UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA E MODELAGEM QUANTITATIVA

ANÁLISE DE VARIÂNCIA MULTIVARIADA APLICADA A DADOS COM MEDIDAS REPETIDAS

MONOGRAFIA DE ESPECIALIZAÇÃO

Mara Rubia Machado Couto

Santa Maria, RS, Brasil

2006

ANÁLISE DE VARIÂNCIA MULTIVARIADA APLICADA A DADOS COM MEDIDAS REPETIDAS

por

Mara Rubia Machado Couto

Monografia apresentada ao Curso de Especialização do Programa de Pós-Graduação em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa.

Orientadora: Prof. MSc. Luciane Flores Jacobi

Santa Maria, RS, Brasil

2006

Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa

A Comissão Examinadora, abaixo assinada,
aprova a Monografia de Especialização

ANÁLISE DE VARIÂNCIA MULTIVARIADA APLICADA A DADOS COM MEDIDAS REPETIDAS

elaborada por
Mara Rubia Machado Couto

como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa

COMISSÃO EXAMINADORA:

Luciane Flores Jacobi, MSc.
(Presidente/Orientadora)

Adriano Mendonça Souza, Dr. (UFSM)

Fernando de Jesus Moreira Júnior, MSc. (UFSM)

Santa Maria, 07 de março de 2006.

Dedico com muito carinho este trabalho ao Rodrigo pela dedicação, compreensão e principalmente pela sua paciência.

AGRADECIMENTOS

A minha orientadora, Profª. MSc Luciane Flores Jacobi, pelo acompanhamento durante este trabalho, pela oportunidade deste aprendizado, por toda a valiosa contribuição nesta pesquisa e sua amizade.

A todos os professores do Departamento de Estatística pela sabedoria que souberam transmitir.

Aos meus colegas em especial: Alícia Bolfoni Dias, Gabriela Bilibio, Gilvete Lírio, Mauricio Lutz e Rejane Cardoso pelas horas de estudos, por todas as suas contribuições e amizade.

RESUMO

Monografia de Especialização
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria

ANÁLISE DE VARIÂNCIA MULTIVARIADA APLICADA A DADOS COM MEDIDAS REPETIDAS
AUTORA: MARA RUBIA MACHADO COUTO
ORIENTADORA: LUCIANE FLORES JACOBI
Data e local da Defesa: Santa Maria, 07 de março de 2006.

Em algumas situações, no processo de análise de dados, existe o interesse de analisar a performance de várias 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 material de suporte teórico-prático para todos os que têm interesse em aplicar as técnicas acima citadas. Para seu desenvolvimento é realizada uma revisão de literatura que reúna os principais aspectos destas metodologias. Após esta revisão, desenvolve-se algumas instruções sobre os procedimentos operacionais necessários para a execução destas técnicas no software *Statistica 7.0*. Com este trabalho proporciona-se, aos usuários de técnicas estatísticas, um referencial básico sobre o uso e aplicações da análise multivariada, particularmente, as técnicas de análise de variância multivariada e análise de perfis de médias.

Palavras-chave: análise multivariada, medidas repetidas, dados longitudinais, análise de variância multivariada, análise de perfis.

ABSTRACT

Specialization Monograph
Post Gratuated Program in Statistics and Quantitative Modelling
Federal University of Santa Maria

MULTIVARIATE ANALYSIS OF VARIANCE: APPLIED THE DATA WITH REPEATED MEASURES

AUTHOR: MARA RUBIA MACHADO COUTO
ADVISOR: LUCIANE FLORES JACOBI
Date and place of defense: Santa Maria, March 7, 2006.

In some situations, in the process of analysis of data, the interest exists to analyze the performance of some variable jointly, and to determine the influence or the importance of each changeable in presence of excessively. For this, the techniques of multivariate analysis are used. When, in particular, this interest is in identifying to the existence of effect between treatments and/or conditions of evaluation and, to verify the existence of interaction between these factors; the job of an analysis of multivariate variance is appropriate, or still, an analysis of profiles of averages; that are methods multivariate applied in the analysis of data with repeated measures and longitudinal data. The present work has for objective to develop a material of theoretician-practical support for all those that will have interest in applying the techniques above cited. For its development an ample revision of literature was carried through that congregated the main aspects of these methodologies. After this revision, developed some instructions on the necessary operational procedures for the execution of these techniques in *Statistica* software. With this work, one provided the users to it of statistical techniques, a basic referencial on the use and applications of the multivariate analysis, particularly, the techniques of multivariate analysis of variance and analysis of profiles of averages.

Key-Words: multivariate analysis, repeated measures, longitudinal data, multivariate analysis of variance, analysis of profiles.

LISTA DE FIGURAS

FIGURA 1-Representação gráfica genérica dos perfis médios de resposta 34
FIGURA 2-Gráfico Q – Q 47
FIGURA 3-Perfis médios de atitudes maternas conforme classe socio-econômica 53
FIGURA 4- Caixa de seleção da análise de variância 54
FIGURA 5-Caixa de seleção para especificação do método. 55
FIGURA 6-Caixa de seleção para especificação dos parâmetros na análise de variância multivariada 55
FIGURA 7-Caixa de seleção das variáveis 56
FIGURA 8-Caixa de seleção para a análise de variância multivariada 56
FIGURA 9-Caixa de seleção para os efeitos da análise de variância multivariada 57
FIGURA 10-Caixa com os resultados da análise 57
FIGURA 11-Caixa de seleção das análises 58
FIGURA 12-Caixa de comandos para obtenção das matrizes 58
FIGURA 13-Caixa de resultados da matriz de variância e covariância 59
FIGURA 14-Caixa de resultados da matriz das soma de quadrados e produtos cruzados dentro dos grupos 59
FIGURA 15-Caixa de resultados da matriz das soma de quadrados e produtos cruzados entre os grupos 60
FIGURA 16-Caixa de resultados da matriz total 60
FIGURA 17-Caixa de seleção para testes multivariados 60
FIGURA 18-Caixa de resultados para os testes multivariados 61
FIGURA 19-Caixa de seleção para as comparações múltiplas 61
FIGURA 20-Caixa de resultados para as comparações múltiplas entre as classes sócio-econômicas 62
FIGURA 21-Caixa de resultados para as comparações múltiplas entre as escalas de atitudes maternas 62
FIGURA 22-Caixa de seleção para a plotagem dos gráficos de perfis de médias 62
FIGURA 23-Gráfico dos perfis de médios 63

LISTA DE TABELAS

TABELA 1 - Estrutura de dados básica para estudos longitudinais 22
TABELA 2 - Análise de variância multivariada 28
TABELA 3 - Análise variância para medidas repetidas.......... 37
TABELA 4 - Distribuição da estatística Λ 40
TABELA 5 - Observações dos perfis 42
TABELA 6 - Valores encontrados para os pares ordenados 46
TABELA 7 - Análise de variância do perfil para o exemplo ... 52

LISTA DE SIGLAS

ANOVA – *Analysis of Variance (análise de variância)*

MANOVA – *Multivariate Analysis of Variance (análise de variância multivariada)*

SUMÁRIO

1 INTRODUÇÃO 12
1.1 Objetivos 13
1.1.1 Objetivo Geral.. 13
1.1.2 Objetivos Específicos 13
1.2 Justificativa.. 14
1.3 Materiais e Métodos 14
1.4 Estrutura do trabalho.. 14

2 REVISÃO DE LITERATURA 16
2.1Conceitos Básicos para Análise Multivariada 17
2.1.1 Medidas Descritivas de Populações Multivariadas.................. 17
2.1.2 Medidas Descritivas para Amostras Multivariadas 18
2.2 Análise Multivariada de Medidas Repetidas.......................... 19
2.3 Medidas Repetidas e Dados Longitudinais........................... 20

3 ANÁLISE DE VARIÂNCIA MULTIVARIADA............................... 23
3.1 Verificação do Pressuposto de Normalidade........................ 24
3.2 Verificação da Igualdade de Matrizes de Variância e Covariância 25
3.3 Teste de Igualdade de Médias para *k* Grupos (Oneway MANOVA).. 27

4 ANÁLISE DE PERFIS DE MÉDIAS... 31
4.1 Análise Multivariada de Perfis 31
4.2 Modelo Matemático para Análise Multivariada de Perfis....... 32
4.3 Alguns Critérios de Teste para a Análise Multivariada de Perfis 39
4.3.1 União-Intersecção de Roy 40
4.3.2 Critério de Wilks 40
4.3.3 Critério de Pillai 41
4.3.4 Critério de Hotelling-Lawley 41

4.4 Algumas Considerações sobre os Critérios de Teste para a Análise Multivariada de Perfis 41

5 APLICAÇÃO DOS MÉTODOS 42
5.1 Aplicação dos Métodos com Saídas do Software *Statistica7.0* 54

6 CONSIDERAÇÕES FINAIS 64

7 REFERÊNCIAS 66

1 INTRODUÇÃO

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 investigar-se o comportamento de uma ou mais variáveis respostas ao longo do tempo, ao longo de uma condição de avaliação; ou 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 de grupo pode-se fazer, em determinadas situações, mais adequada do que isolar cada variável e estudá-la separadamente. Para este tipo de estudo a literatura técnico-científica utiliza os métodos de análise multivariada.

Somente os métodos de estatística multivariada permitem que se analise a performance conjunta das variáveis e se determine a influência ou importância de cada variável na presença das restantes.

O presente trabalho enfoca as técnicas de análise multivariada que englobam medidas repetidas e dados longitudinais completos, balanceados e regulares no tempo, dando-se particular atenção aos métodos de análise de variância multivariada (MANOVA) e análise de perfis de médias.

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.

Uma das principais desvantagens na análise de estudos com medidas repetidas refere-se à existência de correlações entre as medidas devido estas serem realizadas no mesmo indivíduo.

A análise de variância multivariada (MANOVA) é uma extensão da análise de variância simples (ANOVA) e, da mesma forma, esta última, é apropriada sempre que se pretenda fazer comparações de médias. A principal diferença entre as duas está em que a MANOVA compara médias para diferentes variáveis simultaneamente, enquanto a ANOVA avalia diferenças de médias apenas em uma variável.

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.

1.1 Objetivos

1.1.1 Objetivo Geral

Desenvolver um trabalho 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.

1.1.2 Objetivos Específicos

i) Reunir os principais aspectos dessas metodologias;

ii) aplicar as metodologias em um exemplo hipotético que sirva como referencial didático aos interessados nas técnicas para aplicações práticas e também trabalhos futuros;

iii) desenvolver instruções para o uso do pacote estatístico *Statistica* 7.0 na solução de problemas usando esta metodologia.

1.2 Justificativa

Este trabalho aborda a importância do desenvolvimento de um referencial básico para os usuários das técnicas multivariadas que necessitem aplicar em seus trabalhos e pesquisas, os métodos de análise de variância multivariada e análise de perfis de médias.

1.3 Material e Métodos

Para o desenvolvimento do estudo será realizada uma ampla revisão de literatura que reuna os principais aspectos da metodologia multivariada os quais abordam as técnicas de análise de variância multivariada e análise de perfis de médias aplicadas a medidas repetidas e dados longitudinais.

A fim de tornar esta pesquisa um referencial didático, será utilizado um exemplo hipotético que demonstre, passo-a-passo, as técnicas multivariadas aos interessados em sua aplicação.

Desenvolveu-se algumas instruções para analise e interpretação destas técnicas no pacote estatístico *Statistica* 7.0.

1.4 Estrutura do trabalho

O referido trabalho mostra as técnicas de análise multivariada de medidas repetidas.

No Capítulo 2, tem-se uma breve revisão de literatura sobre conceitos básicos necessários para a análise multivariada com enfoque em medidas repetidas e dados longitudinais.

No Capítulo 3, trata-se da análise de variância multivariada (MANOVA) que é uma técnica, no âmbito multivariado, para analisar dados com medidas repetidas. Neste capítulo descreve-se as pressuposições e hipóteses de interesse na análise de dados longitudinais. Tem-se, também, alguns métodos de verificação dos pressupostos de normalidade multivariada e de igualdade de matrizes de variância e covariância.

No Capítulo 4, refere-se à análise de perfis de médias, no contexto multivariado; também conhecida como análise multivariada de perfis. Neste capítulo o modelo multivariado é descrito, juntamente com as hipóteses para análise. Também são apresentados alguns critérios para a análise de perfis, como o *Critério de Roy*, *Critério de Wilks*, *Critério de Hotelling* e o *Critério de Pillai*.

No Capítulo 5, as técnicas, tratadas nos capítulos anteriores, são exemplificadas através de um exemplo hipotético propoto. Os resultados e interpretações da análise serão apresentados com as “saídas” do software *Statistica 7.0*, e detalhadas em cada passo.

2 REVISÃO DE LITERATURA

O presente capítulo faz considerações sobre a análise multivariada de medidas repetidas. Embora a análise multivariada tenha como origem a estatística uni e bivariada, a extensão no domínio multivariado introduz conceitos adicionais e estes têm resultados de particular relevância. Esses conceitos vão desde a decisão para julgar a construção básica de um conjunto multivariado até as conclusões estatísticas dos testes de significância.

2.1 Conceitos Básicos para Análise Multivariada

Nesta seção tem-se, por objetivo, fornecer conceitos básicos para a análise de fenômenos que envolvam mais de uma variável aleatória. Esta introdução à estatística multivariada é relevante, pois inclui métodos de análise das relações de múltiplas variáveis dependentes e/ou múltiplas variáveis independentes.

Portanto, serão tratadas aqui as populações e amostras multivariadas e as respectivas medidas descritivas: vetores de médias, matriz de variância e covariância e de correlação.

Seja $X_1, X_2, \ldots, X_i$ um conjunto de I variáveis aleatórias, que pode ser apresentado sob a forma de um vetor $\underset{\sim}{X}' = [X_1, X_2, \ldots, X_i]$.

Pretende-se estudar estas I características para um conjunto de N indivíduos (no caso de uma população ou *n* no caso de uma amostra).

A matriz dos dados observados terá a seguinte forma:

$$X = (X_{ij}) = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1n_i} \\ X_{21} & X_{22} & \ldots & X_{2n_i} \\ \ldots & \ldots & \ldots & \ldots \\ X_{i1} & X_{i2} & \ldots & X_{in_i} \end{bmatrix}$$

Sendo X_{ij} o valor da variável i $(I = 1,2,\ldots,i)$ para o indivíduo j $(J = 1,2,\ldots,n_i)$.

2.1.1 Medidas Descritivas de Populações Multivariadas

Seja uma população de dimensão N, caracterizada por um vetor de p variáveis, o vetor de médias é definido como sendo:

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_i \end{bmatrix}$$, em que μ_i será a média da variável i, isto é, $\mu_i = \frac{1}{N}\sum_{J=1}^{N} \underset{\sim}{X}_{ij}$.

O cálculo da variância para cada variável da população, é o seguinte:

$$\sigma_i^2 = \frac{1}{N}\sum_{J=1}^{N}\left(X_{ij} - \mu_i\right)^2 \text{ ou } \sigma_i^2 = E\left[\left(X_{ij} - \mu_i\right)^2\right]$$

O desvio-padrão é a raiz quadrada positiva da variância: $\sigma_i = \sqrt{\sigma^2{}_i}$.

Para dois atributos diferentes, um outro parâmetro a calcular é a covariância, definida como a média do produto dos desvios dos valores das variáveis relativamente às respectivas médias, e que mede o grau de relação linear entre cada dois atributos:

$$\sigma_{ij}^2 = \frac{1}{N}\sum_{J=1}^{N}\left[\left(X_{ij} - \mu_i\right)\left(X_{ij} - \mu_j\right)\right] \text{ ou } \sigma_{ij}^2 = E\left[\left(X_{ij} - \mu_i\right)\left(X_{ij} - \mu_j\right)\right]$$

Note que quando $i = j$, a covariância reduz-se à variância e também, que $\sigma_{ij} = \sigma_{ji}$, quaisquer que sejam i e j.

Tomando todas as variáveis simultaneamente, constrói-se uma matriz de variância/covariância da população, Σ.

$$\Sigma_{(i \times i)} = \left(\sigma_{ij}\right) = \begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ \ldots & \ldots & \ldots & \\ \sigma_{i1} & \sigma_{i2} & \ldots & \sigma_i^2 \end{bmatrix}$$

Σ é uma matriz simétrica que tem na diagonal principal as variâncias das p variáveis, enquanto os outros elementos são as covariâncias entre cada duas variáveis. Esta matriz pode ser calculada da seguinte forma:

$$\Sigma = \frac{1}{N}\sum_{J=1}^{N}\left[\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)'\right]$$

A partir da matriz de variância e covariância pode-se atribuir uma matriz de correlações, dada por:

$$\rho_{ij} = \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \ldots & \ldots & \ldots & \\ \rho_{k1} & \rho_{k2} & \ldots & 1 \end{bmatrix} \text{ onde } \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 . \sigma_j^2}} .$$

2.1.2 Medidas Descritivas para Amostras Multivariadas

Seja n o número de observações de um subconjunto de todas as observações possíveis para a população em estudo, isto é, uma amostra de tamanho n. Pode-se estimar para esta amostra, os seguintes estimadores dos parâmetros indicados na subseção 2.1.1:

I) Vetor de médias amostrais

$$\overline{\underset{\sim}{X}} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \ldots \\ \overline{X}_i \end{bmatrix} \text{ sendo } \overline{X}_i = \frac{1}{n}\sum_{J=1}^{n} X_{ij} \quad (I = 1,2,\ldots,i) \text{ a média amostral da variável i.}$$

II) Matriz de variância/covariância amostral

$$S = \begin{bmatrix} S_1^2 & & & \\ S_{21} & S_2^2 & & \\ \ldots & \ldots & \ldots & \\ S_{i1} & S_{i2} & \ldots & S_i^2 \end{bmatrix} \text{ sendo } S = \frac{1}{n-1}\sum_{J=1}^{n}\left[\left(\underset{\sim}{X}_J - \overline{X}\right)\left(\underset{\sim}{X}_J - \overline{X}\right)'\right] \text{ e}$$

$$S_{ij} = \frac{1}{n}\sum_{J=1}^{n}\left[\left(X_{ij} - \overline{X}_i\right)\left(X_{ij} - \overline{X}_j\right)\right] \text{ a covariância amostral entre as variáveis i e j.}$$

III) Matriz de correlações amostrais

$$R = \begin{bmatrix} 1 & & & \\ r_{21} & 1 & & \\ \cdots & \cdots & \cdots & \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}$$ sendo $r_{ij} = \dfrac{S_{ij}}{\sqrt{S_i^2 . S_j^2}}$, a correlação amostral entre as variáveis i e j.

Conforme Reis (1997), um dos objetivos em que consiste a estatística multivariada é a simplificação de dados (vetores de médias, matrizes de variância e covariância e de correlação), descrevendo a informação através de um número reduzido de dimensões de análises.

2.2 Análise Multivariada de Medidas Repetidas

A análise de medidas repetidas é uma metodologia em que o interesse é direcionado à avaliação do comportamento de uma ou mais variáveis respostas classificadas em diferentes subpopulações segundo um ou mais fatores (tratamentos) ao longo do tempo ou de diversas condições de avaliação, que representam as unidades de observação.

Dois enfoques têm sido discutidos na literatura para a análise de medidas repetidas sobre unidades experimentais. O primeiro adota procedimentos univariados, considerando as observações como subdivisão de parcelas, ou seja, subparcelas. E um segundo utiliza o procedimento multivariado, que considera cada unidade experimental como um vetor de observações.

Segundo Singer, Rocha e Nobre (2004), uma das características mais importantes para a análise de medidas repetidas é o processo de obtenção dos dados, existindo basicamente duas formas de obtenção de medidas repetidas. A primeira corresponde à atribuição aleatória da ordem, quando as medidas sob diferentes condições de avaliação são realizadas na mesma unidade de investigação; abrangendo estudos obtidos por meio de delineamentos com parcelas subdivididas (“split-plot”) ou delineamentos com intercâmbio (“crossover”). A segunda ocorre quando as medidas realizadas sob diferentes condições de avaliação estão dispostas ao longo de uma escala ordenada, isto é, são obtidas seqüencialmente e, neste caso, os estudos são denominados longitudinais.

Conforme Riboldi et al.(2003), a análise de medidas repetidas tem os seguintes objetivos:

i. comparação das diferentes subpopulações quanto ao padrão de variação das respectivas distribuições de respostas ao longo das diferentes condições de avaliação, verificação da existência de interação entre o fator ou fatores (tratamentos) que define(m) as subpopulações e aquele que define as condições de avaliação, verificar se os perfis são paralelos;

ii. comparação das diferentes subpopulações quanto às respectivas distribuições médias (em relação às diversas condições de avaliação) de resposta, verificação da existência de efeito do fator ou fatores (tratamentos) que define(m) as subpopulações, verificar se os perfis são coincidentes;

iii. comparação das diferentes condições de avaliação quanto às respectivas distribuições médias (em relação às diversas subpopulações) de respostas, verificação da existência de efeito do fator que define as condições de avaliação, verificar se os perfis são horizontais;

iv. cjuste de modelos (geralmente na forma de curvas polinomiais) para explicar a variação das respostas médias (ou de outras características das distribuições de respostas) como função do fator que define as condições de avaliação.

2.3 Medidas Repetidas e Dados Longitudinais

Um tipo comum de medida repetida são os dados longitudinais, medidas repetidas 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.

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.

As técnicas clássicas de dados longitudinais são geralmente dirigidas para o caso de dados completos e balanceados em relação ao tempo. Na prática três tipos de análise são comumente usados para medidas repetidas:

(1) análise de variância univariada como se fosse um experimento em parcela subdividida (spit-plot) com tratamentos como o fator da parcela e o tempo como o fator da subparcela;

(2) análises univariada e multivariada para transformações lineares das medidas repetidas, tais como inclinações e outras tendências em curvas de regressão, diferenças entre respostas de diferentes pontos de tempo;

(3) métodos baseados em modelos mistos com estruturas paramétricas especiais nas matrizes de covariâncias.

A análise de medidas repetidas requer atenção especial na estrutura de covariância devido à natureza seqüencial dos dados em cada unidade experimental.

Dos três métodos considerados:

(1) a análise de variância univariada ignora a estrutura de covariância o que pode resultar em conclusões incorretas;

(2) análises univariada e multivariada para transformações lineares das medidas repetidas evitam a estrutura de covariância o que pode resultar em análises ineficientes, equivalente a desperdícios dos dados;

(3) metodologia de modelos mistos que permite eficientemente conduzir a questão diretamente na modelagem da estrutura de covariância.

A análise de medidas repetidas utilizando estas metodologias segue duas etapas:

1º) modela a estrutura de covariância;

2º) analisa as tendências de tempo para os tratamentos, estimando e comparando médias.

Em situações de medidas repetidas, em que se tem uma estrutura balanceada e completa, com a seguinte configuração:

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}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
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}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
2	n_2	$Y_{2n_2 1}$	$Y_{2n_2 2}$	...	$Y_{2n_2 k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
I	1	Y_{i11}	Y_{i12}	...	Y_{i1k}
I	2	Y_{i21}	Y_{i22}	...	Y_{i2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
I	n_i	$Y_{in_i 1}$	$Y_{in_i 2}$	...	$Y_{in_i k}$

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

3 ANÁLISE DE VARIÂNCIA MULTIVARIADA

Um dos métodos para analisar dados com medidas repetidas é a análise de variância multivariada.

A análise de variância 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. Esta análise de variância aplicada a todas as variáveis simultaneamente é denominada de Análise de Variância Multivariada.

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 tirar conclusões sobre possíveis diferenças nas médias dos grupos.

A análise de variância multivariada é indicada quando existe uma estreita correlação entre as variáveis, pois, em caso contrário, as análises separadas resolveriam o problema, isto é, a análise de variância multivariada considera simultaneamente todas as variáveis de interesse.

No caso univariado, uma “única medida” dependente é testada quanto à igualdade das médias através dos grupos, enquanto que na multivariada os vetores das médias são testados quanto a igualdade.

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 são generalizações dos pressupostos da análise de variância simples (ANOVA):

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

3.1 Verificação do Pressuposto de Normalidade

Para a aplicação da maioria dos testes estatísticos multivariados deve-se verificar se cada vetor de observações provém de uma população normal multivariada, isto é, detectar se os dados se desviam de modo significativo do comportamento esperado.

Segundo Reis (1997), pode-se afirmar, com base nas propriedades da distribuição normal, que qualquer combinação linear de variáveis normais é também normal e que todos os contornos da função distribuição normal multivariada formam uma elipse.

De acordo com a literatura, não existe um teste que permita testar se um conjunto de dados segue ou não uma distribuição normal conjunta para duas ou mais variáveis. Deste modo, o que mais se faz é testar a normalidade de cada variável separadamente;embora não se possa ter certeza de que as variáveis, mesmo sendo normais quando analisadas separadamente, mantenham essa normalidade em dimensões mais elevadas.

O pressuposto da normalidade pode ser verificado na prática se as seguintes condições estão presentes nos dados observados:

- as distribuições marginais para o conjunto de observações são normalmente distribuídas;
- as representações gráficas de pares de observações representam uma elipse;
- não existem observações muito diferentes (outliers), isto é, que sejam resultado apenas de erros de introdução de dados.

Uma boa alternativa, para verificar se determinada população segue distribuição normal, são as representações gráficas, em particular, o gráfico Q-Q que pode ser utilizado para avaliar a normalidade de determinada distribuição marginal.

Este tipo de gráfico representa os seguintes pares de valores: os quantis da distribuição amostral e os correspondentes valores esperados.

Para o caso multivariado, este gráfico tem por base o conceito de distância de Mahalanobis entre dois vetores de observações, que pode ser calculada da seguinte forma:

$$\underset{\sim}{d_j^2} = \left(\underset{\sim}{x}_j - \underset{\sim}{\overline{x}}\right)' S^{-1} \left(\underset{\sim}{x}_j - \underset{\sim}{\overline{x}}\right) \quad (1)$$

A construção do gráfico Q-Q segue os seguintes passos:

1) ordenação das distâncias $d^2_{(j)}$ por ordem crescente;
2) cálculo das probabilidades acumuladas $p_{(j)} = \frac{j - 0{,}5}{n_j}$ para cada um dos valores anteriores;
3) definição do percentil de ordem 100. $p_{(j)}$ na distribuição χ^2 com p graus de liberdade;
4) representação gráfica dos pares ordenados $\left(d^2_{(j)};\ \chi^2_{(p_{(j)})}\right)$.

Quanto mais a distribuição das observações se aproximar da linha reta, mais aproximadas estarão de uma distribuição normal multivariada. E em caso contrário, um afastamento sistemático do comportamento linear sugere a rejeição da hipótese de normalidade multivariada.

3.2 Verificação da Igualdade de Matrizes de Variância e Covariância

Para testar a igualdade de matrizes de variância e covariância o pressuposto de normalidade multivariada deve ser primeiramente verificado.

Box (1950 apud REIS,1997 p.186) desenvolveu um teste (teste M) que é uma generalização do teste univariado de igualdade de variâncias de Bartlett, e muito sensível a violações do pressuposto da normalidade. Esta sensibilidade atribui-se ao fato de que o teste utiliza as variâncias generalizadas, os determinantes das matrizes de covariância. E deste modo, a hipótese de igualdade de matrizes pode ser rejeitada apenas pela violação do pressuposto da normalidade e não por se tratarem de matrizes significativamente diferentes.

Box (1950 apud REIS,1997 p.186) definiu o teste M utilizando o método do quociente de verossimilhanças e pressupondo que os vetores de médias dos grupos são desconhecidos. As hipóteses a serem testadas são:

$$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_i \text{ com } \hat{\Sigma} = S \text{ e } \underset{\sim}{\hat{m}_j} = \underset{\sim}{\overline{X}_j}$$

$$H_1 : \Sigma_i \neq \Sigma_j \quad i \neq j \text{ com } \hat{S}_j = S_j \quad \text{e } \underset{\sim}{\hat{m}_j} = \underset{\sim}{\overline{X}_j}$$

Seja *n* a dimensão total da amostra, $v_i = n_i - 1$ os graus de liberdade associados a cada grupo, S_i a matriz de covariância do grupo *i* e *S* a matriz de covariância total. O teste M de Box é definido da seguinte maneira:

$$M = (n - I)\ln|S_c| - \sum_{I=1}^{i} v_i \ln|S_i| \qquad (2)$$

onde:

$$n = \sum_{I=1}^{i} n_i \text{ dimensão total da amostra;} \qquad (3)$$

$$S_c = \sum_{I=1}^{i} \frac{n_i S_i}{\sum ni} \text{ é a matriz de variância e covariância comum.} \qquad (4)$$

Segundo Reis (1997), Box sugeriu duas aproximações para o teste M. A primeira é uma aproximação pela distribuição c^2; dada por:

$$M.C \sim \chi^2_{\left[\frac{1}{2}K(K+1)(I-1)\right]} \qquad (5)$$

sendo $C = 1 - \frac{2K^2 + 3K - 1}{6(K+1)(I-1)}\left(\sum_{I=1}^{i} \frac{1}{v_i} - \frac{1}{n-I}\right)$ para grupos com tamanhos de amostras diferentes.

E , $C = 1 - \frac{(2K^2 + 3K - 1)(I+1)}{6(K+1)In}$ para grupos com tamanhos de amostras iguais.

Aconselha-se utilizar essa aproximação quando as dimensões dos grupos são superiores a 20 e o número de variáveis e de grupos inferior a 6.

A segunda maneira é uma aproximação pela distribuição F; designada por:

$$\frac{M\left(1 - a_1 - \frac{v}{v_0}\right)}{v} \sim F_{(v,v_0)} . \qquad (6)$$

Onde :

$$a_1 = 1 - C$$

$$a_2 = \frac{(k-1)(k+2)}{6(I-1)}\left[\sum_{I=1}^{i} \frac{1}{v_i^2} - \frac{1}{(n-I)^2}\right]$$

$$v = \frac{K(K+1)(I-1)}{2} \quad \text{e} \quad v_0 = \frac{v+2}{a_2 + a_1^2}$$

Aconselha-se utilizar essa aproximação em todas as outras situações.

3.3 Teste de Igualdade de Médias para *K* Grupos (Oneway MANOVA)

Na análise de variância multivariada (MANOVA) a hipótese nula a ser testada é a igualdade de médias para um conjunto de *p* variáveis simultaneamente.

As hipóteses a serem testadas são as seguintes :

$$H_0: \ \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \ldots = \underset{\sim}{\mu}_i \text{ com } \underset{\sim}{\mu}_i = \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \\ \mu_{ki} \end{bmatrix} \quad I = 1,2,\ldots,i$$ ⇒ Todos os grupos têm vetores de médias iguais;

$$H_1: \ \underset{\sim}{\mu}_1 \neq \underset{\sim}{\mu}_j \text{ com } i \neq j$$ ⇒ Pelo menos dois grupos têm vetores de médias diferentes.

Assim como no caso univariado pode-se fazer um quadro para análise de variância multivariada.

Tabela 2: Análise de variância multivariada

Causas de Variação	**Graus de liberdade**	**Soma de Quadrados e Produtos**
Tratamentos	I - 1	B
Resíduos	n - I	W
Total	n - 1	T = B+W

Onde:

$$\underset{\sim}{\overline{X}} = \frac{\sum_{I=1}^{i} \sum_{J=1}^{n_{ij}} \underset{\sim}{X}_{ij}}{n}$$ vetor de médias amostrais para todas as observações; (7)

$$\overline{X}_i = \frac{\sum_{J=1}^{n_{ij}} X_{ij}}{n_i}$$ vetor de médias amostrais para o grupo i; (8)

$$B = \sum_{I=1}^{i} n_i \left(\underset{\sim}{\overline{X}}_i - \underset{\sim}{\overline{X}} \right) \left(\underset{\sim}{\overline{X}}_i - \underset{\sim}{\overline{X}} \right)'$$ matriz das somas dos quadrados e produtos cruzados entre os grupos; (9)

$$W = \sum_{I=1}^{i} \sum_{j=1}^{n_i} \left(\overline{\underset{\sim}{X}}_{ij} - \overline{\underset{\sim}{X}}_{i} \right) \left(\overline{\underset{\sim}{X}}_{ij} - \overline{\underset{\sim}{X}}_{i} \right)'$$ matriz das somas dos quadrados e produtos cruzados dentro dos grupos. (10)

O teste de significância mais usado na análise de variância multivariada é o *Critério de Wilks*, indicado pela letra grega Λ (lambda maiúsculo), definido pelo quociente do determinante da matriz das Somas de Quadrados e Produtos de Resíduos (W) e o determinante da matriz das Somas de Quadrados e Produtos Total (T).

Logo,

$$\Lambda = \frac{|W|}{|B+W|} \tag{11}$$

Duas diferenças importantes se notam em relação à análise de variância tradicional:

- na análise tradicional, antes de aplicar o teste F, faz-se a divisão das somas de quadrados pelo número de graus de liberdade; na análise multivariada essa divisão não é feita;
- no teste F, a contribuição dos tratamentos vai no numerador, no teste Λ ela aparece no denominador.

Para avaliar a significância do valor de Λ obtido, pode-se usar tabelas especiais, bastante extensas, mas o mais comum é transformar o valor de Λ num valor correspondente de F, e usar as tabelas de F conhecidas.

Para isso utiliza-se a seguinte fórmula:

$$F_{crit} = \left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \right) \left(\frac{\sum n_i - I - 1}{I - 1} \right) \tag{12}$$

Utiliza-se a seguinte estatística tabelada :

$$F_{tab} = F_{\alpha, v_1, v_2}$$

onde:

$v_1 = 2(I-1)$ e $v_2 = 2\left(\sum n_i - I - 1\right)$ são os graus de liberdade para F e α o nível de significância do teste.

Obtido o quadro de análise de variância, é preciso se aplicar um teste de significância, semelhante ao teste F utilizado na análise tradicional.

Da mesma forma que na análise de variância tradicional aceita-se a hipótese nula se $F_{crit} < F_{tab}$, em caso contrário rejeita-se H_0.

4 ANÁLISE DE PERFIS DE MÉDIAS

No estudo de dados longitudinais um dos objetivos é fazer comparações entre os diversos componentes dos perfis médios de respostas. Para tanto, utiliza-se a análise de perfis de médias que é uma técnica bastante difundida na literatura estatística.

Sob o enfoque multivariado esta técnica é uma derivação da técnica de análise de variância multivariada (MANOVA), onde as hipóteses a serem testadas basicamente são aquelas mencionadas na seção 2.2 (i – iii).

4.1 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 Andrade e Singer (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 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 et al. (1960 apud CASTRO,1997 p.53), afirmam que os dois procedimentos, univariado e multivariado, são assintoticamente equivalentes para tamanhos de amostras encontrados na prática, devido às 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 em situações onde o número total de observações menos o número de tratamentos é maior ou igual ao número de dados longitudinais menos um, isto é, $n - I \geq K - 1$. 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.

4.2 Modelo Matemático para Análise Multivariada de Perfis

Como na análise de variância univariada, a análise multivariada de perfis pode ser representada por um modelo matemático. Este modelo pode ser usado para comparar a influência de diferentes tratamentos no comportamento de variáveis provenientes de estudos longitudinais.

Ao tratar-se de dados longitudinais deve-se considerar a existência de uma possível correlação entre as observações, devido ao fato de as medidas serem realizadas em uma mesma unidade experimental, e desta forma, a aplicação de modelos clássicos de análise de variância e regressão não é apropriada.

Segundo Andrade e Singer (1986), o modelo estatístico utilizado na análise multivariada de perfis talvez seja o mais simples dentre aqueles usualmente adotados para a análise de dados longitudinais. Neste modelo pode-se representar o conjunto de todas as observações na forma matricial:

$$\underset{(n\times k)}{Y} = \begin{bmatrix} y_{111} & y_{112} & \cdots\cdot & y_{11k} \\ y_{121} & y_{122} & \cdots & y_{21k} \\ \cdots & \cdots & \cdots & \cdots \\ y_{1n_i1} & y_{1n_12} & \cdots & y_{1n_1k} \\ \cdots & \cdots & \cdots & \cdots \\ y_{i11} & y_{i12} & \cdots & y_{i1k} \\ \cdots & \cdots & \cdots & \cdots \\ y_{in_i1} & y_{in_i2} & \cdots & y_{in_ik} \end{bmatrix} = \begin{bmatrix} y'_{11} \\ y'_{12} \\ \cdots \\ y'_{1n_1} \\ \cdots \\ y'_{i1} \\ \cdots \\ y'_{in_i} \end{bmatrix}$$

e o modelo linear para a análise multivariada de perfis, considerando um estudo balanceado e completo pode ser representado por:

$$\underset{(n\,x\,k)}{E(Y)} = \underset{(n\,x\,i)}{X}\ \underset{(i\,x\,k)}{\beta} \tag{13}$$

Em que:

$\underset{\sim}{Y} = \left(y_{11}, y_{12}, \ldots, y_{ni}\right)^T$ é uma matriz $n \times i$, sendo cada y_{ij} o perfil individual para a j-ésima unidade amostral da subpopulação i;

$\underset{\sim}{X} = I \otimes 1_i$, 1_i um vetor de uns com número de linhas igual ao número de unidades amostrais na subpopulação i;

$\underset{\sim}{\beta} = \left(\mu_1, \mu_2, \ldots, \mu_i\right)^T$ tem dimensão $i \times k$ e cada μ_i corresponde ao perfil médio da i-ésima subpopulação.

E a matriz $\underset{\sim}{X}$ é representada por:

$$\underset{(n\,x\,i)}{X} = \begin{bmatrix} 1_{n_1} & 0 & \cdots & 0 \\ 0 & 1_{n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1_{n_i} \end{bmatrix}$$

é uma matriz de especificação (planejamento) e β

$$\underset{(i\,x\,k)}{\beta} = \begin{bmatrix} \mu_{11} & & & \\ \mu_{21} & \mu_{22} & & \\ \vdots & \vdots & \cdots & \\ \mu_{i1} & \mu_{i2} & \cdots & \mu_{ik} \end{bmatrix} = \begin{bmatrix} \mu'_1 \\ \mu'_2 \\ \vdots \\ \mu'_i \end{bmatrix}$$

é uma matriz de parâmetros composta por vetores que definem a estrutura de variação do comportamento médio da resposta. Neste caso, a j-ésima linha da matriz de parâmetros β representa o vetor (perfil) médio de respostas para as unidades experimentais do i-ésimo tratamento (I=1, 2,..., i).

Para efeitos de inferência supõe-se que os perfis de resposta y'_{ij} sigam distribuições normais k-variadas e que as matrizes de variâncias e covariâncias correspondentes são todas iguais e seguem a forma geral.

$$\sum = \begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{12} & \sigma_2^2 & & \\ \vdots & \vdots & \ldots & \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_k^2 \end{bmatrix}$$

onde σ_{ij} é a covariância entre as observações medidas em condições de avaliação distintas e σ_i^2 é a variância das observações medidas na mesma condição de avaliação.

A fim de facilitar a interpretação da análise baseada no modelo (13) pode-se considerar a Figura 1 que corresponde à representação gráfica dos perfis médios observados de resposta em um caso particular com I = 4 tratamentos e K = 3 condições de avaliação.

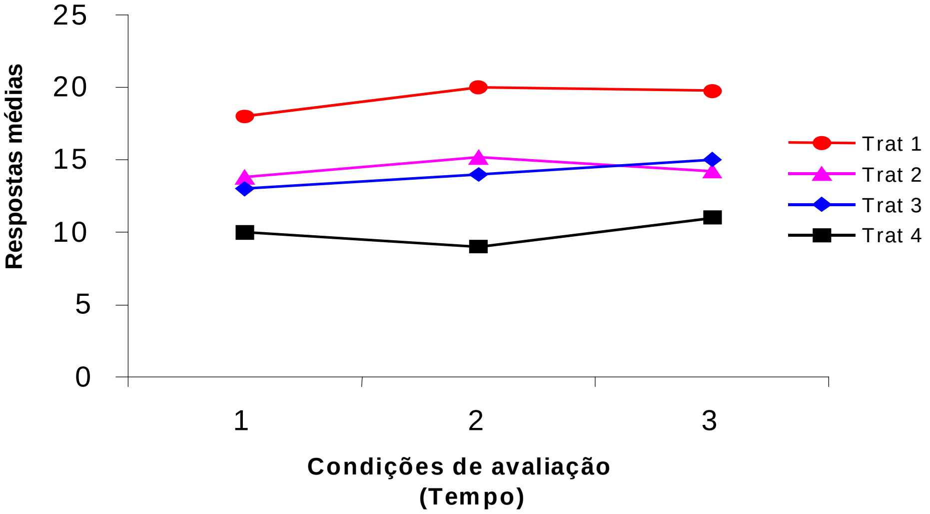

Figura 1: Representação gráfica genérica dos perfis médios de resposta.

Relativamente à configuração da Figura 1, os objetivos da análise de perfis de médias foram apresentados na seção 2.2 (i - iii) e podem ser traduzidos através das seguintes 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.

Em termos dos parâmetros do modelo (13), as hipóteses podem ser expressas por:

1) Perfis Paralelos.

$$H_0^{(1)}: \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(k-1)} - \mu_{1k} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(k-1)} - \mu_{2k} \end{bmatrix} = \mathbf{L} = \begin{bmatrix} \mu_{i1} - \mu_{i2} \\ \mu_{i2} - \mu_{i3} \\ \vdots \\ \mu_{i(k-1)} - \mu_{ik} \end{bmatrix}$$

Na ausência de interação, quando não se rejeita $H_0^{(1)}$, define-se as seguintes hipóteses:

2) Perfis Coincidentes

$$H_0^{(2)}: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1k} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2k} \end{bmatrix} = \mathbf{L} = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ik} \end{bmatrix}$$

3) Perfis Horizontais

$$H_0^{(3)}: \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{i1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{i2} \end{bmatrix} = \mathbf{L} = \begin{bmatrix} \mu_{1k} \\ \mu_{1k} \\ \vdots \\ \mu_{ik} \end{bmatrix}$$

Sob a abordagem multivariada, as hipóteses já apresentadas podem ser expressas na forma de hipótese linear geral:

$$H : D\beta U = 0 \tag{14}$$

em que $D_{d\times i}$ e $U_{u\times i}$ são matrizes conhecidas em que os postos são respectivamente d e u. A matriz D é responsável pelas comparações das respostas entre as unidades amostrais, isto é, relacionada às comparações entre as respostas médias das diferentes subpopulações e a matriz U está relacionada às comparações intra-unidades amostrais, ou seja, ao longo das diferentes condições de avaliação. Deste modo, podemos reescrever as hipóteses $H_0^{(1)}$, $H_0^{(2)}$ e $H_0^{(3)}$ na forma linear geral. Essas matrizes não são determinadas de modo único. Para as hipóteses de interesse, pode-se defini-las como:

$$H_0^{(1)} : D^1_{(i-1)\times i} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

e

$$U^1_{k\times(k-1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$H_0^{(2)} : D = D^1, \quad U = 1_k \qquad \text{e} \qquad H_0^{*(2)} : D = D^1, \quad U = I_k$$

$$H_0^{(3)} : D = 1_i^T, \quad U = U^1 \qquad \text{e} \qquad H_0^{*(3)} : D = I_i, \quad U = U^1$$

Pode-se obter estatísticas de testes para a hipótese linear geral a partir de vários critérios. Estas estatísticas geralmente são funções das raízes características da matriz BW^{-1}, onde $B = U'\hat{\beta}D'\left[D(X'X)^{-1}D'\right]^{-1}D\hat{\beta}U$, com $\hat{\beta} = (X'X)^{-1}X'Y$, é a matriz de somas de quadrados e produtos cruzados devido à hipótese nula, e $W = U'Y'\left[I - X(X'X)^{-1}X'\right]YU$ é a matriz da soma de quadrados e produtos cruzados devido ao erro. O estimador $\hat{\beta}$ é o estimador de mínimos quadrados ou de máxima verossimilhança de β.

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 apud MORRISON,1990 p.194) investigaram a análise de perfil de variância debaixo da suposição de distribuição multinormal dos resíduos com uma matriz de covariância geral e propuseram uma análise aproximada de variância e testes "conservadores " para as diferenças de resposta e a hipótese de paralelismo cujo verdadeiro nível α não pode exceder alguns valores especificados.

Tabela 3 - 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 – I
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			

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:

$$SQ_1 = \frac{1}{n}\sum_{K=1}^{k} G_k^2 - \frac{G^2}{nk} \tag{15}$$

$$SQ_2 = \frac{1}{K}\sum_{I=1}^{i} \frac{1}{n_i} C_i^2 - \frac{G^2}{nk} \tag{16}$$

$$SQ_3 = \frac{1}{k}\sum_{I=1}^{i}\sum_{j=1}^{n_j} R_{ij}^2 - \frac{1}{k}\sum_{I=1}^{i}\frac{1}{n_i} C_i^2 \tag{17}$$

$$SQ_4 = \sum_{I=1}^{i}\sum_{K=1}^{k}\frac{1}{n_i} T_{ik}^2 - \frac{G^2}{nk} - S_1 - S_2 \tag{18}$$

$$SQ_5 = S_6 - S_1 - S_2 - S_3 - S_4 \tag{19}$$

$$SQ_6 = \sum_{K=1}^{k}\sum_{I=1}^{i}\sum_{j=1}^{n_j} x_{ijk}^2 - \frac{G^2}{nk} \tag{20}$$

Greenhouse e Geisser (1959 apud MORRISON,1990 p.194) 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 $\varepsilon = \frac{K^2(\overline{\sigma}_d - \overline{\sigma}_{..})^2}{(K-1)\left(\sum\sum \sigma_{rs}^2 - 2K\sum \overline{\sigma}_{r.}^2 + K^2\overline{\sigma}_{..}^2\right)}$, sendo calculado para os elementos σ_{rs} da matriz de variância e covariância populacional onde $\overline{s}_d$ é a média dos k elementos da diagonal principal, $\overline{s}_{..}$ é a média de todas as variâncias e covariâncias, e $\overline{s}_{r.}$ 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 apud MORRISON,1990 p.194) 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 – l, l – 1 e n – l, respectivamente. Devido a isso estes testes são chamados de testes conservativos, porque eles são baseados na máxima redução dos graus de liberdade.

4.3 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 através dos princípios da União-intersecção de Roy e da Razão de verossimilhança de Wilks. Ainda outros dois testes importantes, devido sua aplicação, são o Critério de Hotelling - Lawley e o Critério de Pillai.

4.3.1 União-intersecção de Roy

O critério de Roy, baseado na metodologia da união-intersecção, propõe como valor crítico a maior raiz característica de θ_i, onde $\theta_i = \lambda_i(1+\lambda_i)^{-1}$ e λ_i é a i-ésima raiz característica de BW^{-1},

$$\theta_s = \max(\theta_i)$$

Uma aproximação para a distribuição dessa estatística é dada por:

$$F = \frac{\theta(r - v + q)}{v}$$

onde $v = \max(k,q)$, k é o número de condições de avaliação, q é o número de comparações independentes e r é o número de graus de liberdade do erro.

Neste caso, tem-se v graus de liberdade para o numerador e $r - v + q$ graus de liberdade para o denominador.

4.3.2 Critério de Wilks

O critério de Wilks, obtido através da razão de verossimilhança, é definido como:

$$\Lambda = \prod_{i=1}^{s} (1 - \theta_i)$$

onde $s = \min(u, d)$ ou s é o número de raízes características não-nulas.

A distribuição Λ de Wilks não se encontra tabelada mas pode ser aproximada à distribuição F, ou à distribuição χ^2. Duas das possíveis aproximações são:

$$\frac{(n - I) - K + 1}{K} . \frac{1 - \Lambda}{\Lambda} \sim F_{(K, n-I-K+1)}$$

ou

$$-\left[(n-1) - \frac{1}{2}(K + I)\right] \ln \Lambda \sim \chi^2_{(K\times(I-1))}$$

onde p é o número de variáveis, n é o número de indivíduos e k o número de grupos.

Podem também ser utilizadas, em determinadas situações, as seguintes distribuições exatas:

Tabela 4 - Distribuição da estatística Λ.

Nº de variáveis	**Nº de grupos**	**Distribuição amostral para dados normais multivariados**
$K = 1$	$I \geq 2$	$\left[\frac{\sum n_i - I}{I - 1}\right]\left[\frac{1 - \Lambda}{\Lambda}\right] \sim F_{(I-1, \sum n_i - I)}$
$K = 2$	$I \geq 2$	$\left[\frac{\sum n_i - I - 1}{I - 1}\right]\left[\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right] \sim F_{(2(I-1), 2(\sum n_i - I - 1))}$
$K \geq 1$	$I = 2$	$\left[\frac{\sum n_i - K - 1}{K}\right]\left[\frac{1 - \Lambda}{\Lambda}\right] \sim F_{(K, \sum n_i - K - 1)}$
$K \geq 1$	$I = 3$	$\left[\frac{\sum n_i - K - 2}{K}\right]\left[\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right] \sim F_{(2K, 2(\sum n_i - K - 2))}$

4.3.3 Critério de Pillai

A estatística de Pillai pode ser definida como

$$P = \sum_{i=1}^{s} \theta_i$$

Devido à complexidade da distribuição exata de P, utiliza-se sua aproximação assintótica dada por:

$$\frac{2n+s+1}{2m+s+1} \cdot \frac{P}{P-v} \sim F_{[s(2m+s+1);s(2n+s+1)]}$$

onde $s = \min(u,d)$ ou s é o número de raízes características não-nulas; e $m = \frac{|I-1-K|-1}{2}$ e $n = \frac{n-I-K-1}{2}$

4.3.4 Critério de Hotelling-Lawley

A estatística de teste de Hotelling-Lawley também é conhecida como Traço de Hotelling-Lawley, é dada por

$$T = \sum_{i=1}^{s} \theta_i (1-\theta_i)^{-1}$$

Uma aproximação para a distribuição dessa estatística é definida por:

$$\frac{2(sn'+1)}{s(2m'+s+1)} T \cap F_{[s(2m'+s+1),2(sn'+1)]}$$

onde $s = \min(u,d)$ ou s é o número de raízes características não-nulas; $m' = \frac{1}{2}[|u-d|-1]$ e $n' = \frac{1}{2}(n-I-u-1)$.

4.4 Algumas Considerações sobre os Critérios de Teste para a Análise Multivariada de Perfis

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 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.

5 APLICAÇÃO DOS MÉTODOS

Suponha que três escalas A, B e C de um inventário que mede determinadas atitudes maternas que foram administradas às mães que participam em um estudo do desenvolvimento da criança. Como uma parte da investigação cada mãe tinha sido atribuída a uma de quatro classes sócio-econômicas. Os dados para esse exemplo hipotético, retirado de Morrison (1990), podem ser observados na Tabela 5.

Tabela 5 – 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. Para tanto utilizou-se a equação (i) em 2.1.2.

$$\overline{\underset{\sim}{X}} = \begin{bmatrix} \frac{305}{21} \\ \\ \frac{328}{21} \\ \\ \frac{333}{21} \end{bmatrix} \Rightarrow \overline{\underset{\sim}{X}} = \begin{bmatrix} 14{,}52 \\ \\ 15{,}62 \\ \\ 15{,}86 \end{bmatrix}$$

Depois determinou-se os vetores de médias amostrais para cada escala (subpopulação) dentro das classes (condições de avaliação). Equação também encontrada em 2.1.2 (i).

$$\overline{\underset{\sim}{X}}_1 = \begin{bmatrix} \frac{144}{8} \\ \frac{160}{8} \\ \frac{158}{8} \end{bmatrix} \Rightarrow \overline{\underset{\sim}{X}}_1 = \begin{bmatrix} 18 \\ 20 \\ 19{,}75 \end{bmatrix} \qquad \overline{\underset{\sim}{X}}_2 = \begin{bmatrix} \frac{69}{5} \\ \frac{76}{5} \\ \frac{71}{5} \end{bmatrix} \Rightarrow \overline{\underset{\sim}{X}}_2 = \begin{bmatrix} 13{,}8 \\ 15{,}2 \\ 14{,}2 \end{bmatrix}$$

$$\overline{\underset{\sim}{X}}_3 = \begin{bmatrix} \frac{52}{4} \\ \frac{56}{4} \\ \frac{60}{4} \end{bmatrix} \Rightarrow \overline{\underset{\sim}{X}}_3 = \begin{bmatrix} 13 \\ 14 \\ 15 \end{bmatrix} \qquad \overline{\underset{\sim}{X}}_4 = \begin{bmatrix} \frac{40}{4} \\ \frac{36}{4} \\ \frac{44}{4} \end{bmatrix} \Rightarrow \overline{\underset{\sim}{X}}_4 = \begin{bmatrix} 10 \\ 9 \\ 11 \end{bmatrix}$$

Logo após determinou-se as matrizes de variância e covariância para as escalas. Usou-se a Equação (ii) de 2.1.2.

$$S_1 = \begin{pmatrix} 19 - 14{,}52 \\ 20 - 15{,}62 \\ 18 - 15{,}86 \end{pmatrix} \begin{pmatrix} 19 - 14{,}52 & 20 - 15{,}62 & 18 - 15{,}86 \end{pmatrix} +$$

$$+ \begin{pmatrix} 20 - 14{,}52 \\ 21 - 15{,}62 \\ 19 - 15{,}86 \end{pmatrix} \begin{pmatrix} 20 - 14{,}52 & 21 - 15{,}62 & 19 - 15{,}86 \end{pmatrix} + \ . \quad . \quad . \ +$$

$$+ \begin{pmatrix} 15 - 14{,}52 \\ 19 - 15{,}62 \\ 19 - 15{,}86 \end{pmatrix} \begin{pmatrix} 15 - 14{,}52 & 19 - 15{,}62 & 19 - 15{,}86 \end{pmatrix} = \begin{pmatrix} 120{,}8832 & 128{,}9392 & 110{,}2976 \\ 128{,}9392 & 169{,}4752 & 137{,}3056 \\ 110{,}2976 & 137{,}3056 & 132{,}5568 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 12 - 14{,}52 \\ 14 - 15{,}62 \\ 12 - 15{,}86 \end{pmatrix} \begin{pmatrix} 12 - 14{,}52 & 14 - 15{,}62 & 12 - 15{,}86 \end{pmatrix} +$$

$$+ \begin{pmatrix} 15 - 14{,}52 \\ 15 - 15{,}62 \\ 17 - 15{,}86 \end{pmatrix} \begin{pmatrix} 15 - 14{,}52 & 15 - 15{,}62 & 17 - 15{,}86 \end{pmatrix} + \ . \quad . \quad . \ +$$

$$+ \begin{pmatrix} 14 - 14{,}52 \\ 16 - 15{,}62 \\ 13 - 15{,}86 \end{pmatrix} \begin{pmatrix} 14 - 14{,}52 & 16 - 15{,}62 & 13 - 15{,}86 \end{pmatrix} = \begin{pmatrix} 9{,}392 & 6{,}712 & 14{,}176 \\ 6{,}712 & 7{,}682 & 6{,}286 \\ 14{,}176 & 6{,}286 & 28{,}578 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} 15 - 14{,}52 \\ 14 - 15{,}62 \\ 17 - 15{,}86 \end{pmatrix} \begin{pmatrix} 15 - 14{,}52 & 14 - 15{,}62 & 17 - 15{,}86 \end{pmatrix} +$$

$$+ \begin{pmatrix} 13 - 14{,}52 \\ 14 - 15{,}62 \\ 15 - 15{,}86 \end{pmatrix} \begin{pmatrix} 13 - 14{,}52 & 14 - 15{,}62 & 15 - 15{,}86 \end{pmatrix} + \ . \quad . \quad . \ +$$

$$+ \begin{pmatrix} 12 - 14{,}52 \\ 13 - 15{,}62 \\ 13 - 15{,}86 \end{pmatrix} \begin{pmatrix} 12 - 14{,}52 & 13 - 15{,}62 & 13 - 15{,}86 \end{pmatrix} = \begin{pmatrix} 15{,}2416 & 9{,}8496 & 11{,}2288 \\ 9{,}8496 & 12{,}4976 & 7{,}5728 \\ 11{,}2288 & 7{,}5728 & 10{,}9584 \end{pmatrix}$$

$$S_4 = \begin{pmatrix} 8-14{,}52 \\ 9-15{,}62 \\ 10-15{,}86 \end{pmatrix} \begin{pmatrix} 8-14{,}52 & 9-15{,}62 & 10-15{,}86 \end{pmatrix} +$$

$$+ \begin{pmatrix} 10-14{,}52 \\ 10-15{,}62 \\ 12-15{,}86 \end{pmatrix} \begin{pmatrix} 10-14{,}52 & 10-15{,}62 & 12-15{,}86 \end{pmatrix} + \quad . \quad . \quad . \quad +$$

$$+ \begin{pmatrix} 11-14{,}52 \\ 7-15{,}62 \\ 12-15{,}86 \end{pmatrix} \begin{pmatrix} 11-14{,}52 & 7-15{,}62 & 12-15{,}86 \end{pmatrix} = \begin{pmatrix} 87{,}7216 & 118{,}6896 & 89{,}8688 \\ 118{,}6896 & 181{,}2976 & 126{,}6928 \\ 89{,}8688 & 126{,}6928 & 98{,}4784 \end{pmatrix}$$

A partir das matrizes de variância e covariância de cada classe, determinou-se a matriz de variância e covariância amostral, cuja fórmula também está referida em (ii) de 2.1.2.

$$S = \frac{1}{20} \cdot (S_1 + S_2 + S_3 + S_4)$$

$$S = \frac{1}{20} \cdot \left[\begin{pmatrix} 120{,}8832 & 128{,}9392 & 110{,}2976 \\ 128{,}9392 & 169{,}4752 & 137{,}3056 \\ 110{,}2976 & 137{,}3056 & 132{,}5568 \end{pmatrix} + \begin{pmatrix} 9{,}392 & 6{,}712 & 14{,}176 \\ 6{,}712 & 7{,}682 & 6{,}286 \\ 14{,}176 & 6{,}286 & 28{,}578 \end{pmatrix} + \begin{pmatrix} 15{,}2416 & 9{,}8496 & 11{,}2288 \\ 9{,}8496 & 12{,}4976 & 7{,}5728 \\ 11{,}2288 & 7{,}5728 & 10{,}9584 \end{pmatrix} + \begin{pmatrix} 87{,}7216 & 118{,}6896 & 89{,}8688 \\ 118{,}6896 & 181{,}2976 & 126{,}6928 \\ 89{,}8688 & 126{,}6928 & 98{,}4784 \end{pmatrix} \right]$$

$$S = \frac{1}{20} \cdot \begin{pmatrix} 233{,}2384 & 264{,}1904 & 225{,}5712 \\ 264{,}1904 & 370{,}9524 & 277{,}8572 \\ 225{,}5712 & 277{,}8572 & 270{,}5716 \end{pmatrix} = \begin{pmatrix} 11{,}6619 & 13{,}2095 & 11{,}2786 \\ 13{,}2095 & 18{,}5476 & 13{,}8928 \\ 11{,}2786 & 13{,}8928 & 13{,}5286 \end{pmatrix}$$

O passo seguinte é a verificação do pressuposto de normalidade multivariada, que para sua verificação utilizou-se o gráfico Q-Q multivariado.

Primeiramente calculou-se as distâncias de Mahalanobis entre dois vetores de observações, equação (1), e em seguida ordenou-se essas distâncias por ordem crescente, e então, calculou-se as probabilidades acumuladas $p_{(j)} = \frac{j-0{,}5}{n_j}$ para cada um dos valores anteriores e desta forma encontrou-se cada percentil na distribuição χ^2;

Tabela 6 – Valores encontrados para os pares ordenados.

j	$d^2_{(j)}$	$p_{(j)}$	$p_{(j)}.100$	χ^2
1	0,295946	0,02381	2,380952	0,208592
2	0,569136	0,071429	7,142857	0,455415
3	0,731672	0,119048	11,90476	0,666934
4	1,135382	0,166667	16,66667	0,867207
5	1,468781	0,214286	21,42857	1,064234
6	2,17391	0,261905	26,19048	1,262337
7	2,745661	0,309524	30,95238	1,464549
8	3,056716	0,357143	35,71429	1,673429
9	3,306616	0,404762	40,47619	1,891456
10	3,922512	0,452381	45,2381	2,121283
11	4,452296	0,5	50	2,365974
12	5,365756	0,547619	54,7619	2,629272
13	5,38434	0,595238	59,52381	2,915978
14	6,468561	0,642857	64,28571	3,232539
15	6,744591	0,690476	69,04762	3,588024
16	7,124279	0,738095	73,80952	3,995923
17	8,077853	0,785714	78,57143	4,477726
18	8,925801	0,833333	83,33333	5,071066
19	11,29971	0,880952	88,09524	5,851795
20	11,81977	0,928571	92,85714	7,014767
21	14,28664	0,97619	97,61905	9,455499

E finalmente representou-se graficamente os pares ordenados.

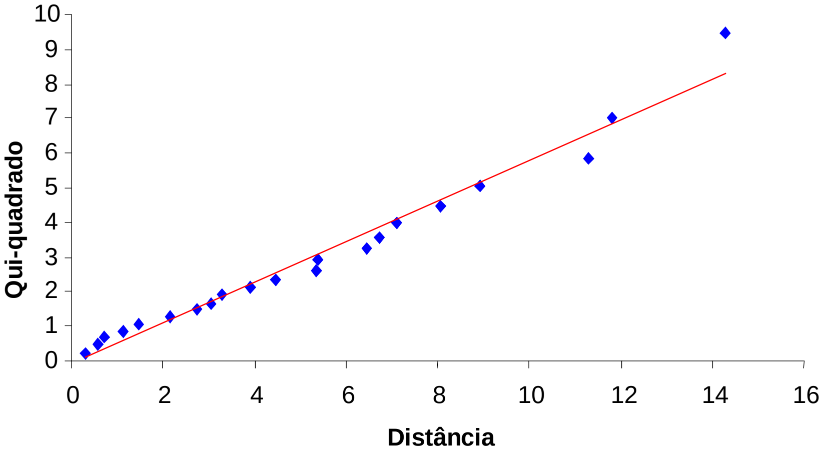

Figura 2 – Gráfico Q-Q.

Conforme a Figura 2 verifica-se que os dados seguem uma distribuição normal multivariada, pois a distribuição das observações se aproxima da linha reta, indicando que estão aproximadas de uma distribuição normal multivariada.

O passo seguinte é a verificação da igualdade de matrizes de variância e covariância.

Para verificar este pressuposto devem ser testadas as seguintes hipóteses:

$$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_I \text{ com } \hat{\Sigma} = \frac{T}{n-1} = S \text{ e } \hat{\underset{\sim}{\mu}}_i = \overline{\underset{\sim}{X}}_i$$

$$H_1 : \Sigma_i \neq \Sigma_j \quad i \neq j \text{ com } \hat{S}_j = S_j \text{ e } \hat{\underset{\sim}{\mu}}_i = \overline{\underset{\sim}{X}}_i$$

Para a realização do teste M, utilizou-se a equação (2).

Determinou-se a matriz de variância e covariância comum, conforme equação (4):

$$S_c = \frac{n_1.S_1}{n} + \frac{n_2.S_2}{n} + \frac{n_3.S_3}{n} + .\frac{n_4.S_4}{n} = \frac{1}{n}.(n_1S_1 + n_2S_2 + n_3S_3 + n_4S_4)$$

$$S_c = \frac{1}{21}\left[8.\begin{pmatrix} 120,8832 & 128,9392 & 110,2976 \\ 128,9392 & 169,4752 & 137,3056 \\ 110,2976 & 137,3056 & 132,5568 \end{pmatrix} + 5.\begin{pmatrix} 9,392 & 6,712 & 14,176 \\ 6,712 & 7,682 & 6,286 \\ 14,176 & 6,286 & 28,578 \end{pmatrix} + \right.$$

$$\left. + 4\begin{pmatrix} 15,2416 & 9,8496 & 11,2288 \\ 9,8496 & 12,4976 & 7,5728 \\ 11,2288 & 7,5728 & 10,9584 \end{pmatrix} + 4\begin{pmatrix} 87,7216 & 118,6896 & 89,8688 \\ 118,6896 & 181,2976 & 126,6928 \\ 89,8688 & 126,6928 & 98,4784 \end{pmatrix}\right]$$

$$S_c = \begin{pmatrix} 54,0932 & 54,4700 & 49,6710 \\ 54,4700 & 71,1520 & 56,6884 \\ 49,6710 & 56,6884 & 61,4767 \end{pmatrix}$$

Encontrada essa matriz, calculou-se seu determinante e o logarítimo natural deste determinante:

$$|S_c| = 11584,4$$

$$\ln|S_c| = 9,357415$$

Após determinou-se a quantidade:

$$\sum_{I=1}^{i} v_i \ln|S_i| = v_1 \ln|S_1| + v_2 \ln|S_2| + v_3 \ln|S_3| + v_4 \ln|S_4|$$

$$\sum_{I=1}^{i} v_i \ln|S_i| = 7\ln|S_1| + 4\ln|S_2| + 3\ln|S_3| + 3\ln|S_4| = 138,7949$$

E, por fim, calculou-se o valor de M, necessário para se efetuar o teste:

$$M = (n-k)\ln|S_c| - \sum_{j=1}^{k} v_j \ln|S_j| = 17.9,357415 - 138,7949 = 159,0761 - 138,7949 = 20,2812$$

Com este valor fez-se uma aproximação à distribuição F, conforme equação (6):

Onde :

$$n = 21 ; K = 3 \text{ e } I = 4$$

$$C = 1 - \frac{2K^2 + 3K - 1}{6(K+1)(I-1)}\left(\sum_{I=1}^{i} \frac{1}{v_i} - \frac{1}{n-I}\right) = 0,638636$$

$$a_1 = 1 - C = 1 - 0,638636 = 0,361364$$

$$a_2 = \frac{(K-1)(K+2)}{6(I-1)}\left[\sum_{I=1}^{i}\frac{1}{v_i^2} - \frac{1}{(n-I)^2}\right] = 0{,}125696$$

$$v = \frac{K(K+1)(I-1)}{2} = 18 \quad \text{e} \quad v_0 = \frac{v+2}{a_2 + a_1^2} = 78{,}03$$

Logo,

$$F_c = \frac{M\left(1 - a_1 - \frac{v}{v_0}\right)}{v} = \frac{20{,}2812.\left(1 - 0{,}3614 - \frac{18}{78{,}03}\right)}{18} = 0{,}417$$

e $F_{(v,v_0)} = F_{(18,148)} = 1{,}57$

Pelo teste M de Box, pode-se concluir que as matrizes de variância e covariância para os k grupos não diferem significativamente (são iguais), ao nível de 0,05 de significância.

Como as pressuposições à ANOVA foram satisfeitas, realizou-se o teste de igualdade de médias para *K* grupos (Oneway MANOVA).

Onde as hipóteses a serem testadas são:

$H_0:\ \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \ldots = \underset{\sim}{\mu}_i$ com $\underset{\sim}{\mu}_i = \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \\ \mu_{ki} \end{bmatrix}$ $\quad I = 1,2,\ldots,i \Rightarrow$ Todos os grupos têm vetores de médias iguais;

$H_1:\ \underset{\sim}{\mu}_i \neq \underset{\sim}{\mu}_j$ com $i \neq j \Rightarrow$ Pelo menos dois grupos têm vetores de médias diferentes.

A análise de variância multivariada foi realizada conforme a Tabela 2.

Primeiramente realizou-se o cálculo para determinar a matriz W, equação (10):

$$W_1 = \begin{pmatrix} 19-18 \\ 20-20 \\ 18-19,75 \end{pmatrix}\begin{pmatrix} 19-18 & 20-20 & 18-19,75 \end{pmatrix} +$$

$$+ \begin{pmatrix} 20-18 \\ 21-20 \\ 19-19,75 \end{pmatrix}\begin{pmatrix} 20-18 & 21-20 & 19-19,75 \end{pmatrix} + \; . \; . \; . \; +$$

$$+ \begin{pmatrix} 15-18 \\ 19-20 \\ 19-19,75 \end{pmatrix}\begin{pmatrix} 15-18 & 19-20 & 19-19,75 \end{pmatrix} = \begin{pmatrix} 24 & 7 & 2 \\ 7 & 16 & 1 \\ 2 & 1 & 11,5 \end{pmatrix}$$

$$W_2 = \begin{pmatrix} 12-13,8 \\ 14-15,2 \\ 12-14,2 \end{pmatrix}\begin{pmatrix} 12-13,8 & 14-15,2 & 12-14,2 \end{pmatrix} +$$

$$+ \begin{pmatrix} 15-13,8 \\ 15-15,2 \\ 17-14,2 \end{pmatrix}\begin{pmatrix} 15-13,8 & 15-15,2 & 17-14,2 \end{pmatrix} + \; . \; . \; . \; +$$

$$+ \begin{pmatrix} 14-13,8 \\ 16-15,2 \\ 13-14,2 \end{pmatrix}\begin{pmatrix} 14-13,8 & 16-15,2 & 13-14,2 \end{pmatrix} = \begin{pmatrix} 6,8 & 5,2 & 8,2 \\ 5,2 & 6,8 & 2,8 \\ 8,2 & 2,8 & 14,8 \end{pmatrix}$$

$$W_3 = \begin{pmatrix} 15-13 \\ 14-14 \\ 17-15 \end{pmatrix}\begin{pmatrix} 15-13 & 14-14 & 17-15 \end{pmatrix} +$$

$$+ \begin{pmatrix} 13-13 \\ 14-14 \\ 15-15 \end{pmatrix}\begin{pmatrix} 13-13 & 14-14 & 15-15 \end{pmatrix} + \; . \; . \; . \; +$$

$$+ \begin{pmatrix} 12-13 \\ 13-14 \\ 13-15 \end{pmatrix}\begin{pmatrix} 12-13 & 13-14 & 13-15 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 6 \\ 0 & 2 & 2 \\ 6 & 2 & 8 \end{pmatrix}$$

$$W_4 = \begin{pmatrix} 8 - 10 \\ 9 - 9 \\ 10 - 11 \end{pmatrix} \begin{pmatrix} 8 - 10 & 9 - 9 & 10 - 11 \end{pmatrix} +$$

$$+ \begin{pmatrix} 10 - 10 \\ 10 - 9 \\ 12 - 11 \end{pmatrix} \begin{pmatrix} 10 - 10 & 10 - 9 & 12 - 11 \end{pmatrix} + \quad . \quad . \quad . \quad +$$

$$+ \begin{pmatrix} 11 - 10 \\ 7 - 9 \\ 12 - 11 \end{pmatrix} \begin{pmatrix} 11 - 10 & 7 - 9 & 12 - 11 \end{pmatrix} = \begin{pmatrix} 6 & -1 & 2 \\ -1 & 6 & -2 \\ 2 & -2 & 4 \end{pmatrix}$$

$$W = \sum_{I=1}^{i} \sum_{J=1}^{n_i} \left(\underset{\sim}{\overline{X}_{ij}} - \underset{\sim}{\overline{X}_{i}} \right) \left(\underset{\sim}{\overline{X}_{ij}} - \underset{\sim}{\overline{X}_{i}} \right)' = W_1 + W_2 + W_3 + W_4 = \begin{pmatrix} 42{,}8 & 11{,}2 & 18{,}2 \\ 11{,}2 & 30{,}8 & 3{,}8 \\ 18{,}2 & 3{,}8 & 38{,}3 \end{pmatrix} \qquad (21)$$

Em seguida determinou-se a matriz B, conforme equação (9):

$$B_1 = 8.\left[\begin{pmatrix} 18 - 14{,}52 \\ 20 - 15{,}62 \\ 19{,}75 - 15{,}86 \end{pmatrix} \begin{pmatrix} 18 - 14{,}52 & 20 - 15{,}62 & 19{,}75 - 15{,}86 \end{pmatrix} \right] = \begin{pmatrix} 96{,}8832 & 121{,}9392 & 108{,}2976 \\ 121{,}9392 & 153{,}4752 & 136{,}3056 \\ 108{,}2976 & 136{,}3056 & 121{,}0568 \end{pmatrix}$$

$$B_2 = 5.\left[\begin{pmatrix} 13{,}8 - 14{,}52 \\ 15{,}2 - 15{,}62 \\ 14{,}2 - 15{,}86 \end{pmatrix} \begin{pmatrix} 13{,}8 - 14{,}52 & 15{,}2 - 15{,}62 & 14{,}2 - 15{,}86 \end{pmatrix} \right] = \begin{pmatrix} 2{,}592 & 1{,}512 & 5{,}976 \\ 1{,}512 & 0{,}882 & 3{,}486 \\ 5{,}976 & 3{,}486 & 13{,}778 \end{pmatrix}$$

$$B_3 = 4.\left[\begin{pmatrix} 13 - 14{,}52 \\ 14 - 15{,}62 \\ 15 - 15{,}86 \end{pmatrix} \begin{pmatrix} 13 - 14{,}52 & 14 - 15{,}62 & 15 - 15{,}86 \end{pmatrix} \right] = \begin{pmatrix} 9{,}2416 & 9{,}8496 & 5{,}2288 \\ 9{,}8496 & 10{,}4976 & 5{,}5728 \\ 5{,}2288 & 5{,}5728 & 2{,}9584 \end{pmatrix}$$

$$B_4 = 4.\left[\begin{pmatrix} 10 - 14{,}52 \\ 9 - 15{,}62 \\ 11 - 15{,}86 \end{pmatrix} \begin{pmatrix} 10 - 14{,}52 & 9 - 15{,}62 & 11 - 15{,}86 \end{pmatrix} \right] = \begin{pmatrix} 81{,}7216 & 119{,}6896 & 87{,}8688 \\ 119{,}6896 & 175{,}2976 & 128{,}6928 \\ 87{,}8688 & 128{,}6928 & 94{,}4784 \end{pmatrix}$$

$$B = B_1 + B_2 + B_3 + B_4 = \begin{pmatrix} 190{,}4384 & 252{,}9904 & 207{,}3712 \\ 252{,}9904 & 340{,}1524 & 274{,}0572 \\ 207{,}3712 & 274{,}0572 & 232{,}2716 \end{pmatrix} \qquad (22)$$

E também determinou-se a matriz T:

$$T = W + B$$

$$T = \begin{pmatrix} 233{,}2384 & 264{,}1904 & 225{,}5712 \\ 264{,}1904 & 370{,}9524 & 277{,}8572 \\ 225{,}5712 & 277{,}8572 & 270{,}5716 \end{pmatrix}$$

Continuando a análise dos dados, foi realizada uma análise multivariada de perfis de médias.

As hipóteses a serem testadas são as seguintes:

$H_0^{(1)}$: os perfis médios de resposta correspondentes as diversas escalas são paralelos, não existe interação escalas x classe social;

$H_0^{(2)}$: os perfis médios de resposta correspondentes as diversas escalas são coincidentes, não existe efeito das escalas;

$H_0^{(3)}$: 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 Tabela 7.

Tabela 7 : Análise de variância do perfil para o exemplo.

Causas de Variação	**Soma de Quadrados**	**Graus de liberdade**	**Quadrado Médio**	**F**	**Graus de Liberdade conservados**	**F_{tab} (α=0,05)**
Escalas (B)	21,24	2	10,62	6,88	1; 17	4,54
Classes (A)	743,90	3	247,97	70,93	2; 17	3,68
Condições de avaliação (dentro das classes)	59,43	17	3,49			
Escalas x Classes (A*B)	18,96	6	3,16	2,05	3; 17	3,29
Condições de Avaliação X Classes (dentro das classes)	52,47	34	1,54			
TOTAL	896	62				

Os cálculos das seis somas de quadrados seguem abaixo, conforme equações (15; 16; 17; 18; 20), respectivamente:

$$SQ_1 = \frac{1}{21}\left[(305)^2 + (328)^2 + (333)^2\right] - \frac{(966)^2}{21.3} = 21{,}2381$$

$$SQ_2 = \frac{1}{3}\left[\frac{1}{8}(462)^2 + \frac{1}{5}(216)^2 + \frac{1}{4}(168)^2 + \frac{1}{4}(120)^2\right] - \frac{(966)^2}{21.3} = 743{,}9$$

$$SQ_3 = \frac{1}{3}\left[(57)^2 + (60)^2 + \ldots + (31)^2 + (30)^2\right] - \frac{1}{3}\left[\frac{1}{8}(462)^2 + \frac{1}{5}(216)^2 + \frac{1}{4}(168)^2 + \frac{1}{4}(120)^2\right] = 59{,}43$$

$$SQ_4 = \frac{1}{8}\left[(144)^2 + (160)^2 + (158)^2\right] + \frac{1}{5}\left[(69)^2 + (76)^2 + (71)^2\right] + \frac{1}{4}\left[(52)^2 + (56)^2 + (60)^2\right]$$

$$+ \frac{1}{4}\left[(40)^2 + (36)^2 + (44)^2\right] - \frac{(996)^2}{21.3} - 21{,}24 - 743{,}9 = 18{,}9619$$

$$SQ_5 = 896 - 21{,}24 - 743{,}9 - 59{,}43 - 18{,}96 = 52{,}47$$

$$SQ_6 = (19)^2 + (20)^2 + \ldots + (12)^2 - \frac{(966)^2}{21{,}3} = 896$$

Pela análise de perfis na Tabela 7, pode-se concluir que os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos, não existe interação escala x classes sociais. Aceita-se $H_0^{(1)}$ em nível de 0,05 de significância, pois $F_c = 2{,}05 < F_{tab} = 3{,}29$;

Para a segunda hipótese analisada $H_0^{(2)}$ pode-se concluir que os perfis médios de resposta correspondentes às diversas escalas não são coincidentes, isto é, existe efeito das escalas. Rejeita-se $H_0^{(2)}$ em nível de 0,05 de significância, pois $F_c = 6{,}88 > F_{tab} = 4{,}54$;

Para a terceira hipótese $H_0^{(3)}$ pode-se concluir que os perfis médios de resposta correspondentes às diversas escalas não são paralelos ao eixo das abscissas, existe efeito das classes sociais. Rejeita-se $H_0^{(3)}$ em nível de 0,05 de significância, pois $F_c = 70{,}93 > F_{tab} = 3{,}68$.

Estes resultados também podem ser verificados com auxílio do gráfico de perfis de médias, a seguir:

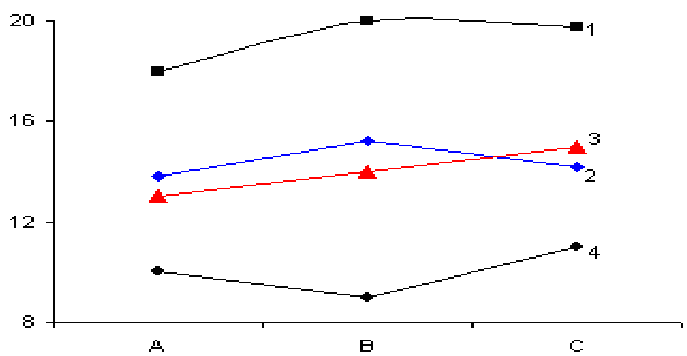

Figura 3: Perfis médios de atitudes maternas conforme classe sócio-econômica.

Na Figura 3, 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 socio-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 terceira não parecem ser diferentes, e a quarta é diferente das demais.

5.1 Aplicação dos Métodos com Saídas do Software *Statistica 7.0*

Utilizando-se o programa *Statistica* para realizar o exemplo anterior, os seguintes passos devem ser efetuados:

(1) Ao iniciar-se o programa, escolhe-se na barra de ferramentas em *Statistics* a opção ANOVA conforme mostra a janela abaixo:

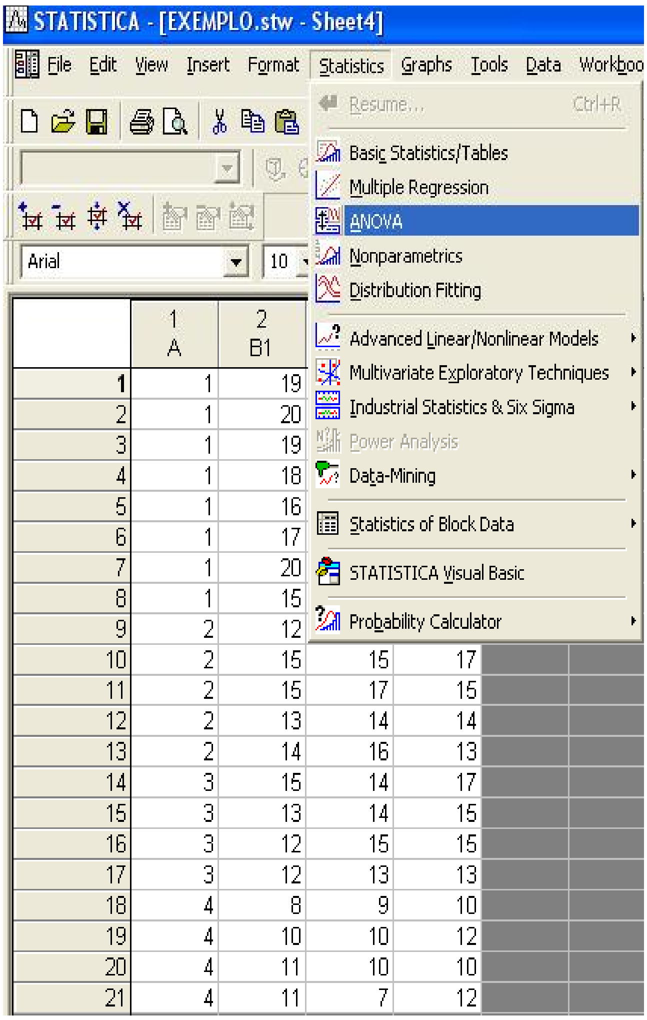

Figura 4 – Caixa de seleção da análise de variância.

(2) Ao clica-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 5:

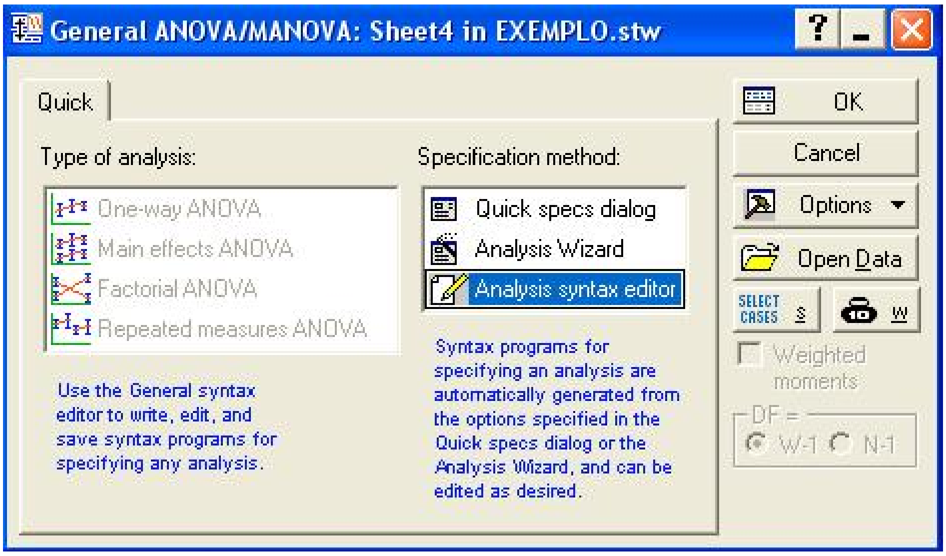

Figura 5 – Caixa de seleção para especificação do método.

(3) A seguir, clica-se no botão *OK* e uma janela como mostrado abaixo (Figura 6) aparecerá:

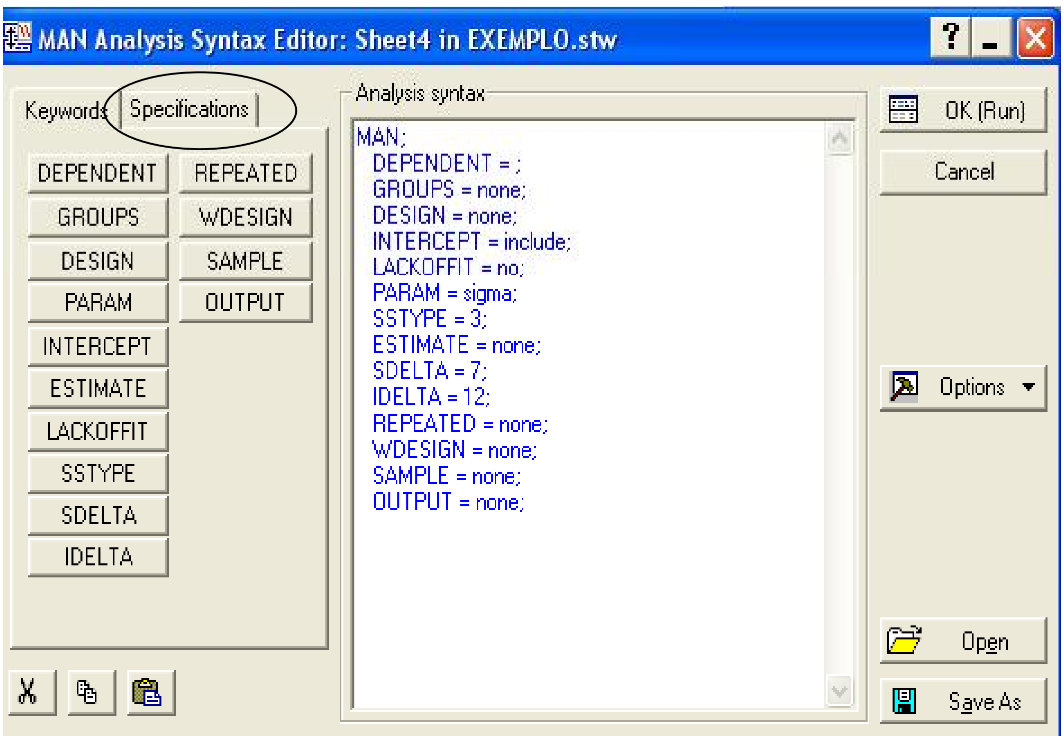

Figura 6 – Caixa de seleção para especificação dos parâmetros na análise de variância multivariada.

(4) Então seleciona-se na Figura 6, *Specifications,* e aparecerá a tela conforme a Figura 8 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 7:

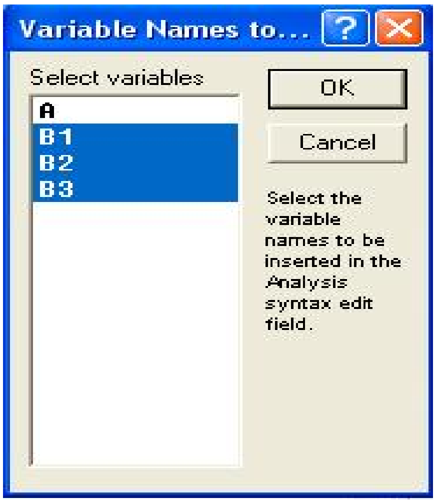

Figura 7 – Caixa de seleção das variáveis.

(5) Após selecionar as variáveis, clica-se em *OK*, e preenche-se o quadro MAN (Figura 8) as escalas (*GROUPS* = A 1 2 3 4), o *DESIGN* (escalas = A), as somas de quadrados (*SSTYPE* = 2) as medidas repetidas (*REPEATD* = B 3) e o (*WDESIGN* = *B*); conforme mostrado a seguir:

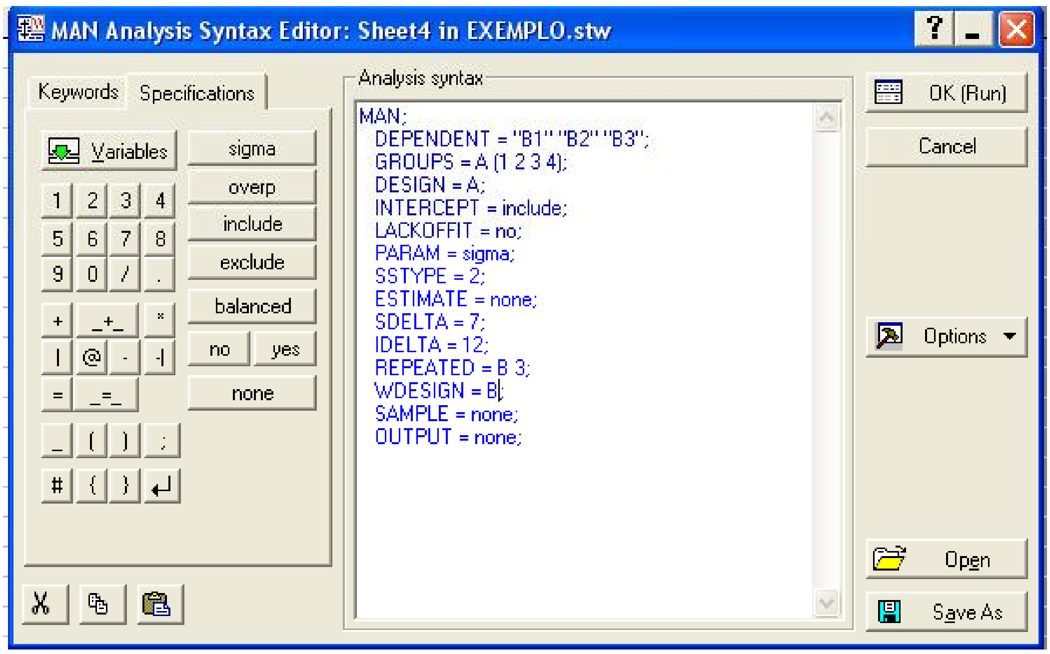

Figura 8 - Caixa de seleção para análise de variância multivariada.

(6) Clica-se em *OK (Run)* e aparecerá a seguinte janela:

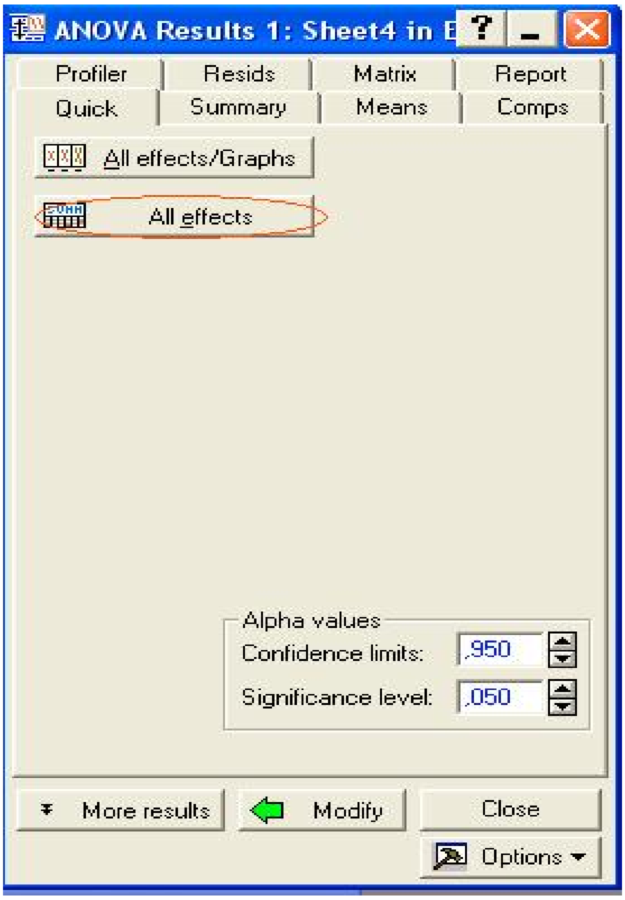

Figura 9 – Caixa de seleção para os efeitos da análise de variância multivariada

(7) Se se clica em *All effects* na Figura 9 irão aparecer os resultados da análise de variância multivariada, mostrados abaixo, que conferem com a análise apresentada na Tabela 7(ver p.59).

	Repeated Measures Analysis of Variance (Sheet4 in EXEMPLO.stw) 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		

Figura 10 – Caixa com os resultados da análise

(8) Se se clica na opção *More results,* mostrada na Figura 9, irá aparecer a seguinte caixa de seleção:

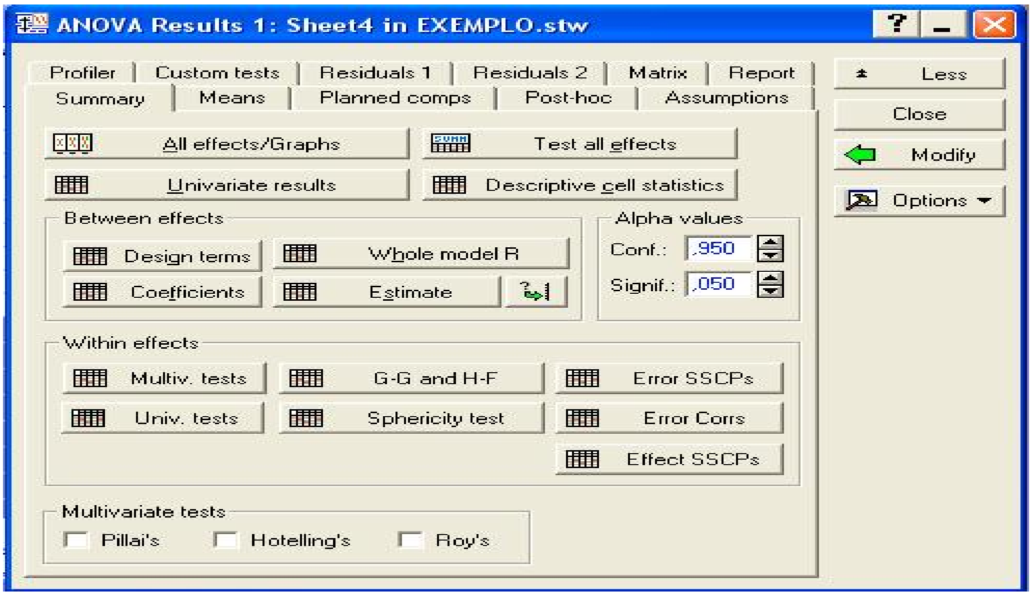

Figura 11.- Caixa de seleção das análises.

(9) Na caixa de seleção, indicada na Figura 11, pode-se selecionar a guia *Summary* onde aparece o botão *Test all effects* e também se conseguirá os resultados do passos (6) e (7). Se se selecionar a guia *Matrix* a seguinte janela aparecerá:

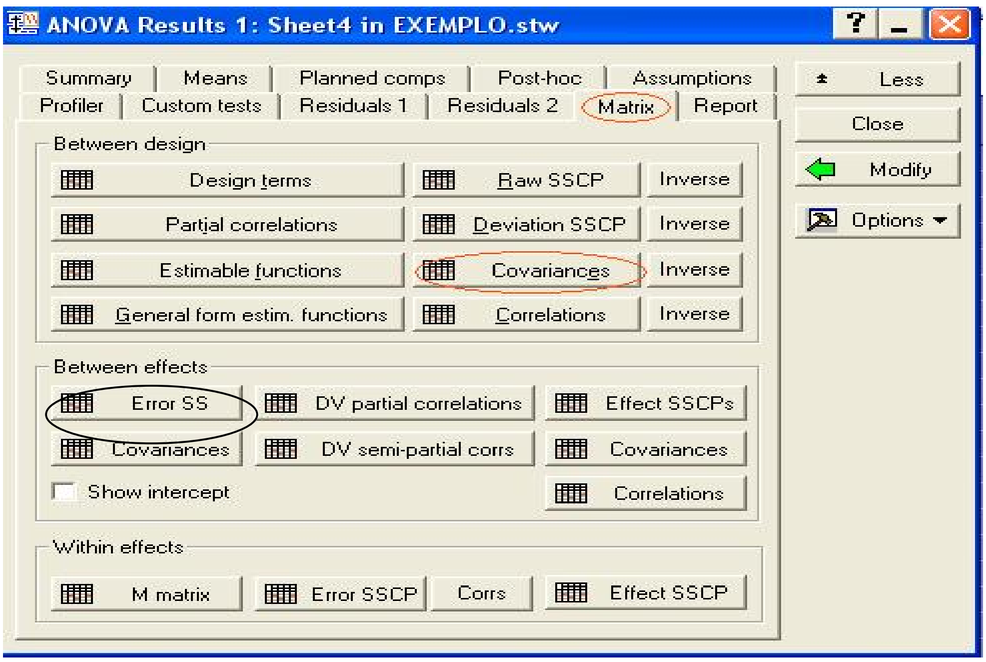

Figura 12 – Caixa de comando para. obtenção das matrizes.

(10) Para se obter a matriz de variância e covariância clica-se, na Figura 12, em *Covariances*, como indicado acima. Após, irá aparecer os seguintes resultados:

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 Intercpt	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

Figura 13 – Caixa de resultados da matriz de variância e covariância.

A matriz em destaque é a matriz de variância e covariância que confere com os resultados encontrados anteriormente.

(11) Para encontrar a matriz W (matriz das somas dos quadrados e produtos cruzados dentro dos grupos) clica-se em *Error SS* no sub-item *Between effects* na Figura 12. Feito isso, Ter-se-á matriz W, que também confere com a matriz encontrada anteriormente(ver p.58). Sendo a seguinte:

Residual Sums of Squares and Cross Proc
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

Figura 14– Caixa de resultados da matriz das somas dos quadrados e produtos cruzados dentro dos grupos

(12) Da mesma forma, pode-se encontrar a matriz B (matriz das somas de quadrados e produtos cruzados entre os grupos). Para isto, na Figura 12 , basta clicar em *Effect SSCP$_S$* também no subitem *Between effects.* Após este procedimento aparecerá a matriz B que confere com a matriz anteriormente encontrada (ver p.58). Cujo resultado é mostrado a seguir:

Effect	Between Effects Sums of Squares and Crc Sigma-restricted parameterization Effective hypothesis decomposition **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

Figura 15 – Caixa de resultados da matriz das somas dos quadrados e produtos cruzados entre os grupos

(13) A matriz T, que é a matriz Total, pode ser encontrada clicando-se, na Figura 12, em *Deviation SSCP* no subitem *Between design*. Aparecendo o seguinte resultado:

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 Intercpt	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

Figura 16 – Caixa de resultados da matriz total.

(14) Na guia *Summary* no sub-item *Withim effects*, você pode encontrar os valores de alguns critérios de teste para análise multivariada de perfis, clicando no botão *Multiv. Tests*. Você pode selecionar mais de um critério, conforme Figura 17:

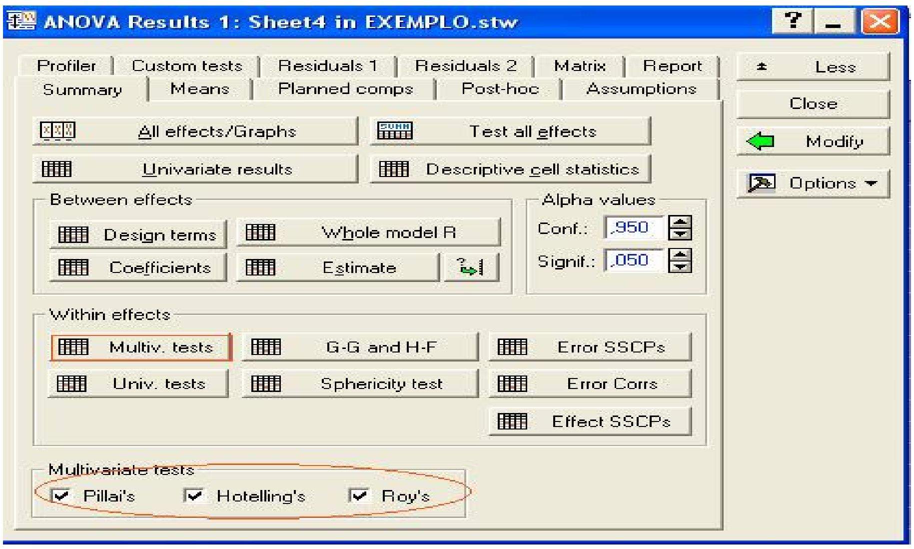

Figura 17 – Caixa de seleção para os testes multivariados.

(15) Ao realizar o passo (14) os resultados dos testes aparecerão, conforme mostra a Figura 18 abaixo:

Effect	Multivariate tests for repeated measure: DV_1 (Sheet4 in EXEI Sigma-restricted parameterization Type II decomposition 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
	Hotellng	0,994855	7,958842	2	16	0,003988
	Roy'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
	Hotellng	0,685343	1,713359	6	30	0,152179
	Roy's	0,508849	2,883477	3	17	0,066180

Figura 18 – Caixa de resultados para os testes multivariados.

Pode-se observar que os resultados destes testes conferem com os encontrados na Tabela 7(ver p.59). 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).

(16) Para realizar as comparações múltiplas seleciona-se a guia *Post-hoc*, como indicado na caixa de seleção seguinte:

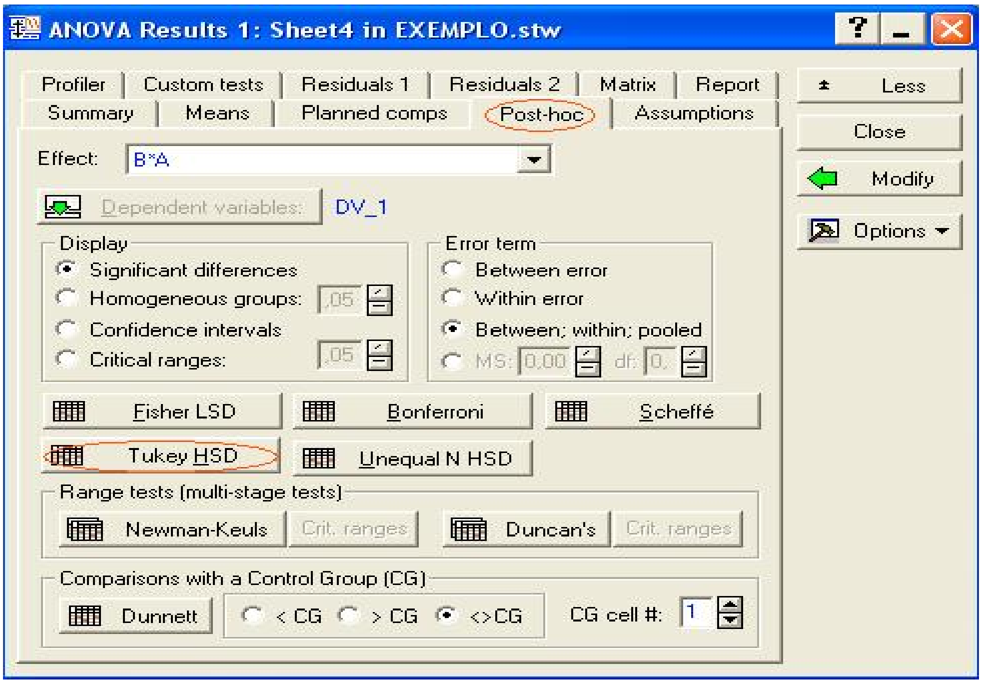

Figura 19 – Caixa de seleção para as comparações múltiplas.

(17) Como pela análise de variância rejeitou-se a hipótese de interação (A*B). Aqui realizou-se 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, mostrada no passo (16), os efeitos A e B na seta ao lado da opção *Effect*, Figura 19. Os resultados destas análises são mostrados, respectivamente, nas Figuras 20 e 21 a seguir

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

Figura 20 – Caixa de resultados para as comparações múltiplas entre as classes sócio-econômicas.

Cell No.	Tukey HSD test; variable DV_1 (Sheet4 Probabilities for Post Hoc Tests Error: Within MS = 1,5431, df = 34,000			
	B	{1} 14,524	{2} 15,619	{3} 15,857
1	B1		0,019441	0,004008
2	B2	0,019441		0,809724
3	B3	0,004008	0,809724	

Figura 21 – Caixa de resultados para as comparações múltiplas entre as escalas de atitudes maternas.

(18) O gráfico de perfis de médias pode-se obter na guia *Means*, Figura 22, clicando-se em *Plot* como o indicado, na caixa de seleção:

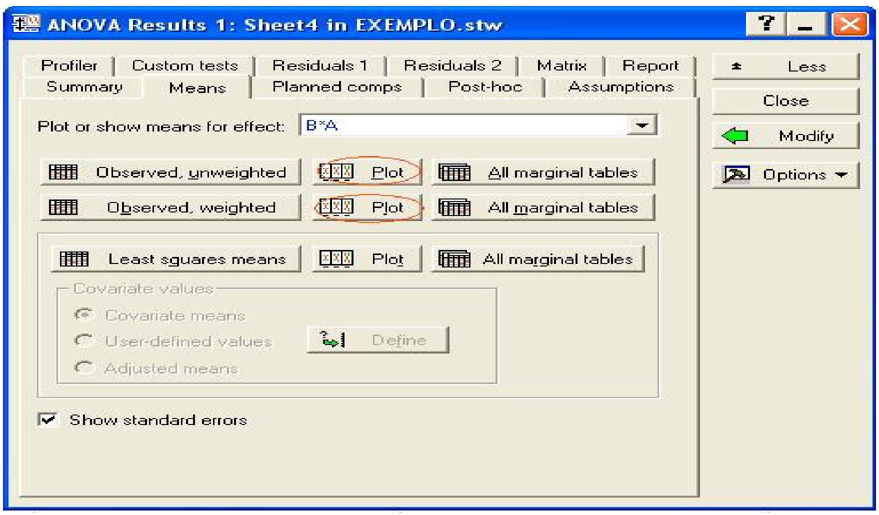

Figura 22 – Caixa de seleção para a plotagem dos gráficos de perfis de médias.

O gráfico abaixo, é o gráfico de perfis de médias gerado após o passo (20)

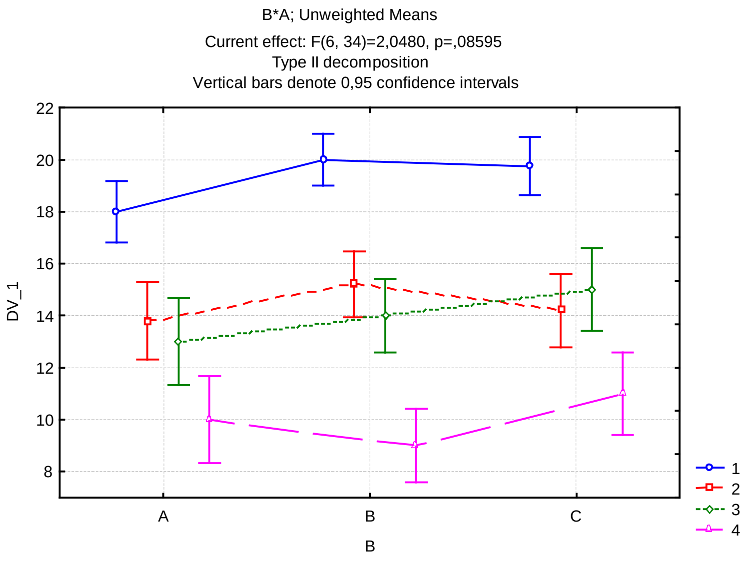

Figura 23 – Gráfico dos perfis médios.

6 CONSIDERAÇÕES FINAIS

Ao final deste trabalho, verificou-se que o estudo das técnicas multivariadas é extremamente útil e importante para a análise do comportamento de um conjunto de várias variáveis e na determinação do papel de cada uma na presença das demais. Observou-se, também, que as técnicas do domínio multivariado que englobam medidas repetidas e dados longitudinais são úteis na avaliação do comportamento dessas várias variáveis ao longo do tempo, ou ainda, em avaliá-las quando submetidas a diferentes condições de avaliação.

Para o desenvolvimento deste trabalho foi realizada uma revisão de literatura que reuniu os principais aspectos sobre as metodologias de análise de variância multivariada e análise multivariada de perfis, que são técnicas multivariadas aplicadas a medidas repetidas e dados longitudinais.

A análise de variância multivariada considera, simultaneamente todas as variáveis de interesse e possibilita analisar as variações globais ou individuais dos dados e as possíveis diferenças entre as médias dos grupos. Como a análise de variância univariada, algumas pressuposições devem ser verificadas. Estas pressuposições também foram abordadas neste trabalho e são as seguintes: as observações devem seguir uma distribuição normal multivariada, e as amostras devem pertencer a grupos populacionais com matrizes de variância e covariância que não difiram significativamente.

A análise multivariada de perfis é uma das técnicas estatísticas aplicada a observações provenientes de experimentos com medidas repetidas. Esta metodologia é uma derivação da análise de variância multivariada, tendo como objetivo comparar os diversos componentes de perfis médios de resposta. É importante lembrar que, para a sua aplicação, o número total de observações menos o número de tratamentos deve ser maior ou igual ao número de dados longitudinais (ou condições de avaliação) menos um.

Para melhor entendimento da técnica estudada, apresentou-se um exemplo hipotético para a aplicação destas metodologias. Em um primeiro momento, este exemplo foi realizado, passo-a-passo, conforme a forma matricial, e, posteriormente, foi executado no software *Statistica7.0*, com as devidas instruções operacionais.

Baseado no trabalho, conclui-se que a utilização destas técnicas proporciona, em determinadas situações, resultados globais sobre as observações em estudo. Desta forma, mais uma opção de análise está disponível, o que talvez seja mais adequado do que estudá-las separadamente. Além disso, observa-se que o software estatístico é uma ferramenta útil neste tipo de análise, pois minimiza o tempo gasto nos cálculos demasiadamente trabalhosos e demorados que este tipo de estudo requer.

Como sugestão para trabalhos futuros, recomenda-se o estudo de curvas de crescimento que são modelos aplicados a dados longitudinais.

7 REFERÊNCIAS

ANDRADE, D. F. de; SINGER, J. da M. **Análise de dados longitudinais.** Campinas: VII Simpósio Nacional de Probabilidade e Estatística, 1986.

CASTRO, S. M. de J. **A metodologia de análise de dados longitudinais.** 1997. 119 f. Monografia (Bacharelado em Estatística) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1997.

FERREIRA, D. F.; **Análise Multivariada.** Lavras, 1996. 389p.

HAIR, J. F. et al. **Multivariate Analysis**. 5th ed. New Jersey: Prentice – Hall, 1998. 730p.

JOHNSON, R. A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis.** 4th ed. New Jersey: Prentice – Hall, 1998. 816p.

LIMA, C. G. de. **Análise de dados longitudinais de experimentos em blocos casualizados.** 1996. 126f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “ Luiz de Queiroz” da Universidade de São Paulo, Piracicaba, 1996.

MARDIA, K. V.; KENT, J. T.; BIBBY, J.M. **Multivariate Analysis.** San Diego: Acadmic Press, 1994.

MORRISON, D. F. **Multivariate Statistical Methods**. 3rd. ed. McGraw-Hill International Editions (Statistics Series), 1990. 495 p.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa: Sílabo, 1997. 343p.

RIBOLDI, J. **Análise de medidas repetidas:** perspectivas e tendências. Botucatu: 8º SEAGRO e 44ª RBRAS, 1999.

RIBOLDI, J. et al. **Medidas repetidas**. 2003. 13f. Notas de aula.

SINGER, J. da M. **Análise de curvas de crescimento.** 1977. 113f. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 1977.

SINGER, J. da M.; ROCHA, F. M. M. da.; NOBRE, J. S. **Análise de medidas repetidas.** Maringá: IV Jornada Regional de Estatística, 2004. 93p.