

Análise de variância multivariada aplicada a dados com medidas repetidas

Mara Rubia Machado Couto^I, Luciane Flores Jacobi^{II},
Guilherme da Silva Machado^{III}

Abstract

Em algumas situações, no processo de análise de dados, existe o interesse de analisar a performance de diversas variáveis conjuntamente e determinar a influência ou a importância de cada variável na presença das demais. Para isto, utilizam-se as técnicas de análise multivariada. Quando, em particular, este interesse está em identificar a existência de efeitos entre tratamentos e/ou condições de avaliação e verificar a existência de interação entre estes fatores; é apropriado o emprego de uma análise de variância multivariada, ou ainda, uma análise de perfis de médias que são métodos multivariados aplicados na análise de dados com medidas repetidas e dados longitudinais. O presente trabalho tem por objetivo desenvolver um referencial básico sobre o uso e aplicações da análise multivariada, e proporcionar instruções sobre os procedimentos operacionais necessários para a execução destas técnicas no software Statistica 9.0.

Palavras-chave: Análise multivariada, Dados longitudinais, Manova, Análise de perfis

Resumo

In some situations, during the process of data analysis, there is the interest both in examining the performance of several variables acting jointly and in determining the influence or relevance of each variable in the presence of the other variables. With this in mind, multivariate analysis techniques were applied to the data. When, in particular, the interest consists of identifying the existence of effects among treatments and/or conditions of evaluation, as well as verifying the occurrence of interaction among these features, it is considered appropriate either a multivariate analysis of variance or an average profile analysis. These are multivariate methods applied to the data analysis with repeated measurement and longitudinal data. The present work aims at developing a basic theoretical framework about the use and functions of multivariate analysis and, in addition, providing instructions about the operational procedures necessary to perform these techniques using the Statistica 9.0 data analysis software system.

Keywords: Multivariate analysis, Longitudinal data, Manova, Profile analysis

^I Instituto Federal Farroupilha. Brasil; mara.couto@iffarroupilha.edu.br

^{II} Universidade Federal de Santa Maria. Brasil. lucianefj8@gmail.com

^{III} Universidade Federal de Santa Maria. Brasil. guilhermesv2015@gmail.com



1 INTRODUCTION

A inerente necessidade de inferir relações entre múltiplas variáveis para analisar e explicar fenômenos sociais e/ou físicos faz com que cada vez mais as técnicas estatísticas estejam presentes no processo de investigação científica. Em muitas situações existe o interesse de se investigar o comportamento de uma ou mais variáveis respostas ao longo do tempo, ao longo de uma condição de avaliação ou ainda de outra escala ordenada qualquer.

Devido à complexidade do estudo dessas várias variáveis que contêm dados simultâneos para explicação de tais fenômenos, a utilização de uma metodologia que estude conjuntamente as variáveis pode, em determinadas situações, ser mais adequada do que isolar cada variável e estudá-la separadamente (FÁVERO et al, 2009). Para este tipo de estudo a literatura técnico-científica utiliza os métodos de análise multivariada, que de acordo com Vicini et al (2018) surgiu por volta de 1901, pois a mesma permite que se analise a performance conjunta das variáveis e se determine a influência ou importância de cada variável na presença das restantes, que vem sendo cada vez mais utilizada devido ao fácil acesso a pacotes computacionais (MANLY, 2008).

Estudos com medidas repetidas têm a vantagem de requerer um número menor de unidades amostrais. Além disso, diminuem a variabilidade decorrente de diferenças individuais e permitem avaliar mudanças que ocorrem dentro e entre as unidades amostrais com mais eficiência, e estudos longitudinais que são um tipo de medidas repetidas são de particular interesse, quando o objetivo é avaliar variações globais ou individuais ao longo das condições de avaliação.

A análise de variância multivariada (MANOVA), conforme Fávero et al (2009) é uma extensão da análise de variância simples (ANOVA), com incorporação de mais de uma variável independente. A principal diferença entre as duas, de acordo com Hair et al (2009) está em que a MANOVA é utilizada para comparar grupos em múltiplas variáveis dependentes, enquanto que a ANOVA avalia diferenças de grupos em uma única variável dependente.

Do mesmo modo que a ANOVA a MANOVA requer cuidados na verificação de algumas pressuposições que são generalizações da ANOVA, tais como a verificação do pressuposto de normalidade multivariada e verificação da igualdade de matrizes de variância e covariância.

A análise de perfis de médias é uma derivação da técnica de análise de variância multivariada (MANOVA) utilizada quando se deseja fazer comparações entre os diversos perfis médios de respostas; facilitando a identificação da existência ou não de interação entre tratamentos, entre grupos e a existência ou não do efeito longitudinal. Para se obter estatísticas que testem a hipótese linear geral, alguns critérios de testes estão disponíveis no contexto multivariado e abordados neste estudo.

Portanto esta pesquisa tem por objetivo desenvolver um referencial básico de suporte teórico-técnico para estudos e pesquisas que exijam o uso das técnicas multivariadas de análise de variância multivariada e análise de perfis de médias.

2 ANÁLISE DE VARIÂNCIA MULTIVARIADA

Um dos métodos para analisar dados com medidas repetidas é a análise de variância multivariada. Esta análise é uma metodologia utilizada para comparar médias de grupos, quando algumas medidas ao longo dos grupos são contínuas e certos pressupostos são cumpridos, podendo-se estender esta metodologia para mais de uma variável dependente.

Da mesma forma que a análise de variância univariada, a análise de variância multivariada também tem o propósito de fazer uma análise de variação dos dados, entre ou dentro dos grupos e, assim, obter conclusões sobre possíveis diferenças nas médias dos grupos. A análise de variância multivariada é indicada quando existe uma estreita correlação entre as variáveis.

No caso univariado, uma “única medida” dependente é testada quanto à igualdade das médias através dos grupos, enquanto na multivariada a hipótese da nulidade refere-se aos vetores das médias. Na análise de variância univariada (ANOVA) a hipótese nula a ser testada é a igualdade das médias através dos grupos.

Na análise de variância multivariada (MANOVA), a hipótese nula a ser testada é a igualdade dos vetores das médias nas múltiplas variáveis dependentes. Uma vantagem da MANOVA sobre a realização de sucessivas ANOVAS, para as diferentes variáveis, é a consideração das covariâncias entre as variáveis, sobretudo se estas apresentarem correlações não-nulas.

Assim como, a análise de variância univariada, a multivariada também segue algumas pressuposições, que Vieira et al (2018), que são generalizações dos pressupostos da análise de variância simples (ANOVA):

- As observações devem se constituir de amostras aleatórias provenientes de populações normais, isto é, devem seguir uma distribuição normal multivariada;
- Essas amostras deverão pertencer a grupos populacionais com idêntica variância, igualdade de matrizes de variância e covariância.

Quando se realiza a ANOVA para medidas repetidas é necessário observar, conforme Field (2009) a suposição de esfericidade, que é referida como à igualdade das variâncias das diferenças entre níveis de tratamento. Porém nessa pesquisa a mesma não será avaliada, pois de acordo com O'Brien e Kaiser (1985) os testes baseados na abordagem MANOVA estão livres de suposições de esfericidade.

2.1. Medidas Repetidas e Dados Longitudinais

Um tipo comum de medida repetida são os dados longitudinais, onde as observações são ordenadas pelo tempo ou pela posição no espaço. Mais especificamente, dados longitudinais são caracterizados pela observação repetida de uma ou mais variáveis respostas em uma mesma unidade experimental. Cada conjunto de unidades de observação pode ser entendido como um perfil individual de respostas, pois contêm os valores da(s) variável(is) resposta(s) em cada uma das ocasiões de observação.

Em situações de medidas repetidas, em que se tem uma estrutura balanceada e completa, a estrutura é a apresentada na Tabela 1.

Tabela 1 – Estrutura de dados básica para estudos longitudinais

Subpopulação (tratamentos)	Unidade amostral	Condições de avaliação			
		1	2	...	K
1	1	Y_{111}	Y_{112}	...	Y_{11k}
1	2	Y_{121}	Y_{122}	...	Y_{12k}
⋮	⋮	⋮	⋮		⋮
1	n_1	$Y_{1n_1,1}$	$Y_{1n_1,2}$...	$Y_{1n_1,k}$
2	1	Y_{211}	Y_{212}	...	Y_{21k}
2	2	Y_{221}	Y_{222}	...	Y_{22k}
⋮	⋮	⋮	⋮		⋮
2	n_2	$Y_{2n_2,1}$	$Y_{2n_2,2}$...	$Y_{2n_2,k}$
⋮	⋮	⋮	⋮		⋮
l	1	Y_{l11}	Y_{l12}	...	Y_{l1k}
l	2	Y_{l21}	Y_{l22}	...	Y_{l2k}
⋮	⋮	⋮	⋮		⋮
l	n_i	$Y_{ln_i,1}$	$Y_{ln_i,2}$...	$Y_{ln_i,k}$

Onde y_{ijk} é o valor da variável resposta dentro da j -ésima unidade experimental do i -ésimo tratamento, sob k -ésima condição de avaliação (tempo, por exemplo) para $l = 1, 2, \dots, i; J = 1, 2, \dots, n_i; K = 1, 2, \dots, k$.

As técnicas clássicas de dados longitudinais são geralmente dirigidas para o caso de dados completos e balanceados em relação ao tempo. Lima (1996) diz que os dados provenientes de estudos longitudinais são chamados de regulares em relação ao tempo se o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do estudo e de balanceados com relação ao tempo, se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais. Se não houver observações perdidas, diz-se que a estrutura dos dados é completa.

2.2. Análise Multivariada de Perfis

A análise multivariada de perfis é uma das técnicas estatísticas utilizadas para analisar observações provenientes de experimentos com medidas repetidas. Esta técnica fundamenta-se no número de unidades experimentais como tamanho da amostra, enquanto o procedimento univariado de análise de perfis utiliza o número total de observações. Segundo Singer e Andrade (1986), duas desvantagens estão associadas a abordagem multivariada: a necessidade de observações completas e o

pequeno poder atribuído aos testes. Desta forma, o procedimento univariado é mais preciso que o multivariado. Por outro lado, a solução univariada exata é mais exigente no que se relaciona com a estrutura de covariância entre as observações dentro de cada unidade experimental. Quando o padrão de covariância exigido não estiver satisfeito, nem ao menos de forma aproximada, pode-se utilizar as técnicas de análise multivariada, pois este tipo de solução é aplicável para uma matriz de variâncias e covariâncias (Σ) qualquer. A única exigência do procedimento multivariado é que Σ deve ser comum a todos os tratamentos. Os procedimentos univariado e multivariado exigem a suposição de que as componentes aleatórias do erro devem seguir uma distribuição normal.

Danford, Hughes e McNee (1960), afirmam que os dois procedimentos, univariado e multivariado, são assintoticamente equivalentes para tamanhos de amostras utilizados na prática. Isso ocorre devido as estimativas na análise univariada convencional serem mais poderosas que na análise multivariada, desde que estejam satisfeitas as suposições sobre a matriz de variância e covariância, pois as estimativas da análise univariada baseiam-se em um número maior de graus de liberdade do que a análise multivariada.

É importante observar que a técnica de análise multivariada de perfis só é aplicável, de acordo com Riboldi et al (2003), em situações onde o número total de observações menos o número de tratamentos é maior ou igual ao número de dados longitudinais menos um. Esta condição se faz necessária para que a matriz das somas de quadrados e produtos cruzados de erro, associada aos efeitos dentro de sujeitos, seja não singular com probabilidade 1.

Os objetivos da análise de perfis de médias podem ser explicados pelas hipóteses:

- $H_0^{(1)}$: os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos, não existe interação Tratamentos x Condições de Avaliação;
- $H_0^{(2)}$: os perfis médios de resposta correspondentes aos diversos tratamentos são coincidentes, não existe efeito de tratamentos;
- $H_0^{(3)}$: os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos ao eixo das abscissas, não existe efeito das condições de avaliação.

Segundo Morrison (1990), a análise de perfis pode ser trabalhada com três hipóteses associadas com o grupo de medidas repetidas onde testa uma análise de variância simples de modelo misto. Como no caso de um único grupo, essa aproximação é estritamente válida só nas condições em que os resíduos dentro de uma unidade experimental tenham a mesma variância e uma correlação comum para todos os pares. Considerando-se que os resíduos devem ser multinormalmente distribuídos para se realizar o teste esta exigência inclui, como um caso especial, o modelo convencional com resíduos independentes.

Por essas razões Greenhouse e Geisser (1959) investigaram a análise de perfil de variância debaixo da suposição de distribuição multinormal dos resíduos com uma matriz de covariância geral. Os autores propuseram uma análise aproximada de variância (Tabela 2) e testes conservadores para as diferenças de resposta e a hipótese de paralelismo cujo verdadeiro nível α não pode exceder alguns valores especificados.

Tabela 2 - Análise de variância para medidas repetidas

Causas de Variação	Soma de quadrados	Graus de liberdade	Quadrado Médio	F	Graus de liberdade conservados
Respostas	SQ_1	$K - 1$	$SQ_1 / K - 1$	$(n - I) \frac{SQ_1}{SQ_5}$	$1, n - 1$
Tratamentos	SQ_2	$I - 1$	$SQ_2 / I - 1$	$\frac{(n - I)SQ_2}{(I - 1)SQ_3}$	
Condições de avaliação	SQ_3	$n - I$	$SQ_3 / n - I$		
Respostas x Tratamentos	SQ_4	$(K - 1)(I - 1)$	$SQ_4 / (K - 1)(I - 1)$	$\frac{(n - I)SQ_4}{(I - 1)SQ_5}$	$I - 1, n - I$
Condições de avaliação x respostas (sem tratamento)	SQ_5	$(K - 1)(n - I)$	$SQ_5 / (K - 1)(n - I)$		
Total	SQ_6	$nK - 1$			

Quadro da MANOVA adaptado de Morrison (1990)

Os testes são baseados em certa aproximação da distribuição das somas de quadrados para um modelo multinormal arbitrário. As seis somas de quadrados são calculadas utilizando-se as seguintes fórmulas:

$$\begin{aligned} SQ_1 &= \frac{1}{n} \sum_{k=1}^k G_k^2 - \frac{G^2}{nk} & SQ_2 &= \frac{1}{K} \sum_{i=1}^i \frac{1}{n_i} C_i^2 - \frac{G^2}{nk} \\ SQ_3 &= \frac{1}{K} \sum_{i=1}^i \sum_{j=1}^{n_i} R_{ij}^2 - \frac{1}{K} \sum_{i=1}^i \frac{1}{n_i} C_i^2 & SQ_4 &= \sum_{i=1}^i \sum_{k=1}^k \frac{1}{n_i} T_{ik}^2 - \frac{G^2}{nk} - S_1 - S_2 \\ SQ_5 &= S_6 - S_1 - S_2 - S_3 - S_4 & SQ_6 &= \sum_{k=1}^k \sum_{i=1}^i \sum_{j=1}^{n_i} x_{ijk}^2 - \frac{G^2}{nk} \end{aligned}$$

Greenhouse e Geisser (1959) mostraram que quando a hipótese de mesmo efeito de resposta é verdadeira, o teste estatístico $F = (n - I) \frac{S_1}{S_5}$ tem aproximadamente distribuição F com graus de liberdade $(K - 1)\varepsilon$ e $(K - 1)(n - I)\varepsilon$, e onde a hipótese de paralelismo for verdadeira $\frac{(n - I)S_4}{(I - 1)S_5}$ tem distribuição F com graus de liberdade $(K - 1)\varepsilon$ e $(K - 1)(n - I)\varepsilon$.

Onde $\varepsilon = \frac{K^2(\bar{\sigma}_d - \bar{\sigma}_r)^2}{(K - 1)(\sum \sum \sigma_{rs}^2 - 2K \sum \bar{\sigma}_r^2 + K^2 \bar{\sigma}_r^2)}$, sendo calculado para os elementos σ_{rs} da matriz de variância e covariância populacional onde $\bar{\sigma}_d$ é a média dos k elementos da diagonal principal, $\bar{\sigma}_r$ é a média de todas as variâncias e covariâncias, e $\bar{\sigma}_r$ é a média dos elementos das r linhas.

O efeito de permitir a estrutura de covariância geral é o de reduzir os graus de liberdade para o primeiro e o terceiro teste pelo fator ε . Entretanto, os elementos da Σ necessária para o cálculo de ε não são conhecidos na prática, e a estimação de ε pela matriz de variância e covariância amostral pode introduzir outra incerteza na aproximação da análise de variância. Greenhouse e Geisser (1959) mostraram que ε deve satisfazer $\varepsilon > \frac{1}{k - 1}$ para qualquer matriz de variância e covariância e, então, os graus de liberdade para os testes podem ser menores que 1 e $n - I$, $I - 1$ e $n - I$, respectivamente. Devido a isso estes testes são chamados de testes conservadores, porque eles são baseados na máxima redução dos graus de liberdade.

2.3. Alguns Critérios de Teste para a Análise Multivariada de Perfis

Dentre as estatísticas de testes disponíveis no contexto multivariado, as mais difundidas são as obtidas pelos princípios da União-intersecção de Roy e da Razão de verossimilhança de Wilks. Ainda outros dois testes importantes, devido a sua aplicação, são o Critério de Hotelling - Lawley e o Critério de Pillai (Fagundes, 2012)

Segundo Ferreira (1996) muitos autores recomendam utilizar o critério de Wilks como referência, por se tratar de um teste baseado na razão de verossimilhança. Outros autores, como Singer e Andrade (1986); Freitas (2007); Cruz e Topa (2009), recomendam que a hipótese nula deva ser rejeitada se pelo menos três dos quatro critérios forem significativos em um nível nominal de significância previamente adotado.

3 APLICAÇÃO DOS MÉTODOS COM SAÍDAS DO SOFTWARE STATISTICA 9.0

Para o desenvolvimento das técnicas no Software serão utilizados os dados de um exemplo hipotético, retirado de Morrison (1990), que podem ser observados na Tabela 3. A seguir serão descritos os procedimentos que devem ser seguidos para às análises desses dados.

Exemplo: Supõe-se que existam três escalas A, B e C de um inventário que avalia determinadas atitudes maternas que foram administradas às mães participantes de um estudo do desenvolvimento da criança. Como uma parte da investigação cada mãe é atribuída a uma de quatro classes sócio-econômicas.

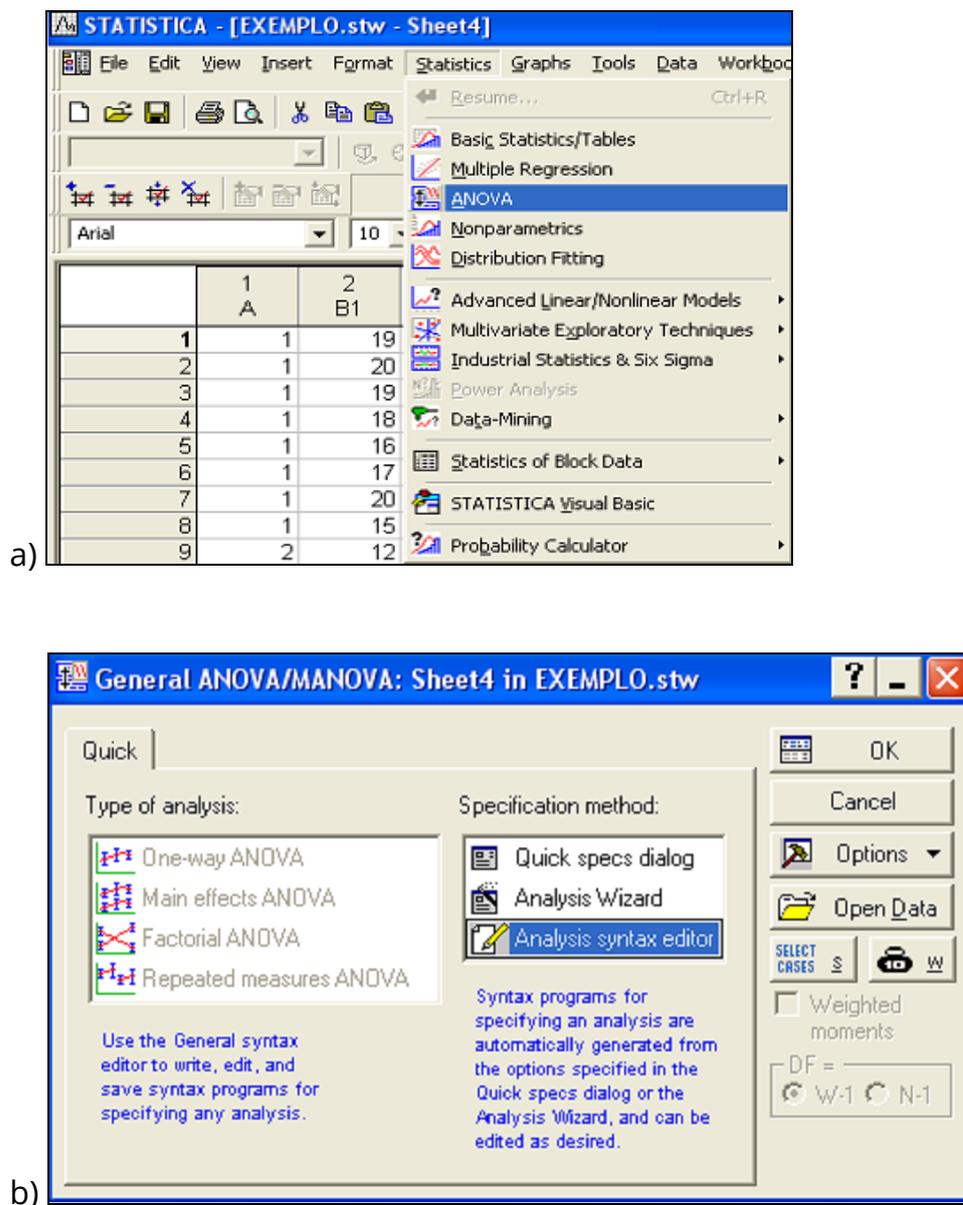
Tabela 3 – Observações dos perfis.

Classe	Escala			Subtotal
	A	B	C	
1	19	20	18	57
1	20	21	19	60
1	19	22	22	63
1	18	19	21	58
1	16	18	20	54
1	17	22	19	58
1	20	19	20	59
1	15	19	19	53
T_{1k}	144	160	158	462
2	12	14	12	38
2	15	15	17	47
2	15	17	15	47
2	13	14	14	41
2	14	16	13	43
T_{2k}	69	76	71	216
3	15	14	17	46
3	13	14	15	42
3	12	15	15	42
3	12	13	13	38
T_{3k}	52	56	60	168
4	8	9	10	27
4	10	10	12	32
4	11	10	10	31
4	11	7	12	30
T_{4k}	40	36	44	120
G	305	328	333	966

A partir dos dados da Tabela 3, determinou-se primeiramente o vetor de médias geral, isto é, a média de cada escala considerando-se todas as classes sociais. Utilizando-se o programa *Statistica 9.0* para realizar o exemplo anterior, os seguintes passos devem ser efetuados:

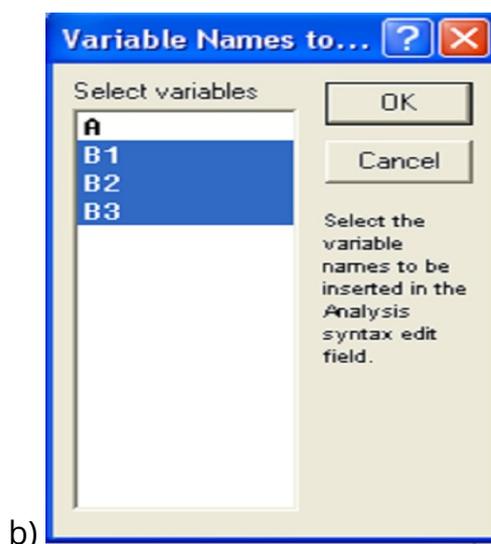
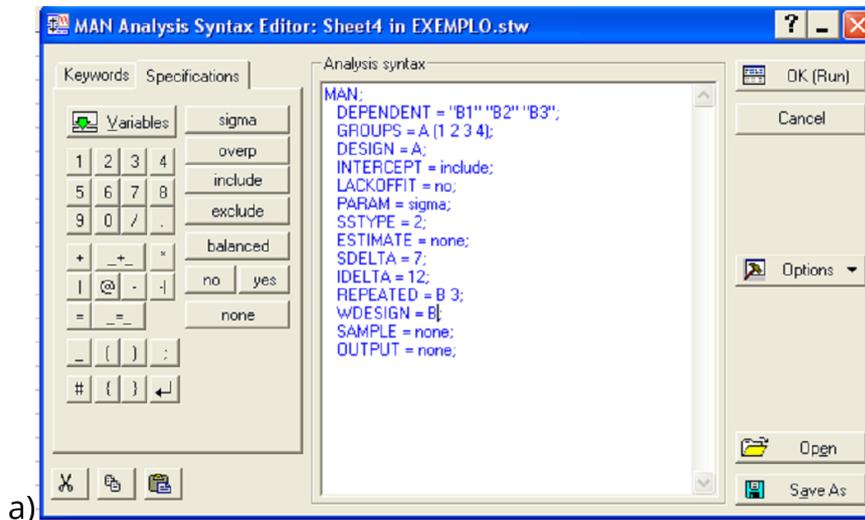
Ao iniciar o programa, escolhe-se na barra de ferramentas em *Statistics* a opção ANOVA conforme mostra a Figura 1 (a). Ao clicar-se em ANOVA abrirá a janela *General ANOVA/MANOVA*. Escolhe-se em *Type of analysis* a opção *Repeated measures ANOVA* e em *Specification method* a *Analysis syntax editor* conforme Figura 1(b).

Figura 1 – figura 1(a) Caixa de seleção da análise de variância, figura 1 (b) Caixa de seleção para especificação do método



A seguir, clica-se no botão OK e uma janela como mostrado abaixo (Figura 2(a)) aparecerá. Então seleciona-se na Figura 2 (a), *Specifications*, e nela seleciona-se *variable* onde aparecerá a caixa para a seleção das variáveis dependentes (no caso são as classes sociais B1, B2 e B3 representando as classes A, B e C, respectivamente) como mostrado na Figura 2 (b).

Figura 2 – figura 2 (a) Caixa de seleção para especificação dos parâmetros na análise de variância multivariada, figura 2 (b) Caixa de seleção das variáveis



Observa-se que para a realização da análise multivariada de perfis de médias deve-se alterar o editor de análises conforme a Figura 3 (a) pois o *default* do software é pré-programado para realizar as análises das somas de quadrados para valores de amostras de mesmo tamanho. Por isso, deve-se programar as somas de quadrados para o tipo 2, que são para amostras de tamanhos diferentes.

Após selecionar as variáveis, clica-se em *OK*, e preenche-se o quadro MAN (Figura 3(a)) as escalas (*GROUPS* = A 1 2 3 4), o *DESIGN* (escalas = A), as somas de quadrados (*SSTYPE* = 2) as medidas repetidas (*REPEATED* = B 3) e o (*WDESIGN* = B); conforme Figura 3 (a). Clica-se em *OK (Run)* e aparecerá a Figura 3 (b). Ao se clicar em

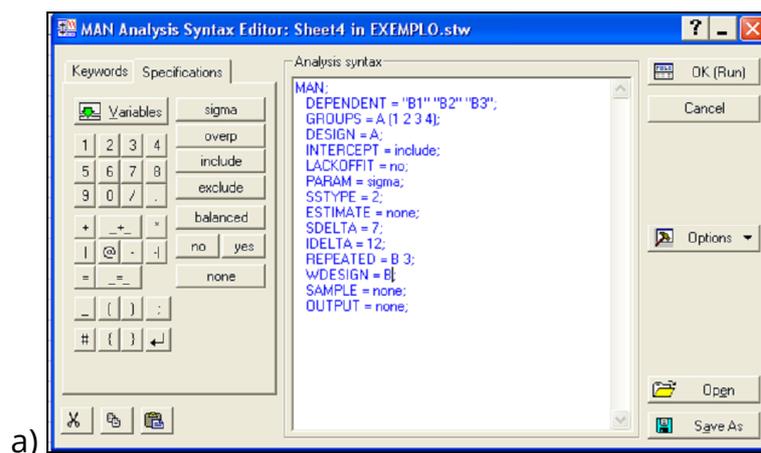
All effects na Figura 3 (b) irão aparecer os resultados da análise de variância multivariada, mostrados na Figura 3 (c). Continuando a análise dos dados, foi realizada uma análise multivariada de perfis de médias.

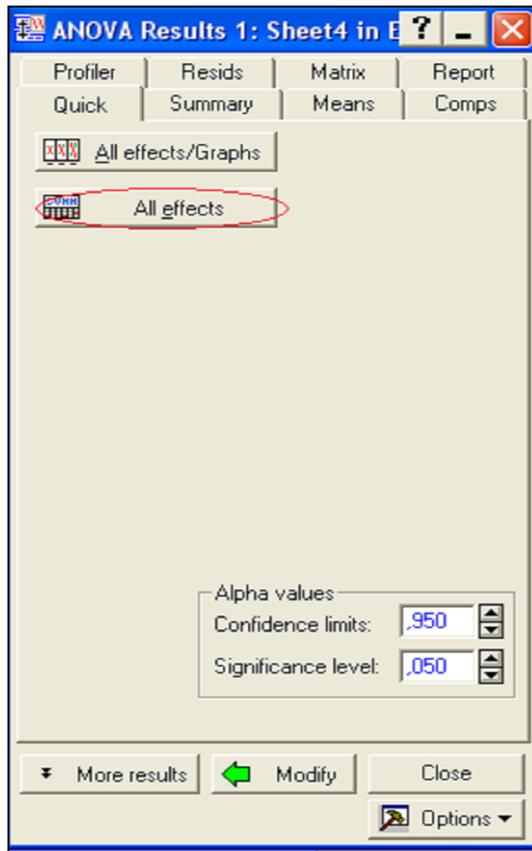
As hipóteses a serem testadas, são representadas das seguintes formas:

- H (1): os perfis médios de resposta correspondentes as diversas escalas são paralelos, não existe interação escalas x classe social;
- H (2): os perfis médios de resposta correspondentes as diversas escalas são coincidentes, não existe efeito das escalas;
- H (3): os perfis médios de resposta correspondentes as diversas escalas são paralelos ao eixo das abscissas, não existe efeito das classes sociais.

A seguinte análise de perfis, segue o modelo da figura 3 (c).

Figura 3 – figura 3 (a) Caixa de seleção para análise de variância multivariada, figura 3 (b) Caixa de seleção para os efeitos da análise de variância multivariada, figura 3(c) Caixa com os resultados da análise





b)

Repeated Measures Analysis of Variance (Sheet4 in EXEMPLO.st)					
Sigma-restricted parameterization					
Type II decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	14812,00	1	14812,00	4236,747	0,0000
A	743,90	3	247,97	70,927	0,0000
Error	59,43	17	3,50		
B	21,24	2	10,62	6,881	0,0031
B*A	18,96	6	3,16	2,048	0,0860
Error	52,47	34	1,54		

c)

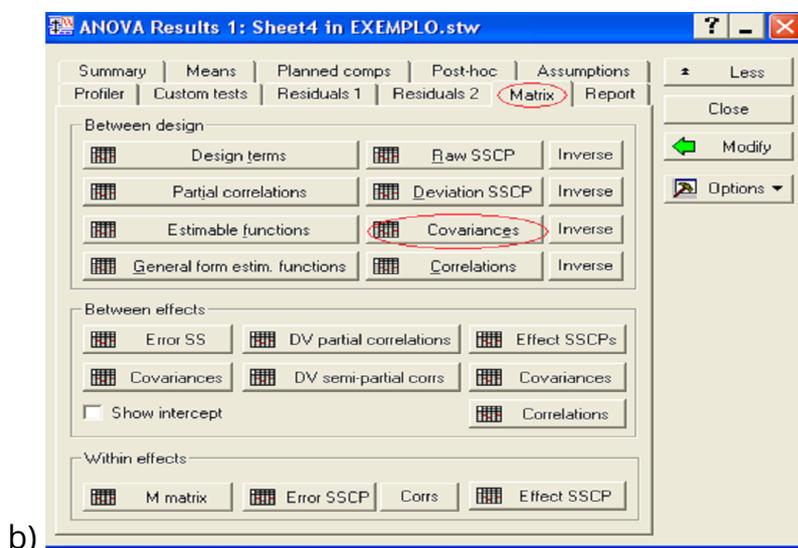
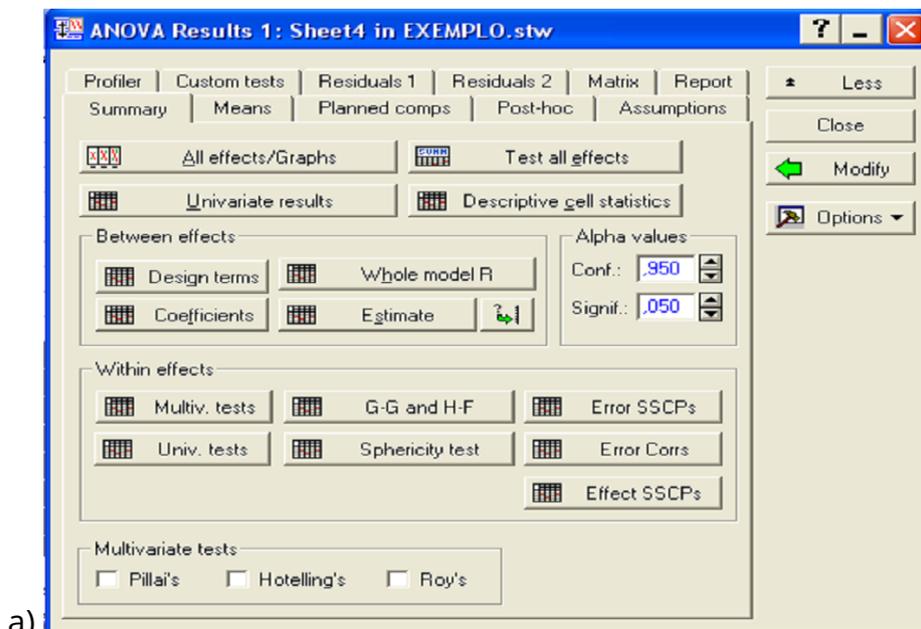
Pela análise de perfis na Figura 3(c), pode-se concluir que os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos, pois não existe interação escala x classes sociais (A*B), pois $p > 0,05$.

Para a segunda hipótese analisada $H_0^{(2)}$ pode-se concluir que os perfis médios de resposta correspondentes às diversas escalas (B) não são coincidentes, isto é, existe efeito das escalas (B). Rejeita-se $H_0^{(2)}$ em nível de 0,05 de significância, pois $p < 0,05$.

Para a terceira hipótese $H_0^{(3)}$ pode-se concluir que os perfis médios de resposta correspondentes às diversas escalas (B) não são paralelos ao eixo das abscissas, existe efeito das classes sociais (A). Rejeita-se $H_0^{(3)}$ em nível de 0,05 de significância, pois $p < 0,05$.

Ao clicar na opção *More results*, mostrada na Figura 3 (b), irá aparecer a caixa de seleção indicada na Figura 4 (a). Nessa caixa de seleção, ao se selecionar a guia *Matrix* a Figura 4 (b), aparecerá:

Figura 4 – figura 4 (a) Caixa de seleção das análises, figura 4 (b) Caixa de comando para obtenção das matrizes



Para obter a matriz de variância e covariância clique na Figura 4 (b) em *Covariances*, como indicado acima. Após, irá aparecer os resultados apresentados nas Figura 5 (a). Clica-se em *Error SS* no subitem *Between effects* na Figura 4 (b) para encontrar a matriz W (matriz das somas dos quadrados e produtos cruzados dentro dos grupos), Figura 5 (b).

Figura 5 – figura 5 (a) Caixa de resultados da matriz de variância e covariância., figura 5 (b) Caixa de resultados da matriz das somas dos quadrados e produtos cruzados dentro dos grupos

a)

Variance/Covariance Matrix (Sheet4 in EXEMPLO.stw)										
Variances and covariances for the vectors in the design matrix X										
Effect	Level	Column	Effect (F/R)	Col. 1 Intercept	Col. 2 A	Col. 3 A	Col. 4 A	Col. 5 B1	Col. 6 B2	Col. 7 B3
Intercept			1 Fixed							
A	1	2	Fixed		0,561905	0,190476	0,200000	2,29524	3,07619	2,52857
A	2	3	Fixed		0,190476	0,447619	0,200000	0,72381	1,21905	0,55714
A	3	4	Fixed		0,200000	0,200000	0,400000	0,60000	1,00000	0,80000
B1			5		2,295238	0,723810	0,600000	11,66190	13,20952	11,27857
B2			6		3,076190	1,219048	1,000000	13,20952	18,54762	13,89286
B3			7		2,528571	0,557143	0,800000	11,27857	13,89286	13,52857

b)

Residual Sums of Squares and Cross Prod				
Sigma-restricted parameterization				
Effective hypothesis decomposition				
Effect	Variable	B1	B2	B3
Error	B1	42,80000	11,20000	18,20000
	B2	11,20000	30,80000	3,80000
	B3	18,20000	3,80000	38,30000

Da mesma forma, pode-se encontrar a matriz B (matriz das somas de quadrados e produtos cruzados entre os grupos). Basta clicar em *Effect SSCP*, Figura 4 (b), no subitem *Between effects*. Após este procedimento aparecerá a matriz B. Cujo resultado é mostrado na Figura 6 (a) A matriz T, que é a matriz Total, pode ser encontrada clicando-se, na Figura 9, em *Deviation SSCP* no subitem *Between design*. Aparecendo o resultado apresentado na Figura 6 (b).

Figura 6 – figura 6 (a) Caixa de resultados da matriz das somas dos quadrados e produtos cruzados entre os grupos, figura 6 (b) Caixa de resultados da matriz total

a)

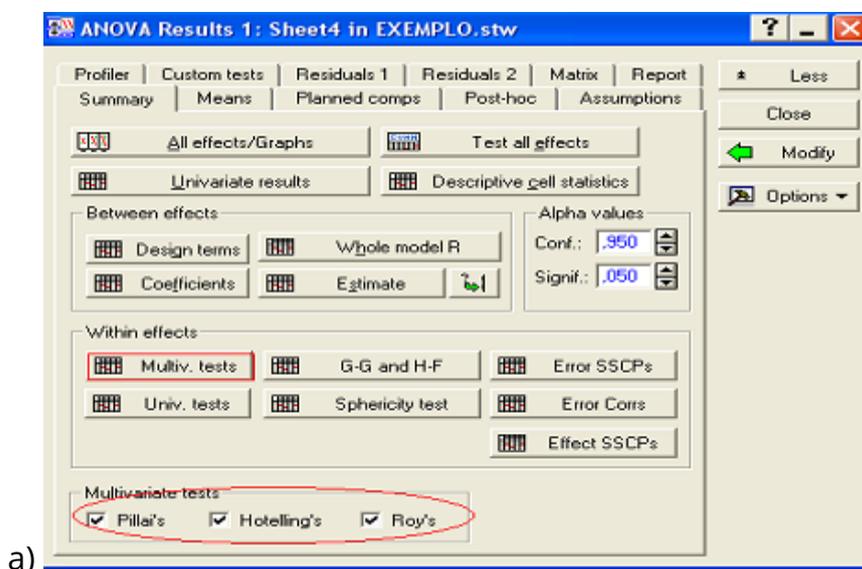
Between Effects Sums of Squares and Cross Products (Sheet4 in EXEMPLO.stw)				
Sigma-restricted parameterization				
Effective hypothesis decomposition				
Effect	Variable	B1	B2	B3
A	B1	190,4381	252,9905	207,3714
	B2	252,9905	340,1524	274,0571
	B3	207,3714	274,0571	232,2714

b)

Deviation Sums of Squares and Cross Products (Sheet4 in EXEMPLO.stw)										
Deviation sums of squares and cross products for the vectors in the design matrix X										
Effect	Level	Column	Effect (F/R)	Col. 1 Intercept	Col. 2 A	Col. 3 A	Col. 4 A	Col. 5 B1	Col. 6 B2	Col. 7 B3
Intercept		1	Fixed							
A	1	2	Fixed		11,23810	3,80952	4,00000	45,9048	61,5238	50,5714
A	2	3	Fixed		3,80952	8,95238	4,00000	14,4762	24,3810	11,1429
A	3	4	Fixed		4,00000	4,00000	8,00000	12,0000	20,0000	16,0000
B1		5			45,90476	14,47619	12,00000	233,2381	264,1905	225,5714
B2		6			61,52381	24,38095	20,00000	264,1905	370,9524	277,8571
B3		7			50,57143	11,14286	16,00000	225,5714	277,8571	270,5714

Na guia *Summary* no subitem *Within effects*, pode-se encontrar os valores de alguns critérios de teste para análise multivariada de perfis, clicando-se no botão *Multiv. Tests*. Pode-se selecionar mais de um critério, conforme Figura 7 (a). Após isso, os resultados dos testes aparecerão, conforme mostra a Figura 7 (b) abaixo.

Figura 7 – figura 7 (a) Caixa de seleção para os testes multivariados, figura 7 (b) Caixa de resultados para os testes multivariados



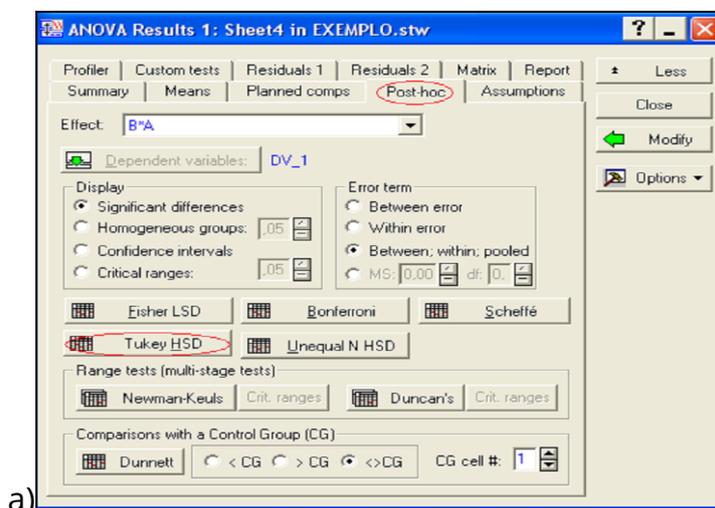
b)

Multivariate tests for repeated measure: DV_1 (Sheet4 in EXEMPLO.stw)						
Sigma-restricted parameterization						
Type II decomposition						
Effect	Test	Value	F	Effect df	Error df	p
B	Wilks	0,501289	7,958842	2	16	0,003988
	Pillai's	0,498711	7,958842	2	16	0,003988
	Hotelling	0,994855	7,958842	2	16	0,003988
	Rov's	0,994855	7,958842	2	16	0,003988
B*A	Wilks	0,563332	1,772525	6	32	0,136442
	Pillai's	0,487260	1,825259	6	34	0,123446
	Hotelling	0,685343	1,713359	6	30	0,152179
	Rov's	0,508849	2,883477	3	17	0,066180

Conforme a Figura 7 (b), pode-se concluir que não existe interação (A*B) entre escalas e classes sociais (pois $p > 0,05$) para todos os critérios; e que existe efeito das classes sociais (pois $p < 0,05$) para todos os critérios.

Para realizar as comparações múltiplas seleciona-se a guia *Post-hoc*, como indicado na caixa de seleção da Figura 8 (a). Como pela análise de variância rejeitou-se a hipótese de interação (A*B). Aqui se realizou o teste *Tukey HSD* somente para efeitos dentro das escalas (A) e dentro das classes sociais (B). Para isto, deve se selecionar na caixa de seleção, Figura 8 (a), os efeitos A e B na seta ao lado da opção *Effect*. Os resultados destas análises são mostrados, respectivamente, nas Figuras 8 (b) e 8 (c) a seguir:

Figura 8 – figura 8 (a) Caixa de seleção para as comparações múltiplas, figura 8 (b) Caixa de resultados para as comparações múltiplas entre as classes sócio-econômicas, figura 8(c) Caixa de resultados para as comparações múltiplas entre as escalas de atitudes maternas



b)

Tukey HSD test; variable DV_1 (Sheet4 in EXEMPLO.stw) Probabilities for Post Hoc Tests Error: Within MS = 1,5431, df = 34,000				
Cell No.	B	{1}	{2}	{3}
1	B 1	14,524	15,619	15,857
2	B 2	0,019441	0,019441	0,004008
3	B 3	0,004008	0,809724	

c)

Tukey HSD test; variable DV_1 (Sheet4 in EXEMPLO.stw) Probabilities for Post Hoc Tests Error: Between MS = 3,4961, df = 17,000					
Cell No.	A	{1}	{2}	{3}	{4}
1	1	19,250	14,400	14,000	10,000
2	2	0,000179	0,000179	0,944703	0,000236
3	3	0,000179	0,944703		0,000506
4	4	0,000178	0,000236	0,000506	

A partir dos resultados mostrados nas figuras 8 (b) e 8 (c) pode-se construir as Tabelas 4 e 5 a seguir, onde as médias são identificadas ou não por mesma letra.

Tabela 4 – Médias estimadas das classes sociais com a classificação pelo teste de Tukey HSD

Classe	Média
1	19,25 a*
2	14,40 b
3	14,00 b
4	10,00 c

*Médias não ligadas por mesma letra não diferiu pelo teste de Tukey HSD em nível de 5% de probabilidade de erro.

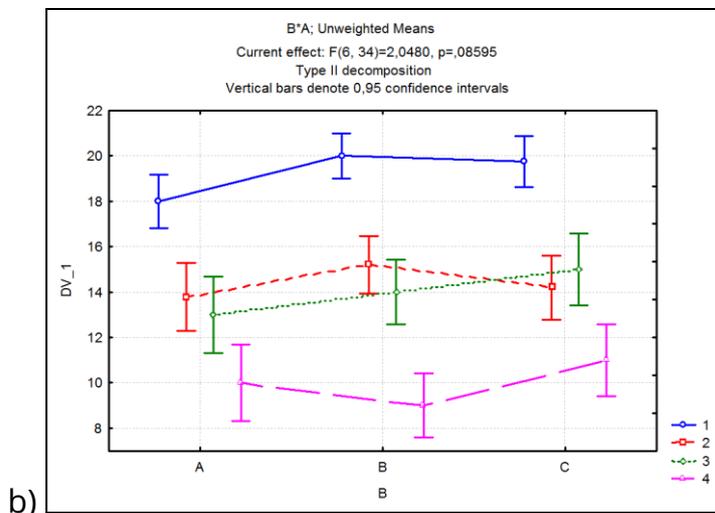
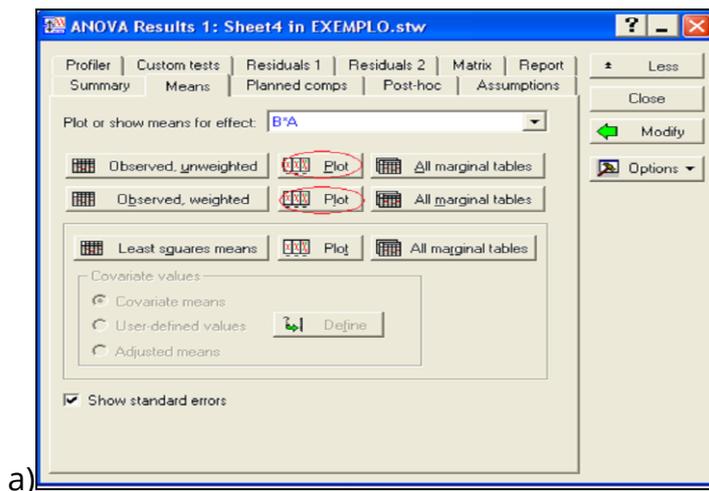
Tabela 5 – Médias estimadas das escalas de atitude com a classificação pelo teste de Tukey HSD

Classe	Média
B3	15,857 a*
B2	15,619 a
B3	14,524 b

*Médias não ligadas por mesma letra não diferiu pelo teste de Tukey HSD em nível de 5% de probabilidade de erro.

O gráfico de perfis de médias pode ser obtido na guia *Means*, Figura 9 (a), clicando em *Plot* como o indicado, na caixa de seleção: O gráfico na Figura 9 (b) é o de perfis de médias.

Figura 9 – figura 9 (a) Caixa de seleção para a plotagem dos gráficos de perfis de médias., figura 9 (b) Gráfico dos perfis médios



Na Figura 9 (b), tem-se a representação gráfica dos perfis médios observados em três escalas de atitudes maternas (A, B, C) e quatro tipos de classes socio-econômicas (1; 2; 3; 4). Observa-se que as classes sócio-econômicas parecem diferir, a classe A parece diferir das classes B e C. Ainda pode-se observar que a primeira escala difere das três restantes, e a segunda e a terceira não parecem ser diferentes, e a quarta é diferente das demais.

4 CONSIDERAÇÕES FINAIS

O software STATISTICA versão 9.0 comprovou ser eficaz na análise de variância multivariada com medidas repetidas, mesmo no caso desbalanceado, realizando os testes para as três hipóteses nulas necessárias, os testes post-hoc e o gráfico de perfis.

A utilização de softwares estatísticos, sejam livres ou não, são de ampla utilidade para realização de análises estatísticas de dados, visto que trazem mais praticidade ao desenvolvimento do trabalho. Sendo assim, o software em estudo é uma alternativa que deve ser considerada na prática, pois, o mesmo foi adquirido pela UFSM, evitando dessa forma o uso softwares sem licença.

REFERENCES

ANDRADE FS, SILVA AM, ARIDE PHR, OLIVEIRA AT. Análise físico-química e da microbiota da água do lago Macurany, Parintins, Amazonas. **Biota Amazônia** (Biote Amazonie, Biota Amazonia, Amazonian Biota). 2016;6(2):132-134.

APHA - American Public Health Association. **Standard Methods for the Examination of Water and Wastewater**, 20 ed., Washington, DC: Apha, 2012.

Barra do Corda, Secretaria Municipal de Educação. **Plano municipal de educação 2015-2025**, 83f, 2015.

BEHBAHANI BA, YAZDI F T, SHAHIDI F, MORTAZAVI S A, MOHEBBI M. Principal component analysis (PCA) for investigation of relationship between population dynamics of microbial pathogenesis, chemical and sensory characteristics in beef slices containing Tarragon essential oil. **Microbial Patogenesis**. 2017;105:37-50.

BELLUTA I, JESUS AS, VIEIRA PM, CORRÊA NM, RALL VLM, VALENTE JPS. Qualidade da Água, Carga Orgânica e de Nutrientes na Foz do Córrego da Cascata: Contribuição da Sub-Bacia para a Represa de Barra Bonita, Rio Tietê (SP) (Water Quality, Organic Loading and Nutrient Loading in the Mouth of Cascata Brook: The Subbasin). **Revista Brasileira de Geografia Física**. 2016;9(1):305-318.

BEZERRA RAR, MARQUES BS, ALMEIDA FB, SAMPAIO WMS, GIONGO P, SILVA AM. Estudo de variação de turbidez (NTU) do Rio santa maria da vitória, entre os municípios de Santa Maria do Jetibá e Santa Leopoldina Estado do Espírito Santo. **Anais Simpac**. 2017;7(1).

Brasil Ministério de Minas e Energia. Secretaria de Geologia, Mineração e Transformação Mineral. Geodiversidade do Estado do Maranhão. **Programa Geologia do Brasil Levantamento da Geodiversidade**, 2013.

CARVALHO FIM, LEMOS VP, DANTAS FILHO HÁ, DANTAS KGF. Avaliação da qualidade das águas subterrâneas de Belém a partir de parâmetros físico-químicos e níveis de elementos traço usando análise multivariada. **Revista Virtual de Química**. 2015;7(6):2221-2241.

CODEVASF - **Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba** (2014). Disponível em: <http://www.codevasf.gov.br/>. Acessado em: 25/09/2017.

COIMBRA CD, CARVALHO G, PHILIPPINI H, SILVA, MFM, NEIVA E. Determinação da concentração de metais traço em sedimentos do estuário do rio Maracápe-PE/BRASIL. **Brazilian Journal of Aquatic Science and Technology**. 2016;19(2):58-75.
DA CONCEIÇÃO FT, BONOTTO DM. Relações hidroquímicas aplicadas na avaliação da qualidade da água e diagnóstico ambiental na bacia do Rio Corumbataí (SP). **Geochimica brasiliensis**. 2017;16(1):1.

DA SILVA MA, VARGAS RR. Análise de fósforo e DQO em amostras de águas naturais no município de Guarulhos. **Revista Educação-UNG-Ser**. 2017;11(3):81.

FERNANDES FMSS. Potencial Químico, Fugacidade e Atividade. **Revista de Ciência Elementar**. 2017;5(4):1-7.

FERREIRA FS, QUEIROZ T M, SILVA TV, ANDRADE ACO. À margem do rio e da sociedade: a qualidade da água em uma comunidade quilombola no estado de Mato Grosso. **Saúde e Sociedade**. 2017;26:822-828.

FREITAS MA. Hidrogeoquímica e isotopia de águas com alta salinidade do Sistema Aquífero Serra Geral na região do Alto Rio Uruguai, Brasil [thesis]. Porto Alegre: **Instituto de Geociências, Universidade Federal do Rio Grande do Sul**. 2016. 195p.

GOIS FA, SOUZA GA, OLIVEIRA MJ, LIMA RS, KOSLOWSKI, LAD. Análise da qualidade da água quanto ao despejo industrial têxtil no Rio dos Índios. **Meio Ambiente e Sustentabilidade**. 2017;9(5).

HAIJGHOLIZADEH M, MELESSE AM. Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. **Catena**; 2017;151:247-258.

HONGYU K. **Comparação do GGEbiplot ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente** [thesis]. Piracicaba: Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo; 2015. 155p.

LORDELO LMK, PORSANI JM, BORJA PC. Qualidade físico-química da água para abastecimento humano em municípios do sertão da Bahia: um estudo considerando diversas fontes de suprimento. **Revista Águas Subterrâneas**; 2018;32(1):97-105.

KOLM HE; SIQUEIRA A, MACHADO EC. Influência da pluviosidade na qualidade da água de dois sangradouros do litoral do Paraná, Brasil. *Braz. J. Aquat. Sci. Technol.*; 2016;20(2):1-11.

MUNIN RM, SUTTI BO, CHIOZZINI VG, BRAGA ES. Relação do fosfato com a formação e degradação da matéria orgânica no sistema hídrico que cruza o município de Mogi das Cruzes (São Paulo, Brasil). **Anais do Encontro Nacional de Pós Graduação**. 2017;1(1):469-473.

RIBEIRO MA, CAMARGO E, FRANCA DT, CALASANS JT, BRANCO MSLC, TRIGO AJ. Gestão da Água e Paisagem Cultural. **Revista da Universidade Federal de Minas Gerais**. 2016;20(2).

SASSOMA ITL, SOUZA IF, AGUIAR NETO AO, CARVALHO CM. Avaliação temporal e espacial de características físico-químicas em águas superficiais do rio Catumbela, Angola. **Revista Brasileira de Agricultura Irrigada-RBAI**. 2015;9(3):113-126.

VASCONCELOS PES, VIVEIROS TB, ARRAZ AA. Caracterização da Situação Socioeconômica e Ambiental do Município de Barra do Corda após o Implemento da Construção Civil Durante os Anos 2002 à 2010. In: **VII CONNEPI-Congresso Norte Nordeste de Pesquisa e Inovação**. 2012.