veem provocando um aumento enorme na quantidade de dados nos últimos anos (Bryant et al., 2008). Relatório da IBM de 2012 afirma que 80% de todos os dados foram gerados nos últimos dois anos, e que 90% desses dados nunca foram analisados (ibm, 2012).

Nas ciências médicas também há um crescimento acelerado na coleta dos dados, provocando um crescimento na demanda por técnicas revolucionárias para gerenciar, analisar e acessar essas informações (Howe et al., 2008). A biologia molecular está intimamente envolvida com o crescimento da quantidade de informação. O GenBank, por exemplo, é uma base de dados pública construída pelo Centro Nacional de Informação de Biotecnologia (*National Center for Biotechnology Information - NCBI*) que recebe diferentes tipos de informação sobre sequenciamento genético de diversas fontes. Assim como a biologia molecular, a neurociência, através do uso de ressonância magnética funcional (*Functional Magnetic Resonance Imaging- FMRI*), enfrenta um aumento acelerado da quantidade de dados.

Nesse contexto as técnicas para análise de dados vem se tornando ainda mais relevantes. Em medicina, por exemplo, a identificação de características que caracterizam sujeitos portadores ou não de determinada doença pode ser a chave para a determinação de procedimentos de diagnóstico, prevenção e tratamento (Rubinov and Sporns, 2010). Frequentemente o diagnóstico de uma determinada doença pode ser interpretado como um teste estatístico no qual compara-se o paciente testado com as populações que apresentam ou não a enfermidade.

O presente trabalho tem como objetivo analisar os resultados do teste estatístico denominado ANOCVA (Fujita et al., 2014b), que tem como uma de suas aplicações a comparação de populações portadoras ou não de determinada enfermidade. Assim como o teste ANOVA compara dois ou mais grupos através das análise da influência de um ou mais fatores na média, o ANOCVA tem como objetivo comparar grupos, porém através da comparação da forma com que as características se

agrupam.

Várias técnicas vem sendo utilizadas para estudar o dados resultantes da Ressonância Magnética Funcional por Imagem, como é o caso do método de entropia de redes (Sato et al., 2013) nos casos de pacientes portadores de Transtorno de Déficit de Atenção e Hiperatividade (TDAH). A entropia de redes indica que a estrutura da rede neuronal dos pacientes diagnosticados com TDAH é diferente. Nessas pessoas, o cérebro apresenta uma desorganização do funcionamento dos circuitos neuronais. O teste ANOCVA, como será apresentado nesse estudo, também pode ser usado para estudar dados de FMRI. Nesse caso ele é usado para identificar a diferença entre populações e regiões específicas do cérebro envolvidas com essa diferença.

O objetivo de técnicas como essas é de ajudar no desenvolvimento de tratamentos e de estabelecer critérios objetivos para o diagnóstico de doenças como o TDAH. Os critérios atuais específicos são determinados pelo comportamento e podem ser considerados subjetivos (De la Peña Olvera et al., 2010) e critérios baseados em imagens poderiam reduzir erros de diagnóstico.

A capítulo 2 revisa os conceitos nos quais o teste ANOCVA se baseia. A seção 2.1 apresenta conceitos sobre Ressonância Magnética Nuclear, que é uma das principais fontes de dados para aplicações do teste ANOCVA. A aplicação desse teste está bastante relacionada com técnicas de clusterização. A seção 2.2 apresenta informações sobre clusterização utilizadas no teste ANOCVA. O teste ANOCVA faz uso da estatística de Silhouette para o cálculo de sua estatística de teste. Na seção 2.3 a estatística de Silhoutte é apresentada.

O capítulo 3 do presente estudo apresenta o teste ANOCVA, quais são as hipóteses nulas e alternativas, como ele é calculado e como se deve proceder para encontrar a distribuição do teste sob hipótese nula. Nesse capítulo uma estatística de teste alternativa é apresentada.

O capítulo 4 apresenta as simulações feitas e analisa seus resultados. A seção 4.1 descreve detalhadamente as simulação feitas anteriormente enquanto a seção 4.3 apresenta uma maneira mais genérica de descrever as simulações. A seção 4.3 analisa e compara os testes através do comportamento das estatísticas de teste quando a variância é nula. A seção 4.4 apresenta os resultados dos testes aplicados em simulações.

O capítulo 5 sumariza o trabalho, aponta seus resultados e indica possibilidades de estudos que podem ser feitos para uma análise mais profunda do teste ANOCVA e possíveis alternativas que podem produzir outros resultados significantes.

O apêndice A apresenta os desenvolvimentos das fórmulas utilizadas nesse estudo. O apêndice C mostra os gráficos gerados através das simulações.

Capítulo 2

Revisão de Conceitos

2.1 FMRI

As propriedades da Ressonância Magnética Nuclear (RMN) foram descritas pela primeira vez por Purcell (Purcell et al., 1946) e Bloch (Bloch, 1946) em 1946 em trabalhos pelos quais eles receberam Prêmio Nobel em 1952. Desde então a RMN se tornou uma ferramenta para investigar a estrutura da matéria, como feito em 1973 por Mansfield (Mansfield and Grannell, 1973).

Os núcleos dos átomos são formados por prótons e nêutrons. Esse núcleo, quando tem número par de prótons e nêutrons, não apresenta momento magnético e por isso não pode ser observado através de RMN (Crosson et al., 2010). Entretanto, os átomos que apresentam momentos magnéticos podem ser alinhados de forma paralela ou anti-paralela quando submetidos a um campo magnético. Se esse campo apresenta uma determinada frequência, que é denominada frequência de Larmor, os spins podem absorver energia necessária para alterar o nível de estado de energia. Esse processo gera uma oscilação no campo magnético aplicado. Quando o campo magnético induzido é cessado, o campo magnético medido demora um intervalo de tempo para cessar. Esse intervalo de tempo está relacionado com o material no qual o campo magnético é aplicado (Babiloni et al., 2009).

Na década de 1980, Behar (Behar et al., 1983), Rothman (Rothman et al., 1984), Hetherington (Hetherington et al., 1985) e Hanstock (Hanstock et al., 1987) demonstraram ser possível obter imagens detalhadas do tecido de cérebro de ratos utilizando ressonância magnética. Nos anos 1990, foram realizados os primeiros experimentos que introduziram a Ressonância Magnética Funcional por Imagem (*Functional Magnetic Resonance Imaging - FMRI*). Esse procedimento, que se caracteriza por ser não invasivo, rapidamente ganhou espaço nas pesquisas em neurociências (Ogawa et al., 2000).

A forma mais comum de utilização da FMRI é através dos contrastes BOLD (*blood oxygen level*

dependent contrast) (Yamazaki et al., 1990), que detecta alterações no conteúdo das hemoglobinas induzidas por alterações no fluxo sanguíneo cerebral e na taxa de consumo de oxigênio que ocorrem em consequência das variações na atividade cerebral. No século XIX já se acreditava que haveria um complexo mecanismo que alteraria o fluxo sanguíneo cerebral para enviar maiores suprimentos de oxigênio às áreas do cérebro que estivessem apresentando maior atividade (Roy and Sherrington, 1890).

A forte correlação entre a ativação neural e taxa de consumo de oxigênio foi demonstrada posteriormente através de outros estudos mais precisos (Raichle, 1987; Kwong et al., 1992; Zhu et al., 1998). Baseados nessa correlação, a amplitude e o período do sinal da ressonância vem sendo extensivamente usadas para detalhamento da informação sobre os eventos cerebrais (Ogawa et al., 2000).

A função básica da FMRI é colher imagens BOLD de um indivíduo enquanto este realiza atividades pré-determinadas (Le Bihan et al., 1995). Tipicamente é colhida uma imagem a cada intervalo de 2 segundos de regiões volumétricas com quantidade na ordem das centenas, sendo que a medição total usualmente varia de 2 a 10 minutos. Há ainda a necessidade de um processamento estatístico de filtragem dos resultados dado a dificuldade de medição da variação do sinal da ordem de 1% a 5% quando medido em um scanner com campo magnético de 1,5T, devido as perturbações naturais (Purdon and Weisskoff, 1998).

As bases de dados utilizadas nas aplicações práticas do presente estudo são FMRI de imagens BOLD extraídas de banco de dados públicos. Maiores detalhes desses dados são apresentados no capítulo 3.

2.2 Clusterização

Clusterização se refere a uma gama de técnicas cujo o objetivo é encontrar subgrupos, ou clusteres, no conjunto de dados (James et al., 2013). Cada um desses subgrupos de dados devem conter observações que sejam similares umas às outras, quando dentro do mesmo subgrupo, ou que sejam pouco similares umas às outras, quando em subgrupos diferentes, segundo alguma métrica de similaridade ou dissimilaridade. É comum em análise de dados recorrer ao que se chama de análise de cluster mesmo quando não existe nenhuma hipótese sobre os dados para auxiliar na fase exploratória da pesquisa.

A seguir é apresentado um exemplo de clusterização de relevância prática para o presente estudo. Como foi apresentado na seção 2.1, os resultados de exames de FMRI apresentam a medida da

quantidade de atividade para determinada região ou *voxel*. Os dados são apresentados através de séries temporais com a mesma quantidade de elementos. Assume-se que o cérebro seja dividido em 20 regiões. Assume-se ainda que os resultados da fMRI para as primeiras 10 regiões sejam séries temporais idênticas entre si e que os resultados para as últimas 10 regiões sejam séries temporais idênticas entre si porém diferente das séries temporais das 10 primeiras regiões. Assim, se a diferença entre as séries for suficiente, espera-se que um algoritmo de clusterização seja capaz de clusterizar as regiões em dois clusters, sendo um deles contendo as 10 primeiras regiões e o outro contendo as 10 últimas regiões. A figura 3.1 ilustra esse exemplo.

Pela praticidade e aplicabilidade, a clusterização se tornou bastante popular (James et al., 2013). O problema de clusterização é classificado como um problema de aprendizado não supervisionado, pois ele não tenta prever uma variável ou característica existente, mas encontrar estrutura nos dados. Isso fez com que surgissem muitos métodos devido à ideia de clusterização ser abstrata.

Dois dos principais métodos de clusterização são o método hierárquico e o método K-Means (Steinbach et al., 2000). No método K-Means, primeiramente define-se o número de clusters e então tenta-se encontrar a clusterização que minimiza as distâncias entre cada observação e o cluster no qual a observação foi determinada. Essa distância é denominada dissimilaridade e será assim tratada nesse estudo. Como a complexidade do problema é alta, usa-se um algoritmo que é capaz de encontrar um solução local partindo de uma inicialização onde se determina de maneira aleatória o cluster no qual cada observação pertence inicialmente. Nesse caso o problema se torna linear com relação ao número de elementos, mas pode ser que a solução encontrada não seja a ótima.

Seja $\Omega = \{\omega_1, \dots, \omega_N\}$ um conjunto no qual será aplicado um algoritmo de clusterização que, segundo algum critério, define os clusters $\mathcal{C} = \{C_1, \dots, C_r\}$ em que cada $C_i, i = 1, \dots, r$ é um conjunto ou cluster não vazio tal que $C_i \cap C_j = \emptyset$ para $i \neq j$ e $\cup_{i=1}^r C_i = \Omega$.

O método K-Means tenta encontrar a clusterização que minimiza a dissimilaridade de cada cluster. Se denotarmos por $d(C_k)$ a medida da dissimilaridade do cluster k , o problema pode ser definido através de:

$$\underset{C_1, \dots, C_r}{\text{minimizar}} \left\{ \sum_{k=1}^r d(C_k) \right\}. \quad (2.1)$$

Para resolver o problema é necessário definir a medida de dissimilaridade de cada cluster. Existem diversas maneiras através das quais isso pode ser feito, sendo a mais comum a distância Euclidiana. A função de dissimilaridade dentro do cluster k é definida por:

$$d(C_k) = \frac{1}{|C_k|} \sum_{\omega_i, \omega_j \in C_k} \|\omega_i - \omega_j\|, \text{ em que } \|x\| = \sqrt{(x^\top x)} \text{ é a norma Euclidiana para vetores. (2.2)}$$

Assim o problema se torna:

$$\text{minimizar}_{C_1, \dots, C_r} \left\{ \sum_{k=1}^r \frac{1}{|C_k|} \sum_{\omega_i, \omega_j \in C_k} \|\omega_i - \omega_j\| \right\}. \quad (2.3)$$

É necessário encontrar algum algoritmo para resolver esse problema. No entanto é fácil perceber que a ordem de magnitude desse problema é K^n , o que torna uma solução analítica muito difícil a não ser que o valor de n seja muito pequeno. Há porém um algoritmo relativamente simples que é capaz de fornecer uma solução ótima local a esse problema de otimização. Esse algoritmo consiste de dois passos:

1. Determina-se inicialmente para cada observação um cluster de 1 até K no qual essa observação será designada.

2. Repete-se os seguintes passos até que a repetição dos passos não produza alterações nos clusters designados a cada observação:

a) Determina-se o centroide de cada cluster, sendo cada característica do centroide definido como a média daquela característica dentre todas as observações pertencentes ao cluster.

b) Designa-se cada observação ao cluster no qual a distância entre a observação e o centroide do cluster calculado em (a) é mínima.

Pode-se provar que esse algoritmo faz com que a medida definida na equação 2.3 seja decrescente a cada passo (James et al., 2013).

O método K-Means exige que o número de clusters seja conhecido previamente, o que pode ser considerado uma desvantagem. Se houver alguma restrição por conta dessa característica outros métodos podem ser usados, dentre eles o método hierárquico.

O método hierárquico tem um algoritmo relativamente simples de ser seguido. É necessário a definição de uma medida de dissimilaridade entre os elementos, que em geral pode ser a distância Euclidiana. Calcula-se então a distância entre todos os possíveis pares de observações. O par de observações que tiver a menor distância entre si formará o primeiro cluster. Para que se possa seguir com o algoritmo, é necessário definir a distância entre uma observação e um cluster, que pode ser calculado através do centroide do cluster da mesma forma que foi feito no método K-Means. Então é calculada novamente a distância entre cada par de observações e/ou cada cluster. O par que tiver

menor distância é novamente fundido em um novo cluster. Nota-se que inicialmente o número de clusteres é igual ao número de observações e que a cada passo o número de clusteres se reduz em uma unidade. Então, após $n - 1$ passos haveria apenas 1 cluster que conteria todas as observações. Para que isso não aconteça é necessário que se use um critério de parada, que pode ser o número de clusteres definido inicialmente, ou um limite para a dissimilaridade média entre os clusteres (Friedman et al., 2001).

Uma característica comum desses dois métodos de clusterização é que em algum momento é preciso determinar o número de clusteres do problema, seja para definir o critério que encerra as iterações, no caso hierárquico, seja no início do problema, no caso do K-Means (Hartigan, 1975). Em alguns casos a análise gráfica da dispersão dos dados pode ajudar na definição do número de clusteres, porém isso só pode ser feito caso a dimensão dos dados seja limitada (Steinbach et al., 2000). Para ajudar na solução desse problema foram sugeridas algumas técnicas, como o método slope (Fujita et al., 2014a), critério de informação Bayesiana (Celeux and Govaert, 1992), índice de Calinski e Harabasz (Caliński and Harabasz, 1974), entre outros.

2.3 Estatística de *Silhouette*

A Estatística de *Silhouette* foi proposta em 1987 (Rousseeuw, 1987) para determinar o número de clusteres mais apropriado para um dado algoritmo de clusterização. Essa estatística quantifica o quanto um elemento está classificado corretamente em uma clusterização. De acordo com Rousseeuw (1987), através do cálculo dessa estatística para diferentes quantidades de clusteres, a quantidade ideal seria a que maximiza a estatística de *Silhouette* média para todos os elementos. Esse método apresenta bons resultados quando os clusteres são homogêneos, ou seja, eles tem aproximadamente a mesma variabilidade interna. Uma solução para problemas com clusteres não homogêneos é apresentada em Fujita et al. (2014a). A seguir é descrito o método da estatística de *Silhouette*.

Seguindo a notação introduzida na seção 2.2, seja uma função de dissimilaridade $d(\omega_i, \omega_{i'})$ entre dois elementos do conjunto Ω que pode ser Euclidiana, Manhattan, etc, define-se a dissimilaridade entre um elemento ω_i e um cluster C_j como

$$d(\omega_i, C_j) = \frac{1}{|C_j|} \sum_{\omega_{i'} \in C_j} d(\omega_i, \omega_{i'}), \quad (2.4)$$

em que $|C_j|$ é o número de elementos no cluster C_j . Define-se por $D_q \in \mathcal{C}$, $q = 1, \dots, N$ o cluster no qual o elemento ω_q foi designado através do algoritmo e por $E_q \in \mathcal{C} - \{D_q\}$, $q = 1, \dots, N$ o cluster

mais próximo no qual o elemento não foi designado. Define-se as quantidades

$$a_q = d(\omega_q, D_q) \text{ e } b_q = \min_{E_q} d(\omega_q, E_q), \quad q = 1, \dots, N. \quad (2.5)$$

A estatística de *Silhouette* é definida então por:

$$s_q = \begin{cases} \frac{b_q - a_q}{\max\{b_q, a_q\}} & \text{se } |D_q| > 1, \\ 0 & \text{se } |D_q| = 1. \end{cases} \quad (2.6)$$

Nota-se que a estatística de *Silhouette* fica contida no intervalo $[-1, 1]$, e que b_q e a_q não podem ser ambos nulos no mesmo caso se existem pelos menos dois clusteres.

Ela é próxima de 1 se b_q é muito maior que a_q . Ou seja, se o elemento ω_q está muito mais próximo ao cluster no qual ele foi designado se comparado à melhor segunda opção. Isso é um indicativo de que o elemento ω_q foi clusterizado de maneira apropriada, visto que o objetivo dos algoritmos de clusterização é agrupar elementos próximos entre si.

A estatística de *Silhouette* tem valor próximo a 0 se os valores b_q e a_q são próximos. Isso indica que o elemento ω_q tem distâncias muito próximas aos clusteres em questão e que a designação do cluster pode ter ocorrido em função de fatores aleatórios. O valor igual a 0 pode ocorrer também quando o número de elementos em um determinado cluster é igual 1, como fica definido na segunda condição da equação (2.6).

Por outro lado, se a estatística de *Silhouette* é um valor próximo a -1 , há uma indicação de que a_q é muito menor em comparação com b_q , ou seja, o elemento ω_q foi clusterizado de maneira não satisfatória segundo esse critério.

Observa-se que os algoritmos de clusterização mais utilizados produzem clusteres com valores de estatística de *Silhouette* em geral positivos, pois os algoritmos mais comuns ou tentam atribuir o cluster ao elemento de forma a diminuir a distância do elemento ao centroide, ou utilizam outro algoritmo que equivale a minimização dessas distâncias.

Outra observação importante pode ser feita com relação ao valor da estatística para elementos com $|D_q| = 1$. Se fosse utilizada a fórmula para elementos com $|D_q| > 1$, o valor da estatística seria 1, pois a medida de distância induziria $a_q = 0$. Se essa definição para elementos com $|D_q| = 1$ não fosse feita, a maximização sempre ocorreria de forma que cada elemento fosse um cluster (a estatística de *Silhouette* seria 1 para todos os elementos). De fato, vários algoritmos criados para determinar o número de clusteres mais apropriado utilizam punição para o caso em que um cluster possui apenas um elemento. No caso da estatística de *Silhouette*, esse valor é arbitrário e outros

valores poderiam ser utilizados, produzindo resultados diferentes.

Capítulo 3

ANOCVA

O teste ANOCVA (*Analysis of Cluster Structure Variability*) (Fujita et al., 2014b) é um teste para comparação da estrutura de clusterização de duas ou mais populações. Ele utiliza séries temporais das características das observações das populações que se pretende comparar para concluir se há diferença significativa entre as populações através dos dados nessa estrutura.

O teste ANOCVA foi desenvolvido a partir de duas ideias estabelecidas, a saber: a estatística de *Silhouette* e o teste ANOVA. Basicamente o teste ANOCVA usa a estatística de *Silhouette* para definir uma medida de variabilidade da estrutura de cluster em cada população. Define-se então um teste estatístico baseado nessa medida para comparar as populações. A seguir será introduzido um exemplo prático para ilustrar o procedimento do teste.

Como foi descrito, a comparação de populações através da análise dos dados gerados pela coleta de FMRI dos indivíduos observados é um exemplo de aplicação desse teste. Sejam duas populações, a primeira sendo classificada como portadora de TDAH e a segunda sendo classificada como controle, os dados coletados são apresentados por $x_{i,j,q}$. Nesse caso temos duas populações, ou seja, $k = 2$. O índice j representa cada uma das k populações a serem testadas, podendo nesse caso ter-se $j = 1$ ou $j = 2$. Cada população apresenta um número n_j indivíduos observados.

Cada indivíduo foi submetido ao FMRI. Os resultados são apresentados através da medida da intensidade de atividade de diferentes regiões do cérebro para aquele indivíduo através do tempo. O índice q nesse caso representa cada uma das N regiões do cérebro nas quais são feitas as medições da intensidade de atividade. Então $x_{10,2,256}$ representa a série temporal da intensidade da atividade medida na região 256 do indivíduo 10 da população 2.

Os dados $x_{i,j,q}$ podem ser de qualquer dimensão, com a restrição de que se possa calcular uma matriz de dissimilaridades para cada observação i da população j e que essa matriz calculada seja

$N \times N$ da seguinte forma:

$$A_{i,j} = \{d(x_{i,j,q}, x_{i,j,q'})\}, \text{ para } i = 1, \dots, n_j, j = 1, \dots, k. \quad (3.1)$$

Outro exemplo de aplicação poderia ser a comparação de diversas lojas (observações, índice j), em duas diferentes cidades (populações, índice i) com relação a quantidade vendida de uma mesma seleção de produtos (características, índice q) no tempo. Nesse caso os dados $x_{i,j,q}$ seriam apresentados como o número de produtos vendidos, sendo uma série temporal. Os dados desse mesmo exemplo poderiam ser apresentados em forma de série temporal, onde poderia ser medida a quantidade de produtos vendida em cada mês de um determinado ano, por exemplo. Nesse caso a observação $x_{i,j,q}$ seria um vetor de tamanho 12.

É interessante notar que o cálculo da matriz de dissimilaridades $N \times N$ não impõe que os dados $x_{i,j,q}$ tenham a mesma dimensão, podendo nesse exemplo que uma loja que tenha vendido produtos por 6 meses ser comparada com uma loja que vendeu produtos por 12 meses. Em geral, para uma observação i na população j os vetores $x_{i,j,q}$ tem a mesma dimensão, pois isso facilita o cálculo da matriz de dissimilaridades nos casos mais comuns, como o da distância euclidiana, por exemplo.

É importante fazer a correta interpretação do teste ANOCVA para entender de que forma ele funciona. Da mesma forma que o teste ANOVA usa a média entre duas populações para comparar se as populações são estatisticamente diferentes, o ANOCVA tem como objetivo verificar se a estrutura de clusterização das características nas populações são estaticamente diferentes.

No primeiro exemplo citado, o teste ANOCVA faz a comparação na forma com que a medida de atividade das regiões do cérebro se agrupam, ou seja, quais regiões do cérebro trabalham de forma similar em função do tempo. Caso as populações apresentem diferenças, o teste ANOCVA é capaz de mostrar quais são as regiões do cérebro que trabalham de forma diferente em cada um dos grupos.

No segundo exemplo, o teste ANOCVA faz a comparação entre as cidades através da forma como os produtos são vendidos no tempo. Caso o teste indique que as cidades apresentam diferenças na forma como vendem os produtos, o teste pode ainda mostrar quais produtos se diferenciam na forma como as lojas de cada uma das cidades vendem os produtos.

Para que se possa fazer uma ilustração do teste, considere a seguinte situação hipotética sem variabilidade: sejam duas populações cujos indivíduos da mesma população apresentam as mesmas matrizes de dados, cada indivíduo apresenta 5 características medidas durante 8 instantes no tempo. Na figura 3.1, para a primeira população, nota-se que as características 1 e 2 podem ser agrupadas em

um cluster enquanto as características 3, 4 e 5 podem ser agrupadas em outro cluster. Com relação à população 2, nota-se que as características 1, 2 e 3 podem ser agrupadas num mesmo cluster enquanto as características 4 e 5 podem ser agrupadas em outro cluster. Nesse caso espera-se que o teste ANOCVA indique que as populações apresentem diferenças estatisticamente significativas na estrutura de clusterização das suas características e que a característica 3 é responsável por diferenciar as duas populações.

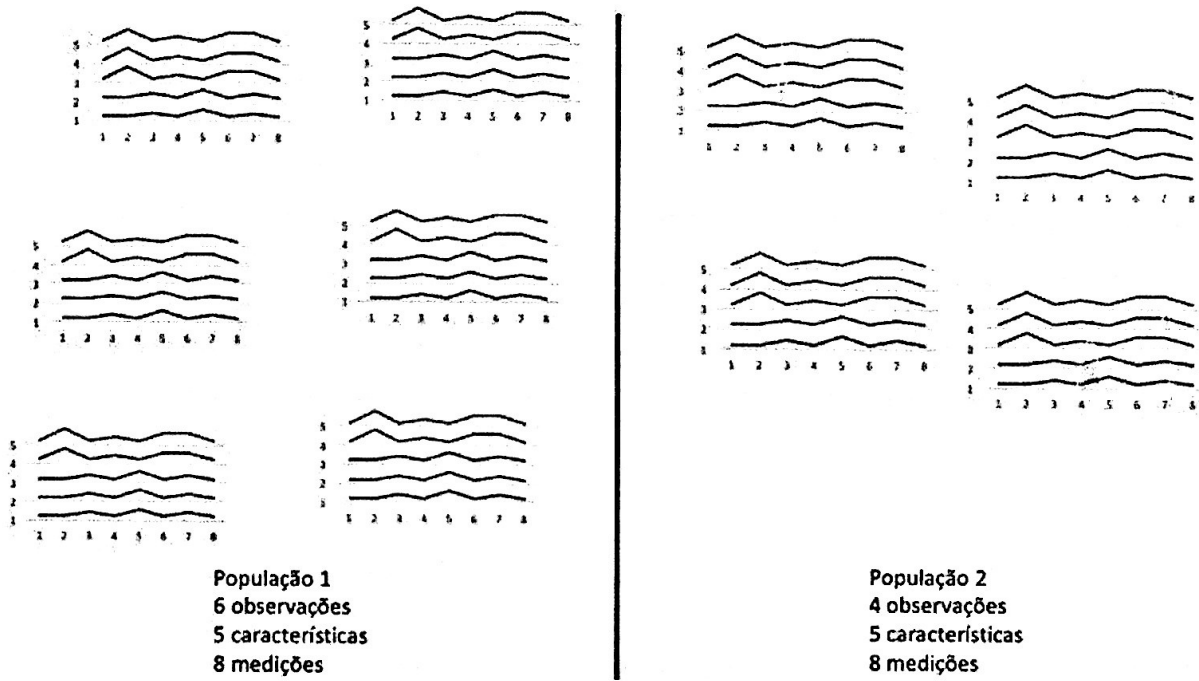


Figura 3.1

A próxima seção aprofunda o exemplo do FRMI para explicar de forma mais detalhada o teste ANOCVA.

3.1 Definição das estatísticas de teste

Conforme descrito na seção anterior, para os valores $x_{i,j,q}$, no exemplo do FMRI, o índice i indica diferentes populações, o índice j representa cada observação dentro da população, o índice q representa cada uma das N regiões do cérebro monitoradas (*voxels*) e cada um dos vetores $x_{i,j,q}$ representa uma série temporal da medida de intensidade da atividade de determinado *voxel* no tempo. Como a dimensão do vetor $x_{i,j,q}$ corresponde ao número de imagens feitas, é possível que cada um dos indivíduos tenha quantidade de imagens diferentes. Isso traz a possibilidade de se utilizar diferentes bancos de dados mesmo que não sejam coletados o mesmo número de imagens de cada indivíduo, conforme foi feito em Fujita et al. (2014b).

É importante salientar que o interesse se encontra na análise das características que diferenciam as populações. Por isso a clusterização é aplicada entre as q características de cada indivíduo, que no caso do FMRI é representado pelos *voxels*, e não entre os indivíduos.

Em seguida fica definida a matriz A_j de dissimilaridades para cada população j que é a média das dissimilaridades entre os elementos da população. Define-se a matriz A de dissimilaridade global, que é a média ponderada das matrizes de dissimilaridades das populações, conforme segue:

$$A_j = \frac{1}{n_j} \sum_{i=1}^{n_j} A_{i,j}, \text{ para } j = 1, \dots, k \text{ e } A = \frac{1}{n} \sum_{j=1}^k n_j A_{i,j}. \quad (3.2)$$

Através da matriz de dissimilaridades A é feita a clusterização através do algoritmo que se julgar mais apropriado, pois ela contém toda a informação necessária para a clusterização. O resultado da clusterização é um vetor l_A com N elementos indicando em qual cluster foi classificado cada elemento $q = 1, \dots, N$. As estatísticas de *Silhouette* são calculadas utilizando o vetor l_A para cada uma das matrizes de dissimilaridades A_j e A conforme definido na seção 2.3. Teremos então $k + 1$ vetores com N elementos cada, a saber:

$$S = (s_1^{(A, l_A)}, \dots, s_N^{(A, l_A)}) \text{ e } S_j = (s_1^{(A_j, l_A)}, \dots, s_N^{(A_j, l_A)}), \quad (3.3)$$

sendo $s_q^{(A, l_A)}$, e estatística de *Silhouette* para o elemento $q = 1, \dots, N$ dada a clusterização l_A usando a matriz de dissimilaridades A e sendo $s_q^{(A_j, l_A)}$, a estatística de *Silhouette* para o elemento $q = 1, \dots, N$ e dada a clusterização l_A usando a matriz de dissimilaridades A_j .

3.2 Teste estatístico

Conforme descrito por Fujita et al. (2014b), dados os vetores S e S_j em (3.3), define-se o vetor

$$\delta S_j = S - S_j. \quad (3.4)$$

Definem-se também as hipóteses nula e alternativa para o teste ANOCVA:

- $H_0 : E(\delta S_1) = \dots = E(\delta S_k) = 0$, que equivale a afirmar que os dados de todas as populações foram clusterizados da mesma forma.
- $H_1 : \text{Ao menos para um dos } j = 1, \dots, k \text{ tem-se } E(\delta S_1) \neq 0$, o que equivale dizer que ao menos um dos itens é classificado de maneira diferente.

A estatística usada para construir o teste é $\Delta S = \sum_{i=1}^k \delta S_j^T \delta S_j$. É esperado que, sob a hipótese nula, o valor resultante da estatística ΔS seja baixo, enquanto que, em caso de hipótese alternativa, o valor da estatística ΔS seja alto. Note que essa estatística tem forma quadrática por ser a soma dos quadrados dos elementos de S_j .

Supondo agora que a hipótese nula foi rejeitada no teste anterior, é necessário identificar os itens que não estão sendo clusterizados da mesma forma em todas as populações. São definidos N testes estatísticos através das estatísticas

$$\delta s_q = s_q^{(A, l_A)} - \frac{1}{k} \sum_{j=1}^k s_q^{(A_j, l_A)} \text{ para } q = 1, \dots, N. \quad (3.5)$$

Note que essa estatística não apresenta forma quadrática, ao contrário da estatística do teste conjunto.

As N hipóteses nulas e alternativas para esses testes são:

- $H_0 : E(\delta s_q) = 0$, que equivale a afirmar que o item q é igualmente clusterizado entre todas as populações.
- $H_1 : E(\delta s_q) \neq 0$, que equivale a afirmar que o item q não é clusterizado igualmente entre todas as populações.

Para testar tais hipóteses os autores definem as estatísticas $\Delta s_q = \delta s_q^2$, $q = 1, \dots, N$. Os autores recomendam que para esse tipo de teste é necessário utilizar-se de fatores de correção para testes múltiplos, tais como Bonferroni (Bonferroni, 1936), por exemplo. No método de Bonferroni, dado um teste com nível de significância global α no qual são realizados M testes consecutivos, cada teste individual deve ser realizado com nível de significância α/M para que seja garantido a condição global.

Os testes conjunto e marginais não identificam em quais populações o item foi clusterizado de maneira diferente. Para obter conclusões dessa natureza seriam necessários testes um contra um, no qual se testa apenas duas populações de cada vez, ou um contra todos, no qual se agrupa na mesma população todas as populações diferentes da qual se está testando.

Devido as dificuldades de determinação das distribuições do teste os autores sugerem o uso de técnicas de *bootstrap* (Efron and Tibshirani, 1994) para a construção empírica de distribuição dos testes sob hipótese nula. Para tal é realizado o processo de reamostragem, que sorteia os n_j elementos para cada uma das j populações dentre todos os elementos da amostra para posteriormente calcular as estatísticas dos testes que estarão sob hipótese nula devido a reamostragem. Dessa maneira a

nova amostra tem o mesmo número de populações, o mesmo número de indivíduos por população para todas as populações, mas está sob a hipótese nula. Repete-se esse processo o número suficiente de vezes e calcula-se a estatística de teste para cada uma dessas novas amostras. O valor-p do teste é então o quantil no qual se encontra a estatística para a amostra original entre todas as amostras de simuladas por *bootstrap*.

3.3 Estatística marginal proposta

Para efeitos de comparação esse trabalho propõe um teste alternativo baseado em uma estatística de teste que tem forma quadrática para verificar se as populações apresentam diferença na forma como suas características são clusterizadas.

Para tal teste o procedimento é bastante similar ao teste ANOCVA. O teste conjunto definido pela estatística (3.4) é mantido da mesma forma. Esse teste indica se as duas populações clusterizam as regiões de forma diferente. Caso essa hipótese seja confirmada, o procedimento proposto sugere que seja utilizada a estatística (3.6) definida por:

$$\delta s_q^* = \sum_{j=1}^k (s_q^{(A_j, I_A)} - s_q^{(A, I_A)})^2 \text{ para } q = 1, \dots, N. \quad (3.6)$$

Note que essa estatística mantém a forma quadrática da estatística conjunta.

As N hipóteses nulas e alternativas para esse teste são:

- $H_0 : E(\delta s_q^*) = 0$, que equivale a afirmar que o item q é igualmente clusterizado entre todas as populações.
- $H_1 : E(\delta s_q^*) \neq 0$, que equivale a afirmar que o item q não é clusterizado igualmente entre todas as populações.

Nota-se que as hipóteses foram alteradas em função das estatísticas para manter o objetivo do teste, que é identificar se um item é clusterizado igualmente entre as populações ou não.

Para esse tipo de teste é necessário utilizar-se de fatores de correção para testes múltiplos, tais como Bonferroni (Bonferroni, 1936), por exemplo.

Portanto, para determinada amostragem com duas populações teremos os seguintes testes:

- Um teste que será denominado “Teste conjunto” que é feito através da estatística (3.4) e compara se há ao menos uma característica dentre as N que diferencia a clusterização entre as populações;

- N testes denominados “Testes marginais originais”, que são feitos através da estatística (3.5) e devem ser feitos caso o “Teste conjunto” rejeite a hipótese nula. Será descrito que o “Teste marginal original” foi realizado para a característica q para $q = 1, \dots, N$.
- N testes denominados “Testes marginais proposto”, que são feitos através da estatística (3.6) e devem ser feitos caso o “Teste conjunto” rejeite a hipótese nula. Será descrito que o “Teste marginal proposto” foi realizado para a característica q para $q = 1, \dots, N$.

Nos próximos capítulos os resultados dos “Testes marginais originais” são comparados aos “Testes marginais propostos”. Para isso são apresentadas diversas simulações de dados com a finalidade de comparar as taxas de rejeição das hipóteses nulas dos testes. Dentro dessas simulações será esclarecido o cenário que motivou a elaboração dos “Testes marginais propostos” e os casos em que ele apresenta resultados mais satisfatórios comparado aos resultados dos “Testes marginais originais”.

Capítulo 4

Simulações

4.1 Simulações originais

O artigo no qual o teste ANOCVA foi publicado cita a realização de uma série de simulações para verificar as propriedades do teste. As simulações foram conduzidas considerando duas populações ($j = 1$ ou $j = 2$) contendo o mesmo número de elementos $n_j = n$, para $j = 1, 2$. Nas simulações, o tamanho das amostras varia de acordo com a seguinte especificação: $n = 10, 20, 30, 40$. Os vetores aleatórios $X_{i,j,q}$ são gerados por distribuições normais bivariadas independentes com a matriz de covariância a matriz identidade 2×2 em todos os casos e a média $\mu_{j,q}$ sendo dada por:

- $\mu_{j,q} = (0, 0)$ para $j = 1$ e $q \in \{1, \dots, 9\}$,
- $\mu_{j,q} = (\Delta, 0)$ para $j = 1$ e $q = 10$,
- $\mu_{j,q} = (2, 0)$ para $j = 1$ e $q \in \{11, \dots, 20\}$,
- $\mu_{j,q} = (0, 0)$ para $j = 2$ e $q \in \{1, \dots, 10\}$,
- $\mu_{j,q} = (2, 0)$ para $j = 2$ e $q \in \{11, \dots, 20\}$,

com Δ variando entre 0 e 2 em incrementos de 0.25. Desta forma apenas o vetor aleatório $X_{i,j,10}$ tem média diferente para as duas populações sendo que na primeira população a média de sua distribuição sofre um deslocamento do ponto $(0, 0)$ para o ponto $(2, 0)$ e na segunda não sofre deslocamento, permanecendo fixa no ponto $(2, 0)$. A figura 4.1, proposta inicialmente em Fujita et al. (2014b), ilustra os esquemas descritos acima.

As simulações foram repetidas 100 vezes com o objetivo de estimar a taxa de rejeição à hipótese nula dos testes conjuntos, marginais originais e marginais propostos descritos no capítulo 3. O

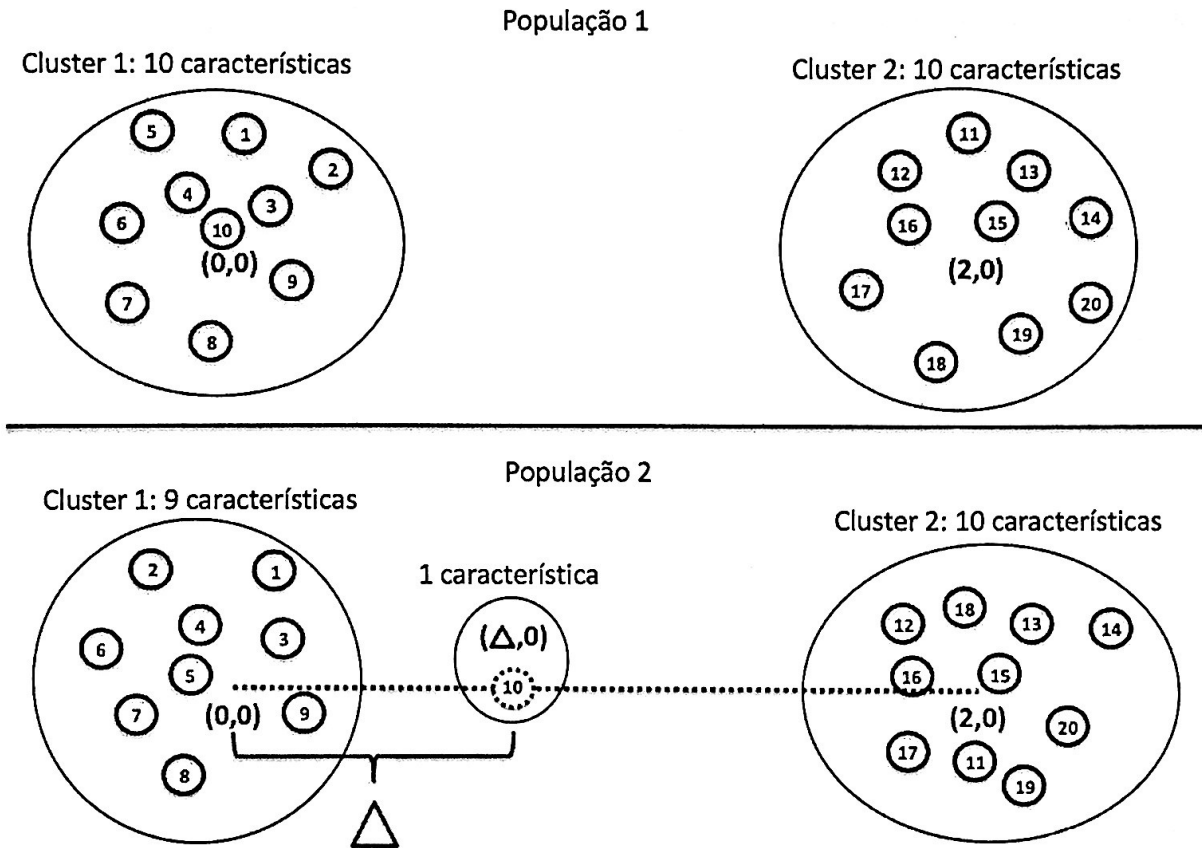


Figura 4.1

método de clusterização adotado foi o procedimento hierárquico de cluster por linkage, mesmo usado em Fujita et al. (2014b).

Foram utilizada duas distâncias nas simulações. Inicialmente foi feito o uso da distância Euclidiana, sem que se chegasse a resultados esperados. Posteriormente foi feito uso da distância Euclidiana com padronização, o qual, para cada característica, subtrai a média e divide pelo desvio absoluto médio (consulte a função *daisy* do pacote *cluster* do software estatístico *R*).

4.2 Simulações propostas

As simulações realizadas em Fujita et al. (2014b) foram conduzidas considerando fixados alguns dos parâmetros do modelo, a saber, a variância da variável aleatória geradora das séries temporais, o número de características e o número de características que não estão sob hipótese nula. O presente trabalho realiza simulações similares porém com diversos valores para cada um dos parâmetros referidos, como já citado anteriormente.

As simulações foram realizadas com o mesmo procedimento hierárquico de cluster por linkage utilizado nas simulações originais e a distância usada foi a Euclidiana. Foram considerados dois e

três clusters para que posteriormente seja possível verificar se o número de clusters influencia nos resultados das simulações. No entanto os vetores bidimensionais $X_{i,j,q}$ e a quantidade de vetores foram alteradas para facilitar generalizações. Os vetores aleatórios $X_{i,j,q}$ continuam distribuídos normalmente e de forma independente. As médias $\mu_{j,q}$ são definidas a seguir. Dados os inteiros N_1 e N_2 ,

- $\mu_{j,q} = (0, 0)$ para $j = 1$ e $q \in \{1, \dots, N_1\}$,
- $\mu_{j,q} = (\Delta, 0)$ para $j = 1$ e $q \in \{N_1 + 1, \dots, N_1 + N_2\}$,
- $\mu_{j,q} = (2, 0)$ para $j = 1$ e $q \in \{N_1 + N_2 + 1, \dots, 2.N_1 + N_2\}$,
- $\mu_{j,q} = (0, 0)$ para $j = 2$ e $q \in \{1, \dots, N_1\}$,
- $\mu_{j,q} = (2, 0)$ para $j = 2$ e $q \in \{N_1 + 1, \dots, 2.N_1 + N_2\}$,

com Δ variando entre 0 e 2 em incrementos de 0.25. A figura 4.2 ilustra a generalização das simulações descritas nesta seção. N_1 é o número de características de cada cluster que não sofrem deslocamento, enquanto N_2 é o número de características que sofrem deslocamento.

Nessas simulações a matriz de variâncias-covariâncias é dada por $\sigma^2.I_2$, sendo que o valor de σ^2 sofrerá variações com o intuito de verificar a influência da variabilidade no resultado dos testes. Em Fujita et al. (2014b) foi adotado $\sigma^2 = 1$. A seção 4.3 trata das simulações para variáveis degeneradas e a seção 4.4 trata das simulações com variáveis não degeneradas.

4.3 Análise das estatísticas para variáveis degeneradas

Antes de serem feitas simulações utilizando-se das definições da seção 4.3, as estatísticas foram calculadas para o caso $\sigma^2 = 0$, ou seja, os dados sendo gerados sem variação. Será comentado aqui o caso $N_1 = 2$ e $N_2 = 1$ onde tem-se todos os elementos da população 1 idênticos entre si e todos os elementos da população 2 idênticos entre si. Todos os cálculos dessa seção (e das subsequentes seções 4.3.1 e 4.3.2) consideram a distância Euclidiana como medida de dissimilaridade. Os cálculos são apresentados no apêndice A.1. No apêndice A.2 foram calculados os valores das estatísticas para o caso geral.

Sendo os elementos $x_{1,j}$ da população 1

$$x_{1,j} = \begin{bmatrix} 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.1)$$

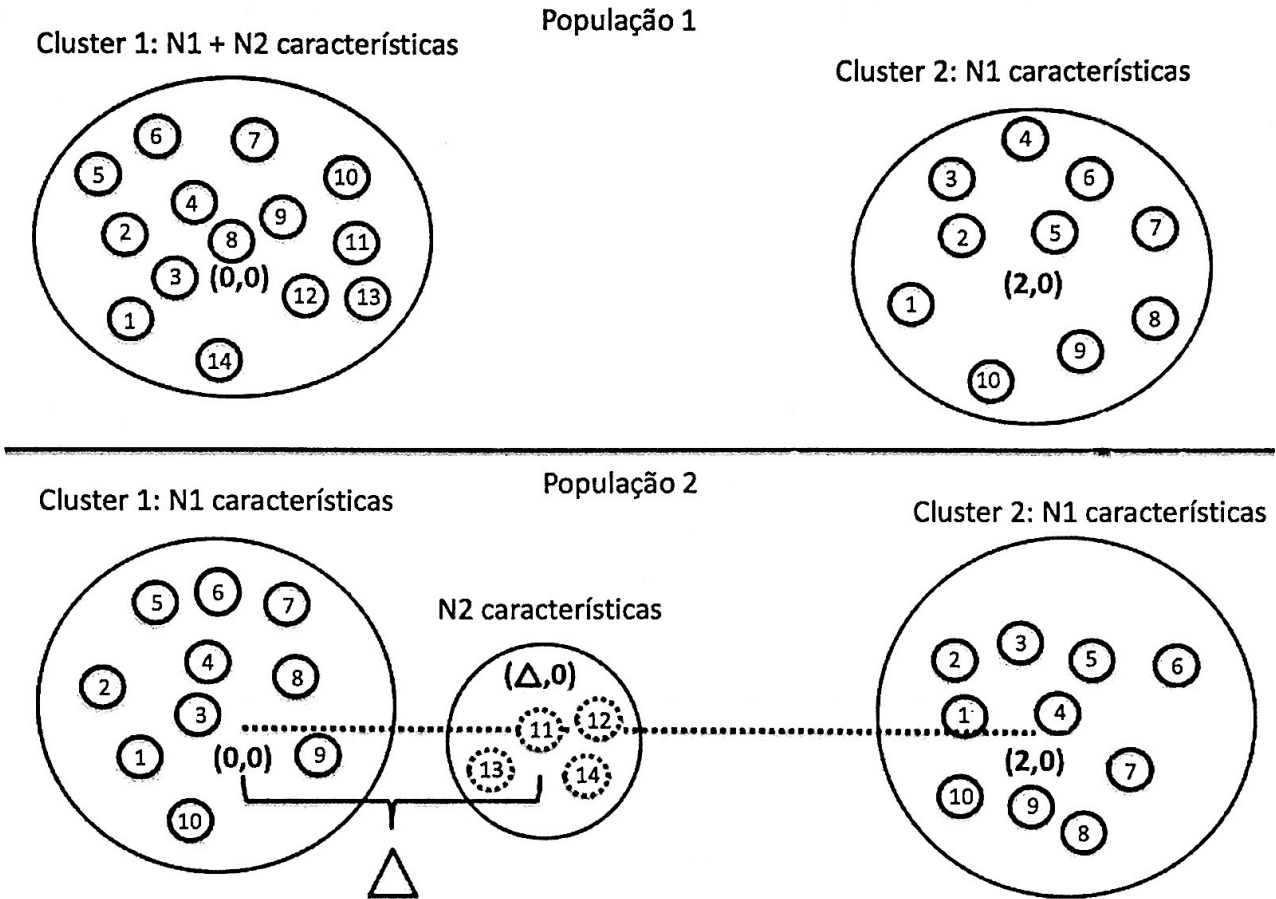


Figura 4.2

e da $x_{2,j}$ da população 2:

$$x_{2,j} = \begin{bmatrix} 0 & 0 & \Delta & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.2}$$

A seguir são apresentados as estatísticas dos testes. Importante salientar que os elementos perturbados se localizam no primeiro cluster, fato que pode ter impacto na taxa de rejeição da hipótese nula para as características que estão nesse cluster.

A estatística para o Teste conjunto é apresentada na equação (4.3).

$$\Delta S = \begin{cases} \frac{\Delta^6 - 12\Delta^5 + 68\Delta^4 - 160\Delta^3 + 192\Delta^2}{16(\Delta - 2)^2(\Delta - 4)^2}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{64(2\Delta^2 - 7\Delta + 4)^2 + \Delta^6 - 8\Delta^5 + 32\Delta^4}{16\Delta^2(\Delta - 4)^2}, & \text{se } 1 < \Delta \leq 2. \end{cases} \tag{4.3}$$

Note que a estatística, para ambos os intervalos, é um quociente entre um polinômio de grau 6 e um polinômio de grau 4. A figura 4.3 apresenta a estatística de teste ΔS para o Teste conjunto em função do Δ . O gráfico é condizente com o que se espera, *i.e.*, a função é monotônica crescente. Dessa forma, quanto mais distante da hipótese nula ($\Delta = 0$), maior será a estatística do teste. Esse

comportamento é esperado em estatísticas de teste (Cox et al., 1977).

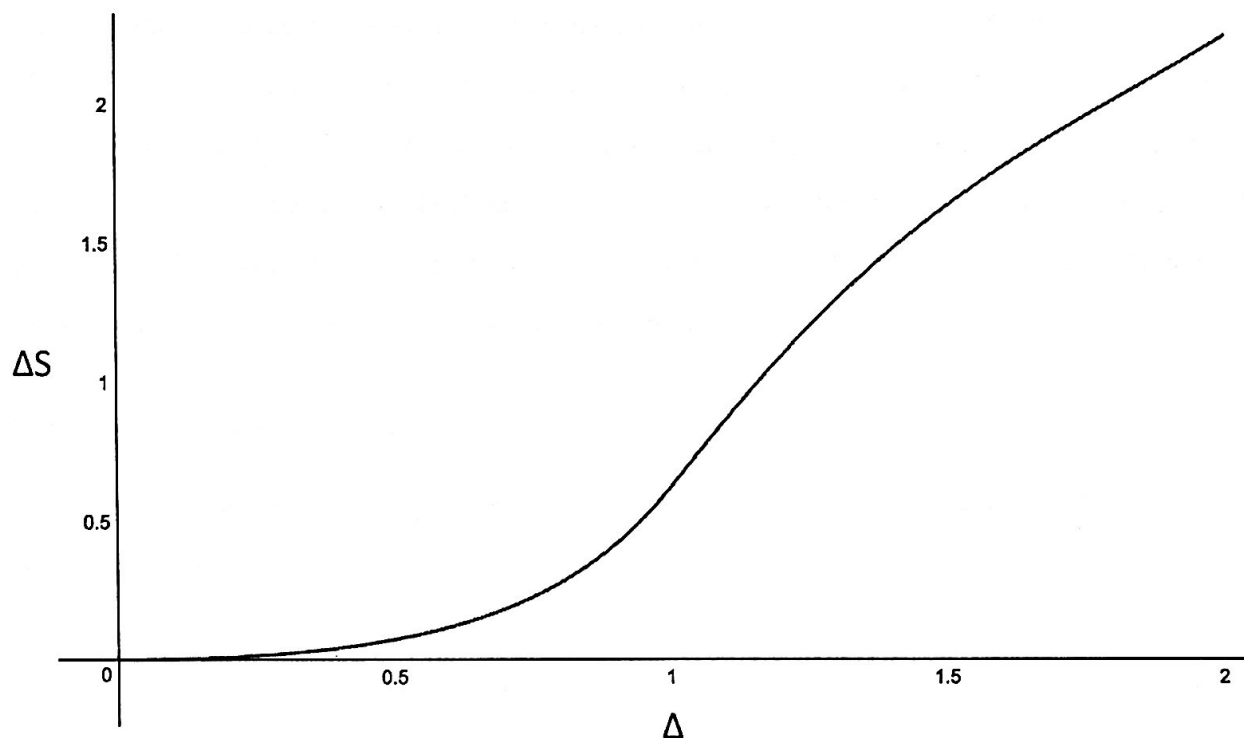


Figura 4.3

4.3.1 Teste marginal original

As estatísticas para o Teste marginal original para as características $q = 1$, $q = 2$, $q = 4$ ou $q = 5$ são nulas para qualquer valor de Δ , ou seja, $\delta s_1 = \delta s_2 = \delta s_4 = \delta s_5 = 0$, o que resulta na não rejeição da hipótese nula para essas características. Essa é uma propriedade desejada, pois essas características não sofrem perturbação por Δ . No entanto, conforme será visto mais adiante na Seção 4.4, as simulações com variabilidade não nula apresentam taxa de rejeição da hipótese nula crescente com relação a Δ .

A estatística para o Teste marginal original para a característica $q = 3$ é apresentada na equação (4.4).

$$\delta s_3 = \begin{cases} \frac{\Delta^2}{(4-\Delta)(4-2\Delta)}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{-5\Delta^2 + 14\Delta - 8}{2\Delta(4-\Delta)}, & \text{se } 1 < \Delta \leq 2. \end{cases} \quad (4.4)$$

A figura 4.4a apresenta o gráfico da estatística de teste δs_3 do Teste marginal original para a característica $q = 3$ em função de Δ . A curva mostra que a estatística de teste para característica nesse caso particular é decrescente em função de Δ para $4/3 < \Delta \leq 2$, chegando a ter valor nulo para $\Delta = 2$. Essa característica é indesejada visto que conforme Δ aumenta, espera-se que o teste

identifique que a característica seja clusterizada de maneira diferente nas duas populações. Esse comportamento, conforme será visto, permanece nas simulações com variabilidade.

Para que o Teste marginal original seja aplicado em uma característica é definida uma região crítica para a estatística de teste. Se a estatística do teste for maior que um determinado valor calculado de acordo com o nível de significância especificado, a hipótese nula é rejeitada. Nota-se que existem valores da estatística de teste δs_3 próximos de $\Delta = 2$ que não teriam a hipótese nula rejeitada pois a estatística de teste é zero quando deveria ser positiva.

4.3.2 Teste marginal proposto

A estatística para o Teste marginal proposto para as características $q = 1$ ou $q = 2$ é apresentada na equação abaixo:

$$\delta s_q^* = \frac{\Delta^2}{32}, q = 1, 2. \quad (4.5)$$

Nota-se que a estatística é não nula. Essa propriedade é não desejada, pois o teste pode concluir por rejeição da hipótese nula quando em fato ela é verdadeira. A explicação para esse efeito é que a estatística de *Silhouette* altera o valor não só da característica que sofre o deslocamento, mas também das características que estão no mesmo cluster da que sofre o deslocamento. Conforme pode ser visto na equação (A.25), o valor da estatística no entanto tende a zero quando N_2/N_1 tende a zero.

A estatística para o Teste marginal proposto para o elemento $q = 3$ é apresentada na equação abaixo:

$$\delta s_3^* = \begin{cases} \frac{\Delta^2(\Delta^2 - 4\Delta + 8)}{(4 - \Delta)^2(2 - \Delta)^2}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{4(2\Delta^2 - 7\Delta + 4) + \Delta^4}{\Delta^2(4 - \Delta)^2}, & \text{se } 1 < \Delta \leq 2. \end{cases} \quad (4.6)$$

A figura 4.4b apresenta a estatística de teste para o Teste marginal proposto para a característica $q = 3$ em função de Δ . A curva ilustra que a estatística de teste é estritamente crescente com relação a Δ . Essa é uma propriedade desejada e indica uma vantagem do Teste marginal proposto sobre o Teste marginal original.

As estatísticas para o Teste marginal proposto para as características $q = 4$ ou $q = 5$ são nulas para qualquer valor de Δ , ou seja, $\delta s_4^* = \delta s_5^* = 0$. Essa é uma propriedade desejada dado que essas características apresentam o mesmo comportamento nas duas populações e é esperado que o teste não conclua pela rejeição da hipótese nula.

Os exercícios feitos nessa seção são limitados principalmente pela restrição de que $\sigma^2 = 0$. Isso

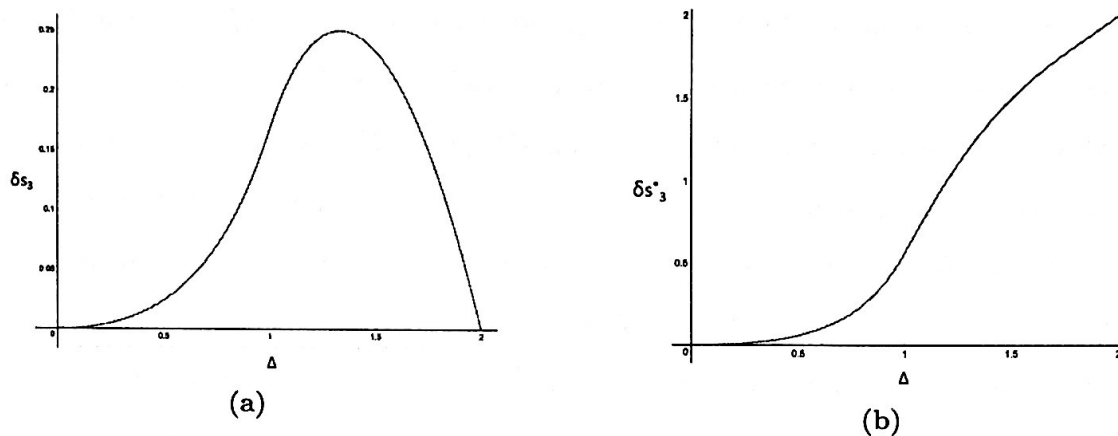


Figura 4.4

foi feito porque as distribuições das estatísticas de teste não são triviais de serem calculadas (Fujita et al., 2014b). A próxima seção apresenta resultados de simulações com $\sigma^2 > 0$ para verificar se as propriedades encontradas no caso degenerado persistem para os não degenerados.

4.4 Simulações

Essa seção apresenta o resultado das simulações produzidas através do *software* estatístico *R*. *R* é um *software* de código aberto desenvolvido para análise estatística e construção de gráficos. O procedimento usado para realizar as simulações dessa dissertação foi feito através da criação de quatro programas auxiliares:

- Programa 1: realiza o cálculo das estatísticas dos testes conjunto, marginal original e marginal proposto para uma coleção de dados;
- Programa 2: calcula o valor-p para esse teste através da técnica de *bootstrap*;
- Programa 3: gera simulações aleatórias com os parâmetros desejados;
- Programa 4: integra os programas anteriores para produzir as simulações com os parâmetros desejados.

Dada a enorme quantidade de simulações realizadas nesse estudo foi utilizada a versão *Server* do *R* instalados em computadores da plataforma *Google Cloud*.

O *R* conta com uma base de pacotes pública mantida por colaboradores independentes conhecida como *CRAN*. O objetivo dessa base é promover a rápida evolução do software sem que para isso seja necessário que o autores do *R* implementem a novas técnicas de forma centralizada. O teste ANOCVA conta com um pacote produzido para que se possa fazer uso do teste sem a necessidade

de sua implementação. Esse pacote encontra-se em sua versão 0.1.0 e pode ser acessado no site do *CRAN* (ANOCVA-R, 2016). Essa dissertação, conforme indicado acima, não fez uso desse pacote.

Para cada combinação das variáveis N_1 , N_2 , Δ , σ^2 e n foram simulados 100 conjuntos de dados. Nesses conjuntos de dados foram aplicados os testes conjunto, marginal original e marginal proposto para que fosse possível então criar os gráficos com as taxas de rejeição das hipóteses nulas.

Foram adotados 4 tamanhos de amostras: $n = n_1 = n_2 = 10$, $n = n_1 = n_2 = 20$, $n = n_1 = n_2 = 30$ e $n = n_1 = n_2 = 40$. Para determinar os valores-p dos testes foi utilizado o método *bootstrap* com 1000 repetições para cada conjunto de dados simulado.

As subseções 4.4.1, 4.4.2 e 4.4.3 utilizam a distância Euclidiana para o cálculo das matrizes de dissimilaridades. A subseção 4.4.4 utilizou a distância Euclidiana com padronização por desvio absoluto médio.

Os resultados dessas simulações são apresentados nas próximas seções. Para cada combinação de N_1 e N_2 são apresentados os resultados para os testes conjunto, marginal original e marginal proposto. No caso dos testes marginais são apresentados os resultados para a característica na qual é aplicada a perturbação e posteriormente para as características nas quais não é aplicada a perturbação.

Como foi citado na seção 4.3, existem $(2N_1 + N_2)$ características que podem ser testadas nos testes marginais. No entanto, conforme demonstrado para o caso geral no Apêndice A.2, as características $q = 1, \dots, N_1$ apresentam resultados bastante similares. O mesmo ocorre com o grupo das características $q = N_1 + 1, \dots, N_1 + N_2$ e com o grupo das características $q = N_1 + N_2 + 1, \dots, 2N_1 + N_2$. Por isso foi escolhido apenas um representante para cada grupo similar. Para o grupo das características $q = 1, \dots, N_1$ é apresentada a característica $q = 1$. Para o grupo das características $q = N_1 + 1, \dots, N_1 + N_2$ é apresentada a característica $q = N_1 + 1$. Para o grupo das características $q = N_1 + N_2 + 1, \dots, 2N_1 + N_2$ é apresentada a característica $q = N_1 + N_2 + 1$. Os resultados para $q = N_1 + 1$ são apresentados primeiramente pois é a característica que apresenta diferença entre as populações.

4.4.1 Simulações: $N_1 = 10$ e $N_2 = 1$ com distância Euclidiana

- Taxas de rejeição do Teste conjunto

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta.

As figuras B.1–B.5 apresentam os resultados das simulações com dois clusteres. Todas as figuras apresentam resultados dentro do esperado para o teste, independentemente da variância usada na

simulação.

As figuras B.6–B.10 apresentam os resultados das simulações com três clusteres. Da mesma forma todos os resultados ficam dentro do esperado, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ .

Pode-se concluir que o Teste conjunto, para o cenário descrito, se comporta conforme esperado, apresentando taxa de rejeição da hipótese nula crescente com relação a Δ .

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + 1 = 11$

Nessas simulações, é esperado que as taxas de rejeição aumentem à medida que Δ aumenta.

Teste marginal original: As figuras B.11–B.15 apresentam os resultados das simulações com dois clusteres. As figuras mostram decrescimento das taxas de rejeição da hipótese nula em função de Δ . Esse efeito contradiz o esperado e é compatível como o resultado obtido na seção 4.3.

Teste marginal proposto: As figuras B.21–B.25 apresentam os resultados das simulações com dois clusteres. As figuras mostram que a taxa de rejeição da hipótese nula cresce em função de Δ . Esse efeito vai ao encontro do que se espera e esse resultado é particularmente importante porque o apresenta o resultado esperado em casos que o Teste marginal original não o faz.

Ao comparar os Testes marginal original e proposto nessas simulações pode-se concluir que o Teste marginal proposto apresenta melhor performance do que o Teste marginal original quando utilizado dois clusteres.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 1$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.31–B.35 apresentam os resultados das simulações com dois clusteres. As figuras B.31 e B.32, com $\sigma^2 = 1$ e $\sigma^2 = 10^{-1}$ respectivamente, apresentam resultados esperados, enquanto que as figuras B.33 e B.34, com $\sigma^2 = 10^{-3}$ e $\sigma^2 = 10^{-5}$ respectivamente, apresentam resultados não esperados, ou seja, as taxas de rejeição da hipótese nula são crescentes com relação ao Δ .

Teste marginal proposto: As figuras B.41–B.50 apresentam os gráficos do Teste marginal proposto para dois e três clusteres. Apenas quando $\sigma^2 = 1$ e dois clusteres são utilizados, conforme mostrado na figura B.41, o Teste marginal proposto apresenta o resultado esperado para a característica. Conforme foi mostrado na Equação (A.25), esse efeito é esperado. Todos os outros casos apresentam resultados não esperados.

Pode-se concluir que nessas simulações nenhum dos testes marginais apresenta o resultado esperado.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + N_2 + 1 = 12$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.51–B.55 apresentam os resultados das simulações com dois clusteres. As taxas de rejeição para simulações com variâncias iguais a 10^{-3} e 10^{-5} não apresentam o comportamento esperado.

Teste marginal proposto: As figuras B.61–B.65 apresentam os gráficos de Teste marginal proposto para dois clusteres. Os resultados para variâncias iguais a 10^{-1} , 10^{-3} , 10^{-5} contradizem o esperado apenas quando $\Delta = 2$.

Pode-se concluir que para esse caso simulado o Teste marginal original e o Teste marginal proposto não apresentam resultados esperados, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ . No entanto o Teste marginal proposto apenas apresenta resultado não esperado quando $\Delta = 2$.

Algo que se nota em todas as simulações feitas com três clusteres nessa seção, como pode ser visto nas figuras B.16–B.20, B.26–B.30, B.36–B.40, B.46–B.50, B.56–B.60 e B.66–B.70 é que nenhum dos testes marginais originais e propostos, respectivamente, apresenta resultado esperado quando aplicado com três clusteres porque a estatística de *Silhouette* penaliza clusteres formados por apenas um elemento, visto que $N_2 = 1$.

4.4.2 Simulações: $N_1 = 10$ e $N_2 = 2$ com distância Euclidiana

- Taxas de rejeição do Teste conjunto

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta.

As figuras B.71–B.75 apresentam os resultados das simulações com dois clusteres. Todas as figuras apresentam resultados dentro do esperado para o teste, independentemente da variância usada na simulação.

As figuras B.76–B.80 apresentam os resultados das simulações com três clusteres. Da mesma forma todos os resultados ficam dentro do esperado, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ .

Pode-se concluir que o Teste conjunto, para o cenário descrito, se comporta conforme esperado, apresentando taxa de rejeição da hipótese nula crescente com relação a Δ .

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + 1 = 11$

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta.

Teste marginal original: As figuras B.81–B.85 apresentam os resultados das simulações com dois clusteres. As figuras mostram decréscimo das taxas de rejeição da hipótese nula em função de Δ . Esse efeito contradiz o esperado e é compatível como o resultado obtido na seção 4.3.

As figuras B.86–B.90 apresentam os resultados das simulações com três clusteres. Elas mostram taxas de rejeição crescentes com relação a Δ . Esses são resultados esperados para essa estatística.

Teste marginal proposto: As figuras B.91–B.95 apresentam os resultados das simulações com dois clusteres. As figuras mostram que a taxa de rejeição da hipótese nula cresce em função de Δ . Esse efeito vai ao encontro do que se espera e esse resultado.

As figuras B.96–B.100 apresentam os resultados das simulações com três clusteres. Os resultados também vão ao encontro do esperado, pois as taxas de rejeição crescem com relação a Δ .

Ao comparar os Testes marginal original e proposto nessas simulações pode-se concluir ambos os testes marginais apresentam os resultados esperados. No caso do Teste marginal original, o resultado esperado é observado apenas com três clusteres. No caso do Teste marginal proposto, o resultado esperado é observado com dois ou três clusteres.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 1$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.101–B.105 apresentam os resultados das simulações com dois clusteres. As figuras B.101 e B.105, com $\sigma^2 = 1$ e $\sigma^2 = 0$, respectivamente, apresentam resultados esperados, enquanto que as figuras B.102–B.104, com $\sigma^2 = 10^{-1}$, $\sigma^2 = 10^{-3}$ e $\sigma^2 = 10^{-5}$, respectivamente, apresentam resultados que vão de encontro com o esperado, ou seja, as taxas de rejeição das hipótese nula são crescentes com relação a Δ .

As figuras B.106–B.110 apresentam os resultados das simulações com três clusteres. Esses resultados mostram taxas de rejeição da hipótese nula que crescem com relação ao Δ , também contradizendo o esperado.

Teste marginal proposto: As figuras B.111–B.120 apresentam os gráficos do Teste marginal proposto para dois e três clusteres. Conforme mostrado nos gráficos B.112–B.120, o Teste marginal proposto apresenta resultados não esperados para o teste, pois apresentam taxas de rejeição crescentes com relação a Δ .

Pode-se concluir que nessas simulações nenhum dos testes marginais apresenta o resultado esperado para esse caso.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + N_2 + 1 = 13$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.121–B.125 apresentam os resultados das simulações com dois clusteres. As taxas de rejeição para simulações com variâncias iguais a 10^{-3} e 10^{-5} não apresentam o comportamento esperado.

As figuras B.126–B.130 apresentam os resultados das simulações com três clusteres. Os resultados mostram taxas crescentes de rejeição para todas as variâncias, o que mais uma vez contradiz o esperado.

Teste marginal proposto: As figuras B.131–B.135 apresentam os gráficos de Teste marginal proposto para dois clusteres. Os resultados para variâncias iguais a 10^{-1} , 10^{-3} , 10^{-5} contradizem o esperado apenas quando $\Delta = 2$.

Quando o teste é realizado com três clusteres os resultados são também contrários ao esperado em todos os casos, conforme pode ser visto nos gráficos B.136–B.140, apresentando taxas de rejeição crescente da hipótese nula.

Pode-se concluir que para esse caso simulado o Teste marginal original e o Teste marginal proposto não apresentam resultados esperados, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ . No entanto o Teste marginal proposto apenas apresenta resultado não esperado quando $\Delta = 2$, quando usados dois clusteres.

4.4.3 Simulações: $N_1 = 10$ e $N_2 = 5$ com distância Euclidiana

- Taxas de rejeição do Teste conjunto

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta.

As figuras B.141–B.145 apresentam os resultados das simulações com dois clusteres. Todas as figuras apresentam resultados dentro do esperado para o teste, independentemente da variância usada na simulação.

As figuras B.146–B.150 apresentam os resultados das simulações com três clusteres. Da mesma forma todos os resultados ficam dentro do esperado, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ .

Pode-se concluir que o Teste conjunto, para o cenário descrito, se comporta conforme esperado, apresentando taxa de rejeição da hipótese nula crescente com relação a Δ .

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + 1 = 11$

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta.

Teste marginal original: As figuras B.151–B.155 apresentam os resultados das simulações com dois clusteres. As figuras B.151 e B.155, com $\sigma^2 = 1$ e $\sigma^2 = 0$, respectivamente, apresentam resultados esperados, enquanto que as figuras B.152–B.154, com $\sigma^2 = 10^{-1}$, $\sigma^2 = 10^{-3}$ e $\sigma^2 = 10^{-5}$, respectivamente, apresentam resultados que vão de encontro com o esperado, ou seja, as taxas de rejeição das hipótese nula são decrescentes com relação a Δ .

As figuras B.156–B.160 apresentam os resultados das simulações com três clusteres. Elas mostram taxas de rejeição crescentes com relação a Δ . Esses resultados vão ao encontro do resultado esperado.

Teste marginal proposto: As figuras B.161–B.170 apresentam os resultados das simulações com dois e três clusteres. As figuras mostram que a taxa de rejeição da hipótese nula cresce em função de Δ . Esse efeito vai ao encontro do que se espera.

Ao comparar os Testes marginal original e proposto nessas simulações pode-se concluir ambos os testes marginais apresentam os resultados esperados. No caso do Teste marginal original, o resultado esperado é visto apenas com três clusteres. No caso do Teste marginal proposto, o resultado esperado é visto com dois ou três clusteres.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 1$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.171–B.175 apresentam os resultados das simulações com dois clusteres. As figuras B.171 e B.175, com $\sigma^2 = 1$ e $\sigma^2 = 0$, respectivamente, apresentam resultados esperados, enquanto que as figuras B.172–B.174, com $\sigma^2 = 10^{-1}$, $\sigma^2 = 10^{-3}$ e $\sigma^2 = 10^{-5}$, respectivamente, apresentam resultados que vão de encontro com o esperado, ou seja, as taxas de rejeição das hipótese nula são crescentes com relação a Δ .

As figuras B.176–B.180 apresentam os resultados das simulações com três clusteres. Esses resultados mostram taxas de rejeição da hipótese nula que crescem com relação ao Δ , também contradizendo o esperado.

Teste marginal proposto: As figuras B.181–B.190 apresentam os gráficos do Teste marginal proposto para dois e três clusteres. Conforme mostrado nos gráficos, o Teste marginal proposto apresenta resultados não esperados para o teste, pois apresentam taxas de rejeição crescentes com relação a Δ .

Pode-se concluir que nessas simulações nenhum dos testes marginais apresenta o resultado esperado para esse caso.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = N_1 + N_2 + 1 = 16$

Nessas simulações, é esperado que as taxas de rejeição se mantenham constantes com relação a Δ .

Teste marginal original: As figuras B.191–B.195 apresentam os resultados das simulações com dois clusteres. As taxas de rejeição para simulações com variâncias iguais a 10^{-1} , 10^{-3} e 10^{-5} não apresentam o comportamento esperado.

As figuras B.196–B.200 apresentam os resultados das simulações com três clusteres. Os resultados mostram taxas crescentes de rejeição para todas as variâncias, o que mais uma vez contradiz o esperado.

Teste marginal proposto: As figuras B.201–B.205 apresentam os gráficos de Teste marginal proposto para dois clusteres. Os resultados para variâncias iguais a 10^{-1} , 10^{-3} , 10^{-5} contradizem o esperado para esse teste.

Quando o teste é realizado com três clusteres os resultados são também contrários ao esperado em todos os casos, conforme pode ser visto nos gráficos B.206–B.210, apresentando taxas de rejeição crescente da hipótese nula.

Pode-se concluir que para esse caso simulado o Teste marginal original e o Teste marginal proposto não apresentam resultados esperados, apresentando taxas de rejeição da hipótese nula crescentes com relação a Δ .

4.4.4 Simulações: $N_1 = 10$ e $N_2 = 1$ com distância Euclidiana padronizada por desvio absoluto médio

- Taxas de rejeição do Teste conjunto

Pode-se concluir que o Teste conjunto, para o cenário descrito, se comporta conforme esperado.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 11$

Nessas simulações, é esperado que a taxa de rejeição aumente à medida que Δ aumenta. Como pode-se notar nos gráficos, os resultados dos testes vão ao encontro do esperado para todos os casos com dois clusteres.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 1$

Nota-se pelos gráficos que as taxa de rejeição da hipótese nula se comportam da maneira esperada quando utilizado dois clusteres.

- Taxas de rejeição dos Testes marginal original e marginal proposto para $q = 12$

Os gráficos mostram que as simulações feitas com dois clusteres apresentam resultados esperados em todos os casos.

Capítulo 5

Conclusões e estudos futuros

5.1 Conclusões

O presente estudo apresentou o teste ANOCVA. Após uma breve introdução do teste e do contexto no qual se aplica, foram revistos os principais conceitos envolvidos, a saber: FMRI, clusterização e a estatística de *Silhouette*. Em seguida o teste ANOCVA foi apresentado através dos Testes conjunto, Teste marginal original e Teste marginal proposto.

O teste ANOCVA foi então estudado através da análise das estatísticas do Teste conjunto, marginal original e marginal proposto para o caso degenerado, ou seja, com variância igual a zero. As estatísticas dos testes foram estudadas inicialmente com valores específicos para as variáveis $N_1 = 2$ e $N_2 = 1$ e posteriormente sem atribuir valores à essas variáveis. Essa análise levou em conta a distância Euclidiana não padronizada como medida de dissimilaridade, deixando o caso em que a distância Euclidiana padronizada é usada como medida de dissimilaridade para estudos futuros.

Em seguida foram desenvolvidas as simulações, onde foi explicado sucintamente o *software* utilizado e a engenharia dos programas criados.

Resultados com distância Euclidiana não padronizada: O trabalho mostrou que o teste conjunto se comporta de forma satisfatória em todos os casos estudados, ou seja, a taxa de rejeição da hipótese nula aumenta à medida em que Δ aumenta.

Com relação aos resultados do teste marginal original, duas particularidades foram encontradas:

- Ocorrência de taxas de rejeição da hipótese nula decrescentes nas características deslocadas por Δ quando Δ está próximo de 2;
- Ocorrência de taxas de rejeição da hipótese nula crescente nas características não deslocadas

por Δ .

O teste marginal proposto, conforme mostrado nessa dissertação não apresenta a primeira particularidade que o teste marginal original apresenta, sendo esta a principal motivação da sua posição por essa dissertação.

Resultados com distância Euclidiana padronizada: O teste marginal proposto, quando aplicado com a distância Euclidiana padronizada por desvio absoluto médio, não apresenta nenhuma das particularidades obtidas no teste original e as propriedades que puderam ser analisadas e comparadas através das simulações realizadas apresentam resultados dentro do esperado. Os resultados devem ser estendíveis para outros casos e podem ser confirmados através de simulações.

Os resultados desse estudo sugerem a aplicação do teste ANOCVA com o uso dos testes marginais propostos e da distância Euclidiana padronizada por desvio absoluto médio como medida de dissimilaridade.

5.2 Estudos futuros

- Desenvolvimento analítico das estatísticas dos testes, como realizado no apêndice A, para a distância Euclidiana padronizada por desvio absoluto médio. Esses resultados poderiam mostrar propriedades que não foram identificadas nas simulações desenvolvidas. Esses resultados não foram obtidos porque a distância Euclidiana padronizada só foi estudada nos últimos dias do desenvolvimento desse trabalho.
- Ampliar as simulações de forma a cobrir valores de parâmetros que não foram aqui estudados. Nesse caso torna-se importante a aplicação de técnicas que permitam reduzir o tempo de processamento e também técnicas de engenharia de software que permitam que o código possa ser facilmente expandido quando necessário.
- Outra opção para estudos futuros é a aplicação dos testes estudados em conjuntos de dados reais como foi feito em Fujita et al. (2014b) com a aplicação do teste ANOCVA para dados com resultados de fMRI para pacientes que apresentam ADHD assim como pacientes controle.

Apêndice A

Fórmulas

A.1 Desenvolvimento das fórmulas da estatística do teste

Supondo $N_1 = 2$ e $N_2 = 1$ teríamos a população 1 com todos os elementos idênticos, sendo os elementos $x_{1,j}$ da população 1

$$x_{1,j} = \begin{bmatrix} 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{A.1})$$

e a população 2 com todos os elementos idênticos, sendo os elementos $x_{2,j}$ da população 2:

$$x_{2,j} = \begin{bmatrix} 0 & 0 & \Delta & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (\text{A.2})$$

Assim, as matrizes de dissimilaridades calculadas com a distância Euclidiana ficariam definidas por:

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 2 & 2 & 2 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \end{bmatrix}, \quad (\text{A.3})$$

$$A_2 = \begin{bmatrix} 0 & 0 & \Delta & 2 & 2 \\ 0 & 0 & \Delta & 2 & 2 \\ \Delta & \Delta & 0 & 2 - \Delta & 2 - \Delta \\ 2 & 2 & 2 - \Delta & 0 & 0 \\ 2 & 2 & 2 - \Delta & 0 & 0 \end{bmatrix}, \quad (\text{A.4})$$

$$A = \begin{bmatrix} 0 & 0 & \Delta/2 & 2 & 2 \\ 0 & 0 & \Delta/2 & 2 & 2 \\ \Delta/2 & \Delta/2 & 0 & 2 - \Delta/2 & 2 - \Delta/2 \\ 2 & 2 & 2 - \Delta/2 & 0 & 0 \\ 2 & 2 & 2 - \Delta/2 & 0 & 0 \end{bmatrix}. \quad (\text{A.5})$$

A clusterização das características:

$$l_A = \{1, 1, 1, 2, 2\}, \text{ para } 0 \leq \Delta \leq 2, \quad (\text{A.6})$$

observando-se que $q = 3$ foi clusterizado no cluster 1 para $\Delta = 2$ por critério de desempate.

E os vetores das estatísticas de *Silhouette*:

$$S_1 = \{1, 1, 1, 1, 1\}, \quad (\text{A.7})$$

$$S_2 = \begin{cases} \{1 - \Delta/4, 1 - \Delta/4, \frac{2-2\Delta}{2-\Delta}, 1, 1\} & , \text{ se } 0 \leq \Delta \leq 1, \\ \{1 - \Delta/4, 1 - \Delta/4, \frac{2-2\Delta}{\Delta}, 1, 1\} & , \text{ se } 1 < \Delta \leq 2, \end{cases} \quad (\text{A.8})$$

e

$$S = \left\{ 1 - \Delta/8, 1 - \Delta/8, \frac{2(2 - \Delta)}{4 - \Delta}, 1, 1 \right\}. \quad (\text{A.9})$$

Podemos então calcular as estatísticas para o Teste conjunto, Teste marginal e Teste marginal proposto conforme descritos na seções 1.3 e 1.4. A estatística para o Teste conjunto é apresentada através da equação A.10.

$$\Delta S = \begin{cases} \frac{\Delta^6 - 12\Delta^5 + 68\Delta^4 - 160\Delta^3 + 192\Delta^2}{16(\Delta-2)^2(\Delta-4)^2}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{64(2\Delta^2 - 7\Delta + 4)^2 + \Delta^6 - 8\Delta^5 + 32\Delta^4}{16\Delta^2(\Delta-4)^2}, & \text{se } 1 < \Delta \leq 2. \end{cases} \quad (\text{A.10})$$

A estatística para o Teste marginal para as características $q = 1, q = 2, q = 4$ ou $q = 5$ é nula para qualquer valor de Δ , ou seja, $\delta s_1 = \delta s_2 = \delta s_4 = \delta s_5 = 0$.

A estatística para o Teste marginal para a característica $q = 3$ é apresentada na equação A.11.

$$\delta s_3 = \begin{cases} \frac{\Delta^2}{(4-\Delta)(4-2\Delta)}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{-5\Delta^2+14\Delta-8}{2\Delta(4-\Delta)}, & \text{se } 1 < \Delta \leq 2. \end{cases} \quad (\text{A.11})$$

A estatística para o Teste marginal proposto para as características $q = 1$ ou $q = 2$ é apresentada na equação A.12:

$$\delta s_q^* = \frac{\Delta^2}{32}, q = 1, 2. \quad (\text{A.12})$$

A estatística para o Teste marginal proposto para o elemento $q = 3$ é apresentada na equação A.13:

$$\delta s_3^* = \begin{cases} \frac{\Delta^2(\Delta^2-4\Delta+8)}{(4-\Delta)^2(2-\Delta)^2}, & \text{se } 0 \leq \Delta \leq 1, \\ \frac{4(2\Delta^2-7\Delta+4)+\Delta^4}{\Delta^2(4-\Delta)^2}, & \text{se } 1 < \Delta \leq 2. \end{cases} \quad (\text{A.13})$$

A estatística para o Teste marginal proposto para as características $q = 4$ ou $q = 5$ é nula para qualquer valor de Δ , ou seja, $\delta s_4 = \delta s_5 = 0$.

A.2 Generalização das fórmulas da estatística do teste

Generalizando as fórmulas da seção anterior para $N_1 \geq 1$ e $N_2 \geq 1$ teríamos a população 1 com todos os elementos idênticos, sendo os elementos $x_{1,j}$ da população 1

$$x_{1,j} = \begin{bmatrix} & N_1 & & N_2 & & N_1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 2 & \dots & 2 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \quad (\text{A.14})$$

e a população 2 com todos os elementos idênticos, sendo os elementos $x_{2,j}$ da população 2:

$$x_{2,j} = \begin{bmatrix} & N_1 & & N_2 & & N_1 \\ 0 & \dots & 0 & \Delta & \dots & \Delta & 2 & \dots & 2 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}. \quad (\text{A.15})$$

Assim, as matrizes de dissimilaridades calculadas com a distância Euclidiana ficariam definidas por:

$$A_1 = \begin{bmatrix} & N_1 & & N_2 & & N_1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 2 & \dots & 2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & 2 & \dots & 2 \\ 0 & \dots & 0 & 0 & \dots & 0 & 2 & \dots & 2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & 2 & \dots & 2 \\ 2 & \dots & 2 & 2 & \dots & 2 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 2 & \dots & 2 & 2 & \dots & 2 & 0 & \dots & 0 \end{bmatrix}, \quad (\text{A.16})$$

$$A_2 = \begin{matrix} & \begin{matrix} N_1 & & N_2 & & N_1 \end{matrix} \\ \begin{matrix} 0 & \dots & 0 & \Delta & \dots & \Delta & 2 & \dots & 2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \Delta & \dots & \Delta & 2 & \dots & 2 \\ \Delta & \dots & \Delta & 0 & \dots & 0 & 2-\Delta & \dots & 2-\Delta \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta & \dots & \Delta & 0 & \dots & 0 & 2-\Delta & \dots & 2-\Delta \\ 2 & \dots & 2 & 2-\Delta & \dots & 2-\Delta & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 2 & \dots & 2 & 2-\Delta & \dots & 2-\Delta & 0 & \dots & 0 \end{matrix} & \begin{matrix} N_1 \\ N_2 \\ N_1 \end{matrix} \end{matrix}, \quad (A.17)$$

$$A = \begin{matrix} & \begin{matrix} N_1 & & N_2 & & N_1 \end{matrix} \\ \begin{matrix} 0 & \dots & 0 & \Delta/2 & \dots & \Delta/2 & 2 & \dots & 2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \Delta/2 & \dots & \Delta/2 & 2 & \dots & 2 \\ \Delta/2 & \dots & \Delta/2 & 0 & \dots & 0 & 2-\Delta/2 & \dots & 2-\Delta/2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta/2 & \dots & \Delta/2 & 0 & \dots & 0 & 2-\Delta/2 & \dots & 2-\Delta/2 \\ 2 & \dots & 2 & 2-\Delta/2 & \dots & 2-\Delta/2 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 2 & \dots & 2 & 2-\Delta/2 & \dots & 2-\Delta/2 & 0 & \dots & 0 \end{matrix} & \begin{matrix} N_1 \\ N_2 \\ N_1 \end{matrix} \end{matrix}. \quad (A.18)$$

A clusterização das características:

$$l_A = \left\{ \begin{matrix} & N_1 & & N_2 & & N_1 \\ 1 & \dots & 1 & 1 & \dots & 1 & 2 & \dots & 2 \end{matrix} \right\} \quad (\text{A.19})$$

observando-se que $q = N_1 + 1, \dots, N_1 + N_2$ foram clusterizados no cluster 1 para $\Delta = 2$ por critério de desempate.

E os vetores das estatísticas de *Silhouette*:

$$S_1 = \left\{ \begin{matrix} & N_1 & & N_2 & & N_1 \\ 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \end{matrix} \right\} \quad (\text{A.20})$$

$$S_2 = \left\{ \begin{matrix} & N_1 & & N_2 & & N_1 \\ \left\{ 1 - \frac{\Delta N_2}{2(N_1+N_2-1)} \dots 1 - \frac{\Delta N_2}{2(N_1+N_2-1)} 1 - \frac{\Delta N_1}{(2-\Delta)(N_1+N_2-1)} \dots 1 - \frac{\Delta N_1}{(2-\Delta)(N_1+N_2-1)} 1 \dots 1 \right\}, & \text{se } 0 \leq \Delta \leq \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \\ \\ & N_1 & & N_2 & & N_1 \\ \left\{ 1 - \frac{\Delta N_2}{2(N_1+N_2-1)} \dots 1 - \frac{\Delta N_2}{2(N_1+N_2-1)} \frac{(2-\Delta) - \frac{\Delta N_1}{(N_1+N_2-1)}}{\frac{\Delta N_1}{(N_1+N_2-1)}} \dots \frac{(2-\Delta) - \frac{\Delta N_1}{(N_1+N_2-1)}}{\frac{\Delta N_1}{(N_1+N_2-1)}} 1 \dots 1 \right\}, & \text{se } \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \leq \Delta \leq 2 \end{matrix} \right. \quad (\text{A.21})$$

e

$$S = \left\{ \begin{matrix} & N_1 & & N_2 & & N_1 \\ 1 - \frac{\Delta N_2}{4(N_1+N_2-1)} \dots 1 - \frac{\Delta N_2}{4(N_1+N_2-1)} 1 - \frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} \dots 1 - \frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} 1 \dots 1 \end{matrix} \right\} \quad (\text{A.22})$$

Podemos então calcular as estatísticas para o Teste conjunto, Teste marginal e Teste marginal proposto conforme descritos na seções 1.3 e 1.4. A estatística para o Teste conjunto é apresentada através da equação A.23.

$$S = \begin{cases} N_2 \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} - \frac{\Delta N_1}{(2-\Delta)(N_1+N_2-1)} \right)^2 + N_2 \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} \right)^2 + 2N_1 \left(\frac{\Delta N_2}{4(N_1+N_2-1)} \right)^2, \text{ se } 0 \leq \Delta \leq \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \\ N_2 \left(1 - \frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} - \frac{(2-\Delta) - \frac{\Delta N_1}{(N_1+N_2-1)}}{\frac{\Delta N_1}{(N_1+N_2-1)}} \right)^2 + N_2 \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} \right)^2 + 2N_1 \left(\frac{\Delta N_2}{4(N_1+N_2-1)} \right)^2, \text{ se } \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \leq \Delta \leq 2 \end{cases} \quad (\text{A.23})$$

A estatística para o Teste marginal original para as características $q = 1, \dots, N_1$ é nula para qualquer valor de Δ , ou seja, $\delta s_1 = \dots = \delta s_{N_1} = 0$. Nota-se também que a estatística para o Teste marginal original para as características $q = N_1 + N_2 + 1, \dots, 2N_1 + N_2$ é nula para qualquer valor de Δ , ou seja, $\delta s_{N_1+N_2+1} = \dots = \delta s_{2N_1+N_2} = 0$.

A estatística para o Teste marginal original para as características $q = N_1 + 1, \dots, N_1 + N_2$ é apresentada na equação A.24.

$$\delta s_q = \begin{cases} \frac{\Delta^2 N_1}{2(4-\Delta)(2-\Delta)(N_1+N_2-1)}, \text{ se } 0 \leq \Delta \leq \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \\ 1 - \frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} - \frac{1}{2} \left(1 + \frac{2-\Delta - \frac{\Delta N_1}{N_1+N_2-1}}{\frac{\Delta N_1}{N_1+N_2-1}} \right), \text{ se } \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \leq \Delta \leq 2 \end{cases}, q = N_1+1, \dots, N_1+N_2. \quad (\text{A.24})$$

A estatística para o Teste marginal proposto para as características $q = 1, \dots, N_1$ é apresentada na equação A.25:

$$\delta s_q^* = \frac{1}{8} \left(\frac{\Delta N_2}{(N_1 + N_2 - 1)} \right)^2, q = 1, \dots, N_1. \quad (\text{A.25})$$

A estatística para o Teste marginal proposto para os elementos $q = N_1 + 1, \dots, N_1 + N_2$ é apresentada na equação A.26:

$$\delta s_q^* = \begin{cases} \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} - \frac{\Delta N_1}{2(2-\Delta)(N_1+N_2-1)} \right)^2 + \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} \right)^2, \text{ se } 0 \leq \Delta \leq \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \\ \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} - 1 - \frac{(2-\Delta) - \frac{\Delta N_1}{(N_1+N_2-1)}}{\frac{\Delta N_1}{(N_1+N_2-1)}} \right)^2 + \left(\frac{\Delta N_1}{(4-\Delta)(N_1+N_2-1)} \right)^2, \text{ se } \frac{2(N_1+N_2-1)}{2N_1+N_2-1} \leq \Delta \leq 2 \end{cases}, q = N_1+1, \dots, N_1+N_2. \quad (\text{A.26})$$

A estatística para o Teste marginal proposto para as características $q = N_1+N_2+1, \dots, 2N_1+N_2$ é nula para qualquer valor de Δ , ou seja, $\delta s_{N_1+N_2+1} = \dots = \delta s_{2N_1+N_2}$.

Apêndice B

Gráficos

B.1 Tabela resumo dos gráficos

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Conjunto	10	1	1	-	2	B.1
Conjunto	10	1	10^{-1}	-	2	B.2
Conjunto	10	1	10^{-3}	-	2	B.3
Conjunto	10	1	10^{-5}	-	2	B.4
Conjunto	10	1	0	-	2	B.5
Conjunto	10	1	1	-	3	B.6
Conjunto	10	1	10^{-1}	-	3	B.7
Conjunto	10	1	10^{-3}	-	3	B.8
Conjunto	10	1	10^{-5}	-	3	B.9
Conjunto	10	1	0	-	3	B.10
Marginal Original	10	1	1	11	2	B.11
Marginal Original	10	1	10^{-1}	11	2	B.12
Marginal Original	10	1	10^{-3}	11	2	B.13
Marginal Original	10	1	10^{-5}	11	2	B.14
Marginal Original	10	1	0	11	2	B.15
Marginal Original	10	1	1	11	3	B.16
Marginal Original	10	1	10^{-1}	11	3	B.17
Marginal Original	10	1	10^{-3}	11	3	B.18
Marginal Original	10	1	10^{-5}	11	3	B.19
Marginal Original	10	1	0	11	3	B.20

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Proposto	10	1	1	11	2	B.21
Marginal Proposto	10	1	10^{-1}	11	2	B.22
Marginal Proposto	10	1	10^{-3}	11	2	B.23
Marginal Proposto	10	1	10^{-5}	11	2	B.24
Marginal Proposto	10	1	0	11	2	B.25
Marginal Proposto	10	1	1	11	3	B.26
Marginal Proposto	10	1	10^{-1}	11	3	B.27
Marginal Proposto	10	1	10^{-3}	11	3	B.28
Marginal Proposto	10	1	10^{-5}	11	3	B.29
Marginal Proposto	10	1	0	11	3	B.30
Marginal Original	10	1	1	1	2	B.31
Marginal Original	10	1	10^{-1}	1	2	B.32
Marginal Original	10	1	10^{-3}	1	2	B.33
Marginal Original	10	1	10^{-5}	1	2	B.34
Marginal Original	10	1	0	1	2	B.35
Marginal Original	10	1	1	1	3	B.36
Marginal Original	10	1	10^{-1}	1	3	B.37
Marginal Original	10	1	10^{-3}	1	3	B.38
Marginal Original	10	1	10^{-5}	1	3	B.39
Marginal Original	10	1	0	1	3	B.40
Marginal Proposto	10	1	1	1	2	B.41
Marginal Proposto	10	1	10^{-1}	1	2	B.42
Marginal Proposto	10	1	10^{-3}	1	2	B.43
Marginal Proposto	10	1	10^{-5}	1	2	B.44
Marginal Proposto	10	1	0	1	2	B.45
Marginal Proposto	10	1	1	1	3	B.46
Marginal Proposto	10	1	10^{-1}	1	3	B.47
Marginal Proposto	10	1	10^{-3}	1	3	B.48
Marginal Proposto	10	1	10^{-5}	1	3	B.49
Marginal Proposto	10	1	0	1	3	B.50

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Original	10	1	1	12	2	B.51
Marginal Original	10	1	10^{-1}	12	2	B.52
Marginal Original	10	1	10^{-3}	12	2	B.53
Marginal Original	10	1	10^{-5}	12	2	B.54
Marginal Original	10	1	0	12	2	B.55
Marginal Original	10	1	1	12	3	B.56
Marginal Original	10	1	10^{-1}	12	3	B.57
Marginal Original	10	1	10^{-3}	12	3	B.58
Marginal Original	10	1	10^{-5}	12	3	B.59
Marginal Original	10	1	0	12	3	B.60
Marginal Proposto	10	1	1	12	2	B.61
Marginal Proposto	10	1	10^{-1}	12	2	B.62
Marginal Proposto	10	1	10^{-3}	12	2	B.63
Marginal Proposto	10	1	10^{-5}	12	2	B.64
Marginal Proposto	10	1	0	12	2	B.65
Marginal Proposto	10	1	1	12	3	B.66
Marginal Proposto	10	1	10^{-1}	12	3	B.67
Marginal Proposto	10	1	10^{-3}	12	3	B.68
Marginal Proposto	10	1	10^{-5}	12	3	B.69
Marginal Proposto	10	1	0	12	3	B.70
Conjunto	10	2	1	-	2	B.71
Conjunto	10	2	10^{-1}	-	2	B.72
Conjunto	10	2	10^{-3}	-	2	B.73
Conjunto	10	2	10^{-5}	-	2	B.74
Conjunto	10	2	0	-	2	B.75
Conjunto	10	2	1	-	3	B.76
Conjunto	10	2	10^{-1}	-	3	B.77
Conjunto	10	2	10^{-3}	-	3	B.78
Conjunto	10	2	10^{-5}	-	3	B.79
Conjunto	10	2	0	-	3	B.80

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Original	10	2	1	11	2	B.81
Marginal Original	10	2	10^{-1}	11	2	B.82
Marginal Original	10	2	10^{-3}	11	2	B.83
Marginal Original	10	2	10^{-5}	11	2	B.84
Marginal Original	10	2	0	11	2	B.85
Marginal Original	10	2	1	11	3	B.86
Marginal Original	10	2	10^{-1}	11	3	B.87
Marginal Original	10	2	10^{-3}	11	3	B.88
Marginal Original	10	2	10^{-5}	11	3	B.89
Marginal Original	10	2	0	11	3	B.90
Marginal Proposto	10	2	1	11	2	B.91
Marginal Proposto	10	2	10^{-1}	11	2	B.92
Marginal Proposto	10	2	10^{-3}	11	2	B.93
Marginal Proposto	10	2	10^{-5}	11	2	B.94
Marginal Proposto	10	2	0	11	2	B.95
Marginal Proposto	10	2	1	11	3	B.96
Marginal Proposto	10	2	10^{-1}	11	3	B.97
Marginal Proposto	10	2	10^{-3}	11	3	B.98
Marginal Proposto	10	2	10^{-5}	11	3	B.99
Marginal Proposto	10	2	0	11	3	B.100
Marginal Original	10	2	1	1	2	B.101
Marginal Original	10	2	10^{-1}	1	2	B.102
Marginal Original	10	2	10^{-3}	1	2	B.103
Marginal Original	10	2	10^{-5}	1	2	B.104
Marginal Original	10	2	0	1	2	B.105
Marginal Original	10	2	1	1	3	B.106
Marginal Original	10	2	10^{-1}	1	3	B.107
Marginal Original	10	2	10^{-3}	1	3	B.108
Marginal Original	10	2	10^{-5}	1	3	B.109
Marginal Original	10	2	0	1	3	B.110

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Proposto	10	2	1	1	2	B.111
Marginal Proposto	10	2	10^{-1}	1	2	B.112
Marginal Proposto	10	2	10^{-3}	1	2	B.113
Marginal Proposto	10	2	10^{-5}	1	2	B.114
Marginal Proposto	10	2	0	1	2	B.115
Marginal Proposto	10	2	1	1	3	B.116
Marginal Proposto	10	2	10^{-1}	1	3	B.117
Marginal Proposto	10	2	10^{-3}	1	3	B.118
Marginal Proposto	10	2	10^{-5}	1	3	B.119
Marginal Proposto	10	2	0	1	3	B.120
Marginal Original	10	2	1	13	2	B.121
Marginal Original	10	2	10^{-1}	13	2	B.122
Marginal Original	10	2	10^{-3}	13	2	B.123
Marginal Original	10	2	10^{-5}	13	2	B.124
Marginal Original	10	2	0	13	2	B.125
Marginal Original	10	2	1	13	3	B.126
Marginal Original	10	2	10^{-1}	13	3	B.127
Marginal Original	10	2	10^{-3}	13	3	B.128
Marginal Original	10	2	10^{-5}	13	3	B.129
Marginal Original	10	2	0	13	3	B.130
Marginal Proposto	10	2	1	13	2	B.131
Marginal Proposto	10	2	10^{-1}	13	2	B.132
Marginal Proposto	10	2	10^{-3}	13	2	B.133
Marginal Proposto	10	2	10^{-5}	13	2	B.134
Marginal Proposto	10	2	0	13	2	B.135
Marginal Proposto	10	2	1	13	3	B.136
Marginal Proposto	10	2	10^{-1}	13	3	B.137
Marginal Proposto	10	2	10^{-3}	13	3	B.138
Marginal Proposto	10	2	10^{-5}	13	3	B.139
Marginal Proposto	10	2	0	13	3	B.140

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Conjunto	10	5	1	-	2	B.141
Conjunto	10	5	10^{-1}	-	2	B.142
Conjunto	10	5	10^{-3}	-	2	B.143
Conjunto	10	5	10^{-5}	-	2	B.144
Conjunto	10	5	0	-	2	B.145
Conjunto	10	5	1	-	3	B.146
Conjunto	10	5	10^{-1}	-	3	B.147
Conjunto	10	5	10^{-3}	-	3	B.148
Conjunto	10	5	10^{-5}	-	3	B.149
Conjunto	10	5	0	-	3	B.150
Marginal Original	10	5	1	11	2	B.151
Marginal Original	10	5	10^{-1}	11	2	B.152
Marginal Original	10	5	10^{-3}	11	2	B.153
Marginal Original	10	5	10^{-5}	11	2	B.154
Marginal Original	10	5	0	11	2	B.155
Marginal Original	10	5	1	11	3	B.156
Marginal Original	10	5	10^{-1}	11	3	B.157
Marginal Original	10	5	10^{-3}	11	3	B.158
Marginal Original	10	5	10^{-5}	11	3	B.159
Marginal Original	10	5	0	11	3	B.160
Marginal Proposto	10	5	1	11	2	B.161
Marginal Proposto	10	5	10^{-1}	11	2	B.162
Marginal Proposto	10	5	10^{-3}	11	2	B.163
Marginal Proposto	10	5	10^{-5}	11	2	B.164
Marginal Proposto	10	5	0	11	2	B.165
Marginal Proposto	10	5	1	11	3	B.166
Marginal Proposto	10	5	10^{-1}	11	3	B.167
Marginal Proposto	10	5	10^{-3}	11	3	B.168
Marginal Proposto	10	5	10^{-5}	11	3	B.169
Marginal Proposto	10	5	0	11	3	B.170

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Original	10	5	1	1	2	B.171
Marginal Original	10	5	10^{-1}	1	2	B.172
Marginal Original	10	5	10^{-3}	1	2	B.173
Marginal Original	10	5	10^{-5}	1	2	B.174
Marginal Original	10	5	0	1	2	B.175
Marginal Original	10	5	1	1	3	B.176
Marginal Original	10	5	10^{-1}	1	3	B.177
Marginal Original	10	5	10^{-3}	1	3	B.178
Marginal Original	10	5	10^{-5}	1	3	B.179
Marginal Original	10	5	0	1	3	B.180
Marginal Proposto	10	5	1	1	2	B.181
Marginal Proposto	10	5	10^{-1}	1	2	B.182
Marginal Proposto	10	5	10^{-3}	1	2	B.183
Marginal Proposto	10	5	10^{-5}	1	2	B.184
Marginal Proposto	10	5	0	1	2	B.185
Marginal Proposto	10	5	1	1	3	B.186
Marginal Proposto	10	5	10^{-1}	1	3	B.187
Marginal Proposto	10	5	10^{-3}	1	3	B.188
Marginal Proposto	10	5	10^{-5}	1	3	B.189
Marginal Proposto	10	5	0	1	3	B.190
Marginal Original	10	5	1	16	2	B.191
Marginal Original	10	5	10^{-1}	16	2	B.192
Marginal Original	10	5	10^{-3}	16	2	B.193
Marginal Original	10	5	10^{-5}	16	2	B.194
Marginal Original	10	5	0	16	2	B.195
Marginal Original	10	5	1	16	3	B.196
Marginal Original	10	5	10^{-1}	16	3	B.197
Marginal Original	10	5	10^{-3}	16	3	B.198
Marginal Original	10	5	10^{-5}	16	3	B.199
Marginal Original	10	5	0	16	3	B.200

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Proposto	10	5	1	16	2	B.201
Marginal Proposto	10	5	10^{-1}	16	2	B.202
Marginal Proposto	10	5	10^{-3}	16	2	B.203
Marginal Proposto	10	5	10^{-5}	16	2	B.204
Marginal Proposto	10	5	0	16	2	B.205
Marginal Proposto	10	5	1	16	3	B.206
Marginal Proposto	10	5	10^{-1}	16	3	B.207
Marginal Proposto	10	5	10^{-3}	16	3	B.208
Marginal Proposto	10	5	10^{-5}	16	3	B.209
Marginal Proposto	10	5	0	16	3	B.210

B.2 Gráficos

B.2.1 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste conjunto

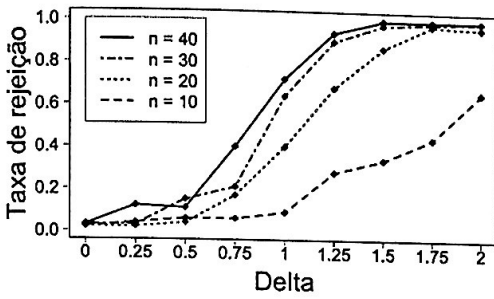


Figura B.1: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

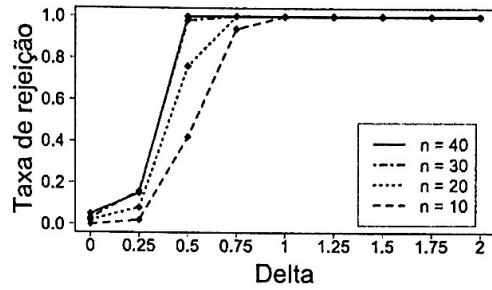


Figura B.2: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

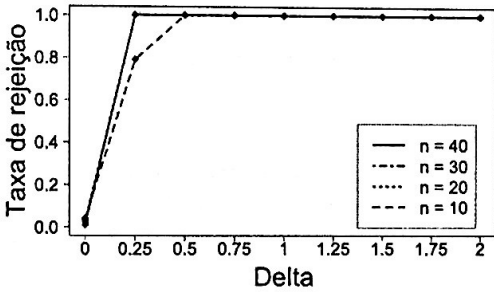


Figura B.3: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

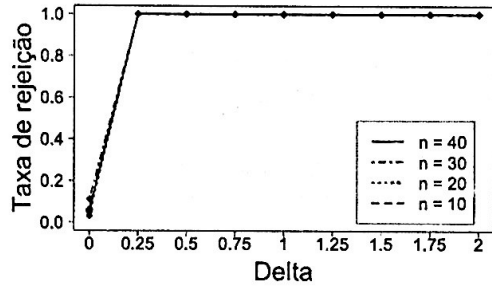


Figura B.4: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

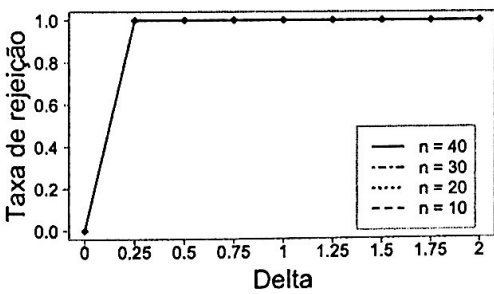


Figura B.5: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

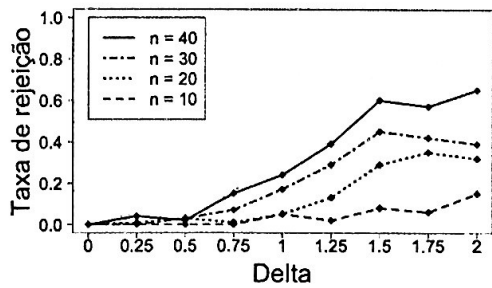


Figura B.6: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

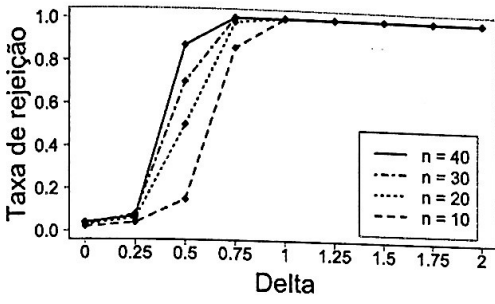


Figura B.7: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

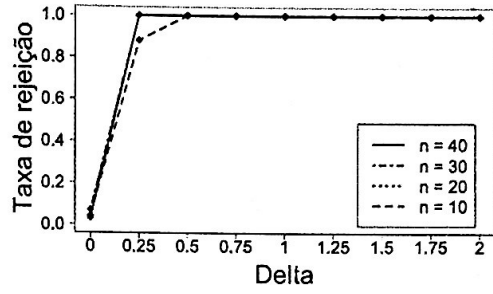


Figura B.8: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

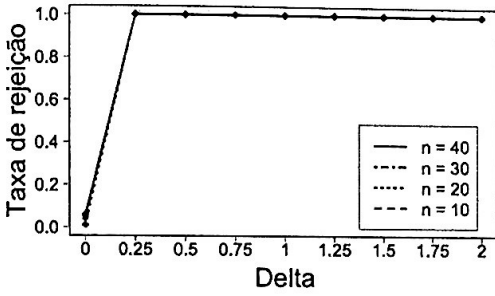


Figura B.9: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

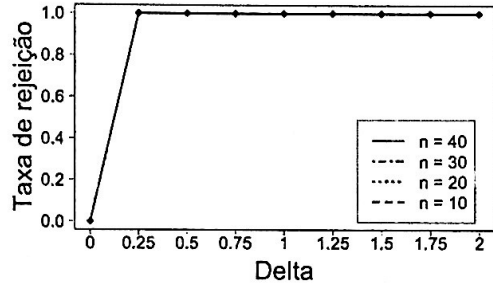


Figura B.10: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.2 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 11$

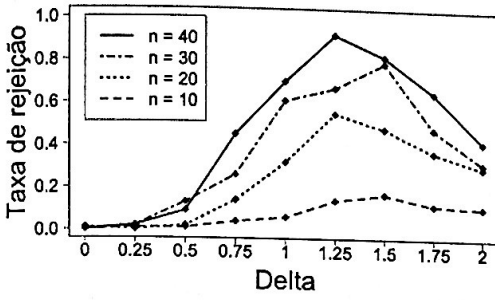


Figura B.11: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusters.

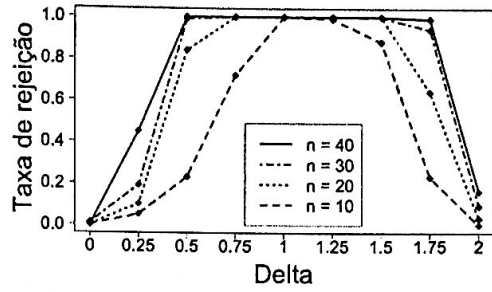


Figura B.12: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusters.

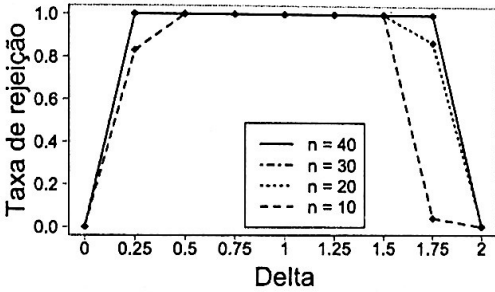


Figura B.13: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusters.

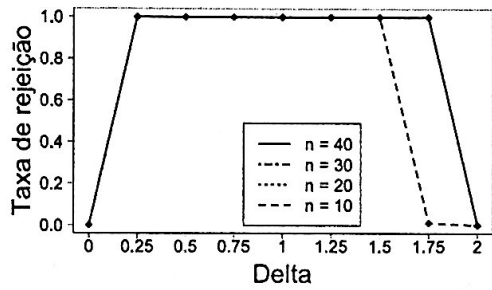


Figura B.14: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusters.

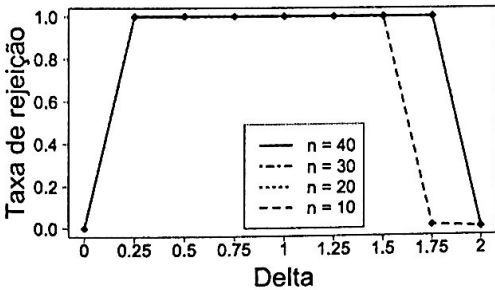


Figura B.15: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusters.

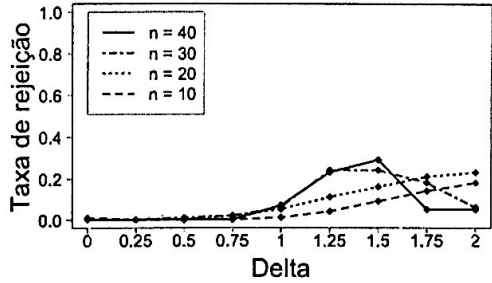


Figura B.16: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusters.

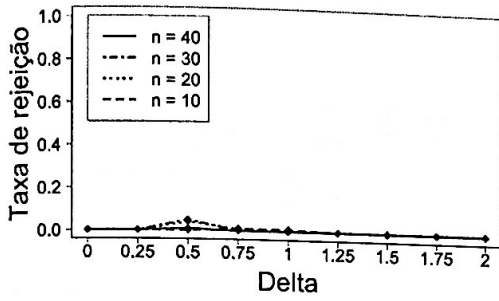


Figura B.17: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

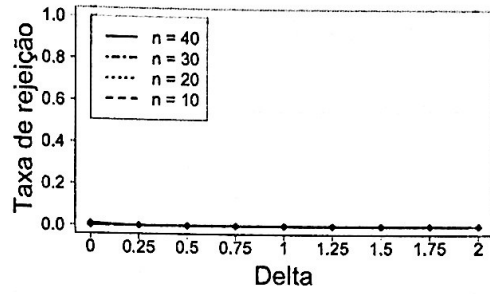


Figura B.18: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

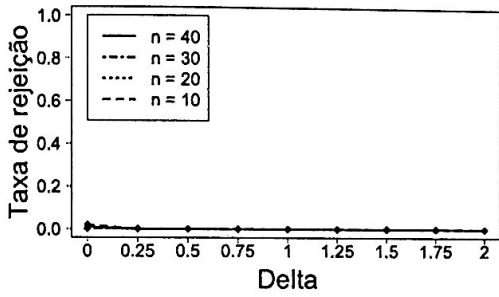


Figura B.19: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

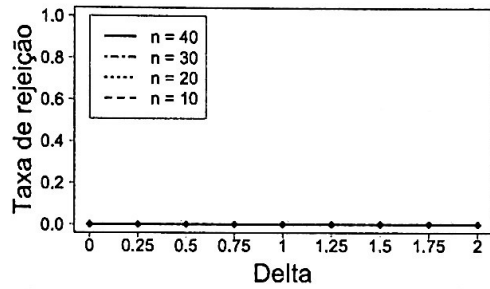


Figura B.20: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.3 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 11$

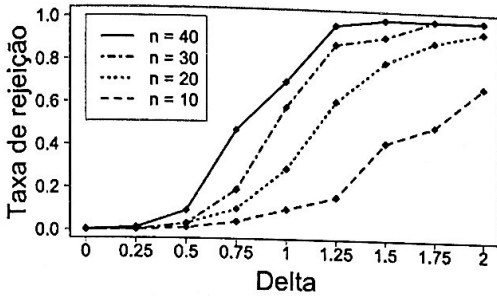


Figura B.21: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusters.

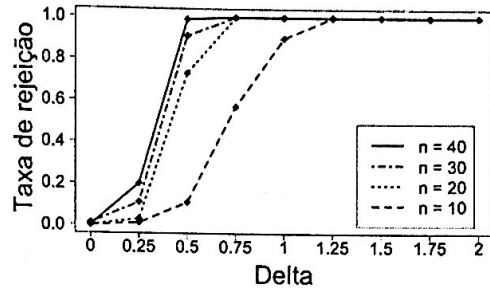


Figura B.22: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusters.

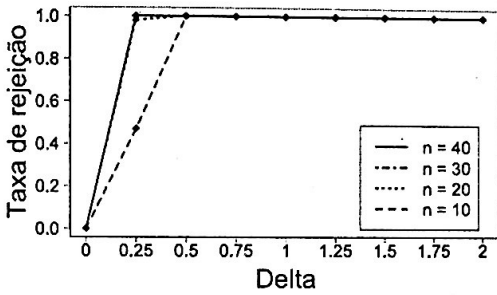


Figura B.23: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusters.

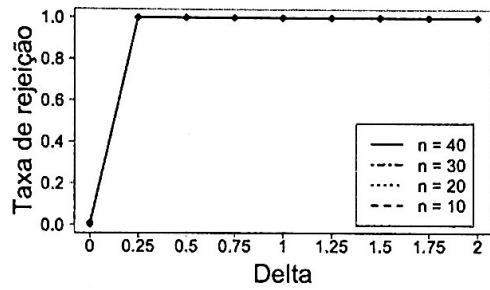


Figura B.24: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusters.

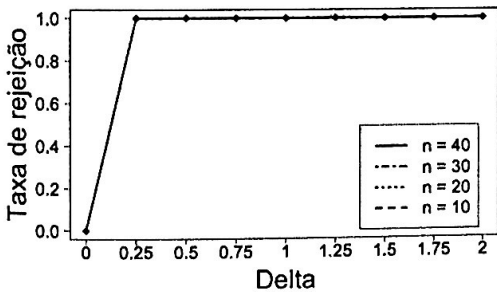


Figura B.25: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusters.

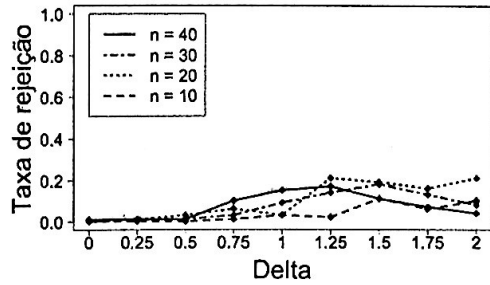


Figura B.26: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusters.

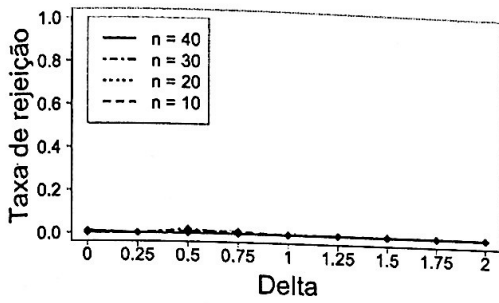


Figura B.27: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

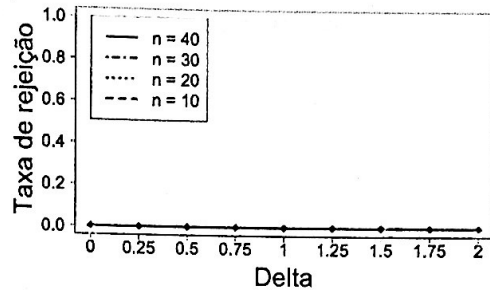


Figura B.28: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

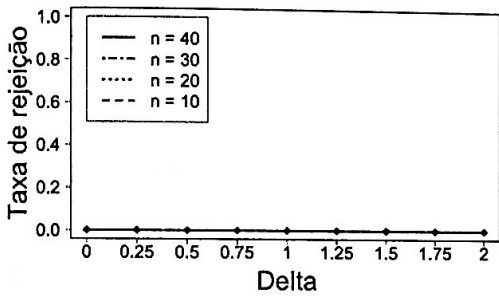


Figura B.29: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

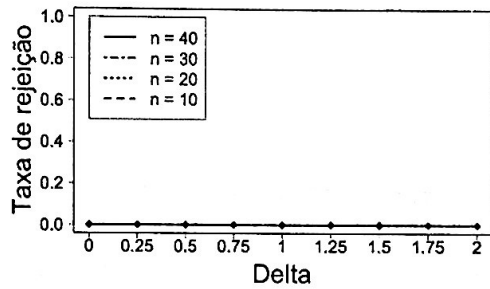


Figura B.30: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.4 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 1$

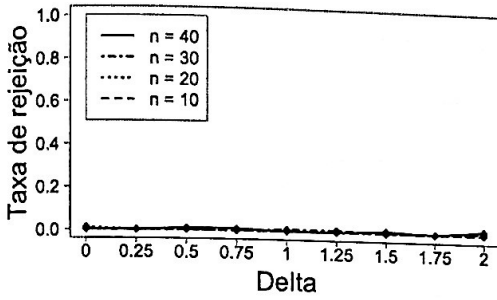


Figura B.31: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusters.

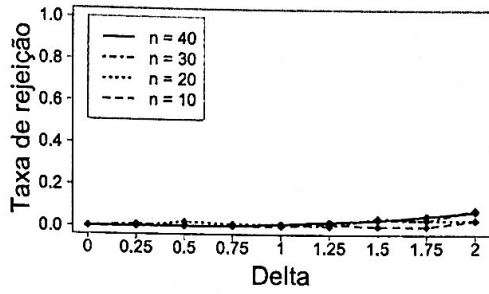


Figura B.32: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusters.

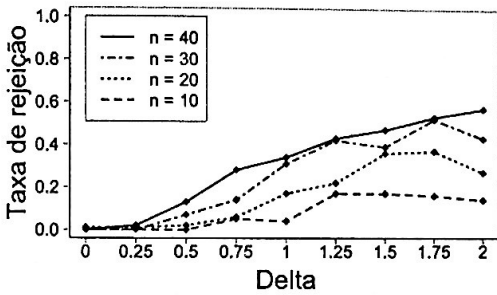


Figura B.33: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusters.

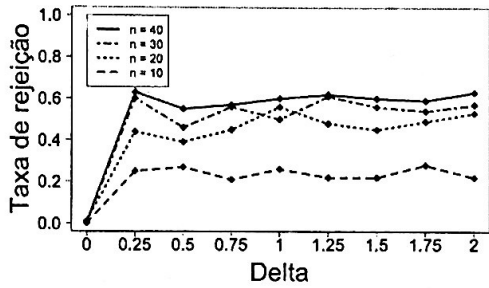


Figura B.34: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusters.

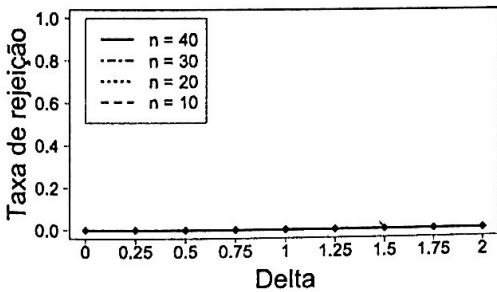


Figura B.35: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusters.

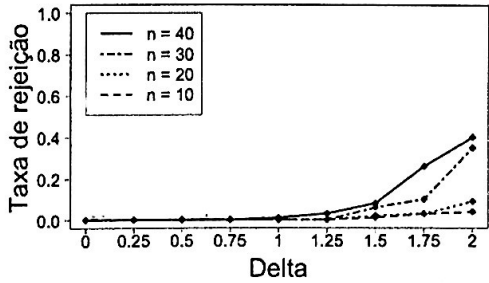


Figura B.36: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusters.

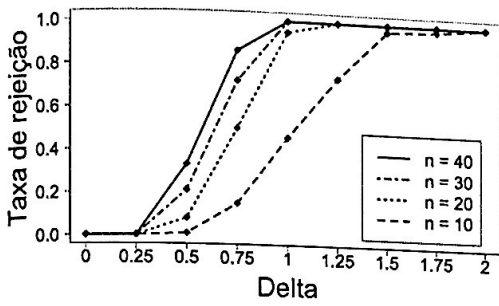


Figura B.37: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

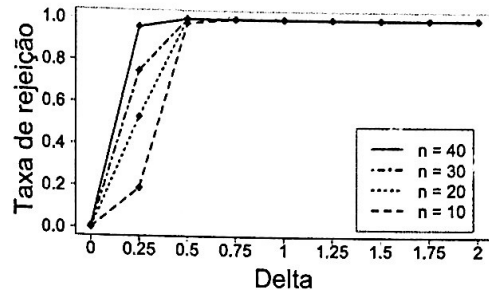


Figura B.38: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

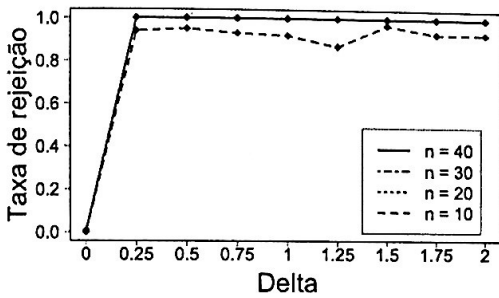


Figura B.39: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

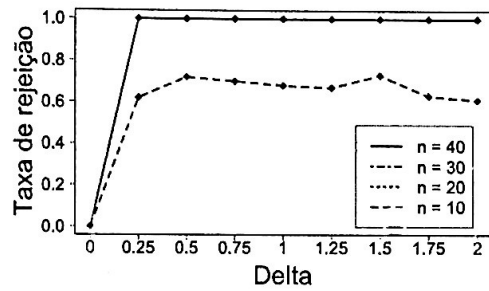


Figura B.40: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.5 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 1$

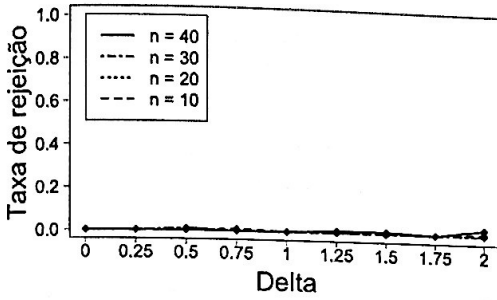


Figura B.41: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

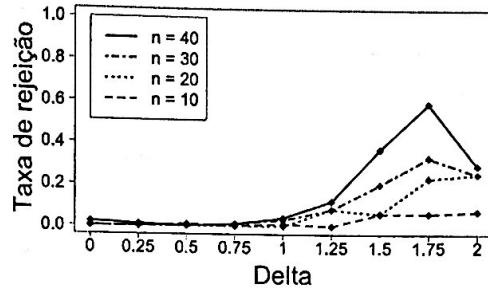


Figura B.42: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

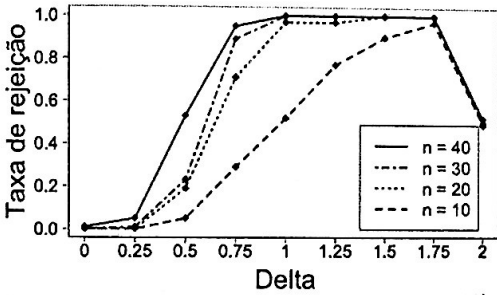


Figura B.43: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

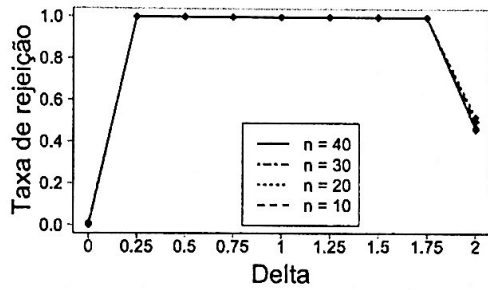


Figura B.44: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

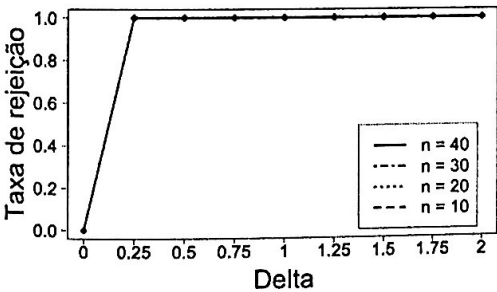


Figura B.45: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

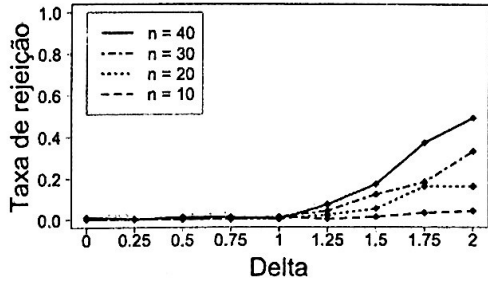


Figura B.46: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

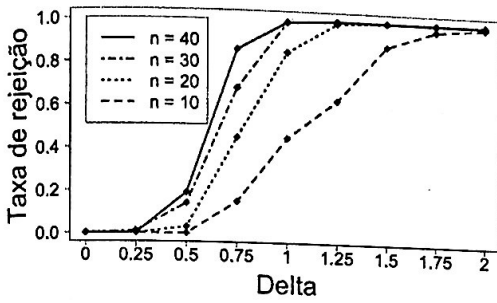


Figura B.47: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusters.

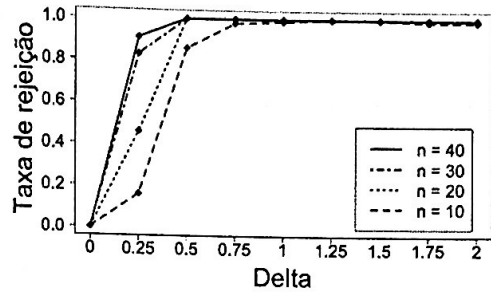


Figura B.48: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusters.

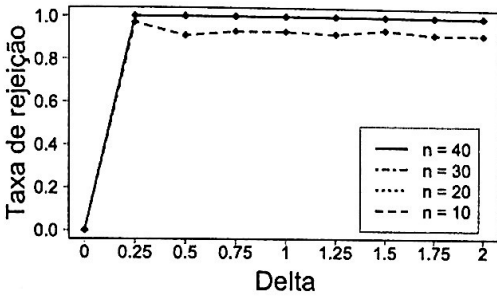


Figura B.49: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusters.

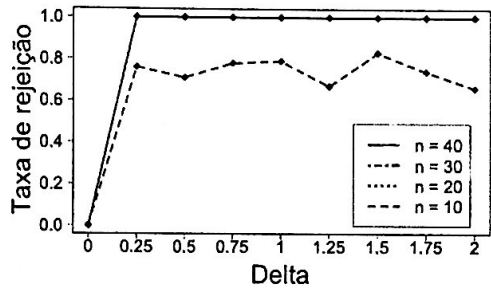


Figura B.50: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusters.

B.2.6 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal original para $q = 12$

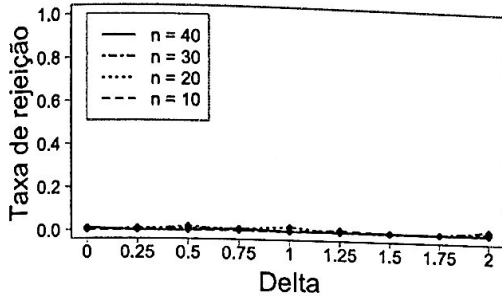


Figura B.51: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

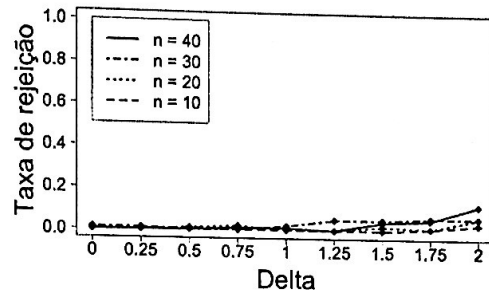


Figura B.52: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

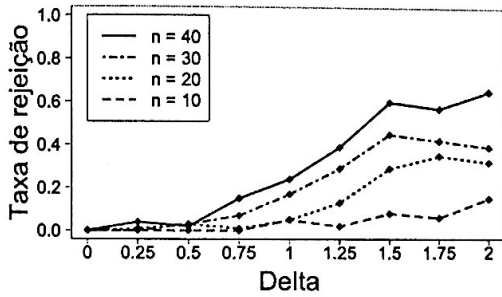


Figura B.53: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

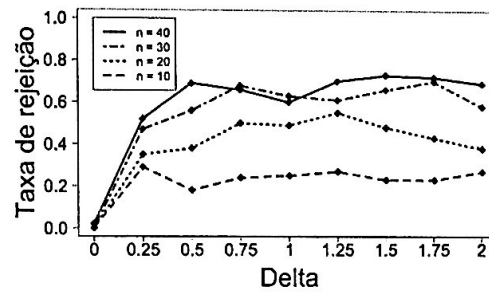


Figura B.54: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

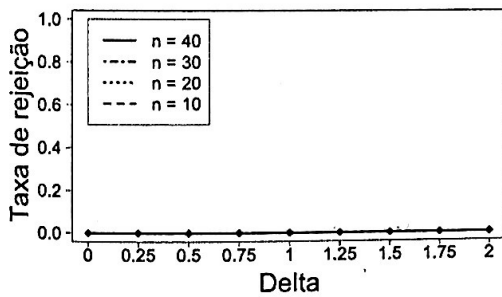


Figura B.55: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

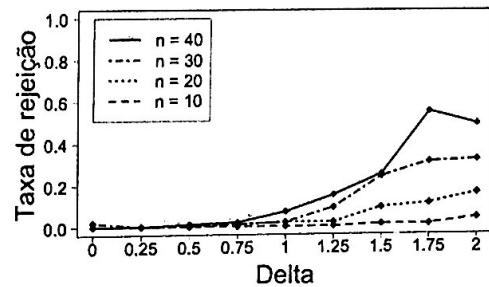


Figura B.56: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

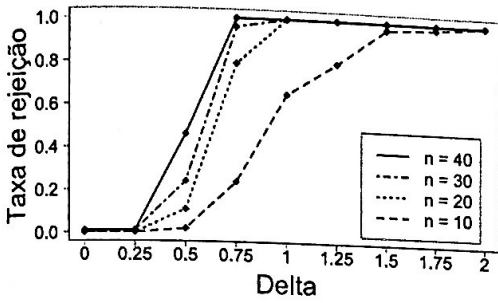


Figura B.57: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

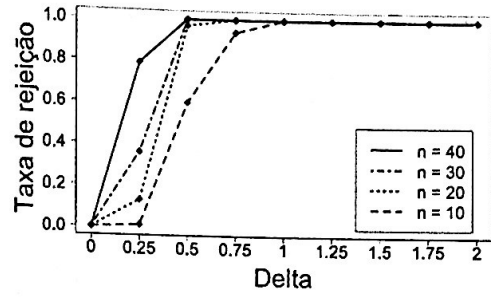


Figura B.58: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

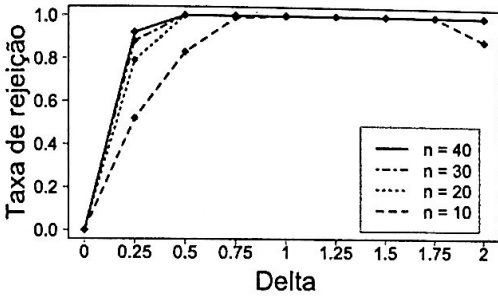


Figura B.59: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

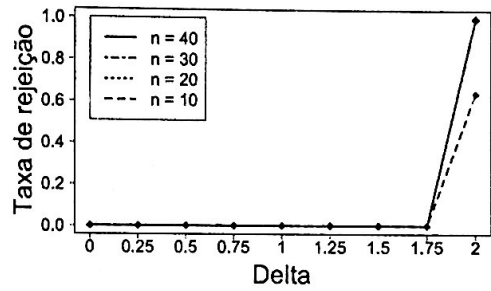


Figura B.60: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.7 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 12$

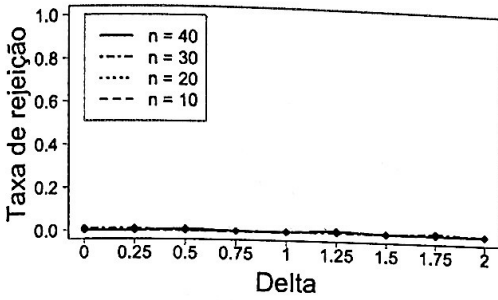


Figura B.61: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusters.

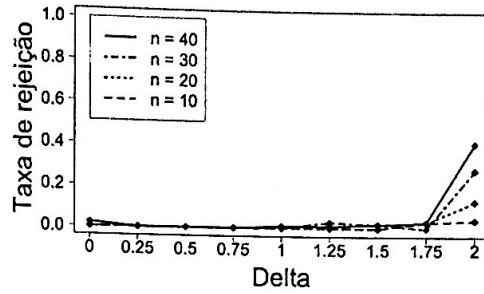


Figura B.62: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusters.

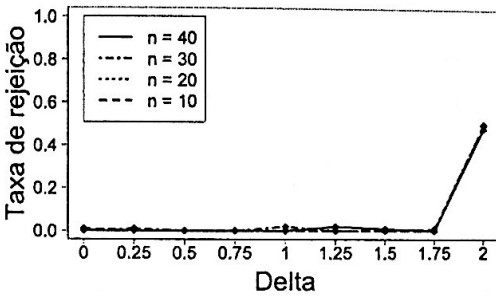


Figura B.63: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusters.

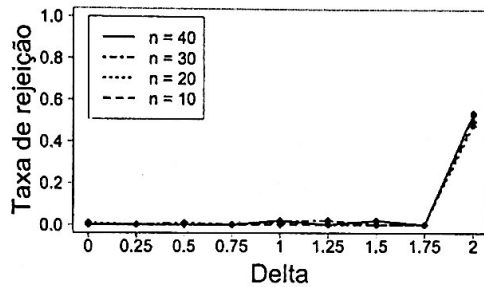


Figura B.64: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusters.

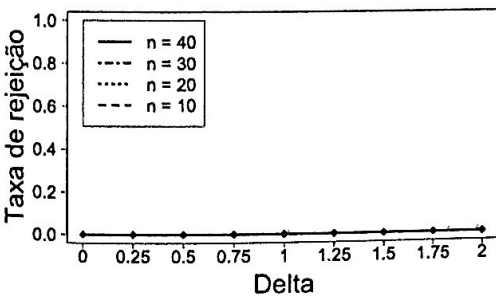


Figura B.65: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusters.

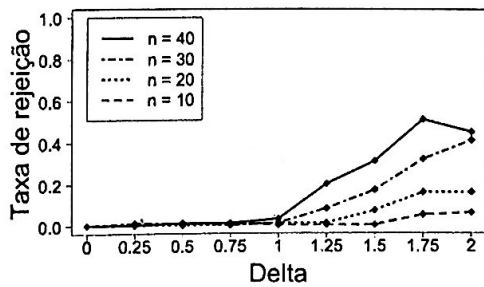


Figura B.66: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusters.

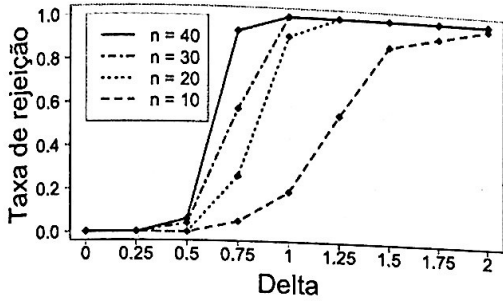


Figura B.67: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

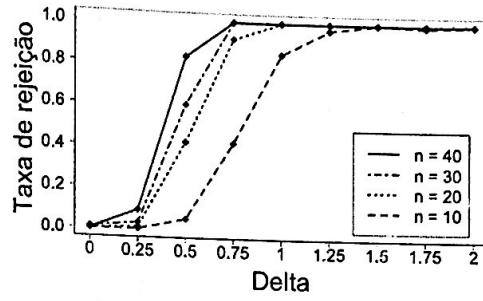


Figura B.68: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

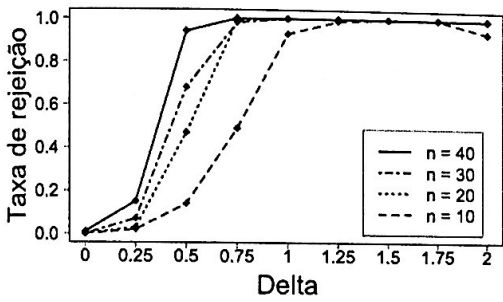


Figura B.69: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

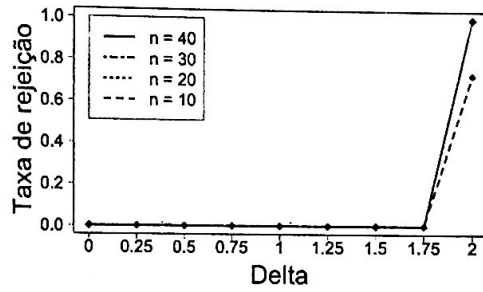


Figura B.70: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

B.2.8 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste conjunto

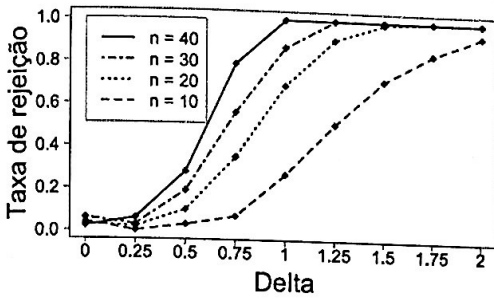


Figura B.71: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusteres.

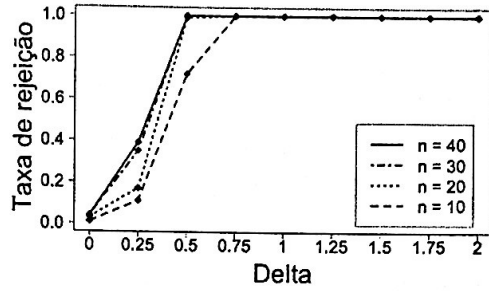


Figura B.72: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusteres.

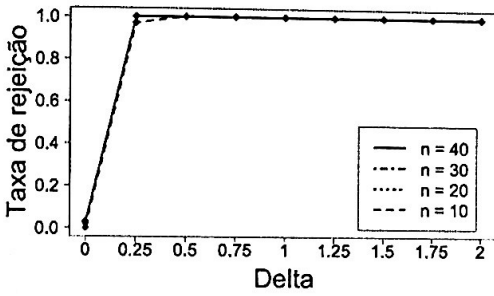


Figura B.73: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusteres.

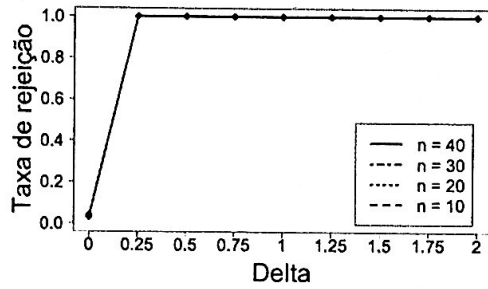


Figura B.74: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusteres.

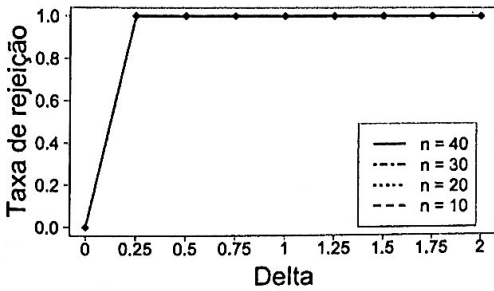


Figura B.75: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusteres.

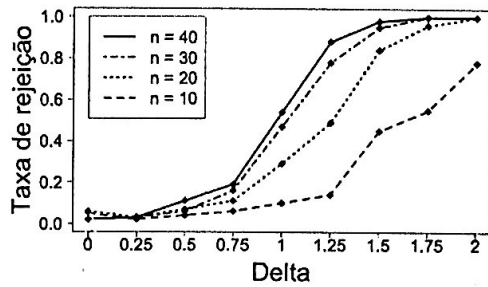


Figura B.76: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusteres.

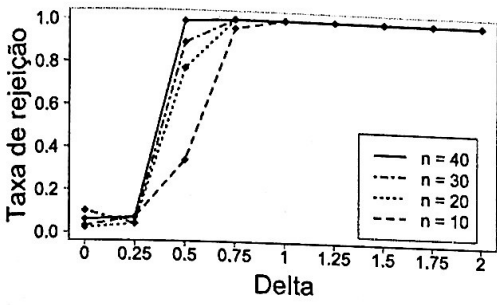


Figura B.77: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

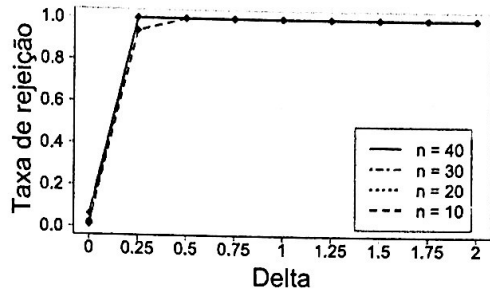


Figura B.78: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

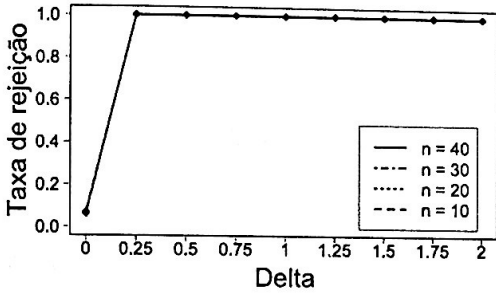


Figura B.79: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

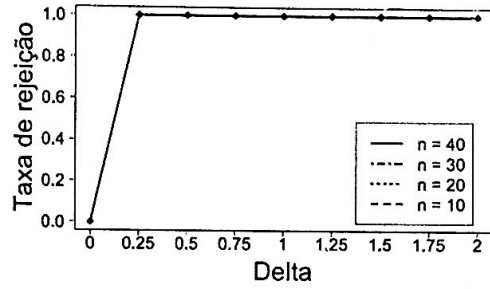


Figura B.80: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.9 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 11$

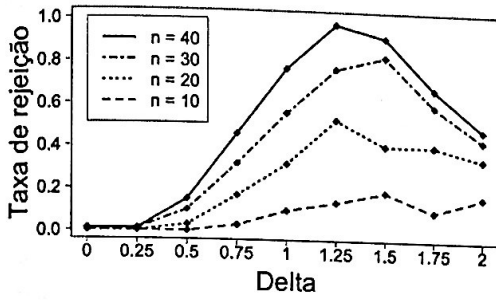


Figura B.81: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusteres.

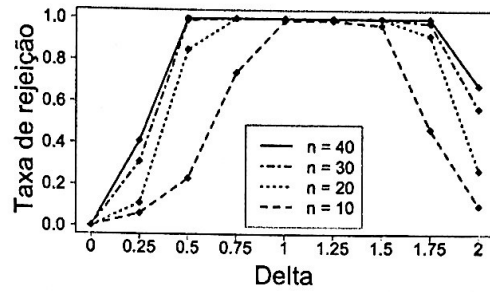


Figura B.82: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusteres.

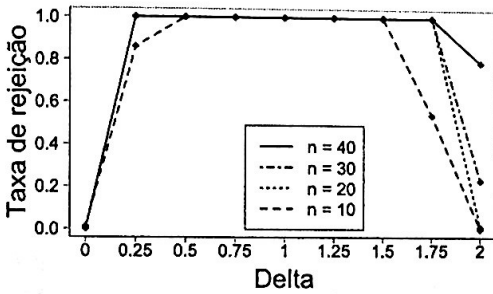


Figura B.83: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusteres.

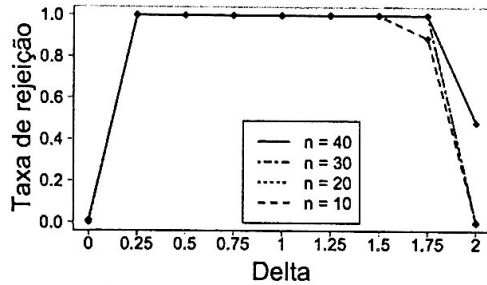


Figura B.84: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusteres.

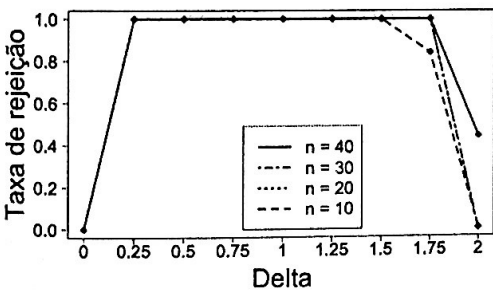


Figura B.85: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusteres.

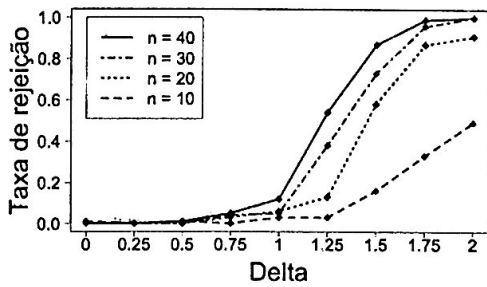


Figura B.86: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusteres.

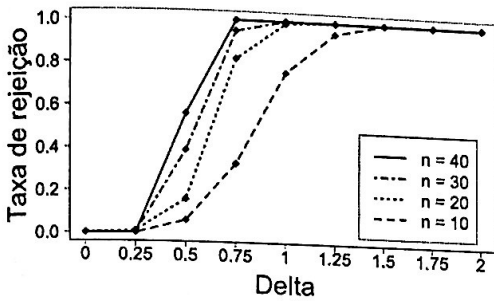


Figura B.87: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

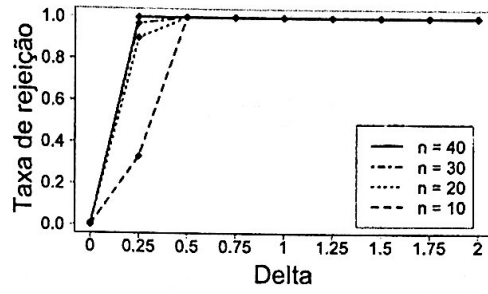


Figura B.88: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

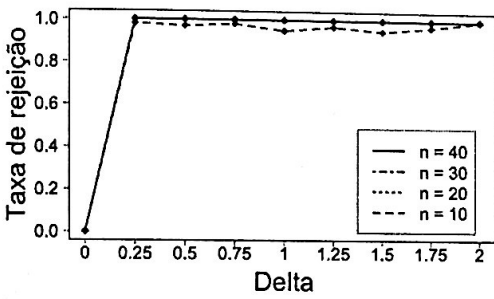


Figura B.89: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

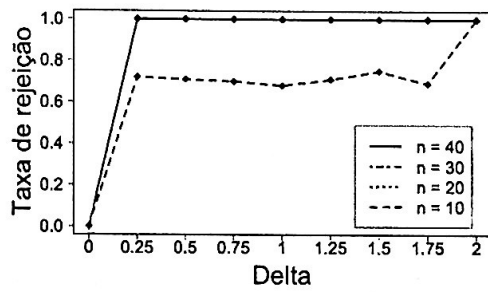


Figura B.90: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.10 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 11$

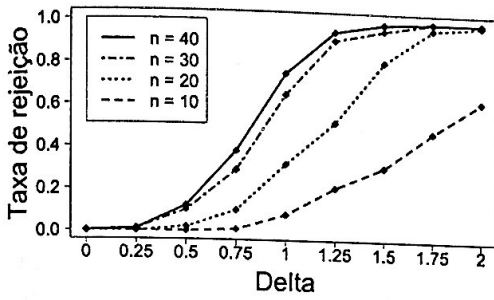


Figura B.91: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusteres.

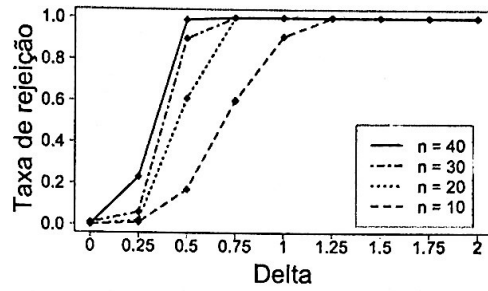


Figura B.92: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusteres.

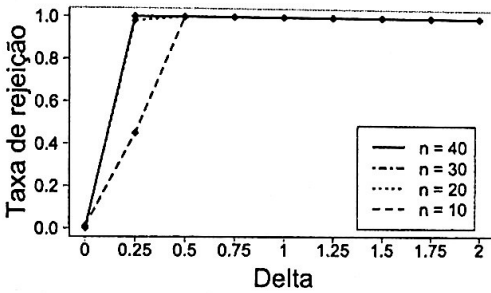


Figura B.93: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusteres.

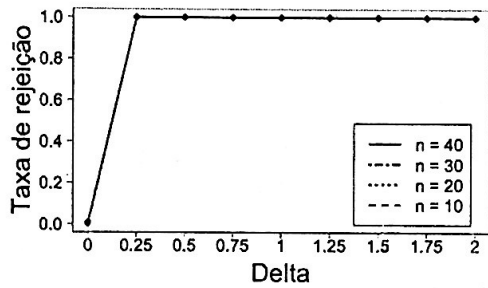


Figura B.94: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusteres.

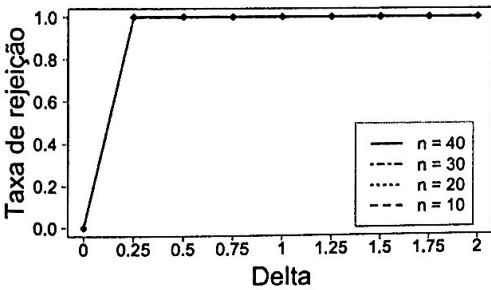


Figura B.95: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusteres.

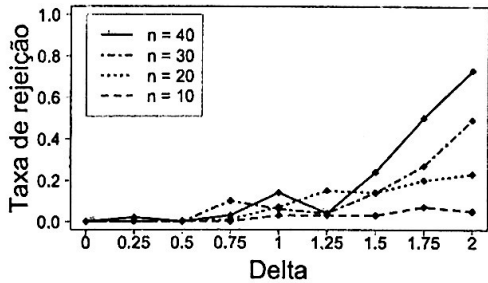


Figura B.96: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusteres.

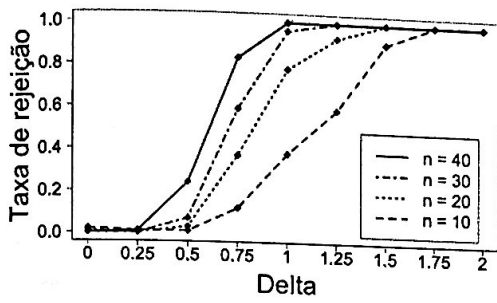


Figura B.97: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

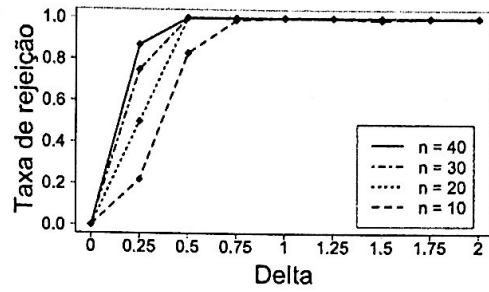


Figura B.98: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

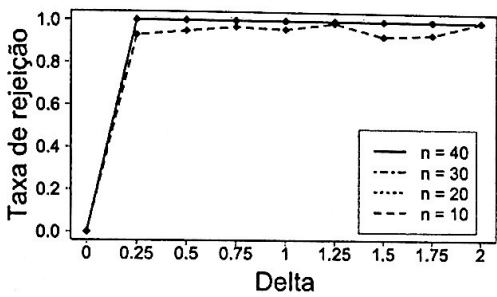


Figura B.99: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

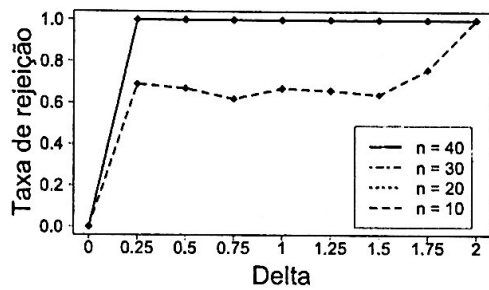


Figura B.100: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.11 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 1$

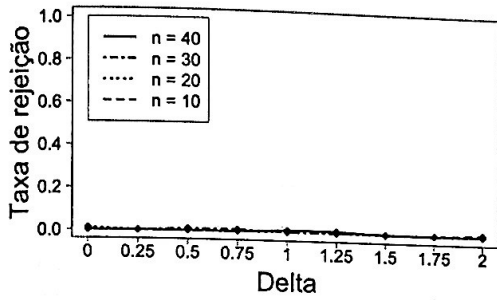


Figura B.101: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusteres.

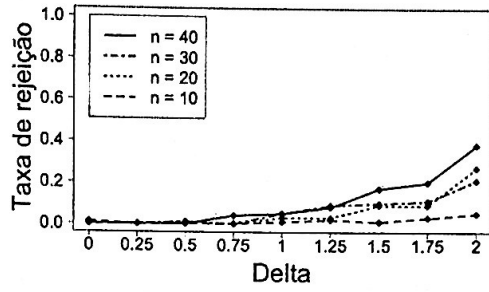


Figura B.102: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusteres.

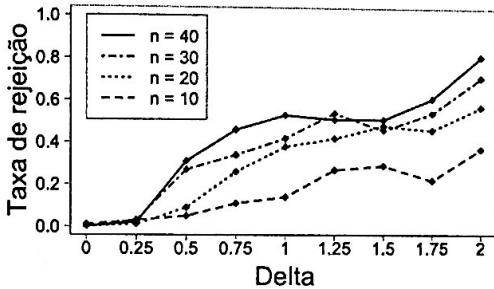


Figura B.103: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusteres.

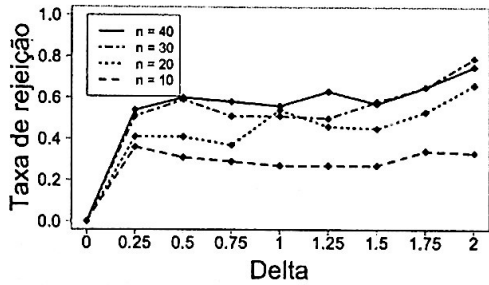


Figura B.104: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusteres.

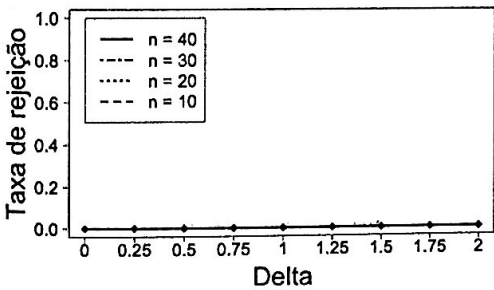


Figura B.105: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusteres.

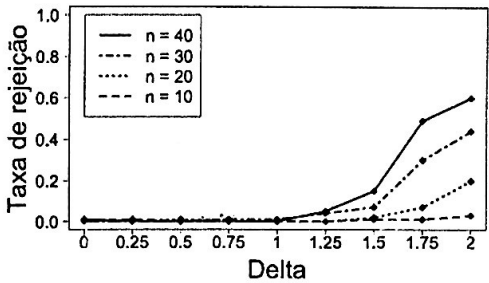


Figura B.106: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusteres.

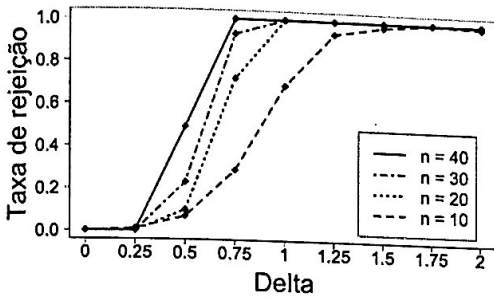


Figura B.107: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

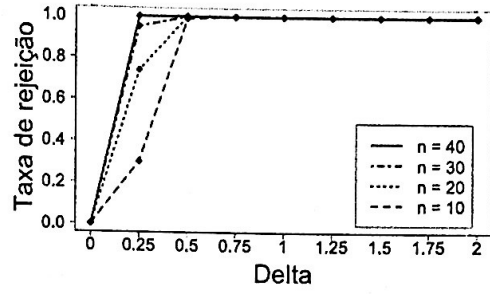


Figura B.108: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

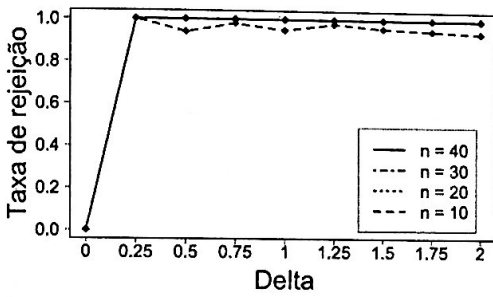


Figura B.109: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

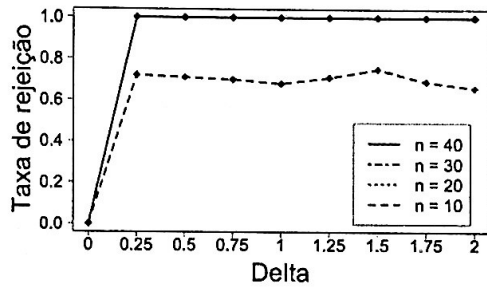


Figura B.110: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.12 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 1$

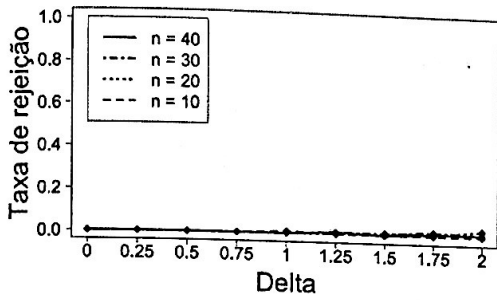


Figura B.111: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusteres.

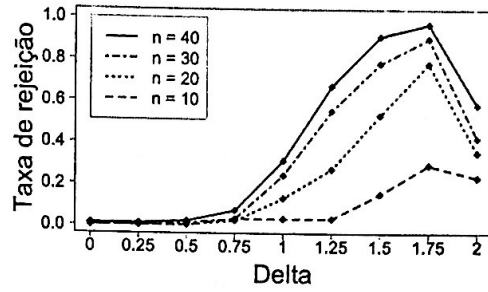


Figura B.112: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusteres.

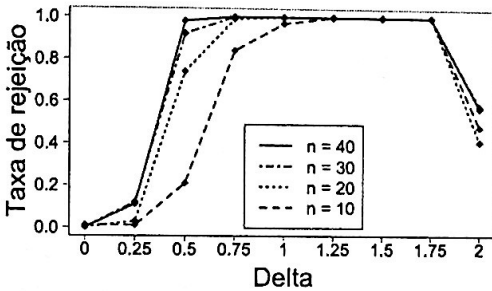


Figura B.113: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusteres.

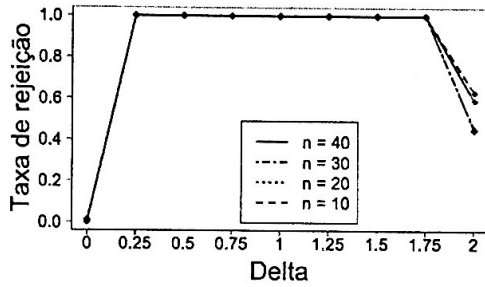


Figura B.114: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusteres.

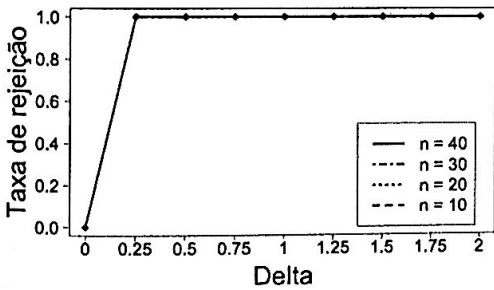


Figura B.115: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusteres.

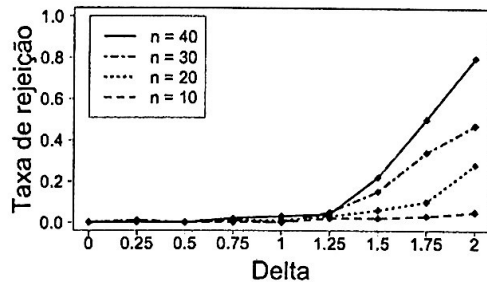


Figura B.116: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusteres.

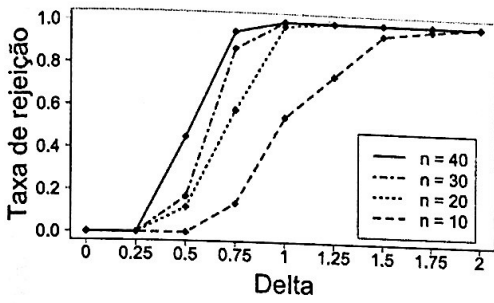


Figura B.117: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

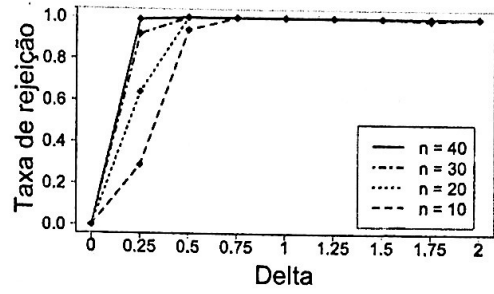


Figura B.118: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

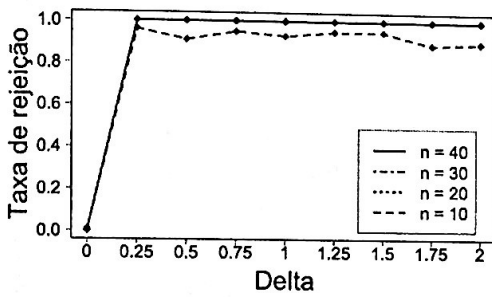


Figura B.119: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

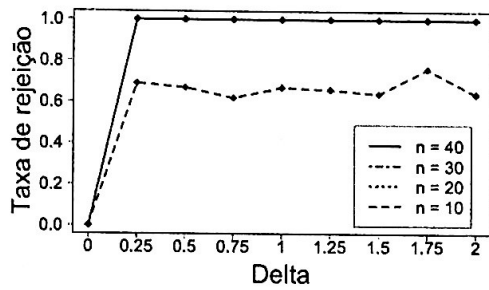


Figura B.120: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.13 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal original para $q = 13$

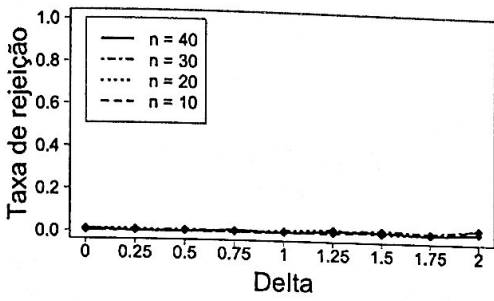


Figura B.121: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusters.

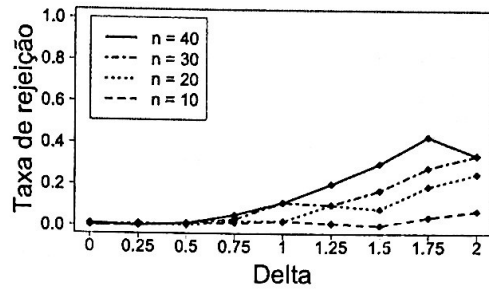


Figura B.122: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusters.

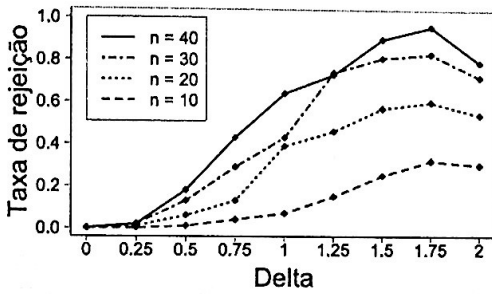


Figura B.123: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusters.

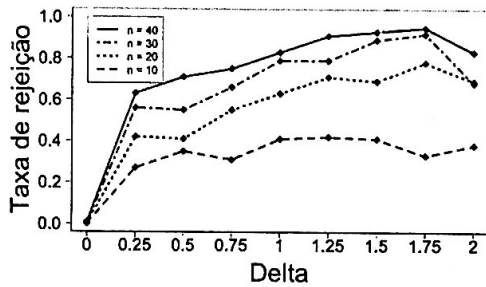


Figura B.124: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusters.

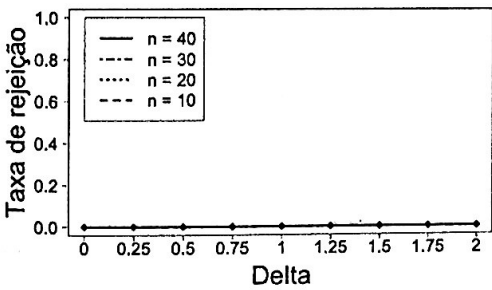


Figura B.125: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusters.

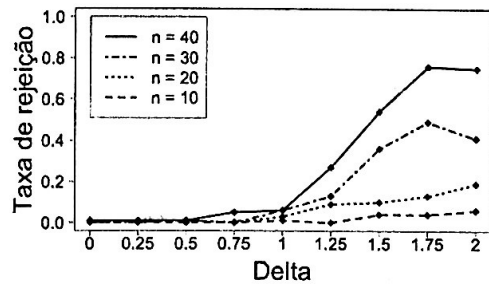


Figura B.126: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusters.

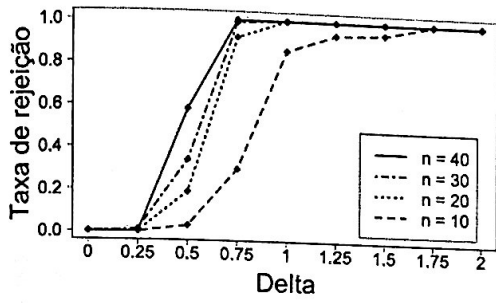


Figura B.127: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

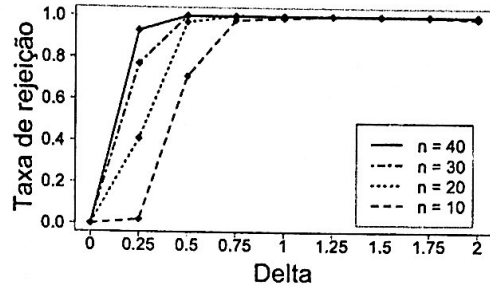


Figura B.128: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

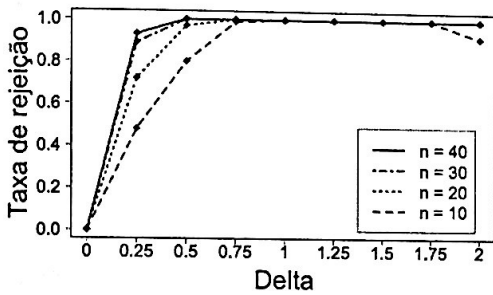


Figura B.129: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

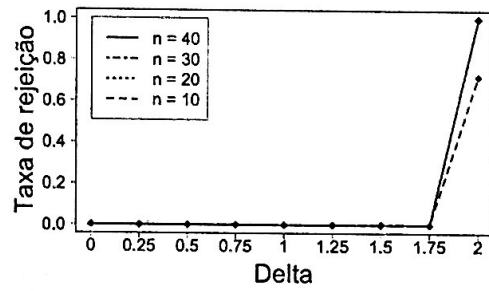


Figura B.130: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.14 Gráficos: $N_1 = 10$ e $N_2 = 2$ - Teste marginal proposto para $q = 13$

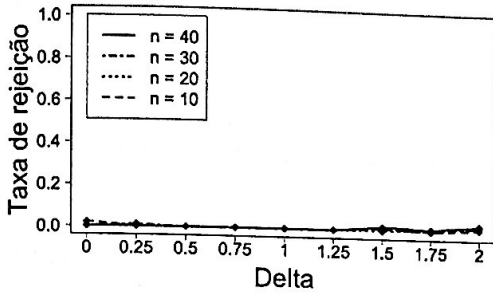


Figura B.131: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e dois clusters.

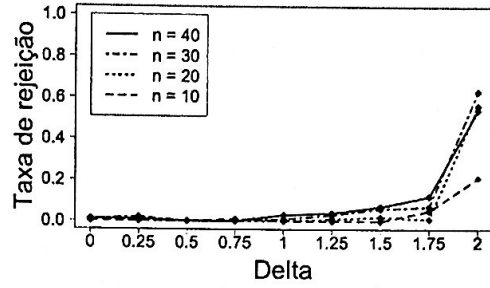


Figura B.132: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e dois clusters.

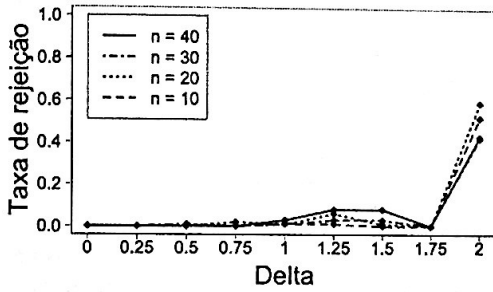


Figura B.133: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e dois clusters.

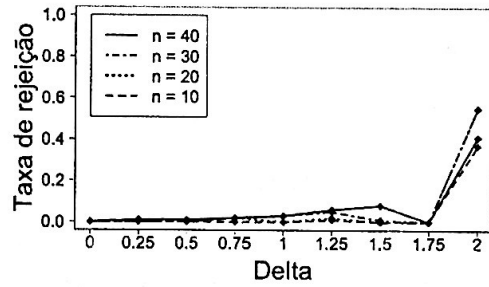


Figura B.134: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e dois clusters.

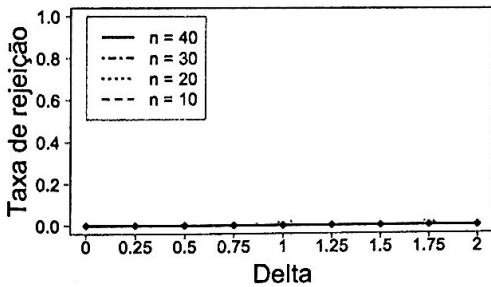


Figura B.135: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e dois clusters.

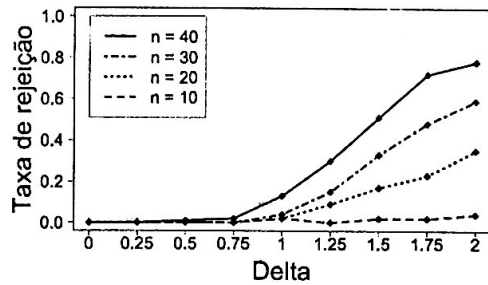


Figura B.136: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 1$ e três clusters.

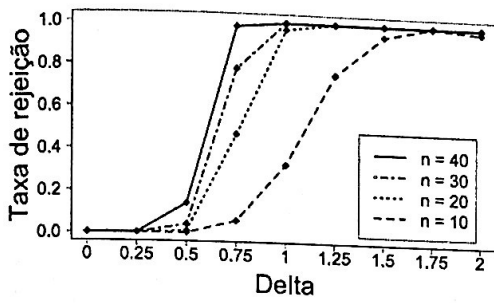


Figura B.137: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-1}$ e três clusteres.

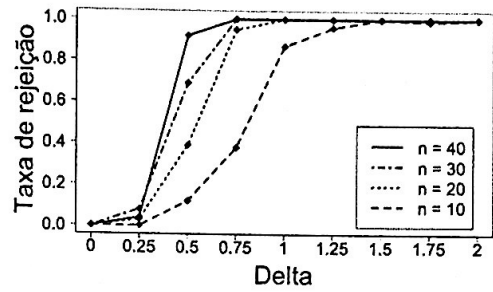


Figura B.138: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-3}$ e três clusteres.

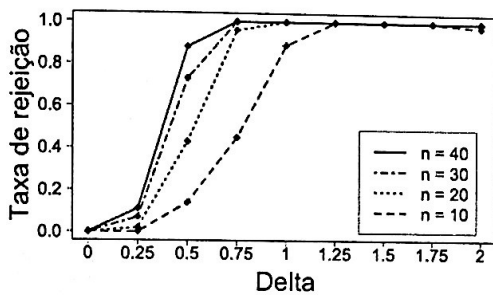


Figura B.139: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 10^{-5}$ e três clusteres.

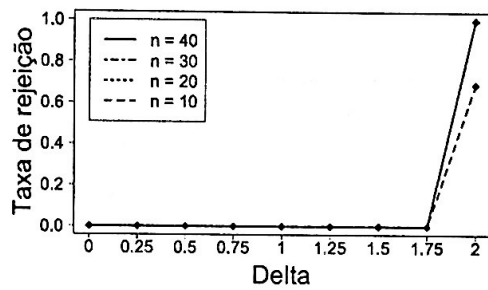


Figura B.140: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 13$ com $N_1 = 10$ e $N_2 = 2$, $\sigma^2 = 0$ e três clusteres.

B.2.15 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste conjunto

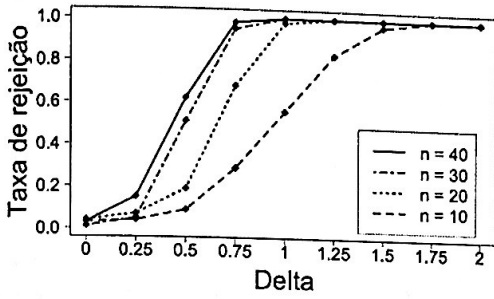


Figura B.141: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusteres.

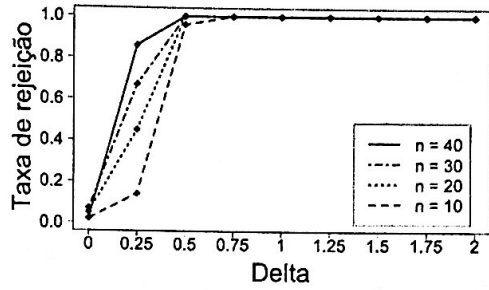


Figura B.142: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusteres.

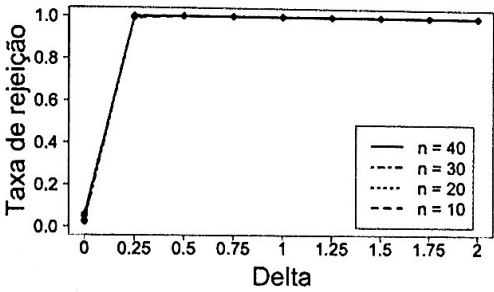


Figura B.143: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusteres.

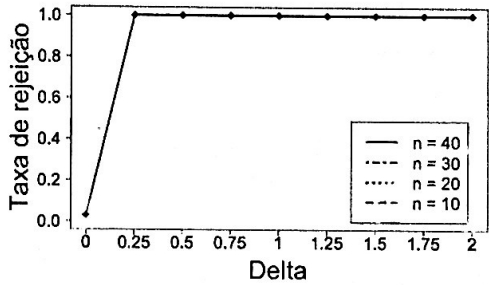


Figura B.144: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusteres.

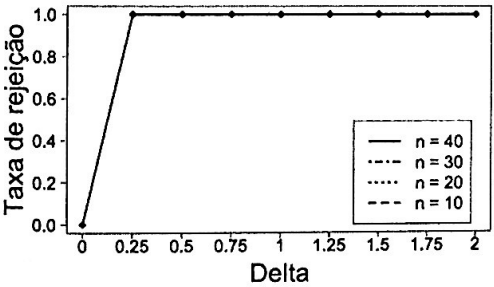


Figura B.145: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusteres.

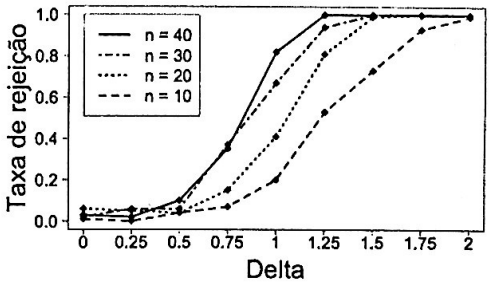


Figura B.146: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusteres.

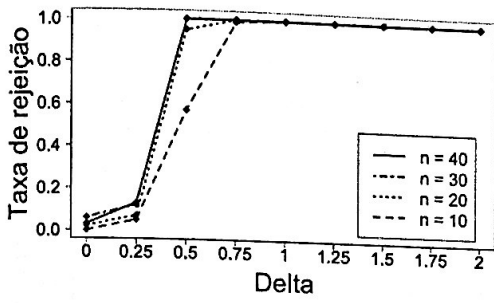


Figura B.147: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

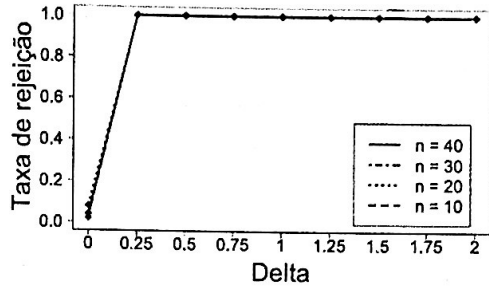


Figura B.148: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

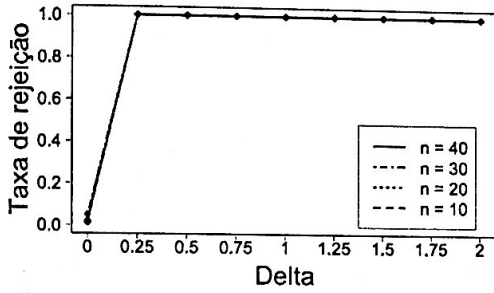


Figura B.149: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

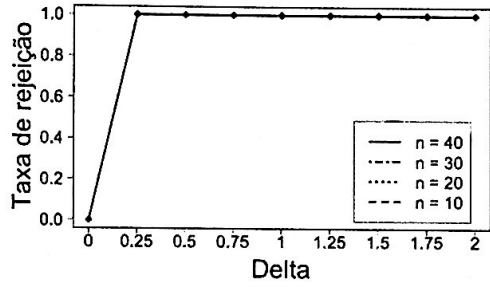


Figura B.150: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

B.2.16 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 11$

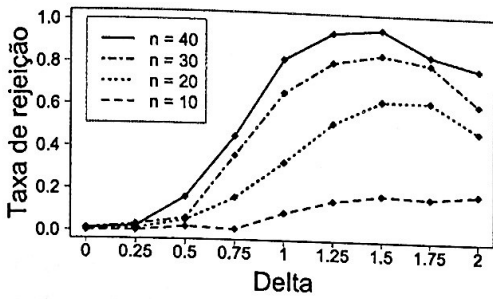


Figura B.151: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusters.

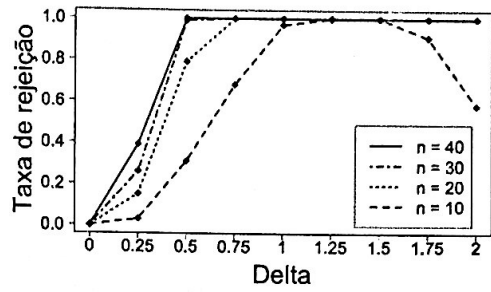


Figura B.152: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusters.

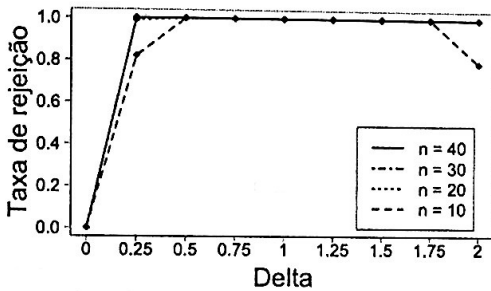


Figura B.153: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusters.

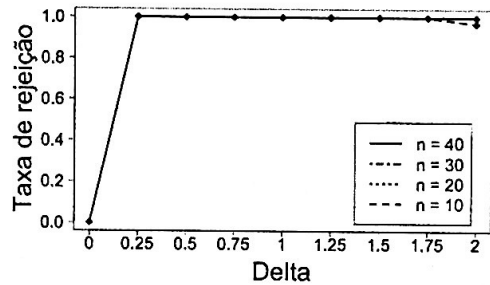


Figura B.154: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusters.

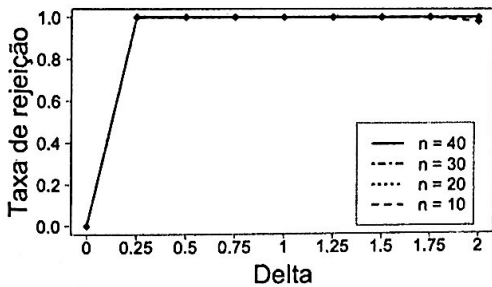


Figura B.155: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusters.

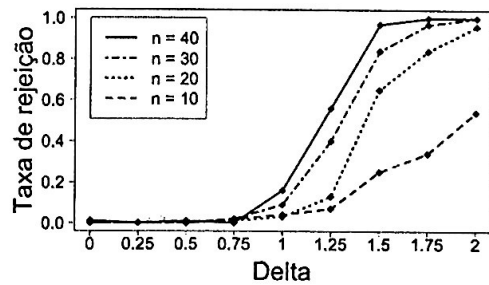


Figura B.156: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusters.

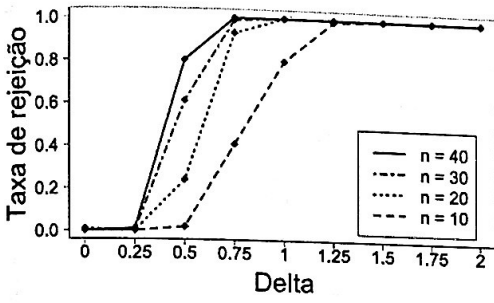


Figura B.157: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusters.

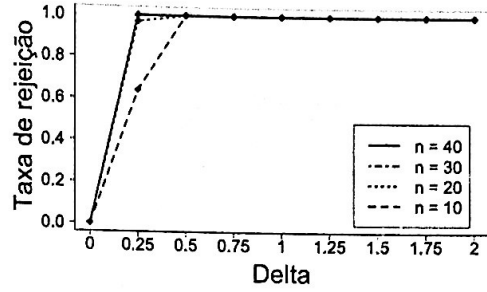


Figura B.158: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusters.

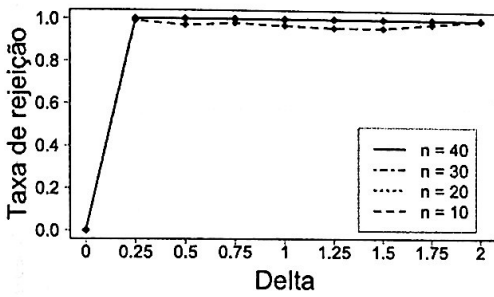


Figura B.159: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusters.

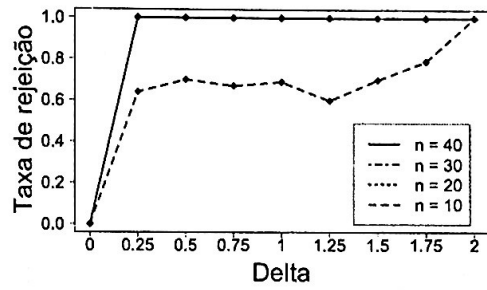


Figura B.160: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusters.

B.2.17 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 11$

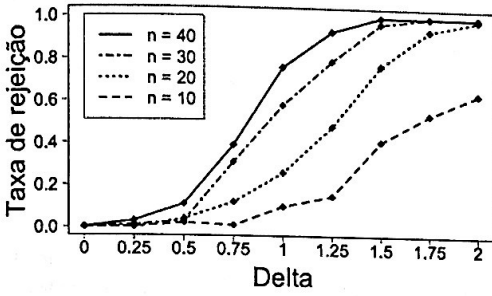


Figura B.161: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusters.

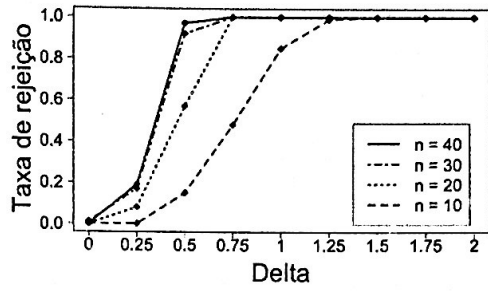


Figura B.162: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusters.

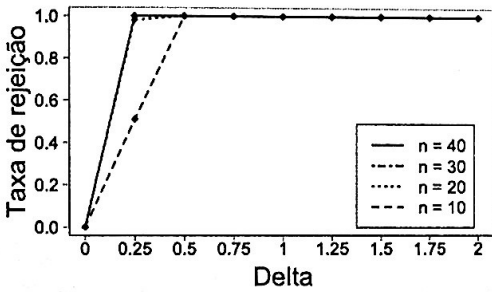


Figura B.163: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusters.

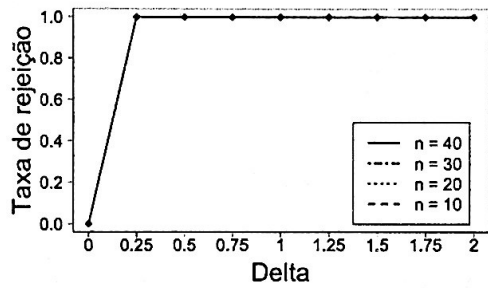


Figura B.164: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusters.

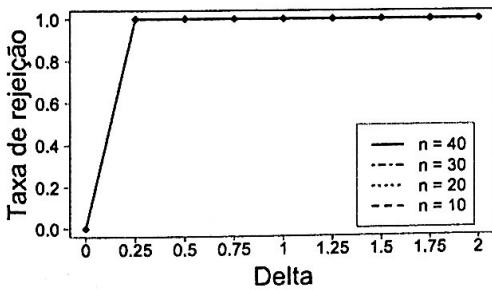


Figura B.165: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusters.

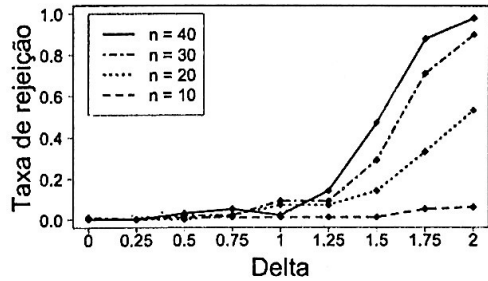


Figura B.166: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusters.

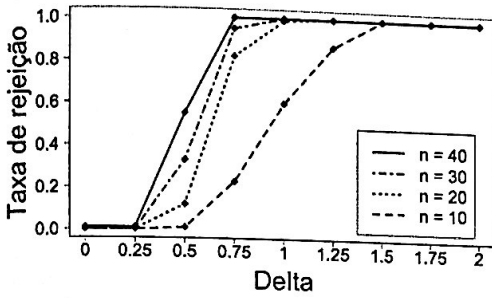


Figura B.167: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

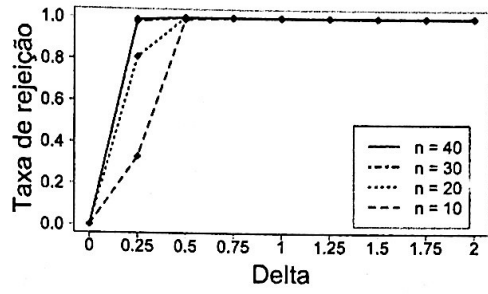


Figura B.168: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

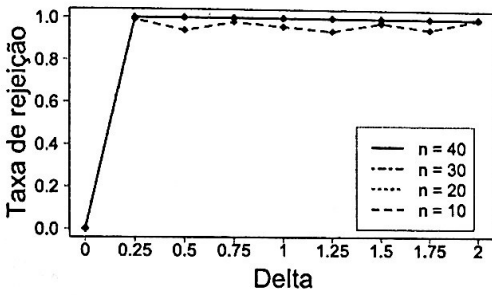


Figura B.169: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

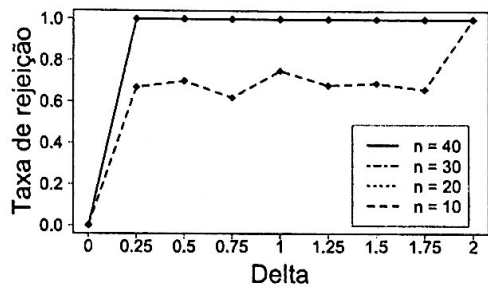


Figura B.170: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

B.2.18 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 1$

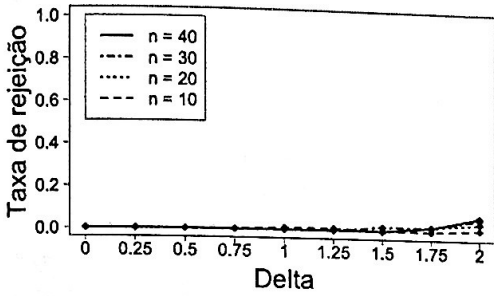


Figura B.171: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusters.

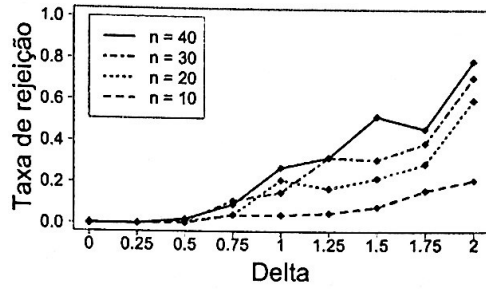


Figura B.172: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusters.

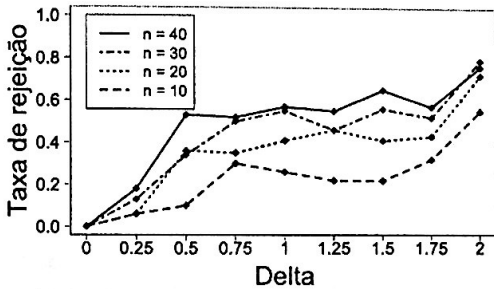


Figura B.173: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusters.

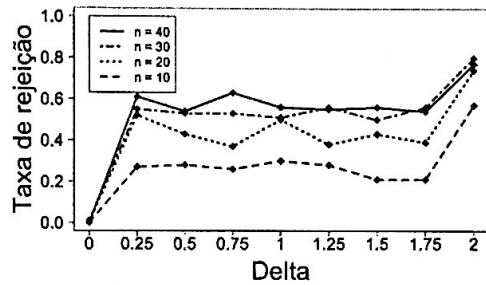


Figura B.174: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusters.

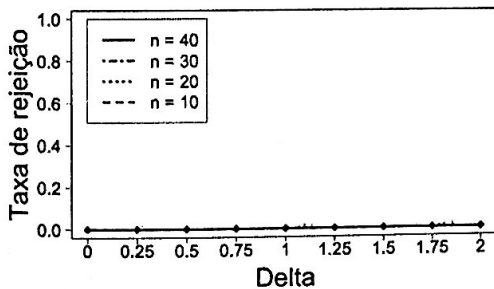


Figura B.175: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusters.

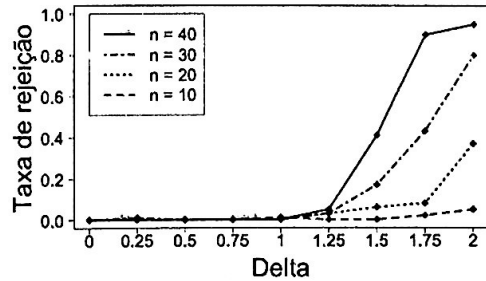


Figura B.176: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusters.

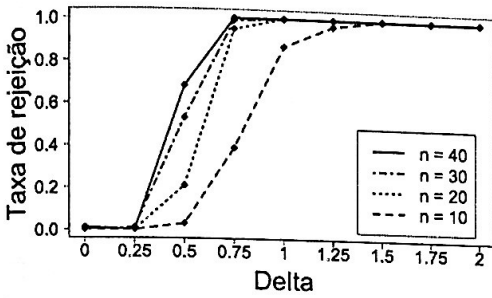


Figura B.177: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

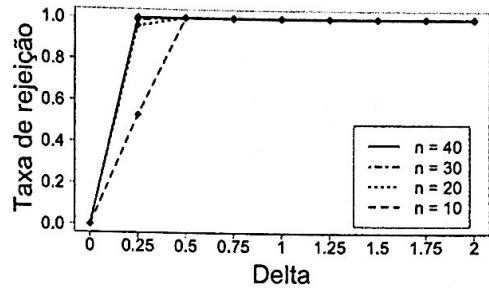


Figura B.178: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

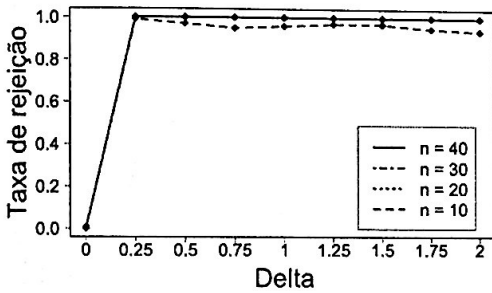


Figura B.179: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

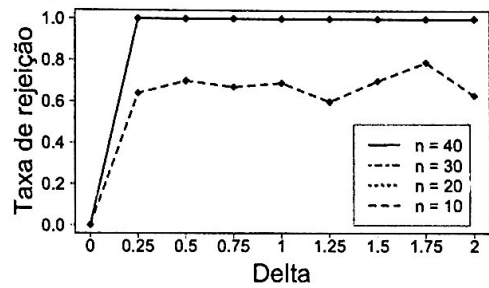


Figura B.180: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

B.2.19 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 1$

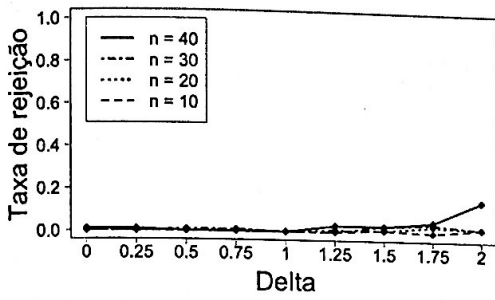


Figura B.181: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusters.

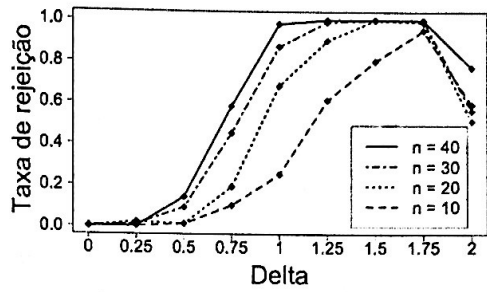


Figura B.182: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusters.

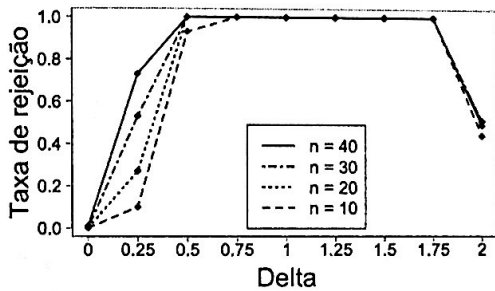


Figura B.183: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusters.

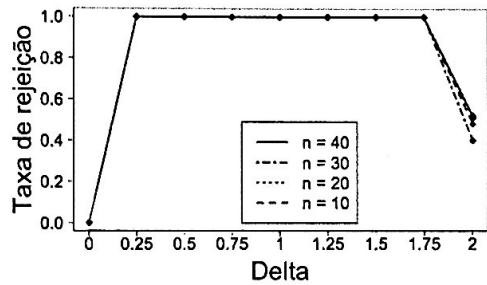


Figura B.184: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusters.

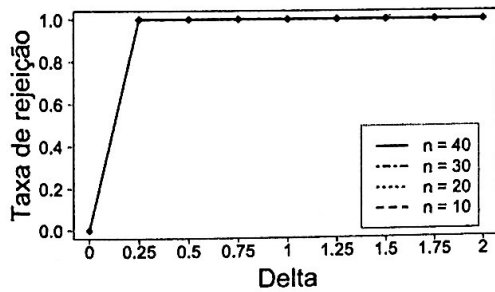


Figura B.185: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusters.

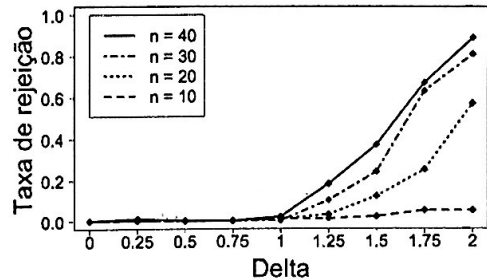


Figura B.186: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusters.

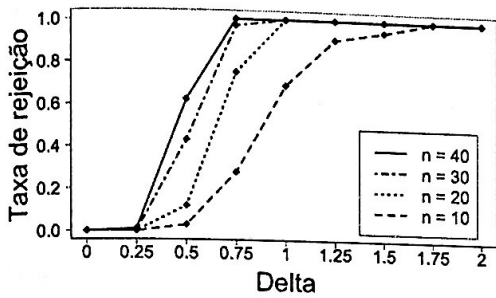


Figura B.187: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

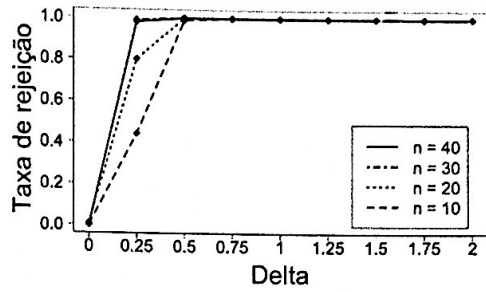


Figura B.188: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

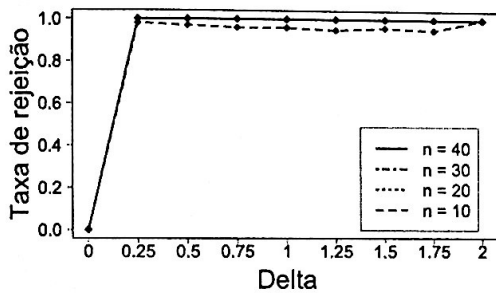


Figura B.189: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

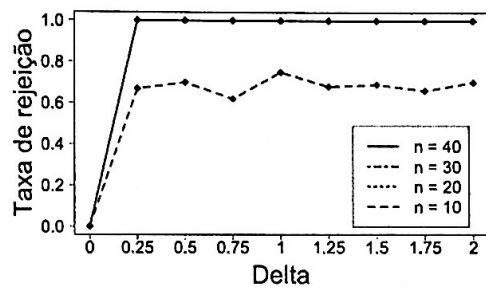


Figura B.190: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

B.2.20 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal original para $q = 16$

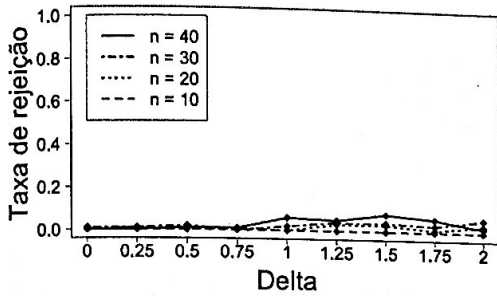


Figura B.191: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusteres.

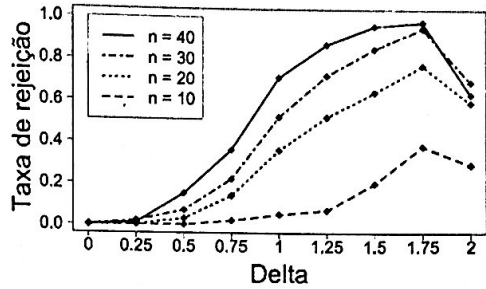


Figura B.192: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusteres.

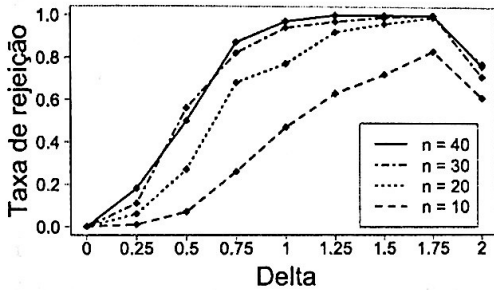


Figura B.193: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusteres.

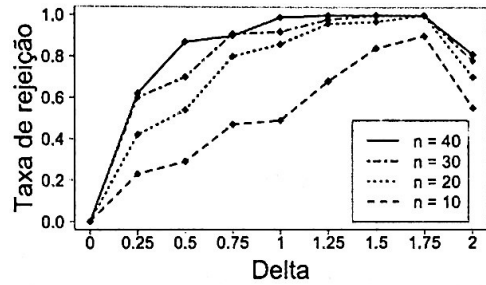


Figura B.194: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusteres.

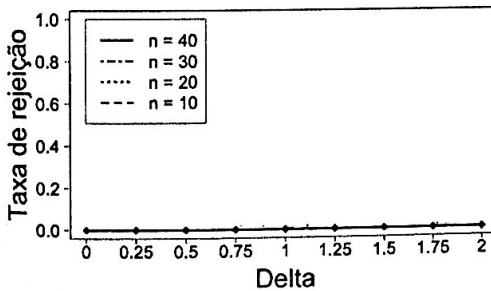


Figura B.195: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusteres.

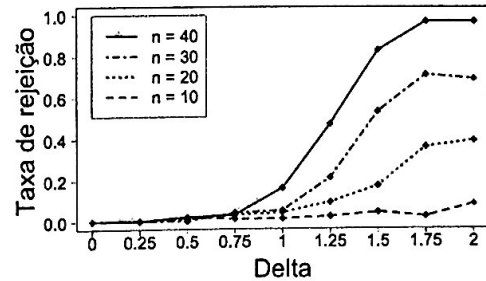


Figura B.196: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusteres.

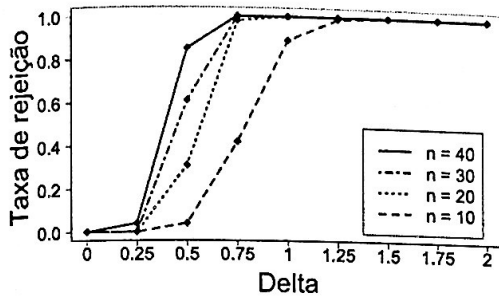


Figura B.197: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

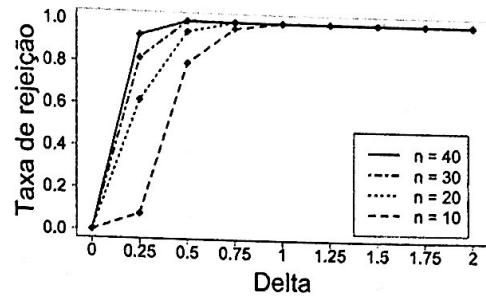


Figura B.198: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

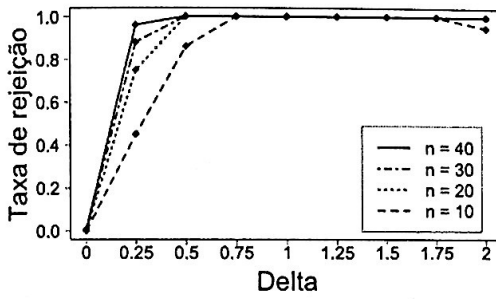


Figura B.199: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

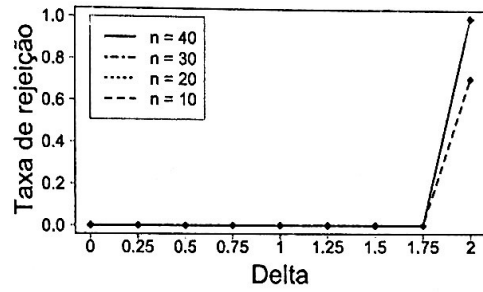


Figura B.200: Taxa de rejeição da hipótese nula no Teste marginal original para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

B.2.21 Gráficos: $N_1 = 10$ e $N_2 = 5$ - Teste marginal proposto para $q = 16$

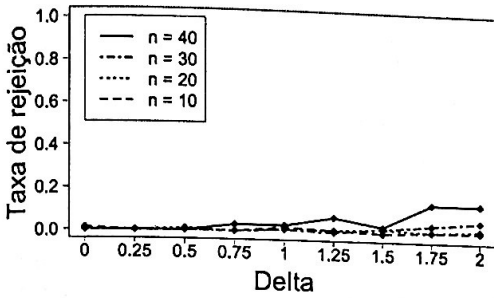


Figura B.201: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e dois clusteres.

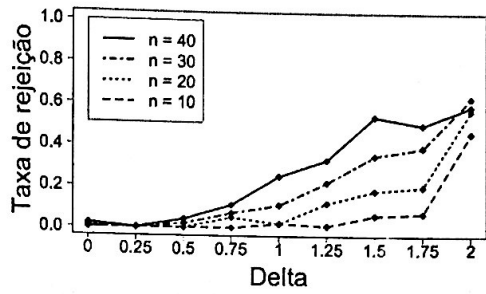


Figura B.202: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e dois clusteres.

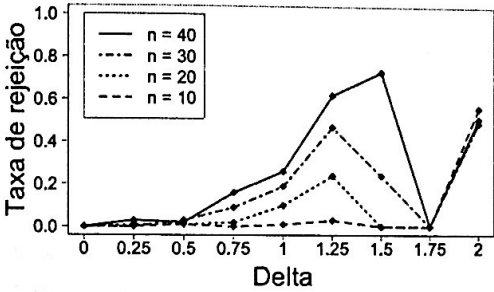


Figura B.203: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e dois clusteres.

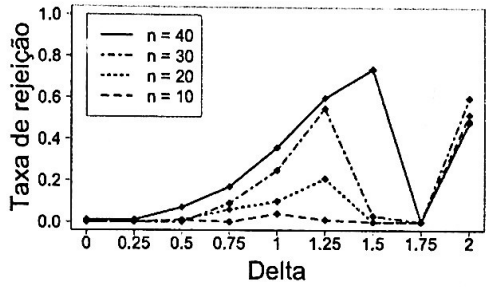


Figura B.204: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e dois clusteres.

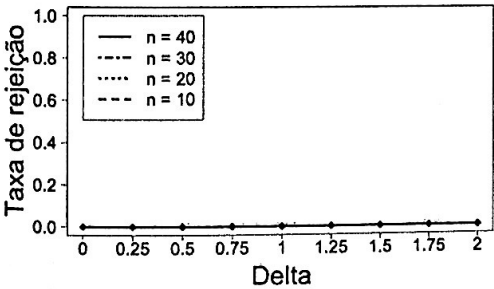


Figura B.205: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e dois clusteres.

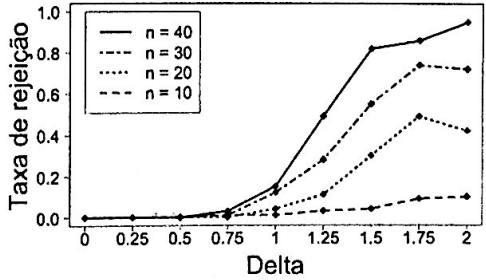


Figura B.206: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 1$ e três clusteres.

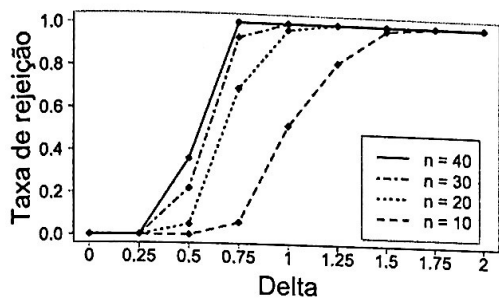


Figura B.207: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-1}$ e três clusteres.

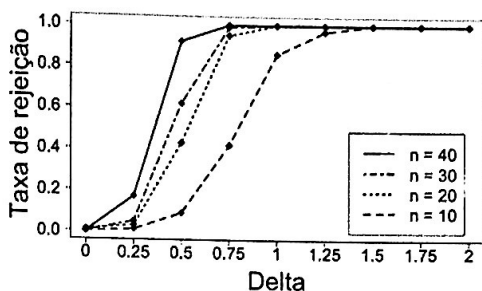


Figura B.208: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-3}$ e três clusteres.

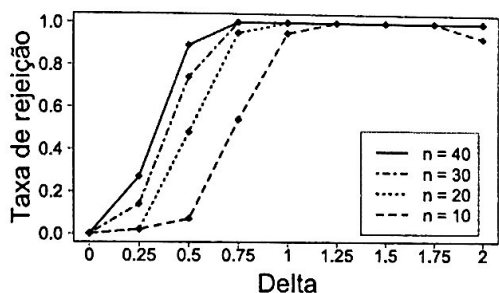


Figura B.209: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 10^{-5}$ e três clusteres.

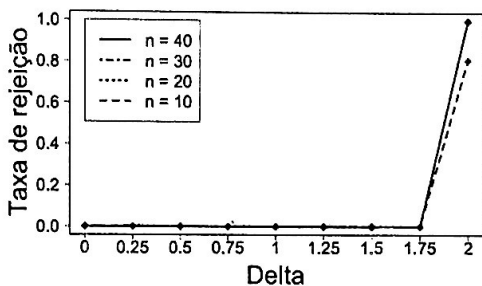


Figura B.210: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 16$ com $N_1 = 10$ e $N_2 = 5$, $\sigma^2 = 0$ e três clusteres.

Apêndice C

Gráficos

C.1 Tabela resumo dos gráficos para as simulações com distância Euclidiana padronizada pelo desvio absoluto médio

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Conjunto	10	1	1	-	2	C.1
Conjunto	10	1	10^{-1}	-	2	C.2
Conjunto	10	1	10^{-3}	-	2	C.3
Conjunto	10	1	10^{-5}	-	2	C.4
Conjunto	10	1	0	-	2	C.5
Conjunto	10	1	1	-	3	C.6
Conjunto	10	1	10^{-1}	-	3	C.7
Conjunto	10	1	10^{-3}	-	3	C.8
Conjunto	10	1	10^{-5}	-	3	C.9
Conjunto	10	1	0	-	3	C.10
Marginal Proposto	10	1	1	11	2	C.11
Marginal Proposto	10	1	10^{-1}	11	2	C.12
Marginal Proposto	10	1	10^{-3}	11	2	C.13
Marginal Proposto	10	1	10^{-5}	11	2	C.14
Marginal Proposto	10	1	0	11	2	C.15
Marginal Proposto	10	1	1	11	3	C.16
Marginal Proposto	10	1	10^{-1}	11	3	C.17
Marginal Proposto	10	1	10^{-3}	11	3	C.18
Marginal Proposto	10	1	10^{-5}	11	3	C.19
Marginal Proposto	10	1	0	11	3	C.20

Teste	N_1	N_2	σ^2	q	Clusteres	Figura
Marginal Proposto	10	1	1	1	2	C.21
Marginal Proposto	10	1	10^{-1}	1	2	C.22
Marginal Proposto	10	1	10^{-3}	1	2	C.23
Marginal Proposto	10	1	10^{-5}	1	2	C.24
Marginal Proposto	10	1	0	1	2	C.25
Marginal Proposto	10	1	1	1	3	C.26
Marginal Proposto	10	1	10^{-1}	1	3	C.27
Marginal Proposto	10	1	10^{-3}	1	3	C.28
Marginal Proposto	10	1	10^{-5}	1	3	C.29
Marginal Proposto	10	1	0	1	3	C.30
Marginal Proposto	10	1	1	12	2	C.31
Marginal Proposto	10	1	10^{-1}	12	2	C.32
Marginal Proposto	10	1	10^{-3}	12	2	C.33
Marginal Proposto	10	1	10^{-5}	12	2	C.34
Marginal Proposto	10	1	0	12	2	C.35
Marginal Proposto	10	1	1	12	3	C.36
Marginal Proposto	10	1	10^{-1}	12	3	C.37
Marginal Proposto	10	1	10^{-3}	12	3	C.38
Marginal Proposto	10	1	10^{-5}	12	3	C.39
Marginal Proposto	10	1	0	12	3	C.40

C.2 Gráficos

C.2.1 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste conjunto com distância Euclidiana padronizada pelo desvio absoluto médio

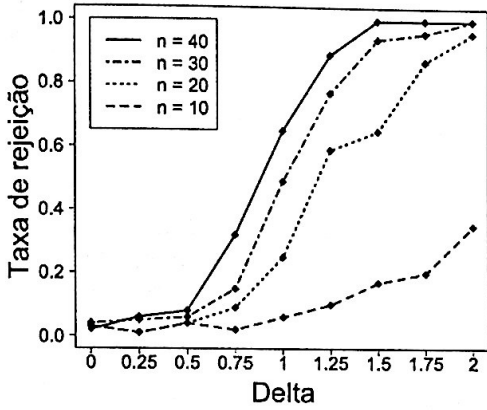


Figura C.1: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

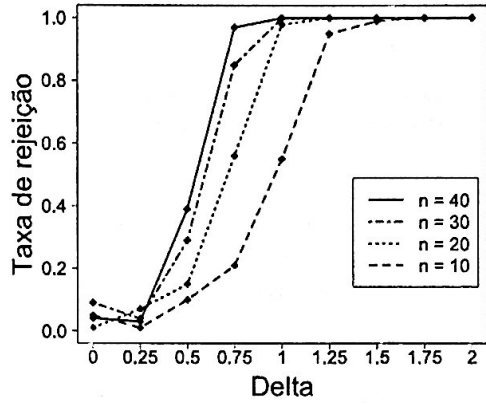


Figura C.2: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

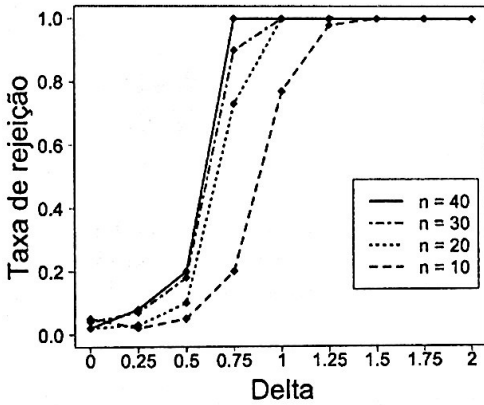


Figura C.3: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

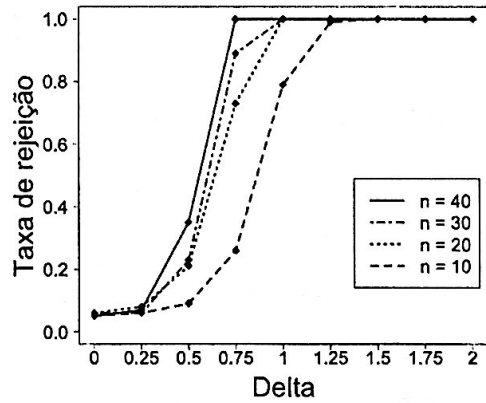


Figura C.4: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

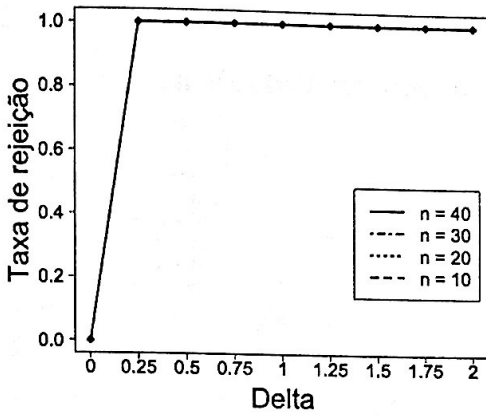


Figura C.5: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

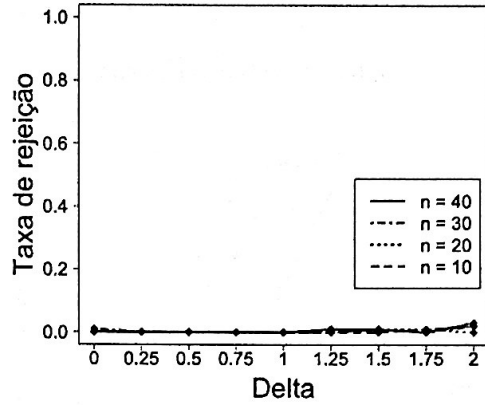


Figura C.6: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

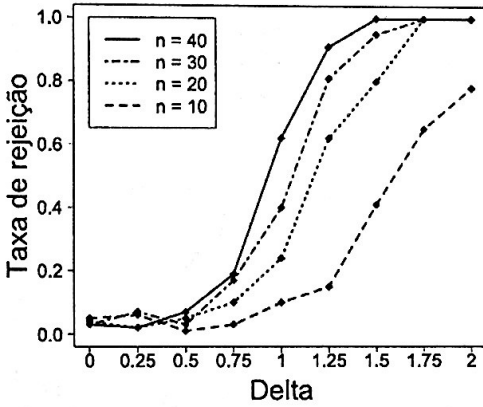


Figura C.7: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

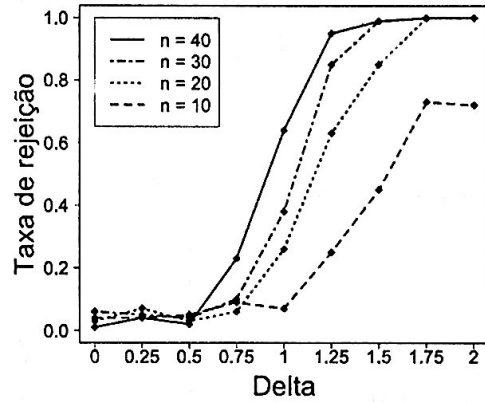


Figura C.8: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

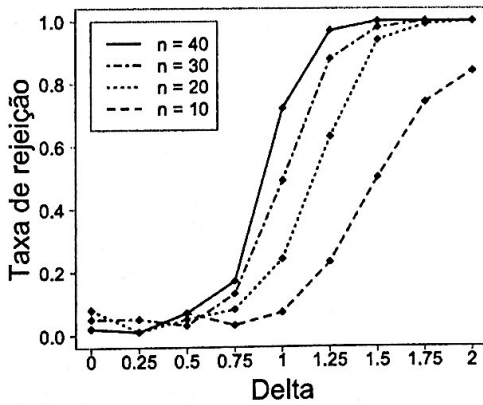


Figura C.9: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

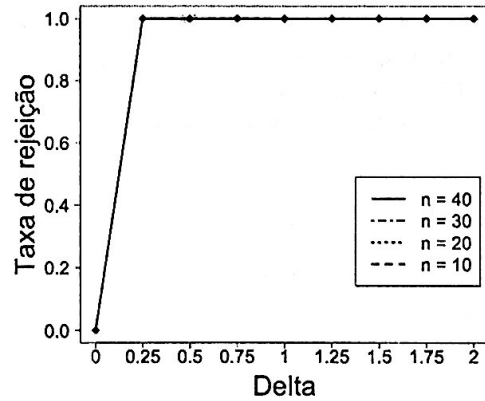


Figura C.10: Taxa de rejeição da hipótese nula no Teste conjunto com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

C.2.2 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 11$ com distância Euclidiana padronizada pelo desvio absoluto médio

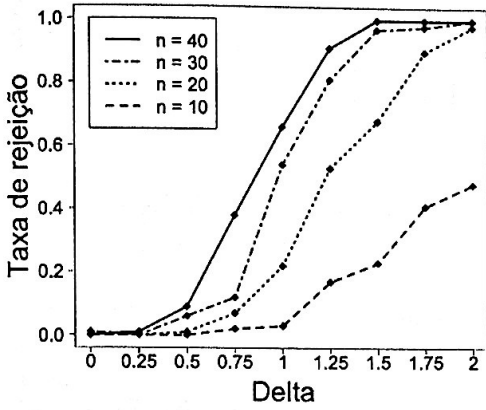


Figura C.11: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

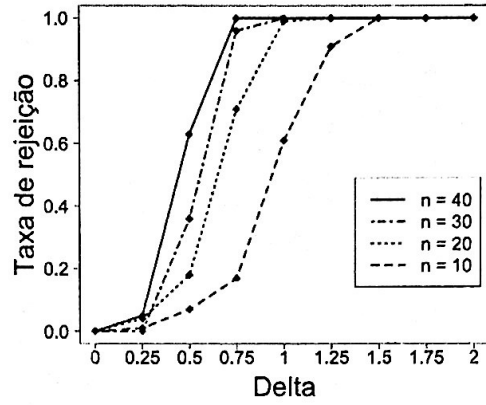


Figura C.12: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

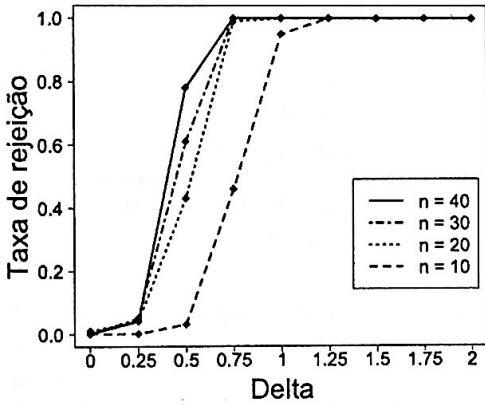


Figura C.13: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

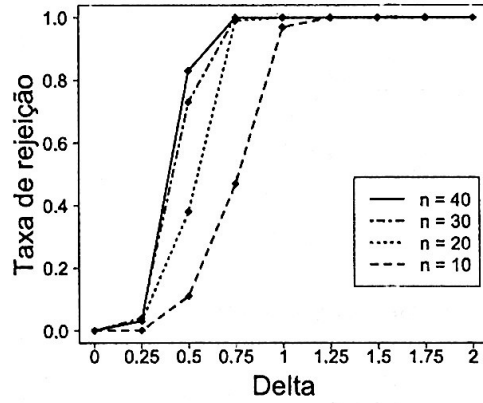


Figura C.14: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

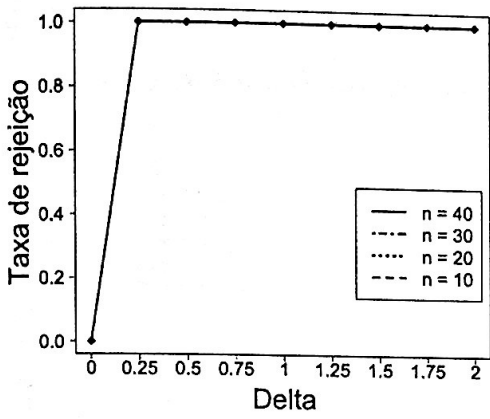


Figura C.15: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

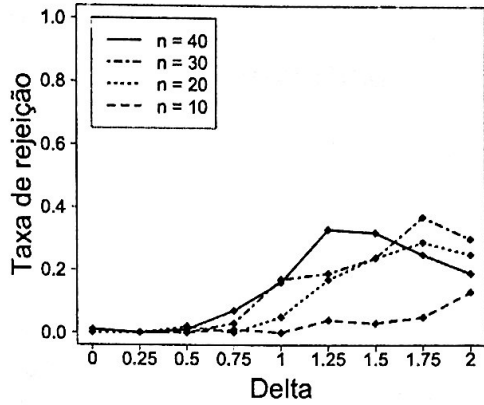


Figura C.16: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

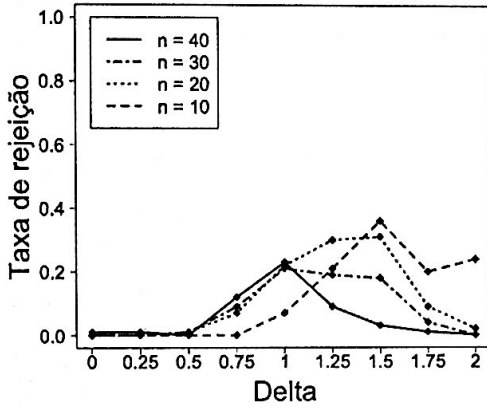


Figura C.17: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

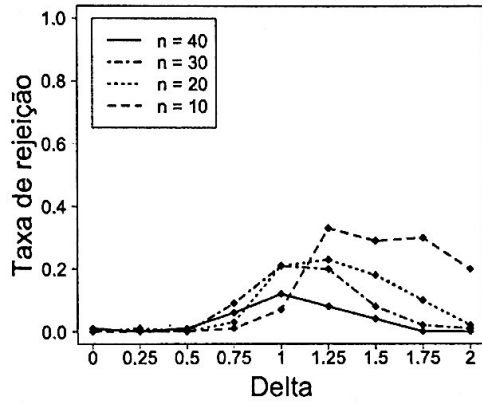


Figura C.18: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

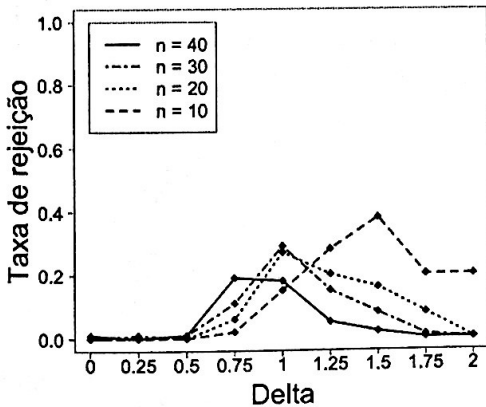


Figura C.19: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

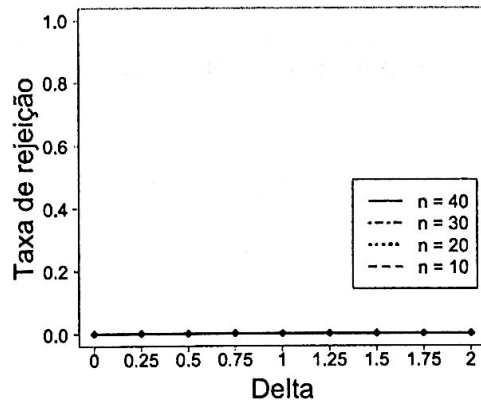


Figura C.20: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 11$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

C.2.3 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 1$ com distância Euclidiana padronizada pelo desvio absoluto médio

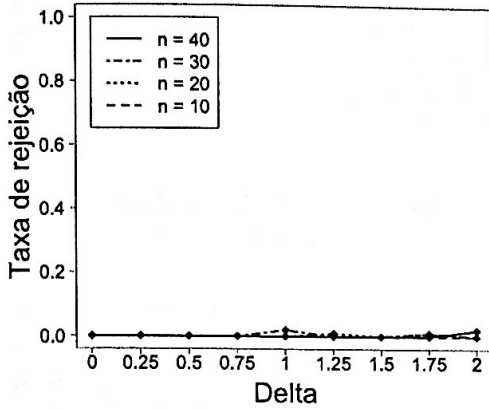


Figura C.21: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusters.

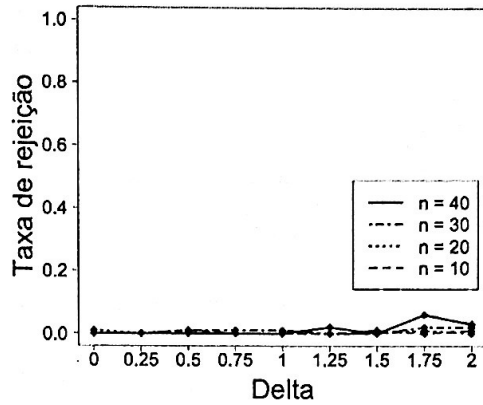


Figura C.22: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusters.

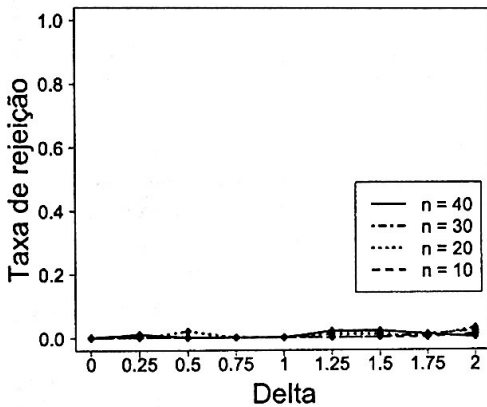


Figura C.23: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusters.

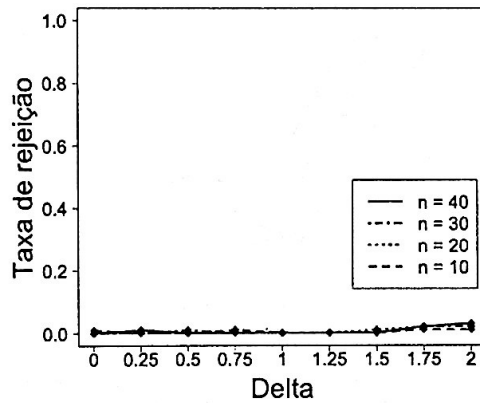


Figura C.24: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusters.

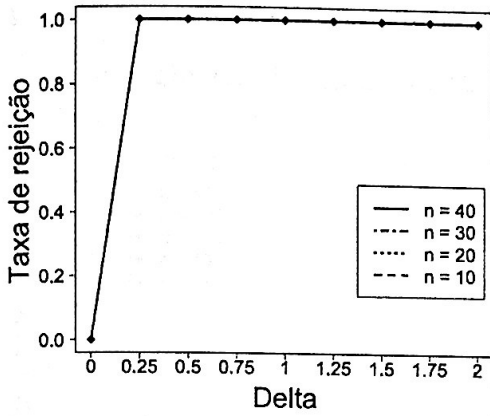


Figura C.25: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

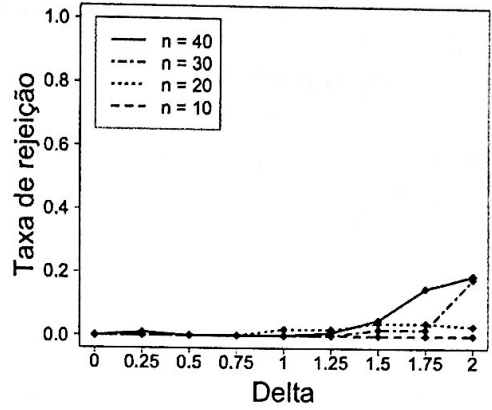


Figura C.26: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

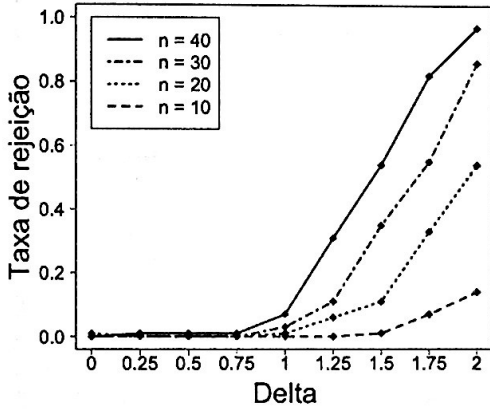


Figura C.27: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

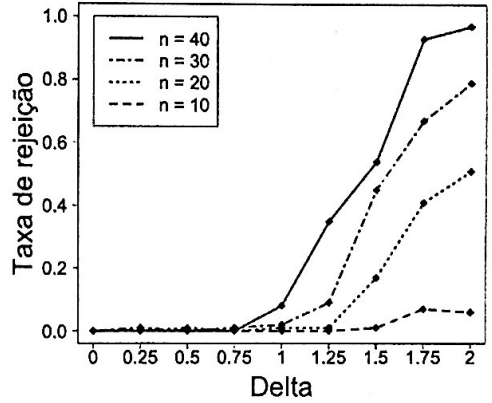


Figura C.28: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

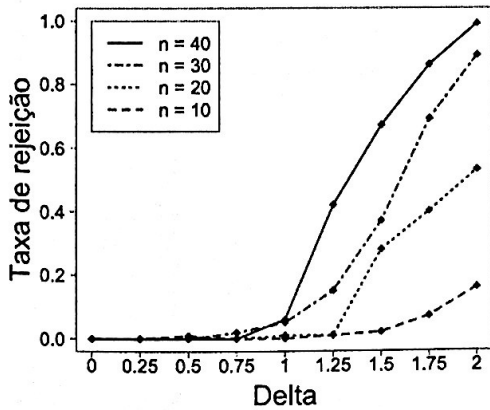


Figura C.29: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

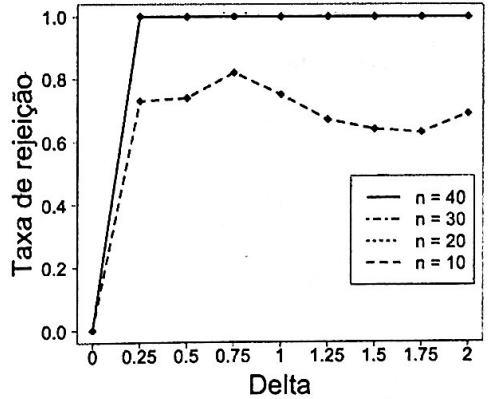


Figura C.30: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 1$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

C.2.4 Gráficos: $N_1 = 10$ e $N_2 = 1$ - Teste marginal proposto para $q = 12$ com distância Euclidiana padronizada pelo desvio absoluto médio

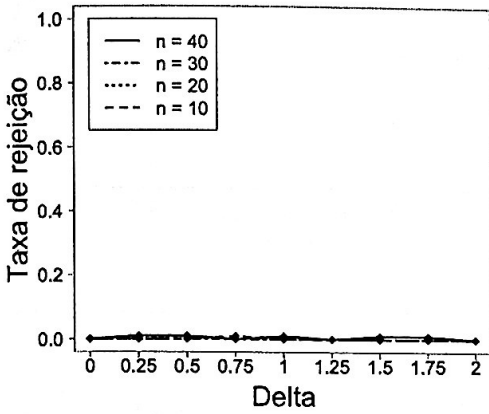


Figura C.31: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e dois clusteres.

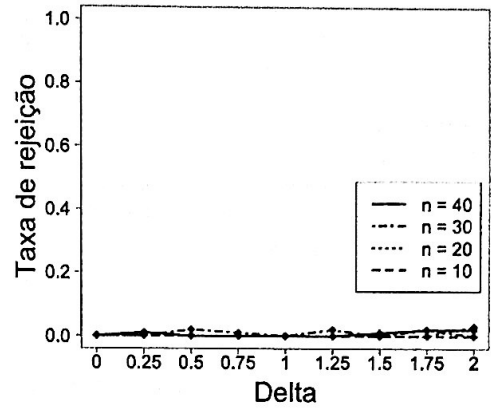


Figura C.32: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e dois clusteres.

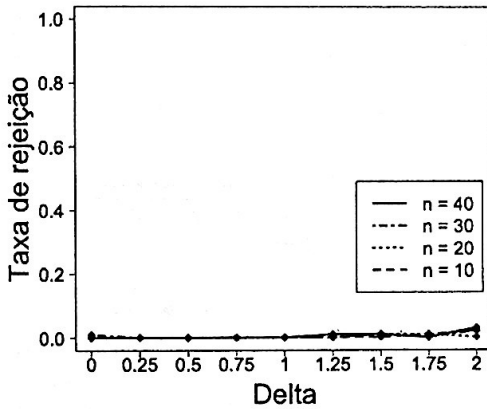


Figura C.33: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e dois clusteres.

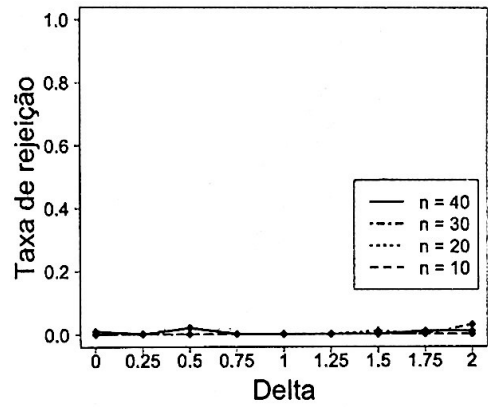


Figura C.34: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e dois clusteres.

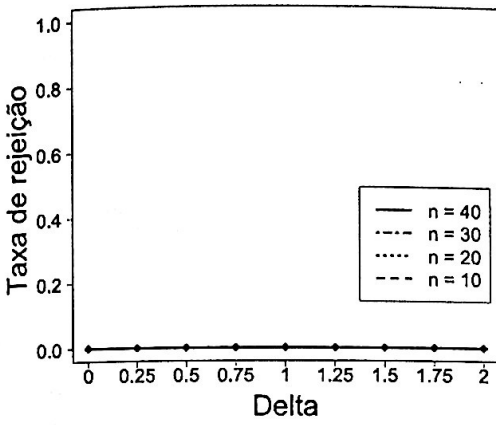


Figura C.35: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e dois clusteres.

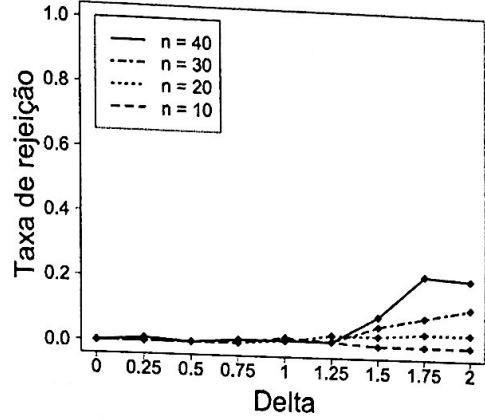


Figura C.36: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 1$ e três clusteres.

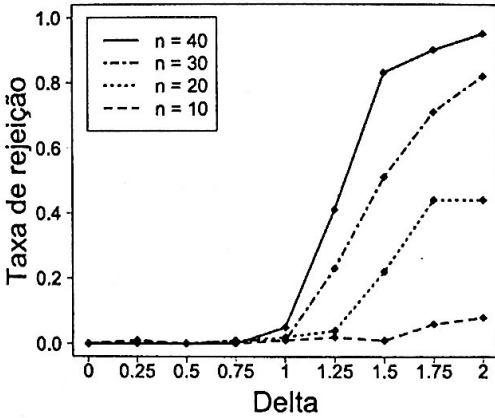


Figura C.37: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-1}$ e três clusteres.

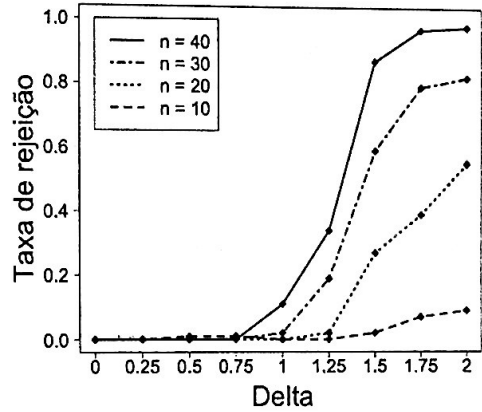


Figura C.38: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-3}$ e três clusteres.

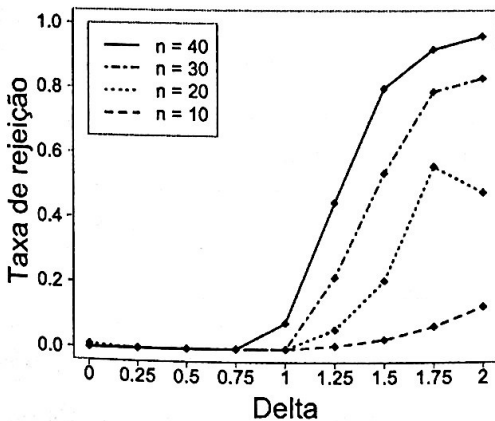


Figura C.39: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 10^{-5}$ e três clusteres.

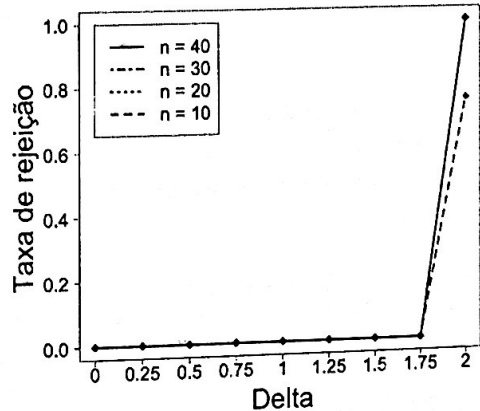
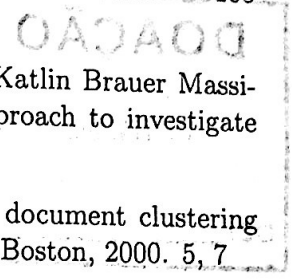


Figura C.40: Taxa de rejeição da hipótese nula no Teste marginal proposto para a característica $q = 12$ com $N_1 = 10$ e $N_2 = 1$, $\sigma^2 = 0$ e três clusteres.

Referências Bibliográficas

- Apply new analytics tools to reveal new opportunities. http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html, 2012. 1
- ANOCVA-R. Package anocva. <https://cran.r-project.org/web/packages/anocva/anocva.pdf>, 2016. 24
- Claudio Babiloni, Vittorio Pizzella, Cosimo Del Gratta, Antonio Ferretti, and Gian Luca Romani. Fundamentals of electroencefalography, magnetoencefalography, and functional magnetic resonance imaging. *International review of neurobiology*, 86:67–80, 2009. 3
- KL Behar, JA Den Hollander, ME Stromski, T Ogino, RG Shulman, OA Petroff, and JW Prichard. High-resolution 1h nuclear magnetic resonance study of cerebral hypoxia in vivo. *Proceedings of the National Academy of Sciences*, 80(16):4945–4948, 1983. 3
- Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946. 3
- Carlo E Bonferroni. *Teoria statistica delle classi e calcolo delle probabilita*. Libreria internazionale Seeber, 1936. 14, 15
- Randal Bryant, Randy H Katz, and Edward D Lazowska. Big-data computing: Creating revolutionary breakthroughs in commerce, science and society, 2008. 1
- Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974. 7
- Gilles Celeux and Gérard Govaert. A classification em algorithm for clustering and two stochastic versions. *Computational statistics & Data analysis*, 14(3):315–332, 1992. 7
- David R Cox, Emil Spjøtvoll, Søren Johansen, Willem R van Zwet, JF Bithell, Ole Barndorff-Nielsen, and M Keuls. The role of significance tests [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 49–70, 1977. 21
- Bruce Crosson, Anastasia Ford, Keith M McGregor, Marcus Meinzer, Sergey Cheshkov, Xiufeng Li, Delaina Walker-Batson, and Richard W Briggs. Functional imaging and related techniques: An introduction for rehabilitation researchers. *Journal of rehabilitation research and development*, 47(2):vii, 2010. 3
- Francisco De la Peña Olvera, Juan David Palacio Ortiz, and Eduardo Barragán Pérez. Declaración de cartagena para el trastorno por déficit de atención con hiperactividad (tdah): rompiendo el estigma. *Revista Ciencias de la Salud*, 8(1), 2010. 2
- Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994. 14
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001. 7
- André Fujita, Daniel Y Takahashi, and Alexandre G Patriota. A non-parametric method to estimate the number of clusters. *Computational Statistics & Data Analysis*, 73:27–39, 2014a. 7

- André Fujita, Daniel Y Takahashi, Alexandre G Patriota, and João R Sato. A non-parametric statistical test to compare clusters with applications in functional magnetic resonance imaging data. *Statistics in medicine*, 33(28):4949–4962, 2014b. 1, 10, 12, 13, 17, 18, 19, 23, 33
- CC Hanstock, MR Bendall, HP Hetherington, DP Boisvert, and PS Allen. Localized in vivo proton spectroscopy using depth-pulse spectral editing. *Journal of Magnetic Resonance (1969)*, 71(2):349–354, 1987. 3
- John A Hartigan. Clustering algorithms. 1975. 7
- HP Hetherington, MJ Avison, and RG Shulman. 1h homonuclear editing of rat brain using semiselective pulses. *Proceedings of the National Academy of Sciences*, 82(10):3115–3118, 1985. 3
- Doug Howe, Maria Costanzo, Petra Fey, Takashi Gojobori, Linda Hannick, Winston Hide, David P Hill, Renate Kania, Mary Schaeffer, Susan St Pierre, et al. Big data: The future of biocuration. *Nature*, 455(7209):47–50, 2008. 1
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013. 4, 5, 6
- Kenneth K Kwong, John W Belliveau, David A Chesler, Inna E Goldberg, Robert M Weisskoff, Brigitte P Poncelet, David N Kennedy, Bernice E Hoppel, Mark S Cohen, and Robert Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679, 1992. 4
- Denis Le Bihan, Peter Jezzard, James Haxby, Norihiro Sadato, Linda Rueckert, and Venkata Mattay. Functional magnetic resonance imaging of the brain. *Annals of internal medicine*, 122(4):296–303, 1995. 4
- Peter Mansfield and Peter K Grannell. Nmr'diffraction'in solids? *Journal of Physics C: solid state physics*, 6(22):L422, 1973. 3
- Seiji Ogawa, Tso-Ming Lee, Ray Stepnoski, Wei Chen, Xiao-Hong Zhu, and Kamil Ugurbil. An approach to probe some neural systems interaction by functional mri at neural time scale down to milliseconds. *Proceedings of the National Academy of Sciences*, 97(20):11026–11031, 2000. 3, 4
- Edward M Purcell, HC Torrey, and Robert V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37, 1946. 3
- Patrick L Purdon and Robert M Weisskoff. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fmri. *Human brain mapping*, 6(4):239–249, 1998. 4
- ME Raichle. Handbook of physiology. the nervous system. higher functions of the brain. 1987. 4
- DL Rothman, KL Behar, HP Hetherington, and RG Shulman. Homonuclear 1h double-resonance difference spectroscopy of the rat brain in vivo. *Proceedings of the National Academy of Sciences*, 81(20):6330–6334, 1984. 3
- Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987. 7
- Charles Smart Roy and Charles S Sherrington. On the regulation of the blood-supply of the brain. *The Journal of physiology*, 11(1-2):85, 1890. 4
- Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010. 1



João Ricardo Sato, Daniel Yasumasa Takahashi, Marcelo Queiroz Hoexter, Katlin Brauer Massirer, and André Fujita. Measuring network's entropy in adhd: A new approach to investigate neuropsychiatric disorders. *NeuroImage*, 77:44–51, 2013. 2

Michael Steinbach, George Karypis, Vipin Kumar, et al. A comparison of document clustering techniques. In *KDD workshop on text mining*, volume 400, pages 525–526. Boston, 2000. 5, 7

T Yamazaki, S Takahashi, K Ishii, K Matsumoto, T Ishibashi, K Sakamoto, and A Ogawa. Meningioma in the pineal region: preoperative diagnosis with ct, mri, and angiography. *Radiation medicine*, 9(1):22–25, 1990. 4

Xiao-Hong Zhu, Seong-Gi Kim, Peter Andersen, Seiji Ogawa, Kamil Uğurbil, and Wei Chen. Simultaneous oxygenation and perfbsion imaging study of functional activity in primary visual cortex at different visual stimulation frequency: Quantitative correlation between bold and cbf changes. *Magnetic resonance in medicine*, 40(5):703–711, 1998. 4

